Miguel Ángel Gómez Villegas

INFERENCIA ESTADÍSTICA



DIAZ DE SANTOS

INFERENCIA ESTADÍSTICA

Miguel Ángel Gómez Villegas
Catedrático de Estadística e Investigación Operativa
Departamento de Estadística e Investigación Operativa
Facultad de Ciencias Matemáticas
Universidad Complutense de Madrid

INFERENCIA ESTADÍSTICA

DIAZ DE SANTOS

© Miguel Ángel Gómez Villegas, 2005

Reservados todos los derechos.

«No está permitida la reproducción total o parcial de este libro, ni su tratamiento informático, ni la transmisión de ninguna forma o por cualquier medio, ya sea electrónico, mecánico por fotocopia, por registro u otros métodos, sin el permiso previo y por escrito de los titulares del Copyright».

Ediciones Díaz de Santos

Internet: http://www.diazdesantos.es
E-Mail: ediciones@diazdesantos.es

ISBN: 84-0000-000-X
Depósito legal: M. 49.000-2004

Diseño de cubierta:
Fotocomposición: Fer, S. A.
Impresión: Edigrafos, S. A.
Impreso en España

A la memoria de mi padre Joaquín,
a mi madre Carmen
y a la memoria de mi tío Angel,
mi primer maestro.

PRÓLOGO

Este libro contiene el material necesario para un curso de Inferencia Estadística de un año de duración. Aunque todavía en algunos planes de estudio, tanto de licenciaturas como de ingenierías, existe un curso sobre Cálculo de Probabilidades y Estadística, cada vez es más frecuente el abordar el estudio de la Inferencia Estadística en un curso propio separado del Cálculo de Probabilidades.

En esta línea este texto está pensado para estudiantes de Estadística, en Matemáticas, Biología, Economía, Informática, Ingenierías, Medicina, Psicología, Sociología, y en definitiva para los que estén interesados en modelizar la incertidumbre mediante la probabilidad para extraer conclusiones de esta modelización. El libro recoge lo que en mi opinión es la naturaleza de la Estadística moderna: un equilibrio entre el análisis de datos -con especial atención a las representaciones gráficas-, y los modelos matemáticos, ambas cosas realizadas con el creciente concurso del ordenador.

Los prerrequisitos incluyen nociones de probabilidad básica, además de los resultados usuales de cálculo, álgebra matricial y análisis real.

Los teoremas y demostraciones se incluyen cuando su enunciado y su demostración son necesarios para una correcta comprensión del texto y siempre que se mantengan dentro del nivel de complejidad y abstracción que se ha querido dar al libro. Al final de cada capítulo se incluye una colección de aproximadamente 25 problemas, cuya solución detallada ha sido realizada en colaboración con la profesora Paloma Maín Yaque.

En cuanto al contenido del texto, los dos primeros capítulos constituyen la introducción a la "Teoría de la Inferencia Estadística", estudian los estadísticos muestrales, los diagramas de tallos y hojas y de cajas, las distribuciones asociadas al muestreo con la distribución normal, los estadísticos ordenados y los principios de reducción de datos. Se recomienda especialmente la atenta lectura de los principios de verosimilitud, suficiencia y condicionalidad; el significado de los mismos, junto al teorema de Birnbaum, son fundamentales para una correcta interpretación de los distintos procedimientos que integran la Inferencia Estadística. También se ha incluido el concepto de intercambiabilidad como una manera más razonable que la independencia para recoger la información suministrada por la muestra. Este concepto es de capital importancia pues justifica la aproximación bayesiana

a la inferencia.

Los Capítulos 3, 4 y 5, se dedican a la estimación por punto, la estimación por regiones y el contraste de hipótesis, que constituyen el núcleo central de la Inferencia Estadística. Un aspecto importante en la manera de desarrollar estos capítulos, es la diferenciación entre los métodos de obtención de las distintas técnicas y los métodos de evaluación de las mismas, que son tratados a continuación. Este tratamiento permite que los tres métodos usados para eliminar la incertidumbre presenten una estructura uniforme.

Los tres Capítulos finales contienen aplicaciones específicas, así el Capítulo 6 es una introducción a la Teoría de la Decisión Estadística, el Capítulo 7 aborda el Análisis de la Varianza y la Regresión Lineal y el Capítulo 8 la Inferencia no Paramétrica.

Tres aspectos hacen singular, en mi opinión, a este libro. El primero es que los métodos bayesianos y los clásicos son desarrollados a la vez sin ningún tipo de menoscabo de unos frente a otros, aunque el autor es partidario de la aproximación bayesiana. Así a continuación de los estimadores por punto clásicos se estudian los estimadores por punto bayesianos. Después de los intervalos de confianza los intervalos creíbles y siguiendo a los contrastes de hipótesis mediante la teoría de Neyman y Pearson, los contrastes bayesianos y los factores Bayes. Además, tras cada método, según corresponda, se estudia su comportamiento frente a la distribución en el muestreo, recogido en el criterio del error cuadrático medio, y frente a la distribución inicial, recogido en la pérdida final esperada. El estudio conjunto de estas dos aproximaciones hace que se haya incluido en el Capítulo 7 el análisis bayesiano de la varianza, que no está en otros textos de nivel parecido, ya que o sólo utilizan la aproximación clásica o sólo la aproximación bayesiana. Así mismo, en el Capítulo 6, sobre la "Teoría de la Decisión Estadística", se incluye la paradoja de Stein, presentada de la manera más elemental posible, para poner de manifiesto las anomalías que presentan los métodos clásicos cuando la dimensión del parámetro es mayor o igual que tres.

El segundo aspecto es que al final de cada Capítulo se incluye una aproximación histórica de la evolución de las ideas que han sido desarrolladas en el mismo, incluyendo unas pinceladas biográficas de quienes más han contribuido al desarrollo de las mismas.

El tercer aspecto es la publicación de la solución detallada de los ejercicios de final de capítulo como un eficaz complemento de las ideas desarrolladas. Con frecuencia, el alumno echa de menos una colección de ejercicios resueltos que le permitan aclarar y comprobar si ha asimilado correctamente los conocimientos expuestos.

No se discute en el texto ningún paquete de programas, pero en la resolución de los problemas se ha utilizado S, SPSS y Statgraphics; pudiendo utilizarse otros paquetes también. El libro ha sido empleado para impartir durante todo un curso 4 horas de clase y una de laboratorio por semana. La utilización de un paquete estadístico permite tratar problemas más reales y

ayuda a aclarar aspectos que de otra manera son más difíciles de comprender.

Quiero expresar mi agradecimiento a todas las personas a las que debo mis conocimientos en el campo de la Estadística, desde aquella primera versión traducida al castellano del libro de Cramér en el que por primera vez leí sobre estas ideas y aquellas primeras clases que, impartidas por el profesor Sixto Ríos, me pusieron en contacto con la Estadística en la Licenciatura de Matemáticas de la Universidad Complutense. A todos los compañeros, tanto los anteriores como los actuales, del Departamento de Estadística e Investigación Operativa de la Complutense de Madrid, muchas gracias, y en particular a los profesores Jesús Artalejo, Antonio Gómez, Eusebio Gómez, Paloma Maín, Javier Montero, Luis Sanz y Teófilo Valdés, que me han sugerido cambios y aclaraciones. Finalmente quiero dar las gracias a Pedro Ortega por su trabajo de verter el manuscrito inicial al Latex.

Madrid, enero de 2005.

Índice general

PRÓLOGO **IX**

1. INTRODUCCIÓN A LA ESTADÍSTICA MATEMÁTICA Y ESTADÍSTICOS MUESTRALES **1**

1.1. Introducción 1
1.2. Estadísticos muestrales 4
1.3. Estadística descriptiva 9
1.3.1. Medidas de centralización 11
1.3.2. Medidas de dispersión 12
1.3.3. Diagramas de tallos y hojas y de caja 14
1.4. Propiedades asintóticas de los momentos muestrales 16
1.5. Distribuciones en el muestreo asociadas a la distribución normal 21
1.5.1. Distribución de la media muestral 21
1.5.2. Distribución del momento muestral respecto al origen de orden dos 22
1.5.3. Distribución del cociente entre la media muestral y el momento respecto al origen de orden dos 25
1.5.4. Distribución del cociente de momentos de orden dos respecto al origen. 29
1.6. Estadísticos ordenados 33
1.6.1. Distribución marginal del estadístico muestral de orden k 33
1.6.2. Distribución conjunta de varios estadísticos ordenados 35
1.6.3. Recubrimientos 37
1.7. Variables aleatorias intercambiables 39
1.8. Aproximación histórica 43
1.9. Ejercicios 44
1.10. Solución a los ejercicios 48

2. REDUCCIÓN DE DATOS: ESTADÍSTICOS SUFICIENTES, ANCILARIOS, COMPLETOS E INVARIANTES **59**

2.1. Introducción 59
2.2. Estadísticos suficientes y minimal suficientes 59

2.3. Estadísticos ancilarios y completos 65
2.4. Estadísticos invariantes 70
2.5. Principios de reducción de datos 72
2.5.1. Principio de verosimilitud 72
2.5.2. Principio de suficiencia 74
2.5.3. Principio de condicionalidad 74
2.5.4. Teorema de Birnbaum 75
2.6. Aproximación histórica 78
2.7. Ejercicios 79
2.8. Solución a los ejercicios 82

3. ESTIMACIÓN PUNTUAL PARAMÉTRICA 95
3.1. Introducción 95
3.2. Propiedades de los estimadores 96
3.2.1. Estimadores centrados 96
3.2.2. Estimadores consistentes 98
3.2.3. Estimadores bayesianos 100
3.3. Criterios de comparación de estimadores 104
3.3.1. Error cuadrático medio 105
3.3.2. Pérdida final esperada 107
3.4. Estimadores centrados de mínima varianza 109
3.5. Cota para la varianza de un estimador 118
3.6. Métodos de construcción de estimadores 127
3.6.1. Método de los momentos 127
3.6.2. Método de la máxima verosimilitud 129
3.6.3. Propiedades asintóticas de los estimadores de máxima verosimilitud 135
3.6.4. Método bayesiano 138
3.6.5. Propiedades asintóticas de los estimadores bayesianos 139
3.7. Aproximación histórica 140
3.8. Ejercicios 142
3.9. Solución a los ejercicios 147

4. ESTIMACIÓN POR REGIONES DE CONFIANZA 169
4.1. Introducción 169
4.2. Métodos de obtención de intervalos de confianza 171
4.3. Intervalos de confianza asociados a la distribución normal . . 177
4.4. Intervalos de confianza para muestras grandes 182
4.5. Regiones de confianza bayesianas 183
4.6. Aproximación histórica 186
4.7. Ejercicios 188
4.8. Solución a los ejercicios 191

5. CONTRASTES DE HIPÓTESIS 207
5.1. Introducción 207
5.2. Métodos de construcción de contrastes de hipótesis 211
5.2.1. Contrastes de la razón de verosimilitudes 211
5.2.2. Contrastes de la razón de verosimilitudes en el caso normal 213
5.2.3. Distribución asintótica de la razón de verosimilitudes . 221
5.2.4. Contrastes bayesianos 225
5.2.5. Contrastes invariantes 228
5.3. Criterios de comparación de contrastes 230
5.4. Contrastes de hipótesis unilaterales y bilaterales 236
5.5. Dualidad entre contrastes de hipótesis y regiones de confianza 241
5.6. Aproximación histórica 245
5.7. Elercicios 247
5.8. Solución a los ejercicios 251

6. TEORÍA DE LA DECISIÓN 271
6.1. Introducción 271
6.2. Métodos de inferencia en la teoría de la decisión 274
6.2.1. Estimación por punto 274
6.2.2. Estimación por regiones de confianza 274
6.2.3. Tests de hipótesis 275
6.3. Obtención de reglas bayes y de reglas minimax 276
6.4. Reglas admisibles y clases completas 280
6.5. Paradoja de Stein 283
6.6. Aproximación histórica 290
6.7. Ejercicios 291
6.8. Solución a los ejercicios 296

7. ANÁLISIS DE LA VARIANZA Y REGRESIÓN LINEAL 315
7.1. Introducción 315
7.2. Análisis de la varianza con un sólo factor 315
7.3. Análisis de la varianza con dos factores 325
7.4. Regresión lineal 329
7.4.1. Regresión lineal simple 330
7.4.2. Regresión lineal múltiple: aproximación matricial . . . 337
7.5. Aproximación bayesiana del análisis de la varianza 349
7.6. Aproximación histórica 355
7.7. Ejercicios 356
7.8. Solución a los ejercicios 363

8. ESTADÍSTICA NO PARAMÉTRICA **389**
8.1. Introducción 389
8.2. Ajustes relativos a la distribución multinomial 390
8.2.1. Bondad del ajuste 390
8.2.2. Homogeneidad entre varias muestras 396
8.2.3. Tablas de contingencia 398
8.3. Aplicaciones de los estadísticos ordenados 401
8.3.1. Intervalos de tolerancia 401
8.3.2. Intervalos de confianza para los cuantiles 404
8.3.3. Contrastes de hipótesis sobre los cuantiles 405
8.4. Problemas no paramétricos relativos a una muestra 407
8.4.1. Contrastes de Kolmogorov-Smirnov para una muestra 407
8.4.2. Contrastes de localización 413
8.4.3. Contrastes de normalidad 421
8.4.4. Valores discordantes 424
8.4.5. Estimación no paramétrica de densidades 425
8.5. Problemas no paramétricos relativos a dos muestras 429
8.5.1. Contrastes de Kolmogorov-Smirnov para dos muestras 429
8.5.2. Test de la mediana 436
8.5.3. Test de Mann-Whitney-Wilcoxon 438
8.5.4. Test de Kruskal-Wallis 442
8.6. Contrastes no paramétricos de independencia 444
8.6.1. Test de Kendall 444
8.6.2. Test del coeficiente de correlación de los rangos de Spearman 449
8.7. Aproximación histórica 452
8.8. Ejercicios 453
8.9. Solución a los ejercicios 458

DISTRIBUCIONES USUALES **477**

TABLAS **481**

REFERENCIAS BIBLIOGRÁFICAS **501**

Capítulo 1

INTRODUCCIÓN A LA ESTADÍSTICA MATEMÁTICA Y ESTADÍSTICOS MUESTRALES

1.1. INTRODUCCIÓN

Dentro del Cálculo de Probabilidades se pretende calcular la probabilidad de ciertos sucesos a partir de probabilidades calculadas de otros sucesos más simples. Así, si se supone que un determinado experimento aleatorio se puede modelizar mediante una variable aleatoria de Bernoulli de parámetro θ conocido, la probabilidad de que el suceso éxito se presente en dos repeticiones consecutivas del experimento es θ^2.

En esta línea, se dice que el Cálculo de Probabilidades utiliza un razonamiento deductivo. Así, tras modelizar la incertidumbre mediante la probabilidad, es posible utilizar dicho razonamiento para llegar a obtener probabilidades de sucesos más complicados a partir de las probabilidades de sucesos más simples.

En cuanto a la Inferencia Estadística, Cramér (1945) dice:

> *El objetivo fundamental de la Teoría Estadística consiste en investigar la posibilidad de extraer de los datos estadísticos inferencias válidas.*

Se admite que los datos estadísticos deben de poseer una característica

que tiene un significado preciso, que es la aleatoriedad. Fisher (1958) dice que la ciencia estadística es esencialmente una rama de la matemática aplicada:

> *La Estadística puede ser mirada como el estudio de poblaciones, como el estudio de la variación y como el estudio de métodos de reducción de datos.*

y añade Fisher que el significado original de la palabra Estadística sugiere que ésta surge con el objetivo de estudiar colectivos humanos viviendo en unión política, aunque más tarde ha ido ampliando su campo de aplicación primero a la Biología y, después, a la Economía y a todas las restantes ciencias experimentales.

Lindley (1965) ve la Estadística como el análisis de cómo los grados de opinión son alterados por los datos.

Kendall y Stuart (1969) afirman:

> *La Estadística es la rama del método científico que se enfrenta con datos obtenidos contando o midiendo elementos de poblaciones asociadas a fenómenos naturales.*

Breiman (1973) dice:

> *La diferencia entre un modelo probabilístico y un modelo estadístico es que en el primero la probabilidad está totalmente especificada, mientras que en el segundo la distribución de probabilidad es desconocida. Todo lo que se conoce son algunas características de esta distribución de probabilidad.*

Cox (1973) establece:

> *La Estadística pretende elegir un elemento de entre una familia $\mathcal{P}$ de modelos de probabilidad, que en la mayor parte de los casos es conocida excepto por un número limitado de parámetros desconocidos, para responder a ¿son los datos consistentes con la familia $\mathcal{P}$? Suponiendo que los datos han sido obtenidos de uno de los modelos de $\mathcal{P}$, ¿qué puede concluirse sobre los valores de los parámetros desconocidos o acerca de los valores de observaciones extraídas del mismo modelo?*

Por último, según Lehmann (1978):

> *La Estadística trata de la recogida de datos así como de su análisis e interpretación.*

Basados en las definiciones anteriores se puede decir que la Inferencia Estadística pretende, dados unos datos sujetos a incertidumbre, obtener el conocimiento de los parámetros, en un sentido amplio, del modelo de Cálculo de Probabilidades que aceptablemente se ajuste a dichos datos. En este sentido, la Inferencia Estadística utiliza un razonamiento de tipo inductivo.

Un proceso de fabricación, a lo largo del tiempo, puede producir a veces algún elemento defectuoso. La probabilidad θ de que un elemento cualquiera sea defectuoso es desconocida. De toda la producción se extraen al azar n elementos y con estos datos se trata de conjeturar el valor de θ.

El ejemplo explicita los elementos con los que se va a trabajar: a saber, una población, una muestra: los n elementos observados; y una distribución de probabilidad. La población se puede considerar que es una variable aleatoria con distribución de Bernoulli de parámetro θ, donde θ es la probabilidad de producir un elemento defectuoso, los n elementos son la repetición con independencia, de n variables aleatorias de Bernoulli y lo que se pretende, basados en la distibución de probabilidad de la muestra $\theta^{\sum_1^n x_i}(1-\theta)^{n-\sum_1^n x_i}$, es conocer el valor de θ. En la misma línea se puede considerar un ejemplo un poco más complicado. Se suponen almacenadas gran cantidad de semillas en condiciones constantes de temperatura y humedad. Las semillas se van deteriorando a lo largo del tiempo, de tal forma que existirá una función $r(t)$ que indique la proporción de semillas que en el instante t continúan siendo válidas. Si en sucesivos instantes de tiempo $t_1, ..., t_s$ se toman cada vez n semillas y se observa que continúan válidas $r_1, ..., r_s$; se trataría de determinar la función $r(t)$ al menos en los instantes t_i con $i = 1, ..., s$. En este caso la distribución de probabilidad de la muestra viene dada por $\prod_{i=1}^{s} \binom{n}{r_i}(r(t_i))^{r_i}(1-r(t_i))^{n-r_i}$. En esta situación no existe el parámetro θ o más bien el parámetro, en sentido amplio, es la función $r(t)$. Por último, un experimentador realiza n observaciones $x_1, ..., x_n$ con vistas a determinar una cantidad θ, que él desconoce, pero supone que estas observaciones pueden ponerse en la forma $x_i = \theta + \varepsilon_i \quad i = 1, ..., n$. Donde ε_i son los errores contenidos en cada observación que se suponen independientes y distribuidos mediante una distribución $N(0, \sigma)$ normal de media cero y de desviación típica σ desconocida. La distribución de probabilidad de la muestra viene dada a partir del producto de n variables $N(0, \sigma)$. En este caso se desea determinar el valor del parámetro desconocido θ. Todas estas situaciones pueden resumirse en que se tiene una v.a. X y sobre la σ-álgebra de Borel de $\mathbb{R}$, está definida una familia de distribuciones de probabilidad $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$. A lo largo de lo que sigue se irá concretando el problema en distintas direcciones.

Así, se hablará de la *estimación por punto* cuando sólo se esté interesado en la elección de un determinado valor para θ, se hablará de la *estimación por intervalo* cuando se quiera obener un subconjunto de Θ al que pertenezca la distribución con la que se está trabajando, y se hablará del *contraste*

de hipótesis cuando se pretenda únicamente elegir entre dos subfamilias de distribuciones de probabilidad $\mathcal{P}$ y $\mathcal{P}'$. A continuación se estudiará la *Teoría de la Decisión* como un modelo más general que unifica las aproximaciones anteriores, para acabar trasladando estos problemas al caso no paramétrico.

1.2. ESTADÍSTICOS MUESTRALES

El modelo matemático de la Inferencia Estadística viene dado por una variable aleatoria X a la que se llama *universo* o *población* cuya función de distribución F viene determinada por algún elemento de una familia $\mathcal{P}$ de distribuciones de probabilidad. En los ejemplos dados en la sección 1.1 F era la distribución de *Bernoulli* de parámetro θ, o la normal de media θ y varianza desconocida, o una $Binomial(n, \pi(t))$ con $\pi(t)$ variable. El problema es determinar el elemento de la familia de distribuciones de probabilidad para X. Para esto de dispone de cierta información que se concre-, mediante la observación de una muestra.

Definición 1.2.1 (muestra aleatoria simple)
Dada una población X se llama muestra aleatoria simple de tamaño n a la repetición de $X_1, ..., X_n$ variables aleatorias independientes con distribución igual a la de X. Es decir, la función de distribución de la muestra $(x_1, ..., x_n)$ es

$$F(x_1, ..., x_n) = \prod_{i=1}^{n} F(x_i)$$

donde $F(x)$ es la función de distribución de la población X que como se ha dicho sigue dada por un elemento de $\mathcal{P}$.
■

Cuando la población X tiene un número finito de elementos, $x_1, ..., x_N$ se dispone, básicamente de dos procedimientos para obtener una muestra de tamaño n, $n \leq N$. En el primero se puede suponer que cada elemento tiene la misma probabilidad de ser escogido y por tanto ésta es $\frac{1}{N}$, es decir sería equivalente a suponer que los N elementos están metidos en una urna y que se extrae sucesivamente con reposición cada elemento de la urna con probabilidad $\frac{1}{N}$, por tanto

$$P\{X_2 = x_2 | X_1 = x_1\} = \frac{1}{N}$$

siendo x_1 y x_2 cualesquiera elementos pertenecientes a la colección $\{x_1, ..., x_N\}$.

En el símil de la urna esto equivale a ir sacando valores *con reeemplazamiento*. Observamos por tanto que se cumple

$$P\{X_1 = x_1, ..., X_n = x_n\} = \frac{1}{N^n}$$

y que esta manera de muestrear está de acuerdo con la Definición 1.2.1. No obstante, utilizar este procedimiento con el fin de conocer la población $\{x_1, ..., x_N\}$ no parece lo más razonable, ya que se podría observar varias veces el mismo elemento. El otro procedimiento posible consiste en suponer que los elementos se extraen *sin reemplazamiento*. Así el primer elemento es escogido con probabilidad $\frac{1}{N}$, y para la obtención del segundo todos los elementos tienen la misma probabilidad pero el elemento a escoger ha de ser diferente del primero; en el símil de la urna sería equivalente a dejar fuera el primer elemento observado y extraer cualquiera de los restantes con equiprobabilidad, con lo que

$$P\{X_2 = x_2 | X_1 = x_1\} = \begin{cases} 0 & \text{si} \quad x_2 = x_1 \\ \frac{1}{N-1} & \text{si} \quad x_2 \neq x_1 \end{cases}$$

En el caso de obtención de la muestra de una población finita *sin reemplazamiento* no se satisfacen las condiciones de la Definición 1.2.1. Es curioso entonces que el procedimiento de obtener una m.a.s. de una población sea el no intuitivo cuando la población es finita. También es importante el observar que aunque las n variables en el muestreo de una población finita sin reemplazamiento no son independientes, sin embargo las variables están idénticamente distribuidas con distribución marginal $P\{X_i = x\} = \frac{1}{N}$ para cada $x \in \{x_1, ..., x_N\}$ y cada $i = 1, ..., n$. En efecto, basta observar que $P\{X_1 = x\} = \frac{1}{N}$ y que por el teorema de la probabilidad total

$$P\{X_2 = x\} = \sum_{i=1}^{N} P\{X_1 = x_i\} P\{X_2 = x | X_1 = x_i\}$$

pero $P\{X_2 = x | X_1 = x_i\} = 0$ si $x = x_i$ y en otro caso $P\{X_2 = x | X_1 = x_i\} = \frac{1}{(N-1)}$ ya que todos los elementos disponibles se escogen con la misma probabilidad, luego

$$P\{X_2 = x\} = \sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{1}{N} \frac{1}{N-1} = \frac{1}{N}.$$

El mismo razonamiento puede aplicarse de forma sucesiva hasta alcanzar n. Si N es suficientemente grande con respecto a n el muestreo *sin reemplazamiento* puede aproximarse por el muestreo *con reemplazamiento* en el sentido de que las probabilidades calculadas tanto en un tipo de muestreo como en el otro, son prácticamente iguales.

Ejemplo 1.2.1
Dada una población finita de $N = 500$ elementos $\{1, 2, ..., 500\}$ de la que se extraen muestras de tamaño $n = 5$, la probabilidad de que todos los valores

muestrales sean mayores que 100 cuando se extraen *con reemplazamiento* es

$$\begin{aligned} P\{X_1 > 100, ..., X_5 > 100\} &= \prod_{i=1}^{5} P\{X_i > 100\} \\ &= \left(\frac{400}{500}\right)^s = 0'32768. \end{aligned}$$

Cuando se extraen *sin reemplazamiento* la probabilidad es

$$P\{X_1 > 100, ..., X_5 > 100\} = \frac{\binom{400}{5}\binom{100}{0}}{\binom{500}{5}} = 0'326035,$$

lo que parece una aproximación aceptable.
■

Existen otros tipos de muestreo, que en determinadas situaciones prácticas pueden parecer más interesantes, de hecho existen libros enteros dedicados al estudio de los distintos tipos de muestreo. A lo largo del resto del libro, se usará la Definición 1.2.1 como definición de muestra aleatoria correspondiente a una población. Frecuentemente no se trabaja con toda la muestra, tanto por resumir la información como para mimetizar ciertos elementos de la población, por lo que se introduce la siguiente definición.

Definición 1.2.2 (estadístico muestral)
Dada $(X_1, ..., X_n)$ una muestra aleatoria simple de la población X, a toda $f : \mathbb{R}^n \to \mathbb{R}^k$ función conocida medible, se la llamará estadístico muestral.

Ejemplo 1.2.2
$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ y $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ son estadísticos muestrales. El primero es, el conocido como la media muestral y al segundo se le denomina la cuasivarianza muestral.
■

A continuación se introducen estadísticos muestrales particulares.

Definición 1.2.3 (momentos centrales)
Se llama *momento muestral respecto al origen* de orden k a

$$a_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \qquad k > 0.$$

El momento de orden uno, es la media muestral, ya citado en el Ejemplo 1.2.2, que se acostumbra a notar por $\overline{X}$. Se llama *momento muestral respecto al centro* de orden k a

$$b_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k \qquad k > 0.$$

El momento respecto al centro más utilizado es b_2, la varianza muestral.
■

Es fácil ver que

$$b_k = a_k - \binom{k}{1}\overline{x}a_{k-1} + ... + (-1)^k\overline{x}^k.$$

Por consiguiente los momentos muestrales respecto al centro, pueden calcularse en función de los momentos muestrales respecto al origen. En particular la varianza muestral puede ponerse como

$$b_2 = a_2 - \overline{X}^2.$$

Intuitivamente los momentos respecto al origen son medidas de centralización, indican alrededor de qué valores se concentran las observaciones, mientras que los momentos respecto al centro son medidas de dispersión, valores altos indican una mayor dispersión de las observaciones.

Proposición 1.2.1
Dada una m.a.s. de tamaño n de una población X con momento de segundo orden finito, se tiene que $E[\overline{X}] = \mu$ y $V[\overline{X}] = \frac{\sigma^2}{n}$; siendo $\mu = E[X]$ y $\sigma^2 = V[X]$.

Demostración:

$$E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu$$

$$\begin{aligned} V[\overline{X}] &= E[(\overline{X}-\mu)^2] = E\left[\left(\tfrac{1}{n}\sum_{i=1}^{n}(X_i-\mu)\right)^2\right] \\ &= \tfrac{1}{n^2}E\left[\sum_{i=1}^{n}(X_i-\mu)^2 + 2\sum_{i=1}^{n}\sum_{\substack{j=1\\ j>i}}^{n}(X_i-\mu)(X_j-\mu)\right] \\ &= \tfrac{1}{n^2}\left[\sum_{i=1}^{n} V[X_i] + 2\sum_{\overline{i}=1}^{n}\sum_{j>i} E[X_i-\mu]E[X_j-\mu]\right] = \tfrac{\sigma^2}{n} \end{aligned}$$

ya que el segundo sumando es nulo. La hipótesis $E[X^2] < +\infty$ es necesaria para asegurar la existencia de μ y σ^2.

Proposición 1.2.2
Dada una m.a.s. de tamaño n de una población con momento poblacional de cuarto orden finito, $\alpha_4 = E[X^4] < +\infty$, se tiene que

$$\begin{aligned} E[b_2] &= \tfrac{n-1}{n}\sigma^2 \\ V[b_2] &= \tfrac{\beta_4-\beta_2^2}{n} - 2\tfrac{\beta_4-2\beta_2^2}{n^2} + \tfrac{\beta_4+3\beta_2^2}{n^3} \end{aligned} \qquad (1.2.1)$$

donde β_k es el momento poblacional de orden k respecto al centro, $\beta_k = E[(X-\alpha_1)^k]$.

Demostración: Sólo se comprobará la primera expresión. Obsérvese que

$$E[b_2] = \frac{1}{n}\sum_{i=1}^{n}\left[E[X_i^2] + E[\overline{X}^2] - 2E[X_i\overline{X}]\right]$$

siendo inmediato el comprobar que

$$E[X_i^2] = \sigma^2 + \mu^2 \qquad E[\overline{X}^2] = \frac{\sigma^2 + n\mu^2}{n} \qquad E[X_i\overline{X}] = \frac{\sigma^2 + n\mu^2}{n}.$$

Ahora es inmediato el comprobar que $E[S^2] = \sigma^2$, la esperanza de la cuasi-varianza muestral S^2 dada por

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \tag{1.2.2}$$

es la varianza poblacional.

■

En la Sección 3.2 se verá la conveniencia de utilizar estadísticos muestrales que tengan de media el parámetro que se pretende estimar. Los momentos muestrales, también pueden introducirse en el caso de muestras de variables n dimensionales. Tan solo se va a abordar en este libro el caso de poblaciones bivariantes. Así, para una m.a.s. $(x_1, y_1), ..., (x_n, y_n)$ se pueden definir los *momentos marginales respecto al origen*

$$a_{k0} = \frac{1}{n}\sum_{i=1}^{n} x_i^k \qquad a_{0k} = \frac{1}{n}\sum_{i=1}^{n} y_i^k \qquad k > 0 \tag{1.2.3}$$

y los *momentos muestrales marginales respecto al centro*

$$b_{k0} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^k \qquad b_{0k} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^k \qquad k > 0 \tag{1.2.4}$$

pero esta vez se pueden definir además *momentos cruzados,* el más utilizado es

$$b_{11} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) \tag{1.2.5}$$

la covarianza muestral y que también acostumbra a notarse como b_{xy}. Se define el *coeficiente de correlación muestral* mediante

$$r = \frac{b_{xy}}{b_x b_y} \tag{1.2.6}$$

donde $b_x^2 = b_{20}$ y $b_y^2 = b_{02}$.

1.3. ESTADÍSTICA DESCRIPTIVA

Antes de continuar con el estudio de los momentos, para inferir a partir de ellos características de la población en estudio, se debe tener en cuenta que también es posible no pasar de la etapa de obtención de la muestra y realizar únicamente lo que se llama *Estadística Descriptiva*. Se entiende ésta como el resumen de los datos, o en terminología más actual el *Análisis de Datos*. Estos procedimientos, muchos de los cuales incluyen representaciones gráficas, son útiles para revelar la estructura de los datos, que frecuentemente vendrán escritos en columnas dentro de una página o registrados en una salida de un ordenador.

Lo primero que se debe tener en cuenta es que, desde este punto de vista, se debe empezar diferenciando entre variable cualitativa, cuando las observaciones sólo pueden clasificarse en categorías no numéricas, como por ejemplo el grupo sanguíneo, tipo de estudios, etc., y variables cuantitativas que son las que toman valores numéricos, como ingresos, número de llamadas a una centralita, altura, etc. De las primeras variables sólo se pueden hacer representaciones gráficas, que permitan hacerse idea de la variable en estudio, existiendo una gran variedad de representaciones; las más frecuentes suelen ser el *diagrama de rectángulos* y el *diagrama de sectores*.

Ejemplo 1.3.1
Se clasifica una colección de 155 coches, 85 europeos, 26 americanos y 44 japoneses, de acuerdo con el lugar de fabricación. Estos datos pueden representarse gráficamente mediante el *diagrama de rectángulos* y el *diagrama de sectores* contenidos en la Figura 1.1

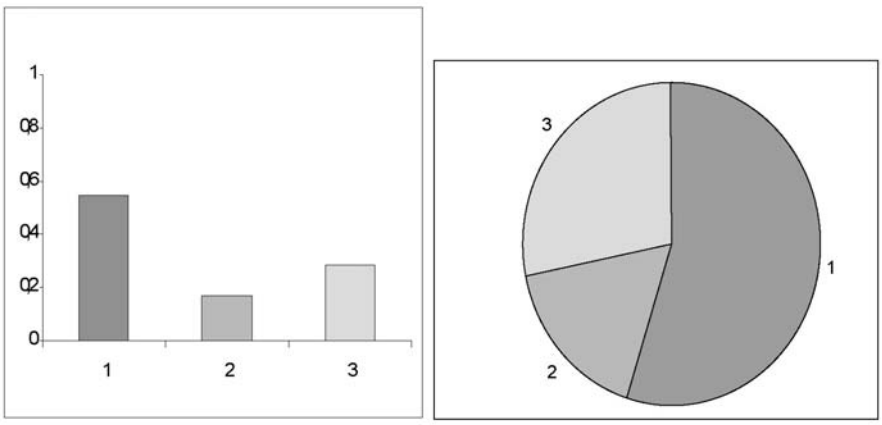


Figura 1.1. Diagrama de rectángulos y diagrama de sectores correspondientes al Ejemplo 1.3.1.

Las variables cuantitativas, por su parte, pueden clasificarse en discretas y continuas, las primeras sólo pueden tomar una cantidad finita o numerable de valores y las segundas pueden tomar cualquier valor en un intervalo acotado o no. Con independencia de si la variable cuantitativas es discreta o continua, y dependiendo básicamente del número de datos de que se disponga, a veces conviene agrupar los n elementos observados en distintas clases $C_1, C_2, ..., C_k$, siendo $x_1, x_2, ..., x_k$ los representantes de las clases y $n_1, n_2, ..., n_k$ las frecuencias absolutas, esto es, el número de observaciones dentro de cada clase. Se denota por $f_1, f_2, ..., f_k$ a las frecuencias relativas dentro de cada clase, donde $f_i = n_i/n$, se verifica que $\sum_{i=1}^{k} n_i = n$ y $\sum_{i=1}^{k} f_i = 1$. Alternativamente si el número de valores es grande se pueden construir las clases, para llegar a la *tabla de frecuencias*.

Otro problema que hay que resolver es el decidir el número de clases distintas a considerar, generalmente igual a $\sqrt{n}$.

También hay que tener presente que cuando los datos se observan pasando antes por una *tabla de frecuencias* se pierde algo de la información que se tiene cuando se consideran los datos directamente, a cambio se facilita el manejo y la interpretación de los mismos.

Al mismo tiempo se deben escoger los límites que definen los intervalos, de forma que las clases sean de la misma longitud, al menos la mayor parte de ellas, y de manera que cada observación pueda ponerse con precisión en una única clase.

Ejemplo 1.3.2
Los siguientes datos, tomados de McPherson (1990), recogen los niveles de colesterol en sangre medidos en miligramos de 135 personas:

145	145	146	158	158	163	168	168	168	168	172	174	175
175	175	175	175	175	178	180	180	180	181	181	187	187
192	194	194	195	196	196	198	198	198	198	201	201	201
204	204	204	205	205	205	205	205	206	206	206	207	208
209	211	211	211	212	212	212	214	214	214	216	217	217
217	217	218	218	218	218	221	221	221	221	223	224	225
225	227	277	227	227	228	230	235	235	235	236	237	238
238	239	241	242	242	243	243	245	245	245	247	248	252
253	253	253	254	254	254	256	256	257	257	261	262	266
267	267	268	268	268	268	273	274	278	283	284	285	289
292	294	296	300	345								

Estos datos se clasifican en 11 clases en la siguiente *tabla de frecuencias:*

Clases	Marcas de clase	frecuencias	frecuencias relativas
$130-150$	140	3	0'022
$150-170$	160	7	0'052
$170-190$	180	16	0'118
$190-210$	200	27	0'200
$210-230$	220	32	0'237
$230-250$	240	18	0'133
$250-270$	260	20	0'148
$270-290$	280	7	0'052
$290-310$	300	4	0'030
$310-330$	320	0	0'000
$330-350$	340	1	0'007

La tabla permite extraer consecuencias elementales de la colección de datos, en primer lugar el único dato contenido en la última clase, parece ser inusual y puede deberse quizás a una errata de transcripción, un tratamiento estadístico de estas observaciones puede verse en la Sección 8.4.4. En segundo lugar, más de la mitad de los valores tienden a concentrarse sobre las 5 clases más bajas, por lo que la distribución parece asimétrica.

■

Algunos autores reservan el nombre de *variables estadísticas* para las colecciones de datos cuando no se hace referencia a la distribución de probabilidad poblacional sobre la que se quieren inferir resultados, no considerándose por lo tanto a éstos datos como una muestra. Aquí no se va a establecer esta distinción.

1.3.1. Medidas de centralización

Una medida de centralización o localización indica el centro de una colección de números. Si los números provienen de diferentes medidas de la misma cantidad una medida de localización se utiliza para *conjeturar* el valor de esa cantidad. En otras situaciones la medida de localización representa un resumen de la colección de números, como cuando se dice, *la nota media de una clase en un examen es 6'5.*

Ya se han visto una colección de medidas de localización, los *momentos muestrales respecto al origen* incluidos en la Definición 1.2.3, el que se usa con mayor frecuencia es el primero $\overline{x} = a_1$ que cuando los datos están agrupados pasa a ser

$$\overline{x} = \sum_{i=1}^{k} f_i x_i.$$

La media muestral presenta el inconveniente de depender de todas las observaciones y se ve muy afectada por los valores extremos. Por este motivo se recurre a definir otras medidas de centralización que sean más *robustas*, es decir menos vulnerables en el sentido antes citado.

Definición 1.3.1.1 (mediana muestral)
Dada una colección finita de elementos la *mediana muestral* M es el valor central, entendiendo como tal el valor $M = x_{([\frac{n}{2}]+1)}$ si n es impar y $M = \frac{1}{2}\left(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}\right)$ si n es par.

Por $[x]$ se entiende la parte entera de x y $x_{(1)} \leq ... \leq x_{(n)}$ es la muestra ordenada.

Definición 1.3.1.2 (media truncada)
Para una colección finita de elementos se llama *media truncada* de orden α con $0 < \alpha < 0'5$, a la media $\overline{x}_\alpha$ de los elementos que quedan de quitar el $100\alpha\,\%$ más bajo y el $100\alpha\,\%$ más alto de los mismos. Formalmente

$$\overline{x}_\alpha = \frac{1}{n - 2[n\alpha]}\left(x_{([n\alpha]+1)} + ... + x_{(n-[n\alpha])}\right)$$

Las medias truncadas más usadas son $\overline{x}_{0'1}$ y $\overline{x}_{0'2}$.

Ejemplo 1.3.1.1
Para la colección de datos

$$35 \quad 69 \quad 72 \quad 77 \quad 79 \quad 79 \quad 84 \quad 85 \quad 86 \quad 87 \quad 95$$

la mediana muestral es $M = 79$ y la *media truncada* correspondiente a $\alpha = 0'2$ es

$$\begin{aligned} \overline{x}_{0'2} &= \tfrac{1}{7}(72 + 77 + ... + 86) \\ &= 80'2857 \end{aligned}$$

1.3.2. Medidas de dispersión

Una medida de dispersión o escala indica lo alejados que están los datos respecto a un valor central. Los *momentos muestrales respecto al centro*, ver Definición 1.2.3, son todos ellos medidas de dispersión. El más usado es la *cuasivarianza muestral*, véase la Expresión 1.2.2, que cuando los datos están agrupados toma la forma

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{k} f_i(x_i - \overline{x})^2$$

donde f_i es la *frecuencia relativa* con que aparece el valor x_i, es decir el cociente entre el número de veces que aparece x_i partido por el número total

de observaciones n. La cuasivarianza muestral también presenta el inconveniente de ser muy sensible a errores en los datos; por lo que en homogeneidad con las medidas de localización se recurre a definir otras medidas de dispersión.

Definición 1.3.2.1 (cuantiles; recorrido intercuartílico)
Dada una colección finita de n números se llama *cuantil muestral* de orden p, con $p \in (0,1)$ al valor x entendiendo como tal el valor $x_{([np]+1)}$ si np no es un número entero y $\frac{1}{2}(x_{(np)} + x_{(np+1)})$ si np es un número entero. Los *cuantiles muestrales* más usados son la *mediana* que se corresponde con $p = 1/2$ (véase la Definición 1.3.1.1) y los tres cuartiles que se corresponden el *primer cuartil* con $p = 1/4$, el *segundo cuartil* con $p = 2/4$ (coincide con la mediana), y el *tercer cuartil* con $p = 3/4$. Se denomina *recorrido intercuartílico muestral* (RIM), a la diferencia entre el tercer y primer cuartil muestrales de la misma forma que se denomina *recorrido intercuartílico poblacional* a la diferencia entre el tercer y primer cuartil poblacional.

Se recuerda que dada una función de distribución F se define *cuantil poblacional* de orden p, al valor x_p tal que se verifique

$$F(x_p) \geq p \qquad \text{y} \qquad P\{X \geq x_p\} \geq 1 - p \tag{1.3.2.1}$$

o de forma equivalente

$$p \leq F(x_p) \leq p + P\{X = x_p\}.$$

La definición de cuantil muestral puede hacerse de la misma forma cambiando la función de distribución poblacional por la función de distribución empírica, (véase la Definición 1.4.1). Aquí se ha preferido dar una definición esencialmente idéntica, pero que cambie todo un intervalo por un único valor, la semisuma de los extremos.

Definición 1.3.2.2 (mediana de las desviaciones absolutas)
Dada la colección de n números se llama *mediana de las desviaciones absolutas* MDA, a la mediana de los números $|x_i - M|$, donde M es la mediana muestral. Es también una medida *robusta* de dispersión en el sentido de que no es demasiado sensible a observaciones erróneas o atípicas.

Ejemplo 1.3.2.1
Para la colección de datos

$$35 \quad 69 \quad 72 \quad 77 \quad 79 \quad 79 \quad 84 \quad 85 \quad 86 \quad 87 \quad 95$$

la *cuasivarianza muestral* es $s^2 = 247'89$, el *primer cuartil* es 72 el *tercer cuartil* es 86, el *recorrido intercuartílico* (RIM) es 14, la *mediana muestral* es 79 y la *mediana de las desviaciones absolutas* (MDA) es 7.

1.3.3. Diagramas de tallos y hojas y de caja

Como ya se ha dicho, la representación más extendida de una colección de datos es mediante la *tabla de frecuencias,* sin embargo existen otras representaciones adecuadas, que están a medio camino entre la representación numérica y gráfica. Éstas son el *diagrama de tallos y hojas* y el *diagrama de caja*, principalmente. El *diagrama de tallos y hojas* se construye realizando los siguientes pasos:

a) se redondea la colección de datos a dos o tres cifras
b) las cifras de orden superior determinan los *tallos* normalmente de una o dos cifras
c) a continuación se escribe una raya vertical y se pone la cifra de las unidades de orden inferior que determina las *hojas*

así los datos 1236, 1278 darían lugar, redondeados a tres cifras, a 124, 128 ambos se corresponderían con el *tallo* 12 y darían lugar a las *hojas* 4 y 8 con lo que su representación en el diagrama sería

12|48

Ejemplo 1.3.3.1
Para los datos

256	256	257	257	261	262	266	267	267
268	268	268	268	273	274	278	283	284

el *diagrama de tallos y hojas* correspondiente viene dado por

4	25	6677
(13)	26	126778888
5	27	348
2	28	34

Siguiendo una práctica usual se han incluido delante de los *tallos* las frecuencias acumuladas hasta llegar al *tallo* que contiene a la mediana, que se indica con paréntesis. Se observa que las *hojas* juegan gráficamente el mismo papel, que las frecuencias absolutas correspondientes a las clases en la tabla de frecuencias, con la ventaja de que informan de los valores que realmente toma la variable.
■

El *diagrama de caja* es también una representación a medio camino entre numérica y gráfica y recoge la mediana, el recorrido intercuartílico, la presencia de posibles valores discordantes o atípicos, junto con una indicación de la posible asimetría de los datos. El *diagrama de caja* se construye de la siguiente forma:

a) se representan tres líneas horizontales correspondientes a la mediana y al primer y al tercer cuartil, que dan lugar a la caja;
b) se trazan dos rectas horizontales, de menor tamaño que las que delimitan la caja y que cubran las observaciones que queden dentro del primer cuartil menos 1'5 RIM y del tercer cuartil más 1'5 RIM;
c) se representan por un punto las observaciones que quedan por debajo del primer cuartil menos 1'5 RIM y por encima del tercer cuartil más 1'5 RIM si es que existen.

Los *diagramas de caja* no están estandarizados, la estructura que se acaba de explicar es la más extendida con modificaciones mínimas. Esta representación da una idea de donde se encuentra el centro de la colección, mediante la mediana; de la dispersión de los datos, mediante el RIM; de la asimetría de los mismos, mediante la distancia de los cuartiles a la mediana y por último informa de la posible existencia de observaciones discordantes, o que desentonan demasiado con los restantes datos en el sentido establecido en c).

Ejemplo 1.3.3.2
Para los datos del Ejemplo 1.3.2 el *diagrama de caja* está recogido en la Figura 1.3.3.1 los datos para su constucción son: la mediana 218, el primer cuartil 198 y el tercer cuartil 248. El dato atípico 345 ha dado lugar al

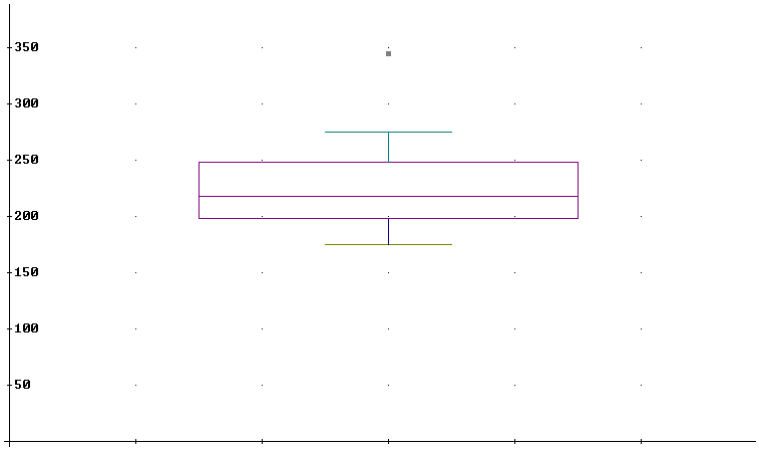


Figura 1.3.3.1. Diagrama de caja correspondiente al Ejemplo 1.3.3.2.

punto aislado por encima de las rectas horizontales. Se observa la ligera asimetría de los puntos por encima de la mediana.

1.4. PROPIEDADES ASINTÓTICAS DE LOS MOMENTOS MUESTRALES

El objetivo de esta sección es demostrar que los momentos muestrales convergen casi seguro a sus correspondientes momentos poblacionales, lo que justificará la utilización de los primeros para estimar los segundos en una población. La aproximación será entonces mayor cuanto mayor sea el tamaño muestral. Además se incluye la función de distribución empírica.

Teorema 1.4.1
Si una población tiene momentos de orden k, el momento muestral respecto al origen de orden k, converge, casi seguro, al momento poblacional respecto al origen de orden k

$$a_k \xrightarrow{c.s.} \alpha_k.$$

Demostración: Basta ver que por la ley fuerte de Kintchine (Loève (1963) pág. 239) dada una sucesión de v.a. $\{Y_n\}_{n\in\mathbb{N}}$ con $E[y_i] = \mu < +\infty$ se sigue que

$$\overline{Y}_n \xrightarrow{c.s.} \mu$$

donde $\overline{Y}_n = \frac{1}{n}\sum Y_i$. El resultado se obtiene tomando $Y_i = X_i^k$.

Teorema 1.4.2
Si una población tiene momentos de orden k, el momento muestral respecto al centro de orden k, converge, casi seguro, al momento poblacional respecto al centro de orden k

$$b_k \xrightarrow{c.s.} \beta_k.$$

Demostración: Si se tiene en cuenta que

$$b_k = a_k - \binom{k}{1}\overline{x}\, a_{k-1} + ... \pm \overline{x}^k$$

y que si $\{X_n^{(i)}\}_{n\in\mathbb{N}}$ $i = 1, ..., p$ converge c.s. a $X^{(i)}$ y es g una función continua de $\mathbb{R}^p$ en $\mathbb{R}$, entonces

$$g(x_n^{(1)}, ..., x_n^{(p)}) \xrightarrow{c.s.} g(x^{(1)}, ..., x^{(p)}),$$

resultado que se demuestra fácilmente, ya que si se denota por $\overrightarrow{X}_n = (X_n^{(1)}, ..., X_n^{(p)})$ y $\overrightarrow{X} = (X^{(1)}, ..., X^{(p)})$, por $\overrightarrow{X}_n \xrightarrow{c.s.} \overrightarrow{X}$ $P\{\lim_{n\to\infty} \overrightarrow{X}_n = \overrightarrow{X}\} = 1$. Pero el suceso $\lim_{x\to\infty} \overrightarrow{X}_n = \overrightarrow{X}$ está contenido en el suceso $\lim_{n\to\infty} g(\overrightarrow{X}_n) = g(\overrightarrow{X})$ puesto que si $\lim_{n\to\infty} \overrightarrow{X}_n = \overrightarrow{X}$ aplicando g se sigue que $g(\lim_{n\to\infty} \overrightarrow{X}_n) = g(\overrightarrow{X})$ y por ser g continua $\lim_{n\to\infty} g(\overrightarrow{X}_n) = g(\lim_{n\to\infty} \overrightarrow{X}_n)$; luego

$$P\{\lim_{n\to\infty} g(\overrightarrow{X}_n) = g(\overrightarrow{X})\} = 1$$

tomando como g la expresión para b_k y como $X_n^{(i)}$ los momentos respecto al origen al variar i, se obtiene que $b_k \xrightarrow{c.s.} \beta_k$.

■

También se puede probar la convergencia en ley de los momentos muestrales hacia la distribución normal.

Teorema 1.4.3
Si la población tiene momentos de orden $2k$ respecto al origen entonces se tiene la siguiente convergencia en ley

$$\frac{a_k - \alpha_k}{\sqrt{\alpha_{2k} - \alpha_k^2}}\sqrt{n} \xrightarrow{\mathcal{L}} Z \sim N(0,1).$$

Demostración: Basta utilizar el teorema central del límite, $\{Y_n\}$ con v.a.i.i.d. con momento de segundo orden y $E[Y_n] = \mu \quad V[Y_n] = \sigma^2 < +\infty$ entonces $\frac{\overline{Y}-\mu}{\sigma}\sqrt{n} \longrightarrow Z \sim N(0,1)$. Se puede considerar entonces $Y_n = X_n^k$, $a_k = \overline{Y}$ con lo que $\mu = \alpha_k$ y $\sigma^2 = \alpha_{2k} - \alpha_k^2$.

■

La convergencia en ley de los momentos muestrales respecto al centro queda recogida en el siguiente resultado:

Teorema 1.4.4
Si existe el momento poblacional respecto al origen de orden $2k$ entonces

$$\sqrt{n}\frac{b_k - \beta_k}{\sqrt{\beta_{2k} - \beta_k^2 - 2k\beta_{k-1}\beta_{k+1} + k^2\beta_{k-1}^2\beta_2}} \xrightarrow{\mathcal{L}} Z\, N(0,1).$$

Demostración: Puede verse en Cramér (1973) pág. 419 y se apoya en el Lema 1.4.1 que tiene interés por sí mismo por lo que se va a demostrar a continuación. Además el siguiente resultado, que relaciona la convergencia en ley con la convergencia en probabilidad, también será de utilidad en el estudio de convergencias a lo largo del texto.

Teorema 1.4.5 (Slutsky)
Si $X_n \xrightarrow{\mathcal{L}} X$ y $Y_n \xrightarrow{P} a$ entonces se tiene que

I) $Y_n X_n \xrightarrow{\mathcal{L}} aX$

II) $X_n + Y_n \xrightarrow{\mathcal{L}} X + a$

Demostración: puede verse en Bickel y Doksun(1977) pág. 461.

Lema 1.4.1
Si $\{a_n\}$ es una sucesión de constantes con $\lim_{n\to\infty} a_n = +\infty$ y a es un número fijo, tal que $a_n(X_n - a) \xrightarrow{\mathcal{L}} X$, entonces para cualquier función $g : \mathbb{R} \to \mathbb{R}$ con derivada continua y no nula en a, se tiene que

$$a_n(g(X_n) - g(a)) \xrightarrow{\mathcal{L}} g'(a)X$$

Demostración: Si se desarrolla $g(X_n)$ alrededor de $g(a)$ en serie de Taylor, se tiene

$$g(X_n) = g(a) + (X_n - a)g'(a'_n)$$

donde $|a'_n - a| \leq |X_n - a|$. Ahora bien de las hipótesis se sigue que $X_n \xrightarrow{P} a$ ya que si se considera $X_n - a = \frac{1}{a_n}[a_n(X_n - a)]$ como la expresión entre corchetes converge en ley a X y $\frac{1}{a_n} \to 0$, por el Teorema de Slutsky se tiene que $X_n - a \xrightarrow{P} 0$. Por tanto $a'_n \xrightarrow{P} a$. Como $\alpha_n(g(X_n) - g(a)) = a_n(X_n - a)g'(a'_n)$ y la función g' es continua en a, entonces $g'(a'_n) \xrightarrow{P} g'(a)$ por lo que aplicando otra vez el Teorema de Slutsky se obtiene el resultado deseado.

■

Debido a su carácter más elemental, se incluye la obtención de la distribución asintótica del momento muestral respecto al centro más sencillo: la varianza muestral.

Teorema 1.4.6
Para m.a.s. de una población con momento α_4 finito, la distribución límite de $b_2 = \frac{1}{n}\sum_1^n (X_i - \overline{X})^2$ es tal que

$$\sqrt{n}(b_2 - \sigma^2) \xrightarrow{\mathcal{L}} Z \sim N(0, \sqrt{\beta_4 - \sigma^4})$$

Demostración: Se puede poner

$$b_2 = \frac{1}{n}\sum_{i=1}^n (X_i - \mu - (\overline{X} - \mu))^2 = \frac{1}{n}\sum_{i=1}^n (Y_i - \overline{Y})^2$$

donde las variables $Y_i = X_i - \mu$ son independientes, tienen de media 0 y de varianza σ^2. Como

$$b_2 = \frac{1}{n}\sum_{i=1}^n Y_i^2 - \overline{Y}^2$$

se tiene que

$$\sqrt{n}(b_2 - \sigma^2) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n Y_i^2 - \sigma^2\right) - \sqrt{n\overline{Y}^2}.$$

Pero

$$\sqrt{n\overline{Y}^2} = (n^{1/4}\overline{Y})^2 \qquad \text{y} \qquad n^{1/4}\overline{Y} = n^{-1/4}\sqrt{n\overline{Y}};$$

como

$$\sqrt{n\overline{Y}} \xrightarrow{\mathcal{L}} Z \sim N(0, \sigma)$$

(en virtud del Teorema 1.4.3) y $n^{-1/4} \xrightarrow{P} 0$; luego por el Teorema de Slutsky, $n^{1/4}\overline{Y} \xrightarrow{\mathcal{L}} 0$ y por tanto converge también en probabilidad, con lo que su cuadrado, $\sqrt{n}\overline{Y}^2$ converge en probabilidad hacia 0. Por el Teorema 1.4.3,

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \sigma^2\right) \xrightarrow{\mathcal{L}} Z \sim N(0, \sqrt{\beta_4 - \sigma^4}),$$

puesto que

$$E\left[\frac{1}{n}\sum_{i=1}^{n} Y_i^2\right] = \sigma^2 \ y \ \ V\left[\frac{1}{n}\sum_{i=1}^{n} Y_i^2\right] = \frac{V[Y_i^2]}{n} = \frac{\beta_4 - \sigma^4}{n}.$$

Otra vez más por el Teorema de Slutsky

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \sigma^2\right) - \sqrt{n}\overline{Y}^2 \xrightarrow{\mathcal{L}} Z \sim N(0, \sqrt{\beta_4 - \sigma^4}).$$

Ejemplo 1.4.1
Si se sabe que

$$\sqrt{n}(Z_n - \theta) \xrightarrow{\mathcal{L}} X$$

donde $X \sim N(0, \sigma(\theta))$, lo que ocurre por ejemplo para $Z_n = \overline{X}$ en una población con momento de orden 2 finito y media θ, entonces

$$\sqrt{n}(g(Z_n) - g(\theta)) \xrightarrow{\mathcal{L}} Y$$

con $Y \sim N(0, |g'(\theta)|\sigma(\theta))$, sin más que exigir que g tenga derivada continua y no nula en el punto θ.

Ejemplo 1.4.2 (estabilización de la varianza)
Dada una v.a., si se conoce la distribución asintótica de una función sencilla de ella, como ocurre en el Ejemplo 1.4.1, se puede tratar de obtener una función cuya varianza se independiente de θ. Por ejemplo, si X sigue una distribución de Poisson $\mathcal{P}(\theta)$, entonces, por el teorema central del límite

$$\sqrt{n}\left(\overline{X} - \theta\right) \xrightarrow{\mathcal{L}} Z$$

con $Z \sim N(0, \sqrt{\theta})$,

$$\sqrt{n}(g(\overline{X}) - g(\theta)) \xrightarrow{\mathcal{L}} Y$$

con $Y \sim N(0, |g'(\theta)|\sqrt{\theta})$. Se trata de obtener una función g tal que $\sqrt{\theta}g'(\theta)$ no dependa de θ. Esto se consigue, por ejemplo, con $g(\theta) = \sqrt{\theta}$, pues $\sqrt{\theta}g'(\theta) = \frac{1}{2}$. Por tanto $g(\overline{X}) = \sqrt{\overline{X}}$ es tal que

$$\sqrt{n}\left(\sqrt{\overline{X}} - \sqrt{\theta}\right) \xrightarrow{\mathcal{L}} Y$$

con $Y \sim N(0, 1/2)$.

Ejemplo 1.4.3
Para una población de Bernoulli de parámetro θ, la distribución asintótica de $\sqrt{n}\overline{X}(1-\overline{X})$ puede calcularse, debidamente centrada, a partir de la distribución de $\overline{X}$. Por el teorema central del límite se sabe que

$$\sqrt{n}\left(\overline{X}-\theta\right) \xrightarrow{\mathcal{L}} Z \sim N(0, \sqrt{\theta(1-\theta)})$$

luego si se considera $g(\theta)=\theta(1-\theta)$ se sigue, aplicando el Lema 1.4.1, que

$$\sqrt{n}(\overline{X}(1-\overline{X})-\theta(1-\theta)) \xrightarrow{\mathcal{L}} Z \in N(0, |1-2\theta|\sqrt{\theta(1-\theta)}).$$

■

Otro estadístico muestral importante es la *función de distribución empírica* o *muestral*.

Definición 1.4.1 (distribución empírica)
Dada una realización particular de una muestra $(x_1, ..., x_n)$ viene definida mediante

$$F_n^*(x) = \begin{cases} 0 & \text{si} \quad x < x_{(1)} \\ \frac{k}{n} & \text{si} \quad x_{(k)} \leq x < x_{(k+1)} \\ 1 & \text{si} \quad x \geq x_{(n)} \end{cases} \qquad k = 1, ..., n-1$$

donde se recuerda que $(x_{(1)}, ..., x_{(n)})$ es la muestra ordenada de menor a mayor.

■

Alternativamene la función $F_n^*(x)$ puede ponerse como

$$F_n^*(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x]}(x_i). \tag{1.4.1}$$

Para todo x se verifica que $F_n^*(x) \xrightarrow{c.s.} F(x)$, siendo $F(x)$ la función de distribución de la población. En efecto, haciendo

$$Y_i = I_{(-\infty,x]}(x_i)$$

se tiene que $E[Y_i] = F(x)$, $V[Y_i] = F(x)[1-F(x)]$ y $F_n^*(x) = \overline{Y}$. Aplicando ahora la ley fuerte de los grandes números se obtiene la convergencia c.s. citada. Además, de la Expresión (1.4.1) se sigue, por el teorema central del límite, que

$$\sqrt{n}\frac{F_n^*(x)-F(x)}{\sqrt{F(x)[1-F(x)]}} \xrightarrow{\mathcal{L}} Z \sim N(0,1).$$

La convergencia casi seguro puede refinarse aún más como indica el siguiente teorema, que es llamado por algunos el teorema central de la estadística.

Teorema 1.4.7 (Glivenko-Cantelli)
Si se tienen m.a.s. de tamaño n de una población X, con función de distribución $F(x)$, para cualquier número real positivo arbitrario ε, se tiene que

$$\lim_{n\to\infty} P\left\{\sup_{x\in\mathbb{R}}|F_n^*(x) - F(x)| \geq \varepsilon\right\} = 0.$$

Demostración: La demostración puede verse en Fisz (1963) pág. 391.
■

Es importante señalar que en este teorema la probabilidad del primer miembro está obtenida a partir de la distribución F; por lo que sería más correcto emplear la notación P_F en lugar de P.

En virtud del Teorema 1.4.7 si se quiere determinar de forma aproximada la función de distribución, se puede obtener una m.a.s. de tamaño n gande, construir a partir de ella la función de distribución empírica F_n^* y esperar que $F_n^*(x)$ sea una buena aproximación de $F(x)$.

Ejemplo 1.4.4
Es inmediato, mediante un programa de ordenador, construir $F_{100}^*(0)$ a partir de una m.a.s. de una población $N(0,1)$. Por ejemplo, simulando una muestra de 100 valores de la distribución $N(0,1)$ se observa que son 52 menores o iguales que 0, por lo tanto es $F_{100}^*(0) = 0,52$. Es sabido que el valor de la función de distribución de la $N(0,1)$ en el punto 0 es $0'5$, por lo que la aproximación es bastante buena.

1.5. DISTRIBUCIONES EN EL MUESTREO ASOCIADAS A LA DISTRIBUCIÓN NORMAL

El Teorema de Glivenko-Cantelli justifica de alguna manera, la sustitución de los parámetros desconocidos de la función de distribución de la población, por sus características análogas calculadas a partir de la muestra. Es por tanto interesante estudiar las distribuciones de los estadísticos muestrales cuando la población de partida es conocida. Debido al papel central de la distribución normal en el Cálculo de Probabilidades, se estudia a continuación la distribución de los momentos muestrales cuando la población es normal.

1.5.1. Distribución de la media muestral

Dada una m.a.s. de tamaño n de una población X con distribución $N(\mu,\sigma)$, se sigue que la distribución de la media muestral $\overline{X}$ es una $N(\mu, \frac{\sigma}{\sqrt{n}})$.

En efecto, la función característica de la media muestal es

$$\begin{aligned}\varphi_{\overline{X}}(t) &= E[e^{i\frac{t}{n}\sum_1^n X_i}] = [\varphi_X(\tfrac{t}{n})]^n = e^{\left(i\frac{t}{n}\mu-\frac{1}{2}\frac{\sigma^2}{n^2}t^2\right)n} \\ &= e^{it\mu-\frac{1}{2}\frac{\sigma^2}{n}t^2}\end{aligned}$$

que es la función característica de la distribución $N(\mu, \frac{\sigma}{\sqrt{n}})$.

1.5.2. Distribución del momento muestral respecto al origen de orden dos

Para m.a.s. de tamaño n de una población $N(0,\sigma)$ la distribución de $a_2 = \frac{1}{2}\sum_1^n X_i^2$ viene dada por

$$g(y) = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})}\left(\frac{n}{\sigma^2}\right)^{n/2} e^{-\frac{n}{2\sigma^2}y} y^{\frac{n}{2}-1} \qquad \text{con} \qquad y \geq 0. \tag{1.5.2.1}$$

Procediendo por pasos sucesivos, si X tiene distribución $N(0,1)$ la distribución de X^2 es una gamma $(a = \frac{1}{2}, p = \frac{1}{2})$, ya que para $y \geq 0$

$$P\{X^2 \leq y\} = P\{-\sqrt{y} \leq X \leq \sqrt{y}\} = 2\Phi(\sqrt{y}) - 1$$

con Φ la función de distribución de la v.a. $N(0,1)$. Derivando respecto de y

$$f_{X^2}(y) = \varphi(\sqrt{y})y^{-1/2}$$

donde φ es la función de densidad de la $N(0,1)$. Por tanto, la densidad de X^2 es $\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x}x^{-\frac{1}{2}}$, con $x \geq 0$. Como la función característica de la distribución *Gamma* (a,p) es $\left(1-\frac{2it}{a}\right)^{-p}$, se sigue que la distribución de $\sum_1^n X_i^2$, cuando X_i se distribuye $N(0,1)$, es una *Gamma* $(a = \frac{1}{2}, p = \frac{n}{2})$, es decir

$$f(x) = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})}e^{-\frac{1}{2}x}x^{\frac{n}{2}-1} \qquad \text{con} \qquad x \geq 0. \tag{1.5.2.2}$$

A una v.a. con esta densidad se la denomina variable aleatoria χ^2_n, donde el parámetro n recibe el nombre de *número de grados de libertad*. Es fácil establecer (ver Ejercicio 1.9.14) que la media de la distribución χ^2_n es n y que su varianza es $2n$. En la Figura 1.5.2.1 pueden verse distintas representaciones de la función de densidad al variar el número de grados de libertad.

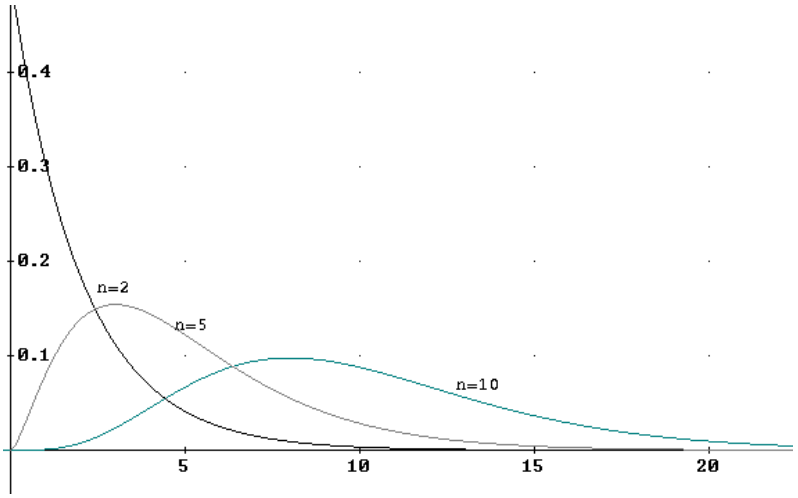


Figura 1.5.2.1. Funciones de densidad de la v.a. χ^2_n para $n = 2, 5, 10$.

Se puede entonces introducir la siguiente distribución.

Definición 1.5.1 (distribución χ^2_n)
Para m.a.s. de tamaño n de una población $N(0,1)$ $(X_1, ..., X_n)$ la v.a. $Y = \sum_{i=1}^{n} X_i^2$ tiene una distribución χ^2_n con n el número de grados de libertad, o de forma equivalente cuando su función de densidad es

$$f_{\chi^2_n}(x) = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} e^{-\frac{1}{2}x} x^{\frac{n}{2}-1} \qquad \text{con} \qquad x \geq 0 \tag{1.5.2.2}$$

Al final del libro, en el Apéndice titulado Tablas, la Tabla 4 recoge cuantiles útiles de la distribución χ^2_n para diferentes valores de n. A título de ejemplo la probabilidad de que la distribución χ^2_{24} tome un valor superior a $36'4$ es $1 - 0'95 = 0'05$. La distribución χ^2_n puede generalizarse al caso en que las variables X_i sean independientes pero tengan distribución $N(\mu_i, 1)$ con $\mu_i \neq 0$. Ya que la función característica de $\sum_{i=1}^{n} X_i^2$ es

$$\begin{aligned} E[e^{it\sum_{i=1}^{n} X_i^2}] &= \int ... \int e^{it\sum_1^n x_j^2} e^{-\frac{1}{2}\sum_{j=1}^{n}(x_j-\mu_j)^2} \frac{1}{(2\pi)^{n/2}} dx_1...dx_n \\ &= \frac{1}{(2\pi)^{n/2}} \int ... \int e^{-\frac{1-2it}{2}\sum_{j=1}^{n}(x_j-\frac{\mu_j}{1-2it})^2 - \sum_{j=1}^{n}\frac{\mu_j^2}{2}(1-\frac{1}{1-2it})} dx_1...dx_n \\ &= \frac{1}{(2\pi)^{n/2}} e^{-\lambda(1-\frac{1}{1-2it})} \int ... \int e^{-\frac{1-2it}{2}\sum_{j=1}^{n}(x_j-\frac{\mu_j}{1-2it})^2} dx_1...dx_n \end{aligned}$$

donde $\lambda = \frac{1}{2}\sum_1^n \mu_j^2$. Como

$$\begin{aligned} e^{-\lambda(1-\frac{1}{1-2it})}(1-2it)^{-n/2} &= e^{-\lambda}(1-2it)^{-n/2}\sum_{j=0}^{\infty}\frac{\lambda^j}{j!(1-2it)^j} \\ &= e^{-\lambda}\sum_{j=0}^{\infty}\frac{\lambda^j}{j!}(1-2it)^{-(\frac{n}{2}+j)} \end{aligned}$$

mediante el teorema de inversión de la función característica, (véase Loève(1963) pág. 188), se obtiene la función de densidad, que resulta ser

$$\begin{aligned} f(x) &= \tfrac{1}{2\pi}\textstyle\int_{-\infty}^{+\infty} e^{-itx}\varphi(t)dt \\ &= \sum_{j=0}^{\infty} e^{-\lambda}\frac{\lambda^j}{j!}\frac{1}{2^{\frac{n}{2}+j}\Gamma(\frac{n}{2}+j)}e^{-\frac{x}{2}}x^{\frac{n}{2}+j-1} \quad \text{con} \quad x \geq 0 \end{aligned} \tag{1.5.2.3}$$

densidad correspondiente a la v.a. χ_n^2 no centrada con *parámetro de no centralidad* $\lambda = \frac{1}{2}\sum_1^n \mu_j^2$ y n número de grados de libertad.

■

Para obtener la distribución de otros momentos muestrales, se necesita el siguiente resultado que recoge la distribución de ciertas funciones de v.a. normales.

Lema 1.5.1
Dada una m.a.s. de tamaño n $(X_1, ..., X_n)$ de una población $N(0, \sigma)$, si se construyen las v.a.

$$Y_i = c_i' X = (c_{i_1}, ..., c_{i_n}) \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \qquad i = 1, ..., p$$

donde p es menor que n, con $c_i' c_j = \delta_{ij}$ ($\delta_{ij} = 1$ si $i = j$ y $\delta_{ij} = 0$ si $i \neq j$), la v.a.

$$Z = \sum_{i=1}^{n} X_i^2 - \sum_{i=1}^{p} Y_i^2,$$

es tal que $\frac{Z}{\sigma^2} \sim \chi_{n-p}^2$, y además la v.a. Z es independiente de las variables $Y_1, ..., Y_p$.

Demostración: Inicialmente se determinan $n - p$ vectores más c_i de dimensión $n \times 1$ para $i = p + 1, ..., n$, que sigan cumpliendo la restricción $c_i' c_j = \delta_{ij}$; esto siempre es posible ya que se puede encontrar en $\mathbb{R}^n$ una base ortonormal de n elementos cuyos p primeros vectores sean los vectores c_i, $i = 1, ..., p$. Se tiene entonces

$$Y_i = c_i' \mathbf{X} \qquad i = 1, ..., n$$

o escrito matricialmente $\mathbf{Y} = C\mathbf{X}$. Es claro que $E[Y_i] = c_i' E[X] = 0$ y por ser las variables Y_i combinaciones lineales de variables normales, tienen distribución normal de media 0, varianza σ^2 y son independientes, pues

$$Cov(Y_i, Y_j) = \sigma^2 \sum_{k=1}^{n} c_{ik} c_{jk} = \sigma^2 \delta_{ij}.$$

Si ahora se expresa la variable aleatoria Z en función de las v.a. Y_i

$$Z = \mathbf{X}'\mathbf{X} - \sum_{i=1}^{p} Y_i^2 = \mathbf{Y}' C C' \mathbf{Y} - \sum_{i=1}^{p} Y_i^2 = \sum_{i=p+1}^{n} Y_i^2$$

con lo que Z es independiente de $Y_1, ..., Y_p$ y además $\frac{Z}{\sigma^2}$, de acuerdo con la Definición 1.5.2.1, tiene distribución χ^2_{n-p}.

Teorema 1.5.1 (Fisher)
Para m.a.s. de tamaño n de una población $N(\mu, \sigma)$ se tiene que b_2, momento muestral respecto al centro de orden 2 y $\overline{X}$, media muestral, son variables aleatorias independientes, además $\overline{X}$ tiene distribución $N(\mu, \frac{\sigma}{\sqrt{n}})$ y $n\frac{b_2}{\sigma^2}$ tiene distribución χ^2_{n-1}.

Demostración: Basta observar que

$$\begin{aligned} nb_2 &= \sum_{i=1}^{n} (X_i - \overline{X})^2 = \sum_{i=1}^{n} (X_i - \mu - (\overline{X} - \mu))^2 \\ &= \sum_{i=1}^{n} (X_i - \mu)^2 - n\left(\overline{X} - \mu\right)^2. \end{aligned}$$

Dado que $Y_i - \mu \sim N(0, \sigma)$, se puede aplicar el Lema 1.5.2.1 con $p = 1$,

$$Y_1 = \left(\frac{1}{\sqrt{n}}, ..., \frac{1}{\sqrt{n}}\right) \begin{pmatrix} X_1 - \mu \\ \vdots \\ X_n - \mu \end{pmatrix}$$

y $Z = nb_2$. Entonces $n\frac{b_2}{\sigma^2} \sim \chi^2_{n-1}$ es independiente de Y_1 y, por tanto, también de $\overline{X}$. En definitiva, en m.a.s. de una población $N(\mu, \sigma)$, la media muestral $\overline{X}$ y la cuasivarianza muestral son v.a. independientes y además son conocidas sus distribuciones.

1.5.3. Distribución del cociente entre la media muestral y el momento respecto al origen de orden dos

Para encontrar la distribución de este cociente es necesario introducir la distribución de Student, la segunda distribución en el muestreo asociada

a poblaciones normales. Se supone que $X, X_1, ..., X_n$ son v.a. independientes con distribución $N(0, \sigma)$, y se pretende encontrar la distribución de la variable

$$T_n = \frac{X}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2}}. \tag{1.5.3.1}$$

Para determinar su función de densidad, se puede suponer que todas las variables son $N(0,1)$ ya que dividiendo por σ numerador y denominador

$$T = \frac{Z}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} Z_i^2}} = \frac{Z}{Y}$$

las v.a. $Z_i \sim N(0,1)$. Con ello las funciones de densidad de Z e Y, véase la Expresión 1.5.2.2, son

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad \text{para} \quad z \in \mathbb{R},$$
$$f(y) = \frac{n^{\frac{n}{2}}}{2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})} e^{-\frac{n}{2}y^2} y^{n-1} \quad \text{para} \quad y > 0,$$

y la densidad conjunta de Z e Y es

$$f(z, y) = f(z) f(y).$$

Haciendo el cambio $T = \frac{Z}{Y}$ $U = Y$, cuyo jacobiano vale U, se obtiene como densidad conjunta de T y U

$$f(t, u) = k e^{-\frac{1}{2}(t^2+n)u^2} u^n \qquad \text{para} \;\; t \in \mathbb{R} \;\; u \geq 0$$

donde $k = n^{n/2}/(2^{\frac{n-1}{2}}\Gamma(\frac{1}{2})\Gamma(\frac{n}{2}))$. La densidad de T se obtiene como marginal mediante

$$f(t) = \int_0^{+\infty} f(t, u)du = k \int_0^{+\infty} e^{-\frac{1}{2}(t^2+n)u^2} u^n du$$

el cambio en la integral $u^2 = x$ hace que ésta quede

$$\int_0^{+\infty} e^{-\frac{1}{2}(t^2+n)u^2} u^n du = \frac{1}{2} \int_0^{+\infty} e^{-\frac{n+t^2}{2}x} x^{\frac{n}{2}-\frac{1}{2}} dx$$

y mediante la expresión

$$\int_0^{+\infty} e^{-ax} x^{p-1} dx = \Gamma(p) a^{-p}$$

con $a = \frac{n+t^2}{2}$ y $p = \frac{n}{2} + \frac{1}{2}$, se tiene que la integral vale

$$\frac{1}{2}\left(\frac{n+t^2}{2}\right)^{-\frac{n+1}{2}} \Gamma\left(\frac{n+1}{2}\right)$$

de donde

$$f(t) = \frac{n^{\frac{n}{2}}}{2^{\frac{n}{2}-\frac{1}{2}}\Gamma(\frac{1}{2})\Gamma(\frac{n}{2})}\frac{1}{2}\left(\frac{n+t^2}{2}\right)^{-\frac{n+1}{2}} \Gamma\left(\frac{n+1}{2}\right). \tag{1.5.3.2}$$

Se llega así a la siguiente definición.

Definición 1.5.3.1 (distribución de Student)
Dadas las v.a. independientes $X, X_1, ..., X_n$ con distribución $N(0,1)$ la v.a. $T_n = X \Big/ \sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2}$ tiene una distribución T_n de Student con n grados de libertad. De forma equivalente, una v.a. T_n tiene esta distribución, cuando su función de densidad es

$$f_{T_n}(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n}\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}} \qquad t \in \mathbb{R}. \tag{1.5.3.3}$$

Es fácil establecer (véase el Ejercicio 1.9.15) que la media de la distribución T_n es 0 y que la varianza es $V[T_n] = \frac{n}{n-2}$ si $n > 2$.

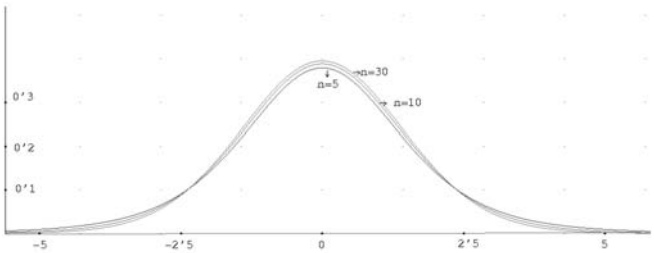


Figura 1.5.3.1. Funciones de densidad de la v.a. de Student para $n = 5, 10, 30$.

En la Figura 1.5.3.1 pueden verse distintas representaciones de la función de densidad (1.5.3.3), al variar el número de grados de libertad n. Al final del libro en el Apéndice titulado Tablas, la Tabla 5 recoge cuantiles útiles de la distribución T_n para diferentes valores de n. A título de ejemplo, para la T_{25} se pueden calcular dos valores, lo más próximos posible, de manera que sea $0'9$ la probabilidad de que la v.a. T_{25} esté entre ellos. Por simetría ha de ser

$$P\{-t < T_{25} < t\} = 0'9$$

o lo que es lo mismo

$$P\{T_{25} \leq t\} = 0'95$$

de donde estos valores son -1′708 y 1′708.

Si en (1.5.3.1) la variable X tiene una distribución $N(\mu, \sigma)$ y las v.a. restantes tienen distribución $N(0, \sigma)$, entonces las v.a. T_n tiene la llamada distribución T_n de Student no central, su función de densidad puede verse en Giri(1993) pág. 202. Además, se observa que el denominador de la expresión que define T_n es la raíz cuadrada de una variable con distribución χ^2_n dividida por su número de grados de libertad, mientras que el numerador tiene distribución $N(0,1)$ e independiente de la variable χ^2_n. Esto permite establecer el siguiente corolario.

Corolario 1.5.3.1
Para m.a.s. de tamaño n de una población $N(\mu, \sigma)$ la distribución de la variable

$$T_{n-1} = \frac{\overline{X} - \mu}{S}\sqrt{n}$$

con S^2 la cuasivarianza muestral, es una distribución de Student con $n-1$ grados de libertad.

Demostración: La variable $\frac{\overline{X}-\mu}{\sigma}\sqrt{n}$ tiene distribución $N(0,1)$ y la variable $(n-1)\frac{S^2}{\sigma^2}$ tiene distribución χ^2_{n-1} como por el teorema de Fisher, ver Teorema 1.5.2.1, ambas son independientes, se sigue el resultado.
■

Una segunda aplicación de la distribución de Student permite encontrar la distribución en el muestreo de ciertos estadísticos en distintas poblaciones normales.

Corolario 1.5.3.2
Si $(X_1, ..., X_m)$ es una m.a.s. de una población $N(\mu_1, \sigma)$ e $(Y_1, ..., Y_n)$ es una m.a.s. de una población $N(\mu_2, \sigma)$ y además son independientes, entonces la distribución de

$$T_{m+n-2} = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{m} + \frac{1}{n}}}, \tag{1.5.3.4}$$

donde

$$S_p^2 = \frac{1}{m+n-2}\left[(m-1)S_1^2 + (n-1)S_2^2\right], \tag{1.5.3.5}$$

recibe el nombre de *cuasivarianza muestral ponderada*, con S_1^2 y S_2^2 las cuasivarianzas muestrales respectivas, y su distribución es una T de Student con $n+m-2$ grados de libertad.

Demostración: La diferencia de medias muestrales $\overline{X}-\overline{Y}$ se distribuye $N(\mu_1-\mu_2,\sigma\sqrt{\frac{1}{m}+\frac{1}{n}})$ tipificando esta variable se obtiene una distribución $N(0,1)$. El Teorema de Fisher y la reproductividad de la distribución χ^2 hace que $(m-1)\frac{S_1^2}{\sigma^2}+(n-1)\frac{S_2^2}{\sigma^2}$ se distribuya como una variable χ^2_{m+n-2}, siendo además independiente de $\overline{X}-\overline{Y}$. Finalmente, de la Definición 1.5.3.1 se sigue el resultado.

■

El razonamiento que se ha seguido es también aplicable aunque las dos varianzas de las poblaciones sean diferentes, en este caso la Expresión (1.5.3.4) depende de los valores σ_1^2 y σ_2^2, hecho que será de capital importancia más adelante, véase la Sección 4.3.

1.5.4. Distribución del cociente de momentos de orden dos respecto al origen.

Dadas dos m.a.s. $(X_1,...,X_m)$ e $(Y_1,...,Y_n)$ independientes y con distribución $N(0,1)$ se trata de calcular la función de densidad del cociente

$$F_{mn}=\frac{\frac{1}{m}\sum_{i=1}^{m}X_i^2}{\frac{1}{m}\sum_{i=1}^{n}Y_i^2}. \tag{1.5.4.1}$$

Como $\sum_{i=1}^{m}X_i^2$ tiene de distribución una χ^2_m y $\sum_{i=1}^{n}Y_i^2$ tiene una distribución χ^2_n, notando por $F_{mn}=U/V$ se tiene que la función de densidad de U es

$$f(u)=\frac{m^{\frac{m}{2}}}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})}u^{\frac{m}{2}-1}e^{-\frac{m}{2}u}\quad\text{para}\quad u\geq 0$$

luego la función de densidad conjunta de U y V, dada su independencia es

$$f(u,v)=ku^{\frac{m}{2}-1}v^{\frac{n}{2}-1}e^{-\frac{1}{2}(mn+nv)}\quad\text{para}\quad u\geq 0, v\geq 0$$

siendo

$$k=\frac{m^{\frac{m}{2}}n^{\frac{n}{2}}}{2^{\frac{m+n}{2}}\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})}$$

haciendo el cambio $S=\frac{U}{V}$ y $T=V$, con lo que $U=ST$ y $V=T$, el jacobiano es T y se sigue que

$$f(s,t)=k(st)^{\frac{m}{2}-1}t^{\frac{n}{2}}e^{-\frac{ms+n}{2}t}$$

con lo que si se utiliza que

$$\int_0^{+\infty}t^{\frac{m+n}{2}-1}e^{-\frac{ms+n}{2}t}dt=\Gamma\left(\frac{m+n}{2}\right)\left(\frac{ms+n}{2}\right)^{-\frac{m+n}{2}}$$

se obtiene que la función de densidad marginal de la variable S es

$$f(s) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} m^{\frac{m}{2}} n^{\frac{n}{2}} s^{\frac{m}{2}-1} (ms+n)^{-\frac{m+n}{2}} \quad \text{para} \quad s>0 \tag{1.5.4.2}$$

Este desarrollo induce la siguiente definición.

Definición 1.5.4.1 (distribución de Snedecor)
Dadas las m.a.s. $(X_1, ..., X_m)$ de una población $N(0,1)$ e $(Y_1, ..., Y_n)$ de una población $N(0,1)$, si ambas muestras son independientes, la v.a.

$$F_{m,n} = \frac{\frac{1}{m}\sum_{i=1}^{m} X_i^2}{\frac{1}{n}\sum_{i=1}^{n} Y_i^2}$$

tiene una distribución F de Snedecor con (m,n) grados de libertad, o de forma equivalente, una v.a. tiene esta distribución cuando su función de densidad es

$$f(x) = \frac{\left(\frac{m}{n}\right)^{\frac{m}{2}} \Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} x^{\frac{m}{2}-1} \left(1+\frac{m}{n}x\right)^{-\frac{m+n}{2}} \quad \text{para} \quad x>0 \tag{1.5.4.2}$$

Es sencillo obtener (véase el Ejercicio 1.9.16) que la media de la v.a. $F_{m,n}$ es $E[F_{m,n}] = n/(n-2)$ si $n>2$ y que la varianza es

$$V[F_{m,n}] = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$$

si $n>4$. En la Figura 1.5.4.1 pueden verse algunas representaciones de la función de densidad de Snedecor con distintos grados de libertad.

De la definición de la v.a. de Snedecor se sigue que si la distribución de X es una $F_{m,n}$, entonces la distribución de X^{-1} es una $F_{n,m}$. Resultado útil a la hora de determinar los cuantiles de la distribución de Snedecor. También es fácil comprobar que si la distribución de la v.a. X es una T_n de Student, la distribución de la v.a. X^2 es la distribución de Snedecor $F_{1,n}$.

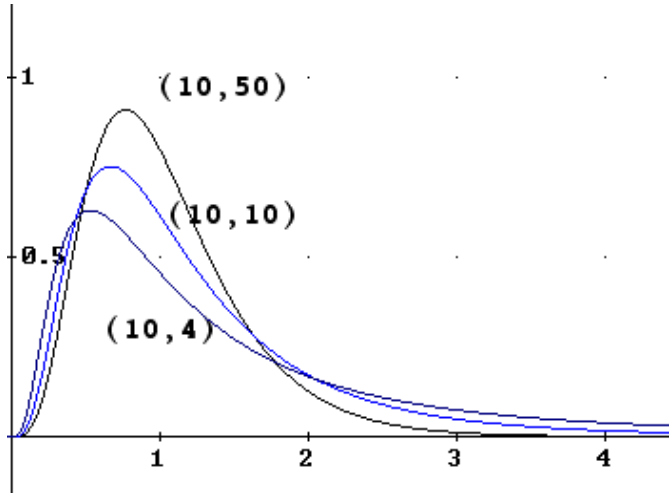


Figura 1.5.4.1. Funciones de densidad de las v.a. $F_{m,n}$ de Snedecor.

Al final del libro, en el Apéndice titulado Tablas, la Tabla 6 recoge los cuantiles de órdenes 0,95 y 0,975 de la distribución $F_{m,n}$ para diferentes valores de m y n. Como utilización de la Tabla, si se quiere calcular para la distribución de Snedecor $F_{5,4}$ la probabilidad

$$P\{F_{5,4} > 9\} = 1 - P\{F_{5,4} \leq 9\}$$

sólo se podría acotar esta probabilidad, de acuerdo con la Figura 1.5.4.2, y se sabría que está entre $0'025 = 1 - 0'95$ y $0'05 = 1 - 0'95$.

Con el concurso de cualquier paquete estándar de estadística se puede aproximar esta probabilidad por $0'0268$.

Si se quiere calcular la probabilidad

$$P\{F_{5,4} \leq 0'16\}$$

se puede utilizar el hecho de que si X tiene distribución de Snedecor $F_{m,n}$ la variable aleatoria X^{-1} tiene de distribución $F_{n,m}$. Por lo tanto

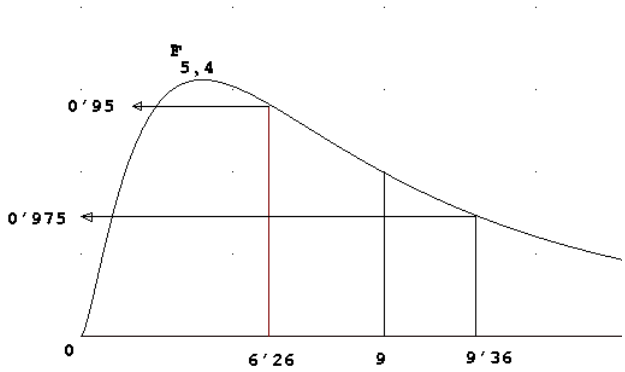


Figura 1.5.4.2. Acotación de la probabilidad $P\{F_{5,4} \leq 9\}$.

$$\begin{aligned} P\{F_{5,4} \leq 0'16\} &= P\{F_{4,5} > 0'16^{-1}\} \\ &= P\{F_{4,5} > 6'25\} \end{aligned}$$

Trabajando como antes, a partir de la Tabla de la $F_{4,5}$ sólo se obtiene que

$$P\{F_{4,5} \leq 5'19\} = 0'95 \quad P\{F_{4,5} \leq 7,39\} = 0'975.$$

Otra vez con el concurso informático puede verse que

$$P\{F_{4,5} \leq 6,25\} = 0'965$$

y la probabilidad deseada es

$$P\{F_{5,4} \leq 0'16\} = 1 - 0'935 = 0'035.$$

La distribución de Snedecor no centrada, se obtiene como el cociente de dos χ^2 independientes entre sus grados de libertad respectivos, y cuando la distribución del numerador únicamente es no centrada. La expresión de su función de densidad puede verse, por ejemplo, en Giri (1993) pág. 203. Se observa que en la definición de la distribución de Snedecor se puede partir de poblaciones $N(0, \sigma)$, pues el valor de la varianza se simplifica al estar en el numerador y el denomindador. Como aplicación, se puede encontrar la distribución del cociente de cuasivarianzas muestrales.

Corolario 1.5.4.1
Si $(X_1, ..., X_m)$ es una m.a.s. de una población $N(\mu_1, \sigma_1)$ e $(Y_1, ..., Y_n)$ es una m.a.s. de una población $N(\mu_2, \sigma_2)$ y ambas muestras son independientes, la v.a.

$$F_{m-1,n-1} = \frac{S_1^2}{S_2^2}\frac{\sigma_2^2}{\sigma_1^2} \tag{1.5.4.3}$$

donde S_1^2 y S_2^2 son las cuasivarianzas muestrales respectivas, tiene una distribución de Snedecor con $(m-1, n-1)$ grados de libertad.

1.6. ESTADÍSTICOS ORDENADOS

Como se ha visto, la función de distribución empírica de la muestra F_n^* converge casi seguro y uniformemente a la función de distribución de la población F. Es intuitivo por tanto, basarse en los estadísticos ordenados $X_{(1)}, ..., X_{(n)}$ para conocer características de ésta distribución. Además ya se ha dicho que existen estadísticos, que son función de los estadísticos muestrales y que dependen poco de observaciones muy grandes o muy pequeñas. Entre los estadísticos muestrales función de los estadísticos ordenados es importante el *recorrido muestral*, $R = X_{(n)} - X_{(1)}$ que es también una medida de la dispersión de la muestra y que por tanto cabe esperar que suministre información de la dispersión de la población. Ya se han definido la *mediana muestral*, véase la Definición 1.3.1.1, y los *cuantiles muestrales*, véase la Definición 1.3.2.1. Como los estadísticos ordenados son funciones de la muestra y los valores muestrales son funciones de la función de distribución de la población, cabe pensar que también los estadísticos muestrales permitan estudiar características de la población.

1.6.1. Distribución marginal del estadístico muestral de orden k

Cuando la distribución de la población es discreta, la distribución del estadístico $X_{(k)}$ puede verse en el Ejercicio 5.9.1. El siguiente teorema estudia el caso en que la distribución de la población sea continua.

Teorema 1.6.1.1
La función de distribución de $X_{(k)}$, el estadístico muestral ordenado de orden k de una población con función de distribución continua $F(x)$ y función de densidad $f(x)$, viene dada por

$$F_{X_{(k)}}(y) = \sum_{i=k}^{n} \binom{n}{i} (F(y))^i (1 - F(y))^{n-i} \tag{1.6.1.1}$$

y su función de densidad

$$f_{X_{(k)}}(y) = \frac{n!}{(k-1)!(n-k)!} (F(y))^{k-1} (1 - F(y))^{n-k} f(y)) \tag{1.6.1.2}$$

Demostración: El suceso $\{X_{(k)} \leq y\}$ es equivalente al suceso {al menos k valores muestrales son menores que y}, de donde se sigue la expresión

(1.6.1.1). La función de densidad se obtiene derivando la función de distribución:

$$\begin{aligned} f_{X_{(k)}(y)} &= \sum_{i=k}^{n} \left[\binom{n}{i} i(F(y))^{i-1} f(y)(1-F(y))^{n-i} - \right. \\ &\quad \left. - \binom{n}{i} (F(y))^{i} (n-i)(1-F(y))^{n-i-1} f(y) \right] \end{aligned}$$

Si se cancelan términos y se tiene en cuenta que para $i = n$ en la Expresión (1.6.1.1) no aparece el término en la potencia $1 - F(y)$, se obtiene el enunciado.

■

Para demostrar el teorema anterior, se puede recurrir al siguiente razonamiento, más directo, que admite una fácil generalización para obtener la densidad conjunta de varios estadísticos ordenados. El razonamiento se basa en la distribución trinomial. De acuerdo con la definición de derivada

$$\begin{aligned} f_{X_{(k)}}(y) &= \lim_{\Delta y \to 0} \frac{F_{X_{(k)}}(y+\Delta y) - F_{X_{(k)}}(y)}{\Delta y} = \\ &= \lim_{\Delta y \to 0} \frac{P\{y < X_{(k)} \leq y + \Delta y\}}{\Delta y} \end{aligned}$$

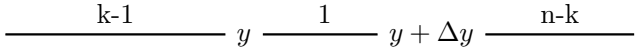


Figura 1.6.1.1. Imagen de la muestra para que el estadístico ordenado sea compatible con $X_{(k)} = y$.

Por lo tanto, para que se dé el suceso $\{X_{(k)} \leq y\}$, de acuerdo con la Figura 1.6.1.1, ha de haber en la muestra $k-1$ observaciones en el intervalo $(-\infty, y]$, una en el intervalo $(y, y+\Delta y)$ y $n-k$ observaciones en el intervalo $(y+\Delta y, +\infty)$. Las probabilidades respectivas de los tres intervalos son $F(y)$, $F(y+\Delta y) - F(y)$ y $1 - F(y+\Delta y)$; luego mediante la distribución trinomial se tiene

$$\begin{aligned} &P\{y \leq X_{(k)} \leq y + \Delta y\} = \\ &= \frac{n!}{(k-1)!1!\,(n-k)!} (F(y))^{k-1} (F(y+\Delta y) - F(y))(1 - F(y+\Delta y))^{n-k} \end{aligned}$$

si se dividen ambos miembros por Δy, se toman límites y se utiliza la continuidad de F, se obtiene que $f_{X_{(k)}}(y)$ puede ponerse como (1.6.1.2).

Teorema 1.6.1.2

Si Z es $X_{(k)}$ el estadístico muestral ordenado de orden k, se nota por z_p el cuantil de orden p de la función de distribución de la variable aleatoria, con

$k = [np]$, y si la función de distribución de la población tiene función de densidad continua y estrictamente positiva en z_p, se tiene que

$$\left(\frac{n}{p(1-p)}\right)^{1/2} f(z_p)(X_{(k)} - z_p) \xrightarrow{\mathcal{L}} Z$$

con Z v.a. $N(0,1)$.

Demostración: Puesto que

$$F_{X_{(k)}}(x) = \sum_{i=k}^{n} \binom{n}{i} (F(x))^j (1 - F(x))^{n-i},$$

si se aproxima la distribución Binomial por la Normal correspondiente, se puede aproximar $F_{X_{(k)}}(x)$ mediante

$$1 - \Phi\left(\frac{k - nF(x)}{\sqrt{nF(x)(1-F(x))}}\right), \tag{1.6.1.3}$$

siendo Φ la función de distribución de la v.a. $N(0,1)$. Por tanto

$$P\left\{\left(\frac{n}{p(1-p)}\right)^{1/2} f(z_p)(X_{(k)} - z_p) \leq \alpha\right\} = P\{X_{(k)} \leq u_p\}$$

donde

$$u_p = \left(x \Big/ \left[\left(\frac{n}{p(1-p)}\right)^{1/2} f(z_p)\right]\right) + z_p$$

y mediante (1.6.1.3)

$$F_{X_{(k)}}(u_p) = 1 - \Phi((p - F(u_p))/[\frac{1}{n}F(u_p)(1 - F(u_p))]^{1/2}).$$

Si se toman límites cuando n tiende a infinito y se elimina la indeterminación por la regla de L'Hôpital, se obtiene que dicho límite vale

$$1 - \Phi\left(\frac{-x(p(1-p))^{1/2}}{f(z_p)} f(z_p)/_{p(1-p)^{1/2}}\right) = 1 - \Phi(-x) = \Phi(x).$$

1.6.2. Distribución conjunta de varios estadísticos ordenados

El razonamiento de la distribución trinomial que se ha seguido para encontrar la distribución marginal del estadístico muestral de orden k, puede generalizarse para hallar la distribución conjunta de cualquier subconjunto de r estadísticos muestrales ordenados

$$X_{(k_1)}, ..., X_{(k_r)} \quad \text{con} \quad 1 \leq k_1 < k_2 < ... < k_r \leq n.$$

La configuración de la muestra puede verse en la Figura 1.6.2.1.

$$\underbrace{\quad}_{k_1-1}\underbrace{y_1}_{k_1}\underbrace{\quad}_{k_2-k_1-1}\underbrace{y_2}_{k_2}\cdots\underbrace{\quad}_{k_{r-1}}\underbrace{y_{r-1}}_{k_{r-1}}\underbrace{\quad}_{k_r-k_{r-1}-1}\underbrace{y_r}_{k_r}\underbrace{\quad}_{n-k_r}$$

Figura 1.6.2.1. Configuración muestral compatible con $X_{(k_1)}=y_1,...,X_{(k_r)}=y_r$.

Si se desprecian los infinitésimos, ya que van a desaparecer al tomar límites, se puede utilizar la distribución multinomial para obtener que la función de densidad conjunta viene dada por

(1.6.2.1)

$$f_{X_{(k_1)}...X_{(k_r)}}(y_1,...,y_r)=\frac{n!}{k_1-1!k_2-k_1-1!...k_r-k_{r-1}-1!n-k_r!}(F(y_1))^{k_1-1}$$

$$(F(y_2)-F(y_1))^{k_2-k_1-1}...(F(y_r)-F(y_{r-1}))^{k_r-k_{r-1}-1}(1-F(y_r))^{n-k_r}$$
$$f(y_1)...f(y_r) \text{ con } -\infty<y_1\leq...\leq y_r<+\infty$$

Esta expresión obviamente se convierte en (1.6.1.2) cuando sólo hay un estadístico ordenado, además toma la forma siguiente cuando el subconjunto incluye a todos los estadísticos ordenados.

(1.6.2.2)

$$f_{X_{(1)},...,X_{(n)}}(y_1,...,y_n)=n!f(y_1)...f(y_n) \text{ para } -\infty<y_1\leq y_2\leq...\leq y_n<+\infty$$

Ejemplo 1.6.2.1

Dada una m.a.s. de una población *Uniforme*$(0,1)$, usando la Expresión (1.6.1.2) se obtiene que la función de densidad del estadístico ordenado muestral de orden k es

$$f_{X_{(k)}}(y)=\frac{n!}{k-1!n-k!}y^{k-1}(1-y)^{n-k} \quad \text{para} \quad y\in(0,1)$$

que corresponde a una distribución $Beta(k,n-k+1)$. Además la función de densidad conjunta del mínimo y del máximo es, según la Expresión (1.6.2.1),

$$f_{X_{(1)}X_{(k)}}(y_1,y_n)=n(n-1)(y_n-y_1)^{n-2} \quad \text{para } 0\leq y_1\leq y_n\leq 1$$

y la función de densidad conjunta de la muestra ordenada es, de acuerdo con la Expresión (1.6.2.2),

$$f_{X_{(1)}...X_{(n)}}(y_1,...,y_n)=n! \quad \text{para} \quad 0\leq y_1\leq...\leq y_n\leq 1.$$

■

La distribución del rango o recorrido muestral, $R = X_{(n)} - X_{(1)}$, puede calcularse a partir de la expresión (1.6.2.1). La densidad conjunta de $X_{(1)}$ y $X_{(n)}$ es

$$f_{X_{(1)}X_{(n)}}(y_1, y_n) = n(n-1)(F(y_n) - F(y_1))^{n-2} f(y_1) f(y_n)$$

para $-\infty < y_1 < y_n < +\infty$. Haciendo el cambio, $R = y_n - y_1$ y $S = y_n$, se obtiene

$$f_{R,S}(r, s) = n(n-1)(F(x) - F(s-r))^{n-2} f(s-r) f(s) \qquad \text{para} \qquad r > 0, s \in \mathbb{R},$$

de donde la distribución marginal del recorrido R es

(1.6.2.3)

$$f_R(r) = n(n-1) \int_{-\infty}^{+\infty} (F(s) - F(s-r))^{n-2} f(s-r) f(s) ds \qquad \text{para} \qquad r > 0.$$

Ejemplo 1.6.2.2

Para m.a.s. de una poblacion *Uniforme*$(0, 1)$, la función de densidad del recorrido muestral $R = X_{(n)} - X_{(1)}$ es

$$\begin{aligned} f_R(r) &= n(n-1) \textstyle\int_r^1 r^{n-2} ds \\ &= n(n-1) r^{n-2} (1-r) \quad \text{para} \quad 0 \leq r \leq 1 \end{aligned}$$

1.6.3. Recubrimientos

Se introduce el concepto de recubrimiento para su utilización en aplicaciones posteriores, relacionadas con la inferencia no paramétrica. La idea es estudiar la probabilidad que la distribución de la población asigna a distintos intervalos, más concretamente. Dado el estadístico muestral ordenado $X_{(1)}, ..., X_{(n)}$, se llaman recubrimientos a los estadísticos C_i, $i = 1, ..., n$ definidos mediante:

$$\begin{aligned} C_1 &= F(X_{(1)}) \\ C_2 &= F(X_{(2)}) - F(X_{(1)}) \\ &\vdots \\ C_n &= F(X_{(n)}) - F(X_{(n-1)}) \end{aligned}$$

Véase la Figura 1.6.3.1. La suma de recubrimientos permite calcular el valor de la función de distribución en los distintos estadísticos ordenados.

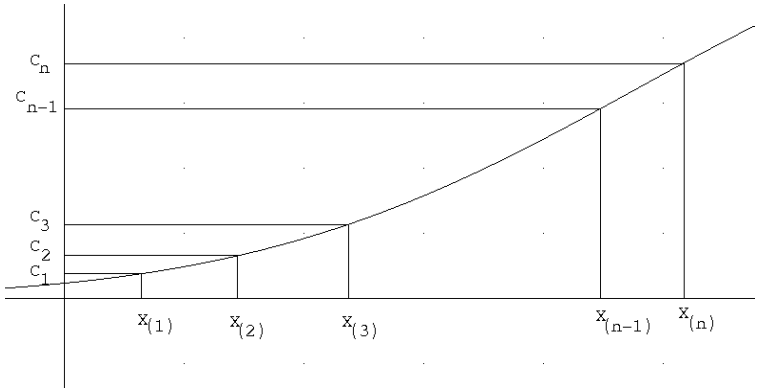


Figura 1.6.3.1. Representación de los recubrimientos.

Es claro que la función de densidad conjunta de $F(X_{(1)}), ..., F(X_{(n)})$, cuando F es continua, es

$$f_{F(X_{(1)}),...,F(X_{(n)})}(u_1, ..., u_n) = n! \quad \text{si} \quad 0 \leq u_1 < u_2 < ... < u_n \leq 1$$

ya que la variable $U = F(X)$ tiene distribución *Uniforme*$(0, 1)$ pues $U = F(X)$ pertenece al intervalo $[0, 1]$ y

$$\begin{aligned} P\{U \leq u\} &= P\{F(X) \leq u\} \\ &= P\{X \leq F^{-1}(u)\} \\ &= FF^{-1}(u) \\ &= u \end{aligned}$$

y la única v.a. que tiene como función de distribución $Q(u) = u$ es la variable *Uniforme*$(0, 1)$. Para el cálculo de la distribución conjunta de los n recubrimientos, como

$$\begin{aligned} F(X_{(1)}) &= C_1 \\ F(X_{(2)}) &= C_1 + C_2 \\ &\vdots \\ F(X_{(n)}) &= C_1 + C_2 + ... + C_n \end{aligned}$$

y el jacobiano de esta transformación vale 1, se sigue que

$$f_{C_1...C_n}(c_1, ..., c_n) = n! \quad \text{si} \quad c_i \geq 0 \ \text{ y } \ \sum_{i=1}^{n} c_i \leq 1. \tag{1.6.3.1}$$

Si ahora se quiere hallar la función de densidad de la suma de un número arbitrario de r recubrimientos $C_{i_1} + ... + C_{i_r}$, debido a la simetría de (1.6.3.1) bastará encontrar la distribución de $C_1 + ... + C_r = F(X_{(r)})$. La función de

densidad $f_{C_{i_1}+...+C_{i_r}}(y)$ coincide con $f_{F(X_r)}(y)$ y, por la expresión (1.6.1.2), se sigue que

$$f_{C_1+...+C_{i_r}}(y) = \frac{n!}{(r-1)!\,(n-r)!} y^{r-1}(1-y)^{n-r} I_{(0,1)}(y).$$

En particular, la función de densidad del recubrimiento C_i es la misma para cualquier i y vale

$$f_{C_i}(y) = n(1-y)^{n-1} I_{(0,1)}(y). \tag{1.6.1.3}$$

Se trata de una distribución $Beta(1,n)$. Como es $E[C_i] = \frac{1}{n+1}$, se concluye que los recubrimientos se reparten con densidad dada por la expresión (1.6.1.3) y su amplitud esperada vale $\frac{1}{n+1}$, lo mismo para todos ellos.

1.7. VARIABLES ALEATORIAS INTERCAMBIABLES

Como ya se ha dicho, la Inferencia Estadística permite modelizar la incertidumbre ayudada por el modelo de probabilidad del que se supone que provienen los datos que se están analizando. Como consecuencia de este proceso hay que determinar a partir de los datos disponibles el valor de ciertos parámetros, que van a caracterizar la población con la que se está trabajando. Más adelante, Sección 8.2.1. se aprenderá a verificar si el modelo que se ha supuesto puede considerarse como correcto, o si por el contrario debería ser cambiado. Se puede pensar que la manera de ajustar el modelo, es siempre a través de una muestra aleatoria simple, sin embargo existen otros tipos de muestras que permiten modelizar la incertidumbre que aparece en los experimentos con los que trata la Estadística. Al estudio de tales nociones se va a dedicar esta sección. Supóngase para fijar ideas que se observa repetidas veces una población X v.a. discreta. Si la muestra es aleatoria simple de tamaño n, se tiene que para cualquier $m < n$

$$\begin{aligned} P\{X_{m+1} = x_{m+1}, ..., X_n = x_n | X_1 = x_1, ..., X_m = x_m\} = \\ = P\{X_{m+1} = x_{m+1}, ..., X_n = x_n\} \end{aligned} \tag{1.7.1}$$

ya que por ser independientes las componentes de la muestra

$$P\{X_{m+1} = x_{m+1}, ..., X_n = x_n, X_1 = x_1, ..., X_m = x_m\} = \prod_{i=1}^{n} P\{X = x_i\}$$

y

$$P\{X_1 = x_1, ..., X_m = x_m\} = \prod_{i=1}^{m} P\{X = x_i\}.$$

Sin embargo (1.7.1) no parece intuitivo, ya que dice que no se aprende de la experiencia pasada hasta m. La noción de aleatoriedad parece mejor capturada si se dice que $P\{X_1 = x_1, ..., X_n = x_n\}$ no depende de como se ordenen las n variables aleatorias. Esta idea, más bien de simetría que de dependencia, es la que pretende recoger la siguiente definición.

Definición 1.7.1 (v.a. intercambiables)
La colección de v.a. $X_1, ..., X_n$ se dice intercambiable si su distribución conjunta verifica

$$P\{X_1 \leq x_1, ..., X_n \leq x_n\} = P\{X_1 \leq x_{r(1)}, ..., X_n \leq x_{r(n)}\}$$

para cualquier permutación de los n índices. Esto equivale, si existe función de densidad, a que

$$f(x_1, ..., x_n) = f(x_{r(1)}, ..., x_{r(n)})$$

y, si existe función de masa, a que

$$p(x_1, ..., x_n) = p(x_{r(1)}, ..., x_{r(n)}).$$

Ejemplo 1.7.1 (tiradas de una chincheta)
Se considera una muestra, en las mismas condiciones siempre, de tiradas de una chincheta, y se hace $X_i = 1$ si en la tirada i-ésima la punta queda arriba y $X_i = 0$ en caso contrario. En esta situación el orden es irrelevante y la sucesión de unos y ceros $x_1, ..., x_n$ es intercambiable en el sentido de la Definición 1.7.1. Si por el contrario se tiran alternativamente una chincheta y una moneda y si al tirar la moneda se hace $X_i = 1$ si sale cara y $X_i = 0$ en caso contrario, la sucesión $x_1....x_{2n}$ de unos y ceros no será intercambiable, y sí que lo será la sucesión de las tiradas impares (correspondiente a la chincheta) entre sí y la sucesión de las tiradas pares (correspondientes a la moneda).

■

Desde luego es evidente que si la muestra $(X_1, ..., X_n)$ es aleatoria simple, la colección de las n variables aleatorias es intercambiable.

$$f(x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i)$$

$$f(x_{r(1)}, ..., x_{r(n)}) = \prod_{i=1}^{n} f(x_{r(i)}) = \prod_{i=1}^{n} f(x_i)$$

ya que todas las densidades son iguales a la densidad de la población. El concepto de intercambiabilidad es más débil, por tanto, que el concepto de independencia. Ambos conceptos no son equivalentes, como pone de manifiesto el siguiente contraejemplo.

Ejemplo 1.7.2
Supongamos tres v.a. X_1, X_2, X_3 cada una de las cuales puede tomar el valor 0 o el valor 1 con función de masa definida mediante

$$\begin{aligned} P\{X_1 = 0, X_2 = 1, X_3 = 1\} &= P\{X_1 = 1, X_2 = 0, X_3 = 1\} \\ &= P\{X_1 = 1, X_2 = 1, X_3 = 0\} \\ &= \tfrac{1}{3}. \end{aligned}$$

Cualquier otra combinación de valores tiene probabilidad cero, tanto de las tres en bloque como de dos o de una de las variables. Es evidente que la colección $\{X_1, X_2, X_3\}$ es intercambiable pero las variables no son independientes, ya que

$$P\{X_1 = 0, X_2 = 1, X_3 = 1\} = \frac{1}{3}$$

pero

$$P\{X_1 = 0\}P\{X_2 = 1\}P\{X_3 = 1\} = 0.$$

Es más, puede probarse que la sucesión no se puede extender a una cuarta v.a. X_4 que tome el valor cero o uno, de manera que la colección $\{X_1, X_2, X_3, X_4\}$ sea intercambiable, véase el Ejercicio 1.9.21.

■

El desarrollo en profundidad del concepto de intercambiabilidad se sale fuera del ámbito de un curso introductorio como éste, pero se recoge el siguiente teorema, sin demostración, para que se tenga un mejor conocimiento de su significado.

Teorema 1.7.1 (representación de v.a. intercambiables 0-1)
Dada una colección de v.a. intercambiables que toman los valores 0 ó 1, existe una función de distribución $Q(\theta)$ tal que la funcion de masa $P\{X_1 = x_1, ..., X_n = x_n\}$ puede ponerse como

$$P\{X_1 = x_1, ..., X_n = x_n\} = \int_0^1 \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} dQ(\theta). \tag{1.7.2}$$

La demostración puede verse en Sudderth (1976).

■

Interesa enfatizar el significado intuitivo del teorema, que está diciendo que si la muestra es de v.a. intercambiables con valores 0 ó 1, se puede modelizar de manera que las v.a. $X_i|\theta$ sean independientes y con distribución de $Bernoulli(\theta)$; y el valor de θ se puede suponer obtenido mediante una función de distribución, lógicamente en el intervalo $[0, 1]$. Es decir que si en una determinada situación práctica, existe incertidumbre sobre variables aleatorias que toman valores 0 ó 1 y las variables son intercambiables, entonces se

puede suponer que se está trabajando con una distribución de probabilidad de la forma

$$P\{X_1 = x_1, ..., X_n = x_n\} = \int_0^1 \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}\pi(\theta)d(\theta),$$

siendo $\pi(\theta)$ una densidad en el intervalo $[0,1]$, es decir, una función positiva tal que

$$\int_0^1 \pi(\theta)d\theta = 1.$$

La noción de v.a. intercambiables, fué introducida por De Finetti (1972) y es central en la aproximación a la inferencia que recibe el nombre de inferencia bayesiana. El siguiente ejemplo, pone de manifiesto la diferencia entre independencia e intercambiabilidad en el contexto de v.a. cero-uno.

Ejemplo 1.7.3
La colección de v.a. $X_i|\theta$ son i.i.d. con distribución $X_i|\theta \sim Bernoulli(\theta)$ y θ es una v.a. con distribución uniforme en el intervalo $(0,1)$ $\theta \sim U(0,1)$. Las *v.a.* $X_1, ..., X_n$ son intercambiables y sin embargo no son independientes, con lo que la colección $(X_1, ..., X_n)$ no es una m.a.s.

En efecto, por el teorema de la probabilidad total en su versión continua

$$P\{X_1 = x_1, ..., X_n = x_n\} = \int_0^1 P\{X_1 = x_1, ..., X_n = x_n|\theta\}d\theta$$

por ser $X_i|\theta$ independientes y con distribución de $Bernoulli(\theta)$

$$P\{X_1 = x_1, ..., X_n = x_n\} = \int_0^1 \theta^{\sum_1^n x_i}(1-\theta)^{n-\sum_1^n x_i}d\theta.$$

Finalmente, por la definición de la función $Beta$

$$\begin{aligned}\int_0^1 \theta^{\sum_1^n x_i}(1-\theta)^{n-\sum_1^n x_i}d\theta &= Beta(\sum_1^n x_i + 1, n - \sum_1^n x_i + 1)\\ &= \frac{t!(n-t)!}{(t-1)!}\end{aligned}$$

siendo $t = \sum_{i=1}^{n} x_i$. Por lo tanto las v.a. $X_1, ..., X_n$ son intercambiables.

Por otro lado

$$\begin{aligned}P\{X_i = x_i\} &= \int_0^1 \theta^{x_i}(1-\theta)^{1-x_i}d\theta\\ &= \frac{x_i!(1-x_i)!}{2}\end{aligned}$$

y evidentemente

$$\frac{t!(n-t)!}{(t+1)!} \neq \prod_{i=1}^{n} \frac{x_i!(1-x_i)!}{2}$$

luego las variables X_i no son independientes.

1.8. APROXIMACIÓN HISTÓRICA

Inicialmente la utilidad de la estadística era únicamente descriptiva. Su objetivo era ordenar datos y representarlos mediante tablas y gráficos. Quizás uno de los primeros en usar representaciones de distribuciones de frecuencias en un par de ejes coordenados fué D'Alembert (1728-1777). La utilización sistemática de estas representaciones de datos en tablas de frecuencias, hubo de esperar hasta Quetelet (1796-1874), a quien también se le debe el término momento.

La representación gráfica en forma de nubes de puntos se debe a los estudios relacionados con la astronomía. El primer recurso a observaciones numerosas de una misma magnitud, aparece en la obra de Tycho Brahe (1546-1601), quien al estudiar un conjunto de datos sobre el movimiento de los planetas recurre por primera vez a la media aritmética para eliminar los errores de las observaciones. No todos los investigadores se mostraron partidarios de esta manera de proceder. Uno de los defensores de la utilización de la media aritmética, y que calculó su distribución bajo normalidad, fue Simpson en 1755. En esta misma época es introducida la mediana por Boscovich, su contribución puede verse en Sheynin (1973). La desviación típica fué introducida por Karl Pearson (1857-1936) en 1892 y la varianza, aunque como momento ya era conocida, no fué utilizada con tal nombre hata 1918 por Fisher (1890-1962), quien la introduce en un trabajo sobre genética.

En cuanto a la historia de la utilización de los estadísticos ordenados, es Laplace (1749-1827) quien, en 1818, en su *Teoría Analítica de las Probabilidades* las emplea por primera vez, al relacionarlos con la estimación de parámetros en un modelo lineal. Galton en 1875 estudia la distribución de la mediana muestral bajo normalidad. Edgeworth obtiene la distribución de la mediana cuando la población tiene colas más altas que la distribución normal. Así, estos estudios se encauzan hacia la robustez, en el sentido de utilizar características muestrales cuya distribución con el muestreo no se vea demasiado afectada si dejan de cumplirse las suposiciones de normalidad. Comentarios relacionados con estos aspectos pueden verse en el capítulo dedicado a la inferencia no paramétrica.

La distribución χ^2 fué introducida por Karl Pearson alrededor de 1900, al estudiar el problema de ajuste entre una distribución poblacional y unos datos muestrales. En 1904, la generalización de este problema al caso en que existan parámetros desconocidos en la distribución, dió lugar a que William Sealy Gosset (1876-1937) introdujera la distribución de Student, así llamada por ser Student el pseudónimo utilizado por Gosset para comunicar los resultados de sus investigaciones con muestras pequeñas, pues la empresa de elaboración de cerveza en la que trabajaba impedía la divulgación de los resultados obtenidos por sus empleados.

Snedecor (1881-1974), estadístico estadounidense, introdujo la distribu-

ción que lleva su nombre alrededor de 1907. Snedecor trabajó en el laboratorio de estadística del *University College* fundado por Karl Pearson, junto a Student y Fisher. En agradecimiento a la importante aportación de Fisher a la Inferencia Estadística, Snedecor le dedicó la distribución del cociente de distribuciones χ^2 y la denotó mediante la letra F.

1.9. EJERCICIOS

1.9.1

Sea X una población de $Bernoulli(1/2)$ y se consideran todas las m.a.s. posibles de tamaño 3. Para cada muestra calcúlense $\overline{X}$ y S^2, la media y la cuasivarianza muestrales y determínense sus distribuciones en el muestreo.

1.9.2

De una población con media θ desconocida y varianza 1, se toma una m.a.s. de tamaño n. ¿Cuál debe ser éste para que la media muestral diste en valor absoluto de la media de la población menos que $0'5$, con una probabilidad mayor o igual que $0'95$?

1.9.3

Para los datos del Ejemplo 1.3.2

a) Constrúyase un histograma de frecuencias.
b) Constrúyase un diagrama de tallos y hojas.
c) Obténgase la media y la mediana.
d) Determínese el recorrido intercuartílico.

1.9.4

Los caudales horarios en m^3 por segundo de dos crecidas del río Ebro en la localidad de Sobrón han sido:

crecida 1

28	29	29	29	29	30	31	32	33	34	38	41
49	64	80	107	139	153	168	183	194	205	220	232
243	258	270	280	290	301	318	324	333	343	354	374
397	406	418	424	424	412	391					

crecida 2

26	27	28	29	31	35	40	46	52	58	68	70
71	73	74	75	76	78	79	81	84	88	92	100
105	109	110	107	104	100	98	95	89	85	80	77
73	70	68	68	66	64	62	61	61	60	60	60

Constrúyase el diagrama de caja de cada crecida.

1.9.5

Sean $(X_1, ..., X_n)$ e $(Y_1, ..., Y_n)$ dos m.a.s. independientes de dos poblaciones $N(0, 4)$ y $N(1, 3)$. Calcúlese $P\{\overline{X} > \overline{Y}\}$.

1.9.6
Dada una m.a.s. de tamaño n, calcúlese la distribución de la media muestral $\overline{X}$, cuando la población es:

a) de Bernoulli.
b) Gamma.
c) Exponencial.
d) De Cauchy.

1.9.7
Sea $(X_1, ..., X_n)$ una m.a.s. de una población $N(\mu, \sigma)$ y X_{n+1} una v.a. $N(\mu, \sigma)$ independiente de la muestra anterior. Calcúlese la distribución de la v.a.

$$Y = \sqrt{\frac{n}{n+1}} \frac{X_{n+1} - \overline{X}}{S},$$

siendo S^2 la cuasivarianza muestral.

1.9.8
Una colección consiste en cuatro elementos numerados $1, 2, 3$ y 4. Se extraen dos elementos sin reemplazamiento y se nota por (X_1, X_2) los valores obtenidos. Se pide:

a) Distribución conjunta de $(X_{(1)}, X_{(2)})$.
b) Distribución de la media muestral $\overline{X}$.

1.9.9
Sea $(X_1, ..., X_n)$ una m.a.s. de una población de $Bernoulli(\theta)$. Obténgase la distribución en el muestreo de la cuasivarianza muestral S^2.

1.9.10
Demuéstrese que dada una m.a.s. $(X_1, ..., X_n)$ de una población $N(\mu, \sigma)$, las v.a. $\overline{X}$ y $X_i - \overline{X}$ son independientes para todo i.

1.9.11
Sea $(X_1, ..., X_6)$ una m.a.s. de una población $N(0, 1)$. Se definen las v.a.

$$Y_1 = (X_1 + X_2 + X_3)^2 + (X_4 + X_5 + X_6)^2$$
$$Y_2 = \frac{X_1 + X_2}{\sqrt{X_3^2 + X_4^2 + X_5^2}}$$

Determínense dos valores k_1 y k_2 tales que

$$P\{Y_1 > k_1\} = P\{Y_2 > k_2\} = 0'1.$$

1.9.12
Dada una m.a.s. $(X_1, ..., X_n)$ de una población $N(\mu, \sigma)$ se construyen

$$\overline{X}_k = \tfrac{1}{k}\sum_{i=1}^{k} X_i \qquad \overline{X}_{n-k} = \tfrac{1}{n-k}\sum_{i=k+1}^{n} X_i$$

$$S_k^2 = \tfrac{1}{k-1}\sum_{i=1}^{k}(X_i - \overline{X}_k)^2 \quad S_{n-k-1}^2 = \tfrac{1}{n-k-1}\sum_{i=k+1}^{n}(X_i - \overline{X}_{n-k})^2$$

Calcúlese la distribución de las v.a.:
a) $\left[(k-1)S_k^2 + (n-k-1)S_{n-k-1}^2\right]/\sigma^2$
b) S_k^2/S_{n-k-1}^2

1.9.13
Dada una sucesión $\{X_n\}$ de v.a. independientes con distribución $N(0,1)$ y un entero positivo k, se define

$$F_{k,m} = \left(\frac{1}{k}\sum_{i=1}^{k} X_i^2\right) \Bigg/ \left(\frac{1}{m}\sum_{i=k+1}^{k+m} X_i^2\right).$$

Pruébese que

$$F_{k,m} \xrightarrow[m \to \infty]{\mathcal{L}} \frac{1}{k}X$$

donde $X \sim \chi_k^2$.

1.9.14
Pruébese que para todo $x \in \mathbb{R}$ la función de distribución empírica $F_n^*(x)$ converge en probabilidad a la función de distribución de la población $F(x)$.

1.9.15
Dadas dos m.a.s. independientes $(X_1, ..., X_m)$ de una población $N(\mu_1, \sigma)$ e $(Y_1, ..., Y_n)$ de una población $N(\mu_2, \sigma)$ respectivamente y dos números reales α y ρ, hállese la distribución de la v.a.

$$Z = \frac{\alpha(\overline{X} - \mu_1) + \beta(\overline{Y} - \mu_2)}{S_p\sqrt{\frac{\alpha^2}{m} + \frac{\beta^2}{n}}},$$

donde

$$S_p^2 = \frac{1}{m+n-2}((m-1)S_1^2 + (n-1)S_2^2)$$

y siendo S_1^2 y S_2^2 las cuasivarianzas muestrales.

1.9.16
Demuéstrese que si X tiene distribución χ_n^2, entonces

$$E[X] = n \qquad V[X] = 2n.$$

1.9.17
Demuéstrese que si X tiene distribución de Student T_n con n grados de libertad, entonces si $n > 2$

$$E[X] = 0 \qquad V[X] = \tfrac{n}{n-2}.$$

1.9.18
Demuéstrese que si X es una v.a. con distribución de Snedecor $F_{m,n}$, entonces si $n > 4$

$$E[X] = \tfrac{n}{n-2} \qquad V[X] = \tfrac{2n^2(m+n-2)}{m(n-2)^2(n-4)}.$$

1.9.19
Demuéstrese que para una m.a.s. de tamaño n de una población con distribución discreta dada por la función de masa

$$P\{X = x_i\} = p_i \qquad i = 1, 2, ...$$

la distribución del estadístico ordenado $X_{(k)}$ es discreta y viene dada por

$$P\{X_{(1)} \leq x_i\} = \sum_{j=k}^{n} \binom{n}{j} (F(x_i))^j (1 - F(x_i))^{n-j}$$

$$P\{X_{(k)}{=}x_i\} = \sum_{j=k}^{n} \binom{n}{j} \left[(F(x_i))^j (1 - F(x_i))^{n-j} - (F(x_{i-1}))^j (1 - F(x_{i-1}))^{n-j} \right].$$

1.9.20
La duración media de una m.a.s. de 10 bombillas de una población de desviación típica 425 horas, fué de 1327 horas. Una m.a.s. independiente de la anterior de tamaño 6 de una población con desviación típica de 375 horas, arrojó una duración media muestral de 1215 horas. Si las medias de las dos poblaciones se suponen iguales, ¿qué probabilidad tiene el obtener una desviación en las medias muestrales menor que la que se ha obtenido?

1.9.21
Dadas las v.a. X_1, X_2 y X_3 que toman los valores cero o uno, se define la distribución de probabilidad

$$P\{X_i = x_i\} = 0 \quad \text{si} \quad x_i = 0, 1 \quad i = 1, 2, 3$$

$$P\{X_i = x_i, X_j = x_j\} = 0 \quad \text{si} \quad x_i, x_j = 0, 1 \quad i, j = 1, 2, 3$$

$$\begin{aligned} P\{X_1 = 0, X_2 = 1, X_3 = 1\} &= P\{X_1 = 1, X_2 = 0, X_3 = 1\} \\ &= P\{X_1 = 1, X_2 = 1, X_3 = 0\} \\ &= \tfrac{1}{3}. \end{aligned}$$

Demuéstrese que la distribución de probabilidad no puede extenderse a una cuarta v.a. X_4 que tome los valores cero o uno, de manera que la colección X_1, X_2, X_3, X_4 sea intercambiable.

1.10. SOLUCIÓN A LOS EJERCICIOS

1.10.1

La distribución muestral queda recogida en la siguiente tabla

muestras			$\overline{x}$	$s^2 = \frac{1}{2}\sum x_i - \frac{3}{2}\overline{x}^2$	P
1	1	1	1	0	1/8
1	1	0	2/3	1/3	1/8
1	0	1	2/3	1/3	1/8
0	1	1	2/3	1/3	1/8
1	0	0	1/3	1/3	1/8
0	1	0	1/3	1/3	1/8
0	0	1	1/3	1/3	1/8
0	0	0	0	0	1/8

por lo tanto

$\bar{x}$	0	1/3	2/3	1
P	1/8	3/8	3/8	1/8

s^2	0	1/3
P	1/4	3/4

1.10.2

De acuerdo con la desigualdad de Tchebycheff

$$P\{|X - E[X] \geq kD[X]\} \leq \frac{1}{k^2}$$

si se aplica con $\overline{X}, E[\overline{X}] = \theta$ y $D[\overline{X}] = \frac{1}{\sqrt{n}}$ se pretende que

$$P\{|\overline{X} - \theta| \leq 0'5\} \geq 0'95$$

o lo que es lo mismo

$$P\{|\overline{X} - \theta| > 0'5\} \leq 0'05$$

de donde $\frac{1}{k^2} = 0'05$ y $\frac{k}{\sqrt{n}} = 0'5$ de manera que n ha de ser por lo menos 80.

1.10.3

a)

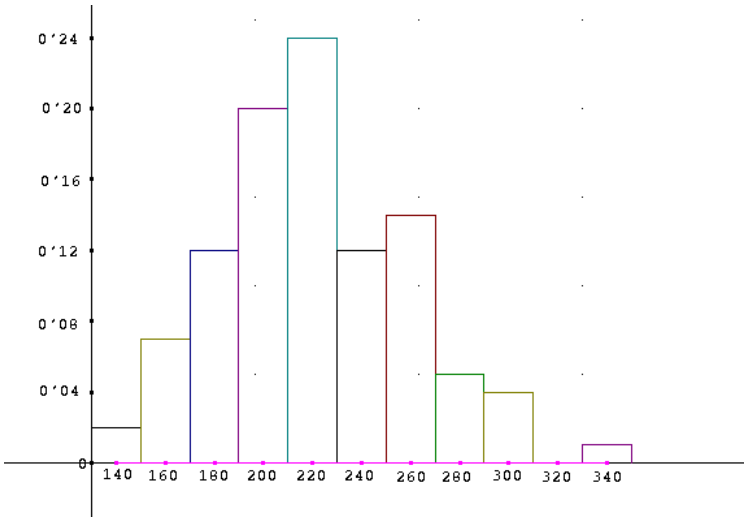


Figura 1.10.1. Histograma correspondiente al Ejercicio 1.10.3.

b) Si se toman dos cifras significativas para los *tallos,* se obtiene el siguiente *diagrama de tallos y hojas*

3	14	5 5 6
5	15	8 8
10	16	3 8 8 8 8
19	17	2 4 5 5 5 5 5 5 8
26	18	0 0 0 1 1 7 7
36	19	2 4 4 5 6 6 8 8 8 8
53	20	1 1 1 4 4 4 5 5 5 5 5 6 6 6 7 8 9
(71)	21	1 1 1 2 2 2 4 4 4 6 7 7 7 7 8 8 8 8
64	22	1 1 1 1 3 4 5 5 7 7 7 7 8
51	23	0 5 5 5 6 7 8 8 9
42	24	1 2 2 3 3 5 5 7 8
33	25	2 3 3 3 4 4 5 6 6 7 7
22	26	1 2 6 7 7 8 8 8 8
13	27	3 4 8
10	28	3 4 5 9
6	29	2 4 6
3	30	0 1
1	34	5

Se puede repetir, redondeando los datos a dos cifras, la de orden superior para los *tallos* y la de orden inferior para las *hojas.*

c) La media $\overline{x} = 135$ y la mediana, al ser un número impar de observaciones, viene dada por $x_{(68)} = 218$.

d) El primer cuartil es $x_{(34)} = 198$ y el tercer cuartil $x_{(102)} = 248$. Por lo que el RIM= 50.

1.10.4

Para construir el *diagrama de cajas* correspondiente a la *crecida 1*, de acuerdo con la Sección 1.3.3, el cuartil de orden $1/4$ es $x_{(12)} = 41$, el cuartil de orden $3/4$ es $x_{(34)} = 343$ y la mediana es $x_{(24)} = 232$. El primer cuartil menos $1'5$ RIM es -412 luego por ese lado no hay observaciones atípicas y la recta horizontal izquierda se situa en $x_{(1)} = 28$, de forma análoga $343 + 1'5\text{RIM}$ es 795 luego la recta horizontal derecha se coloca en $x_{(45)} = 424$. Procediendo de manera análoga con la *crecida 2* el cuartil de orden $1/4$ es $1/2(y_{(12)} + y_{(13)}) = 1/2(60 + 60) = 60$ el cuartil de orden $3/4$ es $1/2(y_{(16)} + y_{(97)}) = 1/2(85 + 88)$ y la mediana es $1/2(y_{(24)}+y_{(25)}) = 1/2(71+73) = 72$. La cota inferior es $60-1'5\cdot 26'5 = 20'25$ luego el valor mínimo del diagrama es $y_{(1)} = 26$ y la cota superior es $86'5+1'5\cdot 26'5 = 126'25$ luego el valor máximo del diagrama es $y_{(48)} = 110$, en esta crecida tampoco hay observaciones atípicas.

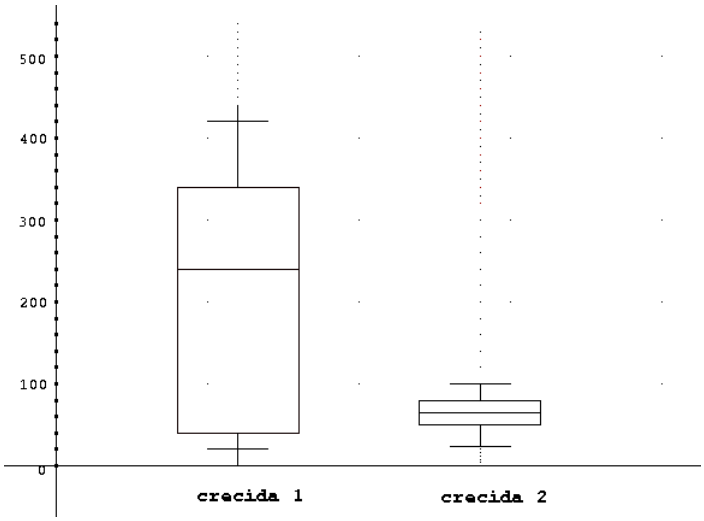


Figura 1.10.2. Representacion de los diagramas de cajas del Ejercicio 1.10.4.

Si se representan los dos diagramas se obtiene la Figura 1.10.2 en la que se observa que la *crecida 1* es asimétrica con más valores por debajo de la mediana que por encima, mientras que los valores de la *crecida 2* son más homogéneos y simétricos.

1.10.5

Como $\overline{X} - \overline{Y} \sim N\left(-1, \sqrt{\frac{16}{25} + \frac{9}{25}}\right) = N(-1, 1)$ se sigue que

$$P\{\overline{X} > \overline{Y}\} = P\{\overline{X} - \overline{Y} > 0\} = P\{\overline{X} - \overline{Y} + 1 > 1\} = P\{Z > 1\}$$

donde $Z \sim N(0,1)$, y de las tablas de la distribución normal se obtiene que dicha probabilidad es $0'1587$.

1.10.6

a) Para la distribución de $Bernoulli(\theta)$, como $\sum_{i=1}^{n} X_i \sim Binomial(n,\theta)$,

$$P\{\overline{X} = t|\theta\} = P\{\frac{1}{n}\sum_{i=1}^{n} X_i = t|\theta\} = P\{\sum_{i=1}^{n} X_i = nt\}$$
$$= \binom{n}{nt}\theta^t(1-\theta)^{n(1-t)}, \quad t = 0, \frac{1}{n}, ..., 1.$$

b) Mediante la función característica

$$\varphi_X(t) = \left(1 - \frac{it}{a}\right)^{-p}$$

se sigue que

$$\varphi_{\sum_{i=1}^{n} X_i/n}(t) = E[e^{i\frac{t}{n}\sum_{i=1}^{n} X_i}] = E[e^{i\frac{t}{n}X_1}]...E[e^{i\frac{t}{n}X_n}]$$
$$= \left[\varphi\left(\frac{t}{n}\right)\right]^n = \left(1 - \frac{it}{an}\right)^{-np}$$

que es la fución característica de la distribución $Gamma(na, np)$ y por tanto ésta es la distribución de $\bar{X}$.

c) Es un caso particular del anterior ya que $X \sim Exponencial(\theta) = Gamma(\theta, 1)$ luego $\bar{X} \sim Gamma(n\theta, n)$.

d) Si $X \sim Cauchy(\theta, \sigma)$, es decir

$$f(x|\theta,\sigma) = \frac{1}{\pi\sigma}\left[1 + \left(\frac{x-\theta}{\sigma}\right)^2\right]^{-1},$$

con función característica $\varphi_X(t) = e^{it\theta}e^{-|t|\sigma}$, entonces

$$\varphi_{\overline{X}}(t) = \left[\varphi_X\left(\frac{t}{n}\right)\right]^n = e^{it\theta}e^{-|t|\sigma}$$

luego curiosamente, $\bar{X} \sim Cauchy(\theta, \sigma)$.

1.10.7

La distribución de $X_{n+1} - \overline{X} \sim N(0, \sigma\sqrt{\frac{n+1}{n}})$ luego la variable tipificada $\frac{X_{n+1}-\overline{X}}{\sigma}\sqrt{\frac{n}{n+1}} \sim N(0,1)$. Por otro lado de acuerdo con el teorema de Fisher, teorema 1.4.2.1, $(n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1}$, de donde al ser esta v.a. independiente de

la anterior, mediante la definición de la distribución de Student, definición 1.4.3.1, se tiene

$$\frac{X_{n+1}-\overline{X}}{S}\sqrt{\frac{n}{n+1}} \sim T_{n-1}.$$

1.10.8

a) Mediante enumeración se sigue que $P\{X_1 = x_1, X_2 = x_2\} = \frac{1}{12}$ y por lo tanto $P\{X_{(1)} = x_{(1)}, X_{(2)} = x_{(2)}\} = \frac{1}{6}$ ya que hay dos parejas (x_1, x_2) y (x_2, x_1) que se aplican en el mismo par $(x_{(1)}, x_{(2)})$.

$(x_{(1)}, x_{(2)})$	P
(1,2)	1/6
(1,3)	1/6
(1,4)	1/6
(2,3)	1/6
(2,4)	1/6
(3,4)	1/6

b) Para obtener la distribución de la media muestral, trabajando con cada par $(x_{(1)}, x_{(2)})$ se obtiene

$\bar{x}$	1'5	2	2'5	3	3'5
P	1/6	1/6	2/6	1/6	1/6

1.10.9

Al desarrollar el cuadrado $(X_i - \overline{X})^2 = X_i^2 - 2X_i\overline{X} + \overline{X}^2$ se obtiene fácilmente que

$$S^2 = \frac{\sum_{i=1}^{n} X_i(n - \sum_{i=1}^{n} X_i)}{(n-1)n}$$

como $\sum_{i=1}^{n} X_i \sim Binomial(n, \theta)$, en general si $\sum_{i=1}^{n} X_i = k$ y $\sum_{i=1}^{n} X_i = n - k$ dan la misma imagen por S^2, $\frac{k(n-k)}{(n-1)n}$ salvo que n sea par, que en ese caso solo hay un valor $k = \frac{n}{2}$ que se transforma en $\frac{n}{4(n-1)}$. Por lo tanto $S^2 = \frac{k(n-k)}{(n-1)n}$ con probabilidades $p_k + p_{n-k}$ si n es impar y caso contrario vale la misma expresión, salvo para $k = \frac{n}{2}$ que $S^2 = \frac{n}{4(n-1)}$ con probabilidad $p_{\frac{n}{2}}$. Donde $p_k = \binom{n}{k}\theta^k(1-\theta)^{n-k}, k = 0, 1, ..., [n/2]$.

1.10.10

Como ambas variables son normales, basta ver que son incorreladas.

$$Cov(\overline{X}, X_i - \overline{X}) = E[\overline{X}(X_i - \overline{X})]$$

ya que

$$E[X_i - \overline{X}] = E\left[-\frac{1}{n}X_1 - ... + (1 - \frac{1}{n})X_i - ... - \frac{1}{n}X_n\right]$$
$$= -\frac{n-1}{n}\mu + \left(1 - \frac{1}{n}\right)\mu = 0,$$

pero

$$E[\overline{X}X_i - \overline{X}^2] = \mu^2 + \frac{\sigma^2}{n} - \left(\frac{\sigma^2}{n} + \mu^2\right) = 0$$

por ser

$$E\left[\left(\frac{1}{n}X_1 + ... + \frac{1}{n}X_i + ... + \frac{1}{n}X_n\right)X_i\right] = \frac{n-1}{n}E[X_1X_i] + \frac{1}{n}E[X_i^2]$$
$$= \frac{n-1}{n}\mu^2 + \frac{1}{n}(\sigma^2 + \mu^2)$$
$$= \mu^2 + \frac{\sigma^2}{n}$$

y la expresión $E[\overline{X}^2] = \frac{\sigma^2}{n} + \mu^2$ puede obtenerse de $V[\overline{X}] = E[\overline{X}^2] - (E[\overline{X}])^2$, luego $Cov(\overline{X}, X_i - \overline{X}) = 0$.

1.10.11

Se tiene que $X_1 + X_2 + X_3 \sim N(0, \sqrt{3})$ y $X_4 + X_5 + X_6 \sim N(0, \sqrt{3})$ con ambas independientes, luego

$$P\{Y_1 > k_1\} = P\left\{\frac{Y_1}{3} > \frac{k_1}{3}\right\}$$
$$= P\left\{\left(\frac{X_1 + X_2 + X_3}{\sqrt{3}}\right)^2 + \left(\frac{X_4 + X_5 + X_6}{\sqrt{3}}\right)^2 > \frac{k_1}{3}\right\}$$
$$= P\left\{\chi_2^2 > \frac{k_1}{3}\right\} = 0'1,$$

de donde $\frac{k_1}{3} = 4'6052$ y por tanto $k_1 = 13'816$.

$$P\{Y_2 > k_2\} = P\left\{\frac{(X_1 + X_2)/\sqrt{2}}{\sqrt{X_3^2 + X_4^2 + X_5^2/3}} > \sqrt{\frac{3}{2}}k_2\right\}$$
$$= P\left\{T_3 > \sqrt{\frac{3}{2}}k_2\right\},$$

en virtud de la definición de la distribución de Student, Definición 1.4.3.1. De las Tablas se sigue que $\sqrt{\frac{3}{2}}k_2 = 1'638$ y por tanto $k_2 = 1'337$.

1.10.12

a) La expresión dada puede ponerse como

$$(k-1)\frac{S_k^2}{\sigma^2}+(n-k-1)\frac{S_{n-k-1}^2}{\sigma^2}$$

cada sumando es una χ_{k-1}^2 y χ_{n-k-1}^2 ambas independientes, luego la suma se distribuye χ_{n-2}^2.

b) Si se observa que

$$\frac{S_k^2}{S_{n-k-1}^2}=\frac{\frac{1}{\sigma^2}\frac{S_k^2}{k-1}}{\frac{S_{n-k-1}^2}{n-k-1}}\frac{k-1}{n-k-1}$$

se tiene que el primer factor tiene distribución de Snedeckor $F_{k-1,n-k-1}$, por lo tanto $\frac{S_k^2}{S_{n-k-1}^2}\sim\frac{k-1}{n-k-1}F_{k-1,n-k-1}$.

1.10.13

Por la ley fuerte de los grandes números, $\frac{1}{m}\sum_{i=k+1}^{k+m}X_i^2$ converge casi seguro a $E[X_i^2]=1$. Si se aplica el teorema de Slutsky, Teorema 1.3.5, el cociente converge en ley a donde converja en ley $\frac{1}{k}\sum_{i=1}^{k}X_i^2$; pero al ser esta expresión constante como función de m y con distribución $\frac{1}{k}\chi_k^2$ se obtiene el resultado.

1.10.14

Basta observar que $F_n^*(x)=\frac{1}{n}\sum_{i=1}^{n}I_{(-\infty,x]}(x_i)=\frac{1}{n}\sum_{i=1}^{n}Y_i$, donde $Y_i=I_{(-\infty,x]}(x_i)$. Por la ley fuerte de los grandes números $F_n^*(x)$ converge casi seguro a $E[Y_i]=P\{X_i\leq x\}=F(x)$, en particular en probabilidad.

1.10.15

Como $\overline{X}\sim N(\mu_1,\frac{\sigma}{\sqrt{m}})$, $\overline{Y}\sim N(\mu_2,\frac{\sigma}{\sqrt{n}})$ se sigue que

$$\frac{\alpha(\overline{X}-\mu_1)+\beta(\overline{Y}-\mu_2)}{\sigma\sqrt{\frac{\alpha^2}{m}+\frac{\beta^2}{n}}}\sim N(0,1).$$

Por otro lado $(m+n-2)\frac{S_p^2}{\sigma^2}=(m-1)\frac{S_1^2}{\sigma^2}+(n-1)\frac{S_2^2}{\sigma^2}$ y estos dos sumandos tienen sendas distribuciones χ^2 e independientes luego su suma tiene distribución χ_{m+n-2}^2. Numerador y denominador son v.a. independientes luego de acuerdo con la definición de la distribución de Student, Definición 1.4.3.1, el cociente $Z\sim T_{m+n-2}$.

1.10.16

Como la densidad de la v.a. χ_n^2 es

$$f_{\chi_n^2}(x)=\frac{1}{2^{\frac{m}{n}}\Gamma(\frac{n}{2})}e^{-\frac{1}{2}x}x^{\frac{n}{2}-1}\quad\text{con}\quad x>0$$

se tiene que

$$E[X] = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} \int_0^{+\infty} x e^{-\frac{1}{2}x} x^{\frac{n}{2}-1} dx = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} \int_0^{+\infty} e^{-\frac{1}{2}x} x^{\frac{n}{2}} dx$$

como

$$\int_0^{+\infty} e^{-ax} x^{p-1} dx = a^{-p}\Gamma(p), \quad a > 0, p > 0$$

la integral tiene esta forma particular con $a = \frac{1}{2}$ $p = \frac{n}{2} + 1$, de donde

$$E[X] = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} \left(\frac{1}{2}\right)^{-\frac{n}{2}-1} \Gamma\left(\frac{n}{2}+1\right).$$

Si se utiliza la expresión recursiva para la función gamma

$$\Gamma(p) = (p-1)\Gamma(p-1)$$

se obtiene que

$$E[X] = n.$$

Para el cálculo de la varianza, el mismo tipo de razonamiento lleva a

$$E[X^2] = 2n\left(\frac{n}{2}+1\right)$$

esta vez con $a = \frac{1}{2}$ y $p = \frac{n}{2}+2$. La sustitución en la expresión de la varianza hace que

$$\begin{aligned} V[X] &= E[X^2] - (E[X])^2 = 2n(\frac{n}{2}+1) - n^2 \\ &= 2n. \end{aligned}$$

1.10.17

Como la densidad de la v.a. de Student con n grados de libertad es

$$f_{T_n}(x) = \frac{1}{\sqrt{n}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma(\frac{1}{2})\Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \quad \text{con} \quad x \in \mathbb{R}$$

se tiene que

$$E[X] = \frac{1}{\sqrt{n}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)} \int_{-\infty}^{+\infty} x(1 + \frac{x^2}{n})^{-\frac{n+1}{2}} dx$$

y por ser el integrando función impar $(f(-x) = -f(x))$ la integral es nula.

Para el cálculo de la varianza

$$\begin{aligned} V[X] = E[X^2] &= \frac{1}{\sqrt{n}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{n}{2})} \int_{-\infty}^{+\infty} x^2(1 + \frac{x^2}{n})^{-\frac{n+1}{2}} dx = \\ &= \frac{1}{\sqrt{n}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)} 2 \int_0^{+\infty} x^2(1 + \frac{x^2}{n})^{-\frac{n+1}{2}} dx \end{aligned}$$

si se hace el cambio $y = \left(1 + \frac{x^2}{n}\right)^{-1}$, la expresión anterior toma la forma

$$V[X] = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)} n \int_0^1 y^{\frac{n}{2}-2}(1-y)^{\frac{1}{2}}\, dy$$

como

$$\int_0^1 x^{p-1}(1-x)^{q-1}dx = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} \quad , \quad p>0 \quad q>0$$

la integral es de esta forma con $p = \frac{n}{2} - 1 \qquad q = \frac{3}{2}$ y simplificando se obtiene

$$V[X] = \frac{n}{n-2}.$$

1.10.18

Como la densidad de la v.a. de Snedecor es

$$f_{F_{m,n}}(x) = \left(\frac{m}{n}\right)^{\frac{m}{2}} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} x^{\frac{m}{2}-1}\left(1+\frac{m}{n}x\right)^{-\frac{m+n}{2}} \qquad \text{con} \qquad x>0$$

se sigue que

$$E[X] = \left(\frac{m}{n}\right)^{\frac{m}{2}} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \int_0^{+\infty} x^{\frac{m}{2}}(1+\frac{m}{n}x)^{-\frac{m+n}{2}}dx$$

esta expresión se puede reducir a la función beta, si se hace el cambio

$$1 - y = \left(1 + \frac{m}{n}x\right)^{-1}$$

con lo que $dx = \frac{n}{m}(1-y)^{-2}dy$ y se obtiene

$$\begin{aligned} E[X] &= \left(\frac{m}{n}\right)^{\frac{m}{2}} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \int_0^1 \left(\frac{n}{m}\right)^{\frac{m}{2}} y^{\frac{m}{2}}(1-y)^{-\frac{m}{2}}(1-y)^{\frac{m+n}{2}}\frac{n}{m}(1-y)^{-2}dy \\ &= \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\frac{n}{m}\int_0^1 y^{\frac{m}{2}}(1-y)^{\frac{n}{2}-2}dy \end{aligned}$$

si se utiliza la definición de la función beta con $p = \frac{m}{2}+1$ y $q = \frac{n}{2}-1$ se obtiene que

$$E[X] = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\frac{n}{m}\frac{\Gamma\left(\frac{m}{2}+1\right)\Gamma\left(\frac{n}{2}-1\right)}{\Gamma\left(\frac{m+n}{2}\right)}$$

y mediante la expresión recursiva de la función Gamma, se llega a que

$$E[X] = \frac{n}{n-2}$$

expresión que curiosamente no depende de m.

Para la obtención de la varianza de la distribución de Snedecor, se necesita calcular

$$E[X^2] = \left(\frac{m}{n}\right)^{\frac{m}{2}} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \int_0^{+\infty} x^{\frac{m}{2}+1} \left(1+\frac{m}{n}x\right)^{-\frac{m+n}{2}} dx$$

el mismo cambio y la definición de la función beta con $p = \frac{m}{n} + 2$ y $q = \frac{n}{2} - 2$, lleva a que

$$\begin{aligned} E[X^2] &= \frac{\Gamma\left(\frac{m}{2}+2\right)\Gamma\left(\frac{n}{2}-2\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{n}{m}\right)^2 \\ &= \frac{\left(\frac{m}{2}+1\right)\frac{m}{2}}{\left(\frac{n}{2}-1\right)\left(\frac{n}{2}-2\right)} \frac{n^2}{m^2} = \frac{(m+2)n^2}{(n-2)(n-4)m} \end{aligned}$$

de donde la varianza es

$$\begin{aligned} V[X^2] &= \frac{(m+2)n^2}{m(n-2)(n-4)} - \frac{n^2}{(n-2)^2} \\ &= \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}. \end{aligned}$$

1.10.19

El suceso $\{X_{(k)} \leq x_i\}$ es equivalente al suceso {en la muestra haya k o más valores menores o iguales que x_i} y este número sigue una distribución $Binomial(n, F(x_i))$, de donde

$$P\{X_{(k)} \leq x_i\} = \sum_{j=k}^{n} \binom{n}{j} (F(x_i))^j (1-F(x_i))^{n-j}.$$

El valor de la función de masa se obtiene ahora de considerar que

$$P\{X_{(k)} = x_i\} = P\{X_{(k)} \leq x_i\} - P\{X_{(k)} \leq x_{i-1}\}.$$

1.10.20

Parece razonable hacer la suposición de que la duración de una bombilla tiene una distribución $N(\mu, \sigma)$, de donde

$$\overline{X}_{10} \sim N(\mu, \frac{\sigma_1}{\sqrt{10}}) \qquad \overline{Y}_6 \sim N(\mu, \frac{\sigma_2}{\sqrt{6}})$$

por lo tanto

$$\overline{X}_{10} - \overline{Y}_6 \sim N(0, \sqrt{\frac{\sigma_1^2}{10} + \frac{\sigma_2^2}{6}} = N(0, 203'71)$$

y se necesita calcular

$$\begin{aligned} P\{\overline{X}_{10} - \overline{Y}_6 \leq 112\} &= P\left\{\frac{\overline{X}_{10} - \overline{Y}_6}{203'71} \leq \frac{112}{203'71}\right\} \\ &= P\{Z \leq 0'55\} \end{aligned}$$

donde $Z \sim N(0,1)$. Por tanto la probabilidad pedida es aproximadamente de $0'71$.

1.10.21

Si se pudiera extender, debería cumplirse que

$$\begin{aligned} P\{X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0\} = \\ = P\{X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 1\} \end{aligned}$$

pero, por las propiedades de la probabilidad, el primer miembro debe valer

$$\begin{aligned} &P\{X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0\} = \\ &\quad = P\{X_1 = 0, X_2 = 1, X_3 = 1\} - P\{X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 1\} \\ &\quad = \frac{1}{3} - P\{X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 0\} \end{aligned}$$

y esta última probabilidad es igual a cero, ya que

$$\begin{aligned} P\{X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0\} \\ \leq P\{X_1 = 1, X_2 = 1, X_3 = 1\} = 0 \end{aligned}$$

de donde se deduce que

$$P\{X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0\} = \frac{1}{3}.$$

Sin embargo si se parte de

$$P\{X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 1\} \leq P\{X_1 = 0, X_2 = 0, X_3 = 1\} = 0$$

entonces ha de ser $P\{X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 1\} = 0$ y no puede darse por tanto la intercambiabilidad en las cuatro variables.

Capítulo 2

REDUCCIÓN DE DATOS: ESTADÍSTICOS SUFICIENTES, ANCILARIOS, COMPLETOS E INVARIANTES

2.1. INTRODUCCIÓN

Hasta ahora se han estudiado procedimientos para determinar la distribución en el muestreo de diversos estadísticos. En un problema paramétrico de estimación, se está interesado en utilizar la información en sentido amplio, para a apartir de una muestra aleatoria simple de una población X con función de distribución perteneciente a una familia paramétrica, conjeturar el valor del parámetro θ desconocido con el que se distribuye la muestra. A tal fin lo primero que se hace es mostrar cómo se resume la información disponible mediante los estadísticos suficientes.

2.2. Estadísticos suficientes y minimal suficientes

Se empieza introduciendo el concepto de estadístico suficiente con un ejemplo. Con m.a.s. de tamaño n de una población de $Bernoulli(\theta)$, la

función de masa de la muestra es

$$P(x_1,...,x_n|\theta) = \theta^{\sum_1^n x_i}(1-\theta)^{n-\sum_1^n x_i}, \quad x_i = 0,1 \text{ para todo } i, \theta \in [0,1].$$

Parece intuitivo, que para determinar θ no es necesario saber las posiciones en que se han producido los éxitos. Bastará saber cuál es el número total $t = \sum_1^n x_i$.

La función de masa puede ponerse como

$$P(x_1,...,x_n|\theta) = \binom{n}{t}\theta^t(1-\theta)^{n-t}\frac{1}{\binom{n}{t}}$$

donde lo importante es que ha podido factorizarse en una parte, que depende de la muestra a través del estadístico $T = \sum_1^n X_i$ y que es la única que depende de θ, y en $\binom{n}{t}^{-1}$ que no depende de θ. Se puede pensar entonces en la muestra como obtenida en dos etapas consecutivas, en la primera se observa $T = t$ con probabilidad $\binom{n}{t}\theta^t(1-\theta)^{n-t}$ y en la segunda se sortean las posiciones en las que van a colocarse los t éxitos, con lo que cada muestra compatible con el valor de $T = t$ tendrá de probabilidad $\binom{n}{t}^{-1}$.

Es claro entonces que el estadístico $T = \sum_1^n X_i$ resume toda la información que respecto a θ nos suministra la muestra, ya que conocido el valor del estadístico se puede construir la muestra mediante un mecanismo independiente del valor de θ.

Definición 2.2.1 (estadístico suficiente)
$T(X_1,...,X_n)$ es un estadístico suficiente para la familia de funciones de distribución $\mathcal{P} = \{F(\overrightarrow{x}|\theta)|\theta \in \Theta\}$ cuando la distribución de probabilidad de la muestra condicionada por el estadístico es independiente de θ.

■

El problema con esta definición es su poca operatividad, ya que se necesita tener previamente el candidato a estadístico suficiente, para después comprobar si lo es. Por este motivo se utiliza el siguiente teorema debido a Halmos y Savage.

Teorema 2.2.1 (factorización)
Un estadístico $T(X_1,...,X_n)$ es suficiente, si existen funciones reales positivas h y g tales que

$$f(x_1,...,x_n|\theta) = h(x_1,...,x_n)g_\theta(T(x_1,...,x_n))$$

donde f es la función de masa o de densidad de la muestra.

Se recoge la demostración en el caso discreto, la demostración general puede verse en Lehmann (1986) páginas 54 y 55.

Si se tiene la factorización basta poner

$$f(x_1, ..., x_n|t) = \frac{f(x_1, ..., x_n|\theta)}{f(t|\theta)}$$

si $T(x_1, ..., x_n) = t$ y cero en caso contrario. Cuando no es cero, sustituyendo queda

$$f(x_1, ..., x_n|t) = \frac{g_\theta(t)h(x_1, ..., x_n)}{\sum\limits_{\overline{y}|T(\overline{y})=t} g_\theta(t)h(y_1, ..., y_n)}$$

simplificando $g_\theta(t)$ se obtiene

$$f(x_1, ..., x_n|t) = \frac{h(x_1, ..., x_n)}{\sum\limits_{\overline{y}|T(\overline{y})=t} h(y_1, ..., y_n)}$$

en las dos opciones cero o esta cociente, son independientes de θ.

Recíprocamente, si $f(x_1, ..., x_n|t)$ no depende de θ, como

$$f(x_1, ..., x_n|\theta) = f(t|\theta)f(x_1, ..., x_n|t, \theta) \tag{2.2.1}$$

se puede tomar

$$g_\theta(t) = f(t|\theta) \qquad \text{y} \qquad h(x_1, ..., x_n) = f(x_1, ..., x_n|t)$$

con lo que se obtiene la descomposición deseada.

Ejemplo 2.2.1

Para encontrar un estadístico suficiente, con m.a.s. de tamaño n de una población de Poisson de parámetro θ, $Poisson(\theta)$, se tiene que

$$f(x_1, ..., x_n|\theta) = e^{-n\theta}\theta^{\sum\limits_1^n x_i} / \prod_{i=1}^n x_i!$$

con lo que

$$T = \sum_{i=1}^n X_i, \quad g_\theta(t) = e^{-n\theta}\theta^{\sum\limits_{i=1}^n X_i} \quad \text{y} \quad h(x_1, ..., x_n) = \left(\prod_1^n x_i!\right)^{-1}.$$

■

Se observa que siempre la propia muestra y el estadístico ordenado $(X_{(1)}, ..., X_{(n)})$ son estadísticos suficientes.

Además si el estadístico T es suficiente para $\mathcal{P}$ cualquier biyección $U = U(T)$ también lo es, pues de acuerdo con el teorema de factorización, si la población tiene función de densidad

$$\begin{aligned} f(x_1, ..., x_n|\theta) &= h(x_1, ..., x_n) g_\theta(T(x_1, ..., x_n)) \\ &= h(x_1, ..., x_n) g_\theta(U) \end{aligned}$$

luego, se obtiene la factorización deseada también en función de U.

Si $S(X_1, ..., X_n)$ es un estadístico suficiente para $\mathcal{P}$ y $T(X_1, ..., X_n)$ es otro estadístico, también es suficiente el estadístico $(S(X_1, ..., X_n), T(X_1, ..., X_n))$ ya que

$$f(x_1, ..., x_n|s, t) = \frac{f(x_1, ..., x_n, s, t|\theta)}{f(s, t|\theta)}$$

ahora bien está última expresión es igual a $\frac{f(x_1,...,x_n,s|\theta)}{f(s|\theta)f(t|\theta)}$ que coincide con $f(x_1, ..., x_n|s)(f(t|s))^{-1}$ donde este último factor no depende de θ para cualquier estadístico T ya que

$$f(t|s) = \sum_{\overline{x}|T(\overline{x})=t} f(x_1, ..., x_n|s).$$

Este último enunciado es por otra parte muy intuitivo, si con S se tiene recogida información suficiente para θ, al añadir T no se gana información, de todas formas es claro que la suficiencia ha sido introducida con el objetivo contrario de resumir la información en lugar de aumentarla.

Una definición más rigurosa de estadístico suficiente puede verse en Garcia Nogales (1991) pág. 102.

Ejemplo 2.2.2
Con m.a.s. de tamaño n de una población $N(\theta, \sigma_0)$ con σ_0 conocida, el estadístico $T(X_1, ..., X_n) = \overline{X}$ es suficiente. Ya que puede escribirse

$$f(x_1, ..., x_n|\theta) = \left(\frac{1}{\sqrt{2\pi\sigma_0}}\right)^n e^{-\frac{1}{2\sigma_0^2}\sum_1^n x_i^2} e^{\frac{n\theta}{\sigma_0^2}(\overline{x}-\frac{\theta}{2})}$$

con lo que

$$h(x_1, ..., x_n) = \left(\frac{1}{\sqrt{2\pi\sigma_0}}\right)^n e^{-\frac{1}{2\sigma_0^2}\sum_1^n x_i^2} \quad \text{y} \quad g_\theta(T(X_1, ..., X_n)) = e^{\frac{n\theta}{\sigma_0^2}(\overline{x}-\frac{\theta}{2})}.$$

Ejemplo 2.2.3
Con m.a.s. de tamaño n de una población $N(\mu_0, \sqrt{\theta})$ con μ_0 conocida, el estadístico $T(X_1, ..., X_n) = \sum_1^n (X_i - \mu_0)^2$ es suficiente, pues

$$f(x_1, ..., x_n|\theta) = \left(\frac{1}{\sqrt{2\pi\theta}}\right)^n e^{-\frac{1}{2\theta}\sum_{i=1}^n (x_i-\mu_0)^2}$$

con lo que

$$h(x_1,...,x_n)=1 \qquad \text{y} \qquad g_\theta(T(x_1,...,x_n))=\left(\frac{1}{\sqrt{2\pi\theta}}\right)e^{-\frac{1}{2\theta}\sum_{i=1}^{n}(x_i-\mu_0)^2}.$$

Ejemplo 2.2.4
Con m.a.s. de tamaño n de una población $N(\mu,\sigma)$, siendo μ y σ desconocidas, puede comprobarse que el estadístico $\overrightarrow{T}(X_1,...,X_n)=\left(\sum_1^n X_i,\sum_1^n X_i^2\right)$ es suficiente.

Ejemplo 2.2.5
Para m.a.s. de tamaño n de una población $U(-\frac{\theta}{2},\frac{\theta}{2})$, compruébese que $\overrightarrow{T}(X_1,...,X_n)=(X_{(1)},X_{(n)})$ es suficiente para θ. Si la población es $U(0,\theta)$, también son suficientes los estadísticos $(X_{(1)},X_{(n)})$ y $X_{(n)}$.

■

Esto lleva al concepto de estadístico minimal suficiente, tratando de encontrar el estadístico suficiente que más resuma la información. Para introducir este concepto es conveniente cambiar el punto de vista desde el que se considera un estadístico. Cualquier estadístico determina una partición del espacio muestral y recíprocamente, dada una partición del espacio muestral, asignándole un valor a cada clase se tiene definido un estadístico.

De esta manera se hablará de partición suficiente, como la asociada a un estadístico suficiente y se introduce la siguiente definición.

Definición 2.2.2 (estadístico minimal suficiente)
Un estadístico T es minimal suficiente si su partición asociada es suficiente y es la menos fina entre todos los estadísticos suficientes S.

Se entiende que la partición asociada a un estadístico T es menos fina que la asociada a otro estadístico S si toda clase de la partición asociada a T es unión de clases de la partición asociada a S.

Teorema 2.2.2 (caracterización de estadísticos minimal suficientes)
La relación de equivalencia que coloca en la misma clase dos puntos muestrales $\overrightarrow{x}=(x_1,...,x_n)$ e $\overrightarrow{y}=(y_1,...,y_n)$ si y sólo si el cociente $f(\overrightarrow{x}|\theta)/f(\overrightarrow{y}|\theta)$ no depende de θ, induce una partición minimal suficiente y por tanto determina un estadístico minimal suficiente.

Demostración: es inmediato ver que la relación definida es de equivalencia, luego si se asigna a cada clase del conjunto cociente un valor del estadístico, queda definido un estadístico suficiente. En efecto si se nota por A_t a la clase en la que $T(\overrightarrow{x})=t$, y $\overrightarrow{x}\in A_t$ se sigue que

$$f(\overrightarrow{x}|t,\theta)=\frac{f(\overrightarrow{x}|\theta)}{f(t|\theta)}=\frac{f(\overrightarrow{x}|\theta)}{\sum_{\overrightarrow{y}\in A_t}f(\overrightarrow{y}|\theta)}=\sum_{\overrightarrow{y}\in A_t}\left(\frac{f(\overrightarrow{y}|\theta)}{f(\overrightarrow{x}|\theta)}\right)^{-1}$$

expresión que es independiente de θ.

Sea cualquier otro estadístico S suficiente, se va a comprobar que si $S(\overrightarrow{x}) = S(\overrightarrow{y}) = s$, es decir $\overrightarrow{x}$ e $\overrightarrow{y}$ están en la misma clase de S, entonces $T(\overrightarrow{x}) = T(\overrightarrow{y})$, con lo que la partición asociada a T, es menos fina que la correspondiente a S. En efecto

$$\frac{f(\overrightarrow{x}|\theta)}{f(\overrightarrow{y}|\theta)} = \frac{\frac{f(\overrightarrow{x}|\theta)}{f(s|\theta)}}{\frac{f(\overrightarrow{y}|\theta)}{f(s|\theta)}} = \frac{f(\overrightarrow{x}|s)}{f(\overrightarrow{y}|s)}$$

donde s es el valor de la clase por S que contiene a $\overrightarrow{x}$ e $\overrightarrow{y}$. Por ser S un estadístico suficiente, numerador y denominador de esta expresión no dependen de θ luego $\overrightarrow{x}$ e $\overrightarrow{y}$ están en la misma clase de la partición asociada a T.

■

Ejemplo 2.2.6

Para m.a.s. de tamaño n de una población de Bernoulli(θ), $\sum_{i=1}^{n} X_i$ es un estadístico minimal suficiente, ya que

$$\frac{f(x_1, ..., x_n|\theta)}{f(y_1, ..., y_n|\theta)} = \theta^{\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i}(1-\theta)^{\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i}$$

y este cociente es independiente de θ cuando $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$.

Teorema 2.2.3

Si la distribución de la muestra es de la familia exponencial k-paramétrica, es decir

$$f(x_1, ..., x_n|\overrightarrow{\theta}) = c(\overrightarrow{\theta})h(\overrightarrow{x})e^{\sum_{i=1}^{k} c_i(\overrightarrow{\theta})T_i(\overrightarrow{x})}$$

y existen valores $\overrightarrow{\theta}^1, ..., \overrightarrow{\theta}^k$ tales que los vectores $\overrightarrow{c}_i = (c_1(\overrightarrow{\theta}^i), ..., c_k(\overrightarrow{\theta}^i))$, $i = 1, ..., k$ son linealmente independientes, entonces el estadístico $\overrightarrow{T}(\overrightarrow{X}) = (T_1(\overrightarrow{X}), ..., T_k(\overrightarrow{X}))$ es minimal suficiente.

Demostración: Basta observar que

$$\frac{f(\overrightarrow{x}|\overrightarrow{\theta})}{f(\overrightarrow{y}|\overrightarrow{\theta})} = \frac{h(\overrightarrow{x})}{h(\overrightarrow{y})}e^{\sum_{1}^{k} c_i(\overrightarrow{\theta})[T_i(\overrightarrow{x}) - T_i(\overrightarrow{y})]}$$

y este cociente es independiente de $\overrightarrow{\theta}$ si y sólo si $\sum_{1}^{k} c_i(\overrightarrow{\theta})[T_i(\overrightarrow{x}) - T_i(\overrightarrow{y})] = 0$ con lo que el sistema homogéneo

$$\sum_{i=1}^{k} c_i(\overrightarrow{\theta}^j)[T_i(\overrightarrow{x}) - T_i(\overrightarrow{y})] = 0 \qquad j = 1, ..., k$$

solo admite la solución trivial $T_i(\overrightarrow{x})-T_i(\overrightarrow{y})=0,\ i=1,...,k$ y el estadístico $\overrightarrow{T}(\overrightarrow{X})=(T_1(\overrightarrow{X}),...,T_k(\overrightarrow{X}))$ es minimal suficiente.

■

Ejemplo 2.2.7
Para una población $N(\mu,\sigma)$ con $\overrightarrow{\theta}=(\mu,\sigma)$ es

$$f(x_1,...,x_n|\overrightarrow{\theta})=\left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n e^{\frac{\mu}{\sigma^2}\sum_{i=1}^{n}x_i-\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2}$$

luego el estadístico $\overrightarrow{T}(X_1,...,X_n)=\left(\sum X_i,\sum X_i^2\right)$ es minimal suficiente, ya que en $\left(\frac{\mu}{\sigma^2},-\frac{1}{2\sigma^2}\right)$ existen $\overrightarrow{\theta}^1=(\mu=0,\sigma^2=1)$ y $\overrightarrow{\theta}^2=(\mu=1,\sigma^2=1)$ con $\overrightarrow{c}_1=\left(0,-\frac{1}{2}\right)$ y $\overrightarrow{c}_2=(1,-\frac{1}{2})$ linealmente independientes.

2.3. ESTADÍSTICOS ANCILARIOS Y COMPLETOS

Intuitivamente el concepto de estadístico ancilario es opuesto al concepto de estadístico suficiente, es decir un estadístico ancilario es el que no da información sobre el parámetro. Posteriormente se verá que las cosas no son tan estrictas, pero la idea puede ser válida como aproximación intuitiva.

Definición 2.3.1 (estadístico ancilario)
El estadístico $U(X_1,...,X_n)$ es ancilario para θ, si su distribución es independiente de θ.

Ejemplo 2.3.1
Para una población $N(\theta,\sigma_0)$ con σ_0 conocida, con m.a.s. de tamaño n se tiene que $U(\vec{X})=X_1-X_2$ tiene de distribución $N(0,\sqrt{2\sigma_0})$, que no depende de θ, por lo que U es un estadístico ancilario para θ.

■

Puesto que la distribución de un estadístico ancilario no depende del parámetro, parece natural esperar que un estadístico ancilario para θ sea independiente de cualquier otro estadístico suficiente, que ésto es cierto pero añadiendo una propiedad más, la de completitud, ha sido probado por Basu (1955). Por lo tanto, antes de demostrar este resultado se pasa a desarrollar la noción de completitud.

Definición 2.3.2 (familia de distribuciones completa)
La familia $\mathcal{P}=\{f(\overrightarrow{x}|\theta)|\theta\in\Theta\}$ es completa si para cualquier función real $g(\overrightarrow{x})$ tal que $E_\theta g(\overrightarrow{X})=0$ para todo $\theta\in\Theta$ se sigue que $g(\overrightarrow{X})$ es nula casi seguro respecto de θ, es decir, $g(\overrightarrow{x})$ es cero salvo en un conjunto de probabilidad cero, calculada a partir de $f(\overrightarrow{x}|\theta)$ para todo θ.

Ejemplo 2.3.2
La familia $\mathcal{P} = \{\theta^x(1-\theta)^{1-x}; x \in \{0,1\}|\theta \in (0,1)\}$ es completa. Ya que si

$$E_\theta g(X) = 0 \quad \text{para todo } \theta \in (0,1)$$

ha de ser $g(1)\theta + g(0)(1-\theta) = 0$ para todo $\theta \in (0,1)$, luego $g(0) = g(1) = 0$.

También la familia $\mathcal{P} = \{Binomial(n,\theta)|\theta \in (0,1)\}$ es completa. Ya que si

$$E_\theta g(X) = 0 \text{ para todo } \theta \in (0,1)$$

ha de ser

$$g(0)(1-\theta)^n + g(1)\binom{n}{1}\theta(1-\theta)^{n-1} + ... + g(i)\binom{n}{i}\theta^i(1-\theta)^{n-i} + ... + g(n)\theta^n$$
$$= (1-\theta)^n \sum_{i=0}^{n} g(i)\binom{n}{i}(\theta/(1-\theta))^i = 0,$$

que es un polinomio de grado n en $\theta/(1-\theta)$, luego para que sea nulo ha de ser $g(0) = ... = g(n) = 0$.

Ejemplo 2.3.3
La familia $\mathcal{P} = \{N(0,\theta)|\theta \text{ desviación típica} \in \mathbb{R}^+\}$ no es completa, ya que $g(X) = X$ es tal que $E_\theta g(X) = 0$ para todo $\theta \in \mathbb{R}^+$.

Definición 2.3.3 (estadístico completo)
El estadístico $T(X_1, ..., X_n)$ es completo si su familia de distribuciones de probabilidad $\mathcal{P} = \{f(t|\theta)|\theta \in \Theta\}$ es completa.

■

Es importante resaltar, que se debe trabajar con la familia de distribuciones del estadístico obtenida a partir de la distribución de la muestra.

Ejemplo 2.3.4
Para muestras de tamaño uno de una población $N(0,\theta)$ con θ la desviación típica, el estadístico $T(X) = X^2$ es completo. Ya que su familia de distribuciones es

$$\mathcal{P} = \left\{ \frac{\theta^{-\frac{1}{2}}}{\sqrt{2}\sqrt{\pi}} e^{-\frac{1}{2\theta^2}t} t^{-\frac{1}{2}} \; ; \; t \in \mathbb{R} \;\middle|\; \theta \in \mathbb{R}^+ \right\}$$

y si

$$\int_0^{+\infty} g(t) e^{-\frac{1}{2\theta^2}t} t^{-\frac{1}{2}} dt = 0 \quad \forall \theta \in \mathbb{R}^+$$

ha de ser $g(t) = 0$. Esto es intuitivo al variar θ, de todas formas un razonamiento riguroso se basa en la unicidad de la transformada de Laplace.

Ejemplo 2.3.5
Para m.a.s. de tamaño n de una población X normal de media θ y de desviación típica θ, con $\theta \in (0, +\infty)$ el estadístico $\overrightarrow{T}(X_1, ..., X_n) = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2 \right)$ no es completo ya que

$$2\left(\sum_{i=1}^{n} X_i \right)^2 - (n+1)\left(\sum_{i=1}^{n} X_i^2 \right)$$

es un estimador insesgado de cero. Para comprobarlo basta utilizar que $E\left[\left(\sum_{i=1}^{n} X_i \right)^2 \right] = n\theta^2(1+n)$, pues $\frac{\left(\sum_{i=1}^{n} X_i - n\theta \right)^2}{n\theta^2}$ se distribuye según una distribución χ_1^2, y que $E\left[\left(\sum_{i=1}^{n} X_i^2 \right) \right] = 2n\theta^2$, pues $\sum_{i=1}^{n} \frac{(X_i - \theta)^2}{\theta^2}$ se distribuye según una distribución χ_n^2.

■

Para aclarar el significado de la completitud, que radica en la cantidad de elementos que contenga la familia $\mathcal{P}$, se tiene el siguiente enunciado.

Ejemplo 2.3.6
La familia $\mathcal{P} = \left\{ \frac{1}{N} I_{\{1,2,...,N\}}(x), x \in \mathbb{N}, N \in \mathbb{N}^+ \right\}$, donde $\mathbb{N}$ es el conjunto de los números naturales, es completa, pero si se quita la distribución correspondiente al valor $N = N_0$, la función

$$g(x) = \begin{cases} c & \text{si} \quad x = N_0 \\ -c & \text{si} \quad x = N_0 + 1 \\ 0 & \text{si} \quad x \notin \{N_0, N_0 + 1\} \end{cases}$$

es un estimador insesgado de cero.

■

De todas formas se dispone de una herramienta muy útil para encontrar estadísticos completos.

Teorema 2.3.1 (existencia de estadísticos completos)
Para la familia de distribuciones exponencial k paramétrica

$$\mathcal{P} = \{c(\overrightarrow{\theta}) h(\overrightarrow{x}) e^{\sum_{1}^{k} c_i(\theta) T_i(\overrightarrow{x})} | \theta \in \Theta \subset \mathbb{R}^k\},$$

si la imagen de la aplicación $\overrightarrow{c} = (c_1(\overrightarrow{\theta}), ..., c_k(\overrightarrow{\theta})) : \Theta \to \mathbb{R}^k$ contiene un rectángulo abierto de $\mathbb{R}^k$, entonces el estadístico $\overrightarrow{T}(\overrightarrow{X}) = (T_1(\overrightarrow{X}), ..., T_k(\overrightarrow{X}))$ es completo.

La demostración puede verse en Lehmann(1986) pág. 142.

Además, en este caso se tiene asegurada la suficiencia del estadístico $\overrightarrow{T}$, mediante el teorema de factorización; y por el Teorema 2.2.3 también se sigue que $\overrightarrow{T}$ es un estadístico minimal suficiente.

Ejemplo 2.3.7
Para m.a.s. de tamaño n de una población $N(\theta, \theta)$ de media y desviación típica θ, con $\theta \in (0, +\infty)$ se tiene que

$$f(x_1, ..., x_n|\theta) = \left(\frac{1}{\sqrt{2\pi}\theta}\right)^n e^{-\frac{n}{2\theta^2}} e^{\frac{1}{\theta}\sum x_i - \frac{1}{2\theta^2}\sum x_i^2}$$

luego $\overrightarrow{c} = (c_1(\theta), c_2(\theta)) = (\frac{1}{\theta}, -\frac{1}{2\theta^2})$. Por tanto cuando $\theta \in (0, +\infty)$ la imagen de la aplicación $\overrightarrow{c}$ puede representarse mediante $x = \frac{1}{\theta}$, $y = -\frac{1}{2}x^2$

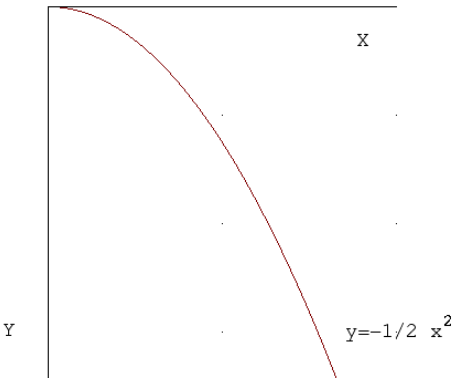


Figura 2.3.1. Representación de $\overrightarrow{c}(\theta)$ para el Ejemplo 2.3.7.

que es una rama de parábola, véase la Figura 2.3.1, y que por tanto no contiene a ningún abierto de $\mathbb{R}^2$. El teorema no permite afirmar que el estadístico $\overrightarrow{T}(\overrightarrow{X}) = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right)$ sea completo. De hecho no lo es; ver el Ejemplo 2.3.5.
■

Ahora se puede enunciar el resultado referido a los estadísticos ancilarios.

Teorema 2.3.2 (Basu)
Si $T(X_1, ..., X_n)$ es un estadístico suficiente y completo y $U(X_1, ..., X_n)$ es un estadístico ancilario, entonces T y U son independientes.

Demostración: Basta ver, utilizando funciones de densidad, que $f(u|t)$ es independiente de θ. Además

$$f(u) = f(t|\theta) f(u|t)$$

luego la esperanza de la función $g(t) = f(u|t) - f(u)$ respecto a la distribución de T es cero, y por la completitud de T se sigue el resultado.

■

El teorema puede utilizarse, para probar la independencia de ciertos estadísticos, para los que sea fácil comprobar su completitud.

Ejemplo 2.3.8

Para una m.a.s. de tamaño n de una población $Uniforme(0, \theta)$ se sigue que $\frac{X_{(1)}}{X_{(n)}}$ y $X_{(n)}$ son independientes.

Puede comprobarse mediante el teorema de factorización que $X_{(n)}$ es suficiente y que, además, es completo. El resultado se sigue, de ver que $\frac{X_{(1)}}{X_{(n)}}$ es ancilario:

$$F_{\frac{X_{(1)}}{X_{(n)}}}(t) = \int_0^\theta f_{X_{(n)}}(y) F_{X_{(1)}}(ty) dy \tag{2.3.1}$$

como es fácil ver que

$$F_{X_{(n)}}(y) = \begin{cases} 0 & \text{si} \quad y \leq 0 \\ \left(\frac{y}{\theta}\right)^n & \text{si} \quad y \in [0, \theta] \\ 1 & \text{si} \quad y \geq 0 \end{cases}$$

se sigue que

$$f_{X_{(n)}}(y) = n\ y^{n-1} \left(\frac{1}{\theta}\right)^n, \quad 0 < y < \theta.$$

Por otra parte

$$\begin{aligned} F_{X_{(1)}}(ty) &= 1 - P\{X_{(1)} > ty\} = 1 - P^n\{X > ty\} \\ &= \begin{cases} 0 & \text{si} \quad ty < 0 \\ 1 - \left(\frac{\theta - ty}{\theta}\right)^n & \text{si} \quad ty \in [0, \theta] \\ 1 & \text{si} \quad ty \geq \theta \end{cases} \end{aligned}$$

por lo que sustituyendo en (2.3.1)

$$F_{\frac{X_{(1)}}{X_{(n)}}}(t) = \frac{n}{\theta^n} \int_0^\theta y^{n-1} dy - \frac{n}{\theta^n} \int_0^\theta \left(1 - t\frac{y}{\theta}\right)^n y^{n-1} dy. \tag{2.3.2}$$

El primer sumando vale 1 y el segundo sumando se reduce a una función que no depende de θ mediante el cambio $\frac{y}{\theta} = x$, por lo que la Expresión (2.3.2) no depende de θ.

■

El siguiente ejemplo muestra cómo, mediante el teorema de Basu, se puede obtener el resultado, ya demostrado, de que para poblaciones normales los estadísticos $\overline{X}$ y S^2 son independientes.

Ejemplo 2.3.9
Para m.a.s. de tamaño n de una población $X \in N(\theta, \sigma)$ se tiene que $\overline{X}$ es suficiente y completo y además $(n-1)\frac{S^2}{\sigma^2}$ es ancilario, pues la distribución χ^2_{n-1} es independiente de θ, luego $\overline{X}$ y S^2 son independientes por el Teorema de Basu.

2.4. ESTADÍSTICOS INVARIANTES

En algunas situaciones, el problema de estimación tiene una estructura que permite también reducir la información muestral. Es intuitivo que si se supone un experimento en el que se busca medir la longitud de una pieza en centímetros a partir de diferentes medidas en centímetros, el resultado debe ser el mismo, mediante un cambio de unidades adecuado, es decir, si las medidas pasan a tomarse en metros.

Otras veces la estructura del problema es tal que sabiendo el valor que debe tener θ en determinados puntos muestrales se tiene unívocamente determinado el valor de una función de θ para otra función de los anteriores puntos muestrales. Así por ejemplo, si para determinar la probabilidad con la que una moneda sale cara, θ, si en vez de conocer el número de caras que se han presentado en n lanzamientos se sabe el número de cruces, lo que se puede estimar ahora es $1-\theta$ en lugar de θ.

Para concretar el concepto de estadístico invariante, hay que definir previamente ciertos conceptos.

Definición 2.4.1
Un conjunto de funciones $\mathcal{G} = \{g : \mathcal{X} \to \mathcal{X}\}$ definidas sobre el espacio muestral forman grupo, respecto a la operación de composición de aplicaciones, cuando se cumplen las dos condiciones:

1) Para toda $g \in \mathcal{G}$ existe $g' \in \mathcal{G}$ tal que $g'(g(x)) = x$, para todo punto muestral. Se dice que g' es la aplicación inversa de g.

2) Dadas dos funciones g y g' de $\mathcal{G}$ existe g" de $\mathcal{G}$ tal que $g'g(x) = g"(x)$ para todo punto muestral. Se dice que $\mathcal{G}$ es cerrado respecto a la composición de aplicaciones.

La pertenencia al grupo $\mathcal{G}$ de la aplicación identidad se sigue de las dos propiedades introducidas.

Ejemplo 2.4.1
Para una población con distribución binomial $Binomial(n, \theta)$ con $\mathcal{X} = \{0, 1, ..., n\}$ el conjunto formado por las dos aplicaciones $\mathcal{G} = \{e(x) = x, g(x) = n - x\}$ es un grupo, ya que cada aplicación del conjunto es inversa de si misma y es cerrada.

Definición 2.4.2 (familia de distribuciones invariante)
La familia de distribuciones $\mathcal{P} = \{f(\overrightarrow{x}|\theta)|\theta \in \Theta\}$ se dice invariante por el grupo de transformaciones $\mathcal{G} = \{g : \mathcal{X} \to \mathcal{X}\}$ si cuando $\overrightarrow{X}$ se distribuye

mediante $f(\overrightarrow{x}|\theta)$, se sigue que para cualquier g perteneciente a $\mathcal{G}$ $g(\overrightarrow{X})$ se distribuye mediante $f(\overrightarrow{x}|\theta')$ con $\theta' \in \Theta$. Se acostumbra a notar $\theta' = \overline{g}(\theta)$ a la aplicación de Θ en Θ que se corresponde con g.

Ejemplo 2.4.2
Para el caso de la distribución $Binomial(n, \theta)$ tratado en el Ejemplo 2.4.1, las aplicaciones correspondientes son $\overline{e}(\theta) = \theta$ y $\overline{g}(\theta) = 1 - \theta$.

■

Ya se puede introducir el concepto de estimador invariante frente a un grupo de transformaciones.

Definición 2.4.3 (estadístico invariante)
Si $\mathcal{G}$ es un grupo de transformaciones que deja invariante a la familia de distribuciones $\mathcal{P} = \{f(\overrightarrow{x}|\theta)|\theta \in \Theta\}$, un estadístico T es invariante, respecto al grupo de transformaciones $\mathcal{G} = \{g : \mathcal{X} \to \mathcal{X}\}$ si para todo g de $\mathcal{G}$ se tiene que $Tg(\overrightarrow{x}) = \overline{g}T(\overrightarrow{x})$.

Ejemplo 2.4.3
Para la población $Binomial(n, \theta)$ con muestras de tamaño 1 el estadístico $T(X) = \frac{X}{n}$ es invariante respecto al grupo $\mathcal{G} = \{e(x) = e, g(x) = n - x\}$, ya que

$$Tg(x) = T(n - x) = 1 - \frac{x}{n}$$
$$\bar{g}T(x) = \bar{g}\left(\frac{x}{n}\right) = 1 - \frac{x}{n}$$

■

Un estadístico es invariante cuando la estimación que se hace al observar $\overrightarrow{x}$ transformada por $\overline{g}$ coincide con la que se hace si se observa $g(\overrightarrow{x})$.

El problema para la consideración de los estadístivos invariantes, es que se requiere descubrir primeramente el grupo $\mathcal{G}$ de transformaciones.

Ejemplo 2.4.4
Para m.a.s. de tamaño n de una población $N(\theta, \sigma_0)$, con σ_0 conocida y $\theta \in \mathbb{R}$, el estadístico $T(\overrightarrow{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i$ es invariante para el grupo de traslaciones $\mathcal{G} = \{g_a : \mathcal{X} \to \mathcal{X},\ g_a(x_1, ..., x_n) = (x_1 + a, ..., x_n + a) \text{ con } a \in \mathbb{R}\}$.

■

Las ideas anteriores son válidas en el caso en que el parámetro θ sea k-dimensional.

Ejemplo 2.4.5
Para m.a.s. de tamaño n, la familia de distribuciones normales de media $\theta_1 = \mu$ y de varianza $\theta_2 = \sigma^2$, con $\mu \in \mathbb{R}$ y $\sigma^2 \in (0, +\infty)$ es invariante por

el grupo de transformaciones $\mathcal{G} = \{g_{a,b}(x_1, ..., x_n) = (ax_1 + b, ..., ax_n + b)\}$. Se obtiene que $\overline{g}_{a,b}(\mu, \sigma^2) = (a\mu + b, a^2\sigma^2)$. El estadístico $\left(\sum_1^n X_i \sum_1^n X_i^2\right)$ no es invariante, mientras que si que lo es $\left(\overline{X}, S^2\right)$.

2.5. PRINCIPIOS DE REDUCCIÓN DE DATOS

En este capítulo se han estudiado distintos procedimientos de resumir la información suministrada por la muestra. Estos procedimientos son; en último término estadísticos y sobre ellos, o a través de ellos, se estiman los valores desconocidos de la población.

Se va a introducir un estadístico más que es la verosimilitud o la función de verosimilitud, que también puede ser utilizado para resumir los datos; finalmente se consideran todos los estadísticos como entes que dan información y a esta información se le va a llamar evidencia. Desde este punto de vista, se cambia, sólo en esta sección, el nombre de estadístico por el de principio y así se habla de principio de suficiencia, de principio de verosimilitud y de principio de invariancia, en lugar de estadístico suficiente, de función de verosimilitud y de estadístico invariante, respectivamente. En este contexto se demuestra el Teorema de Birnbaum, un resultado que relaciona los estadísticos suficientes y la función de verosimilitud y que pone de manifiesto las implicaciones de los distintos métodos de inferencia.

2.5.1. Principio de verosimilitud

La idea es considerar la distribución de probabilidad de la muestra, no como función de $x = (x_1, ..., x_n)$ sino como función del parámetro θ desconocido.

Definición 2.5.1.1. (función de verosimilitud)
Dada la familia $\mathcal{P} = \{f(\overrightarrow{x}|\theta)|\theta \in \Theta\}$ de distribuciones de probabilidad donde $f(.|\theta)$ es una función de densidad o de masa y supuesto que se ha observado el valor muestral $\overrightarrow{x} = (x_1, ..., x_n)$ la función de θ definida mediante

$$L(\theta|\overrightarrow{x}) = f(\overrightarrow{x}|\theta)$$

se llama función de verosimilitud.

Ejemplo 2.5.1.1.
Para una muestra de tamaño 1 de una población $Uniforme(0, \theta)$, la función de densidad y de verosimilitud están representadas en la Figura 2.5.1.1 donde se observa la drástica diferencia entre ambas.

■

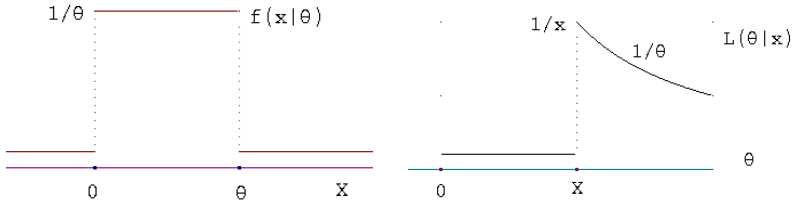


Figura 2.5.1.1. Función de densidad y de verosimilitud para la población Uniforme$(0, \theta)$.

Como se anticipaba, el principio de verosimilitud especifica como debe usarse la función de verosimilitud para la reducción de datos.

Definición 2.5.1.2. (principio de verosimilitud)
Si $\overrightarrow{x}$ e $\overrightarrow{y}$ son dos puntos muestrales, tales que, existe una constante $c(\overrightarrow{x}, \overrightarrow{y})$ que no depende de θ verificando

$$L(\theta|\overrightarrow{x}) = c(\overrightarrow{x}, \overrightarrow{y})L(\theta|\overrightarrow{y}) \text{ para todo } \theta,$$

entonces la evidencia estadística que suministran ambos puntos debe ser idéntica, es decir, $Ev(E_1, \overrightarrow{x}) = Ev(E_2, \overrightarrow{y})$.

■

Dos aspectos son importantes en esta definición. El primero es que la evidencia estadística se toma en sentido amplio y no se define, así puede ser ésta un estadístico muestral, un estadístico suficiente, un intervalo de confianza etc. El segundo es que las dos funciones de verosimilitud no tienen por qué estar obligatoriamente definidas en el mismo espacio muestral. Realmente la evidencia estadística debe expresarse como

$$Ev(E, \overrightarrow{x}) \text{ donde } E = (\mathcal{X}, \theta, \mathcal{P} = \{f(\overrightarrow{x}|\theta)\})$$

la evidencia depende del experimento bajo estudio E y del punto muestral observado. Con lo que del principio de verosimilitud, Definición 2.5.1.2, se concluye que $Ev(E_1, \overrightarrow{x}) = Ev(E_2, \overrightarrow{y})$ con

$$E_1 = (\mathcal{X}, \theta, \mathcal{P}_1 = \{f(\overrightarrow{x}|\theta)\}) \text{ y } E_2 = (\mathcal{Y}, \theta, \mathcal{P}_2 = \{g(\overrightarrow{y}|\theta)\}).$$

Ejemplo 2.5.1.2.
La evidencia que se obtiene de un modelo $Binomial(n, \theta)$ sobre θ, cuando se observan t éxitos en n repeticiones, es la misma que la obtenida de un modelo *Binomial negativo*(t, θ), siempre que se suponga que se han realizado n repeticiones hasta obtener t éxitos. En efecto en el primer caso

$$f(t|\theta) = \binom{n}{t}\theta^t(1-\theta)^{n-t}$$

y en el segundo

$$f(\text{n repeticiones hasta obtener } t \text{ éxitos}|\theta) = \binom{n-1}{t-1}\theta^t(1-\theta)^{n-t}$$

por tanto se puede aplicar el principio de verosimilitud con $c(t, n) = \frac{n}{t}$. Es decir, si para estimar la probabilidad con la que sale cara una moneda, se tira ésta n veces y es t el número de caras que se han presentado (modelo binomial), se obtiene la misma evidencia para θ, que si se repite el experimento hasta observar t caras y para ello ha habido que realizar n tiradas.

2.5.2. Principio de suficiencia

El concepto de estadístico suficiente, también puede trasladarse a principio.

Definición 2.5.2.1 (principio de suficiencia)
En un experimento $E = (\mathcal{X}, \theta, \mathcal{P} = \{f(\overrightarrow{x}|\theta)|\theta \in \Theta\})$ si $T(\overrightarrow{x})$ es un estadístico suficiente para θ y se tiene que $T(\overrightarrow{x}) = T(\overrightarrow{y})$, entonces los dos puntos muestrales, $\overrightarrow{x}$ e $\overrightarrow{y}$, suministran la misma evidencia estadística, es decir $Ev(E, \overrightarrow{x}) = Ev(E, \overrightarrow{y})$.

Ejemplo 2.5.2.1
En un experimento consistente en la repetición de n pruebas de Bernoulli, la evidencia estadística que se obtiene de dos puntos muestrales con el mismo número de éxitos es la misma.

2.5.3. Principio de condicionalidad

Dados dos experimentos $E_1 = (\mathcal{X}_1, \theta, \mathcal{P} = \{f_1(x_1|\theta)|\theta \in \Theta\})$ y $E_2 = (\mathcal{X}_2, \theta, \mathcal{P} = \{f_2(x_2|\theta)|\theta \in \Theta\})$ considérese el experimento mixto, en el que dependiendo de una v.a. J, que toma el valor 1 ó 2 con probabilidad 1/2, se observa el experimento E_J. Formalmente se nota al experimento mixto por $E^* = (\mathcal{X}^*, \theta, \{f^*(x^*|\theta)|\theta \in \Theta\})$ donde $x^* = (j, x_j)$ y $f^*(j, x_j|\theta) = \frac{1}{2}f_j(x_j|\theta)$, entonces el principio de condicionalidad permite afirmar que

$$Ev(E^*, (j, x_j)) = Ev(E_j, x_j).$$

El principio de condicionalidad dice algo bastante intuitivo: mecanismos aleatorios que no dependan del valor a determinar θ, no proporcionan evidencia sobre él. Posteriormente, al tratar los contrastes de hipótesis, se consideran problemas en los que a veces se sugiere aleatorizar, para conseguir un test de tamaño determinado; pues bien según el principio de condicionalidad este tipo de actuación no suministra información sobre θ y no necesita ser considerado.

El valor $\frac{1}{2}$ (probabilidad con la que toma valores la v.a. J) puede ser sustituido por otro cualquiera sin modificar el significado del principio de condicionalidad; lo sustancial es que la distribución de la v.a. J no dependa de θ, el parámetro a determinar.

2.5.4. Teorema de Birnbaum

El siguiente enunciado relaciona los tres principios, de verosimilitud, suficiencia y condicionalidad, y tiene un significado importante para establecer consecuencias de las distintas aproximaciones que pueden darse en la inferencia estadística.

Teorema 2.5.4.1 (Birnbaum)
El principio de verosimilitud es equivalente a los principios de suficiencia y condicionalidad.

Demostración: Dados dos puntos $\overrightarrow{x}$ e $\overrightarrow{y}$ en dos experimentos E_1, E_2 tales que $L_1(\theta|\overrightarrow{x}) = c(\overrightarrow{x}, \overrightarrow{y})L_2(\theta|\overrightarrow{y})$, sobre el experimento mixto E^* construido a partir de E_1 y E_2 mediante el principio de condicionalidad, se define el estadístico

$$T(j, \overrightarrow{x}_j) = \begin{cases} (1, \overrightarrow{x}) & \text{si } \ j = 1 \ \text{ y } \ \overrightarrow{x}_1 = \overrightarrow{x} \\ (1, \overrightarrow{x}) & \text{si } \ j = 2 \ \text{ y } \ \overrightarrow{x}_2 = \overrightarrow{y} \\ (j, \overrightarrow{x}_j) & \text{en el resto} \end{cases} \qquad j = 1, 2$$

es decir, el estadístico T asigna el mismo valor a los puntos $(1, \overrightarrow{x})$ y $(2, \overrightarrow{y})$ y es la identidad para el resto. El estadístico T es suficiente ya que $f((1, \overrightarrow{x})|T = (1, \overrightarrow{x})) = (1+c(\overrightarrow{x}, \overrightarrow{y}))^{-1}$, $f((2, \overrightarrow{y})|T = (1, \overrightarrow{x})) = c(\overrightarrow{x}, \overrightarrow{y})/(1+c(\overrightarrow{x}, \overrightarrow{y}))$, $f((j, \overrightarrow{x}_j)|T = (j, \overrightarrow{x}_j)) = 1$ y $f((j, \overrightarrow{x}'_j)|T = (j, \overrightarrow{x}_j)) = 0$ esto es, no depende del parámetro. Siendo f la distribución condicionada por el estadístico y donde se ha tenido en cuenta que

$$\begin{aligned} f(T = (1, \overrightarrow{x})|\theta) &= f^*((1, \overrightarrow{x})|\theta) + f^*((2, \overrightarrow{y})|\theta) \\ &= \tfrac{1}{2} f_1(\overrightarrow{x}|\theta) + \tfrac{1}{2} f_2(\overrightarrow{y}|\theta) \\ &= \tfrac{1}{2} L_1(\theta|\overrightarrow{x}) + \tfrac{1}{2} L_2(\theta|\overrightarrow{y}) \\ &= \tfrac{1}{2} L_1(\theta|\overrightarrow{x})[1 + c(\overrightarrow{x}, \overrightarrow{y})] \end{aligned}$$

con lo que

$$\begin{aligned} f((1,\overrightarrow{x})|T=(1,\overrightarrow{x})) &= \frac{f^*((1,\overrightarrow{x})|\theta)}{f(T=(1,\overrightarrow{x})|\theta)} \\ &= \frac{\frac{1}{2}L_1(\theta|\overrightarrow{x})}{\frac{1}{2}L_1(\theta|\overrightarrow{x})+\frac{1}{2}L_2(\theta|\overrightarrow{y})} \\ &= \frac{\frac{1}{2}L_1(\theta|\overrightarrow{x})}{\frac{1}{2}L_1(\theta|\overrightarrow{x})+\frac{1}{2}C(\overrightarrow{x},\overrightarrow{y})L_1(\theta|\overrightarrow{x})} \\ &= [1+c(\overrightarrow{x},\overrightarrow{y})]^{-1} \end{aligned}$$

análogamente para los otros valores.

Si se aplica el principio de suficiencia, ha de ser

$$Ev(E^*,(1,\overrightarrow{x})) = Ev(E^*,(2,\overrightarrow{y}))$$

y mediante el principio de condicionalidad

$$\begin{aligned} Ev(E^*,(1,\overrightarrow{x})) &= Ev(E_1,\overrightarrow{x}) \\ Ev(E^*,(2,\overrightarrow{y})) &= Ev(E_2,\overrightarrow{y}) \end{aligned}$$

por lo tanto se tiene demostrado que

$$Ev(E_1,\overrightarrow{x}) = Ev(E_2,\overrightarrow{y})$$

con lo que se obtiene el principio de verosimilitud.

Recíprocamente, para el principio de condicionalidad. Sea E^* el experimento mixto de E_1 y E_2 . Como

$$\begin{aligned} L(\theta|j,x_j) &= f^*(j,x_j|\theta) \\ &= \tfrac{1}{2}f_j(\overrightarrow{x}_j|\theta) \\ &= \tfrac{1}{2}L_j(\theta|\overrightarrow{x}_j), \end{aligned}$$

se tiene que en los dos experimentos E^* y E_j es

$$L(\theta|j,\overrightarrow{x}_j) = \frac{1}{2}L_j(\theta|\overrightarrow{x}_j)$$

y por el principio de verosimilitud

$$Ev(E^*,(j,\overrightarrow{x}_j)) = Ev(E_j,\overrightarrow{x}_j).$$

Por último, para demostrar que se cumple el principio de suficiencia considérese un experimento E y dos puntos muestrales tales que $T(\overrightarrow{x}) = T(\overrightarrow{y})$, siendo T un estadístico suficiente. Por el teorema de factorización, Teorema 2.2.1, se tiene

$$\begin{aligned} f(\overrightarrow{y}|\theta) &= h(\overrightarrow{y})g_\theta(T(\overrightarrow{y})) \\ &= h(\overrightarrow{y})g_\theta(T(\overrightarrow{x})) \\ &= \frac{h(\overrightarrow{y})}{h(\overrightarrow{x})}f(\overrightarrow{x}|\theta) \end{aligned}$$

Es decir,

$$L(\theta|\overrightarrow{y}) = c(\overrightarrow{x}, \overrightarrow{y})L(\theta|\overrightarrow{x})$$

con $c(\overrightarrow{x}, \overrightarrow{y}) = h(\overrightarrow{y})|h(\overrightarrow{x})$. Por el principio de verosimilitud

$$Ev(E, \overrightarrow{x}) = Ev(E, \overrightarrow{y})$$

y por tanto se cumple el principio de suficiencia.
■

El significado del teorema de Birnbaum es importante desde el punto de vista de los fundamentos de la Estadística, muchos de los procedimientos estadísticos usuales violan el principio de verosimilitud, en concreto los procedimientos que se basan en la distribución muestral del estadístico pueden hacerlo.

Si se pasa de un modelo binomial a uno binomial negativo, la función de masa cambia y por tanto las propiedades del estimador también pueden cambiar. Sin embargo a la luz del teorema esto significa contradecir el principio de suficiencia, que es compartido por toda aproximación a la inferencia, o el principio de condicionalidad que parece bastante aséptico. El teorema de Birnbaum constituye uno de los motivos por los que el principio de verosimilitud no es universalmente aceptado, a pesar de que como se verá, la función de verosimilitud posee muchas buenas propiedades estadísticas. Los procedimientos bayesianos, que serán introducidos en la Sección 3.2.3, por estar basados en la distribución de probabilidad final satisfacen el principio de verosimilitud, ya que si $\overrightarrow{x}$ e $\overrightarrow{y}$ cumplen

$$L_2(\theta|\overrightarrow{y}) = c(\overrightarrow{x}, \overrightarrow{y})L_1(\theta|\overrightarrow{x})$$

Las distribuciones finales $\pi_1(\theta|\overrightarrow{x})$ y $\pi_2(\theta|\overrightarrow{y})$, cuando la inicial viene dada por $\pi(\theta)$, son iguales:

$$\begin{aligned}\pi_2(\theta|\overrightarrow{y}) &= \frac{\pi(\theta)f_2(\overrightarrow{y}|\theta)}{\int_\Theta \pi(\theta)f_2(\overrightarrow{y}|\theta)d\theta} \\ &= \frac{\pi(\theta)f_1(\overrightarrow{x}|\theta)}{\int_\Theta \pi(\theta)f_1(\overrightarrow{x}|\theta)d\theta} \\ &= \pi_1(\theta|\overrightarrow{x}).\end{aligned}$$

Por tanto la evidencia estadística que se obtiene mediante $\overrightarrow{x}$ del primer experimento

$$Ev(E_1, \overrightarrow{x}) = \pi_1(\theta|\overrightarrow{x})$$

coincide con la que se obtiene mediante $\overrightarrow{y}$ del segundo experimento

$$Ev(E_2, \overrightarrow{y}) = \pi_2(\theta|\overrightarrow{y})$$

2.6. APROXIMACIÓN HISTÓRICA

Las primeras propiedades de los estadísticos, están al menos en germen, en los trabajos de Laplace, pero quienes las estudian de una forma más sistemática son Karl Pearson y Fisher.

Karl Pearson (1857-1936) nace en Londres. Hijo de un abogado, estudia leyes para satisfacer a su padre, aunque simultanea estos estudios con los de matemáticas, en el *King Colleg*e de Cambridge. Con 22 años se marcha a Alemania a hacer su doctorado y estudia física y metafísica, a su vuelta es profesor de matemáticas y mecánica en el *University College* de Londres. A partir de 1911 se convierte en profesor de ingeniería en la misma institución. Fué un partidario convencido de la teoría de la evolución de las especies, introducida por Darwin, y también un ferviente socialista. En 1889 Galton publica su libro *Herencia Natural* y Weldon se incorpora a la cátedra de zoología en el *University College*. La necesidad de verificar numéricamente lo expuesto por Galton y de dar respuestas a las cuestiones de clasificación planteadas por el zoólogo, llevan a Karl Pearson a desarrollar diferentes distribuciones para resolver las cuestiones que le van planteando.

Inicialmente Karl Pearson recurre a representaciones gráficas de los datos, para posteriormente desarrollar su célebre familia de curvas e introducir procedimientos de ajustes. Así entre 1893 y 1906 publica unos 100 artículos acerca de la teoría de la estadística y sus aplicaciones, entre estos trabajos, están incluidas las *Tablas para estadísticos y biometristas*, destinadas a ayudar a realizar ajustes de curvas a nubes de puntos, y sus estudios sobre la correlación.

Retrata la forma de ser de Karl Pearson, que en su primera época como profesor, al hacer una verificación de la falta de aleatoriedad de los resultados de una ruleta, escribe al gobierno francés, recomendándole que cierre los casinos y envíe sus fondos a la Academia de Ciencias para la fundación de un laboratorio de probabilidad. Entre 1895 y 1897, Karl Pearson realiza sus estudios sobre correlación y regresión y obtiene la distribución del coeficiente de correlación muestral para una distribución normal bivariante.

El célebre método de ajuste de la χ^2 está incluido en una memoria, publicada en el año 1900; posteriormente lo generaliza a tablas de contingencia en el año 1904. Mantuvo una polémica con Fisher respecto al número de grados de libertad de la χ^2 cuando parte de los parámetros desconocidos de la población son estimados de la muestra.

Entre los años 1921 y 1933 da en el *University College* de Londres una serie de conferencias que se han convertido en un clásico para cualquier persona interesada en el estudio de la historia de la estadística.

Con una gran simplificación podría decirse que a Karl Pearson se le debe el fuerte auge del método de los momentos para la determinación de los parámetros, y un estudio bastante completo de las familias de curvas

que obtuvo para ajustar a distintos fenómenos, como solución de ecuaciones diferenciales para distintos valores de las constantes; así clasificó la solución en cinco tipos de curvas que incluían a la beta simétrica, la beta asimétrica, la gamma y la normal.

El resto de las propiedades de los estimadores, junto con otros muchos resultados, se deben a Fisher (1890-1962), sin duda el estadístico más influyente y que más contribuciones de conceptos y resultados ha realizado. Una sucinta biografía de Fisher puede verse en la Sección 4.6. A él se deben los conceptos de consistencia, eficacia, suficiencia y ancilariadad, así como el método de la máxima verosimilitud, el diseño de experimentos y el análisis de la varianza, (véase el Capítulo 7). Un amplio estudio de la biografía y las contribuciones estadísticas de Fisher se encuentra en Girón y Gómez Villegas (1998).

Con referencia a los fundamentos de la inferencia estadística se debe citar a Allan Birnbaum (1923-1976), quien estudió en Berkeley, ciencias, filosofía y matemáticas, trabajó en el *Instituto Courant de Ciencias Matemáticas* e impartió clases como profesor visitante en varias universidades europeas y americanas, como Stanford, Londres, Cambridge etc. Su resultado de 1962 que relaciona los conceptos de verosimilitud, suficiencia y condicionalidad, es de capital importancia para entender los fundamentos de las principales aproximaciones a la inferencia estadística.

2.7. EJERCICIOS

2.7.1

Sea X una observación de una población $N(0,\sigma)$. ¿Es $|X|$ un estadístico suficiente?

2.7.2

Encuéntrese un estadístico suficiente para m.a.s. de tamaño n de cada una de las poblaciones siguientes:

a) $Beta(\alpha,\beta)$, $\alpha>0,\beta>0$.

b) $Gamma(a,p), a>0,p>0$.

c) $p_{N_1N_2}(x)=\frac{1}{N_2-N_1}I_{\{N_1+1,N_1+2,\ldots,N_2\}}(x)$ con $N_1<N_2$, enteros positivos.

d) $f(x|\theta)=e^{-x+\theta}I_{(\theta,+\infty)}(x),\ \ \theta\in\mathbb{R}$.

e) $f(x|\mu,\sigma)=\frac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}(\ln x-\mu)^2}I_{(0,+\infty)}(x),\ \ \mu\in\mathbb{R},\sigma>0$.

f) $p(x|\theta,p)=(1-p)p^{x-\theta}I_{\{\theta,\theta+1,\ldots\}}(x),\ \ \theta\in\mathbb{R}\qquad 0<p<1$.

2.7.3

Para m.a.s. de tamaño n de una población $Uniforme(\theta,4\theta),\theta\in\mathbb{R},$ pruébese que $(X_{(1)},X_{(n)})$ es un estadístico suficiente pero no completo.

2.7.4
Para m.a.s. de tamaño n de una población con función de densidad

$$f(x|\theta) = e^{-(x-\theta)} I_{(\theta,+\infty)}(x), \quad \theta \in \mathbb{R}.$$

a) Pruébese que $X_{(1)}$ es un estadístico suficiente y completo.

b) Mediante el teorema de Basu, pruébese que $X_{(1)}$, el mínimo, y S^2, la cuasivarianza muestral, son independientes.

2.7.5
Para una población discreta dependiente de un parámetro θ, ¿es $T(X_1, ..., X_n) = (X_1, ..., X_{n-1})$ un estadístico suficiente para θ?

2.7.6
Para m.a.s. de tamaño n de una población con función de densidad

$$f(x|\theta) = \frac{x}{\theta} e^{-\frac{1}{2\theta}x^2} \quad x > 0, \ \theta > 0,$$

pruébese que el estadístico $\sum_{i=1}^{n} X_i$ no es suficiente.

2.7.7
Para m.a.s. de tamaño n de una población con parámetro de centralización μ y de escala σ,

$$f(x|\mu, \sigma) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$$

y dos estadístico T_1 y T_2 que verifiquen

$$T_i(ax_1 + b, ..., ax_n + b) = aT_i(x_1, ..., x_n) \quad i = 1, 2$$

para cualesquiera valores a, b con $a > 0$.

a) Pruébese que $\frac{T_1}{T_2}$ es un estadístico ancilario para (μ, σ).

b) Sea $R = X_{(n)} - X_{(1)}$ y S la cuasidesviación típica muestral. Verifíquese que $\frac{R}{S}$ es un estadístico ancilario.

2.7.8
Sea N una variable aleatoria que toma los valores $1, 2, ...$ con probabilidades $p_1, p_2, ...$ con $\sum_{i=1}^{\infty} p_i = 1$. Si primero se observa $N = n$ y a continuación se realizan n experimentos de $Bernoulli(\theta)$ obteniéndose x éxitos. Pruébese, que, de acuerdo con el principio de verosimilitud, se debería obtener la misma evidencia estadística del par (n, x) que del resultado x.

2.7.9
Para una población, con función de masa dada mediante

$$P\{X = -1|\theta\} = \theta \quad P\{X = x|\theta\} = (1-\theta)^2 \theta^x I_{\{0,1,...\}}(x)$$

con $\theta \in (0, 1)$, y una muestra de tamaño 1, ¿es X un estadístico completo?

2.7.10

Para m.a.s. de tamaño n de una población $N(\alpha\sigma, \sigma)$ donde α es un número real conocido, ¿es $T(X_1, ..., X_n) = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2 \right)$ un estadístico suficiente y completo para σ?

2.7.11

Para m.a.s. de tamaño n de una población $Uniforme(\theta, \theta + 1)$ con $\theta > 0$.

a) Encuéntrese un estadístico suficiente para θ.

b) ¿Es el recorrido muestral $R = X_{(n)} - X_{(1)}$ un estadístico ancilario?

c) ¿Se podría generalizar la parte b) a una familia de distribuciones con parámetro de localización θ, es decir, si $F(x|\theta) = F(x-\theta)$ es $R = X_{(n)} - X_{(1)}$ un estadístico ancilario?

2.7.12

Para m.a.s. de tamaño n de una población, con función de densidad dada mediante $f(x|\theta) = f(x - \theta)$, se considera el grupo de transformaciones $\mathcal{G} = \{g_a(x_1, ..., x_n) = (x_1 + a, ..., x_n + a) | a \in \mathbb{R}\}$. Pruébese que si $E_0 X = 0$, entonces el estimador $\overline{X}$ es insesgado e invariante para θ.

2.7.13

Sea $X_1, ..., X_n$ una m.a.s. de tamaño n de una población de parámetros de centralización θ y de escala σ conocida, es decir $f(x|\theta, \sigma) = \frac{1}{\sigma} f\left(\frac{x-\theta}{\sigma}\right)$. Se quiere estimar el parámetro θ y se consideran los dos grupos de transformaciones

$$\mathcal{G}_1 = \{g_{a,c}(x_1, ..., x_n) = (cx_1 + a, ..., cx_n + a) | a \in \mathbb{R} \quad c > 0\}$$

y

$$\mathcal{G}_2 = \{g_a(x_1, ..., x_n) = (x_1 + a, ..., x_n + a) | a \in \mathbb{R}\ \}$$

Pruébese que los estimadores de la forma $T(\overline{X}) = \overline{X} + k$, con k distinta de cero, no son invariantes respecto a $\mathcal{G}_1$ y sí lo son respecto a $\mathcal{G}_2$.

2.7.14

Dada una observación $X = x$ de una población con función de masa

$$p(x|\theta) = \left(\frac{\theta}{2}\right)^{|x|} (1 - \theta)^{1-|x|} \ I_{\{-1,0,1\}}(x) \quad \text{con} \quad 0 \leq \theta \leq 1$$

a) ¿Es X un estadístico completo?

b) ¿Es $|X|$ un estadístico minimal suficiente y completo?

c) ¿Es el estadístico Z ancilario?

$$Z = \begin{cases} X/|X| & \text{si } X \neq 0 \\ 0 & \text{si } X = 0 \end{cases}$$

2.7.15
Se considera una m.a.s. de tamaño n de una población con función de densidad

$$f(x|\theta) = \frac{(\ln\theta)\theta^x}{\theta - 1}\, I_{(0,1)}(x) \quad \text{con} \quad \theta > 1$$

a) Encuéntrese un estadístico suficiente.
b) ¿Es el estadístico de a) minimal suficiente? ¿Es completo?

2.7.16
Se considera una m.a.s. de tamaño n de una población con función de densidad

$$f(x|\theta) = \frac{2x}{\theta^2} I_{(0,\theta)}(x) \quad \text{con} \quad \theta > 0$$

a) ¿Es el estadístico $\left(\prod_{i=1}^{n} X_i, X_{(n)}\right)$ suficiente para θ? ¿Es minimal suficiente?
b) ¿Es completo el estadístico $X_{(n)}$?

2.8. SOLUCIÓN A LOS EJERCICIOS

2.8.1
Como la función de densidad es

$$f(x|\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}|x|^2}$$

por el Teorema 2.2.1 (factorización) se tiene que

$$f(x|\sigma) = h(x)g(|x||\sigma)$$

con $h(x) = \frac{1}{\sqrt{2\pi}}$ y $g(|x||\sigma) = \frac{1}{\sigma} e^{-\frac{1}{2}|x|^2}$; luego el estadístico $|x|$ es suficiente.

2.8.2
a) Como $X \sim Beta(\alpha, \beta)$ su función de densidad es,

$$f(x|\alpha,\beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} \quad 0 < x < 1, \alpha > 0, \beta > 0$$

de donde

$$f(x_1,...,x_n|\alpha,\beta) = \frac{1}{B(\alpha,\beta)} \left(\prod_{i=1}^{n} x_i\right)^{\alpha-1} \left(\prod_{i=1}^{n}(1-x_i)\right)^{\beta-1}$$

con $0 < x_i < 1$ para todo $i = 1, ..., n$. Por el teorema 2.2.1 (factorización) se tiene que $\overrightarrow{T}(x_1,...,x_n) = \left(\prod_{i=1}^{n} X_i, \prod_{i=1}^{n}(1-X_i)\right)$ es un estadístico suficiente.

b) Como $X \sim Gamma(a,p)$ la densidad de la población es,

$$f(x|a,p) = \frac{a^p}{\Gamma(p)} e^{-ax} x^{p-1} \quad x > 0, a > 0, p > 0$$

de donde

$$f(x_1,...,x_n|a,p) = \frac{a^{np}}{\Gamma^n(p)} e^{-a\sum_{i=1}^{n} x_i} \left(\prod_{i=1}^{n} x_i\right)^{p-1}$$

con $x_i > 0$ para todo $i = 1,...,n$. Por el teorema 2.2.1 (factorización) el estadístico suficiente es

$$\overrightarrow{T}(X_1,...,X_n) = \left(\prod_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i\right).$$

c) Si X tiene como función de masa la de la uniforme discreta entre N_1 y N_2, la función de masa de la muestra es

$$P\{x_1,...,x_n|N_1,N_2\} = \left(\frac{1}{N_2 - N_1}\right)^n I_{\{N_1+1,...,+\infty\}}(x_{(1)}) I_{\{-\infty,...,N_2\}}(x_{(n)})$$

luego por el teorema de factorización el estadístico suficiente es

$$\overrightarrow{T}(X_1,...,X_n) = (X_{(1)}, X_{(n)}).$$

d) Como la función de densidad de la muestra es

$$f(x_1,...,x_n|\theta) = e^{-\sum_{i=1}^{n} x_i + n\theta} I_{[\theta,+\infty)}(x_{(1)})$$

el estadístico suficiente, es por el teorema de factorización

$$T(X_1,...,X_n) = X_{(1)}.$$

e) La función de densidad de la muestra es

$$f(x_1,...,x_n|\mu,\sigma) = \frac{1}{\prod_{i=1}^{n} x_i \sigma^n (\sqrt{2\pi})^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} (\ln x_i - \mu)^2}$$

con $x_i \geq 0$ para todo $i = 1,...,n$. Por el Teorema 2.2.1 (factorización), si se desarrolla el cuadrado del exponente para separar la dependencia del parámetro μ y de la muestra, se obiene que el estadístico suficiente es

$$\overrightarrow{T}(X_1,...,X_n) = \left(\sum_{i=1}^{n} \ln^2 X_i, \sum_{i=1}^{n} \ln X_i\right).$$

f) Al ser la función de masa de la muestra

$$P\{x_1,...,x_n|\theta,p\} = (1-p)^n p^{\sum_{i=1}^{n} x_i - n\theta} I_{[\theta,+\infty)}(x_{(1)}) \cdot \\ \cdot I_{Z^+}(x_{(1)}-\theta) \prod_{i=1}^{n} I_{Z^+}(x_i - x_{(1)})$$

el estadístico suficiente es

$$\overrightarrow{T}(X_1,...,X_n) = (X_{(1)}, \sum_{i=1}^{n} X_i).$$

2.8.3

La función de densidad de la muestra es

$$f(x_1,...,x_n|\theta) = \frac{1}{(3\theta)^n} I_{[\theta,+\infty)}(x_{(1)}) I_{(-\infty,4\theta]}(x_{(n)})$$

por tanto, de acuerdo con el Teorema 2.2.1 (factorización), es un estadístico suficiente para θ

$$\overrightarrow{T}(X_1,...,X_n) = (X_{(1)}, X_{(n)}).$$

Para ver que este estadístico no es completo, se comprueba que la función

$$f(X_{(1)}, X_{(n)}) = \frac{n+1}{n+4} X_{(1)} - \frac{n+1}{4n+1} X_{(n)}$$

tiene de media cero y no es idénticamente nula. Para ello la función de distribución de $X_{(1)}$ viene dada por

$$\begin{aligned} F_{X_{(1)}}(u) &= P\{X_{(1)} \leq u\} = 1 - P^n\{X_1 > u\} \\ &= 1 - \left(\int_{\theta}^{4\theta} \frac{1}{3\theta} du \right)^n \\ &= 1 - \left(\frac{4\theta - u}{3\theta} \right)^n \end{aligned}$$

con lo que la función de densidad de $X_{(1)}$ es

$$f_{X_{(1)}}(u) = n \left(\frac{4\theta - u}{3\theta} \right)^{n-1} I_{[\theta,4\theta]}(u)$$

y por lo tanto

$$E[X_{(1)}] = \int_{\theta}^{4\theta} n \frac{u}{(3\theta)^n} (4\theta - u)^{n-1} du$$

expresión que puede calcularse haciendo el cambio $z = 4\theta - n$, con lo que

$$\begin{aligned} E[X_{(1)}] &= \frac{n}{(3\theta)^n} \int_0^{3\theta} (4\theta - z) z^{n-1} dz \\ &= \theta \frac{n+4}{n+1}. \end{aligned}$$

Para el cálculo de la función de densidad del máximo $X_{(n)}$ se utiliza

$$\begin{aligned} F_{X_{(n)}}(v) &= P\{X_{(n)} \leq v\} = P^n\{X \leq v\} \\ &= \left(\int_\theta^v \frac{1}{3\theta} dv\right)^n = \left(\frac{v-\theta}{3\theta}\right)^n \end{aligned}$$

de donde

$$f_{X_{(n)}}(v) = \frac{n}{(3\theta)^n}(v-\theta)^{n-1} I_{[\theta,4\theta]}(v)$$

por lo tanto la esperanza de $X_{(n)}$ es

$$\begin{aligned} E[X_{(n)}] &= \int_\theta^{4\theta} v \frac{n}{(3\theta)^n}(v-\theta)^{n-1} dv \\ &= \frac{n}{(3\theta)^n} \int_\theta^{3\theta} v(v-\theta)^{n-1} dv \end{aligned}$$

que con el cambio $z = v - \theta$ puede ponerse

$$E[X_{(n)}] = \theta \frac{4n+1}{n+1}.$$

Por lo que se sigue que la función $g(X_{(1)}, X_{(n)})$ tiene de media cero.

2.8.4

a) La función de densidad de la muestra es

$$f(x_1, ..., x_n|\theta) = e^{-\sum_{i=1}^n x_i + n\theta} I_{(-\infty, x_{(1)}]}(\theta)$$

de donde de acuerdo con el Teorema 2.2.1 (factorización) es $X_{(1)}$ un estadístico suficiente. Para comprobar que es completo, su función de distribución es

$$\begin{aligned} F_{X_{(1)}}(y) &= P\{X_{(1)} \leq y\} = 1 - P\{X_{(1)} > y\} = 1 - P^n\{X > y\} \\ &= 1 - (1 - F(y))^n \end{aligned}$$

con $F(y) = 1 - e^{-y+\theta}$, para $y \geq \theta$, de donde $f(y) = ne^{n(\theta-y)} I_{[\theta,+\infty)}(y)$. Se debe comprobar si de la expresión

$$\int_\theta^{+\infty} g(y) f(y) dy = 0$$

para todo $\theta > 0$ se sigue que la función g es nula casi seguro P_θ. Pero dicha expresión implica que ha de ser

$$\int_\theta^{+\infty} g(y) e^{-ny} dy = 0 \text{ para todo } \theta,$$

sin más que aplicar, que en general si la función $f(x,\theta)$ es integrable respecto a x

$$\frac{d}{d\theta}\int_{a(\theta)}^{b(\theta)} f(x,\theta)dx = f(b(\theta),\theta)b'(\theta) - f(a(\theta),\theta)a'(\theta) + \int_{a(\theta)}^{b(\theta)} \frac{\partial}{\partial\theta} f(x,\theta)dx.$$

De donde se sigue que $g(\theta)e^{-n\theta} = 0$ para todo θ, por lo que $g(\theta)$ ha de ser cero. Se hace notar que se ha hecho la comprobación solo para las funciones $g(y)$ que sean integrables Riemann, desde un punto de vista práctico esta matización no tiene interés, ya que las funciones que son integrables Riemann son virtualmente cualquier función que se pueda ocurrir, una demostración más rigurosa ha de basarse en la Teoría de la Medida y se saldría fuera de los prerrequisitos previstos para este libro.

b) En virtud de la parte anterior bastará comprobar que la cuasivarianza muestal es un estadístico ancilario para θ, ver definición 8.3.1, ya que de acuerdo con el Teorema 2.3.2 (de Basu) un estadístico suficiente y completo $X_{(1)}$ y un estadístico ancilario S^2 , han de ser independientes.

Si se expresa la cuasivarianza muestral como

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \theta - (\overline{X} - \theta))^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \overline{Z})^2$$

las variables aleatorias Z_i tienen distribución independiente de θ ya que es

$$f(z|\theta) = f(x|0) = e^{-x}I_{[0,+\infty)}(x)$$

como se comprueba fácilmente.

2.8.5

Mediante la definición de suficiencia, basta ver que

$$f(x_1,...,x_n|x_1,...,x_{n-1}) = \frac{f(x_1,...,x_n)f(x_1,...,x_{n-1}|x_1,...,x_n)}{f(x_1,...,x_{n-1})} = \frac{f(x_1|\theta)...f(x_n|\theta)}{f(x_1|\theta)...f(x_{n-1}|\theta)} = f(x_n|\theta)$$

la distribución condicionada depende de θ; luego el estadístico no es suficiente.

2.8.6

De acuerdo con el Teorema 2.2.2 (caracterización de estadísticos minimal

suficientes), el cociente

$$\frac{f(x_1,...,x_n|\theta)}{f(y_1,...,y_n|\theta)} = \frac{\prod_{i=1}^{n} x_i e^{-\frac{1}{2\theta}\sum_{i=1}^{n} x_i^2}}{\prod_{i=1}^{n} y_i e^{-\frac{1}{2\theta}\sum_{i=1}^{n} y_i^2}}$$

es independiente de θ cuando $\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i^2$. Luego es

$$T(X_1,...,X_n) = \sum_{i=1}^{n} X_i^2$$

el estadístico minimal suficiente, como el estadístico $\sum_{i=1}^{n} X_i$ no es función de $\sum_{i=1}^{n} X_i^2$, ya que los puntos $\overrightarrow{x} = (-1,0,...,0)$ e $\overrightarrow{y} = (0,1,0,...,0)$ dan la misma imagen por el segundo estadístico y son distintos para el primero, se tiene que el estadístico $\sum_{i=1}^{n} X_i$ no es suficiente.

2.8.7

a) Por las propiedades de T_1 y T_2, se sigue que

$$\frac{T_1(x_1,...,x_n)}{T_2(x_1,...,x_n)} = \frac{T_1(ax_1+b,...,ax_n+b)}{T_2(ax_1+b,...,ax_n+b)}$$

si se hace el cambio a las variables $Z_i = \frac{X_i-\mu}{\sigma}$ cuya densidad es $f(x|0,1)$ se tiene que el cociente

$$\frac{T_1(a(\sigma z_1+\mu)+b,...,a(\sigma z_n+\mu)b)}{T_2(a(\sigma z_1+\mu)+b,...,a(\sigma z_n+\mu)+b)} = \frac{a\sigma T_1(z_1,...,z_n)}{a\sigma T_2(z_1,...,z_n)}$$

es ancilario para (μ,σ).

b) Mediante la notación

$$R(X_1,...,X_n) = X_{(n)} - X_{(1)}$$
$$S(X_1,...,X_n) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}$$

se cumple la parte a) para $T_1 = R \quad T_2 = S$, ya que

$$\begin{aligned} R(aX_1+b,...,aX_n+b) &= aX_{(n)} + b - (aX_{(1)}+b) \\ &= a(X_{(n)} - X_{(1)}) = aR(X_1,...,X_n) \end{aligned}$$

$$S(aX_1+b,...,aX_n+b)=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(aX_i-a\overline{X})^2}$$
$$=a\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}=aS(X_1,...,X_n)$$

y por lo tanto R/S es ancilario para (μ,σ).

2.8.8
Si primero se observa $N=n$ y a continuación se realizan n experimentos de $Bernoulli(\theta)$, es $p(n,x)=p_n\theta^x(1-\theta)^{n-x}$ donde $n\geq\alpha$ y $p_n=P\{N=n\}$. De otro lado, si directamente se observan x éxitos en n experimentos de $Bernoulli(\theta)$, es $p(x)=\binom{n}{x}\theta^x(1-\theta)^{n-x}$ se sigue que ambas funciones de verosimilitud difieren en $\binom{n}{x}p_n^{-1}$ que no depende de θ, luego $Ev(E_1,(x,n))=Ev(E_2,x)$.

2.8.9
El estadístico $g(X)=X$ tiene de media cero y no es nulo

$$E_\theta[X]=-\theta+(1-\theta)^2\theta\sum_{i=1}^{\infty}i\theta^{i-1}=-\theta+(1-\theta)^2\theta S'(\theta)$$

donde

$$S(\theta)=\sum_{i=1}^{\infty}\theta^i=\frac{\theta}{1-\theta}$$

por lo que $S'(\theta)=\frac{1}{(1-\theta)^2}$

$$E_\theta[X]=-\theta+(1-\theta)^2\theta\frac{1}{(1-\theta)^2}=0$$

luego X no es un estadístico completo.

2.8.10
Es suficiente, ya que

$$f(x_1,...,x_n|\theta)=\frac{1}{(\sqrt{2\pi}\sigma)^n}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\alpha\sigma)^2}$$
$$=\frac{1}{(\sqrt{2\pi}\sigma)^n}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2-\frac{n\alpha^2}{2}+\frac{\alpha}{\sigma}\sum_{i=1}^{n}x_i}$$

y por el Teorema 2.2.1 (factorización)

$$\overrightarrow{T}(X_1,...,X_n)=\left(\sum_{i=1}^{n}X_i,\sum_{i=1}^{n}X_i^2\right)$$

es un estadístico suficiente.

Sin embargo no es completo para σ, ya que como

$$E\left[\sum_{i=1}^{n} X_i\right] = n\alpha\sigma$$

$$E\left[\left(\sum_{i=1}^{n} X_i\right)^2\right] = E\left[\sum_{i=1}^{n} X_i^2 + \sum_{\substack{i=1\\ i\neq j}}^{n} X_i X_j\right] = (1+n\sigma^2)\sigma^2 n$$

la función

$$g(\overrightarrow{T}) = \frac{1}{1+n\alpha^2}\left(\sum_{i=1}^{n} X_i\right)^2 - \frac{1}{1+\alpha^2}\sum_{i=1}^{n} X_i^2$$

no es nula y tiene de media cero para todo σ.

2.8.11

a) Como la función de densidad de la muestra es

$$f(x_1,...,x_n|\theta) = I_{[\theta,x_{(n)}]}(x_{(1)})I_{[x_{(1)},\theta+1]}(x_{(n)})$$

es claro, por el teorema de factorización, Teorema 2.2.1 que el estadístico $T(X_1,...,X_n) = (X_{(1)}, X_{(n)})$ es suficiente.

b) Se debe encontrar la distribución del recorrido muestral, esto puede hacerse calculando la distribución conjunta de $(X_{(1)}, X_{(n)})$ y por ejemplo un cambio a $(R, X_{(1)})$, para pasar a obtener la distribución marginal de R. Sin embargo es mas cómodo observar que

$$P\{R \leq r\} = \int_{\theta}^{\theta+1} f_{X_{(n)}}(x) P\left\{x - X_{(1)} \leq r\right\} dx$$

como

$$\begin{aligned} P\{X_{(n)} \leq x\} &= P^n\{X \leq x\} \\ &= (x-\theta)^n \qquad \text{si} \quad x \in (\theta, \theta+1) \end{aligned}$$

resulta

$$f_{X_{(n)}}(x) = n(x-\theta)^{n-1} I_{[\theta,\theta+1]}(x)$$

además

$$\begin{aligned} P\{X_{(1)} \geq x-r\} &= P^n\{X \geq x-r\} \\ &= (\theta+1-x+r)^n \quad \text{si} \quad x \in (\theta, \theta+1) \end{aligned}$$

es

$$P\{R \leq r\} = n\int_{\theta}^{\theta+1} (x-\theta)^{n-1}(\theta+1-x+r)^n dx$$

expresión que mediante el cambio $y = x - \theta$ se convierte en

$$P\{R \leq r\} = n\int_{0}^{1} y^{n-1}(1-y+)^n dy$$

que no depende de θ, luego el estadístico recorrido muestral es ancilario.

c) Desde luego que en el caso de la distribución uniforme $U(\theta, \theta+1)$ con $\theta > 0$ es $F(x|\theta) = I_{[0,1]}(x-\theta)$, luego θ es un parámetro de localización. En el caso general en que $F(x|\theta) = F(x-\theta)$, se puede observar que

$$P\{R \leq r|\theta\} = P\{X_{(n)} - X_{(1)} \leq r|\theta\}$$

si se hace el cambio $Z_i = X_i - \theta$, se tiene que $Z_{(n)} - Z_{(1)} = X_{(n)} - X_{(1)}$ y evidentemente que la distribución de las variables Z_i es independientes de θ ya que

$$P\{Z \leq z\} = P\{X - \theta \leq z\} = P\{X \leq z + \theta\} = F_X(z+\theta) = F(z)$$

donde esta última distribución es $F(x|\theta)$, que obviamente es independiente de θ. Por último si la distribución de las v.a. Z_i es independientes de θ también lo es la distribución de $Z_{(1)}$ y de $Z_{(n)}$.

2.8.12

Para ver que $\overline{X}$ es un estimador insesgado, como $E_\theta[\overline{X}] = E_\theta[X]$ se tiene

$$E_\theta[X] = \int x f(x-\theta) dx$$

si se hace el cambio $y = x - \theta$

$$\begin{aligned} E_\theta[X] &= \int (y+\theta) f(y) dy \\ &= \theta \end{aligned}$$

al ser $E_0[X] = 0$.

Para ver que $\overline{X}$ es invariante se observa que si X se distribuye según θ, $g_a(X) = X + a$ se distribuye mediante $\overline{g}_a(\theta) = a + \theta$, luego

$$\begin{aligned} T g_a(\overrightarrow{X}) &= T(X_1 + a, ..., X_n + a) \\ &= \overline{X}_n + a \end{aligned}$$

y por otro lado

$$\begin{aligned} \overline{g}_a T(\overrightarrow{X}) &= \overline{g}_a(\overline{X}) \\ &= \overline{X}_n + a. \end{aligned}$$

2.8.13

La transformación que se corresponde con $g_{a,c}(x) = cx + a$ es $\overline{g}_{a,c}(\theta) = a + \theta c$, ya que la densidad de $y = cx + a$ es

$$\begin{aligned} g(y) &= f(x(y)) \left| \frac{dx}{dy} \right| \\ &= \frac{1}{\sigma} f\left(\frac{y - (a + \theta c)}{\sigma c} \right) \frac{1}{|c|} \\ &= \frac{1}{\sigma c} f\left(\frac{y - (a + \theta c)}{\sigma c} \right). \end{aligned}$$

Por lo tanto, para ver si el estimador $\overline{X}+k$ es invariante, se tiene que

$$Tg_{a,c}(\overrightarrow{X}) = c\overline{X} + a + k,$$

por otro lado

$$\begin{aligned}\overline{g}_{a,c}T(\overrightarrow{X}) &= \overline{g}_{a,c}(\overline{X}+k)\\ &= a + (\overline{X}+k)c\end{aligned}$$

con lo que el estimador $\overline{X}+k$ no es invariante si k es distinto de cero.

Por el contrario, la transformación $g_a(x) = x + a$, es claro que $\overline{g}_a(\theta) = a + \theta$, de donde

$$\begin{aligned}Tg_a(\overrightarrow{X}) &= T(X_1 + a, ..., X_n + a)\\ &= \overline{X} + k + a,\end{aligned}$$

mientras que

$$\begin{aligned}\overline{g}_a T(\overrightarrow{X}) &= \overline{g}_a(\overline{X}+k)\\ &= a + \overline{X} + k\end{aligned}$$

y se sigue, que los estimadores de la forma $\overline{X}+k$, son invariantes por el grupo $\mathcal{G}_2$.

2.8.14

a) X no es un estadístico completo, ya que

$$g(-1)\frac{\theta}{2} + g(0)(1-\theta) + g(1)\frac{\theta}{2} = 0$$

es equivalente, a que para todo θ ha de ser

$$2g(0) + \theta[g(-1) - 2g(0) + g(1)] = 0$$

con lo que $g(0) = 0$, $g(-1) = -g(1)$. Por tanto la función $g(x) = |x|$ no es nula y tiene de media cero.

b) Por el Teorema 2.2.2 (caracterización de estadísticos minimal suficientes), la relación que permite colocar a dos puntos x, y en la misma clase, debe ser

$$\frac{\left(\frac{\theta}{2}\right)^{|x|}(1-\theta)^{1-|x|}}{\left(\frac{\theta}{2}\right)^{|y|}(1-\theta)^{1-|y|}} = \theta^{|x|-|y|}(1-\theta)^{|y|-|x|}$$

y esta expresión es independiente de θ, si y solamente si $|x| = |y|$, luego el estadístico $|X|$ es minimal suficiente. Además este estadístico es completo, ya que toma los valores

$$|X| = \begin{cases} 0 & \text{con probabilidad } 1-\theta \\ 1 & \text{con probabilidad } \theta \end{cases}$$

luego si existe $g(|x|)$ de media cero para todo θ, ha de verificar

$$g(0)(1-\theta)+g(1)\theta=0$$

de donde $g(0)=0$ y $g(1)=g(0)$; g es la función cero.

c) El estadístico Z no es ancilario ya que su distribución es

$$Z=\begin{cases} -1 & \text{con la} \quad P\{X=-1\}=\frac{\theta}{2} \\ 1 & \text{con la} \quad P\{X=1\}=\frac{\theta}{2} \\ 0 & \text{con la} \quad P\{X=0\}=1-\theta \end{cases}$$

que obviamente depende de θ.

2.8.15

a) La función de densidad muestral viene dada por

$$f(x_1,...,x_n|\theta)=\frac{(\ln\theta)^n\theta^{\sum_1^n x_i}}{(\theta-1)^n}\prod_{i=1}^n I_{(0,1)}(x_i)$$

pertenece a la familia exponencial uniparamétrica, es decir es de la forma

$$c(\theta)h(\overrightarrow{x})e^{c_1(\theta)T_1(\overrightarrow{x})}$$

con $c_1(\theta)=\ln\theta$ y $T_1(\overrightarrow{x})=\sum_{i=1}^n x_i$, por lo que de acuerdo con el Teorema 2.2.1 (factorización), T_1 es un estadístico suficiente.

b) El estadístico $T_1(\overrightarrow{x})=\sum_{i=1}^n X_i$ es minimal suficiente, ya que por pertenecer la distribución muestral a la exponencial uniparamétrica con $c_1(\theta)=\ln\theta \quad \theta>1$, contiene un vector de dimensión uno linealmente independiente, es decir un escalar distinto de cero, véase el Teorema 2.2.3.

Por último, T_1 es completo, ya que de acuerdo con el Teorema 2.3.1 (existencia de estadísticos completos), la imagen de $\theta>1$ por $\ln\theta$ contiene un rectángulo de $\mathbb{R}^1$ es decir, un intervalo.

2.8.16

a) La función de densidad muestral es

$$\begin{aligned} f(x_1,...,x_n|\theta) &= \left(\frac{2}{\theta^2}\right)^n \prod_{i=1}^n x_i \prod_{i=1}^n I_{(0,\theta)}(x_i) \\ &= \left(\frac{2}{\theta^2}\right)^n \prod_{i=1}^n x_i I_{(X_{(n)},+\infty)}(\theta) \end{aligned}$$

luego de acuerdo con el Teorema 2.2.1 (factorización), $X_{(n)}$ es un estadístico suficiente y si se añade el estadístico $\prod_{i=1}^n X_i$ también es suficiente.

De todas formas parece intuitivo que la información que suministra este último es redundante y que el estadístico minimal suficiente va a ser $X_{(n)}$. En efecto si se aplica la condición necesaria y suficiente de construcción de la partición minimal suficiente, consiste en hacer que la siguiente expresión sea independiente de θ

$$\frac{f(\overrightarrow{x}|\theta)}{f(\overrightarrow{y}|\theta)} = \frac{\prod_{i=1}^{n} x_i}{\prod_{i=1}^{n} y_i} \frac{I_{(X_{(n)},+\infty)}(\theta)}{I_{(Y_{(n)},+\infty)}(\theta)}$$

que obliga a que $X_{(n)} = Y_{(n)}$.

b) Para ver la completitud, al depender el soporte de la distribución muestral de θ, no se está en la familia exponencial, luego se procede a comprobarlo directamente pasando por la distribución de $T = X_{(n)}$

$$F_{X_{(n)}}(t) = \begin{cases} 0 & \text{si} \quad t \leq 0 \\ \left(\frac{t}{\theta}\right)^{2n} & \text{si} \quad t \in (0,\theta) \\ 1 & \text{si} \quad t \geq \theta \end{cases}$$

luego se trata de ver, si de la expresión

$$\int_0^{\theta} g(t) f_{X_{(n)}}(t) dt = 0 \quad \forall \theta > 0$$

se sigue que g es la función cero. Es decir se tiene que

$$\int_0^{\theta} t^{2n-1} g(t) dt = 0 \quad \forall \theta > 0.$$

Si se deriva respecto de θ y se utiliza la regla de derivación dada en el Ejercicio 2.8.4, parte a), se obtiene que ha de ser

$$g(\theta)\, \theta^{2n-1} = 0 \quad \forall \theta > 0.$$

Ahora bien, como θ^{2n-1} es un polinomio de grado $2n-1$, a lo más tiene $2n-1$ soluciones luego para que el producto sea cero para todo θ positivo, ha de ser $g(\theta) = 0$, por lo que $X_{(n)}$ es un estadístico completo.

Capítulo 3

ESTIMACIÓN PUNTUAL PARAMÉTRICA

3.1. INTRODUCCIÓN

En los capítulos anteriores se ha visto, cómo reducir adecuadamente la información suministrada por la muestra. En éste capítulo se va a estudiar cómo se puede estimar el valor de θ desconocido con el que ésta se distribuye. Para ello dada la familia de distribuciones de probabilidad mediante $\mathcal{P} = \{f(\vec{x}|\theta)\,|\theta \in \Theta\}$ se introduce la siguiente definición:

Definición 3.1.1 (estimador por punto)
Dada la muestra $(X_1, ..., X_n)$ de una población con distribución de probabilidad que se supone pertenece a $\mathcal{P} = \{f(x_1, ..., x_n|\theta)|\theta \in \Theta\}$, se llama estimador por punto de θ a cualquier función medible T que aplique el espacio muestral $\mathcal{X}$, $\mathcal{X} = (X_1, ..., X_n)$ de muestras posibles, en el espacio paramétrico, es decir

$$T : \mathcal{X} \longrightarrow \Theta$$

■

En algún contexto se puede estar interesado en una función del parámetro $h(\theta)$ en lugar del propio valor θ, en estos casos se supondrá automáticamente la sustitución de θ por $h(\theta)$.

Un estimador por punto, es pues un estadístico que se emplea para determinar el valor desconocido θ.

Ejemplo 3.1.1
Si la familia de distribuciones de probabilidad es $\mathcal{P} = \{\theta^x(1-\theta)^x I_{\{0,1\}}(x)|\theta \in (0,1)\}$, con m.a.s. de tamaño n, un estimador por punto para θ es

$$T(X_1, ..., X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

■

La Definición 3.1.1. es válida para el caso en que el parámetro de interés sea k-dimensional, sin más que considerar T y Θ como k-dimensionales.

Ejemplo 3.1.2
Si $\mathcal{P} = \left\{ \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n e^{-\frac{1}{2\sigma^2}\sum(X_i-\mu)^2} | \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+ \right\}$ un estimador puntual para $\overrightarrow{\theta} = (\mu, \sigma^2)$ viene dado por $\overrightarrow{T}(X_1, ..., X_n) = (\overline{X}, S^2)$, con

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{y} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

■

En muchos casos existirán candidatos obvios a ser estimadores, pero con frecuencia también se dispondrá de distintos estimadores para un mismo problema.

3.2. PROPIEDADES DE LOS ESTIMADORES

En esta sección se introducen una serie de propiedades, que en determinados contextos, parecen razonables que tengan los estimadores. No obstante, no hay una mejor propiedad a solicitar y en general distintos marcos reales aconsejan demandar distintas propiedades.

3.2.1. Estimadores centrados

Parece deseable que si se utiliza un estimador varias veces en una misma situación, la media del valor que se utiliza respecto a la distribución de la muestra, sea el valor que se quiere estimar.

Definición 3.2.1.1 (estimador centrado)
T es un estimador centrado para $h(\theta)$ cuando $E_\theta T(\overrightarrow{X}) = h(\theta)$ para todo $\theta \in \Theta$.

En general se llama sesgo de un estimador para θ a la diferencia

$$b_\theta(T) = E_\theta T(\overrightarrow{X}) - \theta$$

Ejemplo 3.2.1.1
Para cualquier población que tenga media, la media muestral $\overline{X}$ es un estimador centrado de $h(\theta) = E[X]$ puesto que $E_\theta \overline{X} = E[X]$.

De hecho se tiene que $E[a_k] = \alpha_k$, luego el momento muestral de orden k respecto al origen, es un estimador centrado del momento poblacional de orden k respecto al origen de la población. En general no es cierto un resultado análogo para los momentos respecto al centro ya que para el momento de orden 2 respecto al centro precisamente se ha introducido la cuasivarianza

muestral $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$, para que $E[S^2] = \sigma^2$, véase la Proposición 1.2.2.

Ejemplo 3.2.1.2
Para muestras de tamaño uno de una población de *Bernoulli*(θ), con $\theta \in (0,1)$, no existe estimador centrado para $h(\theta) = \theta^2$, pues debería ser

$$T(1)\theta + T(0)(1-\theta) = \theta^2$$

para todo θ, lo que es obviamente imposible, ya que un polinomio de grado dos puede tener a lo más dos raíces distintas.

Ejemplo 3.2.1.3
Para muestras de tamaño uno de una población de *Bernoulli*(θ), con $\theta \in [0,1]$ existe un estimador centrado para $h(\theta)$=θ, ya que ha de ser,

$$T(1)\theta + T(0)(1-\theta) = \theta$$

para todo θ, y esto lo satisface el estimador $T(0) = 0$, $T(1) = 1$.

■

Los dos últimos ejemplos, ponen de manifiesto que pueden no existir estimadores centrados y que si T es un estimador centrado para θ, $h(T)$ no tiene por qué ser un estimador centrado para $h(\theta)$.

El siguiente ejemplo muestra que, en determinadas situciones, un estimador centrado puede resultar absurdo.

Ejemplo 3.2.1.4
Para una población de Poisson(θ), con $\theta \in (0,\infty)$, si se quiere estimar la función $h(\theta) = e^{-3\theta}$ con muestras de tamaño uno mediante un estimador centrado T, éste debe cumplir

$$\sum_{i=0}^{\infty} e^{-\theta}\frac{\theta^i}{i!}T(i) = e^{-3\theta}$$

o lo que es lo mismo

$$\sum_{i=0}^{\infty}\frac{\theta^i}{i!}T(i) = \sum_{i=0}^{\infty}\frac{\theta^i}{i!}(-2)^i$$

con lo que $T(i) = (-2)^i$ es el estimador centrado que es negativo para los valores impares de i, con lo que de hecho T no es un estimador de $h(\theta)$

■

El que un estimador sea insesgado depende de su distribución en el muestreo, sin embargo cuando se emplea el valor particular del estimador, éste sólo depende de la muestra particular observada. La concreción matemática de esta idea, lleva a la noción del *principio de verosimilitud*, véase la Sección 2.5.1, que dice básicamente, que un estimador sólo debe depender del

valor $\overrightarrow{x}$ observado. Se recoge a continuación el siguiente ejemplo relacionado con estas ideas.

Ejemplo 3.2.1.5
Para estimar el voltaje promedio θ de un circuito eléctrico se realiza una observación de una v.a. X con distribución $N(\theta, \sigma_0)$ con σ_0 conocida, observándose que la lectura es de 200 voltios. Un estimador centrado de θ es $\overline{X}_n$ por tanto cuando el tamaño es $n = 1$ la propia observación X es el estimador centrado. Supongamos que al examinar el aparato más detenidamente se ve que no marcaba por encima de 250 voltios, es decir que la lectura tiene una distribución normal truncada por $X \leq 250$, al haber cambiado la distribución de la población, X ya no es un estimador centrado de θ, a pesar de que seguro que la lectura de 200 ha sido hecha correctamente.

■

Por otro lado es conveniente en algunas situaciones trabajar con estimadores centrados o faltos de sesgo. Así en una situación en la que se disponga de los estimadores $T_1, T_2, ..., T_n$ si se reúne la estimación de θ, que da cada uno de ellos, en el estimador $\overline{T} = \frac{1}{n}\sum_{i=1}^{n} T_i$, y todos son insesgados para θ, $\overline{T}$ también lo es. Como $V[\overline{T}] = \frac{1}{n^2}\sum_{i=1}^{n} V[T_i]$ si todas las varianzas de los estimadores están acotadas, por ejemplo por σ^2, se sigue que $V[\overline{T}] < \frac{\sigma^2}{n}$ expresión que tiende a cero cuando n aumenta, por lo que la distribución de $\overline{T}$ se concentra en el valor θ que se quiere estimar; por el contrario si los estimadores hubieran tenido sesgo, el estimador $\overline{T}$ se hubiera concentrado alrededor de su media $\theta + b(\theta)$ y habría sido inadecuado.

3.2.2. Estimadores consistentes

Otra propiedad deseable que puede exigirse a un estimador, es que tienda al valor que se quiere estimar, cuando el tamaño muestral crece. Es importante subrayar dos aspectos relacionados con esta propiedad, el primero, es que se trata de una propiedad asintótica y el segundo, que depende de la manera en que se concrete el tipo de convergencia a utilizar. Es usual utilizar la convergencia en probabilidad y en este sentido se utiliza la siguiente definición:

Definición 3.2.2.1 (estimador consistente)
Una sucesión de estimadores $T_n = T(X_1, ..., X_n)$ es consistente para el parámetro θ si para todo $\theta \in \Theta$ la sucesión T_n converge en probabilidad θ a θ, es decir, si para cualquier ε positivo se tiene

$$\lim_{n\to\infty} P\{|T_n - \theta| \leq \varepsilon|\theta\} = 1$$

para todo $\theta \in \Theta$.

Ejemplo 3.2.2.1
Si el parámetro θ es el momento poblacional respecto al origen de orden k, es decir $\theta = \alpha_k$, entonces el momento muestral respecto al origen de orden k $a_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ es un estimador consistente.

■

Una condición suficiente de consistencia viene dada por el siguiente enunciado:

Proposición 3.2.2.1
Si T_n es una sucesión de estimadores tales que para todo θ

$$\lim_{n\to\infty} E_\theta[T_n] = \theta \qquad \lim_{n\to\infty} V_\theta[T_n] = 0$$

entonces la sucesión de estimadores es consistente.

Demostración: Ya que para cualquier ε se tiene

$$P\{|T_n - \theta| > \varepsilon|\theta\} \leq \varepsilon^{-2} E_\theta[(T_n - \theta)^2]$$

y si se suma y se resta $E_\theta[T_n]$ en la expresión del cuadrado del segundo miembro, se obtiene que

$$P\{|T_n - \theta| > \varepsilon|\theta\} \leq \varepsilon^{-2}[V_\theta[T_n] + (ET_n - \theta)^2]$$

expresión que, en virtud de las hipótesis, tiende a cero cuando n tiende a infinito.

■

Se observa que la proposición dice que si la sucesión de estimadores tiene asintóticamente sesgo cero, es decir, $\lim_{n\to\infty} b(T_n, \theta) = 0$, y la varianza de la sucesión tiende a cero; entonces la sucesión de estimadores es consistente.

Ejemplo 3.2.2.2
Con m.a.s. de tamaño n de una población de Bernoulli(θ), el estimador

$$T_n = \frac{1}{n+2}\left(\sum_{1}^{n} X_i + 1\right)$$

es consistente para θ. Ya que

$$\lim_{n\to\infty} E[T_n] = \lim_{n\to\infty} \frac{n\theta + 1}{n+2} = \theta$$

y

$$\lim_{n\to\infty} V[T_n] = \lim_{n\to\infty} \frac{n\theta(1-\theta)}{(n+2)^2} = 0$$

por lo que aplicando la Proposición 3.2.2.1 el estimador es consistente.

3.2.3. Estimadores bayesianos

En algunas situaciones, se puede tener información suficiente sobre el parámetro θ, como para poder considerar a este una variable aleatoria. Inicialmente la incertidumbre sobre θ puede modelizarse por una distribución de probabilidad, que recibe el nombre de *distribución inicial* o *a priori*, y que se denotará por $\pi(\theta)$, una función de masa o de densidad según proceda. Cuando se observa una muestra, se calcula la *distribución final* o *a posteriori* mediante

$$\pi(\theta|x_1,...,x_n) = \frac{\pi(\theta)f(x_1,...,x_n|\theta)}{m(x_1,...,x_n)} \tag{3.2.3.1}$$

donde m es la llamada *distribución predictiva*, dada por

$$m(x_1,...,x_n) = \int_\Theta \pi(\theta)f(x_1,...,x_n|\theta)d\theta$$

La idea es que antes de tomar la muestra, la información sobre θ viene dada por $\pi(\theta)$ y tras la experimentación se debe utilizar $\pi(\theta|x_1,...,x_n)$. El estimador bayesiano de θ es toda la distribución final y por extensión cualquier medida de centralización correspondiente a esta distribución.

Ejemplo 3.2.3.1
Si la información sobre θ para una población de Bernoulli(θ) viene recogida por $\pi(\theta) = I_{(0,1)}(\theta)$, tras de observar una m.a.s. de tamaño n la distribución final viene dada por

$$\begin{aligned}\pi(\theta|x_1,...,x_n) &= \frac{\theta^{\sum_1^n x_i}(1-\theta)^{n-\sum_1^n x_i}I_{(0,1)}\theta}{\int_0^1 \theta^{\sum_1^n x_i}(1-\theta)^{n-\sum_1^n x_i}d\theta} \\ &= \frac{\theta^{\sum_1^n x_i}(1-\theta)^{n-\sum_1^n x_i}I_{(0,1)}(\theta)}{B(\sum_1^n x_i+1, n-\sum_1^n x_i+1)}\end{aligned}$$

luego la distribución a posteriori es una $Beta(\theta|\sum_1^n x_i + 1, n - \sum_1^n x_i + 1)$.

■

El sentido de los estimadores bayesianos es que toda la información sobre θ, una vez que se observa la nuestra está contenida en la distribución final. Es decir, para los estimadores bayesianos la distribución en el muestreo no tiene demasiado interés, el protagonista es la distribución final y se trabaja condicionado por la muestra $(x_1,...,x_n)$ particular obtenida. Véase el principio de verosimilitud en la Sección 2.5.1.

Todos los elementos que han sido estudiados en el capítulo anterior estaban referidos a la distribución en el muestreo, por lo tanto no se aplican

a los estimadores bayesianos. Lo único que hay que añadir es que, dentro de los estadísticos, el concepto de estadístico suficiente admite una definición diferente más adecuada al contexto bayesiano pero que es equivalente en condiciones muy generales al concepto de suficiencia desarrollado en la Sección 2.2, véase Lehman (1986) pág. 21, y que se pasa a estudiar.

Definición 3.2.3.1 (estadístico suficiente bayesiano)
$T(X_1, ..., X_n)$ es un estadístico suficiente bayesiano para la familia $\mathcal{P} = \{f(\overrightarrow{x}|\theta)|\theta \in \Theta\}$ si, cualquiera que sea la distribución inicial $\pi(\theta)$, se tiene que la distribución final dada por la muestra y por el valor del estadístico son la misma. Es decir, $\pi(\theta|x_1, ..., x_n) = \pi(\theta|t)$ con $T(x_1, ..., x_n) = t$.

■

Así, en el Ejemplo 2.2.6 para el modelo de Bernoulli(θ) el estadístico $T(x_1, ..., x_n) = \sum_{i=1}^{n} x_i$ es suficiente en el sentido bayesiano respecto a la distribución inicial $\pi(\theta) = I_{(0,1)}(\theta)$, pues

$$\begin{aligned} \pi(\theta|t) &= \frac{\pi(\theta)f(t|\theta)}{m(t)} \\ &= \frac{I_{(0,1)}(\theta)\binom{n}{t}\theta^t(1-\theta)^{n-t}}{\int_0^1 \binom{n}{t}\theta^t(1-\theta)^{n-t}d\theta} \end{aligned}$$

expresión que coincide con $\pi(\theta|x_1, ..., x_n)$.

De todas formas ambas definiciones de suficiencia son equivalentes.

Proposición 3.2.3.1
Decir que $T(x_1, ..., x_n)$ es suficiente para θ en el sentido de que la distribución de probabilidad de la muestra condicionada por el estadístico es independiente de θ, es equivalente a que para cualquier distribución inicial $\pi(\theta)$ sea $\pi(\theta|x_1, ..., x_n) = \pi(\theta|t)$ con $T(x_1, ..., x_n) = t$.

Demostración: Si T es suficiente de acuerdo con la Definición 2.2.1, teniendo en cuenta que $f(\overrightarrow{x}|\theta) = f(t|\theta)f(\overrightarrow{x}|t, \theta)$ y que $f(\overrightarrow{x}|t, \theta)$ no depende de θ, se tiene

$$\pi(\theta|x_1, ..., x_n) = \frac{\pi(\theta)f(t|\theta)f(\overrightarrow{x}|t, \theta)}{\int_\Theta \pi(\theta)f(t|\theta)f(\overrightarrow{x}|t, \theta)d\theta}$$

de donde

$$\pi(\theta|x_1, ..., x_n) = \frac{\pi(\theta)f(t|\theta)}{\int_\Theta \pi(\theta)f(t|\theta)d\theta} = \pi(\theta|t).$$

Recíprocamente

$$\begin{aligned} f(x_1, ..., x_n|\theta) &= \frac{\pi(\theta|x_1,...,x_n)m(x_1,...,x_n)}{\pi(\theta)} \\ &= \frac{\pi(\theta|t)}{\pi(\theta)}m(x_1, ..., x_n) \end{aligned}$$

el primer factor de este producto depende de θ y de la muestra sólo a través de t y el segundo depende únicamente de la muestra, luego se tiene la factorización en función de T, con lo que T es suficiente.

■

A continuación se recoge el cálculo de las distribuciones finales para algunos modelos usuales.

Ejemplo 3.2.3.2

Para m.a.s. de tamaño n de una población de $Bernoulli(\theta)$, si se supone que la información inicial sobre θ viene dada por la función de densidad $Beta(\alpha_0, \beta_0)$, es decir

$$\pi(\theta) = \frac{1}{\beta(\alpha_0, \beta_0)} \theta^{\alpha_0 - 1}(1-\theta)^{\beta_0 - 1} \quad \text{con} \quad \theta \in (0,1),$$

como

$$f(x_1, ..., x_n|\theta) = \theta^{\sum_1^n x_i}(1-\theta)^{n - \sum_1^n x_i},$$

se obtiene como densidad final

$$\pi(\theta|x_1, ..., x_n) = \frac{\theta^{\alpha_0 + \sum x_i - 1}(1-\theta)^{\beta_0 + n - \sum x_i - 1}}{B(\alpha_0 + \sum x_i, \beta_0 + n - \sum x_i)}$$

que es una $Beta(\alpha_1, \beta_1)$ con parámetros

$$\alpha_1 = \alpha_0 + \sum x_i$$
$$\beta_1 = \beta_0 + n - \sum x_i$$

Ejemplo 3.2.3.3

Si la población es $N(\theta, \sigma)$ con σ conocida y la distribución inicial sobre θ es $N(\mu_0, \sigma_0)$, se tiene que

$$\pi(\theta) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{1}{2}\frac{(\theta - \mu_0)^2}{\sigma_0^2}}$$

$$f(x_1, ..., x_n|\theta) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum_1^n (x_i - \theta)^2}$$

con lo que la función de densidad final será, de acuerdo con la expresión

$$\pi(\theta|x_1, ..., x_n) = \frac{e^{-\frac{1}{2\sigma_0^2}(\theta - \mu_0)^2} e^{-\frac{1}{2\sigma^2}\sum (x_i - \theta)^2}}{\int_{-\infty}^{+\infty} e^{-\frac{1}{2\sigma_0^2}(\theta - \mu_0)^2} e^{-\frac{1}{2\sigma_0^2}\sum (x_i - \theta)^2} d\theta}.$$

Como por otra parte, dada una v.a. $N(E, \sqrt{V})$ se sigue que

$$\int_{-\infty}^{+\infty} e^{-\frac{1}{2V}\theta^2 + \frac{E}{V}\theta} d\theta = \sqrt{2\pi}\sqrt{V} e^{\frac{1}{2}\frac{E^2}{V}} \tag{3.2.3.2}$$

expresión que permite calcular la integral que aparece en el denominador de la distribución final, y como

$$\pi(\theta|\overrightarrow{x}) = \frac{e^{-\frac{1}{2\sigma_0^2}(\theta^2-2\mu_0\theta)} e^{-\frac{1}{2\sigma^2/n}(\theta^2-2\overline{x}\theta)^2}}{\int_{-\infty}^{+\infty} e^{-\frac{1}{2}(\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2/n})\theta^2+(\frac{\mu_0}{\sigma_0^2}+\frac{\overline{x}}{\sigma^2/n})\theta} d\theta}$$

se obtiene, con el concurso de (3.2.3.1), que

$$\frac{1}{V} = \frac{1}{\sigma_0^2} + \frac{1}{\sigma^2/n}$$
$$\frac{E}{V} = \frac{\mu_0}{\sigma_0^2} + \frac{\overline{x}.}{\sigma^2/n}$$

de donde la función de densidad final es $N(\mu_1, \sigma_1)$ con

$$\mu_1 = \frac{\frac{\mu_o}{\sigma_o^2} + \frac{\overline{x}}{\sigma^2/n}}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2/n}}$$
$$\sigma_1 = \frac{1}{\left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2/n}\right)^{1/2}}.$$

La media μ_1 de la distribución final se puede poner como media ponderada de la media de la distribución inicial y de la media muestral, en la forma

$$\mu_1 = \frac{\sigma^2}{\sigma^2 + n\sigma_0^2}\mu_0 + \frac{n\sigma_0^2}{\sigma^2 + n\sigma_0^2}\overline{x}.$$

Además si σ_0 tiende a infinito la media de la distribución final tiende a ser $\overline{x}$, es decir cuando la información inicial es muy pequeña, el efecto de la distribución a priori prácticamente no se refleja en la distribución final.

Cuando la varianza poblacional σ^2 tiende a infinito la media muestral, informa poco del valor de θ y la media de la distribución final, tiende a aproximarse a la media de la distribución inicial.

■

Como se ha visto en estos ejemplos, facilita bastante los cálculos, el hecho de que la distribución inicial y final pertenezcan a la misma familia de distribuciones. Por este motivo se introduce el siguiente concepto.

Definición 3.2.3.2 (familia conjugada)
Una familia de distribuciones iniciales $P = \{\pi(\theta|\overrightarrow{\alpha})|\overrightarrow{\alpha} \in \Lambda\}$ se dice conjugada de la familia de funciones de masa o de densidad de la población $\mathcal{P} = \{f(\overrightarrow{x}|\theta)|\theta \in \Theta\}$, cuando para cualquier distribución inicial perteneciente a P, la distribución final también pertenece a P cualquiera que sea la muestra obtenida. Es decir, si la distribución inicial es $\pi(\theta) = \pi(\theta|\overrightarrow{\alpha}_0)$ la distribución final es $\pi(\theta|\overrightarrow{x}) = \pi(\theta|\overrightarrow{\alpha}_1)$ con $\overrightarrow{\alpha}_0$ y $\overrightarrow{\alpha}_1$ pertenecientes a Λ.

■

De acuerdo con la definición, la familia Beta es conjugada de la familia de Bernoulli y la familia normal es conjugada de la familia normal. De todas formas, la elección de una familia conjugada para un problema particular, no debe apoyarse en la facilidad de cálculo, sino en ver si algún miembro de la familia conjugada recoge correctamente la información que se tiene sobre el parámetro. En la práctica el problema con la aproximación bayesiana suele ser que la utilización de distintas distribuciones iniciales, con la misma información muestral, lleva a diferentes distribuciones finales. Frecuentemente, si el tamaño muestral es grande, las distribuciones finales serán muy parecidas. Como pone de manifiesto el ejemplo siguiente.

Ejemplo 3.2.3.4
La proporción de enfermos θ en una determinada colectividad, puede ser modelizada mediante una v.a. con distribución Beta. Si de estudios previos puede suponerse que θ tiene distribución $Beta(\alpha = 5, \beta = 10)$ y al observar una m.a.s. de tamaño $n = 100$ se observan 20 individuos enfermos, el estimador bayesiano sería, de acuerdo con el Ejemplo 3.2.3.2, $\pi(\theta|\sum_1^{100} x_i = 20)$ una distribución $Beta(\theta|\alpha_1 = 25, \beta_1 = 90)$. Nótese que la distribución inicial $Beta(\theta|\alpha = 5, \beta = 10)$ cumple que $P\{\theta \leq 0'5\} = 0'91$, $P\{\theta \leq 0'3\} = 0'416$ y $P\{\theta \leq 0'8]\} = 0'199$. Si se hubieran escogido unos parámetros iniciales de $\alpha = 5$ y $\beta = 15$ en ese caso la distribución inicial hubiera dado $P\{\theta \leq 0'5\} = 0'99$, $P\{\theta \leq 0'3\} = 0'718$ y $P\{\theta \leq 0'8\} = 1$, con lo que la distribución final hubiera sido $Beta(\theta|\alpha_1 = 25, \beta_1 = 95)$, en ambos casos la probabilidad final $P\{\theta \leq 0'5|\sum_1^{100} x_i = 20\}$ es aproximadamente uno. Con frecuencia suele ocurrir que la información muestral domina a la información inicial, como ha ocurrido aquí, de cualquier forma la información inicial ha de ser escogida cuidadosamente de acuerdo con la información disponible. Si se resume la distribución final por su media $E\left[\theta|\sum_{i=1}^{100} X_i = 20\right]$ se obtiene $0'217$ para la primera distribución inicial y $0'208$ para la segunda, dos valores muy parecidos.

3.3. CRITERIOS DE COMPARACIÓN DE ESTIMADORES

Las propiedades estudiadas de los estimadores sirven para restringir la búsqueda de los adecuados a las distintas situaciones. En general el conjunto de los estimadores disponibles suele ser amplio, por lo que se hace necesario disponer de criterios para poder comparar y eventualmente quedarse con un estimador.

La rama de la estadística que se ocupa de manera extensiva de evaluar los distintos procedimientos estadísticos, es la teoría de la decisión, sin embargo se van a ver aquí como introducción dos criterios de comparación de estimadores: el del error cuadrático medio, adecuado para los estimadores contemplados de acuerdo con la distribución en el muestreo y el de la pérdida final esperada, adecuado para los estimadores bayesianos, aquellos en los que el parámetro es considerado como una variable aleatoria con distribución inicial.

3.3.1. Error cuadrático medio

El criterio del error cuadrático medio compara estimadores promediando alguna medida de precisión respecto a todas las muestras posibles.

Definición 3.3.1 (error cuadrático medio)
Dado el estimador $T(X_1, ..., X_n)$ de θ se llama error cuadrático medio (ECM) a la expresión función de θ

$$ECM_T(\theta) = E_\theta[(T-\theta)^2] \tag{3.3.1.1.}$$

El ECM es la media de los cuadrados de las diferencias entre los valores que se estiman para θ y el valor del parámetro. En general cualquier función positiva sería válida para ser utilizada en la Expresión (3.3.1.1), pero la cuadrática, por la facilidad de su tratamiento analítico, y porque puede ser interpretada en función de la varianza del estimador y de su sesgo, es la más usada.

■

En efecto, se comprueba fácilmente que

$$ECM_T(\theta) = Var_\theta[T] + b_\theta^2[t]$$

el ECM es la suma de la varianza del estimador y del cuadrado del sesgo.

Por lo tanto si el estimador es centrado el ECM coincide con la varianza del estimador.

Ejemplo 3.3.1.1
Para una población de $Bernoulli(\theta)$, con m.a.s. de tamaño n el ECM de $\overline{X}$ coincide con su varianza y vale

$$ECM_{\overline{X}}(\theta) = \frac{\theta(1-\theta)}{n}$$

Ejemplo 3.3.1.2
Con m.a.s. de tamaño n de una población $N(\mu, \sigma)$ se sabe que los estimadores insesgados de ambos parámetros son $\overline{X}$ para μ y S^2 para σ^2 respectivamente.

De donde

$$ECM_{\overline{X}}(\mu) = Var[\overline{X}] = \frac{\sigma^2}{n}$$
$$ECM_{S^2}(\sigma^2) = Var[S^2] = \frac{2\sigma^4}{n-1}$$

Si se emplea como estimador para σ^2 la varianza muestral $b_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ que es sesgado, con sesgo

$$b_\theta[b_2] = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{\sigma^2}{n}$$

se tiene que

$$ECM_{b_2}(\sigma^2) = \frac{2(n-1)\sigma^4}{n^2} + \frac{\sigma^4}{n^2} = \frac{2n-1}{n^2}\sigma^4$$

con lo que el estimador sesgado b_2 tiene menor ECM que el estimador insesgado S^2, cualquiera que sea el valor de σ^2.

■

El ejemplo anterior muestra, que en algunos casos, de acuerdo con este criterio, puede ser mejor utilizar estimadores con un pequeño sesgo siempre que permitan una mayor reducción de la varianza.

Ejemplo 3.3.1.3

Para el caso normal de varianza conocida, se pueden comparar los siguientes estimadores para la media θ, mediante el criterio del ECM.

$$T_1(\overrightarrow{X}) = \overline{X} \quad T_2(\overrightarrow{X}) = \sum_{i=1}^{n} c_i X_i \quad T_3(\overrightarrow{X}) = \tfrac{1}{2} \quad T_4(\overrightarrow{X}) = \tfrac{1}{2}\overline{X} + \tfrac{1}{4}$$
$$\sum_{i=1}^{n} c_i = 1$$

con lo que se obtiene

$$ECM_{T_1}(\theta) = \tfrac{\sigma^2}{n} \qquad ECM_{T_2}(\theta) = \sigma^2 \sum_{i=1}^{n} c_i^2$$
$$ECM_{T_3}(\theta) = \left(\tfrac{1}{2} - \theta\right)^2 \qquad ECM_{T_4}(\theta) = \tfrac{1}{4}\tfrac{\sigma^2}{n} + \tfrac{1}{4}(\tfrac{1}{2} - \theta)^2$$

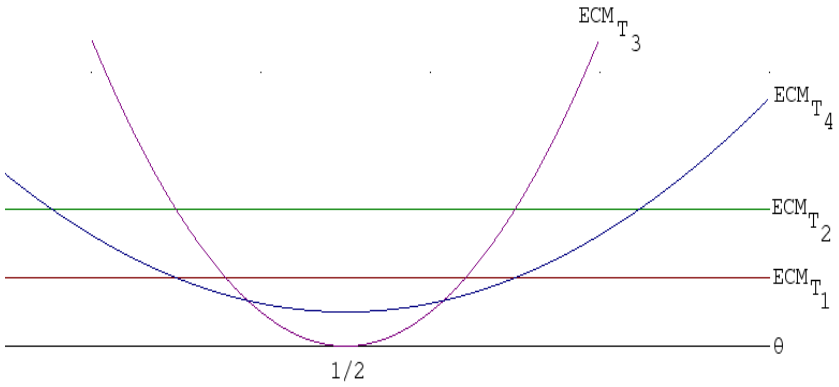


Figura 3.3.1.1. Comparación de los ECM correspondientes a varios estimadores en el Ejemplo 3.3.1.3.

El estimador T_1 lleva a excluir al estimador T_2, de acuerdo con la Figura 3.3.1.1, sobre todo a medida que el tamaño muestral aumenta.

■

En general, una vez que se introduce un criterio, éste permite excluir a los *estimadores inadmisibles*. De acuerdo con el criterio del ECM un estimador T' es *inadmisible* si existe un estimador T tal que $ECM_T(\theta) \leq ECM_{T'}(\theta)$ para todo θ perteneciente a Θ y para algún valor de θ la desigualdad es estricta. No obstante suele ocurrir, como en el Ejemplo 3.3.1.3, que no es posible quedarse con un único estimador como el mejor, por este motivo lo que se hace es utilizar el criterio en la clase más pequeña de los estimadores que tengan una propiedad que sea deseable para el problema con que se esté trabajando.

3.3.2. Pérdida final esperada

El criterio de la pérdida final esperada compara estimadores promediando las diferencias entre el valor del estimador y el valor del parámetro respecto a la distribución final sobre los distintos valores de θ. Es adecuado por tanto cuando se tenga información sobre el parámetro que permita considerar a éste como una variable aleatoria con información inicial dada por $\pi(\theta)$ es decir, esta indicado este criterio para estudiar los estimadores desde el punto de vista bayesiano. A tal fin se introduce la siguiente definición.

Definición 3.3.2.1 (pérdida final esperada)
Dado el estimador $T(X_1, ..., X_n)$, la distribución inicial $\pi(\theta)$ y definida la función de pérdida $L(t, \theta)$, donde t son los valores que toma el estimador, es

decir $T(x_1, ..., x_n) = t$. Se define la PFE_T o el riesgo a posteriori como

$$PFE_T = E[L(t, \theta)|X_1 = x_1, ..., X_n = x_n] \tag{3.3.2.1}$$

donde dicha esperanza está calculada respecto a la distribución final dada por $\pi(\theta|x_1, ..., x_n)$.

Se observa que el valor de la muestra es fijo y que la variable aleatoria es θ, a diferencia de lo que ocurre con el ECM; donde la variable aleatoria es el estadístico, véase la Expresión (3.3.1.1).

■

Cualquier función positiva es válida como función de pérdida, pero suele trabajarse con la pérdida cuadrática. De hecho cuando no se diga que función de pérdida se toma para el cálculo de la PFE se entenderá la pérdida cuadrática.

En general mediante la PFE se obtiene un único número, por lo que siempre es posible comparar los estimadores, téngase en cuenta que el punto muestral es fijo, por este motivo en gran cantidad de situaciones puede determinarse el estimador óptimo mediante este criterio, que recibe el nombre, por extensión, de estimador Bayes, o estimador bayesiano, véase la Sección 3.2.3.

Ejemplo 3.3.2.1
Para una población de $Bernoulli(\theta)$, con m.a.s. de tamaño n, pérdida cuadrática e información inicial dada por una distribución $Beta(\alpha_0, \beta_0)$ se sabe, por el Ejemplo 3.2.3.1, que la densidad final para θ viene dada por una $Beta(\alpha_1, \beta_1)$ donde los parámetros son

$$\alpha_1 = \alpha_0 + \sum_1^n x_i \qquad \text{y} \qquad \beta_1 = \beta_0 \sum_1^n x_i + 1$$

por lo tanto

$$PFE_T = E[(t - \theta)^2|\overrightarrow{x}]$$

expresión que se hace mínima como función de t cuando es $t = E[\theta|\overrightarrow{x}]$, la media de la distribución final, véase el Ejercicio 3.8.11, y para dicho valor es

$$PFE_T = \frac{\alpha_1\beta_1}{(\alpha_1 + \beta_1)^2(\alpha_1 + \beta_1 + 1)}$$

con lo que el estimador Bayes es

$$T(X_1, ..., X_n) = \frac{\alpha_0 + \sum_1^n X_i}{\alpha_0 + \beta_0 + n}$$

la esperanza de la distribución final.

Ejemplo 3.3.2.2
Para una población $N(\theta, \sigma)$ con σ conocida con m.a.s. de tamaño n y función de pérdida cuadrática, si la información inicial viene dada por una $N(\mu_0, \sigma_0)$, se sigue, de acuerdo con el Ejemplo 3.2.3.1, que la densidad final es $N(\mu_1, \sigma_1)$ donde

$$\mu_1 = \frac{\mu_0/\sigma_0^2 + n\overline{x}/\sigma^2}{1/\sigma_0^2 + n/\sigma^2} \qquad \sigma_1^2 = \frac{1}{1/\sigma_0^2 + n/\sigma^2}$$

por lo tanto el estimador bayesiano es μ_1 y la $PFE_{\mu_1} = \sigma_1^2$. Si se toma como función de pérdida $L(t, \theta) = |t - \theta|$, entonces

$$PFE_T = E[|t - \theta||\overrightarrow{x}]$$

expresión que se hace mínima para la mediana de la distribución final, véase el Ejercicio 3.8.14, por lo que el estimador bayesiano es

$$T(X_1, ..., X_n) = \mu_1$$

la mediana de la distribución final, véase la Expresión (1.3.2.1).

3.4. ESTIMADORES CENTRADOS DE MÍNIMA VARIANZA

Se va a tratar ahora un caso, históricamente el primero, en el que si se restringe la clase de estimadores a utilizar, el criterio del error cuadrático medio, permite escoger un estimador óptimo. La clase va a ser la de los estimadores centrados y por tanto el ECM coincide con la varianza del estimador.

Desgraciadamente, y con frecuencia, la clase de estimadores en que se busca el óptimo queda muy reducida.

Definición 3.4.1
Si se nota por $h(\theta)$ la función a estimar y se considera la clase de estimadores

$$\mathcal{U}_h = \{T(X_1, ..., X_n) | E_\theta T = h(\theta) \text{ con } E_\theta T^2 < +\infty \quad \forall \theta \in \Theta\}$$

Se llama estimador centrado de mínima varianza ($ECMV$) al estimador T de esta clase tal que para cualquier otro T' perteneciente a $\mathcal{U}_h$ cumpla que $V_\theta[T] \leq V_\theta[T']$ para todo θ perteneciente a Θ.

■

El siguiente teorema recoge una caracterización del $ECMV$.

Teorema 3.4.1
Dada la clase $\mathcal{U}_h$ de la Definición (3.4.1) que para $h = \theta$ se convierte en la clase

$$\mathcal{U}_h = \{f(X_1, ..., X_n) | E_\theta f = \theta \quad \text{con} \quad E_\theta f^2 < +\infty \quad \forall \theta \in \Theta\}$$

entonces, el estimador T perteneciente a la clase $\mathcal{U}_\text{h}$ es el estimador óptimo con el criterio del ECM si y solamente si $E_\theta[Tf] = 0$ para todo θ de Θ y para toda función f perteneciente a $\mathcal{U}_0$.

Demostración: Sea T óptimo. Si $f(\overrightarrow{x}) \in \mathcal{U}_0$ y es $E_\theta[f^2] = 0$, entonces f es cero casi seguro y por tanto sea o no T el estimador óptimo es $E_\theta[Tf] = 0$. Se puede entonces suponer que $E_\theta[f^2]$ es estrictamente mayor que cero. Si $E_\theta[Tf]$ es distinta de cero para algún valor θ_0, se puede construir un estimador con varianza, en el punto θ_0 estrictamente menor que T. Sea $T + \lambda f$.

$$\begin{aligned} V_{\theta_0}[T + \lambda f] &= E_{\theta_0}[(T + \lambda f - E_{\theta_0}T - \lambda E_{\theta_0} f)^2] = E_{\theta_0}[(T - E_{\theta_0}T + \lambda f)^2] = \\ &= V_{\theta_0}[T] + \lambda^2 E_{\theta_0} f^2 + 2\lambda E_{\theta_0}[(T - E_{\theta_0}T)f] \end{aligned}$$

y por ser $E_{\theta_0}[f] = 0$ se puede poner

$$V_{\theta_0}[T + \lambda f] = V_{\theta_0}[T] + \lambda^2 E_{\theta_0} f^2 + 2\lambda E_{\theta_0}[Tf].$$

Como $E_{\theta_0}[Tf]$ es distinta de cero se tiene que si $E_{\theta_0}[Tf] < 0$, entonces para cualquier $\lambda \in (0, -2\frac{E_{\theta_0}Tf}{E_{\theta_0}f^2})$ se sigue que

$$\lambda^2 E_{\theta_0} f^2 + 2\lambda E_{\theta_0}[Tf] < 0 \tag{3.4.1}$$

con lo que, al menos para ese θ_0, es

$$V_{\theta_0}[T + \lambda f] < V_{\theta_0}[T].$$

Si por el contrario es $E_{\theta_0}[Tf] > 0$, entonces para cualquier $\lambda \in (-2\frac{E_{\theta_0}Tf}{E_{\theta_0}f^2}, 0)$ se obtiene también (3.4.1) y la conclusión es la misma.

Recíprocamente, si se supone que para cualquier $f \in \mathcal{U}_0$ y cualquier θ es $E_\theta[Tf] = 0$ para algún estimador T de $\mathcal{U}_h$, se sigue que T es óptimo en $\mathcal{U}_h$. En efecto para T' cualquier otro estimador perteneciente a $\mathcal{U}_h$, $T - T'$ pertenece a la clase $\mathcal{U}_0$ ya que

$$E_\theta[T - T'] = h(\theta) - h(\theta) = 0$$

y

$$E_\theta[(T - T')^2] = E_\theta[T^2] + E_\theta[T'^2] - 2E_\theta[TT'], \tag{3.4.2.}$$

el segundo miembro de (3.4.2) está acotado, ya que T y T' pertenece $\mathcal{U}_h$ y $E_\theta[TT']$ está acotada también ya que si se denota por ρ el coeficiente de correlación lineal es

$$\rho_\theta(T,T') = \frac{E_\theta[TT'] - h^2(\theta)}{(V_\theta[T]V_\theta[T'])^{1/2}} \leq 1.$$

En virtud de la hipótesis, si se considera como f a la diferencia $T - T'$, habrá de ser

$$E_\theta[T(T-T')] = 0$$

o lo que es lo mismo

$$E_\theta[T^2] = E_\theta[TT']$$

de donde se sigue que $V_\theta[T] = Cov_\theta(T,T')$ y por lo tanto

$$\rho_\theta(T,T') = \frac{Cov_\theta(T,T')}{(V_\theta[T]V_\theta[T'])^{1/2}} = \left(\frac{V_\theta[T]}{V_\theta[T']}\right)^{1/2} \leq 1$$

luego $V_\theta[T] \leq V_\theta[T']$.

■

Como subproducto de esta demostración se tiene el siguiente resultado.

Corolario 3.4.1
El $ECMV$ es único.

Si existieran dos T y T' por la última parte de la demostración del Teorema 3.4.1 sería $E_\theta[T(T-T')] = 0$ y por lo tanto el coeficiente de correlación lineal habría de ser

$$\rho_\theta(T,T') = 1 \qquad \text{para todo } \theta \in \Theta$$

o lo que es equivalente $T' = a(\theta)T + b(\theta)$, por lo que

$$E_\theta[T'] = E_\theta[T] + b(\theta)$$

o bien

$$h(\theta) = a(\theta)h(\theta) + b(\theta).$$

De donde ha de ser $b(\theta) = 0$ y $a(\theta) = 1$, con lo que T ha de coincidir casi seguro con T' para todo θ.

Corolario 3.4.2
Si $T_1(\overrightarrow{x})$ y $T_2(\overrightarrow{x})$ son $ECMV$ para $h_1(\theta)$ y $h_2(\theta)$ respectivamente, entonces $b_1T_1(\overrightarrow{x}) + b_2T_2(\overrightarrow{x})$ es el $ECMV$ para la función de $\theta \quad b_1h_1(\theta) + b_2h_2(\theta)$.

Demostración: Basta observar que para cualquier $f \in \mathcal{U}_0$ se tiene

$$E_\theta[(b_1T_1 + b_2T_2)f] = b_1E_\theta[T_1f] + b_2E_g[T_2f] = 0.$$

■

Mediante el Teorema 3.4.1, es posible obtener el $ECMV$ como ponen de manifiesto los dos ejemplos que siguen.

Ejemplo 3.4.1
Para una población de $Bernoulli(\theta)$, puede verse que

$$U_0 = \{0(x) = 0\}.$$

Es decir, el único estimador insesgado de cero es la función cero; luego cualquier estimador insesgado de θ, por ejemplo $T(X) = X$, es el $ECMV$ para θ. Además no hay otro por ser éste único.

Ejemplo 3.4.2
Con muestras de tamaño uno de la distribución uniforme discreta y para $\theta = N$, es fácil ver que $T(X) = 2X - 1$ es un estimador insesgado de N, como la clase $\mathcal{U}_0$ solo contiene a la función cero, T es también el $ECMV$ para N. Si ahora se elimina de la familia de distribuciones discretas la distribución correspondiente al valor n_0 de N, puede verse que

$$U_0 = \{\text{cero}, \ \ f_c \ \ \text{con} \ \ c \in \mathbb{R}\}$$

donde

$$f_c(x) = \begin{cases} 0 & \text{si} \quad x \neq n_0, n_0 + 1 \ \ \text{y} \ \ x = 1, 2, 3, ... \\ c & \text{si} \quad x = n_0 \\ -c & \text{si} \quad x = n_0 + 1 \end{cases}$$

Luego $T(X) = 2X - 1$ no es el $ECMV$ para N, ya que

$$E_{n_0+1}[(2X - 1)f_c(X)] = -\frac{2c}{n_0 + 1}$$

y de acuerdo con el Teorema 3.4.1, si fuera óptimo, esta esperanza debería haber sido nula. Sin embargo si se define $W(n_0) = W(n_0 + 1) = 2n_0$ y se deja igual que antes en los restantes valores. $W(X) = 2X - 1$, se comprueba que es óptimo, para la nueva clase de distribuciones (las uniformes discretas menos la correspondiente al valor n_0). Este es un ejemplo de la existencia de estimador óptimo sin que exista estadístico completo, véase la Definición 2.3.3.

■

El siguiente enunciado permite excluir a ciertos estimadores como candidatos a óptimos.

Teorema 3.4.2
El $ECMV$ ha de ser función simétrica de las observaciones.

Demostración: en otro caso el estimador

$$\overline{T} = \frac{1}{n!} \sum_{\pi} T_{\pi}(X_1, ..., X_n),$$

donde π es una permutación de los n primero número naturales y entendiendo como $T_{\pi}(X_1, ..., X_n)$ el valor del estadístico T en la muesta obtenida al permutar lo índices de acuerdo con π, es un estimador *mejor* que T.

Es claro que $E[\overline{T}] = E[T] = h(\theta)$ y que

$$V_{\theta}[\overline{T}] = \frac{1}{(n!)^2} \left\{ \sum_{\pi} V_{\theta}[T_{\pi}] + \sum_{\pi \neq \pi'} Cov_{\theta}(T_{\pi}, T_{\pi'}) \right\}$$

y como

$$\rho_{\theta}(T_{\pi}, T_{\pi'}) = \frac{Cov_{\theta}(T_{\pi}, T_{\pi'})}{(V_{\theta}[T_{\pi}] V_{\theta}[T_{\pi'}])^{1/2}} \leq 1$$

se sigue que

$$Cov_{\theta}(T_{\pi}, T_{\pi'}) \leq V_{\theta}[T]$$

con la desigualdad estricta a no ser que T sea simétrico. Entonces se tiene que

$$\begin{aligned} V_{\theta}[\overline{T}] &< \frac{1}{(n!)^2} \{n! V_{\theta}[T] + n!(n! - 1) V_{\theta}[T]\} \\ &= \frac{1}{n!} V_{\theta}[T](1 + n! - 1) = V_{\theta}[T] \end{aligned}$$

y T no sería óptimo.
■

Un método para encontrar el $ECMV$ consiste en restringirse a los estadísticos que sean función del estadístico suficiente, como demuestra el siguiente enunciado.

Teorema 3.4.3 (Rao-Blackwell)
Para estimar $h(\theta)$ se nota por $\mathcal{U}_h$, como en la Definición 3.4.1, a la clase de los estimadores centrados de $h(\theta)$ con varianza finita, si S es un estadístico suficiente y T pertenece a la clase $\mathcal{U}_h$, entonces $E[T|S]$ es un estimador de la clase $\mathcal{U}_h$ tal que

$$V_{\theta}(E[T|S]) \leq V_{\theta}[T] \tag{3.4.3}$$

para todo θ perteneciente a Θ. Además la igualdad en (3.4.3) se verifica si y sólo si $E[T|S]$ coincide casi seguro con T respecto a todo θ perteneciente a Θ.

Demostración: se tiene que

$$E_\theta[T] = E_\theta E[T|S] = h(\theta), \tag{3.4.4}$$

donde se observa que en (3.4.4), la esperanza respecto a la distribución de T condicionada por S es independiente de θ, por ser S un estadístico suficiente.

De acuerdo con la desigualdad de Cauchy-Schwarz que dice que para dos v.a. X e Y se tiene que

$$E^2[XY] \leq E[X^2]E[Y^2]$$

(véase Rao (1973), pág. 54), se sigue que $E[T|S]$ es un estimador de la clase $\mathcal{U}_h$ pues

$$E^2\,[T|S] = E^2[T \times 1|S] \leq E[T^2|S]$$

de donde $E_\theta E^2[T|S] \leq E_\theta E[T^2|S] = E_\theta T^2$ expresión que es finita por pertenecer T a la clase $\mathcal{U}_h$. Además como $E[T|S]$ y T tienen ambos de media $h(\theta)$, se sigue, de la última desigualdad restando $h^2(\theta)$, que la Expresión (3.4.3) es cierta.

Por último si es $E[T|S] = T$ se da la igualdad en (3.4.3) y recíprocamente, si es

$$V_\theta(E[T|S]) = V_\theta(T)$$

es porque

$$E^2[T|S] = E[T^2|S].$$

Entonces si se efectúa el desarrollo de

$$\begin{aligned} E_\theta[(T - E[T|S])^2] &= E_\theta T^2 + E_\theta E^2[T|S] - 2E_\theta[TE[T|S]] \\ &= 2E_\theta T^2 - 2E_\theta[TE[T|S]] \end{aligned} \tag{3.4.5}$$

como

$$\begin{aligned} E_\theta[TE[T|S]] &= E_\theta E[TE[T|S]|S] = E_\theta E^2[T|S] \\ &= E_\theta E[T^2|S] = E_\theta T^2. \end{aligned}$$

Luego (3.4.5) es cero y por tanto T ha de coincidir casi seguro con $E[T|S]$. ■

Decir que T y $E[T|S]$ coinciden, es equivalente a decir que T es función de S.

Por otro lado la hipótesis de suficiencia del estimador S sólo se utiliza para poder considerar a $E[T|S]$ como un estimador. Realmente el teorema dice que, dado un estimador T, al tomar la esperanza condicionada por un estimador, en este caso S, la varianza de T disminuye.

Además, salvo tecnicismos matemáticos, todo se ha reducido a ver que

$$Var_\theta[T] = Var(E[T|S]) + E_\theta(VarE[T|S]) \tag{3.4.6}$$

donde el segundo sumando del segundo miembro es evidentemente no negativo y por tanto la varianza del estimador $E[T|S]$ es siempre menor o igual que la varianza del estimador T.

La relación entre la ausencia de sesgo y la propiedad de completitud viene recogida en el siguiente enunciado.

Teorema 3.4.4 (Lehmann-Scheffé)
Dada la familia de distribuciones de probabilidad de la muestra $\mathcal{P} = \{f(\overrightarrow{x}|\theta)|\theta \in \Theta\}$, si para estimar $h(\theta)$ la clase $\mathcal{U}_h$ es no vacía y existe un estadístico $S(X_1, ..., X_n)$ suficiente y completo, entonces $E[T|S]$, para cualquier T perteneciente a $\mathcal{U}_h$, es el $ECMV$ para $h(\theta)$.

Demostración: Basta observar que para cualquier estadístico T' perteneciente a $\mathcal{U}_h$, por el teorema de Rao-Blackwell

$$V_\theta[E[T'|S]] \leq V_\theta[T']$$

y, como $E[T'|S]$ es un estadístico función de S, que es completo, ha de coincidir con $E[T|S]$, con lo que el estimador $E[T|S]$ es óptimo.
■

Se observa entonces que siempre que exista un estadístico completo y suficiente, el estadístico función de él que sea centrado para la función a estimar, es el estimador óptimo en la clase de los estimadores centrados, pero es a costa de que la clase de los centrados función del suficiente y completo sólo contiene a un elemento, que obviamente es el óptimo.

Ejemplo 3.4.3
Con m.a.s. de tamaño n de una población de $Poisson(\theta)$, para encontrar el estimador óptimo para θ, en la clase de los centrados, se tiene que $S = \sum_{i=1}^{n} X_i$ es un estadístico suficiente que también es completo, pues

$$P\{X_1 = x_1, ..., X_n = x_n|\theta\} = \frac{1}{\prod_{i=1}^{n} x_i!} e^{-n\theta} e^{(\ln\theta)\sum x_i}$$

es una familia exponencial uniparamétrica, por el Teorema 2.2.1 S es suficiente y $\ln\theta$ contiene un abierto de $\mathbb{R}$, luego por el Teorema 2.3.1 es completo.

Además como

$$E_\theta\left[\sum_{i=1}^{n} X_i\right] = n\theta,$$

resulta que $\overline{X}$ es función del estadístico suficiente y completo, luego es el óptimo para su media, que es θ.

Si se quiere el $ECMV$ para $h(\theta) = e^{-\theta}$, es decir para la probabilidad de que la variable aleatoria valga cero, se debe encontrar un estimador insesgado

$T(\overrightarrow{X})$ de $h(\theta)$, para luego calcular $E[T|S]$ que será el estimador óptimo de acuerdo con el Teorema 3.4.4.

Un mecanismo simple para encontrar un estimador insesgado en estas situaciones, consiste en utilizar estimadores de la forma

$$T(X_1, ..., X_n) = \begin{cases} 1 & \text{si} \quad X_1 = 0 \\ 0 & \text{si} \quad X_1 \neq 0 \end{cases}$$

con lo que obviamente

$$E_\theta T = P_\theta\{X_1 = 0\} = e^{-\theta}$$

y todo se reduce a obtener $E\left[T \left| \sum_1^n X_i = s \right.\right] = P\left\{X_1 = 0 \left| \sum_1^n X_i = s \right.\right\}$.

Pero

$$P\left\{X_1 = 0 \left| \sum_1^n X_i = s \right.\right\} = \frac{P\{X_1 = 0\} P\left\{\sum_2^n X_i = s\right\}}{P\left\{\sum_1^n X_i = s\right\}}$$

y como $\sum_{i=1}^k X_i$ tiene distribución de $Poisson(k\theta)$, se llega a que

$$E\left[T \left| \sum_1^n X_i = s \right.\right] = \left[\frac{n-1}{n}\right]^{\sum_1^n X_i}$$

es el $ECMV$ para $h(\theta) = e^{-\theta}$.

Ejemplo 3.4.4
Para m.a.s. de tamaño n de una población exponencial de parámetro θ

$$f(x|\theta) = \theta e^{-\theta x}, \;\; x > 0, \;\; \text{con } \theta > 0$$

si se pretende encontrar el estimador óptimo entre los insesgados para θ, puede comprobarse que $S(X_1, ..., X_n) = \sum_{i=1}^n X_i$ es un estadístico suficiente y completo, y que su esperanza es $\frac{n}{\theta}$. Por lo tanto $\overline{X}$ es el $ECMV$ para su media, $h(\theta) = \frac{1}{\theta}$. Sin embargo como se necesita un estimador insesgado para θ, puede ponerse que

$$E_\theta\left[\frac{1}{\sum_{i=1}^n X_i}\right] = \int_0^{+\infty} s^{-1} \frac{\theta^n}{\Gamma(n)} e^{-\theta s} s^{n-1} ds$$

ya que $\sum_{i=1}^{n} X_i$ tiene como distribución una $Gamma(a = \theta, p = n)$. Si se utiliza la igualdad

$$\int_0^{+\infty} e^{-ax} x^{p-1} dx = a^{-p}\Gamma(p)$$

se sigue que

$$E\left[\frac{1}{\sum_{i=1}^{n} X_i}\right] = \frac{\theta^n}{\Gamma(n)}\theta^{-(n-1)}\Gamma(n-1) = \frac{\theta}{n-1},$$

de donde $\frac{n-1}{\sum_{i=1}^{n} X_i}$ es el $ECMV$ para θ.

Ejemplo 3.4.5

Para determinar con m.a.s. de tamaño n de una población de $Bernoulli(\theta)$ el $ECMV$ primero para la función $h_1(\theta) = \theta$ y después para $h_2(\theta) = \theta(1-\theta)$; se observa que

$$P\{X_1 = x_1, ..., X_n = x_n|\theta\} = (1-\theta)^n \exp\left[(\ln\frac{\theta}{1-\theta})\sum_{i=1}^{n} x_i\right]$$

con lo que es fácil verificar que $S = \sum_{i=1}^{n} X_i$ es un estadístico suficiente y completo. Luego $\overline{X}$ es un estimador centrado para $h_1(\theta)$ y por tanto el $ECMV$ para θ. Para la función $h_2(\theta) = P\{X_1 = 1, X_2 = 0|\theta\}$ todo se reduce a encontrar $E[T|S]$ con

$$T(X_1, ..., X_n) = \begin{cases} 1 & \text{si } X_1 = 1 \text{ y } X_2 = 0 \\ 0 & \text{en el resto} \end{cases}$$

como

$$E\left[T\left|\sum_{1}^{n} X_i = s\right.\right] = \frac{P\{X_1 = 1, X_2 = 0|\theta\}P\{\sum_{3}^{n} X_i = s-1|\theta\}}{P\{\sum_{1}^{n} X_i = s|\theta\}}$$

se sigue, utilizando que $\sum_{3}^{n} X_i$ se distribuye según una $Binomial(n-2, \theta)$, que

$$E\left[T\left|\sum_{i=1}^{n} X_i = s\right.\right] = \frac{s(n-s)}{n(n-1)}$$

con lo que $\frac{\sum_{i=1}^{n} X_i(n-\sum_{i=1}^{n} X_i)}{n(n-1)}$ es el $ECMV$ para $h_2(\theta) = \theta(1-\theta)$.

Ejemplo 3.4.6
En el caso de m.a.s. de tamaño n de una población normal, la función de densidad de la muestra es

$$f(x_1,...,x_n|\mu,\sigma^2)=\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum_1^n(x_i-\mu)^2} \tag{3.4.7}$$

por lo tanto, si la media μ es conocida, el estadístico $\sum_1^n(X_i-\mu)^2$ es suficiente y completo para σ^2; luego $\frac{1}{n}\sum_1^n(X_i-\mu)^2$ que es centrado y función del suficiente y completo, es el estimador óptimo.

Si el parámetro de interés continua siendo σ^2, pero la media μ es desconocida, se puede poner (3.4.7) en la forma

$$f(x_1,...,x_n|\mu,\sigma^2)=\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{n}{2\sigma^2}} e^{\frac{\mu}{\sigma^2}\sum x_i-\frac{1}{2\sigma^2}\sum x_i^2} \tag{3.4.8}$$

Se trata de la familia exponencial biparamétrica y se tiene que el estadístico $\overrightarrow{T}(X_1,...,X_n)=\left(\sum_1^n X_i,\sum_1^n X_i^2\right)$ es suficiente y completo para $\overrightarrow{\theta}=(\mu,\sigma^2)$, verifíquese. Luego por el Teorema 3.4.4, la cuasivarianza muestral, que puede escribirse como

$$S^2=\frac{1}{n-1}\left[\sum_1^n X_i^2-\left(\sum_1^n X_i\right)^2/n\right],$$

es el estimador óptimo para σ^2.

Si el parámetro de interés es ahora μ, con σ^2 conocida puede ponerse (3.4.7) en la forma

$$f(x_1,...,x_n|\mu,\sigma^2)=\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{n}{2\sigma^2}-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2} e^{\frac{\mu}{\sigma^2}\sum_{i=1}^n x_i}, \tag{3.4.9}$$

que es una familia exponencial uniparamétrica. Se sigue que $T(X_1,...,X_n)=\sum_1^n X_i$ es un estadístico suficiente y completo para μ, luego $\overline{X}$ que, es función de T y es insesgado, es el estimador óptimo para μ. Por último, cuando ambos parámetros son desconocidos, la Expresión (3.4.8) permite afirmar que $(\overline{X},S^2)$, que es función del estadístico suficiente y completo, y que es insesgado para (μ,σ^2) es el estimador óptimo.

3.5. COTA PARA LA VARIANZA DE UN ESTIMADOR

Se va a estudiar en esta sección un procedimiento indirecto para encontrar el estimador óptimo, mediante el criterio del ECM en la clase de los

estimadores insesgados. La idea es lograr una cota alcanzable para la varianza de cualquier estimador insesgado, de manera que cuanto más próxima sea la varianza de un estimador a la cota, mejor será el estimador con que se está trabajando.

Con el fin de obtener esta cota, es necesario introducir ciertas condiciones de regularidad, sobre la familia de distribuciones de probabilidad; la primera, básicamente, dice que el recorrido de la población no puede depender del parámetro θ a estimar y la segunda, relativa a la clase de estimadores a utilizar, exige que la derivación e integración conmuten. Más concretamente, han de satisfacerse las dos condiciones siguientes:

R1) El conjunto

$$S = \{(x_1, ..., x_n) \in \mathcal{X} | f(x_1, ..., x_n|\theta) > 0\}$$

sea independiente de θ y que para todo θ exista $\frac{\partial}{\partial\theta} f(x_1, ..., x_n|\theta)$. Al conjunto S se le denomina soporte de la distribución f.

R2) Para la clase $\mathcal{U}_h$, de los estimadores insesgados de la función de interés $h(\theta)$ que tienen varianza finita, ha de cumplirse que

$$\frac{d}{d\theta} \int_{\mathcal{X}} T(x_1, ..., x_n) f(x_1, ..., x_n|\theta) dx_1 ... dx_n$$
$$= \int_{\mathcal{X}} T(x_1, ..., x_n) \frac{\partial}{\partial\theta} f(x_1, ..., x_n|\theta) dx_1 ... dx_n$$

y que esta expresión sea válida también cuando $T(x_1, ..., x_n) = 1$, aunque obviamente la unidad no es un estimador centrado de la función de interés.

Se llama información de Fisher, que sobre θ suministran m.a.s. de tamaño n a la expresión

$$I_n(\theta) = E_\theta \left[\left(\frac{\partial}{\partial\theta} \ln f(X_1, ..., X_n|\theta \right)^2 \right] \tag{3.5.1}$$

que alternativamente puede ponerse como

$$I_n(\theta) = \int_{\mathcal{X}} \left(\frac{\partial}{\partial\theta} f(x_1, ..., x_n|\theta) \right)^2 \frac{1}{f(x_1, ..., x_n|\theta)} dx_1 ... dx_n.$$

Se ha empleado notación de función de densidad, los mismos requisitos han de cumplirse para funciones de masa.

En la práctica es interesante, sin comprobarlo directamente, saber que las condiciones R1) y R2) son satisfechas. Así si la función de masa o de densidad de la muestra es exponencial uniparamétrica, es decir

$$f(x_1, ..., x_n|\theta) = c(\theta) h(x_1, ..., x_n) e^{\pi(\theta) t(x_1, ..., x_n)}$$

siendo $\pi(\theta)$ una función derivable en Θ tal que $\pi'(\theta) \neq 0$ para todo $\theta \in \Theta$, entonces se cumplen las dos condiciones de regularidad, (puede verse en Bickel y Doksum (1977) pág. 127).

Teorema 3.5.1 (cota de Fréchet-Cramér-Rao)
Si se verifican las condiciones de regularidad R1) y R2) y la información de Fisher es $I_n(\theta) < +\infty$, se tiene que la función $h(\theta)$ a estimar es derivable y que para cualquier estimador de la clase $\mathcal{U}_h = \{T(\overrightarrow{X}) | E_\theta T(\overrightarrow{X}) = h(\theta)$ con $E_\theta T^2 < +\infty\}$ se cumple

$$[h'(\theta)]^2 \leq V_\theta[T] I_n(\theta) \qquad \forall \theta \in \Theta.$$

Además, la igualdad se da, si y solamente si, existe una función $K(\theta)$ tal que

$$\frac{\partial}{\partial \theta} \ln f(\overrightarrow{x}|\theta) = K(\theta)[T(\overrightarrow{x}) - h(\theta)] \qquad \forall \theta \in \Theta.$$

Demostración: Para ver que $h(\theta)$ es derivable como es

$$E_\theta[T(\overrightarrow{X})] = h(\theta)$$

por R2) se tiene que

$$h'(\theta) = \int_{\mathcal{X}} T(\overrightarrow{x}) \frac{\partial}{\partial \theta} f(\overrightarrow{x}|\theta) d\overrightarrow{x} \tag{3.5.2}$$

bastará acotar el valor

$$E_\theta \left[|T(\overrightarrow{X})| \left| \frac{\partial}{\partial \theta} \ln f(\overrightarrow{X}|\theta) \right| \right] \leq \left(E_\theta[T^2(\overrightarrow{X})] \right)^{1/2} \left(E_\theta[(\frac{\partial}{\partial \theta} \ln f(\overrightarrow{X}|\theta))^2] \right)^{1/2}$$

y en virtud de las hipótesis, estos dos productos están acotados.

Por otro lado, a partir de la igualdad

$$\int_{\mathcal{X}} f(x_1, ..., x_n|\theta) dx_1 ... dx_n = 1$$

se sigue, derivando en los dos miembros y usando la condición R2), que

$$E_\theta \left[\frac{\partial}{\partial \theta} \ln f(\overrightarrow{X}|\theta) \right] = 0. \tag{3.5.3}$$

Obsérvese que también es necesaria la condición R1) para la obtención de esta expresión. Si se calcula la covarianza entre $T(\overrightarrow{x})$ y $\frac{\partial}{\partial \theta} \ln f(\overrightarrow{x}|\theta)$, se tiene que

$$Cov_\theta \left(T(\overrightarrow{X}), \frac{\partial}{\partial \theta} \ln f(\overrightarrow{X}|\theta) \right) = E_\theta \left[T(\overrightarrow{X}) \frac{\partial}{\partial \theta} \ln f(\overrightarrow{x}|\theta) \right] \tag{3.5.4}$$

ya que por la Expresión (3.5.3) la media de la segunda variable es nula.

Por la Expresión (3.5.9) se sigue que (3.5.4) puede ponerse como

$$Cov_\theta\left(T(\overrightarrow{X}), \frac{\partial}{\partial\theta}\ln f(\overrightarrow{X}|\theta)\right) = h'(\theta)$$

luego utilizando que el cuadrado del coeficiente de correlación es siempre menor o igual que 1, ha de ser por (3.5.3)

$$[h'(y)]^2 \leq V_\theta[T]\ .\ E_\theta[(\frac{\partial}{\partial\theta}\ln f(\overrightarrow{X}|\theta))^2]$$

Con lo que se tiene la primera parte del teorema.

Además, por las propiedades del coeficiente de correlación, la desigualdad se convierte en igualdad si y solamente si, para cada θ las variables $\frac{\partial}{\partial\theta}\ln f\,(\overrightarrow{x}|\theta)$ y $T(\overrightarrow{x})$ están relacionadas linealmente. Es decir

$$\frac{\partial}{\partial\theta}\ln f(\overrightarrow{x}|\theta) = A(\theta) + B(\theta)T(\overrightarrow{x})$$

y ya que el primer miembro tiene de media 0, véase (3.5.3), se puede tomar $K(\theta) = B(\theta)$ con lo que se obtiene la expresión deseada.

■

Se observa que si la información de Fisher $I_n(\theta)$ es cero también lo es $h'(\theta)$ y en caso contrario, dividiendo en el enunciado de la Cota de FCR por $I_n(\theta)$ se sigue que

$$V_\theta[T] \geq \frac{[h'(\theta)]^2}{I_n(\theta)}.$$

Es claro, que cuando $I_n(\theta)$ sea grande la expresión del segundo miembro de esta acotación será pequeña y por tanto se puede interpretar, que en ese caso, los estimadores de $\mathcal{U}_h$ tendrán menor varianza, al menos como cota inferior, por lo que se puede decir que el experimento da mayor información sobre la función del parámetro $h(\theta)$. Esto constituye una primera justificación de por qué se llama información a la expresión $I_n(\theta)$.

Corolario 3.5.1

Bajo las hipótesis del teorema anterior, si $h(\theta) = \theta$, es decir se toman los estimadores insesgados para θ, se tiene que

$$V_\theta[T] \geq \frac{1}{I_n(\theta)}.$$

El cálculo de la información de Fisher, $I_n(\theta)$ puede hacerse mediante

$$I_n(\theta) = nI_1(\theta) \tag{3.5.3}$$

ya que

$$\frac{\partial}{\partial\theta}\ln f(x_1, ..., x_n|\theta) = \sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln f(x_i|\theta).$$

Si se eleva esta última expresión al cuadrado y se emplea (3.5.3) para cada i, se sigue que al ser la m.a.s., la esperanza de los productos cruzados es nula y por lo tanto

$$E_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f(X_1,...,X_n|\theta)\right)^2\right]=\sum_{i=1}^{n}E_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f(X_i|\theta)\right)^2\right].$$

Esta expresión, completa el significado de $I_n(\theta)$, ya que parece lógico pensar que la información suministada por m.a.s. de tamaño n sea n veces la información que suministra una muestra de tamaño uno.

Ejemplo 3.5.1
Con m.a.s. de tamaño n de una población de $Bernoulli(\theta)$, para el cálculo de la Cota de FCR para los estimadores de θ, se tiene que con $y=\sum_{i=1}^{n}x_i$ dichos estimadores han de verificar

$$\sum_{y=0}^{n}T(y)\theta^y(1-\theta)^{n-y}=\theta.$$

Se sigue, que por ser ambos miembros polinomios en θ

$$\frac{d}{d\theta}\sum_{y=0}^{n}T(y)\theta^y(1-\theta)^{n-y}=\sum_{y=0}^{n}T(y)\frac{\partial}{\partial\theta}\left[\theta^y(1-\theta)^{n-y}\right],$$

luego se satisfacen las dos condiciones de regularidad. Alternativamente se trata de la familia exponencial uniparamétrica con $\pi'(\theta)$ no nula.

$$I_1(\theta)=E_\theta\left[\left(\frac{\partial}{\partial\theta}\ln P\{X|\theta\}\right)^2\right]$$

se convierte en

$$E_\theta\left[\left(\frac{X-\theta}{\theta(1-\theta)}\right)^2\right]=\frac{V_\theta[X]}{\theta^2(1-\theta)^2}$$

con lo que es $I_1(\theta)=\frac{1}{\theta(1-\theta)}$ y por lo tanto $I_n(\theta)=\frac{n}{\theta(1-\theta)}$.

Entonces

$$\text{Cota de } FCR=\frac{\theta(1-\theta)}{n},$$

de donde se deduce que el $ECMV$ para θ es $\overline{X}_n$, que es el estimador que tiene dicha varianza.

Si se puede derivar otra vez la Expresión (3.5.2) (más adelante se verán condiciones bajo las que esto es posible) se tiene que

$$I_n(\theta)=-E_\theta\left[\frac{\partial^2}{\partial\theta^2}\ln f(\overrightarrow{X}|\theta)\right] \tag{3.5.6}$$

expresión que con frecuencia es más fácil de calcular.

En efecto, para obtener (3.5.6) basta utilizar que

$$\frac{\partial}{\partial\theta}\ln f(\overrightarrow{x}|\theta)=\frac{1}{f(\overrightarrow{x}|\theta)}\frac{\partial}{\partial\theta}f(\overrightarrow{x}|\theta) \tag{3.5.7}$$

si se deriva (3.5.7) respecto a θ, y se multiplica por $f(\overrightarrow{x}|\theta)$, queda

$$\left[\frac{\partial^2}{\partial\theta^2}\ln f(\overrightarrow{x}|\theta)\right]f(\overrightarrow{x}|\theta)=\frac{\partial^2}{\partial\theta^2}f(\overrightarrow{x}|\theta)-\left(\frac{\partial}{\partial\theta}f(\overrightarrow{x}|\theta)\right)^2(1/f(\overrightarrow{x}|\theta))$$

e integrando en $\overrightarrow{x}$ y utilizando 3.5.7

$$E_\theta\left[\frac{\partial^2}{\partial\theta^2}\ln f(\overrightarrow{X}|\theta)\right]=E_\theta\left[\frac{\partial^2}{\partial\theta^2}f(\overrightarrow{X}|\theta)\right]-E_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f(\overrightarrow{X}|\theta)\right)^2\right]$$

pero por la hipótesis de regularidad, el primer sumando del segundo miembro es nulo.

Ejemplo 3.5.2

Para una población de $Poisson(\theta)$, con m.a.s. de tamaño n, si se quiere estimar θ, se tiene

$$\ln p(x|\theta)=-\theta+\mathrm{n}\ln\theta-\ln x!$$

con lo que

$$\frac{\partial}{\partial\theta}\ln p(x|\theta)=-1+\frac{x}{\theta}$$

y

$$\frac{\partial^2}{\partial\theta^2}\ln p(x|\theta)=-\frac{1}{\theta^2}x$$

de donde se deduce que

$$E_\theta\left[\frac{\partial^2}{\partial\theta^2}\ln p(X|\theta)\right]=-\frac{1}{\theta^2}E_\theta[X]=-\frac{1}{\theta}$$

y por tanto es $I_1(\theta)=\frac{1}{\theta}$. Véase que existe regularidad suficiente.

Se ha obtenido entonces que $I_n(\theta)=\frac{n}{\theta}$ y la Cota de $FCR=\frac{\theta}{n}$. Luego el estimador $\overline{X}$ cuya varianza coincide con la Cota es óptimo. Véase el Ejemplo 3.4.3.

Corolario 3.5.2

Bajo las hipótesis de regularidad del Teorema 3.5.1, suponiendo válida la Expresión (3.5.6) y si existe algún estimador $T(\overrightarrow{X})$ de la clase $\mathcal{U}_h$ cuya varianza alcance la Cota de FCR, ha de poderse factorizar

$$\frac{\partial}{\partial\theta}\ln f(\overrightarrow{x}|\theta)=\frac{nI_1(\theta)}{h'(\theta)}[T(\overrightarrow{x})-h(\theta)]. \tag{3.5.8}$$

Demostración: por el Teorema 3.5.1 la Cota es alcanzable si

$$\frac{\partial}{\partial\theta}\ln f(\overrightarrow{x}|\theta) = K(\theta)[T(\overrightarrow{x}) - h(\theta)],$$

derivando respecto de θ se tiene

$$\frac{\partial^2}{\partial\theta^2}\ln f(\overrightarrow{x}|\theta) = K'(\theta)[T(\overrightarrow{x}) - h(\theta)] - K(\theta)h'(\theta)$$

y si se toman esperanzas

$$E_\theta\left[\frac{\partial^2}{\partial\theta^2}\ln f(\overrightarrow{X}|\theta)\right] = -K(\theta)h'(\theta),$$

o lo que es lo mismo

$$K(\theta) = \frac{I_n(\theta)}{h'(\theta)} = \frac{nI_1(\theta)}{h'(\theta)}.$$

■

La Expresión (3.5.8) permite construir directamente el estimador T, la función $h(\theta)$ y el valor de la Cota a menos de constantes aditivas y multiplicativas.

En el Ejemplo 3.5.1 es

$$\frac{\partial}{\partial\theta}\ln P\{\overrightarrow{X} = \overrightarrow{x}|\theta\} = \sum_{i=1}^{n} x_i\frac{1}{\theta} - \left(n - \sum_{i=1}^{n} x_i\right)\frac{1}{1-\theta}$$

sacando factor común $\sum_{i=1}^{n} X_i$, se puede poner

$$\frac{\partial}{\partial\theta}\ln P\{\overrightarrow{X} = \overrightarrow{x}|\theta\} = \frac{n}{\theta(1-\theta)}[\overline{X} - \theta]$$

luego $\overline{X}$ es el $ECMV$ para $h(\theta) = \theta$ y la Cota de FCR vale $\frac{\theta(1-\theta)}{n}$. Se observa que se pueden multiplicar y restar $\overline{X}$ y θ por sendas constantes con lo que se mantendría la factorización dada en (3.5.7).

Todavía se puede concretar más respecto a poder alcanzar la Cota de FCR, pero ésto está relacionado con que la familia de distribuciones de probabilidad sea la exponencial uniparamétrica con mínimas suposiciones adicionales, lo que resta importancia a este método para construir $ECMV$.

Teorema 3.5.2
Si la familia de distribuciones de probabilidad verifica las condiciones de regularidad R1 y R2 y la Cota de FCR es alcanzable, entonces la familia es exponencial uniparamétrica, es decir se puede escribir como

$$f(\overrightarrow{x}|\theta) = c(\theta)h(\overrightarrow{x})e^{\pi(\theta)T(\overrightarrow{x})} \tag{3.5.9}$$

y además es T el estadístico que alcanza la Cota.

Recíprocamente, si la muestra se distribuye mediante (3.5.9) con $\pi'(\theta)$ no nula, entonces T alcanza la Cota de FCR.

Demostración: si se alcanza la Cota, por el Teorema 3.5.1, ha de ser

$$\frac{\partial}{\partial\theta}\ln f(\overrightarrow{x}|\theta) = K(\theta)[T(\overrightarrow{x}) - h(\theta)]$$

integrando respecto de θ se obtiene

$$\ln f(\overrightarrow{x}|\theta) = A(\theta)T(\overrightarrow{x}) + B(\theta)R(\overrightarrow{x})$$

que equivale a

$$f(\overrightarrow{x}|\theta) = e^{B(\theta)}e^{R(\overrightarrow{x})}e^{A(\theta)T(\overrightarrow{x})}.$$

Recíprocamente, si $f(\overrightarrow{x}|\theta)$ es de la forma (3.5.9) se sigue que

$$\frac{\partial}{\partial\theta}\ln f(\overrightarrow{x}|\theta) = \frac{c'(\theta)}{c(\theta)} + \pi'(\theta)T(\overrightarrow{x})$$

que es de la forma

$$K(\theta)[T(\overrightarrow{x}) - h(\theta)]$$

donde $K(\theta) = \pi'(\theta)$ y $h(\theta) = -\frac{c'(\theta)}{c(\theta)\pi'(\theta)}$.
■

Para el Ejemplo 3.5.1 con m.a.s. de tamaño n de una población de Bernoulli es

$$P\{X_1 = x_1, ..., X_n = x_n|\theta\} = \theta^{\sum_1^n x_i}(1-\theta)^{n-\sum_1^n x_i}$$

expresión que puede ponerse alternativamente como

$$(1-\theta)^n e^{\left[\ln\left(\frac{\theta}{1-\theta}\right)\right]\sum x_i}$$

luego $\pi(\theta) = \ln\frac{\theta}{1-\theta}$, $\pi'(\theta) = \frac{1}{\theta(1-\theta)}$ que no es nula y $h(\theta) = -n\theta$. Se trata pues de la familia exponencial uniparamétrica, $\pi(\theta)$ tiene derivada no nula, luego es $\sum X_i$ el $ECMV$ que alcanza la Cota para $h(\theta) = n\theta$.

Condiciones más generales que las dadas por R1 y R2 y que aseguren que se satisfacen las dos hipótesis de regularidad, pueden verse en Lehman (1959) pág. 52-53. En esencia si $\phi(\overrightarrow{x})$ es una función acotada, la función de θ dada mediante la integral

$$\int_{\mathcal{X}} \phi(\overrightarrow{x})e^{\theta T(\overrightarrow{x})}h(\overrightarrow{x})d\overrightarrow{x}$$

es una función analítica. El teorema recogido en Lehmann de hecho es más general, pero aquí sólo se necesita para la familia exponencial uniparamétrica, que es para la que se puede alcanzar la Cota de FCR.

Ejemplo 3.5.3
Para m.a.s. de tamaño n de una distribución de $Poisson(\theta)$, por pertenecer la distribución de Poisson a la familia exponencial y darse las condiciones de regularidad, pues $\pi'(\theta) = \frac{1}{\theta}$ con

$$P\{X_1 = x_1, ..., X_n = x_n|\theta\} = e^{-n\theta}\frac{1}{\prod_1^n x_i!}e^{(\ln\theta)\sum_{i=1}^n x_i}$$

Al tomar logaritmos y derivar se obtiene

$$\frac{\partial}{\partial\theta}\ln P\{\overrightarrow{X} = \overrightarrow{x}|\theta\} = \frac{n}{\theta}[\overline{X} - \theta]$$

luego, a menos de constantes, $\overline{X}$ es el $ECMV$ para θ, y

$$\frac{nI_1(\theta)}{h'(\theta)} = nI_1(\theta)$$

vale $\frac{n}{\theta}$; luego la Cota de FCR para los estimadores insesgados de θ es $I_n^{-1}(\theta) = \frac{\theta}{n}$.

■

Aún existiendo $ECMV$ puede ocurrir que éste no alcance la Cota.

Ejemplo 3.5.4
Para m.a.s. de tamaño n de una distribución exponencial de parámetro θ, $f(x|\theta) = \theta e^{-\theta x}$ con $x > 0 \quad \theta > 0$, se puede ver que se satisfacen las condiciones de regularidad y que $I_n(\theta) = \frac{n}{\theta^2}$. Sin embargo es $T = \frac{n-1}{\sum X_i}$ el $ECMV$ para θ con varianza $\frac{\theta^2}{n-2}$. Lo que ocurre es que de acuerdo con el Corolario 3.4.4, la Cota de FCR es alcanzable para la función $\frac{1}{\theta}$ por el estimador $\overline{X}$ y vale $\frac{1}{n\theta^2}$.

■

Pese a estas limitaciones se introduce el siguiente concepto.

Definición 3.5.1 (estimador Eficiente)
T lo es para una familia $\mathcal{P} = \{f(\overrightarrow{x}|\theta)|\theta \in \Theta\}$ y para la función $h(\theta)$, si su varianza alcanza la Cota de FCR y el estimador tiene de media $h(\theta)$.

■

Realmente sólo tendría sentido su cálculo en la familia exponencial uniparamétrica y para la función $h(\theta)$ conveniente.

A pesar de todo, y principalmente por motivos históricos, se sigue usando el concepto.

Definición 3.5.2 (eficiencia de un estimador)
Se llama eficiencia de un estimador insesgado para la función $h(\theta)$, a la expresión

$$ef_\theta(T) = \frac{[h'(\theta)]^2}{I_n(\theta)V_\theta[T]}$$

Es decir, es el cociente entre la varianza mínima y la varianza del estimador.

■

Claramente la eficiencia es siempre menor o igual que 1 y un estimador eficiente, alcanza la Cota de FCR y tiene de eficiencia 1.

Ejemplo 3.5.5
Para la distribución de $Poisson(\theta)$, se tiene que $\overline{X}$ para estimar θ tiene $E_\theta[\overline{X}] = \theta$, $V_\theta[\overline{X}] = \frac{\theta}{n}$, y que $I_n(\theta) < \frac{n}{\theta}$, véase el Ejemplo 4.5.3, luego

$$ef_\theta(\overline{X}) = \frac{1}{\frac{n}{\theta}\frac{\theta}{n}} = 1$$

y el estimador $\overline{X}$ es eficiente.

3.6. MÉTODOS DE CONSTRUCCIÓN DE ESTIMADORES

Anteriormente se han estudiado las propiedades que deseablemente ha de tener un estimador, se van a estudiar ahora métodos que permitan construir estimadores en situaciones bastante generales y que en contextos amplios den lugar a estimadores con buenas propiedades. Se supondrá el esquema habitual, una familia de distribuciones de probabilidad sobre una población, de manera que se desconozca un parámetro $\overrightarrow{\theta} \in \Theta \subset \mathbb{R}^r$. Se trata de obtener estimadores $\overrightarrow{T}(X_1, ..., X_n)$ para $\overrightarrow{\theta}$ o para alguna función vectorial $\overrightarrow{h}(\overrightarrow{\theta})$.

3.6.1. Método de los momentos

Este método es debido a Karl Pearson e históricamente fué el primero que se utilizó de forma sistemática para ajustar parámetros de distribuciones a colecciones de datos. Se recuerda que $\alpha_k = E[X^k]$ es el momento poblacional de orden k respecto al origen y que $a_k = \frac{1}{n}\sum_1^n X_i^k$ es el momento muestral de orden k respecto al origen. Para poner de manifiesto la dependencia del momento α_k respecto del parámetro, se escribirá $\alpha_k = \alpha_k(\theta_1, ..., \theta_r)$. El método de los momentos para construir estimadores consiste en resolver en $\theta_1, ..., \theta_r$ el sistema

$$\left.\begin{array}{c} \alpha_k(\theta_1, ..., \theta_r) = a_k \\ k = 1, ..., r \end{array}\right\} \tag{3.6.1.1}$$

cuyas soluciones se denotan por $\widehat{\theta}_1(X_1,...,X_n),...,\widehat{\theta}_r(X_1,...,X_n)$.

Ejemplo 3.6.1.1
Para m.a.s. de tamaño n de una población de *Bernoulli*(θ) en este caso $r=1$, $\alpha_1(\theta)=\theta$ y $a_1=\overline{X}$, por lo tanto $\widehat{\theta}=\overline{X}$ es el estimador por el método de los momentos.

Ejemplo 3.6.1.2
Para una población con distribución de *Poisson*(θ) se necesita una sóla ecuación, $\alpha_1(\theta)=\theta$ y $a_1=\overline{X}$ luego el estimador es $\widehat{\theta}=\overline{X}$.

Ejemplo 3.6.1.3
Para una población X con distribución $N(\mu,\sigma)$ con ambos parámetros desconocidos $\overrightarrow{\theta}=(\mu,\sigma)$ es $r=2$ y se debe resolver el sistema

$$\left.\begin{array}{c}\mu=\overline{X}\\ \sigma^2+\mu^2=\frac{1}{n}\sum\limits_{i=1}^{n}X_i^2\end{array}\right\}$$

lo que lleva a $\widehat{\mu}=\overline{X}$ y $\widehat{\sigma}=\sqrt{b_2}$ la desviación típica muestral $\sqrt{\frac{1}{n}\sum(X_i-\overline{X})^2}$.

■

Algunas veces, por comodidad, se igualan r momentos centrales en lugar de los momentos respecto al origen, por lo que el procedimiento de construcción dado por las ecuaciones (3.6.1.1) no ha de ser tomado de forma estricta.

Si se está interesado en una parametrización diferente, es decir en lugar de en $\overrightarrow{\theta}=(\theta_1,...,\theta_r)$ en $\overrightarrow{\varphi}(\overrightarrow{\theta})=(\varphi_1(\theta_1,...,\theta_r),...,\varphi_r(\theta_1,...,\theta_r))$ se pueden estimar previamente los r parámetros mediante $\widehat{\theta}_1,...,\widehat{\theta}_r$ y tomar como estimadores de $\vec{\varphi}$ $(\varphi_1(\widehat{\theta}_1,...,\widehat{\theta}_r),...,\varphi_r(\widehat{\theta}_1,...,\widehat{\theta}_r))$, o bien se pueden expresar los momentos poblacionales α_k en función de $(\varphi_1,...,\varphi_r)$ y resolver las ecuaciones en $\varphi_1,...,\varphi_r$. Es decir (3.6.1.1) tomaría la forma

$$\left.\begin{array}{c}\alpha_k(\varphi_1,...,\varphi_r)=a_k\\ k=1,...,r\end{array}\right\} \qquad (3.6.1.2)$$

con lo que se obtendrían los estimadores

$$(\widehat{\varphi}_1(X_1,...,X_n),...,\widehat{\varphi}_r(X_1,...,X_n))\,.$$

Así en el Ejemplo 3.6.1.3 se podría haber puesto $\overrightarrow{\varphi}=(\varphi_1,\varphi_2)$ con

$$\begin{aligned}\varphi_1(\mu,\sigma)&=\mu\\ \varphi_2(\mu,\sigma)&=\sigma^2\end{aligned}$$

con lo que puede verse que se obtiene $\widehat{\varphi}_1(X_1, ..., X_n) = \overline{X}$ y $\widehat{\varphi}_2(X_1, ..., X_n) = b_2$. Sin embargo esta especie de invariancia no ocurre en general.

Frecuentemente los estimadores, obtenidos por el método de los momentos son consistentes, véase la Sección 3.2.2. Esto se debe a que los momentos muestrales convergen casi seguro a sus respectivos momentos poblacionales; entonces, si la transformación que pasa de $(\theta_1, ..., \theta_r)$ a $(\alpha_1, ..., \alpha_r)$ es biyectiva con funciones inversas continuas $\theta_i = \varphi_i(\alpha_1, ..., \alpha_r)$, se sigue que $\widehat{\theta}_i = \varphi_i(a_1, ..., a_r)$ converge casi seguro a $\varphi_i(\alpha_1, ..., \alpha_r) = \theta_i$.

No pueden establecerse propiedades generales para los estimadores obtenidos por el método de los momentos, no obstante en algunos casos pueden tener eficiencia pequeña para tamaño muestral fijo e incluso de manera asintótica, puede verse en Fisz(1963) pág. 484.

3.6.2. Método de la máxima verosimilitud

Este método fué propuesto y desarrollado por Fisher (1921, 1925); las referencias pueden verse en el ya citado libro de Fisz. Es el método más importante por estar basado en un principio de reducción de datos bastante natural, véase la Sección 2.5.1, y es el que goza de mejores propiedades, sobre todo para tamaños muestrales grandes.

El concepto se introduce mediante un ejemplo. Una urna contiene bolas blancas y negras, hasta un total de 4 pero se desconoce θ la proporción de blancas, con lo que el espacio paramétrico es $\Theta = \{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$. Se supone que se observan dos bolas con reemplazamiento con lo que al ser el estadístico suficiente $T = X_1 + X_2$, el número de bolas blancas, se puede construir la Tabla 6.3.2.1

	$T=0$	$T=1$	$T=2$
$\theta = \frac{1}{4}$	$\left(\frac{3}{4}\right)^2 = 0'5625$	$2\frac{1}{4}\frac{3}{4} = 0'375$	$\left(\frac{1}{4}\right)^2 = 0'0625$
$\theta = \frac{1}{2}$	$\left(\frac{1}{2}\right)^2 = 0'25$	$2\frac{1}{2}\frac{1}{2} = 0'5$	$\left(\frac{1}{2}\right)^2 = 0'25$
$\theta = \frac{3}{4}$	$\left(\frac{1}{4}\right)^2 = 0'0625$	$2\frac{1}{4}\frac{3}{4} = 0'375$	$\left(\frac{3}{4}\right)^2 = 0'5625$

Tabla 6.3.2.1. Obtención del EMV.

El estimador de máxima verosimilitud, en lo sucesivo EMV, será aquel que dado $T(X_1, ..., X_n)$ asigne a θ el valor, de los tres posibles, que dé mayor probabilidad al suceso observado, es decir

$$\widehat{\theta}(0) = \frac{1}{4} \qquad \widehat{\theta}(1) = \frac{1}{2} \qquad \widehat{\theta}(2) = \frac{3}{4}.$$

Para el caso en que se tenga una variable aleatoria con función de densidad $f(\overrightarrow{x}|\theta)$ se escogerá el que haga máximo $f(\overrightarrow{x}|\cdot)$ como función de θ.

Definición 3.6.2.1 (estimador de máxima verosimilitud)
Dada una realización $(x_1, ..., x_n)$ de una m.a.s. de $\mathcal{P} = \{f(\overrightarrow{x}|\theta)|\theta \in \Theta\}$, se representa por $L(\theta|x_1, ..., x_n)$ a la función de densidad o de masa de la

muestra, como función de θ, a la que se le da el nombre de función de verosimilitud. Un estimador $\widehat{\theta} : \mathcal{X} \longrightarrow \Theta$ se denomina de máxima verosimilitud si $L(\widehat{\theta}|x_1, ..., x_n) = \underset{\theta \in \Theta}{\text{máx}} L(\theta|x_1, ..., x_n)$.

■

El EMV dada la muestra $(x_1, ..., x_n)$ es el valor de θ que hace más probable el punto muestral que se ha observado.

Lo primero que se señala es que puede ocurrir que el máximo no sea alcanzable, se convierta en un supremo y entonces no existiría EMV.

Como el máximo de una función y su logaritmo neperiano coinciden, con frecuencia se tabaja con el logaritmo neperiano de la función de verosimilitud para facilitar los cálculos.

Ejemplo 3.6.2.1
Para una población de $Bernoulli(\theta)$ con m.a.s. de tamaño n la función de verosimilitud es

$$L(\theta|x_1, ..., x_n) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i} \qquad \theta \in [0, 1]$$

y su logaritmo neperiano es

$$\ln L(\theta|x_1, ..., x_n) = \sum x_i \ln \theta + \left(n - \sum x_i\right) \ln(1-\theta)$$

con lo que derivando respecto a θ se obtiene

$$\frac{\partial}{\partial \theta} \ln L(\theta|x_1, ..., x_n) = \frac{1}{\theta} \sum x_i - \frac{n - \sum x_i}{1-\theta}$$

e igualando a cero se llega a que $\widehat{\theta} = \overline{X}$ que obviamente está contenido en el intervalo $[0, 1]$. La derivada segunda respecto a θ, particularizada en $\overline{X}$ lleva a

$$\begin{aligned} \frac{\partial^2}{\partial \theta^2} \ln L(\theta|x_1, ..., x_n)\Big|_{\overline{X}} &= -\frac{1}{\theta^2} \sum x_i - \frac{n - \sum x_i}{(1-\theta)^2}\Big|_{\overline{X}} \\ &= -\frac{n\overline{X}}{\overline{X}^2} - \frac{n - n\overline{X}}{(1-\overline{X})^2} \end{aligned}$$

y los sumandos de esta expresión son obviamente negativos, por serlos sus numeradores, luego $\overline{X}$ es un máximo local que al ser interior al recinto y el único máximo local, es el máximo global. Por tanto $\widehat{\theta} = \overline{X}$ es el EMV.

■

En general si $\theta \in \mathbb{R}^r$, se considera el sistema

$$\left.\begin{array}{c} \frac{\partial}{\partial \theta_j} \ln L(\theta_1, ..., \theta_r|x_1, ..., x_r) = 0 \\ j = 1, ..., r \end{array}\right\} \tag{3.6.2.1}$$

Estas ecuaciones reciben el nombre de *ecuaciones de verosimilitud* y a sus soluciones se les llama EMV en sentido débil.

Se advierte que no siempre ambos estimadores tienen que coincidir. Además, y curiosamente, las buenas propiedades que poseen los estimadores de máxima verosimilitud radican más bien en satisfacer la Ecuación (3.6.2.1) que en su caracter de ser máximos, como se verá en la Sección 3.6.3.

Si la función de verosimilitud es una función cóncava de θ para todo punto $\overrightarrow{x} = (x_1, ..., x_n)$ y su máximo no se alcanza en la frontera de Θ, entonces el sistema (3.6.2.1) tiene como solución el EMV.

Ejemplo 3.6.2.2
Para m.a.s. de tamaño n de una población $N(\mu, \sigma)$ con $\overrightarrow{\theta} = (\mu, \sigma)$ es

$$L(\overrightarrow{\theta}\,|\,\overrightarrow{x}) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum_1^n (x_i-\mu)^2}$$

para maximizar esta función, se puede utilizar que $\sum_1^n (x_i-a)^2 \geq \sum_1^n (x_i-\overline{x})^2$, con lo que el problema queda reducido a una dimensión con $\widehat{\mu} = \overline{x}$. La obtención del máximo en σ de la función

$$\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum_1^n (x_i-\overline{x})^2}$$

puede hacerse tomando logaritmos neperianos y derivando respecto a σ, con lo que se obtiene

$$\frac{\partial}{\partial\theta} \ln L(\overrightarrow{\theta}\,|\,\overrightarrow{x}) = -\frac{n}{\sigma} + \frac{\sum (x_i - \overline{x})^2}{\sigma^3}$$

expresión que igualada a cero lleva a $\widehat{\sigma} = \sqrt{b_2}$. Como la derivada segunda es

$$\frac{\partial^2}{\partial\theta^2} = \ln L(\overrightarrow{\theta}\,|\,\overrightarrow{x}) = \frac{n}{\sigma^2} - \frac{3n}{\sigma^4} b_2$$

su particularización en $\sigma = b_2^{1/2}$ es -$\frac{2n}{b_2}$ que al ser negativa, hace que $\widehat{\sigma} = b_2^{1/2}$ sea un mínimo local, pero como cuando σ tiende a $+\infty$ y a $-\infty$ la función a maximizar tiende a cero, se convierte en un mínimo global. Luego el EMV es $\overrightarrow{\theta} = (\overline{x}, \sqrt{b_2})$ donde $b_2 = \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})^2$ es la varianza muestal.

Ejemplo 3.6.2.3
Para m.a.s. de tamaño n de una población $Uniforme(0, \theta)$ con $\theta \geq 0$, la función de densidad es $f(x|\theta) = \frac{1}{\theta} I_{[0,\theta]}(x)$, que se puede poner como

$$L(\theta|x) = \frac{1}{\theta} I_{[x,+\infty)}(\theta) I_{(-\infty,x]}(0)$$

si se emplea ahora que

$$I_{(a,b)}(x)I_{(c,d)}(c) = I_{(\text{máx}\{a,c\},\text{mín}\{b,d\})}(x)$$

se sigue que

$$\begin{aligned} L(\theta|\overrightarrow{x}) &= \frac{1}{\theta^n}\prod_1^n I_{[x_i,+\infty)}(\theta)\prod_1^n I_{(-\infty,x_i]}(0) \\ &= \frac{1}{\theta^n} I_{[x_{(n)},+\infty)}(\theta) I_{(-\infty,x_{(1)}]}(0) \end{aligned}$$

con lo que la función de verosimilitud es la de la Figura 3.6.1

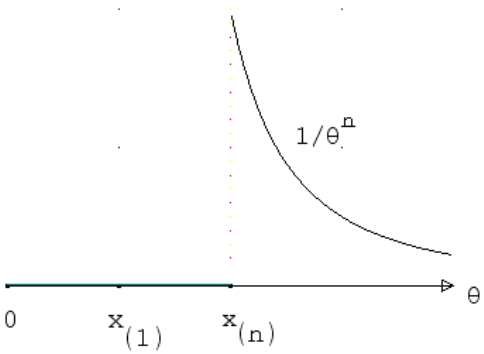


Figura 3.6.1. Función de verosimilitud para una población $U(0,\theta)$.

El máximo en $[0,+\infty)$ se obtiene en $\widehat{\theta} = x_{(n)}$. Este estimador es siempre menor que θ, y no es centrado ya que

$$E[X_{(n)}] = \frac{n}{n+1}\theta.$$

■

Este ejemplo, pone de manifiesto la drástica diferencia existente en considerar la distribución de probabilidad como función de θ en lugar de como función de las observaciones $\overrightarrow{x}$.

Ejemplo 3.6.2.4
Para m.a.s. de tamaño n de una población $U(\theta-\frac{1}{2},\theta+\frac{1}{2})$ se tiene, análogamente al ejemplo anterior, que

$$\begin{aligned} L(\theta|x) &= I_{[\theta-\frac{1}{2},\theta+\frac{1}{2}]}(x) \\ &= I_{[x-\frac{1}{2},x+\frac{1}{2}]}(\theta) \end{aligned}$$

con lo que

$$L(\theta|x_1,...,x_n) = I_{[x_{(n)}-\frac{1}{2},x_{(1)}+\frac{1}{2}]}(\theta)$$

y por tanto cualquier valor comprendido en el intervalo $[x_{(n)} - \frac{1}{2}, x_{(1)} + \frac{1}{2}]$ es EMV, en particular lo es $\frac{1}{2}(x_{(1)} + x_{(n)})$.

Por consiguiente, no es cierto que el EMV sea siempre un estadístico suficiente.

■

Se pasa a estudiar ahora algunas propiedades del EMV.

Proposición 3.6.2.1
Dada la familia $\mathcal{P}$ de distribuciones de probabilidad de la muestra, el EMV siempre es función del estadístico suficiente.

Demostración: se emplea notación de función de densidad, lo mismo es válido para el caso discreto. Se tiene que por el teorema de factorización, teorema 2.2.1, ha de ser

$$f(x_1,...,x_n|\theta) = h(x_1,...,x_n)g_\theta(T(x_1,...,x_n))$$

donde ambas funciones son no negativas. Luego maximizar en θ el primer miembro, es equivalente a maximizar en θ la función g del segundo, con lo que se obtiene $\widehat{\theta} = \widehat{\theta}(T)$.

■

La proposición no dice que el EMV sea el mismo suficiente, sino que es función del suficiente, véase el Ejemplo 3.5.2.4 donde el estadístico suficiente es $(X_{(1)}, X_{(n)})$ no siéndolo $\frac{1}{2}(X_{(1)} + X_{(n)})$.

La siguiente propiedad va a ser la de invariancia del EMV, pero invariancia en sentido diferente al empleado anteriormente para los estimadores, véase la Sección 2.4. En concreto se trata de estudiar cómo se comporta el EMV cuando se realiza un cambio en el parámetro θ mediante una función $h(\theta) = \lambda$. Supóngase que se está interesado inicialmente en estudiar la desviación típica de la población y posteriormente se quiere encontrar el EMV de la varianza, lo que se va a obtener es que el EMV para $\lambda = \sigma^2$ es el cuadrado del EMV para $\theta = \sigma$.

Proposición 3.6.2.2
Dada la familia $\mathcal{P} = \{f(\overrightarrow{x}|\theta)|\theta \in \Theta\}$ de funciones de densidad o de masa, si se nota por $L(\theta|\overrightarrow{x})$ la función de verosimilitud para θ, mediante una reparametrización dada por la función $h : \Theta \longrightarrow \Lambda$ se llama verosimilitud inducida por la función h a

$$M(\lambda|\overrightarrow{x}) = \sup_{\theta\in\Theta_\lambda} L(\theta|\overrightarrow{x})$$

donde

$$\Theta_\lambda = \{\theta \in \Theta|h(\theta) = \lambda\}.$$

Se sigue que el EMV para λ con la verosimilitud inducida, viene dado por $\widehat{\lambda} = h(\widehat{\theta})$, donde $\widehat{\theta}$ es el EMV para θ con la verosimilitud L.

Demostración: basta observar que

$$\begin{aligned} M(\widehat{\lambda}|\overrightarrow{x}) = & \sup_{\theta|h(\theta)=\widehat{\lambda}} L(\theta|\overrightarrow{x}) \geq L(\widehat{\theta}|\overrightarrow{x}) = \\ & = \sup_{\theta} L(\theta|\overrightarrow{x}) = \sup_{\lambda\in\Lambda} \sup_{\theta\in\Theta_\lambda} L(\theta|\overrightarrow{x}) \geq \\ & \geq M(\widehat{\lambda}|\overrightarrow{x}) \end{aligned}$$

luego todas las desigualdades se convierten en igualdades y se obtiene el resultado deseado.

■

Obsérvese que en realidad, siempre que se haga una reparametrización, se debería construir la función de verosimilitud inducida M. En la práctica esto no se hace, aunque se supone para que la proposición sea válida.

Ejemplo 3.6.2.5
Ciertas componentes de un sistema tienen duración exponencial de media θ

$$f(x|\theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}, \qquad x > 0 \quad \text{con} \quad \theta > 0.$$

Para m.a.s. de tamaño n se desea estimar por el método de la máxima verosimilitud la probabilidad de que una componente falle antes del tiempo t, con t un valor conocido.

El parámetro de interés es

$$h(\theta) = P\{X \leq t|\theta\} = 1 - \exp\left\{-\frac{t}{\theta}\right\}.$$

Para obtener el EMV para θ, la verosimilitud viene dada por

$$L(\theta|\overrightarrow{x}) = \theta^{-n} \exp\left\{-\frac{1}{\theta}\sum_{i=1}^{n} x_i\right\}$$

de donde

$$\ln L(\theta|\overrightarrow{x}) = -n \ln\theta - \frac{1}{\theta}\sum_{i=1}^{n} x_i$$

con lo que

$$\frac{\partial}{\partial\theta} \ln L(\theta|\overrightarrow{x}) = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i \tag{3.6.2.2}$$

y se obtiene la solución $\widehat{\theta} = \overline{X}$ al igualar a cero la expresión anterior.

Como la derivada segunda es

$$\frac{\partial^2}{\partial\theta^2}\ln L(\theta|\overrightarrow{x}) = \frac{n}{\theta^2} - \frac{2}{\theta^3}\sum x_i,$$

que particularizada en $\overline{X}$ toma el valor $-n\overline{X}^{-2}$ entonces la Expresión (3.6.2.2) tiene un máximo local, que se convierte en global al tender a cero la verosimilitud cuando θ tiende a $\pm\infty$, con lo que $\overline{X}$ es el EMV para θ. De acuerdo con la Proposición 3.6.2.2 el EMV para $h(\theta)$ es

$$h(\hat{\theta}) = 1 - \exp\left\{-t\overline{X}^{-1}\right\}.$$

3.6.3. Propiedades asintóticas de los estimadores de máxima verosimilitud

Esta sección está dedicada a ver, que bajo condiciones bastante generales, los estimadores de máxima verosimilitud (EMV) son consistentes y asintóticamente normales.

Teorema 3.6.3.1 (consistencia del EMV)
Si Θ es un subconjunto de $\mathbb{R}$, todas las funciones $f(\overrightarrow{x}|\theta)$ tienen soporte común y $\ln f(\overrightarrow{x}|\theta)$ es derivable en un conjunto abierto A de Θ, existe una sucesión de raíces $\widehat{\theta}_n(\overrightarrow{x})$ de la ecuación de verosimilitud

$$\frac{\partial}{\partial\theta}\ln f(\overrightarrow{x}|\theta) = 0 \tag{3.6.3.1}$$

tal que para todo θ de A, $\widehat{\theta}_n$ converge en probabilidad a θ. Es decir la sucesión $\widehat{\theta}_n$ es consistente.

Demostración: se trata de comprobar que para todo $\varepsilon > 0$

$$\lim_{n\to\infty} P\{\overrightarrow{x}_n; |\widehat{\theta}_n(\overrightarrow{x}_n) - \theta| > \varepsilon|\theta\} = 0. \tag{3.6.3.2}$$

Para cualquier valor θ de A y cualquier δ positivo se sigue que

$$\int_{\mathcal{X}} f(x|\theta)\ln\frac{f(x|\theta\pm\delta)}{f(x|\theta)}dx \leq 0 \tag{3.6.3.3}$$

ya que $\ln y \leq y - 1$. Por tanto

$$\frac{1}{n}\sum_{i=1}^{n}(\ln f(x_i|\theta\pm\delta) - \ln f(x_i|\theta))$$

converge en probabilidad θ a

$$E_\theta[\ln f(X|\theta\pm\delta) - \ln f(X|\theta)] = E_\theta\left[\ln\frac{f(X|\theta\pm\delta)}{f(X|\theta)}\right]$$

que es una cantidad negativa por (3.6.3.3). De manera que con probabilidad uno bajo θ, ha de ser

$$\ln L(\theta \pm \delta|\overrightarrow{x}) - \ln L(\theta|\overrightarrow{x}) \leq 0.$$

En resumen la función $\ln L(\theta|\overrightarrow{x})$ es cóncava en el intervalo $(\theta-\delta, \theta+\delta)$ como indica la Figura (3.6.3.1) y al ser una función continua en θ, ha de haber un punto en el abierto $\widehat{\theta}_n$ con tangente horizontal con lo que se cumple (3.6.3.2).

■

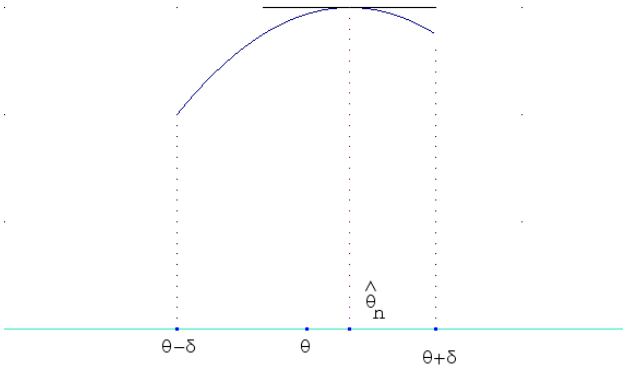


Figura 3.6.3.1. Representación del logaritmo de la función de verosimilitud.

De todas formas el teorema es útil, sólo cuando existe una única raíz de la ecuación de verosimilitud, ya que entonces esa raíz ha de ser el estimador consistente, en caso contrario no se sabría cuál de las raíces al variar n, va a dar lugar a la sucesión consistente.

Teorema 3.6.3.2 (distribución asintótica del EMV)
Para m.a.s. de una población con función de densidad $f(x|\theta)$, si se verifican las dos condiciones siguientes:

H1) En un intervalo abierto A de $\Theta \subset \mathbb{R}$ existe $\frac{\partial^3}{\partial\theta^3}\ln f(x|\theta)$ y está acotada por una función $M(x)$ integrable, es decir $\left|\frac{\partial^3}{\partial\theta^3}\ln f(x|\theta)\right| < M(x)$ con $E_\theta[M(X)] < k$ para cualquier θ perteneciente a A.

H2) Existe al menos un valor θ_0 de A con $E_{\theta_0}\left[\frac{\partial}{\partial\theta}\ln f(X|\theta_0)\right] = 0$, $E_{\theta_0}\left[\frac{1}{f(X|\theta_0)}\frac{\partial^2}{\partial\theta^2}f(X|\theta_0)\right] = 0$ y $0 < I_1(\theta_0) < +\infty$, donde $I_1(\theta_0)$ es la información de Fisher, véase (3.5.1).

Entonces cualquier sucesión consistente $\widehat{\theta}_n$ de raíces de la ecuación de verosimilitud verifica la siguiente convergencia en ley

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} Z \sim N\left(0, \frac{1}{\sqrt{I_1(\theta_0)}}\right) \tag{3.6.3.5}$$

Demostración: si se denota por $s(\theta) = \frac{\partial}{\partial\theta} \ln f(\overrightarrow{x}|\theta)$ y se desarrolla esta función alrededor del valor θ_0 citado en H2), se obtiene

$$s(\widehat{\theta}_n) = s(\theta_0) + s'(\theta_0)(\widehat{\theta}_n - \theta_0) + s"(\theta_n^*)\frac{(\widehat{\theta}_n - \theta_0)^2}{2}$$

donde $|\widehat{\theta}_n - \theta_n^*| \leq |\widehat{\theta}_n - \theta_0|$, y al ser $s(\widehat{\theta}_n) = 0$ se puede poner:

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) = -\frac{\frac{1}{\sqrt{n}}s(\theta_0)}{\frac{1}{n}\left[s'(\theta_0) + s"(\theta_n^*)\frac{\widehat{\theta}_n - \theta_0}{2}\right]} \tag{3.6.3.6}$$

En esta última expresión, se tiene que

$$\frac{1}{n}s'(\theta_0) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\ln f(x_i|\theta_0)$$

que converge con probabilidad uno, bajo θ_0, a $-I_1(\theta_0)$. Por otra parte

$$\frac{1}{n}|s"(\theta_n^*)| \leq \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\partial^3}{\partial\theta^3}\ln f(x_i|\theta_n^*)\right| \leq \frac{1}{n}\sum_{i=1}^{n}M(x_i)$$

y de acuerdo con H1) $\frac{1}{n}|s"(\theta_n^*)| < k$. Como además $\widehat{\theta}_n$ converge en probabilidad a θ_n, se sigue que $s"(\theta_n^*)\frac{(\widehat{\theta}_n - \theta_0)}{2}$ converge en ley a cero, por lo tanto el denominador de la expresión (3.6.3.6) converge en probabilidad, bajo θ_0, a $-I_1(\theta_0)$. Para el numerador, mediante el teorema central del límite aplicado a $\frac{1}{n}s(\theta_0) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln f(x_i|\theta_0)$ se tiene la convergencia en ley bajo θ_0

$$\frac{1}{\sqrt{n}}s(\theta_0) \xrightarrow{\mathcal{L}} Z \sim N\left(0, \sqrt{I_1(\theta_0)}\right)$$

de donde se cumple (3.6.3.5), sin más que aplicar el Teorema de Slutsky (véase el Teorema 1.4.5).

■

Una condición que permite asegurar que se cumple H2) puede verse en Lehmann y Casella (1998), pág. 450.

Ejemplo 3.6.3.1

Para m.a.s. de tamaño n de una población $N(\theta, \sigma)$ con $\theta \in \mathbb{R}$, la distribución asintótica de la media muestral $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n}X_i$ es tal que $\sqrt{n}\frac{\overline{X}_n - \theta}{\sigma}$ converge en ley bajo θ a una v.a. $N(0,1)$. Este resultado, que ya se sabía por el teorema central del límite, puede obtenerse mediante la aplicación del Teorema 3.6.3.2; ya que como existe $\frac{\partial^3}{\partial\theta^3}\ln f(x|\theta)$ y vale cero, se da la

acotación H1) con $M(x) = 0$. También se verifica H2) pues para todo $\theta < 0$ $E_\theta\left[\frac{\partial}{\partial\theta}\ln f(X|\theta)\right] = \frac{1}{\sigma^2}E_\theta[X-\theta] = 0$.

$$E_\theta\left[\frac{1}{f(x|\theta)}\frac{\partial^2}{\partial\theta^2}f(X|\theta)\right] = E_\theta[X-\theta] = 0$$

y el valor de la información de Fisher es $I_1(\theta) = \frac{1}{\sigma^2}$, por lo tanto al ser $\overline{X}_n$ el EMV se tiene la convergencia enunciada.

■

En general se estará interesado en estimar una función $h(\theta)$, si ésta tiene derivada continua en θ_0 y $\widehat{\theta}_n$ es el EMV para θ, puede probarse que se obtiene la convergencia en ley bajo θ_0 de

$$\sqrt{n}\frac{h(\widehat{\theta}_n) - h(\theta_0)}{\frac{|h'(\theta_0)|}{\sqrt{I_1(\theta_0)}}} \longrightarrow Z \sim N(0,1) \tag{3.6.3.7}$$

(véase el Ejercicio 3.7.31)

Ejemplo 3.6.3.2
Para m.a.s. de tamaño n de una población $X \sim Bernoulli(\theta)$ se puede obtener un estimador del cociente entre la probabilidad de éxito y fracaso $h(\theta) = \frac{\theta}{1-\theta}$; como el EMV para θ es $\overline{X}_n$, se tiene que el estimador para $h(\theta)$ es $h(\overline{X}) = \frac{\overline{X}}{1-\overline{X}}$ y es tal que $\sqrt{n}\frac{\frac{\overline{X}}{1-\overline{X}} - \frac{\theta}{1-\theta}}{\sqrt{\frac{\theta}{1-\theta}}} \longrightarrow Z \sim N(0, 1-\theta)$ en ley bajo θ.

■

Como se verá en el Capítulo 4 al estudiar las regiones de confianza, la utilización de la distribución asintótica de los EMV, permitirá la determinación de regiones de confianza aproximadas.

3.6.4. Método bayesiano

Como ya se ha dicho en la Sección 3.2.3, en situaciones donde tenga sentido considerar a θ como una variable aleatoria con función de densidad o de masa dada por $\pi(\theta)$, toda la información sobre θ suministada por la m.a.s. de tamaño n queda recogida en la distribución final $\pi(\theta|x_1, ..., x_n)$. Esta distribución es realmente el estimador bayesiano es decir, es el resultado de utilizar el método de estimación bayesiano.

Si se quiere utilizar un valor de θ en lugar de toda la distribución de probabilidad final, se puede utilizar $E[\theta|x_1, ..., x_n]$ la media de la distribución final. La justificación para esta decisión está en utilizar la PFE con función de pérdida cuadrática y tomar como estimador Bayes el que minimiza la pérdida final esperada, véase la Sección 3.3.2, a este estimador por extensión se le llama también estimador Bayes o estimador bayesiano, como se dijo

en la sección citada. Si se utiliza otra función de pérdida el mécanismo es el mismo, así por ejemplo, si la función de pérdida es el valor absoluto, el estimador bayesiano que minimiza la PFE es la mediana de la distribución final, véase el Ejercicio 3.8.14.

Ejemplo 3.6.4.1
Para una población de $Bernoulli(\theta)$, con m.a.s. de tamaño n, si la información inicial sobre θ viene recogida por la distribución uniforme

$$\pi(\theta) = I_{(0,1)}(\theta)$$

el estimador bayesiano, la distribución final es

$$\pi(\theta|x_1,...,x_n) = \frac{\theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i} I_{(0,1)}(\theta)}{\int_0^1 \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i} d\theta}$$

la distribución $Beta(\theta|\sum x_i + 1, n - \sum x_i + 1)$, y por extensión el estimador bayesiano es la media de la distribución final

$$T(x_1,...,x_n) = (\sum x_i + 1)/(n+1).$$

3.6.5. Propiedades asintóticas de los estimadores bayesianos

Al igual que se ha hecho al estudiar el método de la máxima verosimilitud y sus propiedades asintóticas, se van a ver condiciones que aseguren que los estimadores bayesianos son consistentes y asintóticamente eficientes.

Teorema 3.6.5.1 (distribución asintótica de la distribución final)
Dada una m.a.s. con función de densidad $f(x|\theta)$, si se verifican las condiciones H1) y H2) del Teorema 3.6.3.1 y además se cumplen las condiciones:

H3) la función de densidad inicial es continua y positiva para todo $\theta \in \Theta$, con

$$\int |\theta|\pi(\theta)d\theta < +\infty$$

H4) para cada $\delta > 0$, existe $\varepsilon > 0$ tal que la probabilidad del suceso

$$\sup_{|\theta-\theta_0|\geq\delta} \left\{ \frac{1}{n}(\ln f(x|\theta) - \ln f(x|\theta_0)) \right\} \leq -\varepsilon$$

tiende a 1 cuando $n \to \infty$.

Entonces si $\widetilde{\theta}_n$ es el estimador bayesiano para $\pi(\theta)$ con pérdida cuadrática, se tiene la siguiente convergencia en ley bajo θ_0

$$\sqrt{n}(\widetilde{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} Z \sim N(0, 1/\sqrt{I_1(\theta_0)}) \tag{3.6.5.1}$$

con lo que $\widetilde{\theta}_n$ es consistente y asintóticamente eficiente.

Demostración: Puede verse en Lehmann y Casella (1998), pág. 490.
■

Es importante señalar que para la familia exponencial uniparamétrica, expresada como

$$f(x_i|\theta) = e^{\theta T(x_i) - A(\theta)} \tag{3.6.5.2}$$

se verifican todas las condiciones de regularidad recogidas en el Teorema 3.6.5.1 y relativas al modelo de muestreo, la comprobación puede verse en Lehmann y Casella (1998), Ejemplo 8.4 pág. 491.

Otras aproximaciones asintóticas pueden verse en Berger (1985) pág. 224.

Es interesante observar la similitud entre la aproximación desde el punto de vista de la distribución en el muestreo del EMV $\widehat{\theta}_n$, que se distribuye asintóticamente $N(\theta_0, (\sqrt{nI_1(\theta_0)})^{-1})$, Teorema 3.6.3.2, y la distribución asintótica del estimador bayesiano $\widetilde{\theta}_n$ que es también $N(\theta_0, (\sqrt{nI_1(\theta_0)})^{-1})$.

Ejemplo 3.6.5.1
Para m.a.s. de tamaño n de una población $N(\theta, 1)$, si la densidad inicial $\pi(\theta)$ es una $N(0, 1)$ se sigue que $\pi(\theta|\overrightarrow{x})$ es una $N\left(\frac{n\overline{x}}{1+n}, \frac{1}{\sqrt{1+n}}\right)$, véase el Ejemplo 3.2.3.2. Como la información de Fisher es $I_1(\theta_0) = 1$ la distribución asintótica de $\sqrt{n}\left(\frac{n\overline{x}}{1+n} - \theta_0\right)$ es una $N(0, 1)$, de acuerdo con el Teorema 3.6.5.1.

3.7. APROXIMACIÓN HISTÓRICA

La emergencia de la estadística matemática puede decirse que se produce entre los años 1885 y 1925. En esta época ya se ha realizado el fuerte desarrollo de la Biología y de la Teoría del Seguro y existen los desarrollos matemáticos de la Teoría de errores y la Teoría de la integración de Riemann, siendo conocidos los trabajos de fundamentación de la matemática debidos a Cantor, Péano y Borel.

Paralelamente, la necesidad de manejar adecuadamente la incertidumbre se hace patente dentro de la Astronomía y la Biología y el campo de los datos se extiende en dominios de la Agronomía, la Biometría, la Psicología y la Economía.

El primer problema de inferencia paramétrica parece que fué planteado por Jacob Bernoulli (1654-1705) en su libro el *Arte de la Conjetura* de 1713. Consistió en tratar de dar un valor numérico al parámetro probabilidad de éxito, asociado a un experimento de Bernoulli.

Algunos años más tarde la obra de Thomas Bayes (1701-1761) publicada póstumamente en 1763 y titulada *Un ensayo hacia la resolución de un*

problema en la doctrina del azar contiene el método de estimación bayesiano que consiste en la utilización de la probabilidad final, véase la Sección 3.6.4 aunque Bayes solo utiliza la distribución inicial uniforme. La primera traducción al castellano del ensayo de Bayes puede verse en Gómez Villegas y otros (2001). El método es redescubierto por Laplace (1749-1827) y generalizado por éste al caso en que la distribución inicial no sea la distribución uniforme.

El método de los momentos es introducido por Karl Pearson (1857-1936) a finales del siglo XIX, para aproximar densidades a nubes de puntos, véase la Sección 2.6.

En 1912 Fisher (1890-1962) introduce la función de verosimilitud y el método de la máxima verosimilitud, para estudiar después el concepto de estimador suficiente y consistente. Más tarde, en 1920 Fisher introduce la noción de información que lleva su nombre y demuestra la independencia de la media y la varianza muestrales en el caso normal. Una exposición algo más extensa acerca de la vida de Fisher y sus contribuciones a la estimación estadística se encuentra en la Sección 4.6.

La sustitución de una colección de puntos por uno sólo, en concreto por la media muestral, se debe inicialmente al astrónomo Tycho Brahe (1546-1601), que trataba de utilizar medidas experimentales para determinar la posición de estrellas en el cielo.

Si se tuviera que citar a alguien como el que más lejos llegó en los aspectos descriptivos de una colección de puntos para obtener conclusiones estadísticas, sin duda sería Adolphe Quetelet (1796-1874). Ligado a la Astronomía en su observatorio de Bruselas, este belga, va a ser el causante, entre otras cosas, del fenómeno llamado *quetelismo*, consistente en aproximar los fenómenos que dan lugar a un número suficiente de datos, por una distribución normal. También se debe a él el concepto de *hombre medio,* que al igual que el de promedio o tipo, permitiría hacer comparaciones entre distintos colectivos. Nace en Gante donde obtiene su doctorado en 1819. Enseña matemáticas en el Ateneo de Bruselas y visita París en 1823; allí recibió instrucción de Laplace sobre probabilidades, de Arago sobre Astronomía y de Bouvard sobre Meteorología. En 1832 funda el observatorio astronómico de Bruselas al que estuvo ligado durante toda su vida. Dos años más tarde fue elegido secretario perpetuo de la Academia Real de Bélgica y en 1851 lanza la idea de un congreso internacional sobre Estadística para aunar y desarrollar trabajos sobre esta ciencia.

En sus *Cartas a su alteza real el duque reinante de Saxe–Coburgo y Gotha sobre la teoría de las probabilidades aplicada a las ciencias morales y polítcas* publicadas en 1846, recoge sus ideas sobre el *hombre medio* y explica su utilidad a la hora de hacer comparaciones. No dejó de tener sus adversarios; asi Cournot afirmaba que el hombe medio, si fuese forzado a existir, podía ser un monstruo, análogamente a como la media de todos los triángulos posibles no tenía por qué ser un triángulo.

De todas formas, el concepto de hombre medio servía para suavizar las desviaciones aleatorias y revelar las regularidades que podían caracterizar las leyes de la llamada en su época *física social*, antecedente de la actual Sociología.

La aproximación de cualquier fenómeno aleatorio por la distribución normal, la otra idea capital de Quetelet, radicaba en el Teorema de límite que había obtenido Laplace, y que de manera informal era interpretado por Quetelet en el sentido de que si las causas accidentales de cualquier fenómeno fueran suficientemente numerosas, el fenómeno en cuestión debería aproximarse a la distribución normal.

De hecho, la conexión entre la distribución normal y la idea de suma de causas independientes y accidentales le llevó a aproximar la distribución normal para una nube de puntos a partir de la distribución binomial.

Lo que es claro es que la contribución de Quetelet dejó allanado el camino, para la modelización de la incertidumbre mediante diferentes modelos estocásticos.

Un estudio más detallado de la vida y de la obra de algunos estadísticos puede consultarse en Gómez Villegas (2004).

3.8. EJERCICIOS

3.8.1

Para m.a.s. de tamaño n de una población de $Bernoulli(\theta)$, si $h(\theta) = \frac{1}{\theta}$ es la función a estimar, demuéstrese que no existe estimador centrado para $h(\theta)$ basado en el estadístico $\sum_{i=1}^{n} X_i$.

3.8.2

Para una m.a.s. de tamaño n de una población de $Bernoulli(\theta)$ encuéntrese un estimador insesgado de $h(\theta) = \theta^2$.

3.8.3

Sea $X = x$ una observación de una población de $Poisson(\theta)$. ¿Existe un estimador insesgado para $h(\theta) = \frac{1}{\theta}$?

3.8.4

Dada una m.a.s. de tamaño n de una población $Uniforme(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, estúdiese si el estimador $X_{(n)} - X_{(1)}$ es centrado para θ.

3.8.5

Para una m.a.s. de una población $N(\theta, 1)$ y siendo $h(\theta) = \theta^2$ la función a estimar, demuéstrese que $T(X_1, ..., X_n) = \overline{X}^2 - \frac{1}{n}$ es el EMV para $h(\theta)$. ¿Es $T(X_1, ..., X_n)$ un estimador consistente?

3.8.6
Supóngase que se tienen dos variables independientes X e Y con media común μ y diferentes varianzas σ_X^2 y σ_Y^2. Entre todos los estimadores centrados de la media del tipo $aX + (1-a)Y$ encuéntrese el valor de a, en términos de las varianzas conocidas σ_X^2, σ_Y^2, que minimice la varianza.

3.8.7
Para una m.a.s. de tamaño n de una población $Beta(\theta, 2)$

$$f(x|\theta) = \theta(\theta+1)x^{\theta-1}(1-x)I_{(0,1)}(x) \quad \text{con} \quad \theta > 0$$

¿Es consistente para θ el estimador $2\overline{X}/(1-\overline{X})$?

3.8.8
Sean $X_1, ..., X_n$ variables aleatorias i.i.d. con distribución de $Poisson(\theta)$ y supuesto que la información inicial sobre θ viene recogida por una distribución $Gamma(\theta|a,p)$, $a > 0$, $p > 0$.

a) Pruébese que la familia *Gamma* es conjugada respecto a la población de *Poisson*.

b) Calcúlese la media y la varianza de la distribución final.

3.8.9
En una población $N(\mu, \sigma)$ con ambos parámetros desconocidos se está interesado en estimar el parámetro σ^2. A partir de m.a.s. de tamaño n, se sabe que si $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ es la cuasivarianza muestral, $(n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1}$.

Si se elige como distribución inicial para σ^2 una densidad de la familia

$$\mathcal{P} = \left\{ \pi(\sigma^2|a,p) = \frac{a^p}{\Gamma(p)} \left(\frac{1}{\sigma^2}\right)^{p+1} e^{-\frac{a}{\sigma^2}},\ 0 < \sigma^2 < +\infty,\ p > 0,\ a > 0 \right\}$$

a) Demuéstrese que la distribución inicial de la v.a. $\frac{1}{\sigma^2}$ es una $Gamma(a,p)$. A la densidad dada por cada elemento de $\mathcal{P}$ se le llama $Gamma\ Inversa(a,p)$.
b) Demuéstrese que la familia $\mathcal{P}$ es conjugada respecto de

$$\mathcal{F} = \left\{ (n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1} \right\}$$

c) Calcúlese la media y la varianza de la distribución final.

3.8.10
Sea $X_1, ..., X_n$ una m.a.s. de una población $N(\mu, \sigma)$. Se definen los estimadores de σ^2

$$T_1(X_1, ..., X_n) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \qquad T_2(X_1, ..., X_n) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

a) Compárese el sesgo de ambos estimadores.

b) Mediante el criterio del ECM compruébese que

$$ECM_{T_1}(\sigma^2) = \frac{(2n-1)\sigma^4}{n^2} \qquad ECM_{T_2}(\sigma^2) = \frac{2\sigma^4}{n-1}$$

c) Pruébese que entre todos los estimadores de la forma $\alpha T_1(X_1, ..., X_n)$, con $\alpha \in \mathbb{R}$, el óptimo, de acuerdo con el criterio del error cuadrático medio, se obtiene para $\alpha = \frac{n}{n+1}$.

3.8.11
Si la información inicial sobre θ viene dada por la función $\pi(\theta)$, bien de densidad o de masa, con momento de orden dos finito y se supone que la pérdida en que se incurre por estimar θ mediante $T(x_1, ..., x_n)$ es la cuadrática $L(\theta, T) = (\theta - T)^2$, demuéstrese que la pérdida final esperada PFE_T es mínima cuando $T(x_1, ..., x_n) = E[\theta|x_1, ..., x_n]$, y que el valor de la pérdida esperada coincide con la varianza de la distribución final.

3.8.12
Calcúlese la PFE para el estimador bayesiano del Eejercicio 3.7.8.

3.8.13
Calcúlese la PFE para el estimador bayesiano del Ejercicio 3.7.9.

3.8.14
Compruébese que si la información inicial sobre θ viene dada por una función de distribución $F(\theta)$, y se supone que la pérdida en que se incurre por estimar θ mediante $T(x_1, ..., x_n)$ es el valor absoluto de la diferencia $L(\theta, T) = |\theta - T|$, entonces la pérdida final esperada PFE_T es mínima, cuando el estimador $T(x_1, ..., x_n) = \theta^*$ es la mediana de la distribución final.

3.8.15
A una persona se le pasa un test de inteligencia, cuyo resultado X se supone que sigue una distribución $N(\theta, \sigma = 10)$, donde θ es su nivel de inteligencia real. Supóngase que en el colectivo a que pertenece la persona, la inteligencia θ se distribuye $N(100, \sigma_0 = 15)$. ¿Cuál sería el estimador bayesiano de θ si se supone pérdida cuadrática? ¿cuánto valdría éste si la persona obtiene como resultado del test el valor 110? ¿cuál es el valor de la PFE correspondiente?

3.8.16
Se tiene la sospecha de que una moneda esta sesgada en el sentido de que su probabilidad de cara es menor que $0'5$. Si esta información se modeliza mediante una distribución $Beta(10, 8)$ como distribución inicial ¿cuál es el estimador bayesiano cuando después de tirar la moneda 50 veces se obtienen 20 caras? Utilícese como función de pérdida el valor absoluto.

3.8.17
La proporción θ de votantes a un determinado partido político en unas elecciones es desconocida y su distribución inicial es una $Beta(4, 10)$.

a) Se toma una m.a.s. de 1000 votantes y se encuentra que 125 van a votar al partido citado. Con pérdida cuadrática, ¿cuál sería el estimador bayesiano para θ?

b) Si se cambia el sistema de muestreo y se necesitan 1000 personas hasta conseguir 125 que voten al partido político ¿cuál sería el estimador bayesiano para θ con pérdida cuadrática?

c) ¿Qué conclusión se obtiene de las respuestas dadas en relación con el principio de verosimilitud?

3.8.18
Sea $X_1, ..., X_n$ una m.a.s. de una población $Exponencial(\theta)$, es decir

$$f(x|\theta) = \theta e^{-\theta x} I_{(0,+\infty)}(x) \quad \text{con} \quad \theta > 0$$

a) Pruébese que la Cota de Frechet-Cramér-Rao para estimadores centrados de θ es $\frac{\theta^2}{n}$.

b) Pruébese que $\frac{n-1}{n\overline{X}}$ es el $ECMV$ para θ, con $ECM = \frac{\theta^2}{n-2}$.

3.8.19
Para m.a.s. de tamaño n de una población con distribución $Binomial\ Negativa(1, \theta)$

$$p(x|r, \theta) = \binom{x+r-1}{x} \theta^r (1-\theta)^x \quad x = 0, 1, 2, ... \ con \quad \theta > 0$$

obténgase el $ECMV$ para $h(\theta) = P\{X = 0|\theta\}$.

3.8.20
Sea $X_1, ..., X_n$ una m.a.s. de una población de $Bernoulli(\theta)$ y sea s un número entero conocido $0 < s < n$. Encuéntrese el $ECMV$ para las funciones:

a) $h(\theta) = \theta^s$.

b) $h(\theta) = \theta^s + (1-\theta)^{n-s}$.

3.8.21
Para m.a.s. de tamaño n de una población $Exponencial(\theta)$

$$f(x|\theta) = \theta e^{-\theta x} I_{(0,+\infty)}(x) \quad \text{con} \quad \theta > 0$$

obténgase el $ECMV$ para $h(\theta) = F(x|\theta) = 1 - e^{-\theta x}$, siendo x un número real fijado de antemano.

3.8.22

Para m.a.s. de tamaño n de poblaciones con las funciones de densidad o masa siguientes, encuéntrese la Cota de Frechet-Cramér-Rao y el estimador eficiente, si existe.

a) $f(x|\theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}} I_{(0,+\infty)}(x) \quad$ con $\theta > 0$.

b) $p(x|\theta) = \theta(1-\theta)^x \quad x = 0, 1, 2, \ldots \quad$ con $\quad 0 < \theta < 1$.

c) $f(x|\theta) = \frac{1}{x\theta\sqrt{2\pi}} e^{-\frac{1}{2\theta^2}(\ln x)^2} I_{(0,+\infty)}(x) \quad$ con $\quad \theta > 0$.

3.8.23

Para m.a.s. de tamaño n de una población con función de densidad dada mediante

$$f(x|\theta) = \frac{x+1}{\theta(\theta+1)} e^{-\frac{x}{\theta}} I_{(0,+\infty)}(x) \quad \text{con} \quad \theta > 0.$$

Obténgase un estimador centrado para $h(\theta) = \frac{(3+2\theta)(2+\theta)}{\theta+1}$ que alcance la Cota de FCR.

3.8.24

Para una población con función de densidad $Gamma(\theta, p)$ con p conocido

$$f(x|\theta) = \frac{\theta^p}{\Gamma(p)} e^{-\theta x} x^{p-1} I_{(0,+\infty)}(x) \quad \text{con} \quad \theta > 0, \quad p > 0$$

calcúlese el estimador de θ para m.a.s. de tamaño n por el método de los momentos.

3.8.25

Dada una observación X de una población $N(0, \sigma)$, se pide:

a) Obténgase un estimador insesgado para σ^2.

b) ¿Cuál es el estimador de máxima verosimilitud para σ?

c) Discútase cómo estimar σ por el método de los momentos.

3.8.26

Para m.a.s. de tamaño n de una población $Uniforme(0, \theta)$ encuéntrense los estimadores de θ mediante el método de la máxima verosimilitud y el método de los momentos.

3.8.27

Determínese, con m.a.s. de tamaño n, el EMV para θ en los siguientes casos:

a) $f(x|\theta) = \frac{1}{\theta} I_{\{1,2,\ldots,\theta\}}(x) \quad$ con θ entero positivo.

b) $f(x|\theta) = \frac{1}{2} e^{-|x-\theta|} \quad$ con $\quad \theta \in \mathbb{R}$.

c) $f(x|\theta) = e^{-x+\theta} I_{[\theta,+\infty)}(x) \quad$ con $\quad \theta \in \mathbb{R}$.

d) $f(x|\theta) = \theta\alpha x^{\alpha-1} e^{-\theta x^\alpha} I_{[0,+\infty)}(x) \quad$ con $\quad \theta > 0 \quad$ y $\quad \alpha$ un número real conocido.

e) $f(x|\theta) = \theta(1-x)^{\theta-1} I_{[0,1]}(x) \quad$ con $\quad \theta \geq 1$.

f) $f(x|\theta) = \frac{\theta}{1-\theta} x^{\frac{2\theta-1}{1-\theta}} I_{(0,1]}(x) \quad$ con $\quad \frac{1}{2} \leq \theta < 1$.

3.8.28
Para m.a.s. de tamaño n de una población $N(\theta, 1)$ sólo se registra si cada observación es menor, igual que cero, o mayor. Si se sabe que ha habido m valores, con $m < n$, menores o iguales que cero, encuéntrese el EMV de θ.

3.8.29
Una distribución normal bivariante tiene de vector de medias (θ_1, θ_2) con $\theta_1 \geq \theta_2$ y matriz de covarianza $\Sigma = I$. A partir de una m.a.s. de tamaño n, obténgase el EMV de (θ_1, θ_2).

3.8.30
Para m.a.s. de tamaño n de una población $N(\theta, \sqrt{\theta})$ con $\theta > 0$, se pide:

a) Demostrar que $\widehat{\theta}$, el EMV para θ, es una raiz de la ecuación de segundo grado $\theta^2 + \theta - a_2 = 0$, donde $a_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$.

b) ¿Cuál es la distribución asintótica de $\widehat{\theta}$?

3.8.31
Demuéstrese que si se dan las condiciones de regularidad del Teorema 3.6.3.2 y se desea estimar una función del parámetro $h(\theta)$ con derivadas continua en θ_0, se tiene la convergencia en ley bajo θ_0

$$\sqrt{n}(h(\widehat{\theta}) - h(\theta_0)) \overset{\mathcal{L}_{\theta_0}}{\to} |h'(\theta_0)|N(0, \frac{1}{\sqrt{I_1(\theta_0)}}).$$

3.8.32
Encuéntrese la distribución final asintótica para m.a.s. de tamaño 50 de una población de $Bernoulli(\theta)$. Particularícese al caso en que en las 50 repeticiones se obtenga 12 éxitos. ¿Cuál sería la distribución final exacta si la densidad inicial es la distribución $Uniforme(0, 1)$?

3.9. SOLUCIÓN A LOS EJERCICIOS

3.9.1
Si existiera un estadístico $T\left(\sum_{i=1}^{n} X_i\right)$ tal que

$$\sum_{x=0}^{n} T(x)\binom{n}{x}\theta^x(1-\theta)^{n-x} = \frac{1}{\theta}$$

el límite del primer miembro de esta igualdad debería coincidir con el límite del segundo miembro, pero esto es imposible ya que

$$\lim_{\theta \to 0} \sum_{x=0}^{n} T(x)\binom{n}{x}\theta^x(1-\theta)^{n-x} = T(0)$$

mientras que

$$\lim_{\theta \to 0} \frac{1}{\theta} = +\infty.$$

3.9.2
Para construir un estimador centrado para θ^2, como $P\{X_i = 0\} = 1 - \theta$ y $P\{X_i = 1\} = \theta$ es claro que

$$T(X_1, ..., X_n) = \begin{cases} 1 & \text{si } X_1 X_2 = 1 \\ 0 & \text{resto} \end{cases}$$

que tiene de media el valor buscado

$$\begin{aligned} E_\theta T &= P\{X_1 X_2 = 1|\theta\} = P\{X_1 = 1|\theta\} P\{X_2 = 1|\theta\} \\ &= \theta^2. \end{aligned}$$

3.9.3
Si existiera debería cumplir

$$\begin{aligned} E_\theta[T(X)] &= T(0)e^{-\theta} + T(1)e^{-\theta}\theta + T(2)e^{-\theta}\frac{\theta^2}{2!} + ... \\ &= \frac{1}{\theta} \end{aligned}$$

multiplicando ambos miembros por θe^θ , debería verificarse

$$T(0)\theta + T(1)\theta + T(2)\frac{\theta^2}{2} + ... = e^\theta$$

lo que es imposible, obviamente, luego no existe estimador insesgado.

3.9.4
Si se denota por $U = X_{(n)}$ se tiene que

$$\begin{aligned} F_{X_{(n)}}(u) &= P\{X_{(n)} \leq u\} = P^n\{X \leq u\} = \\ &= \begin{cases} 0 & \text{si } u < \theta - \frac{1}{2} \\ (u - \theta + \frac{1}{2})^n & \text{si } \theta - \frac{1}{2} \leq u \leq \theta + \frac{1}{2} \\ 1 & \text{si } u \geq \theta + \frac{1}{2} \end{cases} \end{aligned}$$

luego la función de densidad del máximo es

$$g(u) = n\left(u - \theta + \frac{1}{2}\right)^{n-1} I_{[\theta - \frac{1}{2}, \theta + \frac{1}{2}]}(u)$$

de donde

$$E[U] = \int_{\theta - \frac{1}{2}}^{\theta + \frac{1}{2}} un\left(u - \theta + \frac{1}{2}\right)^{n-1} du$$

que haciendo el cambio $z = u - \theta + \frac{1}{2}$ puede verse que

$$\int_0^1 n\left(z + \theta - \frac{1}{2}\right) z^{n-1} dz = \frac{n}{n+1} + \theta - \frac{1}{2}.$$

De manera análoga, si se denota por $V = X_{(1)}$ se sigue que

$$\begin{aligned} F_{X_{(1)}}(v) &= P\{X_{(1)} \leq v\} = 1 - P\{X_{(1)} > v\} \\ &= 1 - P^n\{X > v\} = 1 - \left(\int_v^{\theta+\frac{1}{2}} dx\right)^n \\ &= 1 - \left(\theta + \frac{1}{2} - v\right)^n \end{aligned}$$

luego la función de densidad del mínimo es

$$h(v) = n(\theta + \frac{1}{2} - v)^{n-1} I_{[\theta-\frac{1}{2},\theta+\frac{1}{2}]}(v)$$

de donde

$$E[V] = \int_{\theta-\frac{1}{2}}^{\theta+\frac{1}{2}} vn(\theta + \frac{1}{2} - v)^{n-1} dv$$

que haciendo el cambio $z = \theta + \frac{1}{2} - v$ puede comprobarse que

$$\int_0^1 n(\theta + \frac{1}{2} - z)z^{n-1} dz = \theta + \frac{1}{2} - \frac{n}{n+1}$$

Por lo tanto

$$\begin{aligned} E[X_{(n)} - X_{(1)}] &= \frac{n}{n+1} + \theta - \frac{1}{2} - (\theta + \frac{1}{2}) + \frac{n}{n+1} \\ &= \frac{n-1}{n+1} \end{aligned}$$

luego $X_{(n)} - X_{(1)}$ no es un estimador centrado de θ.

3.9.5

La función de densidad de la muestra es exponencial uniparamétrica

$$\begin{aligned} f(x_1, ..., x_n|\theta) &= \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}\sum_{i=1}^n (x_i-\theta)^2} \\ &= \frac{1}{(\sqrt{2\pi})^n} e^{\theta \sum_{i=1}^n x_i} e^{-\frac{n}{2}\theta^2} e^{-\frac{1}{2}\sum_{i=1}^n x_i^2}. \end{aligned}$$

Mediante el Teorema 2.3.1, como $\theta \in \mathbb{R}$ y contiene un abierto de $\mathbb{R}$ el estadístico $\sum_{i=1}^n X_i$ es completo. El estimador del enunciado es función del estadístico suficiente y completo, la suficiencia se sigue del Teorema 2.2.1, (factorización) y es centrado

$$\begin{aligned} E_\theta\left[\overline{X}^2 - \frac{1}{n}\right] &= E_\theta[\overline{X}^2] - \frac{1}{n} \\ &= \frac{\sigma^2}{n} + \theta^2 - \frac{1}{n} \end{aligned}$$

ya que $V[\overline{X}] = E[\overline{X}^2] - (E[\overline{X}])^2$, luego

$$E_\theta\left[\overline{X}^2 - \frac{1}{n}\right] = \theta^2$$

y según el Teorema 3.4.4 (Lehmann-Scheffé) es el ECMV para θ^2.

Por último es un estimador consistente, ya que $\frac{1}{n}$ tiende a cero y como $\overline{X} \sim N(\theta, \frac{1}{\sqrt{n}})$ se tiene la convergencia en ley

$$\sqrt{n}(\overline{X} - \theta) \xrightarrow{\mathcal{L}} Z \sim N(0,1)$$

luego mediante el Lema 1.4.1 con $h(\overline{X}) = \overline{X}^2, h(\theta) = \theta^2$ se sigue la convergencia

$$\sqrt{n}(\overline{X}^2 - \theta^2) \xrightarrow{\mathcal{L}} |h'(\theta)|Z \sim N(0, 2\theta).$$

Mediante la expresión

$$\overline{X}^2 - \theta^2 = \frac{\overline{X}^2 - \theta^2}{2\theta/\sqrt{n}} \frac{2\theta}{\sqrt{n}}$$

y el Teorema 1.4.5 (Slutsky) se tiene la convergencia en ley

$$\overline{X}^2 - \theta^2 \xrightarrow{\mathcal{L}} 0$$

y por converger en ley a una constante, se tiene también la convergencia en probabilidad.

3.9.6

Si se denota por $T = aX + (1-a)Y$ se tiene

$$V[T] = a^2\sigma_x^2 + (1-a)^2\sigma_y^2$$

ya que X e Y son variables aleatorias independientes. Al imponer

$$\frac{d}{da}V[T] = 2a\sigma_x^2 - 2(1-a)\sigma_y^2 = 0$$

se sigue que ha de ser $a = \sigma_y^2/(\sigma_x^2+\sigma_y^2)$. Como al particularizar en la derivada segunda el valor de a se obtiene un valor positivo, se corresponde con un mínimo.

3.9.7

Por el teorema central del límite, se tiene la convergencia en ley

$$\sqrt{n}(\overline{X} - \frac{\theta}{\theta+2}) \xrightarrow{\mathcal{L}} Z \sim N\left(0, \frac{1}{2+\theta}\sqrt{\frac{\theta}{3+\theta}}\right)$$

si se parametriza en función de $\varphi = \frac{\theta}{\theta+2}$, se tiene alternativamente

$$\sqrt{n}(\overline{X} - \varphi) \xrightarrow{\mathcal{L}} Z \sim N\left(0, \frac{1-\varphi}{2}\sqrt{\frac{2\varphi}{3-\varphi}}\right)$$

mediante el Lema 1.3.1 con $h(\overline{X}) = \frac{2\overline{X}}{1-\overline{X}}$ se tiene la convergencia en ley

$$\sqrt{n}\left(\frac{2\overline{X}}{1-\overline{X}} - \frac{2\varphi}{1-\varphi}\right) \xrightarrow{\mathcal{L}} |h'(\varphi)|N\left(0, \frac{1-\varphi}{2}\sqrt{\frac{2\varphi}{3-\varphi}}\right)$$

y como $h'(\varphi) = 2/(1-\varphi)^2$ se sigue que

$$\sqrt{n}\left(\frac{2\overline{X}}{1-\overline{X}} - \frac{2\varphi}{1-\varphi}\right) \xrightarrow{\mathcal{L}} \frac{1-\varphi}{2}\sqrt{\frac{2\varphi}{3-\varphi}}N(0,1)$$

que si se escribe en función de $\theta = 2\varphi/(1-\varphi)$ toma la forma

$$\sqrt{n}\left(\frac{2\overline{X}}{1-\overline{X}} - \theta\right) \xrightarrow{\mathcal{L}} \frac{\theta+2}{2}\sqrt{\frac{\theta}{\theta+3}}N(0,1).$$

Luego, mediante la aplicación del Teorema de Slutsky, $\frac{2\overline{X}}{1-\overline{X}}$ converge en probabilidad a θ y, por tanto, el estimador es consistente para θ.

Un método de prueba alternativo consiste en utilizar que el EMV para θ es precisamente $2\overline{X}/(1-\overline{X})$ y por lo tanto se conoce su distribución asintótica.

3.9.8

a) De acuerdo con la definición de familia conjugada, Definición 3.2.3.2, se trata de ver que si la información inicial es un miembro de $\mathcal{P} = \{\pi(\theta|\overrightarrow{\alpha}) = \frac{a^p}{\Gamma(p)}e^{-a\theta}\theta^{p-1}$ con $\overrightarrow{\alpha} = (a,p)$ $a>0, p>0\}$ y la distribución de la población es $f(\overrightarrow{x}|\theta) = e^{-n\theta}\frac{\theta^{\Sigma x_i}}{\prod_{i=1}^{n} x_i!}$, $\theta > 0$ la distribución final también es de $\mathcal{P}$.

$$\begin{aligned}
\pi(\theta|x_1,...,x_n) &= \frac{\pi(\theta|a_0,p_0)p(x_1,...,x_n|\theta)}{\int_0^{+\infty}\pi(\theta|a_0,p_0)p(x_1,...,x_n|\theta)d\theta} \\
&= \frac{\frac{a^p}{\Gamma(p_0)}e^{-a_0\theta}\theta^{p_0-1}\frac{1}{\prod_{i=1}^{n} x_i!}e^{-n\theta}\theta^{\sum_{i=1}^{n} x_i}}{\int \frac{a^p}{\Gamma(p_0)}e^{-a_0\theta}\theta^{p_0-1}\frac{1}{\prod_{i=1}^{n} x_i!}e^{-n\theta}\theta^{\sum_{i=1}^{n} x_i}d\theta} \\
&= \frac{e^{(-a_0+n)\theta}\theta^{p_0+\Sigma x_i-1}}{\Gamma(a_0+n, p_0+\Sigma x_i)} = \pi(\theta|\overrightarrow{\alpha}_1)
\end{aligned}$$

con $a_1 = a_0 + n \quad p_1 = p_0 + \Sigma x_i$, con lo que se obtiene como densidad final una en $\mathcal{P}$ y por tanto la familia *Gamma* es conjugada respecto a la población de *Poisson*.

b) De acuerdo con las propiedades de la distribución *Gamma* se tiene

$$E[\theta|x_1,...,x_n] = \frac{p_0 + n\overline{x}}{a_0 + n} \qquad V[\theta|x_1,...,x_n] = \frac{p_0 + n\overline{x}}{(a_0+n)^2}.$$

3.9.9

a) La densidad de la v.a. $x = \frac{1}{\sigma^2}$, al ser $\left|\frac{d\sigma^2}{dx}\right| = x^{-2}$ viene dada por

$$\pi(x) = \frac{a^p}{\Gamma(p)} x^{p+1} e^{-ax} x^{-2} \sim Gamma(a,p).$$

b) Si la densidad inicial es $\pi(\sigma^2|a,p)$, la densidad final viene dada por

$$\begin{aligned}\pi(\sigma^2|S^2) &= \frac{\pi(\sigma^2|a,p) f_{\chi^2_{n-1}}((n-1)\frac{s^2}{\sigma^2}|\sigma^2)}{\int_0^{+\infty} \pi(\sigma^2|a,p) f_{\chi^2_{n-1}}((n-1)\frac{s^2}{\sigma^2}|\sigma^2) d\sigma^2} \\ &= \frac{(\sigma^{-2})^{p+\frac{n-1}{2}} e^{-\sigma^{-2}(a+\frac{n-1}{2}s^2)}}{\int_0^{+\infty} (\sigma^{-2})^{p+\frac{n-1}{2}} e^{-\sigma^{-2}(a+\frac{n-1}{2}s^2)} d\sigma^2} \\ &= \frac{a_1^{p_1}}{\Gamma(p_1)} (\sigma^{-2})^{p_1+1} e^{-\sigma^{-2}a_1}\end{aligned}$$

con $p_1 = p + \frac{n-1}{2} - 1$ y $a_1 = a + \frac{n-1}{2} S^2$. Luego es de la forma $\pi(\sigma^2|a_1,p_1)$.

c) Para el cálculo de la media de la distribución final, se observa que

$$\begin{aligned}E[\sigma^{2k}] &= \frac{a^p}{\Gamma(p)} \int_0^{+\infty} \sigma^{2k} (\sigma^{-2})^{p+1} e^{-a\sigma^{-2}} d\sigma^2 \\ &= \frac{a^p}{\Gamma(p)} \int_0^{+\infty} (\sigma^{-2})^{p-k+1} e^{-a\sigma^{-2}} d\sigma^2\end{aligned}$$

mediante el cambio $x = \sigma^{-2}$ con $\left|\frac{d\sigma^2}{dx}\right| = x^{-2}$, puede ponerse

$$E[\sigma^{2k}] = \frac{a^p}{\Gamma(p)} \int_0^{+\infty} x^{p-k-1} e^{-ax} dx$$

y por propiedades de la función *Gamma*

$$\begin{aligned}E[\sigma^{2k}] &= \frac{a^p}{\Gamma(p)} a^{-p+k} \Gamma(p-k) \\ &= a^k \frac{\Gamma(p-k)}{\Gamma(p)}\end{aligned}$$

con lo que $E[\sigma^2] = \frac{a}{p-1}$, $V[\sigma^2] = \frac{a^2}{(p-1)^2(p-2)}$. Por lo tanto para obtener la distribución final, basta sustituir a_1 y p_1 en estas expresiones.

3.9.10

a) Para el cálculo del sesgo como $T_1 = \frac{n-1}{n}S^2$ es $E[T_1] = \frac{n-1}{n}\sigma^2$ mientras que $E[T_2] = \sigma^2$, luego

$$b_{\sigma^2}(T_1) = -\frac{\sigma^2}{n}\ ,\ \ b_{\sigma^2}(T_2) = 0.$$

b) Como el ECM es igual a la varianza más el cuadrado del sesgo, se tiene

$$ECM_{T_1}(\sigma^2) = V_{\sigma^2}(T_1) + \frac{\sigma^4}{n^2}$$

y como $(n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1}$ es $V[(n-1)\frac{S^2}{\sigma^2}] = 2(n-1)$, de donde $V_{\sigma^2}(T_1) = 2\sigma^4\frac{n-1}{n^2}$ y como $E[T_1] = \frac{n-1}{n}\sigma^2$, se sigue que

$$\begin{aligned} ECM_{T_1}(\sigma^2) &= \frac{\sigma^4}{n^2}2(n-1) + \frac{\sigma^4}{n^2} \\ &= \frac{2n-1}{n^2}\sigma^4. \end{aligned}$$

Al ser $T_2 = S^2$, del hecho de que la varianza de la distribución χ^2_{n-1} sea $2(n-1)$ se sigue que

$$\begin{aligned} ECM_{T_2}(\sigma^2) = V_{\sigma^2}(T_2) &= V_{\sigma^2}\left[\frac{nT_1}{n-1}\right] \\ &= V_{\sigma^2}\left[\frac{\sigma^2}{n-1}\chi^2_{n-1}\right] \\ &= \frac{\sigma^4}{(n-1)^2}2(n-1) \\ &= \frac{2\sigma^4}{n-1}. \end{aligned}$$

c) Si se denota por $T_\alpha = \alpha T_1$, como

$$ECM_{T_\alpha}(\sigma^2) = V_{\sigma^2}(T_\alpha) + b^2_{\sigma^2}(T_\alpha)$$

y

$$\begin{aligned} V_{\sigma^2}(T_\alpha) &= \alpha^2 V_{\sigma^2}(T_1) = \alpha^2 2\sigma^4\frac{n-1}{n^2} \\ b_{\sigma^2}(T_\alpha) &= E_{\sigma^2}(T_\alpha) - \sigma^2 = E_{\sigma^2}(\alpha T_1) - \sigma^2 \\ &= \alpha\frac{n-1}{n}\sigma^2 - \sigma^2 \end{aligned}$$

al sustituir se obtiene

$$\begin{aligned} ECM_{T_\alpha}(\sigma^2) &= \alpha^2 2\sigma^4\frac{n-1}{n^2} + \left(\alpha\frac{n-1}{n}\sigma^2 - \sigma^2\right)^2 \\ &= 2\alpha^2\sigma^4\frac{n-1}{n^2} + \sigma^4\left(\alpha\frac{n-1}{n} - 1\right)^2. \end{aligned}$$

Derivando respecto de α e igualando a cero, se obtiene que $\alpha = \frac{n}{n+1}$, comprobándose fácilmente que se corresponde con un mínimo. Por tanto el estimador con ECM mínimo de la forma αT_1 es $\frac{n}{n+1}T_1 = \frac{1}{n+1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ y su error cuadrático medio es $\frac{2n-1}{(n+1)^2}\sigma^4$.

3.9.11

Si se denota por $T^*(x_1, ..., x_n) = E[\theta|x_1, ..., x_n]$, para cualquier estimador $T(x_1, ..., x_n)$ la pérdida final esperada (PFE) es

$$PFE_T = E[(\theta - T)^2|\overrightarrow{x}] = \int_\Theta (\theta - T)^2 \pi(\theta|\overrightarrow{x})d\theta$$
$$= E[(\theta - T^*)^2|\overrightarrow{x}] + E[(T^* - T)^2|\overrightarrow{x}]$$

ya que

$$2(T^* - T)\int_\Theta (\theta - T^*)\pi(\theta|\overrightarrow{x})d\theta = 0.$$

Como $E[(T^* - T)^2|\overrightarrow{x}] = (T^* - T)^2$ se obtiene que

$$PFE_T = V[\theta|\overrightarrow{x}] + (T^* - T)^2$$

por lo tanto el valor mínimo se obtiene para $T = T^*$ y vale $V[\theta|\overrightarrow{x}]$. Lo que se afirma, es que las desviaciones cuadráticas de una variable aleatoria respecto a un punto se hacen mínimas, cuando el punto es la media de la distribución.

3.9.12

De acuerdo con el Ejercicio 3.8.11, será $PFE_{T^*} = E[\theta|\overrightarrow{x}]$ la media de la distribución final, y como de acuerdo con el Ejercicio 3.8.8 la distribución final es una *Gamma* $(a_1 = a_0 + n, p_1 = p_0 + \sum_{i=1}^{n} x_i)$, el estimador bayesiano es $T^*(\overrightarrow{x}) = \frac{p_0 + \sum_{i=1}^{n} x_i}{a_0 + n}$ y $PFE_{T^*} = V[\theta|\overrightarrow{x}] = \frac{p_0 + \sum_{i=1}^{n} x_i}{(a_0+n)^2}$.

3.9.13

De acuerdo con el Ejercicio 3.8.11, será $PFE_{T^*} = E[\theta|\overrightarrow{x}]$ y por el Ejercicio 3.8.9 apartado c), la media de la distribución *Gamma-Inversa* es $T^*(\overrightarrow{x}) = \frac{a + \frac{n-1}{2}S^2}{p - 2 + \frac{n-1}{2}}$ y

$$PFE_{T^*} = \frac{\left(a + \frac{n-1}{2}S^2\right)^2}{\left(p + \frac{n-1}{2} - 2\right)^2\left(p + \frac{n-1}{2} - 3\right)}.$$

3.9.14

En efecto sea θ^* tal que $F(\theta^*|x_1, ..., x_n) = \frac{1}{2}$, la mediana de la distribución final. Si se supone inicialmente que $\theta^* < T$, con $T = T(x_1, ..., x_n)$ cualquier

valor, se tiene

$$|\theta-\theta^*|-|\theta-T|=\begin{cases}\theta^*-T & \text{si } \theta\le\theta^*\\ 2\theta-(\theta^*+T) & \text{si } \theta^*\le\theta\le T\\ T-\theta^* & \text{si } \theta\ge T\end{cases}$$

luego se obtiene la siguiente acotación

$$|\theta-\theta^*|-|\theta-T|\le(\theta^*-T)I_{(-\infty,\theta^*]}(\theta)+(T-\theta^*)I_{[\theta^*,+\infty)}(\theta)$$

Si se integra en los dos miembros respecto a la distribución final

$$\int_\Theta(|\theta-\theta^*|-|\theta-T|)dF(\theta|\overrightarrow{x})\le\int_{-\infty}^{\theta^*}(\theta^*-T)dF(\theta|\overrightarrow{x})+\int_{\theta^*}^{+\infty}(T-\theta^*)dF(\theta|\overrightarrow{x})$$

pero el segundo miembro vale $(\theta^*-T)P\{\theta\le\theta^*|\overrightarrow{x}\}+(T-\theta^*)P\{\theta\ge\theta^*|\overrightarrow{x}\}$ y por la definición de mediana vale cero, por lo tanto siempre es cierto que

$$\int_\Theta|\theta-\theta^*|dF(\theta|\overrightarrow{x})\le\int_\Theta|\theta-T|dF(\theta|\overrightarrow{x}).$$

Se tiene entonces que el mínimo de las desviaciones en valor absoluto respecto a una constante se alcanza para la mediana.

Se observa que si $T<\theta^*$

$$|\theta-\theta^*|-|\theta-T|=\begin{cases}\theta^*-T & \text{si } \theta\le\theta^*\\ (\theta^*+T)-2\theta & \text{si } T\le\theta\le\theta^*\\ T-\theta^* & \text{si } \theta\ge\theta^*\end{cases}$$

y se puede seguir el razonamiento anterior con modificaciones obvias.

3.9.15

La información muestral viene dada por $f(x|\theta)$ una distribución $N(\theta,10)$ y la inicial por $\pi(\theta)$ con distribución $N(100,15)$, entonces, de acuerdo con el Ejemplo 3.3.2.2, la información final $\pi(\theta|x)$ se distribuye $N(\mu_1,\sigma_1)$ con

$$\mu_1=\frac{\frac{\mu_0}{\sigma_0^2}+n\frac{\overline{x}}{\sigma^2}}{\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}}=\frac{\frac{100}{225}+\frac{x}{100}}{\frac{1}{225}+\frac{1}{100}}=\frac{400+9x}{13}$$

$$\sigma_1^2=\left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right)^{-1}=\left(\frac{1}{225}+\frac{1}{100}\right)^{-1}=69'23.$$

El valor del estimador bayesiano con pérdida cuadrática, véase el Ejercicio 3.8.11, es la media de la distribución final μ_1 que cuando $x=110$ vale $(400+9\cdot110)/13=106'923$ y la PFE es la varianza de la distribución final $69'23$.

3.9.16

La información muestral viene dada por $p(x|\theta)$ una distribución de $Bernoulli(\theta)$ y la inicial por $\pi(\theta)$ con distribución $Beta(10,8)$, entonces de acuerdo con el Ejemplo 3.3.2.1 la información final $\pi(\theta|\overrightarrow{x})$ tiene distribución $Beta(\alpha_1,\beta_1)$ donde

$$\alpha_1=\alpha_0+\sum_1^{50}x_i=10+20=30 \qquad \beta_1=\beta_0+n-\sum_1^{50}x_i=8+50-20=38.$$

El valor del estimador bayesiano con pérdida el valor absoluto, véase el Ejercicio 3.8.14, es la mediana de la distribución final. Con el concurso de un paquete estadístico se obtiene que $\theta^*=0'44$.

3.9.17

a) La información inicial viene dada por $\pi(\theta)$, una $Beta(4,10)$ y la información muestral $f(x|\theta)$ es una distribución de $Bernoulli(\theta)$, por tanto la distribución final $\pi(\theta|\overrightarrow{x})$ se distribuye mediante una $Beta(\alpha_0+\sum_{i=1}^{1000}x_i,\beta_0+1000-\sum_{i=1}^{1000}x_i)=Beta(129,885)$. De donde el estimador bayesiano es la media de la distribución final y vale $0'127$.

b) La información inicial es la misma que en la parte a) y la distribución muestral es $P\{y|r,\theta\}=\binom{y-1}{r-1}\theta^r(1-\theta)^{y-r}$ con $y=r,r+1,...$ donde y es el número de votantes examinados hasta obtener por primera vez 125 que voten al partido en estudio; esta variable tiene una distribución *Binomial Negativa*$(125,\theta)$.

La distribución final

$$\pi(\theta|y=1000)=\frac{\frac{1}{B(\alpha_0,\beta_0)}\theta^{\alpha_0-1}(1-\theta)^{\beta_0-1}\binom{y-1}{r-1}\theta^r(1-\theta)^{1-r}}{\int_0^1\frac{1}{B(\alpha_0,\beta_0)}\theta^{\alpha_0-1}(1-\theta)^{\beta_0-1}\binom{y-1}{r-1}\theta^r(1-\theta)^{1-r}d\theta}$$

de donde el estimador bayesiano es el mismo de la parte a).

c) Las dos funciones de verosimilitud coinciden en la parte que dependen de θ luego la evidencia que se obtiene, mediante el estimador bayesiano, es idéntica, de donde se sigue que el estimador bayesiano cumple el principio de verosimilitud.

3.9.18

a) Al tratarse de la familia exponencial uniparamétrica

$$f(x_1,...,x_n|\theta)=\theta^n e^{-\theta\sum_{i=1}^n x_i}$$

como el soporte es independiente de θ y la función $\pi(\theta)=-\theta$ tiene derivada no nula y continua, se verifican las dos condiciones de regularidad, ver comentario tras el Teorema 3.5.2. Además

$$I_1(\theta)=-E\left[\frac{\partial^2}{\partial\theta^2}\ln f(X|\theta)\right]=\frac{1}{\theta^2}$$

ya que $\frac{\partial^2}{\partial\theta^2}\ln f(X|\theta) = -\frac{1}{\theta^2}$, por tanto la cota para $h(\theta) = \theta$ es

$$Cota\ de\ FCR = \frac{1}{I_n(\theta)} = \frac{\theta^2}{n}.$$

b) Por el Teorema 2.3.1 es $\overrightarrow{c}(\theta) = -\theta$, y como esta imagen contiene un rectángulo abierto de $\mathbb{R}$, luego $\sum_{i=1}^{n} X_i$ es un estadístico completo. El resto se sigue del Ejemplo 3.4.4.

Por el Teorema 3.5.2 se sigue que como en este caso se está en la familia exponencial uniparamétrica, se alcanza la Cota de FCR pero para la función $h(\theta) = \frac{1}{\theta}$ ya que

$$\frac{\partial}{\partial\theta}\ln f(\overrightarrow{x}|\theta) = -n\left(\overline{X} - \frac{1}{\theta}\right)$$

de ahí la aparente contradicción entre las partes a) y b).

3.9.19

La función de masa de la muestra es

$$P\{x_1, ..., x_n|\theta\} = \theta^n(1-\theta)^{\sum_{i=1}^{n} x_i}$$

de donde el estadístico $\sum_{i=1}^{n} X_i$ es suficiente.

Además, puesta la función de masa en la forma

$$P\{x_1, ..., x_n|\theta\} = \theta^{n\sum_{i=1}^{n} x_i \ln(1-\theta)}$$

y como la imagen de $\ln(1-\theta)$ contiene un abierto de $\mathbb{R}$, el estadístico $\sum_{i=1}^{n} X_i$ es también completo, por el Teorema 2.3.1.

Como $E\left[\sum_{i=1}^{n} X_i\right] = n\frac{1-\theta}{\theta}$ se puede utilizar el teorema de Lehmann-Scheffé, Teorema 3.4.4. Para ello un estimador centrado, sencillo, para θ es

$$T(X_1, ..., X_n) = \begin{cases} 1 & \text{si } X_1 = 0 \\ 0 & \text{resto} \end{cases}$$

todo se reduce a calcular $E[T|S=s]$ que será el $ECMV$ para θ.

$$E[T|S=s] = P\{X_1=0|\sum_{i=1}^{n} X_i = s\} = \frac{P\{X_1=0, \sum_{i=1}^{n} X_i = s\}}{P\{\sum_{i=1}^{n} X_i = s\}}$$

$$= \frac{P\{X_1=0\}P\{\sum_{i=2}^{n} X_i = s\}}{P\{\sum_{i=1}^{n} X_i = s\}} = \frac{\theta\binom{s+n-2}{s}\theta^{n-1}(1-\theta)^s}{\binom{n+s-1}{s}\theta^n(1-\theta)^s}$$

$$= \frac{n-1}{n+s-1}$$

ya que $\sum_{i=2}^{n} X_i$ tiene distribución $Binomial\ Negativa(n-1,\theta)$ y $\sum_{i=1}^{n} X_i$ tiene distribución $Binomial\ Negativa(n,\theta)$ con lo que $E[T|S=s] = \frac{n-1}{n+s-1}$. Por lo que el estimador pedido es

$$T^*(X_1,...,X_n) = \frac{n-1}{n+\sum_{i=1}^{n} X_i - 1}.$$

3.9.20

La función de masa de la muestra es

$$P\{x_1,...,x_n|\theta\} = \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}$$

$$= (1-\theta)^n e^{\sum_{i=1}^{n} x_i \ln(\theta/1-\theta)}$$

por el Teorema 2.2.1 (de factorización) el estadístico $\sum_{i=1}^{n} X_i$ es suficiente y por el Teorema 2.3.1 la imagen, cuando θ recorre el intervalo [0,1], de $\ln(\theta/1-\theta)$ contiene un abierto de $\mathbb{R}$, luego el estadístico $\sum_{i=1}^{n} X_i$ es además completo.

a) Un estimador centrado para $h(\theta)=\theta^s$ es

$$T_1(X_1,...,X_n) = \begin{cases} 1 & \text{si } X_1=...=X_s=1 \\ 0 & \text{resto} \end{cases}$$

luego por el teorema de Lehmann-Scheffé, Teorema 3.4.4, el $ECMV$ para θ^s es $E\left[T_1|\sum_{i=1}^{n} X_i = t\right]$. Esta esperanza es igual a

$$P\{X_1=...=X_s=1|\sum_{i=1}^{n} X_i = t\} = \frac{P\{X_1=...=X_s=1, \sum_{i=1}^{n} X_i = t\}}{P\{\sum_{i=1}^{n} X_i = t\}}$$

pero esta última expresión vale

$$\frac{P\{X_1 = ... = X_s = 1, \sum_{i=s+1}^{n} X_i = t\}}{P\{\sum_{i=1}^{n} X_i = t\}} = \frac{\theta^s \binom{n-s}{t-s} \theta^{t-s}(1-\theta)^{n-t}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} = \frac{\binom{n-s}{t-s}}{\binom{n}{t}}$$

luego el estimador óptimo para θ^s es

$$T_1^*(X_1, ..., X_n) = \binom{n-s}{\sum_{i=1}^{n} X_i - s} \Big/ \binom{n}{\sum_{i=1}^{n} X_i}$$

b) De manera análoga un estimador centrado para $h(\theta) = (1-\theta)^{n-s}$ es

$$T_2(X_1, ..., X_n) = \begin{cases} 1 & \text{si } X_{s+1} = 0, ..., X_n = 0 \\ 0 & \text{resto} \end{cases}$$

luego el $ECMV$ para $(1-\theta)^{n-s}$ es $E\left[T_2 \middle| \sum_{i=1}^{n} X_i = t\right]$. Es fácil comprobar que esta esperanza vale

$$T_2^*(X_1, ..., X_n) = \binom{s}{\sum_{i=1}^{n} X_i} \Big/ \binom{n}{\sum_{i=1}^{n} X_i}$$

Aplicando el Corolario 3.4.2, el $ECMV$ para $h(\theta) = \theta^s + (1-\theta)^{n-s}$ es

$$\frac{\binom{n-s}{\sum_{i=1}^{n} X_i - s} + \binom{s}{\sum_{i=1}^{n} X_i}}{\binom{n}{\sum_{i=1}^{n} X_i}}$$

3.9.21

Por ser la función de densidad de la muestra

$$f(x_1, ..., x_n|\theta) = \theta^n e^{-\theta \sum_{i=1}^{n} X_i}$$

el estimador $T = \sum_{i=1}^{n} X_i$ es suficiente, Teorema 2.2.1 (factorización) y por contener $-\theta$ un abierto en $\mathbb{R}$, es además completo, Teorema 2.3.1. El estadístico siguiente es centrado para $h(\theta) = 1 - e^{-\theta x}$

$$T(X_1, ..., X_n) = \begin{cases} 1 & \text{si } X_1 \leq x \\ 0 & \text{resto} \end{cases}$$

luego por el Teorema 3.4.4 (Lehmann-Scheffé), el $ECMV$ es $E\left[T \left| \sum_{i=1}^{n} X_i = t \right.\right]$.

Pero esta esperanza vale

$$\begin{aligned} \int_0^x f(x_1| \sum_{i=1}^{n} X_i = t)dx_1 &= \int_0^x \frac{f(x_1, \sum_{i=1}^{n} X_i = t)}{f(\sum_{i=1}^{n} X_i = t)} dx_1 \\ &= \int_0^x f(x_1|\theta) f_{\sum_{i=2}^{n} X_i}(t - x_1|\theta) \frac{1}{f_{\sum_{i=1}^{n} X_i}(t)} dx_1 \end{aligned}$$

y en virtud de la reproductividad de la función *Gamma*

$$\begin{aligned} \sum_{i=1}^{n} X_i &\sim Gamma(\theta, n) \\ \sum_{i=2}^{n} X_i &\sim Gamma(\theta, n-1) \end{aligned}$$

por lo que haciendo operaciones se obtiene que

$$E\left[T \left| \sum_{i=1}^{n} X_i = t \right.\right] = 1 - \left(\frac{\sum_{i=1}^{n} X_i - x}{\sum_{i=1}^{n} X_i} \right)^{n-1}$$

es el $ECMV$ para $1 - e^{-\theta x}$.

3.9.22

a) Por pertenecer $f(x_1, ..., x_n|\theta)$ a la familia exponencial uniparamétrica, si existe estimador T eficiente para $h(\theta)$ debe poderse factorizar

$$\frac{\partial}{\partial \theta} \ln f(x_1, ..., x_n|\theta) = K(\theta)(T - h(\theta))$$

donde

$$K(\theta) = \frac{I_n(\theta)}{h'(\theta)},$$

véase el Corolario 3.5.3. Como es

$$\ln f(x_1,...,x_n|\theta) = -n\ln\theta - \sum_{i=1}^{n} x_i \Big/ \theta$$
$$\frac{\partial}{\partial\theta}\ln f(x_1,...,x_n|\theta) = \frac{n}{\theta^2}(\overline{x}-\theta)$$

de donde el estimador $\overline{X}$ es eficiente para la función $h(\theta)=\theta$. Además $I_n(\theta)=n/\theta^2$ y la Cota de FCR es θ^2/n.

b) Análogamente

$$\ln f(x_1,...,x_n|\theta) = n\ln\theta + \left(\sum_{i=1}^{n} x_i\right)\ln(1-\theta)$$

de donde

$$\frac{\partial}{\partial\theta}\ln f(x_1,...,x_n|\theta) = -\frac{n}{1-\theta}\left(\overline{x}-\frac{1-\theta}{\theta}\right)$$

por el Corolario 3.5.3 al pertenecer la distribución a la exponencial uniparamétrica, si existe estimador T eficiente ha de poderse factorizar, con lo que es $\overline{X}$ el estimador eficiente para la función $h(\theta)=\frac{1-\theta}{\theta}$ y

$$\frac{I_n(\theta)}{h'(\theta)} = -\frac{n}{1-\theta}$$

de donde la Cota de FCR es $\frac{1-\theta}{n\theta^2}$.

c) En este caso la función de densidad de la muestra es

$$f(x_1,...,x_n|\theta) = \frac{1}{\left(\prod_{i=1}^{n} x_i\right)\theta^n(\sqrt{2\pi})^n} e^{-\frac{1}{2\theta^2}\sum_{i=1}^{n}(\ln x_i)^2}$$

luego es exponencial uniparamétrica. Además

$$\frac{\partial}{\partial\theta}\ln f(x_1,...,x_n|\theta) = \frac{n}{\theta^3}\left(\frac{1}{n}\sum_{i=1}^{n}(\ln x_i)^2 - \theta^2\right)$$

por lo que el estimador $\frac{1}{n}\sum_{i=1}^{n}(\ln X_i)^2$ es eficiente para la función de θ $h(\theta)=\theta^2$ y

$$\frac{I_n(\theta)}{h'(\theta)} = \frac{n}{\theta^3}.$$

Por lo tanto es $I_n(\theta) = \frac{2n}{\theta^2}$ y la Cota de FCR es $\frac{2\theta^4}{n}$.

3.9.23
La función de densidad de la muestra es

$$f(x_1,...,x_n|\theta)=\frac{\prod_{i=1}^{n}(x_i+1)}{\theta^n(\theta+1)^n}e^{-\frac{1}{\theta}\sum_{i=1}^{n}x_i};$$

derivando respecto a θ su logaritmo se obtiene

$$\frac{\partial}{\partial\theta}\ln f(x_1,...,x_n|\theta)=\frac{n}{\theta^2}\left(\overline{x}-\frac{\theta(2\theta+1)}{\theta+1}\right)$$

si existe estimador eficiente para alguna función $h(\theta)$, de acuerdo con el Teorema 3.5.2, debe ser para la función $\frac{\theta(2\theta+1)}{\theta+1}$ a menos de una constante; es decir

$$h(\theta)=\frac{\theta(2\theta+1)}{\theta+1}+k$$

e igualando $h(\theta)=\frac{(3+2\theta)(2+\theta)}{\theta+1}$ se obtiene $k=6$. Por lo tanto el estimador eficiente es $\overline{X}+6$.

3.9.24
De acuerdo con lo expuesto en la Sección 3.6.1 $\alpha_1(\theta)=E_\theta[X]=\frac{p}{\theta}$ ha de igualarse al momento muestral de orden uno respecto al origen $\overline{X}$, de donde $\widehat{\theta}=\frac{p}{\overline{X}}$.

3.9.25
a) El estimador insesgado para σ^2 es X^2 ya que $E[X^2]=E[(X-\mu)^2]$ con $\mu=0$ luego $E[X^2]=\sigma^2$.

b) La función de verosimilitud es

$$L(\sigma|x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\frac{x^2}{\sigma^2}}$$

y se trata de maximizarla en σ para x fijo. Para ello si se toman logaritmos

$$\ln L(\sigma|x)=-\ln\sqrt{2\pi}-\ln\sigma-\frac{1}{2}\frac{x^2}{\sigma^2}$$

de donde

$$\frac{\partial}{\partial\sigma}\ln L(\sigma|x)=-\frac{1}{\sigma}+\frac{x^2}{\sigma^2}$$

al igualar a cero se obtiene que $\widehat{\sigma}=|x|$ que si se sustituye en la segunda derivada

$$\frac{\partial^2}{(\partial\sigma)^2}\ln L(\sigma|x)=\frac{1}{\sigma^2}-3\frac{x^2}{\sigma^4}$$

vale $-\frac{2}{x^2}$ luego es un máximo ya que es interior al recinto, salvo que sea $|x|=0$ en cuyo caso no existiría EMV.

c) Según (3.6.1.1) se debe igualar $\alpha_1(\sigma) = a_1$ pero $\alpha_1(\sigma)$ no depende de σ, luego estrictamente no tendría sentido el método. No obstante parece natural escoger el momento de orden dos respecto al origen $\alpha_1(\sigma) = E[X^2] = \sigma^2$ e igualar con el correspondiente momento muestral respecto al origen de orden dos, que al ser la muestra de tamaño uno es X^2, por tanto la ecuación es $\sigma^2 = X^2$ y $\widehat{\sigma} = |X|$.

3.9.26

El EMV es $\widehat{\theta} = X_{(n)}$, véase el Ejemplo 3.6.2.3. Para el método de los momentos hay que igualar $\alpha_1(\theta) = \frac{\theta}{2}$ con $\overline{X}$ de donde $\widehat{\theta} = 2\overline{X}$.

3.9.27

a) La función de verosimilitud es

$$\begin{aligned} L(\theta|x_1,...,x_n) &= \frac{1}{\theta^n}\prod_{i=1}^{n} I_{\{1,2,...,\theta\}}(x_i) \\ &= \frac{1}{\theta^n}\prod_{i=1}^{n} I_Z(x_i) = I_{\{x_{(n)},x_{(n)}+1,...\}}(\theta) \end{aligned}$$

a medida que θ aumenta uno a uno desde $x_{(n)}$ esta función disminuye, luego es $\widehat{\theta} = X_{(n)}$.

b) Se trata de maximizar en θ

$$L(\theta|x_1,...,x_n) = \frac{1}{2^n}e^{-\sum_{i=1}^{n}|x_i-\theta|}$$

o lo que es equivalente, obtener θ que minimice $\sum_{i=1}^{n}|x_i-\theta|$. Pero este valor es la mediana muestral; el valor central si n es impar y cualquier valor comprendido entre los dos centrales si n es par, véase el Ejercicio 3.8.14.

c) La función de verosimilitud es

$$\begin{aligned} L(\theta|x_1,...,x_n) &= e^{-\sum_{i=1}^{n}x_i+n\theta}\prod_{i=1}^{n} I_{[\theta,+\infty)}(x_i) \\ &= e^{-\sum_{i=1}^{n}x_i+n\theta} I_{(-\infty,x_{(1)}]}(\theta) \end{aligned}$$

expresión que, como función de θ, es máxima cuando θ es lo mayor posible, es decir es $\widehat{\theta} = X_{(1)}$.

d) Como la función de verosimilitud es

$$L(\theta|x_1,...,x_n) = (\theta\alpha)^n\left(\prod_{i=1}^{n} x_i^{\alpha-1}\right)e^{-\theta\sum_{i=1}^{n}x_i^{\alpha}}$$

su logaritmo neperiano

$$\ln L(\theta|x_1,...,x_n) = n\ln\theta + n\ln\alpha + \sum_{i=1}^{n}(\alpha-1)\ln x_i - \theta\sum_{i=1}^{n} x_i^{\alpha}$$

si se deriva y se iguala a cero

$$\frac{\partial}{\partial\theta}\ln L(\theta|x_1,...,x_n) = \frac{n}{\theta} - \sum_{i=1}^{n} x_i^{\alpha}$$

queda que es un máximo local en θ el valor $\widehat{\theta} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^{\alpha}\right)^{-1}$, ya que la segunda derivada en él es negativa y vale $-\frac{n}{\theta^2}$, que se convierte en un máximo global al ser interior al recinto.

e) De manera análoga

$$L(\theta|x_1,...,x_n) = \theta^n \prod_{i=1}^{n}(1-x_i)^{\theta-1}$$

su logaritmo neperiano es

$$\ln L(\theta|x_1,...,x_n) = n\ln\theta + (\theta-1)\sum_{i=1}^{n}\ln(1-x_i)$$

con derivada

$$\frac{\partial}{\partial\theta}\ln L(\theta|x_1,...,x_n) = \frac{n}{\theta} + \sum_{i=1}^{n}\ln(1-x_i)$$

su solución al igualar a cero es $\left(-\frac{1}{n}\sum_{i=1}^{n}\ln(1-x_i)\right)^{-1}$, si esta expresión es mayor que 1 es el EMV ya que la derivada segunda es $-\frac{n}{\theta^2}$ y el punto pertenece a la región $\theta \geq 1$ luego es un máximo global. En caso contrario este valor es menor estricto que 1 y por ser la función decreciente el máximo se alcanza en 1, luego en cualquier caso

$$\widehat{\theta} = \text{máx}\left\{1, \left(-\frac{1}{n}\sum_{i=1}^{n}\ln(1-x_i)\right)^{-1}\right\}.$$

f) La función de verosimilitud es

$$L(\theta|x_1,...,x_n) = \left(\frac{\theta}{1-\theta}\right)\prod_{i=1}^{n} x_i^{\frac{2\theta-1}{1-\theta}}$$

su logaritmo neperiano es

$$\ln L(\theta|x_1,...,x_n) = n\ln\theta - n\ln(1-\theta) + \frac{2\theta-1}{1-\theta}\sum_{i=1}^{n}\ln x_i$$

cuya derivada vale

$$\frac{\partial}{\partial\theta}\ln L(\theta|x_1,...,x_n)=\frac{n}{\theta}+\frac{n}{1-\theta}+\frac{1}{(1-\theta)^2}\sum_{i=1}^{n}\ln x_i$$

que tiene como raiz al igualarla a cero $\widehat{\theta}=\left(1-\frac{1}{n}\Sigma\ln x_i\right)^{-1}$. Como la derivada segunda es

$$\frac{\partial^2}{\partial\theta^2}\ln L(\theta|x_1,...,x_n)=-\frac{n}{\theta^2}+\frac{n}{(1-\theta)^2}+\frac{2}{(1-\theta)^3}\sum\ln x_i$$

que en ese punto es negativa, será el estimador si satisface la restricción de estar en el intervalo $[1/2,1)$. Desde luego $\widehat{\theta}<1$ ya que $-\frac{1}{n}\sum\ln x_i>0$, pero en general no tiene por qué ser $\frac{1}{2}\leq\widehat{\theta}$, luego el EMV es

$$\widehat{\theta}=\text{máx}\left\{\frac{1}{2},\left(1-\frac{1}{n}\sum_{i=1}^{n}\ln X_i\right)^{-1}\right\}.$$

3.9.28

En lugar de conocer la muestra inicial, se informa de la suma hasta n, de variables aleatorias Y_i tales que

$$Y_i=\begin{cases}1 & \text{si } X_i\leq 0\\ 0 & \text{si } X_i>0\end{cases}$$

por lo tanto

$$\begin{aligned}P\{Y_i=1|\theta\}&=P\{X_i\leq 0|\theta\}=P\{X_i-\theta\leq-\theta|\theta\}\\&=P\{Z\leq-\theta\}=\Phi(-\theta)=1-\Phi(\theta)\end{aligned}$$

con Φ la función de distribución de la v.a. $N(0,1)$. Por la propiedad de invariancia del EMV si se nota por $\tau=1-\Phi(\theta)$ se puede estimar $\widehat{\tau}$ a partir de $Y_1,...,Y_n$, con lo que el EMV de θ será $\widehat{\theta}=\Phi^{-1}(1-\widehat{\tau})$. De acuerdo con el Ejemplo 3.6.2.1 es $\widehat{\tau}=\frac{m}{n}$, de donde

$$\widehat{\theta}=\Phi^{-1}(1-\frac{m}{n}).$$

Por ejemplo, con $n=10$ y $m=3$ sería $\widehat{\theta}=0'524$ aproximadamente.

3.9.29

La función de verosimilitud vale

$$\begin{aligned}L(\theta_1\theta_2|x_1y_1,...,x_ny_n)&=\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta_1)^2}\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n}(y_i-\theta_2)^2}\\&=\left(\frac{1}{\sqrt{2\pi}}\right)^{2n} e^{-\frac{1}{2}\left[\sum_{i=1}^{n}(x_i-\theta_1)^2+\sum_{i=1}^{n}(y_i-\theta_2)^2\right]}\end{aligned}$$

por lo tanto maximizarla en θ_1, θ_2 es equivalente a minimizar el exponente

$$\sum_{i=1}^{n}(x_i-\theta_1)^2+\sum_{i=1}^{n}(y_i-\theta_2)^2 = \sum_{i=1}^{n}(x_i-\overline{x})^2+n(\overline{x}-\theta_1)^2+\sum_{i=1}^{n}(y_i-\overline{y})^2+n(\overline{y}-\theta_2)^2$$

lo que se logra haciendo $\widehat{\theta}_1 = \overline{x}$, $\widehat{\theta}_2 = \overline{y}$. Por lo tanto si $\overline{x} \geq \overline{y}$ ya se ha obtenido el EMV, caso contrario el máximo se encontrará en la frontera de la región $\theta_1 \geq \theta_2$, es decir habrá que buscar el mínimo de

$$n(\overline{x}-\theta)^2 + n(\overline{y}-\theta)^2$$

que se obtiene, si se deriva y se iguala a cero, en $\widehat{\theta} = \frac{\overline{x}+\overline{y}}{2}$, la segunda derivada vale uno luego $\frac{\overline{x}+\overline{y}}{2}$ se corresponde con un mínimo. En resumen el EMV es

$$(\widehat{\theta}_1, \widehat{\theta}_2) = \begin{cases} (\overline{x}, \overline{y}) & \text{si } \overline{x} \geq \overline{y} \\ \left(\frac{\overline{x}+\overline{y}}{2}, \frac{\overline{x}+\overline{y}}{2}\right) & \text{si } \overline{x} < \overline{y} \end{cases}$$

3.9.30

La función de verosimilitud es

$$L(\theta|x_1, ..., x_n) = \left(\frac{1}{2\pi\theta}\right)^{n/2} e^{-\frac{1}{2\theta}\sum_{i=1}^{n}(x_i-\theta)^2}$$

de donde

$$\ln L(\theta|x_1, ..., x_n) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\theta - \frac{1}{2\theta}\sum_{i=1}^{n}(x_i-\theta)^2.$$

Por lo tanto la ecuación de verosimilitud es

$$\begin{aligned}\frac{\partial}{\partial\theta}\ln L(\theta|x_1, ..., x_n) &= -\frac{n}{2}\frac{1}{\theta} + \frac{1}{2\theta^2}\sum_{i=1}^{n}(x_i-\theta)^2 + \frac{1}{\theta}\sum_{i=1}^{n}(x_i-\theta) \\ &= -\frac{n}{2} - \frac{n}{2\theta} + \frac{1}{2\theta^2}\sum_{i=1}^{n}x_i^2\end{aligned}$$

que al igualarla a cero y dividirla por n queda

$$\theta^2 + \theta - a_2 = 0.$$

Las dos raíces son $-\frac{1}{2} + \frac{\sqrt{1-4a_2}}{2}$ y $-\frac{1}{2} - \frac{\sqrt{1-4a_2}}{2}$ de donde por ser negativa la segunda, la sucesión consistente de raíces que es el EMV es $\widehat{\theta}_n = -\frac{1}{2} + \frac{\sqrt{1-4a_2}}{2}$.

Para encontrar su distribución asintótica, por el Teorema 3.6.3.2 se observa que se cumple H1) ya que en cualquier abierto de $\theta > 0$ existe

$$\frac{\partial^3}{\partial\theta^3}\ln f(x|\theta) = 3\frac{(x-\theta)^2}{\theta^4} + \frac{1}{\theta^3}(4x - 3\theta - \frac{1}{2}) + \frac{1}{\theta^2}$$

además

$$E_\theta\left[\frac{\partial^3}{\partial\theta^3}\ln f(X|\theta)\right]=\frac{3}{\theta^3}+\frac{1}{\theta^3}(\theta-\frac{1}{2})+\frac{1}{\theta^2}$$

función que está acotada al ser un polinomio en $\frac{1}{\theta}$ en cualquier intervalo que no contenga al cero.

También se verifica H2) ya que para cualquier θ

$$\begin{aligned}E_\theta\left[\frac{\partial}{\partial\theta}\ln f(X|\theta)\right]&=-\frac{1}{2\theta}+\frac{1}{2\theta^2}E_\theta[(X-\theta)^2]+\frac{1}{\theta}E_\theta[X-\theta]\\&=0\end{aligned}$$

es $E_\theta\left[\frac{1}{f(X|\theta)}\frac{\partial^2}{\partial\theta^2}f(X|\theta)\right]=0$ con $I_1(\theta)$ finito ya que

$$\frac{\partial^2}{\partial\theta^2}\ln f(X|\theta)=-\frac{1}{\theta^2}(x-\frac{1}{2})-\frac{1}{\theta^3}(x-\theta)^2-\frac{1}{\theta^2}(x-\theta)$$

luego

$$I_1(\theta)=-E_\theta\left[\frac{\partial^2}{\partial\theta^2}\ln f(X|\theta)\right]=\frac{2\theta+1}{2\theta^2}$$

de donde de acuerdo con el Teorema 3.6.3.2 se tiene la convergencia en ley, $\sqrt{n}(\widehat{\theta}_n-\theta)\xrightarrow{\mathcal{L}}Z\sim N(0,I_1^{-1/2}(\theta))$ es decir $Z\sim N(0,\frac{1}{\theta}\sqrt{\theta+\frac{1}{2}})$.

3.9.31

Basta ver que en virtud del Teorema 3.6.3.2 se tiene la convergencia en ley $\sqrt{n}(\widehat{\theta}_n-\theta_0)\xrightarrow{\mathcal{L}}N(0,\frac{1}{\sqrt{I_1(\theta_0)}})$ por lo que de acuerdo con el Lema 1.3.1 con $g(\theta)=h(\theta)$ y $a_n=\sqrt{n}$ se sigue la convergencia en ley $\sqrt{n}(h(\widehat{\theta}_n)-h(\theta_0))\xrightarrow{\mathcal{L}}h'(\theta_0)Z$ donde $Z\sim N(0,\frac{1}{\sqrt{I_1(\theta_0)}})$.

3.9.32

Tras comprobar las condiciones de regularidad del Teorema 3.6.5.1, la distribución final es $N(\widehat{\theta}_n,(nI_1(\theta_n))^{-1/2})$, luego como $\ln p(x|\theta)=x\ln\theta+(1-x)\ln(1-\theta)$ se sigue que $\frac{\partial}{\partial\theta}\ln p(x|\theta)=\frac{x}{\theta}-\frac{1-x}{1-\theta}$ y $\frac{\partial^2}{\partial\theta^2}\ln p(x|\theta)=-\frac{x}{\theta^2}-\frac{1-x}{(1-\theta)^2}$ de donde

$$\begin{aligned}I_1(\theta)&=E_\theta\left[\frac{X}{\theta^2}+\frac{1-X}{(1-\theta)^2}\right]\\&=\frac{1}{\theta(1-\theta)}\end{aligned}$$

y por tanto $nI_1(\widehat{\theta}_n)=\frac{n}{\overline{X}(1-\overline{X})}$ ya que el EMV de θ es $\overline{X}$. De donde la distribución final!asintótica es $N\left(\overline{X},\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}\right)$, que particularizada en

$n = 50$ y $\sum_{i=1}^{50} x_i = 12$ se convierte en $N(0'24, 0'06)$. Si la distribución inicial es $\pi(\theta) \sim Beta(1,1)$, como la distribución muestral es $p(x_1, ..., x_n|\theta) = \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}$ la distribución final es $\pi\left(\theta|\sum_{i=1}^{n} x_i\right)$ una distribución $Beta\left(\sum_{i=1}^{n} x_i + 1, n - \sum_{i=1}^{n} x_i\right)$ que se reduce a una $Beta(12, 38)$. Mediante el concurso del ordenador puede comprobarse la perfecta aproximación de la distribución Normal a esta Beta.

Capítulo 4

ESTIMACIÓN POR REGIONES DE CONFIANZA

4.1. INTRODUCCIÓN

Hasta ahora se ha estudiado cómo eliminar la incertidumbre en la determinación de un parámetro mediante la utilización de un estimador por punto. Este procedimiento parece demasiado drástico y es más natural tratar de determinar una región, función de la muestra, dentro de la cual se pueda asegurar, de alguna forma que esté contenido el parámetro. Como la incertidumbre se modeliza por la probabilidad, se busca una región que tenga probabilidad alta de contener al parámetro y para facilitar su manejo e interpretación se tratará de conseguir regiones que sean un intervalo, si el parámetro es unidimensional, o un rectángulo, en más dimensiones.

Para introducir la estimación por *regiones de confianza* se empieza por el siguiente ejemplo introductorio. Se quiere determinar la distancia θ a la que un lanzador puede arrojar su jabalina, para lo cual se dispone de una sucesión particular de lanzamientos $x_1, ..., x_n$. Parece razonable suponer que los datos constituyen una m.a.s. de tamaño n de una población X con distribución $N(\theta, \sigma)$, de esta forma cada x_i distará de θ una cantidad aleatoria con distribución $N(0, \sigma)$. Para simplificar la exposición se supondrá que la desviación típica σ es conocida. Dado que $\overline{X} \sim N(\theta, \frac{\sigma}{\sqrt{n}})$, se pueden determinar dos valores de las tablas de la distribución $N(0, 1)$, tales que

$$P\{-1'96 \leq Z \leq 1'96\} = 0'95$$

donde $Z = \frac{\overline{X}-\theta}{\sigma}\sqrt{n}$, o alternativamente se cumple que

$$P\{-1'96 \leq \frac{\overline{X} - \theta}{\sigma}\sqrt{n} \leq 1'96|\theta\} = 0'95$$

independientemente del valor que tenga la media θ de la población.

Gráficamente, véase la Figura 4.1.1, el conjunto de puntos $-1'96 \leq Z \leq 1'96$ es equivalente a

$$\{\overline{X} - \frac{\sigma}{\sqrt{n}}1'96 < \theta < \overline{X} + \frac{\sigma}{\sqrt{n}}1'96\} = \{\underline{\theta}(\overline{X}) < \theta < \overline{\theta}(\overline{X})\}$$

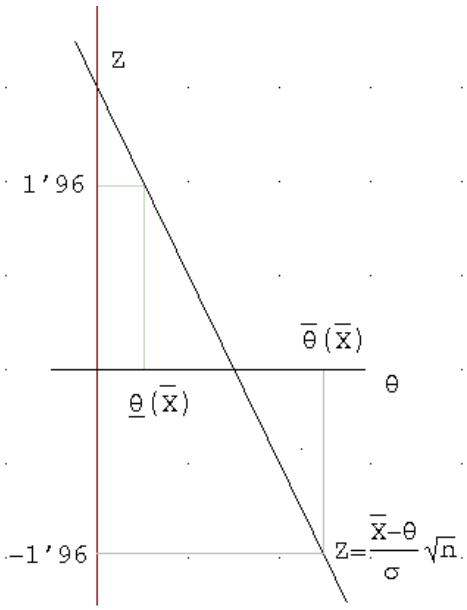


Figura 4.1.1. Esquema de obtención de la región de confianza para θ en el caso $N(\theta, \sigma)$ con σ conocida.

Es decir que para cualquier valor θ el intervalo aleatorio

$$\left(\overline{X} - \frac{\sigma}{\sqrt{n}}1'96, \overline{X} + \frac{\sigma}{\sqrt{n}}1'96\right)$$

tendrá una probabilidad de contener a θ igual a $0'95$. Entonces se toma como estimador de confianza dicho intervalo basado en esta propiedad.

El problema se presenta al tratar de interpretar, una vez tomada una muestra particular, el intervalo $\left(\overline{X} - \frac{\sigma}{\sqrt{n}}1'96, \overline{X} + \frac{\sigma}{\sqrt{n}}1'96\right)$ diciendo que tiene una probabilidad $0'95$ de contener a θ. Si se supone que los lanzamientos, seis veces consecutivas, son $89'90, 90'02, 89'92, 91'10, 88'71$ y $90'51$ con lo que $\overline{X} = 90'03$, y que el valor de σ es $0'8$, el intervalo de confianza para θ es $\left(90'03 - \frac{0'8}{\sqrt{6}}, 90'03 + \frac{0'8}{\sqrt{6}}1'96\right) = (89'39, 90'67)$.

Es claro que una vez tomada la muestra particular, dicho intervalo contendrá o no, puesto que es constante, al valor de θ correspondiente a la media

de la población. Interpretando la probabilidad como límite de frecuencias, se suele decir que si en N repeticiones que se hicieran tomando grupos de 6 observaciones con distribución $N(\theta, \sigma)$, si se determina cada vez el intervalo particular para lo resultados obtenidos, el número de veces n que el intervalo cubriría el valor θ, estaría en la proporción $\frac{n}{N}$ y este valor se aproximaría a $0'95$.

Con base en estas ideas se introduce la siguiente definición.

Definición 4.1.1 (región de confianza)
Dada una familia de distribuciones de probabilidad $\mathcal{P} = \{f(\overrightarrow{x}|\theta)|\theta \in \Theta\}$ se llama región de confianza de grado $1-\alpha$ basada en una m.a.s. de tamaño n, a una región $S(X_1, ..., X_n)$ del espacio paramétrico Θ tal que $P\{S(X_1, ..., X_n) \ni \theta|\theta\} \geq 1 - \alpha \quad \forall\theta \in \Theta$ con $\alpha \in (0, 1)$ y normalmente pequeño. Si Θ es un subconjunto de $\mathbb{R}$ y $S(X_1, ..., X_n) = (\underline{\theta}(X_1, ..., X_n), \overline{\theta}(X_1, ..., X_n))$, se habla entonces de un intervalo de confianza. Pudiendo ser uno de los extremos infinito.

■

En esta definición se admite tácitamente que cuando la distribución de $S(\overrightarrow{X})$ sea continua se puede alcanzar el valor $1-\alpha$ y que en caso contrario la probabilidad del primer miembro es el valor más próximo a $1-\alpha$ por exceso que sea posible, con el fin de no aumentar innecesariamente la probabilidad y por tanto el tamaño de la región de confianza.

Ejemplo 4.1.1
Si X es $N(\theta, \sigma_0)$ con σ_0 conocida $\left(\overline{X} - \frac{\sigma_0}{\sqrt{n}} z_{\frac{\alpha}{2}}, \overline{X} + \frac{\sigma_0}{\sqrt{n}} z_{\frac{\alpha}{2}}\right)$ es un intervalo de grado de confianza $1-\alpha$ para la media θ y con $z_{\frac{\alpha}{2}}$ tal que $\Phi(z_{\frac{\alpha}{2}}) = 1-\frac{\alpha}{2}$, siendo Φ la función de distribución de la v.a. $N(0,1)$.

■

Siempre es deseable hacer que la medida de la región de confianza sea mínima, entre todas las del mismo grado de confianza $1 - \alpha$. Esto lleva, en el Ejemplo 4.1.1, a tomar unos valores de $z_{\frac{\alpha}{2}}$ y de $-z_{\frac{\alpha}{2}}$ opuestos, es fácil comprobar que cualesquiera otros dan una amplitud mayor, debido a la simetría de la distribución normal.

Frecuentemente puede conseguirse que $P\{S(X_1, ..., X_n) \ni \theta|\theta\} = 1 - \alpha$ $\forall\theta \in \Theta$, pero si la distribución de $S(X_1, ..., X_n)$ es discreta esto puede no ser posible, por lo que hay que contentarse, con el mayor estricto.

4.2. MÉTODOS DE OBTENCIÓN DE INTERVALOS DE CONFIANZA

Como se ha visto en el Ejemplo 4.1.1, el método de construcción del intervalo de confianza se ha basado en que, para cada θ, $\frac{\overline{X}-\theta}{\sigma_0}\sqrt{n}$ es una

varaible aleatoria cuya distribución no depende de θ, el parámetro desconocido, y que en su expresión todo es conocido, salvo el parámetro θ. Esta variable aleatoria, ha permitido trasladar un enunciado de probabilidad sobre la v.a. $T = \frac{\overline{X}-\theta}{\sigma_0}\sqrt{n}$, a un enunciado en términos de confianza de que θ se encuentre entre dos extremos $\left(\overline{X} - \frac{\sigma_o}{\sqrt{n}} z_{\frac{\alpha}{2}}, \overline{X} + \frac{\sigma_0}{\sqrt{n}} z_{\frac{\alpha}{2}}\right)$. Esta acción de pivotar la probabilidad admite una formulación bastante general y es la que ha dado el nombre a la variable $T(X_1, ..., X_n; \theta)$ de *variable pivotal.*

Teorema 4.2.1 (variable pivotal)
Sea $(X_1, ..., X_n)$ una m.a.s. correspondiente a una distribución $\mathcal{P} = \{f(\overrightarrow{x}|\theta)|\theta \in \Theta\}$ con Θ un intervalo de $\mathbb{R}$, si existe $T(X_1, ..., X_n : \theta)$ tal que:

1) Para todo θ $T(\cdot : \theta)$ es medible, está perfectamente determinada salvo θ y con distribución independiente de θ.

2) Para todo $x_1, ..., x_n$ $T(x_1, ..., x_n : \cdot)$ es una función de θ continua y estrictamente monótona.

Entonces se puede construir un intervalo de grado de confianza $1 - \alpha$ para θ con $\alpha \in [0, 1]$.

Demostración: Fijado $1 - \alpha$ se determinan los números $\gamma_1(\alpha)$ y $\gamma_2(\alpha)$ tales que

$$P\{\gamma_1(\alpha) < T(X_1, ..., X_n; \theta) < \gamma_2(\alpha)|\theta\} \geq 1 - \alpha \qquad \forall\theta \in \Theta$$

al no depender la distribución de T del valor de θ, γ_1 y γ_2 solo son función de α. La resolución en θ de las ecuaciones

$$\left.\begin{array}{l} T(x_1, ..., x_n; \theta) = \gamma_1(\alpha) \\ T(x_1, ..., x_n; \theta) = \gamma_2(\alpha) \end{array}\right\} \tag{4.2.1}$$

permite obtener $\theta_1(x_1, ..., x_n; \alpha)$ y $\theta_2(x_1, ..., x_n; \alpha)$, donde una posible configuración puede verse en la Figura 4.2.1.

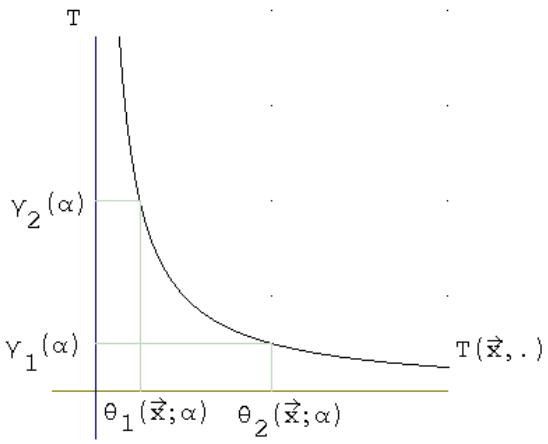


Figura 4.2.1. Esquema del método de la variable pivotal.

Por lo tanto

$$P\{\theta_1(X_1,...,X_n;\alpha) < \theta < \theta_2(X_1,...,X_n;\alpha)|\theta\} \geq 1-\alpha \qquad \forall\theta \in \Theta$$

con lo que un intervalo de confianza de grado $1-\alpha$ vendrá dado por

$$S(X_1,...,X_n) = (\theta_1(X_1,...,X_n;\alpha), \theta_2(X_1,...,X_n;\alpha)).$$

■

Obviamente, a distintas variables pivotales les corresponden intervalos diferentes.

En situaciones bastante generales existirá al menos una variable pivotal cumpliendo las hipótesis del enunciado; en concreto $T = -2\sum_{i=1}^{n} \ln F_\theta(X_i)$ lo es si como función de θ, la función de distribución es continua y estrictamente monótona, ya que en ese caso T se distribuye como una χ^2_{2n}, como es fácil de comprobar, véase el Ejercicio 4.7.3.

Ejemplo 4.2.1
Si la función de densidad es $f(x|\theta) = \theta x^{\theta-1} I_{(0,1)}(x)$ con $\theta \geq 0$ entonces

$$F(x|\theta) = \begin{cases} 0 & \text{si } x \leq 0 \\ x^\theta & \text{si } x \in (0,1) \\ 1 & \text{si } x \geq 1 \end{cases}$$

y si para simplificar los cálculos, se toman dos valores de las tablas de la distribución χ^2_{2n} tales que

$$F(\chi^2_{2n;1-\frac{\alpha}{2}}) = \frac{\alpha}{2} \qquad F(\chi^2_{2n;\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$$

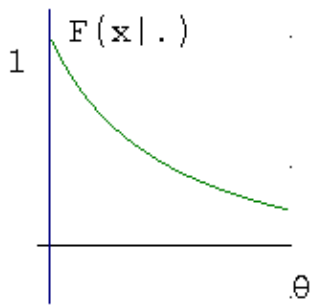


Figura 4.2.2. $F(x|\theta)$ como función de θ para el Ejemplo 4.2.1.

se sigue que

$$P\{\chi^2_{2n;1-\frac{\alpha}{2}} < -2\theta \sum_{i=1}^{n} \ln X_i < \chi^2_{2n;\frac{\alpha}{2}} | \theta\} = 1-\alpha \quad \forall \theta \in \Theta$$

y por tanto

$$P\left\{ \frac{\chi^2_{2n;1-\frac{\alpha}{2}}}{-2\sum_{i=1}^{n} \ln X_i} < \theta < \frac{\chi^2_{2n;\frac{\alpha}{2}}}{-2\sum_{i=1}^{n} \ln X_i} \middle| \theta \right\} = 1-\alpha \quad \forall \theta \in \Theta$$

con lo que se tiene el intervalo deseado.

■

El método puede generalizarse al caso en el que $T(X_1, ..., X_n, \theta)$ sea un estadístico $T(X_1, ..., X_n)$ si se introducen restricciones adicionales; mediante el siguiente enunciado.

Teorema 4.2.2 (método de Neyman)
Sea $(X_1, ..., X_n)$ una m.a.s. correspondiente a una distribución de $\mathcal{P} = \{f(\overrightarrow{x}|\theta)\theta \in \Theta\}$ con Θ un intervalo de $\mathbb{R}$. Si para cualquier $\alpha \in [0,1]$ se pueden encontrar dos funciones $\gamma_1(\theta, \alpha)$ y $\gamma_2(\theta, \alpha)$ y un estadístico $T(X_1, ..., X_n)$ verificando que para todo $\theta \in \Theta$

$$P\{\gamma_1(\theta, \alpha) \leq T(X_1, ..., X_n) \leq \gamma_2(\theta, \alpha)|\theta\} \geq 1-\alpha,$$

si además las funciones γ_1 y γ_2 son estrictamente monótonas en θ, del mismo sentido y las ecuaciones

$$\gamma_1(\theta, \alpha) = T(X_1, ..., X_n)$$
$$\gamma_2(\theta, \alpha) = T(X_1, ..., X_n)$$

se pueden invertir para resolverlas en θ como función de $T = t$, entonces se puede construir un intervalo para θ de grado de confianza $1-\alpha$.

Demostración: Supóngase que $\gamma_1(\theta, \alpha)$ y $\gamma_2(\theta, \alpha)$ son estrictamente crecientes, como en la Figura 4.2.3. Para cada θ_0 fijo el suceso $\gamma_1(\theta_0; \alpha) \leq T \leq \gamma_2(\theta_0, \alpha)$ moviéndose T, es el mismo, que el suceso $\underline{\theta}(T) \leq \theta_0 \leq \overline{\theta}(T)$ moviéndose T, donde son $\underline{\theta}(T)$ y $\overline{\theta}(T)$ soluciones de

$$\left.\begin{array}{l} \gamma_2(\underline{\theta}(T); \alpha) = T \\ \gamma_1(\overline{\theta}(T); \alpha) = T \end{array}\right\} \tag{4.2.2}$$

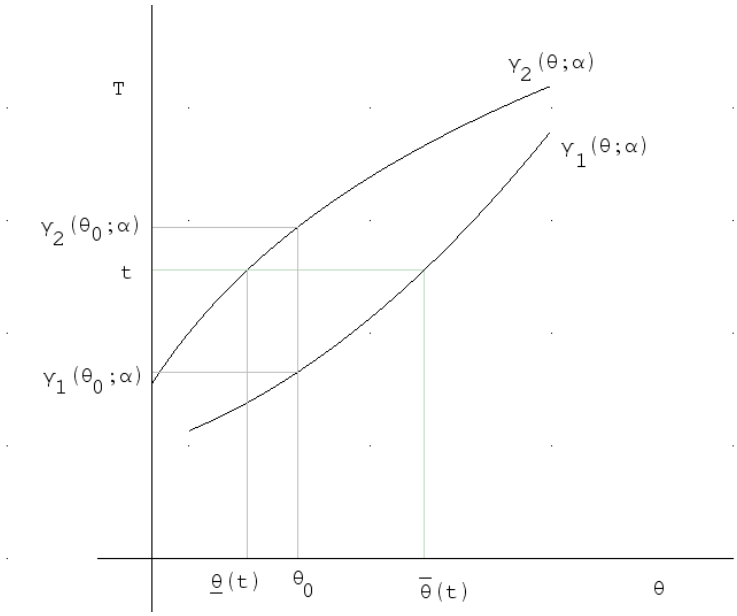


Figura 4.2.3. Representación del Teorema 4.2.2 (Método de Neyman).

Por tanto se tiene que

$$P\{\underline{\theta}(T) < \theta < \overline{\theta}(T)|\theta\} \geq 1-\alpha \qquad \forall\theta \in \Theta$$

o lo que es lo mismo $(\underline{\theta}(T), \overline{\theta}(T))$ es un intervalo de grado de confianza $1-\alpha$.

■

Igual que en el caso de la variable pivotal, es claro que este procedimiento depende del estadístico T empleado.

Las suposiciones de monotonía estricta en los dos teoremas sirven para la determinación de la región de confianza en forma de intervalo, por último la hipótesis de continuidad en θ de las funciones $\gamma_i(\theta;\alpha)$, $i = 1, 2$ aseguran la existencia de soluciones del Sistema 4.2.2.

Ejemplo 4.2.2

Dada una m.a.s. de tamaño n de una población $Uniforme(0, \theta)$, para determinar un intervalo de grado de confianza $1-\alpha$ para θ, se puede emplear el estadístico suficiente $T = X_{(n)}$, cuya función de distribución es

$$F(t|\theta) = \begin{cases} 0 & \text{si } t \leq 0 \\ (\frac{t}{\theta})^n & \text{si } 0 \leq t \leq \theta \\ 1 & \text{si } t \geq \theta \end{cases}$$

siguiendo el *método de Neyman*, se deben encontrar dos funciones $\gamma_1(\theta;\alpha)$ y $\gamma_2(\theta,\alpha)$ de manera que

$$P\{\gamma_1(\theta;\alpha) \leq T \leq \gamma_2(\theta;\alpha)|\theta\} \geq 1-\alpha, \qquad \alpha \in [0,1].$$

En este caso se puede conseguir la igualdad. Bastará hacer

$$F(\gamma_1(\theta,\alpha)|\theta) = \alpha_1$$
$$1 - F(\gamma_1(\theta,\alpha)|\theta) = \alpha - \alpha_1$$

con $\alpha_1 \in [0,\alpha]$, de donde

$$\gamma_1(\theta;\alpha) = \theta\alpha_1^{\frac{1}{n}}$$
$$\gamma_2(\theta;\alpha) = \theta(1-(\alpha-\alpha_1))^{1/n}.$$

Gráficamente, véase la Figura 4.2.4, las funciones γ_i son rectas que pasan por el origen; si se igualan

$$\left.\begin{array}{l}\gamma_1(\theta;\alpha) = T(X_1,...,X_n)\\ \gamma_2(\theta\,;a) = T(X_1,...,X_n)\end{array}\right\}$$

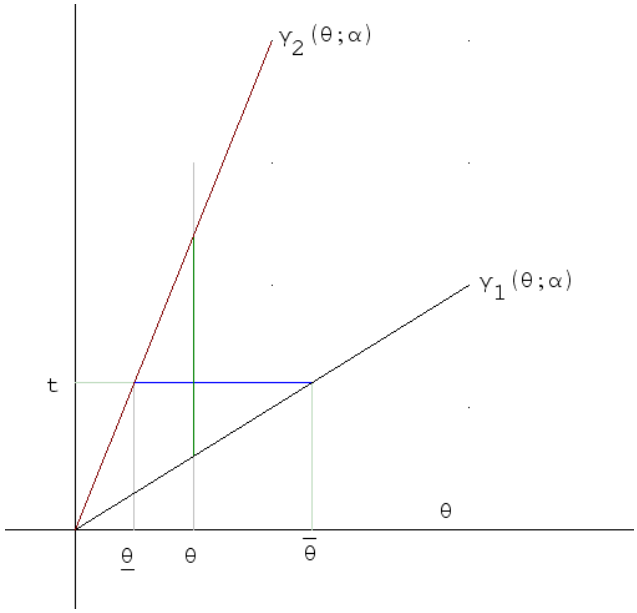


Figura 4.2.4. Representación del intervalo de confianza para la población $U(0,$

queda

$$\underline{\theta}(T(X_1,...,X_n)) = X_{(n)}(1-(\alpha-\alpha_1))^{-1/n}$$
$$\overline{\theta}(T(X_1,...,X_n)) = X_{(n)}\alpha_1^{-1/n}$$

En definitiva $(X_{(n)}(1-(\alpha-\alpha_1))^{-1/n}, X_{(n)}\alpha_1^{-1/n})$ es un intervalo para θ de grado de confianza $1-\alpha$.

Es fácil comprobar que al variar $\alpha_1 \in [0,\alpha]$ el intervalo de amplitud mínima viene dado por $\alpha_1 = \alpha$, y por lo tanto, es

$$IC_{1-\alpha}(\theta) = (X_{(n)},\ X_{(n)}\alpha^{-1/n}).$$

4.3. INTERVALOS DE CONFIANZA ASOCIADOS A LA DISTRIBUCIÓN NORMAL

En los casos siguientes se supone una población $X \sim N(\mu, \sigma)$ y que se trabaja con m.a.s. de tamaño n.

Intervalo para la media con varianza conocida:

Si se toma $\theta = \mu$ se puede emplear que $\frac{\overline{X}-\theta}{\sigma}\sqrt{n}$ tiene distribución $N(0,1)$ y al ser σ conocida se puede utilizar como variable pivotal

$$T(X_1, ..., X_n; \theta) = \frac{\overline{X} - \theta}{\sigma}\sqrt{n} \tag{4.3.1}$$

determinando $z_{\frac{\alpha}{2}}$ para que $P\left\{\frac{|\overline{X}-\theta|}{\sigma}\sqrt{n} \leq z_{\frac{\alpha}{2}}\right\} = 1-\alpha$ se obtiene que un intervalo de confianza de grado $1-\alpha$ para la media de la población es

$$IC_{1-\alpha}(\theta) = \left(\overline{X} - \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}}, \overline{X} + \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}}\right).$$

Además al ser la distribución de la normal simétrica; éste es el intervalo de amplitud mínima entre todos los de grado de confianza $1-\alpha$.

Intervalo para la media con varianza desconocida:

En este caso la variable T de la Expresión (4.3.1) no puede ser empleada como variable pivotal, por lo que se recurre a

$$T(X_1, ..., X_n; \theta) = \frac{\overline{X} - \theta}{S}\sqrt{n}, \tag{4.3.2}$$

con $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ la cuasivarianza muestral, ya que dicha expresión se distribuye según una T de Student con $n-1$ grados de libertad. Esto conduce a que un intervalo de grado de confianza $1-\alpha$ para la media es

$$IC_{1-\alpha}(\theta) = \left(\overline{X} - t_{n-1;\frac{\alpha}{2}}\frac{S}{\sqrt{n}}, \overline{X} + t_{n-1;\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right)$$

donde por el mismo motivo que antes, el intervalo es de amplitud mínima entre los de probabilidad $1-\alpha$ basados en la variable pivotal (4.3.2).

Intervalo para la varianza cuando la media es desconocida:

Si el parámetro es $\theta = \sigma^2$ se puede utilizar como variable pivotal

$$T(X_1, ..., X_n; \theta) = (n-1)\frac{S^2}{\theta} \tag{4.3.3}$$

que se distribuye según una χ^2_{n-1}, con lo que se obtiene que un intervalo de grado $1-\alpha$ para la varianza es

$$IC_{1-\alpha}(\theta) = \left(\frac{(n-1)S^2}{\chi^2_{n-1;\frac{\alpha}{2}}}, \frac{(n-1)S^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}} \right)$$

donde

$$P\{\chi^2_{n-1} \leq \chi^2_{n-1;\frac{\alpha}{2}}\} = 1 - \frac{\alpha}{2}$$
$$P\{\chi^2_{n-1} \leq \chi^2_{n-1;1-\frac{\alpha}{2}}\} = \frac{\alpha}{2}$$

que no son los valores correspondientes a la amplitud mínima, sino valores que dejan colas iguales a $\frac{\alpha}{2}$ bajo la distribución χ^2_{n-1} y que son más fáciles de calcular. Cuando μ es conocida es más correcto utilizar como variable pivotal

$$T(X_1, ..., X_n; \theta) = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\theta} \right)^2 \tag{4.3.4}$$

que se distribuye como χ^2_n.

Intervalo para la diferencia de medias:

Si se dispone de dos m.a.s. de tamaños n_1 y n_2 de dos poblaciones, X con distribución $N(\mu_1, \sigma_1)$ e Y con distribución $N(\mu_2, \sigma_2)$, respectivamente. Se puede emplear, si se conocen las dos varianzas y es $\theta = \mu_1 - \mu_2$

$$T(X_1, ..., X_{n_1}, Y_1, ..., Y_{n_2}; \theta) = \frac{\overline{X} - \overline{Y} - \theta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \tag{4.3.5}$$

como variable pivotal, ya que se distribuye $N(0,1)$ y es conocida salvo θ.

Esto lleva a que un intervalo de grado de confianza $1-\alpha$ para la diferencia de medias es

$$IC_{1-\alpha}(\theta) = \left(\overline{X} - \overline{Y} - z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},\ \overline{X} - \overline{Y} + z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$$

intervalo de amplitud mínima basado en (4.3.5), donde

$$\Phi(z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2},$$

con Φ la función de distribución de la $N(0,1)$.

Si las dos varianzas son desconocidas pero iguales, $\sigma_1^2 = \sigma_2^2 = \sigma^2$ entonces se utiliza como variable pivotal

$$T(X_1, ..., X_{n_1}, Y_1, ..., Y_{n_2}; \theta) = \frac{\overline{X} - \overline{Y} - \theta}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{4.3.6}$$

que se distribuye según una distribución de Student $T_{n_1+n_2-1}$, donde

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

es una ponderación de las cuasivarianzas de cada muestra y que había sido llamada *cuasivarianza muestral ponderada*, véase la expresión (1.5.3.5). Esto lleva al intervalo de confianza

$$IC_{1-\alpha}(\theta) = \left(\overline{X} - \overline{Y} - t_{n_2+n_2-1;\frac{\alpha}{2}} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},\ \ \overline{X} - \overline{Y} + t_{n_2+n_2-1;\frac{\alpha}{2}} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right)$$

que es de amplitud mínima y donde $t_{n_1+n_2-1;\frac{\alpha}{2}}$ es el cuantil de orden $1 - \frac{\alpha}{2}$ de la distribución de Student con $n_1 + n_2 - 2$ grados de libertad.

En el caso en que σ_1^2 y σ_2^2 sean desconocidas y distintas la distribución de

$$\frac{\overline{X} - \overline{Y} - \theta}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \tag{4.3.7}$$

depende de $\frac{\sigma_1^2}{\sigma_2^2}$. A dicha distribución se la conoce con el nombre de distribución de Behrens-Fisher, por lo que (4.3.7) no puede utilizarse como variable pivotal.

Existen distintas soluciones al problema, no totalmente satisfactorias, aquí se utilizará la debida a Hsu que consiste en aproximar la distribución, por una T con $f = \text{mín}\{n_1, n_2\} - 1$ como número de grados de libertad, con lo que se obtiene el intervalo de confianza aproximado

$$IC_{1-\alpha}(\theta) \simeq \left(\overline{X} - \overline{Y} - t_{f;\frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}},\ \overline{X} - \overline{Y} + t_{f;\frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right)$$

soluciones alternativas pueden verse en Dudewicz y Mishra (1988) pág. 567.

Intervalo de confianza para el cociente de varianzas:

Si el parámetro es ahora $\theta = \frac{\sigma_1^2}{\sigma_2^2}$ se puede emplear como variable pivotal

$$T(X_1, ..., X_{n_1}, Y_1, ..., Y_{n_2}; \theta) = \frac{S_1^2}{S_2^2} \frac{1}{\theta} \tag{4.3.8}$$

que se distribuye como una distribución de Snedecor F_{n_1-1,n_2-1}; sin más que utilizar que $(n_i-1)\frac{S_i^2}{\sigma_i^2}$ se distribuyen como $\chi^2_{n_i-1}$ $i=1,2$. Esto lleva al intervalo de grado de confianza $1-\alpha$ para el cociente de varianzas

$$IC_{1-\alpha}(\theta)=\left(\frac{S_1^2}{S_2^2}f^{-1}_{n_1-1,n_2-1;\frac{\alpha}{2}}\ ,\ \frac{S_1^2}{S_2^2}f_{n_2-1,n_1-1;\frac{\alpha}{2}}\right)$$

donde $f_{n_1-1,n_2-1;\frac{\alpha}{2}}$ y $f_{n_2-1,n_1-1;\frac{\alpha}{2}}$ son los cuantiles de orden $1-\frac{\alpha}{2}$ de la distribución F correspondiente.

Este intervalo no tiene amplitud mínima, ya que con el fin de facilitar los cálculos se han tomado colas de probabilidad iguales en lugar de ordenadas iguales.

Región de confianza para la media y la varianza:

Si en el muestreo de una población $N(\mu,\sigma)$ se está interesado en encontrar una región de confianza para el parámetro vectorial $\overrightarrow{\theta}=(\mu,\sigma^2)$, se puede utilizar que $\frac{\overline{X}-\mu}{\sigma}\sqrt{n}$ y $(n-1)\frac{S^2}{\sigma^2}$ son v.a. independientes que se distribuyen $N(0,1)$ y χ^2_{n-1} respectivamente, véase el Teorema 1.4.2.1 (Fisher), con lo que si λ, c_1 y c_2 se determinan para que cumplan

$$P\{-\lambda<\frac{\overline{X}-\mu}{\sigma}\sqrt{n}<\lambda\}=\sqrt{1-\alpha}\qquad P\{c_1<(n-1)\frac{S^2}{\sigma^2}<c_2\}=\sqrt{1-\alpha},$$

lo que puede lograrse a partir de las tablas de la distribución $N(0,1)$ y de la χ^2_{n-1}, entonces se sigue, en virtud de la independencia, que

$$P\{-\lambda<\frac{\overline{X}-\mu}{\sigma}\sqrt{n}<\lambda, c_1<(n-1)\frac{S^2}{\sigma^2}<c_2\}=1-\alpha \tag{4.3.9}$$

La zona rayada de la Figura 4.3.1 determinada por la parábola $\sigma^2=\frac{n}{\lambda^2}(\overline{X}-\mu)^2$ y las rectas $\sigma^2=(n-1)\frac{S^2}{c_1}$ y $\sigma^2=(n-1)\frac{S^2}{c_2}$, constituye una región de grado de confianza $1-\alpha$. Nótese que en la Expresión (4.3.9), las variables aleatorias son $\overline{X}$ y S^2.

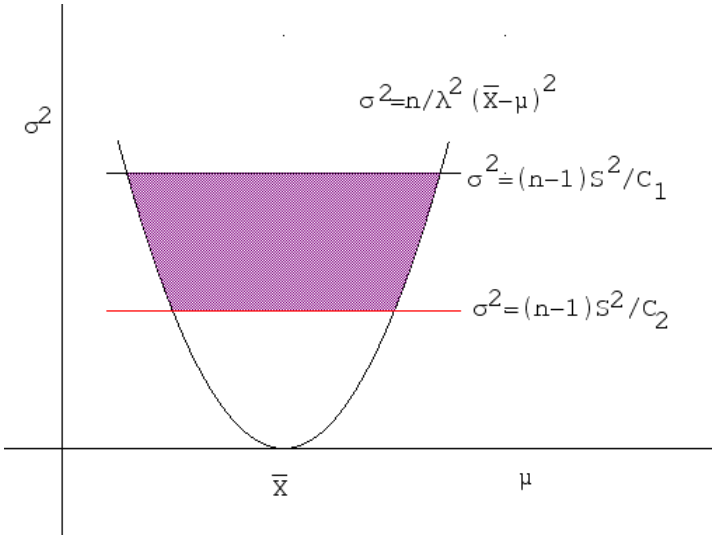


Figura 4.3.1. Región de confianza para el caso $N(\mu, \sigma)$.

Sin embargo esta región no es un rectángulo. No obstante, en esta situación se puede conseguir otra que si lo sea; por poderse aplicar un resultado más general que se recoge a continuación.

Proposición 4.3.1
Si dada $(X_1, ..., X_n)$ una m.a.s. de una distribución perteneciente a la familia $\mathcal{P} = \{f(\overrightarrow{x}|\,\overrightarrow{\theta})|\,\overrightarrow{\theta} = (\theta_1, ..., \theta_k) \in \mathbb{R}^k\}$ y para todo $\alpha \in [0, 1]$ se pueden determinar regiones de confianza

$$P\{S_i(X_1, ..., X_n) \ni \theta_i|\,\overrightarrow{\theta}\} \geq 1 - \frac{\alpha}{k} \qquad i = 1, ..., k \tag{4.3.10}$$

entonces $S(X_1, ..., X_n) = \prod_{i=1}^{k} S_i(X_1, ..., X_n)$ es una región de grado de confianza $1 - \alpha$.

Demostración: En primer lugar, obsérvese que si los conjuntos $S_i(\overrightarrow{X})$ son intervalos, entonces $S(\overrightarrow{X})$ es un rectángulo.

Además

$$P\{S(\overrightarrow{X}) \ni \overrightarrow{\theta}\,|\,\overrightarrow{\theta}\} = P\{\bigcap_{i=1}^{k}(S_i(\overrightarrow{X}) \ni \theta_i)|\,\overrightarrow{\theta}\}$$

ahora bien, pasando a los complementarios

$$P\{\bigcap_{i=1}^{k}(S_i(\overrightarrow{X}) \ni \theta_i)|\,\overrightarrow{\theta}\} = 1 - P\{\bigcup_{i=1}^{k}(S_i(\overrightarrow{X}) \not\ni \theta_i)|\,\overrightarrow{\theta}\}$$

y si se aplica que la probabilidad de la unión de sucesos es menor o igual que la suma de las probabilidades de los sucesos; se obtiene que

$$1 - \sum_{i=1}^{k} P\{S_i(\overrightarrow{X}) \not\ni \theta_i | \overrightarrow{\theta}\} \geq 1 - \sum_{i=1}^{k} - \frac{\alpha}{k}$$

de donde se sigue el resultado.

■

Es claro, que en el caso de una población normal, basta hacer $k = 2$ y utilizar que $\frac{\overline{X}-\mu}{S}\sqrt{n}$ tiene de distribución T_{n-1} y que $(n-1)\frac{S^2}{\sigma^2}$ tiene de distribución χ^2_{n-1}, de donde

$$S_1(\overrightarrow{X}) = \left(\overline{X} - t_{n-1;\frac{\alpha}{4}} \frac{S}{\sqrt{n}},\ \overline{X} + t_{n-1;\frac{\alpha}{4}} \frac{S}{\sqrt{n}}\right)$$

y

$$S_2(\overrightarrow{X}) = \left(\frac{(n-1)s^2}{\chi^2_{n-1;\frac{\alpha}{4}}},\ \frac{(n-1)s^2}{\chi^2_{n-1;1-\frac{\alpha}{4}}}\right)$$

permiten construir el rectángulo.

4.4. INTERVALOS DE CONFIANZA PARA MUESTRAS GRANDES

Se ha visto al hablar de los EMV, véase la Sección 3.6.3, que en ciertos casos se pueden encontrar sucesiones de estimadores T_n y constantes $\sigma_n(\theta)$ tales que para todo θ $\frac{T_n-\theta}{\sigma_n(\theta)} \xrightarrow{\mathcal{L}_\theta} N(0,1)$. Por lo tanto, si el tamaño muestral es suficientemente grande, se puede determinar $z_{\frac{\alpha}{2}}$ para que

$$P\{-z_{\frac{\alpha}{2}} < \frac{T_n - \theta}{\sigma_n(\theta)} < z_{\frac{\alpha}{2}} | \theta\} \simeq 1 - \alpha \tag{4.4.1}$$

con $z_{\frac{\alpha}{2}}$ tal que $\Phi(z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$ donde Φ es la función de distribución de la v.a. $N(0,1)$. Si la relación anterior puede invertirse, se podrá utilizar para determinar un intervalo de confianza para θ.

Nótese que si la ecuación de verosimilitud tiene una única raíz $\widehat{\theta}_n$ puede tomarse $T_n = \widehat{\theta}_n$ la sucesión consistente de raices y $\sigma_n(\theta) = \frac{1}{\sqrt{nI_1(\theta)}}$ que, sí es una función continua, como $\widehat{\theta}_n \xrightarrow{c.s.} \theta$, puede ser aproximada por $\sigma_n(\widehat{\theta}_n)$ lo que facilita la inversión de la Expresión (4.4.1).

Ejemplo 4.4.1
Para construir mediante muestras grandes, un intervalo de grado de confianza $1 - \alpha$ para el parámetro θ de una población con función de densidad $f(x|\theta) = \theta e^{-x\theta}$, $x > 0$, $\theta > 0$. Como la ecuación de verosimilitud es

$$\frac{n}{\theta} - \sum_{i=1}^{n} x_i = 0$$

que tiene como única solución $\widehat{\theta}_n = \frac{1}{\overline{X}_n}$. Haciendo cálculos se obtiene que $\sigma_n(\theta) = \frac{\theta}{\sqrt{n}}$, lo que permite escribir

$$P\left\{-z_{\frac{\alpha}{2}} < \frac{\widehat{\theta}_n - \theta}{\theta}\sqrt{n} < z_{\frac{\alpha}{2}}|\theta\right\} \simeq 1-\alpha$$

donde $\phi(z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$ con Φ función de distribución de la $N(0,1)$.

Expresión que puede invertirse para obtener el intervalo de grado de confianza $1-\alpha$ para θ:

$$IC_{1-\alpha}(\theta) \simeq \left(\frac{\overline{X}_n^{-1}}{1+z_{\frac{\alpha}{2}}/\sqrt{n}},\ \frac{\overline{X}_n^{-1}}{1-z_{\frac{\alpha}{2}}/\sqrt{n}}\right)$$

■

La desigualdad de Tchebycheff también puede emplearse en casos muy generales para construir intervalos de confianza. Claro está que los intervalos que se obtienen mediante este método, son más amplios, que los construidos mediante procedimientos específicos a cada modelo de probabilidad.

Ejemplo 4.4.2
Para construir mediante muestras grandes, un intervalo de confianza $1-\alpha$ para el parámetro θ de una población de $Poisson(\theta)$, con $\theta > 0$ la expresión

$$P\{|\overline{X} - y| < k\sqrt{\theta}|\theta\} \geq \frac{1}{k^2},$$

permite obtener

$$P\{|\overline{X} - \theta| < k\sqrt{\theta}|\theta\} \geq 1 - \frac{1}{k^2}.$$

Así si se hace $1 - \frac{1}{k^2} = 1-\alpha$ y se aproxima θ por $\overline{X}$ se obtiene el intervalo de confianza aproximado para θ

$$IC_{1-\alpha}(\theta) \simeq \left(\overline{X} - \sqrt{\frac{\overline{X}}{\alpha}},\ \overline{X} + \sqrt{\frac{\overline{X}}{\alpha}}\right).$$

4.5. REGIONES DE CONFIANZA BAYESIANAS

En el ejemplo que ha servido de introducción a este capítulo, se tenía que el intervalo $\left(\overline{X} - 1'96\frac{\sigma}{\sqrt{n}}, \overline{X} + 1'96\frac{\sigma}{\sqrt{n}}\right)$, cuyos extremos son aleatorios, cubría a la media θ de la distribución normal con probabilidad $0'95$. Sin embargo, tomada una muestra particular no puede hablarse de la probabilidad de que $(89'39, 90'67)$ contenga a θ, ya que θ es un parámetro y no una variable aleatoria.

Mediante la aproximación bayesiana, véase la Sección 3.6.4, si se dispone de información inicial sobre θ que se pueda concretar en una distribución inicial $\pi(\theta)$, será posible construir la distribución final para θ, $\pi(\theta|\overrightarrow{x})$ y con ella determinar una región que tenga probabilidad final alta de contener a θ. Para ello se introduce la siguiente definición.

Definición 4.5.1 (región creíble)
Dada una familia de distribuciones de probabilidad $\mathcal{P} = \{f(\overrightarrow{x}|\theta)|\theta \in \Theta\}$ si la información inicial sobre Θ viene dada mediante la función de densidad o de masa $\pi(\theta)$, la región $C(x_1, ..., x_n) \subset \Theta$ es una región creíble de probabilidad $1 - \alpha$ si

$$P\{\theta \in C(x_1, ..., x_n)|x_1, ..., x_n\} \geq 1 - \alpha \tag{4.5.1}$$

donde esta probabilidad se calcula mediante la distribución final, es decir

$$\int_{C(x_1,...,x_n)} \pi(\theta|x_1, ..., x_n)d\theta \geq 1 - \alpha.$$

■

Se ha empleado notación con función de densidad, y en ese caso se puede conseguir la igualdad en la Expresión 4.5.1. Lo mismo es válido con función de masa y entonces, posiblemente, la desigualdad en la Expresión 4.5.1 será estricta.

Ejemplo 4.5.1
Para X una población de $Bernoulli(\theta)$, si la información inicial acerca de θ viene dada por $\pi(\theta) = I_{(0,1)}(\theta)$, con m.a.s. de tamaño n la distribución final es

$$\pi(\theta|x_1, ..., x_n) = \frac{\theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i} I_{(0,1)}(\theta)}{\int_0^1 \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i} d\theta} = \frac{\theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i} I_{(0,1)}(\theta)}{B\left(\sum_{i=1}^n x_i + 1, n - \sum_{i=1}^n x_i + 1\right)}$$

una distribución $Beta$ de parámetros $p = \sum_{i=1}^n X_i + 1$ y $q = n - \sum_{i=1}^n X_i + 1$.

Mediante las tablas de la distribución $Beta$, para el caso $n = 10$ $\sum_{i=1}^{10} x_i = 3$ y $1 - \alpha = 0'90$ se obtiene que el intervalo creíble es $(0'135, 0'564)$ donde se ha tomado $F(0'135) = 0'05$ y $F(0'564) = 0'95$, de la distribución $Beta$ de parámetros 4 y 8.

■

Como se ve en este ejemplo, en la definición de región creíble hay una indeterminación a la hora de escoger el intervalo, aquí se ha optado por lo más sencillo, que consiste en determinar las colas de la distribución de probabilidad final iguales y por tanto de valor $\frac{\alpha}{2}$. Lo más correcto es escoger la región de manera que los puntos en ella tengan la máxima probabilidad final posible, con lo que además se consigue que cuando la región sea un intervalo tenga amplitud mínima. En esta situación se habla de región creíble de más alta distribución final, y lo será $C(x_1, ..., x_n)$ si es de la forma $C(x_1, ..., x_n) = \{\theta | \pi(\theta | x_1, ..., x_n) \geq c\}$ donde c se determina para que $P\{\theta \in C(x_1, ..., x_n) | x_1, ..., x_n\} \geq 1 - \alpha$. Obsérvese que a medida que c disminuye esta probabilidad va aumentando, cuando c vale 0 $P(C(x_1, ..., x_n) | \overrightarrow{x}) = 1$.

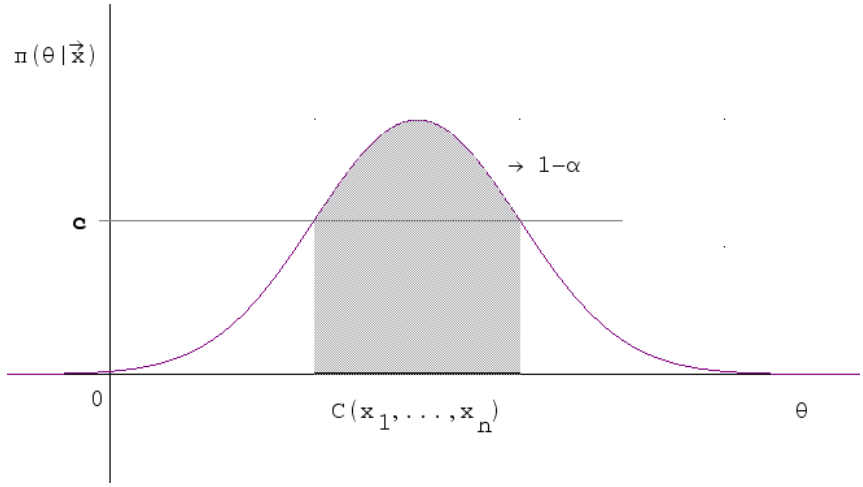


Figura 4.5.1. Determinación de la región creíble de más alta probabilidad final.

Así en el Ejemplo 4.5.1, el intervalo creíble de más alta probabilidad viene dado por $(0'117, 0'542)$ lo que puede comprobarse a partir de la distribución *Beta*. Se observa que este intervalo tiene menor amplitd que el obtenido en el Ejemplo 4.5.1.

Ejemplo 4.5.2

Para m.a.s. de tamaño n de una población $N(\theta, 1)$, cuando la información inicial sobre θ puede suponerse $N(0, 1)$, al ser la distribución final $N\left(\frac{n\overline{x}}{n+1}, \frac{1}{\sqrt{n+1}}\right)$, véase el Ejemplo 3.2.3.3, el intervalo creíble viene dado por

$$\left(\frac{n\overline{x}}{n+1} - z_{\frac{\alpha}{2}} \frac{1}{\sqrt{n+1}},\ \frac{n\overline{x}}{n+1} + z_{\frac{\alpha}{2}} \frac{1}{\sqrt{n+1}}\right).$$

Recordemos que el intervalo de confianza obtenido desde el punto de vista clásico es

$$\left(\overline{x} - z_{\frac{\alpha}{2}} \frac{1}{\sqrt{n}},\ \overline{x} + z_{\frac{\alpha}{2}} \frac{1}{\sqrt{n}}\right)$$

que tiene mayor amplitud, aparte de su diferente interpretación.
■

El estudio de las regiones de confianza óptimas descansa, básicamente, en las propiedades de óptimo de los contrastes estadísticos por lo que ha de retrasarse a la Sección 5.5.

4.6. APROXIMACIÓN HISTÓRICA

Desde el punto de vista histórico, los primeros trabajos sobre regiones de confianza se debieron a Fisher alrededor de 1925. Ronald A. Fisher, nació el 17 de Febrero de 1890 en East Finchley (Londres) era el más jóven de ocho hermanos. Desde muy pequeño mostró una gran habilidad para los razonamientos matemáticos, en particular para la geometría.

Al haber tenido muy mala visión, el médico le prohibió estudiar con luz artificial, lo cual pudo ayudarle, al no poder escribir demasiado, a desarrollar su gran intuición geométrica.

En 1909 obtuvo una beca para el *Caius College* de Cambridge donde se diplomó en 1913. Allí estudió matemáticas y mecánica y se apasionó por la Biometría y la Genética. En 1912 publicó el primero de sus 395 artículos, en el que introduce un criterio para ajustar curvas de frecuencias, en él utiliza, sin nombrarlo así, el método de la máxima verosimilitud. Trabajó en una compañía mercantil en Londres hasta 1915. En este año publicó un trabajo sobre la distribución del coeficiente de correlación. Este artículo le dió fama entre la comunidad estadística y le puso en contacto con Student. Tras cuatro años de dedicarse a la enseñanza, aceptó en 1919 un puesto en la Estación Experimental de Rothamsted, 50 km al noroeste de Londres, que prefirió al que le ofreció Karl Pearson para que se incorporara al Laboratorio Galton del *University College* de Londres, que Karl Pearson dirigía. En esta estación experimental se dedicó hasta 1933 a su otra vocación aparte de la estadística: la genética. A la jubilación de Karl Pearson en 1933, Fisher se ocupó del Departamento de Eugenesia en el *University College*. Su último puesto en Inglaterra es de catedrático de genética en Cambridge, desde 1943 hasta 1957. En este año se traslada a Adelaida, en Australia, donde muere en 1962.

Durante su estancia en Rothamsted desarrolló el diseño de experimentos, pero también se debe a él prácticamente todo el desarrollo de la teoría de la estimación estadística. Estableció claramente la distinción entre muestra y población e introdujo los conceptos de suficiencia y ancilaridad, definió

la cantidad de información que lleva su nombre, desarrolló el método de estimación de la máxima verosimilitud y obtuvo las propiedades asintóticas del mismo.

Lo abundante de su producción científica no le impidió participar en varias instituciones estadísticas; fué presidente de la *Royal Statistical Society*, de la Sociedad de Biometría y presidente del Instituto Internacional de Estadística. Además de esto, tuvo tiempo para polemizar duramente con algunos de sus colegas.

Una reedición de los tres principales libros de Fisher, puede verse en Bennet (1990).

En cuanto a la determinación de las regiones de confianza, su deducción se apoyó en el concepto de función de verosimilitud y le llevó a obtener la distribución que él llamaba *fiducial*. De forma resumida si la función de verosimilitud $L(\theta|\overrightarrow{x}) = f(\overrightarrow{x}|\theta)$ es integrable sobre todo el espacio paramétrico, la distribución fiducial viene dada por

$$\frac{L(\theta|\overrightarrow{x})}{\int_{\Theta} L(\theta|\overrightarrow{x})d\theta}$$

A partir de esta distribución determinaba Fisher dos valores que contuvieran con probabilidad alta, $1-\alpha$, a θ y que constituían el intervalo fiducial.

Esta aproximación no ha sido continuada, sin embargo, con demasiado éxito, pues presenta algunos inconvenientes, básicamente de interpretación.

La otra figura que contribuyó al desarrollo de las regiones de confianza fué Neyman alrededor de 1937 como consecuencia del desarrollo de los contrastes de hipótesis, que serán estudiados en el siguiente capítulo.

Jerzy Neyman nació el 16 de Abril de 1894 en Bendery, en Rusia, en el seno de una familia de origen polaco. Cursó estudios superiores de matemáticas y física en la universidad de Kharkov. Allí se interesó por los desarrollo de la teoría de la medida de Lebesgue, dominio en el que publicó sus primeros trabajos.

Durante la revolución rusa de 1917 a 1921 la vida de Neyman corre peligro, por lo que se traslada a Polonia, donde obtiene su doctorado en 1923. En 1924 consigue una beca para trabajar en el *University College* de Londres bajo la dirección de Karl Pearson, y allí coincide con Fisher, Student y Egon Pearson, hijo de Karl Pearson. En colaboración con Egon Pearson y con Fisher inicialmente, desarrolla la teoría de los contrastes de hipótesis y de la estimación por regiones de confianza, hasta que choca con este último. En 1925 viaja a París y asiste a los cursos impartidos por Lebesgue y Borel.

De vuelta a Polonia, dirige el Departamento de Estadística del Instituto Nencki de Biología Experimental hasta 1934, en que Egon Pearson le propone un puesto de profesor en el *University College*, que Neyman ocupa durante cuatro años. En agosto de 1938 se traslada a la Universidad

de California, en Berkeley, para crear un Departamento de Estadística y verosímilmente también, huyendo de la Segunda Guerra Mundial que veía acercarse.

Se quedó en Berkeley hasta el fin de sus días impulsando el Departamento de Estadística. Participó en el Instituto Internacional de Estadística y en la Sociedad Bernoulli, su rama europea. Murió el 5 de Agosto de 1981.

Se interesó en teoría de la probabilidad, física estadística y procesos estocásticos; y quizás sus contribuciones en estadística más importantes fueron la creación de los tests estadísticos, en colaboración con Egon Pearson, y la elaboración de la moderna teoría de muestras.

Las diferencias entre la aproximación fiducial y la de regiones de confianza, produjo la ruptura entre Fisher y Neyman. Se debe apuntar en el haber de Neyman, que éste no respondió jamás a las críticas de Fisher, salvo en 1961, año en que celebró, con humor por cierto, las bodas de plata de su disidencia con Fisher.

Fisher y el binomio Neyman y Egon Pearson fueron los que contribuyeron a la creación y desarrollo de la estimación mediante regiones de confianza.

4.7. EJERCICIOS

4.7.1

Para m.a.s. de tamaño n, de una población con función de densidad

$$f(x|\theta) = \frac{5x^4}{\theta^5} I_{(0,\theta]}(x) \quad \text{con} \quad \theta > 0$$

a) Hállese el estimador de máxima verosimilitud T.

b) Determínese la región de confianza de grado $0'95$ para θ, de la forma $(\lambda T, +\infty)$ para λ conveniente.

4.7.2

Para m.a.s. de tamaño n, de una población con función de densidad

$$f(x|\theta) = (\theta+1)x^\theta I_{(0,1)}(x) \quad \text{con} \quad \theta > 0$$

constrúyase la región de confianza de grado $1-\alpha$ basada en la variable pivotal $T(X_1, ..., X_n; \theta) = -\sum_{i=1}^n \ln X_i(\theta+1)$ tomando probabilidades colas iguales.

4.7.3

Demuéstrese que para m.a.s. de tamaño n si la función de distribución de la población $F(x|\theta)$, es continua en θ, y como función de θ es estrictamente monótona, entonces

$$T = -\sum_{i=1}^{n} \ln F(x_i|\theta) \sim \chi^2_{2n}$$

y, por tanto, constituye una variable pivotal.

4.7.4
Para una población con función de densidad

$$f(x|\theta) = \frac{4}{\theta^2} x \; I_{(0,\frac{\theta}{2}]}(x) - \frac{4}{\theta^2}(x-\theta) I_{(\frac{\theta}{2},\theta]}(x) \quad \text{con} \;\; \theta > 0,$$

y a partir de una observación

a) Hállese el estimador de máxima verosimilitud T para θ.

b) Constrúyase una región de confianza de grado $1-\alpha$ de la forma $(\frac{T}{2}, \lambda T)$.

4.7.5
Para una población con función de densidad

$$f(x|\theta) = e^{-(x-\theta)} I_{[\theta,+\infty)}(t) \qquad \theta \in \mathbb{R},$$

a partir de m.a.s. de tamaño n, determínese un intervalo de confianza de grado $1-\alpha$ mediante el método de la variable pivotal con probabilidades de colas iguales.

4.7.6
Determínese a partir de m.a.s. de tamaño n de una población con función de densidad

$$f(x|\theta) = 3x^2\theta^{-3} I_{(0,\theta)}(x)$$

un intervalo de grado de confianza $1-\alpha$ mediante el método de Neyman. Tómense probabilidades de colas iguales.

4.7.7
La cantidad de lluvia, en litros, caída en dos regiones diferentes A y B arroja los siguientes resultados:

A	81'7	77'5	121'4	76'4	79'8	105'1	86'2	72'2	103'2	130'8
B	90'3	50'7	77'1	96'4	95'7	107'4	60'8	106'9	64'7	102'4

Se puede suponer que los datos constituyen m.a.s. de poblaciones normales e independientes.

a) Determínese un intervalo de confianza de grado $0'95$ para la varianza de la primera población.

b) Supuesto que las dos varianzas poblacionales son desconocidas pero iguales determínese un intervalo de confianza de grado $0'95$ para la diferencia de medias poblacionales.

c) Repítase la parte anterior si se supone que las dos varianzas son desconocidas y diferentes.

4.7.8
Para estudiar la conveniencia de aumentar sus instalaciones, una empresa desea estimar la demanda que va a tener durante el próximo año. Para ello

selecciona al azar a 10 de sus clientes y observa que el número de unidades demandadas en el último año se distribuye con arreglo a la siguiente tabla:

nº de unidades	1000	1002	1004	1006	1008	1010	1012
nº de clientes que las demandan	1	2	1	2	1	2	1

Si se supone que la demanda va a seguir comportándose igual y que la varianza poblacional es 16, determínese una región de grado de confianza $1-\alpha$ para la media poblacional:

a) Sin efectuar hipótesis sobre la distribución de la población.

b) Suponiendo que la población es $N(\theta, \sigma)$.

4.7.9

Con vistas a comparar la duración de dos marcas de televisores, se observan las duraciones hasta la primera avería, de dos m.a.s. de tamaño 10, en miles de horas, obteniéndose los siguientes resultados:

tipo A	19'9	19'4	20'2	19'2	20'5	10'2	15'1	10'5	19'4	20'6
tipo B	27'5	27'2	21'2	23'9	29'5	31'8	38'1	37'2	29'8	31'3

Si se supone que el número de horas que funciona un televisor hasta estropearse, sigue una distribución normal, determínese un intervalo para el cociente de varianzas de grado 0'95.

4.7.10

Determínese un rectángulo de grado de confianza 0'95 para la media y la varianza de la cantidad de lluvia caida en la región A, con los datos del ejercicio 4.7.7.

4.7.11

Obténgase un intervalo asintótico de grado de confianza $1-\alpha$ para la probabilidad de éxito de una variable aleatoria de $Bernoulli(\theta)$. Aplíquese para una moneda con $1-\alpha = 0'95$, si se sabe que en 100 lanzamientos se han obtenido 40 caras.

4.7.12

Obténgase un intervalo asintótico de grado de confianza $1-\alpha$ para el parámetro θ en una v.a. de $Poisson(\theta)$. Se supone que el número de insectos en una fila sembrada de un cultivo determinado, se distribuye mediante una distribución de Poisson. Se toman diez filas al azar y se obtienen los siguientes números de insectos

$$160, 130, 36, 42, 53, 70, 45, 50, 104, 29.$$

Determínese un intervalo asintótico de grado de confianza $0'9$ para el número medio de insectos por fila.

4.7.13
Determínese un intervalo creíble de probabilidad $1-\alpha$ para la probabilidad de éxito θ en una distribución de *Bernoulli*. Supóngase que la información inicial queda recogida por una distribución $Beta(1,1)$ y que en 10 repeticiones del experimento se han observado 5 éxitos.

4.7.14
Determínese un intervalo creíble de probabilidad $1-\alpha$ para el parámetro θ de una distribución de $Poisson(\theta)$, cuando la información inicial para θ viene dada por una distribución $Gamma(a,p)$.

Para los datos del Ejercicio 4.7.12 determínese un intervalo creíble de probabilidad $0'9$ para el número medio de insectos por fila, si la información inicial viene recogida por la distribución $Gamma(0'5, 30)$.

4.7.15
Un individuo realiza un test de inteligencia cuyo resultado X se supone que sigue una distribución $N(\theta, \sigma = 10)$, siendo θ su nivel de inteligencia real. La información inicial viene recogida porque en el colectivo a que pertenece el individuo, la inteligencia θ tiene una distribución $N(100, \sigma_0 = 15)$. Determínese una región creíble de probabilidad $0'95$ para su nivel de inteligencia cuando el resultado del test ha sido 110. Calcúlese un intervalo de grado de confianza $0'95$ y compárense ambos intervalos.

4.7.16
Sean $X_1, ..., X_n$ v.a. i.i.d. con distribución $N(\theta, \sigma)$, donde ambos parámetros son desconocidos. Se considera la información inicial dada mediante $\pi(\mu, \sigma^2) = 1/\sigma^2$ para $\mu \in \mathbb{R}$ $\sigma^2 \in \mathbb{R}^+$. Calcúlese la distribución final $\pi(\mu, \sigma^2 | x_1, ..., x_n)$.

a) Demuéstrese que $\pi(\mu | x_1, ..., x_n)$ es tal que $\frac{\overline{x}-\mu}{S/\sqrt{n}} | x_1, ..., x_n \sim T_{n-1}$, y obténgase para μ una región creíble de más alta probabilidad $1-\alpha$, compárese con la región de grado de confianza $1-\alpha$.

b) Demuéstrese que $\pi(\sigma^2 | x_1, ..., x_n)$ es tal que la distribución $(n-1)\frac{S^2}{\sigma^2} | x_1, ..., x_n \sim \chi^2_{n-1}$, y obténgase para σ^2 una región creíble de probabilidad $1-\alpha$, compárese con la región de grado de confianza $1-\alpha$.

4.8. SOLUCIÓN A LOS EJERCICIOS

4.8.1

a) La función de densidad de la muestra es

$$f(x_1, ..., x_n | \theta) = 5^n \frac{\prod_{i=1}^{n} x_i^4}{\theta^{5n}} I_{[x_{(n)}, +\theta)}(\theta)$$

por lo que el EMV es $\widehat{\theta} = X_{(n)}$.

b) La región de confianza, de la forma $(\lambda X_{(n)}, +\infty)$ estará determinada, cuando se obtenga el valor de λ tal que

$$P\{\lambda X_{(n)} \leq \theta|\theta\} = 1-\alpha \quad \text{para todo } \theta > 0$$

para lo que ha de cumplirse que

$$P\{X_{(n)} \leq y|\theta\} = P^n\{X \leq y|\theta\} = \left(\frac{y}{\theta}\right)^{5n}$$

de donde ha de ser $\lambda^{5n} = 1-\alpha$; o lo que es lo mismo $\lambda = (1-\alpha)^{\frac{1}{5n}}$ y

$$RC_{1-\alpha}(\theta) = \left((1-\alpha)^{\frac{1}{5n}} X_{(n)}, +\infty\right).$$

4.8.2

Por el Teorema 4.2.1 (variable pivotal) se debe obtener la distribución de la variable $y = -(\theta+1)\ln x$. Si se hace el cambio $x = e^{-y/(\theta+1)} \quad \left|\frac{dx}{dy}\right| = \frac{1}{\theta+1}e^{-\frac{y}{\theta+1}}$ resulta que $g(y|\theta) = e^{-y}$ con $y > 0$, luego es independiente de θ. Haciendo el cambio a $z = 2y$ con lo que $y = \frac{1}{2}z$, $\left|\frac{dy}{dz}\right| = \frac{1}{2}$ y $g(z) = \frac{1}{2}e^{-\frac{1}{2}z}$, con $z > 0$ y se tiene que $Z \sim Gamma(a = \frac{1}{2}, p = 1)$. En resumen que

$$-2(\theta+1)\sum_{i=1}^{n} \ln X_i \sim \chi^2_{2n}.$$

Mediante las tablas de la distribución χ^2_{2n}, es posible determinar $\chi^2_{2n;1-\frac{\alpha}{2}}$ y $\chi^2_{2n;\frac{\alpha}{2}}$ tales que

$$P\left\{\chi^2_{2n;1-\frac{\alpha}{2}} \leq -2(\theta+1)\sum_{i=1}^{n} \ln X_i \leq \chi^2_{2n;\frac{\alpha}{2}}\right\} = 1-\alpha$$

y por tanto el intervalo de confianza de grado 1-α es

$$IC_{1-\alpha}(\theta) = \left(\frac{\frac{1}{2}\chi^2_{2n;1-\frac{\alpha}{2}} - \sum_{i=1}^{n} \ln X_i}{-\sum_{i=1}^{n} \ln X_i}, \frac{\frac{1}{2}\chi^2_{2n;\frac{\alpha}{2}} - \sum_{i=1}^{n} \ln X_i}{-\sum_{i=1}^{n} \ln X_i}\right)$$

4.8.3

Efectivamente, por ser $F(\cdot|\theta)$ continua, $Y = F(X_i|\theta) \sim U(0,1)$ y $-2\ln F(X|\theta)$ ~ χ^2_2 ya que si se hace el cambio $z = -2\ln y$, con lo que $y = e^{-\frac{1}{2}z}$, $\left|\frac{dy}{dz}\right| = \frac{1}{2}e^{-\frac{1}{2}z}$ se obtiene $g(z) = \frac{1}{2}e^{-\frac{1}{2}z}$ que es la densidad χ^2_2, de donde la distribución de

$$T = -2\sum_{i=1}^{n} \ln F(X_i|\theta) \sim \chi^2_{2n}.$$

Por lo tanto se pueden determinar los cuantiles de la distribución χ^2_{2n} para que

$$P\{\chi^2_{2n;1-\frac{\alpha}{2}} \leq T(X_1, ..., X_n; \theta) \leq \chi^2_{2n;\frac{\alpha}{2}}\} = 1 - \alpha,$$

la monotonía estricta en θ y la invertibilidad de las ecuaciones

$$\chi^2_{2n;1-\frac{\alpha}{2}} = T(x_1, ..., x_n; \underline{\theta}(x_1, ..., x_n))$$
$$\chi^2_{2n;\frac{\alpha}{2}} = T(x_1, ..., x_n; \overline{\theta}(x_1, ..., x_n))$$

permite obtener el intervalo de confianza $(\underline{\theta}(X_1, ..., X_n), \overline{\theta}(X_1, ..., X_n))$, de acuerdo con el Teorema 4.2.1 (variable pivotal).

4.8.4

a) La función de densidad viene dada en la Figura 4.8.1

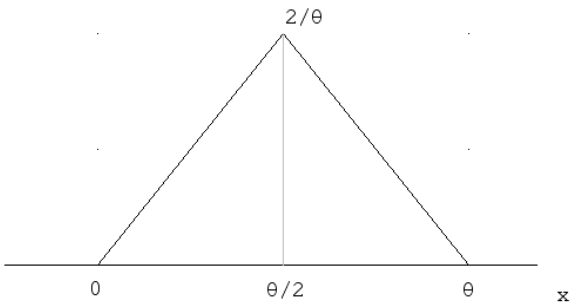


Figura 4.8.1. Función de densidad triangular correspondiente al Ejercicio 4.7.4.

como hay que maximizarla como función de θ para un valor de x fijo

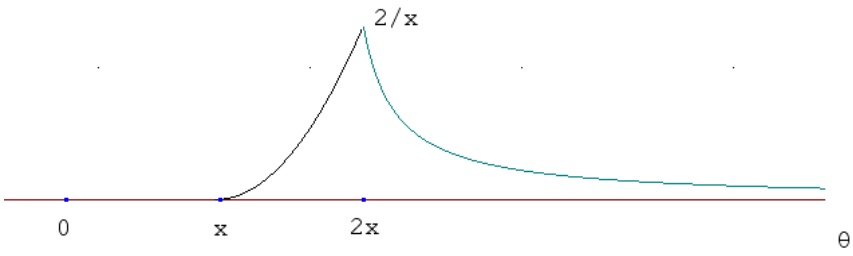


Figura 4.8.2. Función de verosimilitud correspondiente a la densidad triangular del Ejercicio 4.7.4.

es claro, que de acuerdo con la Figura 4.8.2 el estimador de máxima verosimilitud es $\widehat{\theta} = 2X$.

b) Se trata de determinar λ de manera que

$$P\{X \leq \theta \leq \lambda 2X|\theta\} = 1 - \alpha \quad \text{para todo } y > 0.$$

Pero esta expresión es equivalente a

$$P\left\{ \frac{\theta}{2\lambda} \leq X \leq \theta \middle| \theta \right\} = 1 - \alpha \ \text{ para todo } \theta > 0$$

siempre que $\lambda \geq 1/2$, véase la Figura 4.8.3, de donde

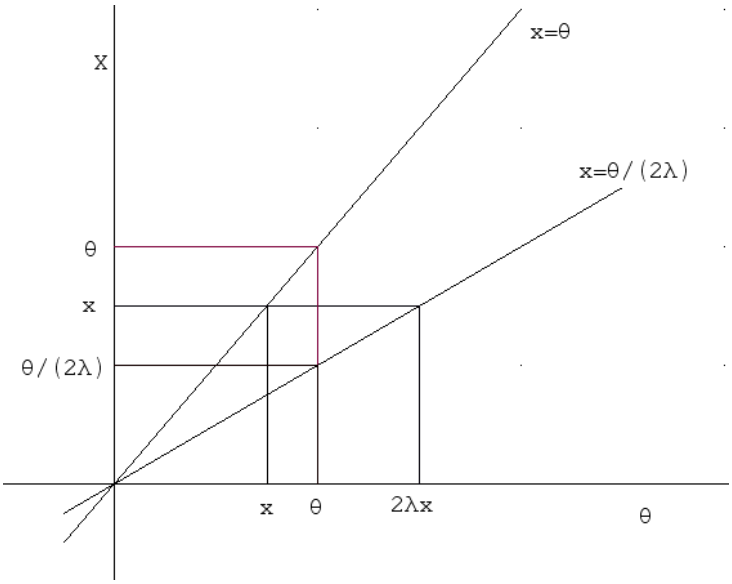


Figura 4.8.3. Obtención del intervalo de confianza para el Ejercicio 4.7.4.

$$\int_{\frac{\theta}{2\lambda}}^{\frac{\theta}{2}} \frac{4}{\theta^2} x dx + \int_{\frac{\theta}{2}}^{\theta} \frac{4}{\theta^2}(\theta - x)dx = \frac{4}{\theta^2}\left[\frac{x^2}{2}\right]_{\frac{\theta}{2\lambda}}^{\frac{\theta}{2}} + \frac{4}{\theta^2}\left[\theta x - \frac{x^2}{2}\right]_{\frac{\theta}{2}}^{\theta}$$
$$= 1 - \frac{1}{2\lambda^2}$$

que al igualar a $1 - \alpha$ permite obtener dos soluciones en λ de las que solo es válida $\lambda = 1/\sqrt{2\alpha}$.

Si se hubiera supuesto que $\lambda \leq 1/2$, habría quedado en la Figura 4.8.3 la recta $x = \theta/(2\lambda)$ por encima de la $x = \theta$; el intervalo hubiera sido $(2\lambda x, x)$, pero este intervalo en θ es equivalente al $(\theta, \theta/(2\lambda))$ en x y es imposible hacer que este tenga de probabilidad $1 - \alpha$, con α distinto de uno, ya que es cero la probabilidad de que x sea mayor que θ.

Por lo tanto el intervalo pedido es $IC_{1-\alpha}(\theta) = \left(X, \sqrt{\frac{2}{\alpha}}X\right)$.

4.8.5

Por ser la función de densidad de la muestra

$$f(x_1,...,x_n|\theta) = e^{-\sum_1^n (x_i-\theta)} \prod_{i=1}^n I_{[\theta,+\infty)}(x_i)$$
$$= e^{-\sum_i^n (x_i-\theta)} I_{(-\infty,x_{(1)}]}(\theta)$$

se puede tratar de ver, si es variable pivotal $T(X_{(1)},\theta) = X_{(1)} - \theta$.

En efecto la función de distribución de esta variable es para t no negativo

$$F(t) = P\{X_{(1)} - \theta \leq t\}$$
$$= 1 - P^n\{X > t + \theta\}$$
$$= 1 - e^{-nt}$$

independiente de θ. Si se aplica el Teorema 4.2.1 (variable pivotal) se cumple que

1) $T(\cdot,\theta)$ tiene distribución independiente de θ y es una función medible

2) $\mathrm{T}(\mathrm{X}_{(1)},\cdot)$ es una función continua y monótona estrictamente decreciente en θ

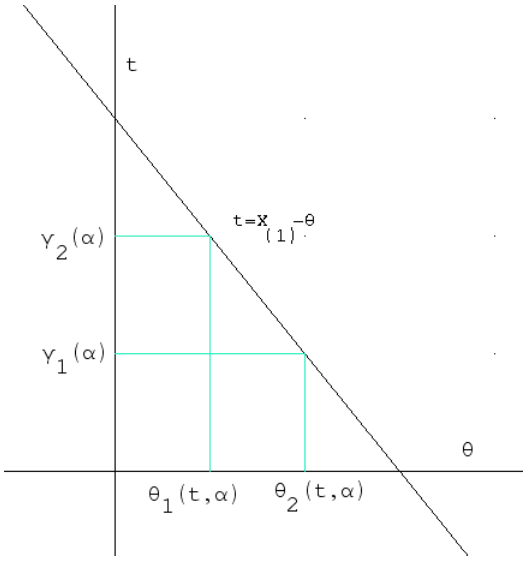


Figura 4.8.4. Representación del método de la variable pivotal.

luego T es una variable pivotal y hay que determinar $\gamma_i(\alpha)$, $i = 1, 2$, para que

$$P\{\gamma_1(\alpha) \leq T \leq \gamma_2(\alpha)\} \geq 1 - \alpha,$$

lo que puede lograrse resolviendo las ecuaciones

$$F(\gamma_1(\alpha)) = \alpha_1 \qquad 1 - F(\gamma_2(\alpha)) = \alpha - \alpha_1.$$

De acuerdo con el enunciado se debe usar $\alpha_1 = \frac{\alpha}{2}$, en este caso es fácil hacer la determinación de $\alpha_1 \in [0, \alpha]$ para que la amplitud del intervalo de confianza sea mínima.

$$1 - e^{n\gamma_1(\alpha)} = \alpha_1 \qquad 1 - (1 - e^{-n\gamma_2(\alpha)}) = 1 - \alpha + \alpha_1$$

lo que lleva, de acuerdo con la Figura 4.8.4, a

$$\theta_1(t, \alpha) = X_{(1)} + \frac{1}{n}\ln(\alpha - \alpha_1) \qquad \theta_2(t, \alpha) = X_{(1)} + \frac{1}{n}\ln(1 - \alpha_1),$$

de donde el intevalo de confianza pedido se obtiene para $\alpha_1 = \frac{1}{2}$ y es

$$IC_{1-\alpha}(\theta) = \left(X_{(1)} + \frac{1}{n}\ln\frac{\alpha}{2}, X_{(1)} + \frac{1}{n}\ln(1 - \frac{\alpha}{2}) \right)$$

la amplitud del intervalo en función de α_1 se hace mínima cuando

$$\min_{0 \le \alpha_1 \le \alpha} \ln\frac{1-\alpha}{\alpha - \alpha_1} = \ln\frac{1-\alpha}{\alpha}$$

es decir si α_1 vale 0 y por lo tanto el intervalo de confianza de amplitud mínima, basado en la variable pivotal es $(X_{(1)} + \frac{1}{n}\ln\alpha, X_{(1)})$.

4.8.6
La función de distribución de la población es, para $\theta > 0$

$$F(x|\theta) = \begin{cases} 0 & \text{si } x < 0 \\ \left(\frac{x}{\theta}\right)^3 & \text{si } 0 \le x \le \theta \\ 1 & \text{si } x \ge \theta \end{cases}$$

Si se aplica el Teorema 4.2.2 (método de Neyman), basado en el estadístico suficiente $X_{(n)}$, ya que

$$f(x_1, ..., x_n|\theta) = 3^n\theta^{-n} I_{[x_{(n)}, +\infty)}(\theta) \prod_{i=1}^{n} \left(\frac{x_i}{\theta}\right)^2.$$

Se trata de determinar dos funciones $\gamma_i(\theta, \alpha)$ $i = 1, 2$ tales que para $\theta > 0$

$$P\{\theta_1(\theta, \alpha) \le X_{(n)} \le \gamma_2(\theta, \alpha)|\theta\} \ge 1 - \alpha$$

para lo cual se obtiene la función de distribución del estadístico

$$P\{X_{(n)} \le t\} = P^n\{X \le t\} = \begin{cases} 0 & \text{si } t < 0 \\ \left(\frac{t}{\theta}\right)^{3n} & \text{si } 0 \le t \le \theta \\ 1 & \text{si } t \ge \theta \end{cases}$$

de donde han de ser

$$F(\gamma_1(\theta,\alpha)) = \alpha_1 \qquad 1 - F(\gamma_2(\theta,\alpha)) = \alpha - \alpha_1$$

con $0 \leq \alpha_1 \leq \alpha$, o lo que es lo mismo

$$\left(\frac{\gamma_1(\theta,\alpha)}{\theta}\right)^{3n} = \alpha_1 \qquad 1 - \left(\frac{\gamma_2(\theta,\alpha)}{\theta}\right)^{3n} = \alpha - \alpha_1,$$

de acuerdo con la Figura 4.8.5 ha de ser

$$\theta_1(t,\alpha) = X_{(n)}(1-\alpha-\alpha_1)^{-\frac{1}{3n}} \qquad \theta_2(t,\alpha) < X_{(n)}\alpha_1^{-\frac{1}{3n}}$$

con lo que el intervalo de colas iguales, para $\alpha_1 = \frac{\alpha}{2}$ viene dado por

$$IC_{1-\alpha}(\theta) = \left(X_{(n)}\left(1-\frac{\alpha}{2}\right)^{-\frac{1}{3n}},\ X_{(n)}\left(\frac{\alpha}{2}\right)^{-\frac{1}{3n}}\right)$$

Aquí también es sencillo determinar el intervalo de amplitud mínima, sin más que observar que el mínimo de la expresión

$$X_{(n)}(\alpha_1^{-\frac{1}{3n}} - (1-\alpha+\alpha_1)^{-\frac{1}{3n}})$$

se alcanza para $\alpha_1 = \alpha$ y vale $(X_{(n)},\ X_{(n)}\alpha^{-\frac{1}{3n}})$.

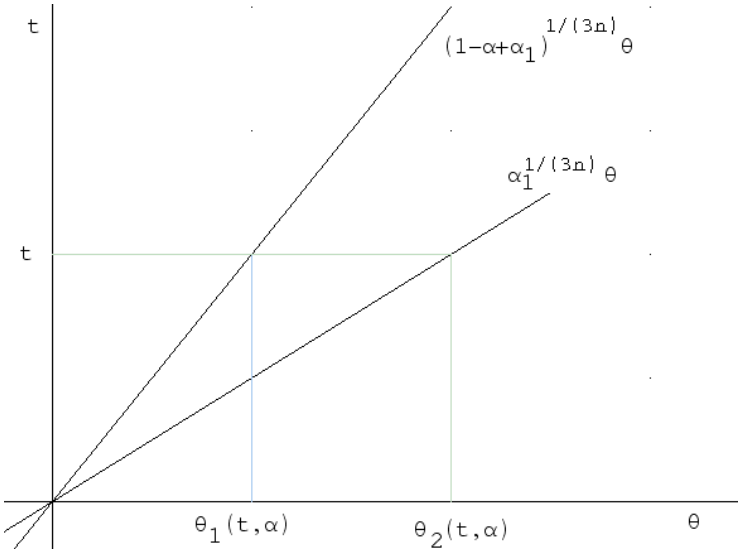


Figura 4.8.5. Representación del teorema de Neyman para el Ejercicio 4.8.6.

4.8.7

a) El intervalo de confianza correspondiente a σ_1^2 viene dado por

$$IC_{1-\alpha}(\sigma_1^2) = \left(\frac{(n-1)S_1^2}{\chi^2_{n-1;\frac{\alpha}{2}}},\ \frac{(n-1)S_1^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}}\right)$$

donde $S_1^2 = 420'069$, $\chi^2_{9;025} = 19$ y $\chi^2_{9;0'975} = 2'7$ por lo que el intervalo es $(198'98, 1400'23)$.

b) El intervalo de confianza para la diferencia de medias, cuando las dos varianzas son desconocidas pero iguales, es de acuerdo con 4.3.6

$$IC_{1-\alpha}(\mu_1 - \mu_2) = \left(\overline{X} - \overline{Y} - t_{n_1+n_2-2;\frac{\alpha}{2}} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \right.$$
$$\left. \overline{X} - \overline{Y} - t_{n_1+n_2-2;\frac{\alpha}{2}} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

en este caso $\overline{X} = 93'48$, $\overline{Y} = 85'24$, $S_1^2 = 420'069$ y $S_2^2 = 421'414$ por lo que la varianza muestral ponderada es $S_p^2 = 420'604$. Mediante las tablas de la distribución de Student, véase la tabla 5 al final del libro, se obtiene $t_{18;0'025} = 2'101$ de donde el intervalo es $(-12'1047, 28'4847)$.

c) Cuando las dos varianzas son desconocidas y diferentes

$$IC_{1-\alpha}(\mu_1 - \mu_2) = \left(\overline{X} - \overline{Y} - t_{f;\frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}},\ \overline{X} - \overline{Y} - t_{f;\frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right)$$

con $f = \text{mín}\{n_1, n_2\} - 1 = 9$. Mediante las tablas de la distribución de Student es $t_{9;0'025} = 2'262$ lo que lleva al intervalo $(-12'5598, 28'9398)$, ligeramente más amplio que el obtenido en la parte anterior.

4.8.8

a) Mediante la acotación de Tchebycheff aplicada a la media muestral se sigue que

$$P\{|\overline{X} - \mu| \leq \varepsilon\} \geq 1 - \frac{1}{\varepsilon^2} V[\overline{X}]$$

de donde si se determina ε para que

$$1 - \frac{1}{\varepsilon^2} \frac{\sigma^2}{n} = 0'95$$

se puede emplear como intervalo de confianza aproximado $(\overline{X} - \varepsilon, \overline{X} + \varepsilon)$, lo que lleva a $\varepsilon^2 = 16/0'5$, $\varepsilon = 5'657$ y como

$$\overline{X} = \frac{1000 \cdot 1 + ... + 1012 \cdot 1}{10} = 1006$$

se sigue que $IC_{0'95}(\mu) \simeq (1000'343, 1011'657)$.

b) Al suponer que la distribución de partida es normal se puede determinar el intervalo de confianza para la media con varianza conocida y de acuerdo con la Expresión 4.3.1

$$IC_{1-\alpha}(\theta) = (\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}},\ \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}})$$

que como $z_{0'025} = 1'96$ es $(1003'52, 1008'479)$. Lógicamente, al tener más información el intervalo es más pequeño.

4.8.9

De acuerdo con la Expresión (4.3.8), el intervalo de confianza de grado $1-\alpha$, viene dado por

$$IC_{1-\alpha}\left(\frac{\sigma_1^2}{\sigma_2^2}\right) = \left(\frac{S_1^2}{S_2^2} f^{-1}_{n_1-1,n_2-1;\frac{\alpha}{2}},\ \frac{S_1^2}{S_2^2} f_{n_2-1,n_1-1;\frac{\alpha}{2}}\right)$$

como $S_1^2 = 2'8689, S_2^2 = 28'056$ y mediante las tablas de la distribución de Snedecor $f_{9,9;0'025} = 4'03$ $f^{-1}_{9,9;0'025} = 0'2481$; por lo que le intervalo es $(0'0238, 0'3861)$.

4.8.10

Mediante la Proposición 4.3.1 se tiene que como

$$P\left\{\left(\overline{X} - t_{n-1;\frac{\alpha}{4}}\frac{S}{\sqrt{n}}, \overline{X} + t_{n-1;\frac{\alpha}{4}}\frac{S}{\sqrt{n}}\right) \ni \mu \,\middle|\, (\mu,\sigma^2)\right\} \geq 1-\frac{\alpha}{2}$$

$$P\left\{\left(\frac{(n-1)S^2}{\chi^2_{n-1;\frac{\alpha}{4}}}, \frac{(n-1)S^2}{\chi^2_{n-1;1-\frac{\alpha}{4}}}\right) \ni \sigma^2 \,\middle|\, (\mu,\sigma^2)\right\} \geq 1-\frac{\alpha}{2}$$

basta obtener $\overline{X} = 93'48$ $S^2 = 42''069$ y $t_{9;0'0125} = 2'686$ con lo que el primer intervalo es $(76'07, 110'89)$. A partir de $\chi^2_{9;0'0125} = 21'0345$ y $\chi^2_{9;0'9875} = 2'219$ se obtiene el segundo $(179'7343, 1703'7499)$.

4.8.11

Si X tiene distribución de $Bernoulli(\theta)$ mediante la distribución asintótica del estimador de máxima verosimilitud (véase la Sección 3.6.3) se tiene que aproximadamente $\overline{X}$ tiene distribución $N(\theta, \frac{\sigma(\theta)}{\sqrt{n}})$ con $\sigma(\theta) = \sqrt{\theta(1-\theta)}$. Se puede estimar esta expresión por $\sigma(\overline{X})$, por lo que el intervalo de confianza asintótico viene dado por

$$IC_{1-\alpha}(\theta) \simeq \left(\overline{X} - z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{X}(1-\overline{X})}{n}},\ \overline{X} + z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}\right)$$

(véase la Sección 4.4) como $z_{0'025} = 1'96$ y $\overline{X} = 0'4$ se obtiene como estimador el intervalo $(0'304, 0'496)$.

4.8.12

El EMV de θ es $\overline{X}$, de manera que por el Teorema central del límite o por el comportamiento asintótico de los EMV (véase la Sección 3.6.3)

$$\overline{X} \sim N\left(\theta, \sqrt{\frac{\theta}{n}}\right),$$

debido a la convergencia c.s. de $\overline{X}$ a θ se puede suponer que aproximadamente

$$\frac{\overline{X}-\theta}{\sqrt{\frac{\overline{X}}{n}}} \simeq N(0,1)$$

por lo que a partir de las tablas de la distribución normal se puede obtener $z_{\frac{\alpha}{2}}$ tal que

$$P\left\{-z_{\frac{\alpha}{2}} \leq \frac{\overline{X}-\theta}{\sqrt{\frac{\overline{X}}{n}}} \leq z_{\frac{\alpha}{2}}\right\} = 1-\alpha$$

lo que lleva al intervalo de confianza aproximado

$$\left(\overline{X} - z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{X}}{n}}, \overline{X} + z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{X}}{n}}\right).$$

Si se aplica a la muestra dada, la media muestral es

$$\overline{X} = \frac{160+130+...+29}{10} = 71'9$$

y como $z_{0'05} = 1'645$ queda $(67'4891, 76'3109)$ como intervalo asintótico de grado de confianza $0'9$ para el número medio de insectos por fila θ. Puede compararse este intervalo con el obtenido en el Ejemplo 4.4.2.

4.8.13

Si la información inicial viene recogida por la distribución $Beta(\alpha_0, \beta_0)$, se sigue que la densidad final es una $Beta(\alpha_1, \beta_1)$ donde

$$\alpha_1 = \alpha_0 + \sum_{i=1}^{n} x_i$$
$$\beta_1 = \beta_0 + n - \sum_{i=1}^{n} x_i$$

(véase el Ejemplo 3.2.3.1) la familia *Beta* es conjugada de la verosimilitud *Binomial*. Por lo tanto el intervalo creíble viene dado por $(z_{1-\frac{\alpha}{2}}, z_{\frac{\alpha}{2}})$ cuantiles de la distribución $Beta(\alpha_1, \beta_1)$.

Con $\alpha_0 = 1, \beta_0 = 1, n = 1$ y $\sum_{i=1}^{10} x_i = 5$ se obtiene para $\alpha = 0'1$ que la distribución final es $Beta(6,6)$; con el concurso de un paquete estadístico, los cuantiles de esta distribución son $z_{0'95} = 0'271$ y $z_{0'05} = 0'729$ con lo que θ estará en el intervalo $(0'271, 0'729)$ con probabilidad $0'9$.

4.8.14

La familia de distribuciones *Gamma* es conjugada para la verosimilitud de *Poisson* (véase el Ejercicio 3.7.8). Por tanto la densidad final es una

$Gamma(a_1, p_1)$ con

$$a_1 = a + n$$
$$p_1 = p + \sum_{i=1}^{n} x_i$$

por lo que el intervalo creíble de probabilidad $1-\alpha$ para θ viene dado por $\left(z_{1-\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right)$ los cuantiles de orden $\frac{\alpha}{2}$ y $1-\frac{\alpha}{2}$, respectivamente, de la distribución $Gamma(a_1, p_1)$. Estos valores pueden determinarse fácilmente con cualquier paquete estadístico, o bien se pueden extraer de las tablas de la distribución $\chi^2_{2p_1}$ con $2p_1$ grados de libertad; ya que puede comprobarse que si $\theta \sim Gamma(a_1, p_1)$ la nueva v.a. $\lambda = 2a_1\theta \sim Gamma(\frac{1}{2}, p_1)$ o lo que es igual $Gamma(\frac{1}{2}, \frac{\nu}{2})$ con $\nu = 2p_1$. Es decir $\lambda = 2a_1\theta \sim \chi^2_{2p_1}$ de donde se sigue que el intervalo creíble puede escribirse alternativamente como

$$\left(\frac{1}{2a_1}\chi^2_{2p_1;1-\frac{\alpha}{2}},\ \frac{1}{2a_1}\chi^2_{2p_1;\frac{\alpha}{2}}\right)$$

en función de los cuantiles de la distribución $\chi^2_{2p_1}$.

Si se sustituyen los valores numéricos se obtiene $a_1 = 10'5$, $p_1 = 749$ y $\alpha = 0'1$. Como p_1 es mayor que 100, mediante la aproximación $\sqrt{2\chi^2_{2p_1}} \simeq N(\sqrt{4p_1 - 1}, 1)$ se sigue que para obtener $c_1 = \chi^2_{2p_1;1-\frac{\alpha}{2}}$ ha de obtenerse c_1 de la expresión

$$\frac{\alpha}{2} = P\{\chi^2_{2p_1} \leq c_1\} = P\{Z \leq \sqrt{2c_1} - \sqrt{4p_1 - 1}\}$$

donde $Z \sim N(0, 1)$; por tanto a partir de las tablas de la distribución normal ha de ser $\sqrt{2c_1} - \sqrt{rp_1 - 1} = -1'645$, de donde $c_1 = 1408'8281$.

Análogamente, para la obtención de $c_2 = \chi^2_{2p_1;\frac{\alpha}{2}}$

$$1 - \frac{\alpha}{2} = P\{\chi^2_2 \leq c_2\} = P\{Z \leq \sqrt{2c_2} - \sqrt{4p_1 - 1}\}$$

con lo que $\sqrt{2c_2} - \sqrt{4p_1 - 1} = 1'645$ y $c_2 = 1588'8786$.

Por lo que el intervalo creíble de probabilidad $0'9$ es

$$\left(\frac{c_1}{2a_1}, \frac{c_2}{2a_1}\right) = (67'087, 75'6609)$$

de amplitud ligeramente menor, que el obtenido asintóticamente en el Ejercicio 4.7.12.

4.8.15

Si la información inicial viene dada por $\pi(\theta)$ una distribución $(100, 15)$ y la verosimilitud es $f(x|\theta)$ una densidad $N(\theta, 10)$ la distribución final

$\pi(\theta|x_1,...,x_n)$ es una distribución $N(\mu_1,\sigma_1)$ (véase el Ejercicio 3.7.15) con

$$\mu_1 = \frac{\mu_0/\sigma_0^2 + n\overline{x}/\sigma^2}{\sigma_0^{-2} + n\sigma^{-2}} = \frac{100/225 + x/100}{225^{-1} + 100^{-1}} = \frac{400 + 9x}{13} = 106'923$$
$$\sigma_1^2 = (\sigma_0^{-2} + n\sigma^{-2})^{-1} = (225^{-1} + 100^{-1})^{-1} = 69'23$$

por lo tanto el intervalo creíble de probabilidad $1-\alpha$ vendrá dado por $\left(c_{1-\frac{\alpha}{2}}, c_{\frac{\alpha}{2}}\right)$ donde estos valores son los cuantiles de orden $\frac{\alpha}{2}$ y $1-\frac{\alpha}{2}$ respectivamente de la distribución $N(\mu_1,\sigma_1)$. Por tanto

$$P\left\{c_{1-\frac{\alpha}{2}} \leq \theta \leq c_{\frac{\alpha}{2}}|x_1,...,x_n\right\} = 1-\alpha$$

de donde $c_{1-\frac{\alpha}{2}} = \mu_1 + \sigma_1 z_{1-\frac{\alpha}{2}}$ y $c_{\frac{\alpha}{2}} = \mu_1 + \sigma_1 z_{\frac{\alpha}{2}}$ con $z_{1-\frac{\alpha}{2}}$ y $z_{\frac{\alpha}{2}}$ los cuantiles de la distribución N(0,1) de órdenes $\frac{\alpha}{2}$ y $1-\frac{\alpha}{2}$ respectivamente.

De manera que el intervalo creíble es

$$\begin{aligned} IC_{1-\alpha}(\theta) &= (\mu_1 - \sigma_1 z_{\frac{\alpha}{2}}, \mu_1 + \sigma_1 z_{\frac{\alpha}{2}}) \\ &= (90'6148, 123'2312) \end{aligned}$$

que tiene una amplitud de $2\cdot\sigma_1 z_{\frac{\alpha}{2}} = 2\cdot 8'3205\cdot 1'96 = 32'6164$.

Para el caso clásico, a partir de la Expresión 4.3.1

$$\begin{aligned} IC_{1-\alpha}(\theta) &= \left(\overline{X} - \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}},\ \overline{X} + \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}}\right) \\ &= (X - \sigma z_{\frac{\alpha}{2}},\ X + \sigma z_{\frac{\alpha}{2}}) \\ &= (90'4,\ 129'6) \end{aligned}$$

cuya amplitud es $2\cdot\sigma z_{\frac{\alpha}{2}} = 2\cdot 10\cdot 1'96 = 39'2$.

4.8.16

a) Lo primero que se observa es que la integral a lo largo de todo el espacio paramétrico es infinita, a tal densidad se la denomina densidad impropia. Vamos a ver que a pesar de esto se obtienen densidades marginales a partir de $\pi(\mu,\sigma^2|x_1,...,x_n)$.

La función de densidad de la muestra es

$$f(x_1,...,x_n|\mu,\sigma^2) = \left(\frac{1}{\sqrt{2\pi}}\right)^n (\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2}\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

que si se emplea

$$\sum_{i=1}^{n}(x_i-\mu)^2 = (n-1)S^2 + n(\overline{x}-\mu)^2$$

puede ponerse alternativamente como

$$f(\overrightarrow{x}|\mu,\sigma^2)=\left(\frac{1}{\sqrt{2\pi}}\right)^n(\sigma^2)^{-\frac{n}{2}}e^{-\frac{1}{2\sigma^2}[(n-1)S^2+n(\overline{x}-\mu)^2]}$$

y por tanto la densidad final es

$$\pi(\mu,\sigma^2|\overrightarrow{x})=\frac{(\sigma^2)^{-1}\left(\frac{1}{\sqrt{2\pi}}\right)^n(\sigma^2)^{-\frac{n}{2}}e^{-\frac{1}{2\sigma^2}[(n-1)S^2+n(\overline{x}-\mu)^2]}}{\int_0^{+\infty}\int_{-\infty}^{+\infty}(\sigma^2)^{-1}\left(\frac{1}{\sqrt{2\pi}}\right)^n(\sigma^2)^{-\frac{n}{2}}e^{-\frac{1}{2\sigma^2}[(n-1)S^2+n(\overline{x}-\mu)^2]}d\mu d\sigma^2}$$

simplificando, el denominador es

$$\int_0^{+\infty}(\sigma^2)^{-\left(\frac{n}{2}+1\right)}e^{-\frac{(n-1)S^2}{2\sigma^2}}\left[\int_{-\infty}^{+\infty}e^{-\frac{1}{2\sigma^2/n}(\overline{x}-\mu)^2}d\mu\right]d\sigma^2$$

la integral entre corchetes vale $\sqrt{\frac{2\pi}{n}}(\sigma^2)^{1/2}$, sin más que emplear que para cualquier E y $D>0$ es

$$\int_{-\infty}^{+\infty}e^{-\frac{1}{2D^2}(\mu-E)^2}d\mu=\sqrt{2\pi}D$$

por lo que

$$\pi(\mu,\sigma^2|\overrightarrow{x})=\frac{(\sigma^2)^{-\left(\frac{n}{2}+1\right)}e^{-\frac{1}{2\sigma^2}\left[(n-1)S^2+n(\overline{x}-\mu)^2\right]}}{\sqrt{\frac{2}{n}}\Gamma\left(\frac{1}{2}\right)\int_0^{+\infty}(\sigma^2)^{-\left(\frac{n}{2}+1\right)+\frac{1}{2}}e^{-\frac{(n-1)S^2}{2\sigma^2}}d\sigma^2}$$

Además como

$$\int_0^{+\infty}x^{-(p+1)}e^{-ax^{-1}}dx=a^{-p}\Gamma(p),$$

lo que se sigue haciendo el cambio $ax^{-1}=y$, la integral del denominador vale $a^{-p}\Gamma(p)$ con $a=\frac{(n-1)S^2}{2}$ y $p=\frac{n-1}{2}$. Por lo que el denominador es

$$\frac{1}{\sqrt{n}}2^{\frac{1}{2}}\Gamma\left(\frac{1}{2}\right)\left[(n-1)\frac{S^2}{2}\right]^{-\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)$$

y la densidad marginal

$$\pi(\mu|\overrightarrow{x})=\frac{\int_0^{\infty}(\sigma^2)^{-\left(\frac{n}{2}+1\right)}e^{-\frac{1}{2\sigma^2}\left[(n-1)S^2+n(\overline{x}-\mu)^2\right]}}{\frac{1}{\sqrt{n}}2^{\frac{1}{2}}\Gamma\left(\frac{1}{2}\right)\left[(n-1)\frac{S^2}{2}\right]^{-\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)},$$

volviendo a aplicar que el numerador es $a^{-p}\Gamma(p)$ con $a=\frac{(n-1)s^2+n(\overline{x}-\mu)^2}{2}$ y $p=\frac{n}{2}$ se obtiene que

$$\pi(\mu|\overrightarrow{x})=\sqrt{n}\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{n-1}{2})}\frac{\left[(n-1)S^2+n(\overline{x}-\mu)^2\right]^{-\frac{n}{2}}}{\left[(n-1)S^2\right]^{-\frac{n-1}{2}}}$$

o alternativamente

$$\pi(\mu|\overrightarrow{x}) = \sqrt{n}\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{n-1}{2})}\left[1+\frac{1}{n-1}\left(\frac{\overline{x}-\mu}{s/\sqrt{n}}\right)^2\right]^{-n/2}\frac{1}{\sqrt{n-1}S}$$

por lo que $\pi(t|\overrightarrow{x})$ es la densidad de una distribución de Student T_{n-1} con $t=\frac{\overline{x}-\mu}{S/\sqrt{n}}$.

La región creíble de más alta probabilidad final viene dada por

$$P\left\{t_{n-1;1-\frac{\alpha}{2}} \leq \frac{\overline{x}-\mu}{s/\sqrt{n}} \leq t_{n-1;\frac{\alpha}{2}}\middle|\overrightarrow{x}\right\} = 1-\alpha$$

ya que al ser la distribución de T simétrica basta hacer que los cuantiles de la distribución dejen colas iguales a $\frac{\alpha}{2}$. Es decir que

$$P\left\{\overline{x}-\frac{S}{\sqrt{n}}t_{n-1;\frac{\alpha}{2}} \leq \mu < \overline{x}+\frac{S}{\sqrt{n}}t_{n-1;1-\frac{\alpha}{2}}\middle|\overrightarrow{x}\right\} = 1-\alpha$$

luego la región creíble de más alta probabilidad $1-\alpha$ es

$$\left((\overline{x}-\frac{S}{\sqrt{n}}t_{n-1;\frac{\alpha}{2}}, \overline{x}+\frac{S}{\sqrt{n}}t_{n-1;1-\frac{\alpha}{2}})\right)$$

Se observa que numéricamente coincide con la región de grado de confianza $1-\alpha$ obtenida a partir de (4.3.2), pero su interpretación es bien distinta. Aquí la probabilidad final de que μ esté contenida entre los dos extremos es igual a $1-\alpha$, en el caso clásico son ambos extremos las v.a. y fijado μ estos extremos aleatorios contienen a μ con probabilidad $1-\alpha$, por eso cuando se obtiene una muestra particular ya no se puede hablar de probabilidad y se dice que es $1-\alpha$ el grado de confianza con que μ está entre ambos extremos.

b) Para la obtención de $\pi(\sigma^2|\overrightarrow{x})$ se parte de $\pi(\mu,\sigma^2|\overrightarrow{x})$, si se integra respecto de μ, se obtiene

$$\pi(\sigma^2|\overrightarrow{x}) = \frac{\int_{-\infty}^{+\infty}(\sigma^2)^{-\frac{n}{2}-1}e^{-\frac{1}{\sigma^2}\frac{\left[(m-1)S^2+n(\overline{x}-\mu)^2\right]}{2}}d\mu}{\frac{1}{\sqrt{n}}\left(\frac{1}{2}\right)^{-\frac{n}{2}}\left[(n-1)S^2\right]^{-\frac{n-1}{2}}\Gamma(\frac{1}{2})\Gamma(\frac{n-1}{2})}$$

pero el numerador puede ponerse como

$$(\sigma^1)^{-\frac{n}{2}-1}e^{-\frac{(n-1)S^2}{2\sigma^2}}\int_{-\infty}^{+\infty}e^{-\frac{1}{2\sigma^2}n(E-\mu)^2}d\mu$$

y usando de nuevo que la integral vale $\sqrt{\frac{2\pi}{n}}\sigma$ se obtiene, tras simplificar, que

$$\pi(\sigma^2|\overrightarrow{x}) = \frac{(\sigma^2)^{-\frac{n}{2}-1}e^{-\frac{(n-1)S^2}{2\sigma^2}}\left(\frac{1}{2}\right)^{-\frac{1}{2}}(\sigma^2)^{\frac{1}{2}}}{\left(\frac{1}{2}\right)^{-\frac{n}{2}}\left[(n-1)S^2\right]^{-\frac{n-1}{2}}\Gamma(\frac{n-1}{2})}$$

con lo que mediante el cambio $y = (n-1)\frac{S^2}{\sigma^2}$ se obtiene que

$$\pi(y^2|\overrightarrow{x}) = \left(\frac{1}{2}\right)^{\frac{n-1}{2}} \frac{1}{\Gamma\left(\frac{n-1}{2}\right)} e^{-\frac{1}{2}y} y^{\frac{n-1}{2}-1}, \quad y > 0$$

con lo que efectivamente $(n-1)\frac{S^2}{\sigma^2}|\overrightarrow{x} \sim \chi^2_{n-1}$ y se obtiene el resultado deseado.

Se puede calcular la región creíble de probabilidad $1-\alpha$ a partir de

$$P\left\{\chi^2_{n-1;1-\frac{\alpha}{2}} \leq (n-1)\frac{S^2}{\sigma^2} \leq \chi^2_{n-1;\frac{\alpha}{2}}|\overrightarrow{x}\right\} = 1-\alpha,$$

ahora ya no es de más alta probabilidad y sí su aproximación con colas iguales, con lo que se obtiene que el intervalo creíble de probabilidad $1-\alpha$ viene dado por

$$\left(\frac{(n-1)S^2}{\chi^2_{n-1;\frac{\alpha}{2}}}, \frac{(n-1)S^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}}\right)$$

que coincide numéricamente con el obtenido en el caso clásico a partir de la Expresión (4.3.3) pero con una interpretación radicalmente diferente; de manera análoga a lo ocurrido para la media μ en la parte a).

Capítulo 5

CONTRASTES DE HIPÓTESIS

5.1. INTRODUCCIÓN

El problema general de la estimación ha sido eliminar la incertidumbre acerca del parámetro que caracteriza a la familia de distribuciones de probabilidad.

Una forma de hacerlo ha sido tomar los estimadores por punto y sustituir el valor desconocido por una función de la muestra; otra segunda ha consistido en la utilización de regiones de confianza o regiones creíbles en las que estará el valor desconocido con probabilidad alta, bien sea interpretada ésta como grado de confianza o bien como probabilidad final.

La manera de proceder en este capítulo va a ser debilitar la petición del problema, tratando de dilucidar si dada la muestra se puede afirmar que el valor de θ pertenece al conjunto paramétrico Θ_0 o bien al conjunto paramétrico Θ_1. Es decir, que ahora se pretende discriminar entre lo que se denomina la hipótesis nula $H_0 : \theta \in \Theta_0$ y la hipótesis alternativa $H_1 : \theta \in \Theta_1$.

Como se ha hecho en los capítulos anteriores, se estudiarán tanto la aproximación clásica como la bayesiana, según se pueda considerar el valor θ como parámetro o como variable aleatoria, dicho de otra manera, según se considere la evaluación del procedimiento que se utiliza bien frente a la distribución en el muestreo, o condicionado a la muestra que se ha observado. Es importante señalar que según la aproximación utilizada, se va a producir una segunda diferencia fundamental; en la aproximación clásica se va a introducir una asimetría entre la hipótesis nula y la alternativa que no existe en la aproximación bayesiana. En este sentido se interpreta, dentro de la primera aproximación, que el contraste estadístico lo que pretende es ver si la evidencia obtenida al observar la muestra es favorable a la hipótesis nula, o por el contrario hay que dirigirse hacia la hipótesis alternativa.

Por esta razón se debe colocar como hipótesis nula, aquella sobre la que se tenga inicialmente un mayor convencimiento de que va a ser cierta, pues como se verá por la forma de construir los contrastes, desde el punto de vista clásico, se actúa tratando de proteger a la hipótesis nula. Esta importante consideración no ha de ser tenida en cuenta en los contrastes bayesianos, ya que en esta última aproximación se consideran ambas hipótesis de manera simétrica.

Situaciones a las que se puede aplicar el modelo del contraste de hipótesis, en sus dos aproximaciones pueden ser: ¿de los resultados de las tiradas de una moneda puede concluirse que ésta es equilibrada?, ¿es μ_0 la talla media de una población?, ¿tiene la vitamina C una repercusión positiva en la curación de la gripe?, ¿la utilización del cinturón de seguridad rebaja la proporción de accidentes mortales?, ¿una central atómica hace aumentar los niveles de radiación en la zona instalada? y un largo etc.

Para explicitar el modelo a emplear se suponen m.a.s. de tamaño n de una distribución de la familia $\mathcal{P} = \{f(x_1, ..., x_n|\theta)|\theta \in \Theta\}$ donde se emplea notación de función de densidad.

Definición 5.1.1 (contraste estadístico)
Es cualquier partición del espacio muestral $\mathcal{X}$ en dos conjuntos RA y RC de tal manera que si el punto muestral $\overrightarrow{x} = (x_1, ..., x_n)$ pertenece a RA se dice que se acepta la hipótesis nula, es decir, se admite $H_0 : \theta \in \Theta_0$, y si $\overrightarrow{x} \in RC$ se dice que se rechaza la hipótesis nula, o lo que es lo mismo, se admite $H_1 : \theta \in \Theta_1$. A RA se le denomina región de aceptación y a RC región crítica.
■

Con frecuencia el contraste de hipótesis es especificado en función de un estadístico $T(x_1, ..., x_n)$ que resume la información muestral, así si se quiere, para una población $N(\theta, 1)$, contrastar la hipótesis nula $H_0 : \theta = 0$ frente a la hipótesis alternativa $H_1 : \theta = 3$, se puede emplear como región crítica $RC = \{\overline{X} \geq 0'392\}$, es decir, se rechaza la hipótesis nula para aquellos puntos muestrales para los que la media muestral iguale o supere al valor $0'392$. En lugar de utilizar toda la muestra se emplea la media muestral.

Algunos autores, por una restricción de tipo matemático que se introducirá más adelante, llaman a los contrastes estadísticos que aquí se han definido contrastes no aleatorizados y admiten la existencia de contrastes aleatorizados como aquellos para los que en determinados puntos se realiza un sorteo para decidir entre la aceptación o el rechazo de la hipótesis nula. En este texto no se van a considerar los contrastes aleatorizados ya que desde el punto de vista práctico no son realistas.

Se llama hipótesis simple, bien sea la nula o la alternativa, a aquella que consta de un único elemento, en otro caso se habla de hipótesis compuesta. Cuando el espacio paramétrico sea toda la recta real, es decir $\Theta = \mathbb{R}$, se

llaman hipótesis unilaterales a aquellas en las que se contrasta un semiintervalo frente a su complementario y se llama hipótesis bilaterales a aquellas en las que se contrasta un intervalo, propio o no frente a su complementario.

Así, si se está trabajando con una población $N(\theta, \sigma_0)$ con σ_0 conocida, es un caso de hipótesis nula simple frente a alternativa simple, el contrastar la hipótesis nula $H_0 : \theta = \theta_0$ frente a $H_1 : \theta = \theta_1$. Un caso de hipótesis nula unilateral frente a alternativa unilateral es el contrastar la hipótesis nula $H_0 : \theta \leq \theta_0$ frente a $H_1 : \theta > \theta_0$. Es una hipótesis bilateral el contrastar $H_0 : \theta \in [a, b]$ frente a $H_1 : \theta < a$ ó $\theta > b$. Salvo en el primer caso, en todos los demás se está trabajando con hipótesis compuestas.

Mediante la utilización de un contraste se pueden cometer dos tipos de errores: rechazar H_0 cuando $\theta \in \Theta_0$, llamado *error de tipo I* o aceptar H_0 cuando $\theta \in \Theta_1$ llamado *error de tipo II*. Por lo tanto lo que puede ocurrir con un test de hipótesis queda recogido en la Tabla 5.1.1.

	$\overrightarrow{x} \in RC$	$\overrightarrow{x} \in RA$
$\theta \in \Theta_0$	error de tipo I	correcto
$\theta \in \Theta_1$	correcto	error de tipo II

Tabla 5.1.1. Consecuencias de un test de hipótesis.

Será deseable construir contrastes de manera que ambos tipos de error sean pequeños. Si se quiere disminuir la probabilidad de cometer error de tipo I se deben llevar puntos de la region crítica a la región de aceptación pero entonces aumenta la probabilidad de cometer error de tipo II, de acuerdo con la Tabla 5.1.1. La solución es llegar a un compromiso entre las probabilidades de error de ambos tipos que se quieran asumir, para ello se definen los siguientes conceptos.

Definición 5.1.2 (nivel de significación y tamaño)
Al contrastar $H_0 : \theta \in \Theta_0$ frente a $H_1 : \theta \in \Theta_1$ mediante el test de región crítica RC y región de aceptación RA, se llama nivel de significación a cualquier número $\alpha \in (0, 1)$ tal que $P(RC|\theta) \leq \alpha \quad \forall \theta \in \Theta_0$.

Se llama tamaño al menor de los niveles de significación, es decir al valor α tal que $\sup_{\theta \in \Theta_0} P(RC|\theta) = \alpha$.

■

Lo que se pretende es emplear tests con nivel de significación α pequeño ya que este valor es la cota de los errores de tipo I.

Definición 5.1.3 (función de potencia)
Al contrastar $H_0 : \theta \in \Theta_0$ frente a $H_1 : \theta \in \Theta_1$ mediante el test de región crítica RC y región de aceptación RA, se llama función de potencia a $\beta(\theta) = P(RC|\theta)$ cuando θ recorre tanto la hipótesis nula como la alternativa.

Se observa que cuando $\theta \in \Theta_0$ la función de potencia es la probabilidad bajo θ de cometer *error de tipo I* y cuando $\theta \in \Theta_1$ la potencia es 1 menos la probabilidad de cometer *error de tipo II*. Por lo tanto interesa obtener contrastes que tengan función de potencia pequeña cuando $\theta \in \Theta_0$ y grande cuando $\theta \in \Theta_1$, de esta manera las probabilidades de cometer ambos tipos de error serán pequeñas.

Ejemplo 5.1.1
Si para una población $N(\mu, \sigma = 1)$, se quiere contrastar la hipótesis nula $H_0 : \mu = 0$ frente a la hipótesis alternativa $H_1 : \mu = 3$, con m.a.s. de tamaño $n = 25$, el contraste de región crítica $RC = \{\overline{X} \geq 0'392\}$ tiene de tamaño $0'025$, ya que

$$\begin{aligned} P\{\overline{X} \geq 0'392|\mu = 0\} &= P\{Z > \tfrac{0'392}{1/\sqrt{25}}\} \\ &= P\{Z \geq 1'96\} \\ &= 0'025 \end{aligned}$$

la función de potencia es

$$\begin{aligned} \beta(\mu) &= P\{\overline{X} \geq 0'392|\mu\} \\ &= P\{Z \geq \tfrac{0'392-\mu}{1/\sqrt{25}}\} \\ &= 1 - \Phi(1'96 - 5\mu) \end{aligned}$$

con Z distribuida $N(0, 1)$. Así cuando $\mu = 3$ la función de potencia es $\Phi(13'04)$ prácticamente igual a 1.
■

Para evitar que al tomar tamaños α diferentes, con la misma evidencia muestral, en unos casos se rechace la hipótesis nula y en otros se acepte, se introduce el concepto de *p-valor*. Con poca precisión pueden decirse que el *p-valor* es la probabilidad del suceso cola que ha sido observado.

Definición 5.1.4 (p-valor)
Si para contrastar $H_0 : \theta \in \Theta_0$ frente a $H_1 : \theta \in \Theta_1$ el test tiene de región crítica $RC = \{\overrightarrow{x}|T(\overrightarrow{x}) \geq c\}$ para T un estadístico conveniente y se observa, la muestra $\overrightarrow{x} = (x_1, ..., x_n)$, se denomina p-valor correspondiente a $\overrightarrow{x}$ al valor

$$p(x_1, ..., x_n) = \sup_{\theta \in \Theta_0} P\{T(\overrightarrow{X}) \geq T(x_1, ..., x_n)|\theta\}. \tag{5.1.1}$$

Lo primero que se observa es que el *p-valor* es un concepto que depende de la observación realizada $\overrightarrow{x}$. Además, si se emplea el test de tamaño α, con α estrictamente menor que el *p-valor* correspondiente a $\overrightarrow{x}$, este valor de $\overrightarrow{x}$ pertenece a la región de aceptación, mientras que para un valor del tamaño mayor que el *p-valor*, $\overrightarrow{x}$ pertenece a la región crítica del test. En este sentido el *p-valor* es el tamaño límite correspondiente al valor observado.

Por último, como a menor tamaño α es menos verosímil que se presente un suceso incompatible con la hipótesis nula, un *p-valor* observado pequeño se interpreta como poca evidencia a favor de H_0, y se dice que el *p-valor* es el valor de la evidencia en contra de la hipótesis nula.

Así en el Ejemplo 5.1.1 si se observa una muestra cuya media muestral es de $\overline{x} = 0'1$, el p-valor es

$$\begin{aligned} p(\overrightarrow{x}) &= P\{\overline{X} \geq 0'1|\mu = 0\} = P\{Z > 0'1 \times 5\} \\ &= P\{Z > 0'5\} = 0'0062. \end{aligned}$$

Es decir hay una fuerte evidencia de que H_0 sea falsa; todavía con un tamaño $\alpha = 0'1$ se habría rechazado H_0, mientras que con $\alpha = 0'001$ se habría aceptado H_0.

5.2. MÉTODOS DE CONSTRUCCIÓN DE CONTRASTES DE HIPÓTESIS

En uniformidad con lo hecho en la estimación por punto y en la estimación por regiones de confianza, se van a estudiar ahora diferentes procedimientos de construcción de contrastes. Así se desarrollan a continuación el método de la razón de verosimilitudes, el método bayesiano y la invariancia.

5.2.1. Contrastes de la razón de verosimilitudes

Este método de obtención de contrastes de hipótesis está intimamente ligado con el método de la máxima verosimilitud que se ha desarrollado para la construcción de estimadores.

Definición 5.2.1.1 (contraste de la razón de verosimilitudes)
Para contrastar $H_0 : \theta \in \Theta_0$ frente a $H_1 : \theta \in \Theta_1$ con m.a.s. de tamaño n el test de la razón de verosimilitudes (RV) es el que tiene de región crítica $RC = \{\overrightarrow{x} = (x_1, ..., x_n)|\lambda(\overrightarrow{x}) \leq k\}$ donde

$$\lambda(\overrightarrow{x}) = \frac{\sup\limits_{\theta\in\Theta_0} f(x_1, ..., x_n|\theta)}{\sup\limits_{\theta\in\Theta} f(x_1, ..., x_n|\theta)} \tag{5.2.1.1}$$

y k es una constante $0 \leq k \leq 1$ determinada para que $\sup\limits_{\theta\in\Theta_0} P\{RC|\theta\} \leq \alpha$, siendo α un valor fijo perteneciente al intervalo $[0, 1]$.

Esta construcción, es bastante razonable. Si, de acuerdo con la Sección 3.6.2, se denota por

$$L(\widehat{\theta}_0|\overrightarrow{x}) = \sup_{\theta\in\Theta_0} f(x_1, ..., x_n|\theta) \quad \text{y} \quad L(\widehat{\theta}|\overrightarrow{x}) = \sup_{\theta\in\Theta_0\cup\Theta_1} f(x_1, ..., x_n|\theta)$$

se tiene que $\widehat{\theta}_0$ y $\widehat{\theta}$ son los estimadores de máxima verosimilitud, el primero bajo la hipótesis nula y el segundo sobre todo el espacio paramétrico. Por lo tanto aquellos $\overrightarrow{x}$ para los que $\lambda(\overrightarrow{x})$ sea pequeño se puede decir que dan poca evidencia de que sea cierta la hipótesis nula; por el contrario aquellos puntos muestrales $\overrightarrow{x}$ para los que $\lambda(\overrightarrow{x})$ sea grande dan evidencia de que la hipótesis nula sea cierta. Esto justifica la construcción de la región crítica contenida en la Expresión (5.2.1.1).

Ejemplo 5.2.1.1
Para contrastar $H_0 : \theta \leq \theta_0$ frente a $H_p : \theta > \theta_0$ con m.a.s. de tamaño n de una población de Bernoulli(θ), mediante el test de la RV se tiene que

$$\lambda(\overrightarrow{x}) = \frac{\sup\limits_{\theta \leq \theta_0} \theta^{\sum\limits_1^n x_i}(1-\theta)^{n-\sum\limits_1^n x_i}}{\sup\limits_{\theta \in [0,1]} \theta^{\sum\limits_1^n x_i}(1-\theta)^{n-\sum\limits_1^n x_i}}$$

El supremo del denominador se alcanza en $\widehat{\theta} = \overline{x}$ y para el numerador si $\overline{x} \leq \theta_0$ es $\widehat{\theta}_0 = \overline{x}$, caso contrario al ser el numerador una función creciente en θ, el supremo se alcanza en θ_0, con lo que se tiene

$$\lambda(\overrightarrow{x}) = \begin{cases} 1 & \text{si} \quad \overline{x} \leq \theta_0 \\ \left(\frac{\theta_0}{\overline{x}}\right)^{\sum\limits_{i=1}^n x_i} \left(\frac{1-\theta_0}{1-\overline{x}}\right)^{n-\sum\limits_{i=1}^n x_i} & \text{si} \quad \overline{x} \geq \theta_0 \end{cases}$$

como la función $\lambda(\overrightarrow{x})$ es una función decreciente de $\overline{x}$, la región crítica viene dada por $\{\overline{x} | \overline{x} \geq c\}$. Dado α, para la determinación de la constante puede utilizarse que $\sum\limits_{i=1}^n X_i$ se distribuye $Binomial(n, \theta)$, por lo que basta hacer que $P\{\overline{X} \geq c | \theta_0\} \leq \alpha$.

Ejemplo 5.2.1.2
Para contrastar $H_0 : \theta = \theta_0$ frente a $H_1 : \theta \neq \theta_0$ con m.a.s. de tamaño n de una población $N(\theta, 1)$ mediante el test de la RV

$$\begin{aligned} \lambda(\overrightarrow{x}) &= \frac{\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum\limits_1^n (x_i-\theta_0)^2}}{\sup\limits_{\theta} \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum\limits_1^n (x_i-\theta)^2}} \\ &= e^{-\frac{n}{2}(\overline{x}-\theta_0)^2} \end{aligned}$$

Esta expresión es simétrica en $\overline{x}$ alrededor de θ_0 y alcanza su máximo en θ_0 donde λ vale 1, por lo que se puede poner como región crítica $\{\overline{x} | |\overline{x} - \theta_0| \geq c\}$. La obtención de la constante se sigue de que $\overline{X} \sim N(\theta_0, \frac{1}{\sqrt{n}})$.
■

En ambos ejemplos se observa que la región crítica es siempre función de la muestra a través del estadístico suficiente. Esto es un resultado general, ya que por el Teorema 2.2.1 (factorización) (véase la Sección 2.2), tanto $L(\widehat{\theta}_0|\overrightarrow{x})$ como $L(\widehat{\theta}|\overrightarrow{x})$ son funciones de dicho estadístico.

El método de la RV da lugar a tests estadísticos con buenas propiedades en casi todos los casos, como se verá más adelante.

5.2.2. Contrastes de la razón de verosimilitudes en el caso normal

Debido al papel central que ocupa la distribución normal en la inferencia estadística, se van a desarrollar con algún detalle, los tests de hipótesis a los que se llega en el caso normal. Siempre que sea posible se obtendrán estos tests mediante la RV, aunque en algunos casos se va a recurrir a aproximaciones, con el fin de facilitar la obtención de las constantes que determinan la región crítica del test.

Contrastes relativos a la media:

Para m.a.s. de tamaño n de una población $N(\theta, \sigma)$, pueden resumirse los distintos contrastes para la media en la Tabla 5.2.2.1

		Región Crítica	
H_0	H_1	σ conocida	σ desconocida
$\mu \leq \mu_0$	$\mu > \mu_0$	$\overline{X} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha$	$\overline{X} \geq \mu_0 + \frac{S}{\sqrt{n}} t_{n-1;\alpha}$
$\mu \geq \mu_0$	$\mu < \mu_0$	$\overline{X} \leq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha}$	$\overline{X} \leq \mu_0 + \frac{S}{\sqrt{n}} t_{n-1;1-\alpha}$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\lvert\overline{X} - \mu_0\rvert \geq \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}}$	$\lvert\overline{X} - \mu_0\rvert \geq \frac{S}{\sqrt{n}} t_{n-1;\frac{\alpha}{2}}$

Tabla 5.2.2.1. Contrastes relativos a la media de la distribución normal, con z y t cuantiles de las distribuciones normal y de Student respectivamente.

Ejemplo 5.2.2.1
Para la obtención del primer contraste con varianza desconocida

$$\lambda(\overrightarrow{x}) = \frac{\sup\limits_{\substack{\sigma^2 \geq 0 \\ \mu \leq \mu_0}} \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum\limits_1^n (x_i - \mu)^2}}{\sup\limits_{\substack{\sigma^2 \geq 0 \\ \mu \in \mathbb{R}}} \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum\limits_1^n (x_i - \mu)^2}}$$

es fácil ver que los estimadores de MV del denominador, véase la Sección 3.6.2, son

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_1^n (x_i - \overline{x})^2 \qquad \widehat{\mu} = \overline{x}$$

y que los estimadores del numerador son

$$\widehat{\sigma}_0^2 = \frac{1}{n}\sum_1^n (x_i - \widehat{\mu}_0)^2 \qquad \widehat{\mu}_0 = \begin{cases} \overline{x} & \text{si} \quad \overline{x} \leq \mu_0 \\ \mu_0 & \text{si} \quad \overline{x} \geq \mu_0 \end{cases}$$

por lo que la razón de verosimilitudes es

$$\lambda(\overrightarrow{x}) = \begin{cases} 1 & \text{si} \quad \overline{x} \leq \mu_0 \\ \left(\dfrac{\sum_1^n (x_i - \overline{x})^2}{\sum_1^n (x_i - \mu_0)^2} \right)^{n/2} & \text{si} \quad \overline{x} > \mu_0 \end{cases}$$

de forma que la razón de verosimilitudes es una función no creciente de $\overline{x}$, cuya representación viene dada en la Figura 5.2.2.1.

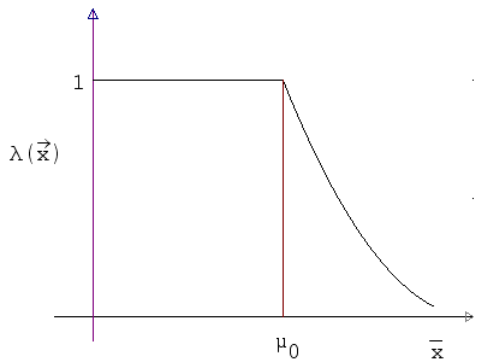


Figura 5.2.2.1. $\lambda(\overrightarrow{x})$ *como función de* $\overrightarrow{x}$*, para el Ejemplo 5.2.2.1.*

La región crítica $RC = \{\lambda(\overrightarrow{x}) \leq k\}$ puede escribirse alternativamente como $RC = \{\overline{x} \geq c\}$. Para la obtención de la constante c se debe suponer que

$$P\{\overline{X} \geq c | \mu, \sigma^2\} \leq \alpha \qquad \forall \mu \leq \mu_0 \quad \forall \sigma^2 > 0. \tag{5.2.2.1}$$

Para ello se determina c de manera que $P\{\overline{X} \geq c | \mu_0, \sigma^2\} = \alpha$, empleando que $\frac{\overline{X} - \mu_0}{S}\sqrt{n} \sim T_{n-1}$ ha de ser

$$P\left\{ \frac{\overline{X} - \mu_0}{S}\sqrt{n} \geq \frac{c - \mu_0}{S}\sqrt{n} \middle| \mu_0, \sigma^2 \right\} = P\{T_{n-1} \geq \frac{c - \mu_0}{S}\sqrt{n}\}$$
$$= \alpha$$

de donde $c = \mu_0 + \frac{S}{\sqrt{n}} t_{n-1;\alpha}$. Para comprobar ahora que con esta constante se verifica (5.5.2.1), se observa que

$$
\begin{aligned}
& P\{\overline{X} \geq \mu_0 + \tfrac{S}{\sqrt{n}} t_{n-1;\alpha} | \mu, \sigma^2\} = \\
& = P\{\tfrac{\overline{X}-\mu}{S}\sqrt{n} \geq \tfrac{\mu_0 - \mu + \frac{S}{\sqrt{n}} t_{n-1;\alpha}}{S/\sqrt{n}} | \mu, \sigma^2\}
\end{aligned}
\tag{5.2.2.2}
$$

y como $\frac{\overline{X}-\mu}{S}\sqrt{n} \sim T_{n-1}$, se sigue que (5.2.2.2) coincide con

$$
P\left\{ T_{n-1} \geq \frac{\sqrt{n}}{S}(\mu_0 - \mu) + t_{n-1;\alpha} \right\}
$$

expresión menor o igual que α siempre que sea $\mu_0 - \mu \geq 0$.
■

El resto de los contrastes de la Tabla 5.2.2.1 se obtienen de forma análoga, salvo el último, que se corresponde con la aproximación de región crítica con colas iguales para facilitar la obtención de las constantes, véase el Ejercicio 5.7.8.

Contrastes relativos a la varianza:

Para m.a.s. de tamaño n de una población $N(\mu, \sigma)$, pueden resumirse los distintos contrastes para la varianza en la Tabla 5.2.2.2.

		Región Crítica	
H_0	H_1	μ conocida	μ desconocida
$\sigma^2 \geq \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\sum_1^n (x_i - \mu)^2 \leq \chi^2_{n;1-\alpha}\sigma_0^2$	$S^2 \leq \chi^2_{n-1;1-\alpha}\frac{\sigma_0^2}{n-1}$
$\sigma^2 \leq \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\sum_1^n (x_i - \mu)^2 \geq \chi^2_{n;\alpha}\sigma_0^2$	$S^2 \geq \chi^2_{n-1;\,\alpha}\frac{\sigma_0^2}{n-1}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\sum_1^n (x_i - \mu)^2 \leq \chi^2_{n;1-\frac{\alpha}{2}}\sigma_0^2$	$S^2 \leq \chi^2_{n-1;1-\frac{\alpha}{2}}\frac{\sigma_0^2}{n-1}$
		o $\sum_1^n (x_i - \mu)^2 \geq \chi^2_{n;\frac{\alpha}{2}}\sigma_0^2$	o $S^2 \geq \chi^2_{n-1;\frac{\alpha}{2}}\frac{\sigma_0^2}{n-1}$

Tabla 5.2.2.2. Contrastes relativos a la varianza de la distribución normal donde $\chi^2_{n;\beta}$ es el cuantil de orden $1-\beta$ de la distribución χ^2_n.

Ejemplo 5.2.2.2
Para la obtención del último contraste, el relativo a la hipótesis puntual con media conocida, la RV es

$$\lambda(\overrightarrow{x}) = \frac{\left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^n e^{-\frac{1}{2\sigma_0^2}\sum_1^n (x_i-\mu_0)^2}}{\sup_\sigma \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum_1^n (x_i-\mu_0)^2}}$$

el estimador de MV para el denominador es

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_1^n (x_i - \mu_0)^2$$

por lo que la RV puede ponerse como

$$\lambda(\overrightarrow{x}) = \left(\frac{\widehat{\sigma}^2}{\sigma_0^2}\right)^{n/2} e^{-\frac{n}{2}\left(\frac{\widehat{\sigma}^2}{\sigma_0^2}-1\right)}$$

cuya representación puede verse en la Figura 5.2.2.2

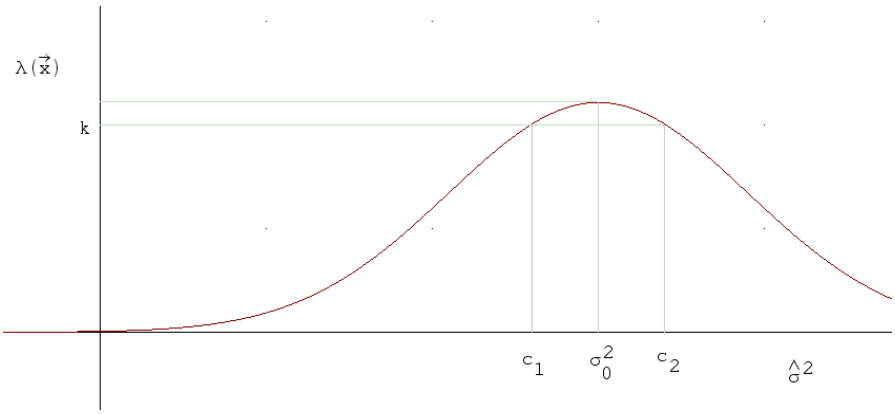


Figura 5.2.2.2. $\lambda(\overrightarrow{x})$ como función de $\widehat{\sigma}^2$, para el Ejemplo 5.2.2.2.

La región crítica $RC = \{\lambda(\overrightarrow{x}) \leq k\}$ puede ponerse alternativamente como $RC = \{\widehat{\sigma}^2 \leq c_1 \text{ ó } \widehat{\sigma}^2 \geq c_2\}$, donde las constantes han de verificar, además de la restricción del tamaño α del test, que

$$k = \left(\frac{c_1}{\sigma_0^2}\right)^{\frac{n}{2}} e^{-\frac{n}{2}\left(\frac{c_1}{\sigma_0^2}-1\right)}$$
$$= \left(\frac{c_2}{\sigma_0^2}\right)^{\frac{n}{2}} e^{-\frac{n}{2}\left(\frac{c_2}{\sigma_0^2}-1\right)}$$

o lo que es lo mismo ha de ser $\ln\frac{c_1}{c_2} = \frac{c_1-c_2}{\sigma_0^2}$.

Para evitar resolver esta ecuación junto a la restricción de tamaño α, se utiliza la aproximación con probabilidad de cola iguales, es decir se determinan las constantes c_1 y c_2 para que

$$\frac{\alpha}{2} = P\{\widehat{\sigma}^2 \leq c_1 | \sigma_0^2\} \qquad \frac{\alpha}{2} = P\{\widehat{\sigma}^2 \geq c_2 | \sigma_0^2\}$$

lo que puede hacerse teniendo en cuenta que $n\frac{\widehat{\sigma}^2}{\sigma_0^2} = \sum_1^n \frac{(x_i-\mu_0)^2}{\sigma_0^2} \sim \chi_n^2$, de donde

$$\begin{aligned} \frac{\alpha}{2} &= P\left\{ n\frac{\widehat{\sigma}^2}{\sigma_0^2} \leq n\frac{c_1}{\sigma_0^2} \middle| \sigma_0^2 \right\} \\ &= P\left\{ \chi_n^2 \leq n\frac{c_1}{\sigma_0^2} \middle| \sigma_0^2 \right\} \end{aligned}$$

por lo que $n\frac{c_1}{\sigma_0^2} = \chi^2_{n;1-\frac{\alpha}{2}}$. Esto hace que la región crítica sea

$$RC = \{\sum_1^n (x_i-\mu)^2 \leq \chi^2_{n;1-\frac{\alpha}{2}}\sigma_0^2 \quad \text{o bien} \quad \sum_1^n (x_i-\mu)^2 \geq \chi^2_{n;\frac{\alpha}{2}}\sigma_0^2\}.$$

Contrastes relativos a la diferencia de medias:

Para una m.a.s. de tamaño m de una población $N(\mu_1, \sigma_1)$ e independiente de otra m.a.s. de tamaño n de una población $N(\mu_2, \sigma_2)$, pueden resumirse los distintos contrastes para la diferencia de medias $\mu_1 - \mu_2$ en la siguiente Tabla.

		Región Crítica	
H_0	H_1	σ_1^2, σ_2^2 conocidas	$\sigma_1^2 = \sigma_2^2$ desconocidas
$\mu_1 - \mu_2 \leq \delta$	$\mu_1 - \mu_2 > \delta$	$\overline{x} - \overline{y} - \delta \geq$ $\geq z_\alpha\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}$	$\overline{x} - \overline{y} - \delta \geq$ $\geq t_{m+n-2;\alpha} S_p \sqrt{\frac{1}{m} + \frac{1}{n}}$
$\mu_1 - \mu_2 \geq \delta$	$\mu_1 - \mu_2 < \delta$	$\overline{x} - \overline{y} - \delta \leq$ $\leq -z_\alpha\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}$	$\overline{x} - \overline{y} - \delta \leq$ $\leq -t_{m+n-2;\alpha} S_p \sqrt{\frac{1}{m} + \frac{1}{n}}$
$\mu_1 - \mu_2 = \delta$	$\mu_1 - \mu_2 \neq \delta$	$\lvert\overline{x} - \overline{y} - \delta\rvert \geq$ $\geq z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}$	$\lvert\overline{x} - \overline{y} - \delta\rvert \geq$ $\geq t_{m+n-2;\frac{\alpha}{2}} S_p \sqrt{\frac{1}{m} + \frac{1}{n}}$

Tabla 5.2.2.3. Contrastes relativos a la diferencia de medias de poblaciones normales, con $z_\alpha, t_{\nu;\alpha}$ cuantiles de las distribuciones $N(0,1)$ y T_ν respectivamente.

Los cálculos para la obtención de la Tabla 5.2.2.3 son análogos a los realizados en las tablas anteriores.

Para la determinación de las constantes se utiliza la distribución del siguiente estadístico

(5.2.2.3)

$$\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}} \Bigg/ \sqrt{\left[(m-1)\frac{S_1^2}{\sigma_1^2}+(n-1)\frac{S_2^2}{\sigma_2^2}\right]\frac{1}{(m+n-2)}} \sim T_{m+n-2}$$

ya que $\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}} \sim N(0,1)$ y $(m-1)\frac{S_1^2}{\sigma_1^2}+(n-1)\frac{S_2^2}{\sigma_2^2} \sim \chi^2_{m+n-2}$.

Si las varianzas son conocidas se pueden obtener las constantes de los tests de la RV, si las varianzas son desconocidas pero iguales se tiene, véase la Expresión (4.3.6), que poniendo $\sigma_1^2=\sigma_2^2=\sigma^2$ este valor desconocido desaparece al dividir numerador y denominador de la Expresión (5.2.2.3). Además el denominador toma la forma de la raíz cuadrada de

(5.2.2.4)
$$S_p^2=\frac{(m-1)S_1^2+(n-1)S_2^2}{m+n-2}$$

que es una poderación de los estimadores de la varianza con cada una de las muestras.

Contrastes relativos al cociente de varianzas:

Como se ha visto en el problema de las muestras de poblaciones normales, es de capital importancia el saber si $\sigma_1^2/\sigma_2^2=1$, los tests de la RV en este marco pueden verse en la Tabla 5.2.2.4

		Región Crítica	
H_0	H_1	μ_1,μ_2 conocidas	μ_1,μ_2 desconocidas
$\frac{\sigma_1^2}{\sigma_2^2}\le 1$	$\frac{\sigma_1^2}{\sigma_2^2}>1$	$\frac{\frac{1}{m}\sum_1^m(x_i-\mu_1)^2}{\frac{1}{n}\sum_1^n(y_i-\mu_2)^2}\ge f_{m,n;\alpha}$	$\frac{S_1^2}{S_2^2}\ge f_{m-1,n-1;\alpha}$
$\frac{\sigma_1^2}{\sigma_2^2}\ge 1$	$\frac{\sigma_1^2}{\sigma_2^2}<1$	$\frac{\frac{1}{n}\sum_1^n(y_i-\mu_2)^2}{\frac{1}{m}\sum_1^n(x_i-\mu_1)^2}\ge f_{n,m;\alpha}$	$\frac{S_2^2}{S_1^2}\ge f_{n-1,m-1;\alpha}$
$\frac{\sigma_1^2}{\sigma_2^2}=1$	$\frac{\sigma_1^2}{\sigma_2^2}\ne 1$	$\frac{\frac{1}{m}\sum_1^m(x_i-\mu_1)^2}{\frac{1}{n}\sum_1^n(y_i-\mu_2)^2}\ge f_{m,n;\frac{\alpha}{2}}$	$\frac{S_1^2}{S_2^2}\ge f_{m-1,n-1;\frac{\alpha}{2}}$ si $\frac{S_1^2}{S_2^2}\ge 1$
		o bien	o bien
		$\frac{\frac{1}{n}\sum_1^n(y_i-\mu_2)^2}{\frac{1}{m}\sum_1^m(x_i-\mu_1)^2}\ge f_{n,m;\frac{\alpha}{2}}$	$\frac{S_2^2}{S_1^2}\ge f_{n-1,m-1;\frac{\alpha}{2}}$ si $\frac{S_1^2}{S_2^2}<1$

Tabla 5.2.2.4. Contrastes relativos al cociente de varianzas de poblaciones normales, con $f_{\nu_1\nu_2;\beta}$ cuantil de la distribución F de Snedecor.

El test correspondiente a la hipótesis nula puntual, es la aproximación con colas iguales al test de la RV.

Ejemplo 5.2.2.3
Para contrastar $H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$ frente a $H_1 : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$ con m.a.s. de tamaño m y n de dos poblaciones $N(\mu_1, \sigma_1)$ y $N(\mu_2, \sigma_2)$ e independientes, el test de la RV viene dado en función de

$$\lambda(\overrightarrow{x}, \overrightarrow{y}) = \frac{\sup\limits_{\mu_1,\mu_1,\sigma} \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^m e^{-\frac{1}{2\sigma^2}\sum\limits_1^n (x_i-\mu_1)^2} \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum\limits_1^n (y_i-\mu_2)^2}}{\sup\limits_{\mu_1,\mu_2,\sigma_1,\sigma_2} \left(\frac{1}{\sqrt{2\pi}\sigma_1}\right)^m e^{-\frac{1}{2\sigma_1^2}\sum\limits_1^m (x_i-\mu_1)^2} \left(\frac{1}{\sqrt{2\pi}\sigma_2}\right)^n e^{-\frac{1}{2\sigma_2^2}\sum\limits_1^m (y_i-\mu_2)^2}}$$

Los estimadores de MV para el denominador son:

$$\widehat{\mu}_1 = \overline{x} \qquad \widehat{\mu}_2 = \overline{y} \qquad \widehat{\sigma}_1^2 = \frac{1}{m}\sum_1^m (x_i - \overline{x})^2 \qquad \widehat{\sigma}_2^2 = \frac{1}{n}\sum_1^n (y_i - \overline{y})^2,$$

de manera análoga los estimadores para el numerador son:

$$\widehat{\mu}_1^0 = \overline{x} \qquad \widehat{\mu}_2^0 = \overline{y} \qquad \widehat{\sigma}_0^2 = \frac{m\widehat{\sigma}_1^2 + n\widehat{\sigma}_2^2}{m+n},$$

por lo que la RV es

$$\begin{aligned}\lambda(\overrightarrow{x}, \overrightarrow{y}) &= \frac{\left(\frac{1}{\widehat{\sigma}_0^2}\right)^{\frac{m+n}{2}} e^{-\frac{1}{2\widehat{\sigma}_0^2}(m\widehat{\sigma}_1^2 + n\widehat{\sigma}_2^2)}}{\left(\frac{1}{\widehat{\sigma}_1^2}\right)^{\frac{m}{2}} e^{-\frac{m}{2}} \left(\frac{1}{\widehat{\sigma}_1^2}\right)^{\frac{n}{2}} e^{-\frac{n}{2}}} \\ &= (n+m)^{\frac{n+m}{2}} \left(\frac{\widehat{\sigma}_1^2}{\widehat{\sigma}_2^2}\right)^{\frac{n}{2}} \left(m\frac{\widehat{\sigma}_1^2}{\widehat{\sigma}_2^2} + n\right)^{-\frac{m+n}{2}}.\end{aligned}$$

De acuerdo con la representación de la Figura 5.2.2.3, la región crítica

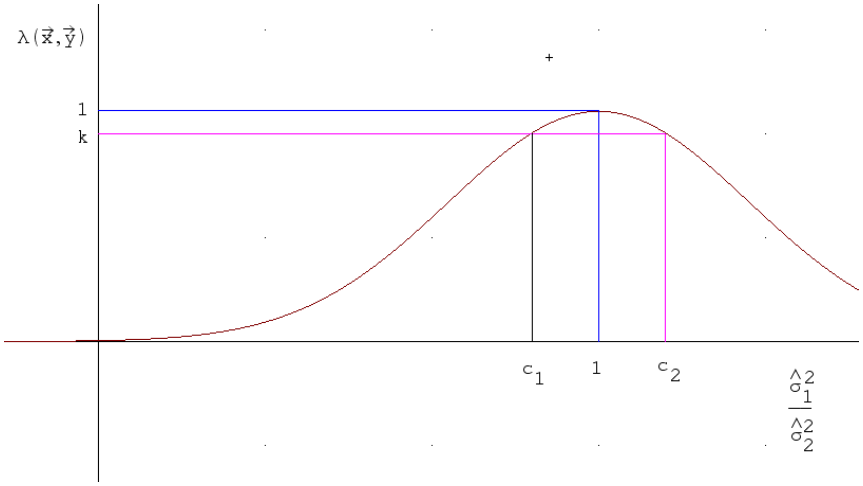


Figura 5.2.2.3. $\lambda(\overrightarrow{x}, \overrightarrow{y})$ *como función de* $\widehat{\sigma}_1^2/\widehat{\sigma}_2^2$, *para el Ejemplo 5.2.2.3.*

del test de la RV es $RC = \{\widehat{\sigma}_1^2/\widehat{\sigma}_2^2 \leq c_1 \text{ o } \widehat{\sigma}_1^2/\widehat{\sigma}_2^2 \geq c_2\}$ donde las constantes han de verificar la ecuación

$$c_1^{\frac{m}{2}}(mc_1+n)^{-\frac{m+n}{2}} = c_2^{\frac{m}{2}}(mc_2+n)^{-\frac{m+n}{2}}$$

y la restricción $P\{RC|\sigma_1^2/\sigma_2^2 = 1\} = \alpha$.

Para facilitar los cálculos, se impone

$$P\{\widehat{\sigma}_1^2/\widehat{\sigma}_2^2 \leq c_1|H_0\} = \frac{\alpha}{2} \qquad P\{\widehat{\sigma}_1^2/\widehat{\sigma}_2^2 \geq c_2|H_0\} = \frac{\alpha}{2}$$

lo que lleva, puesto que $(m-1)\frac{S_1^2}{\sigma_1^2} \sim \chi_{m-1}^2$ y $(n-1)\frac{S_2^2}{\sigma_2^2} \sim \chi_{n-1}^2$, a las constantes de la Tabla 5.2.2.4, donde la condición de comprobar primeramente si es $S_1^2 \geq S_2^2$ se introduce para facilitar la obtención, a partir de las tablas, de los cuantiles de la distribución de Snedecor.

■

El método de la razón de verosimilitudes, como el resto de los que se van a estudiar, tiene la propiedad de proporcionar tests que poseen buenas propiedades en contextos suficientemente generales, no obstante, como con casi todos los procedimientos estadíticos, siempre pueden encontrarse situaciones en las que da lugar a contrastes que no tienen un comportamiento adecuado, cómo pone de manifiesto el siguiente ejemplo.

Ejemplo 5.2.2.4

Se pretende contrastar la hipótesis nula de que una v.a. discreta se distribuye con una función de masa

$$P\{x\} = \begin{cases} \alpha/2 & \text{si} \quad x = \pm 2 \\ \frac{1-2\alpha}{2} & \text{si} \quad x = \pm 1 \\ \alpha & \text{si} \quad x = 0 \end{cases}$$

frente a la hipótesis alternativa de que es algún miembro de la familia de distribuciones

$$P\{x|\theta\} = \begin{cases} \theta c & \text{si} \quad x = -2 \\ \frac{1-c}{1-\alpha}\frac{1-2\alpha}{2} & \text{si} \quad x = \pm 1 \\ \alpha\frac{1-c}{1-\alpha} & \text{si} \quad x = 0 \\ (1-\theta)c & \text{si} \quad x = 2 \end{cases}$$

cuando $\theta \in [0,1]$, donde las constantes cumplen $0 < \alpha < \frac{1}{2}$ y $\frac{\alpha}{2-\alpha} < c < \alpha$.

La razón de verosimilitud en los distintos puntos muestrales vale

$$\lambda(2) = \frac{\frac{\alpha}{2}}{\sup\limits_{\theta\in[0,1]}\{\frac{\alpha}{2},(1-\theta)c\}} = \frac{\frac{\alpha}{2}}{\text{máx}\{\frac{\alpha}{2},c\}} = \frac{\alpha}{2c}$$
$$\lambda(-2) = \frac{\frac{\alpha}{2}}{\sup\limits_{\theta\in[0,1]}\{\frac{\alpha}{2},\theta c\}} = \frac{\frac{\alpha}{2}}{\text{máx}\{\frac{\alpha}{2},c\}} = \frac{\alpha}{2c}$$
$$\lambda(1) = \lambda(-1) = \frac{1-\alpha}{1-c}$$
$$\lambda(0) = \frac{1-\alpha}{1-c},$$

por lo que la RV queda representada en la Figura 5.2.2.4

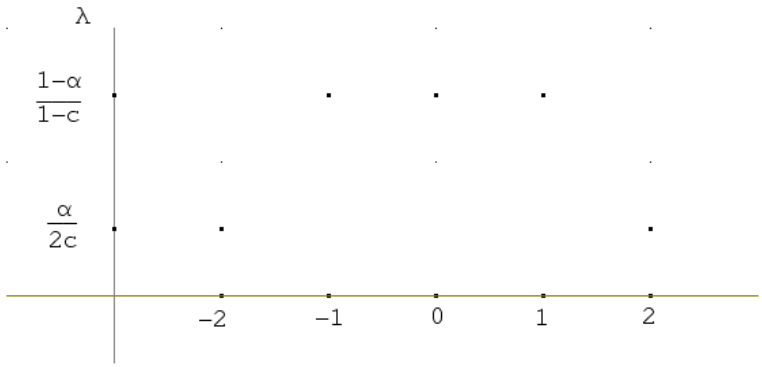


Figura 5.2.2.4. $\lambda(x)$ para el Ejemplo 5.2.2.4.

y por lo tanto la $RC = \{x|\lambda(x) \leq k\} = \{-2, 2\}$ si $k \in \left(\frac{\alpha}{2c}, \frac{1-\alpha}{1-c}\right)$. Este test tiene de tamaño $P\{-2, 2|H_0\} = \alpha$ y de potencia $P\{-2, 2|\theta\} = c < \alpha$, es decir el test que se obtiene por el método de la RV para la restricción de k tomada, tiene peor potencia que el que siempre rechaza con probabilidad α independientemente del valor muestral observado x.

5.2.3. Distribución asintótica de la razón de verosimilitudes

Hasta ahora se ha visto que la distribución de la razón de verosimilitudes o de una función de ella, tiene distribución conocida cuando la población se distribuye $N(\mu, \sigma)$. Como se ha dicho el método de la razón de verosimilitudes está pensado para poder construir contrastes de hipótesis en situaciones

bastantes generales, por lo que será necesario conocer su distribución bajo la hipótesis nula con el fin de poder determinar la constante.

Se va a demostrar, que bajo determinadas hipótesis, la distribución asintótica de $-2\ln\lambda(X_1,...,X_n)$ es una distribución χ^2_ν, en la que el número ν de grados de libertad es la diferencia entre la dimensión de todo el espacio paramétrico Θ y la dimensión del espacio paramétrico Θ_0, obtenido de Θ cuando la hipótesis nula se supone cierta.

Teorema 5.2.3.1 (distribución asintótica de la RV)
Para constrastar $H_0 : \theta = \theta_0$ frente a $H_1 : \theta \neq \theta_0$ cuando $\theta \in \Theta$ es un subconjunto de $\mathbb{R}$, si se denota por

$$\lambda_n(x_1,...,x_n) = \frac{f(x_1,...,x_n|\theta_0)}{\underset{\theta\in\Theta}{\max} f(x_1,...,x_n|\theta)} = \frac{L(\theta_0|x_1,...,x_n)}{L(\widehat{\theta}|x_1,...,x_n)} \tag{5.2.3.1}$$

y se cumplen las siguientes hipótesis:

H1) En un intervalo abierto A contenido en Θ existe $\frac{\partial^3}{\partial\theta^3}\ln f(x|\theta)$ con $|\frac{\partial^3}{\partial\theta^3}\ln f(x|\theta)| < M(x)$ y $E_\theta M(X) < K$ para todo $\theta \in A$.

H2) En A se tiene $E_{\theta_0}\left[\frac{\partial}{\partial\theta}\ln f(X|\theta_0)\right] = 0$, $E_{\theta_0}\left[\frac{1}{f(X|\theta_0)}\frac{\partial^2}{\partial\theta^2}f(X|\theta_0)\right] = 0$ y $0 < I_1(\theta_0) < +\infty$.

Entonces se tiene la siguiente convergencia en ley

$$-2\ln\lambda_n(x_1,...,x_n) \overset{\mathcal{L}_{\theta_0}}{\to} \chi^2_1.$$

Demostración: Recordando que H1) y H2) son las condiciones introducidas para que exista una sucesión consistente de raíces $\widehat{\theta}_n$ de la ecuación de verosimilitud, véase la Sección 3.6.3.

$$\sum_1^n \frac{\partial}{\partial\theta}\ln f(x_i|\theta) = 0$$

se cumple que

$$\widehat{\theta}_n \overset{c.s.}{\to} \theta_0 \quad \text{y} \quad \sqrt{n}(\widehat{\theta}_n - \theta_0) \overset{\mathcal{L}_{\theta_0}}{\to} Z \in N\left(0, \frac{1}{\sqrt{I_1(\theta_0)}}\right)$$

Por otro lado con $\overrightarrow{x} = (x_1,...,x_n)$ se tiene

$$-2\ln\lambda_n(\overrightarrow{x}) = -2[\ln f(\overrightarrow{x}|\theta_0) - \ln f(\overrightarrow{x}|\widehat{\theta}_n)]$$

por lo tanto si se desarrolla $l(\theta_0|\overrightarrow{x}) = \ln f(\overrightarrow{x}|\theta_0)$ alrededor de $\widehat{\theta}_n$ se obtiene

$$l(\theta_0|\overrightarrow{x}) = l(\widehat{\theta}_n|\overrightarrow{x}) + (\theta_0 - \widehat{\theta}_n)l'(\widehat{\theta}_n|\overrightarrow{x}) + \frac{(\theta_0 - \widehat{\theta}_n)^2}{2}l"(\theta^*_n|\overrightarrow{x})$$

donde $|\widehat{\theta}_n - \theta_0| > |\widehat{\theta}_n - \theta^*_n|$ y por tanto cuando $\widehat{\theta}_n$ converge en probabilidad a θ_0 también θ^*_n convergerá a θ_0.

Como $l'(\widehat{\theta}_n|\overrightarrow{x}) = 0$

$$\begin{aligned} -2\ln\lambda(\overrightarrow{x}) &= -2\left[\frac{(\theta_0-\widehat{\theta}_n)^2}{2}l"(\theta_n^*|\overrightarrow{x})\right] \\ &= -\left[\left(\sqrt{nI_1(\theta_n)}(\theta_0-\widehat{\theta}_n)\right)^2\frac{l"(\theta_n^*|\overrightarrow{x})}{nI_1(\theta_0)}\right] \end{aligned}$$

y al tenerse que

$$\sqrt{nI_n(\theta_0)}(\theta_0-\widehat{\theta}_n) \overset{\mathcal{L}_{\theta_0}}{\to} Z \sim N(0,1)$$

y que por la ley de los grandes números

$$\frac{1}{n}l"(\theta_n^*, \overrightarrow{x}) = \frac{1}{n}\sum_1^n \frac{\partial^2}{\partial\theta^2}\ln f(x_i|\theta_n^*) \overset{c.s.}{\to} -I_1(\theta_0).$$

Se sigue, tras elevar al cuadrado y aplicar el Teorema 1.4.5 (Slutsky), la siguiente convergencia

$$nI_1(\theta_0)(\theta_0-\widehat{\theta}_n)^2\frac{l"(\theta_n^*;\overrightarrow{x})}{-nI_1(\theta,0)} \overset{\mathcal{L}_{\theta_0}}{\to} Z^2 \sim \chi_1^2$$

con Z v.a. con distribución $N(0,1)$, por lo tanto

$$-2\ln\lambda_n(x_1,...,x_n) \overset{\mathcal{L}_{\theta_0}}{\to} \chi_1^2.$$

Nótese que los grados de libertad de la distribución χ_ν^2 son

$$\begin{aligned} \nu &= \dim\Theta - \dim\Theta_0 \\ &= 1 \end{aligned}$$

■

El siguiente teorema generaliza este resultado al caso k-dimensional. Su demostración puede verse en Rao (1973) pág. 418.

Teorema 5.2.3.2 (distribución asintótica de la RV, caso general) Para contrastar $H_0 : \theta_1 = \theta_1(\beta_1,...,\beta_q)$,...,$\theta_k = \theta_k(\beta_1,...,\beta_q)$ frente a $H_1 : \theta_1,...,\theta_k$ con $q \leq k$ y donde todos los parámetros son reales, si las derivadas $\frac{\partial\theta_i}{\partial\beta_j}$ son funciones continuas y se verifican condiciones análogas a H1) y H2) del Teorema 5.2.3.1 para el caso k-dimensional (véase Rao (1973) pág. 391) se tiene que notando por

$$\left.\begin{array}{l} \frac{\partial}{\partial\theta_j}\ln f(x_1,...,x_n|\overrightarrow{\theta}) = \sum_{i=1}^n \frac{\partial}{\partial\theta_j}\ln f(x_i|\overrightarrow{\theta}) = 0 \\ j = 1,...,k \end{array}\right\}$$

a la ecuación de verosimilitud bajo la hipótesis alternativa; y si sus soluciones respectivas $\widehat{\theta}_0$ y $\widehat{\theta}_n$ son únicas. Entonces, se tiene que

$$-2\ln\lambda(x_1,...,x_n) \overset{\mathcal{L}_\theta}{\to} \chi^2_{k-q}.$$

Obsérvese que al igual que en el teorema anterior, los grados de libertad de la distribución χ^2 son

$$k-q = \dim\Theta - \dim\Theta_0.$$

Ejemplo 5.2.3.1 (modelos log-lineales)
Se considera un experimento aleatorio en el que las n observaciones se pueden clasificar, de forma exhaustiva, en 4 clases disjuntas de probabilidades respectivas $(\theta_1,\theta_2,\theta_3,\theta_4)$. Mediante la aplicación de la distribución multinomial, se trata de contrastar

$$H_0 : \theta_1 = \theta_2 \;\; y \;\; \theta_3 = \theta_4 \quad \text{frente a} \quad H_1 : (\theta_1,\theta_2,\theta_3,\theta_4) \text{ arbitrarios.}$$

Aplicando el test de la RV, bajo H_0 la distribución multinomial es

$$\frac{n!}{x_1!x_2!x_3!x_4!}\theta_1^{x_1+x_2}\theta_3^{x_3+x_4}$$

tomando logaritmos y derivando se obtiene $\widehat{\theta}_1^0 = (x_1+x_2)/2n$ y $\widehat{\theta}_3^0 = (x_3+x_4)/2n$. Bajo H_1 la distribución multinomial es

$$\frac{n!}{x_1!x_2!x_3!x_4!}\theta_1^{x_1}\theta_2^{x_2}\theta_3^{x_3}\theta_4^{x_4}$$

de donde $\widehat{\theta}_i = \frac{x_i}{n}$.

Entonces

$$\begin{aligned}(5.2.3.2)\qquad -2\ln\lambda(\overrightarrow{x}) &= 2\sum_{i=1}^{4} x_i \ln\frac{\widehat{\theta}_i}{\widehat{\theta}_i^0} \\ &= 2x_1\ln\left(\frac{2x_1}{x_1+x_2}\right) + 2x_2\ln\left(\frac{2x_2}{x_1+x_2}\right) \\ &+ 2x_3\ln\left(\frac{2x_3}{x_3+x_4}\right) + 2x_4\ln\left(\frac{2x_4}{x_3+x_4}\right)\end{aligned}$$

Expresión lineal en los logaritmos de los cocientes $\widehat{\theta}_i/\widehat{\theta}_i^0$, cuya distribución asintótica, de acuerdo con el Teorema 5.2.3.2 es una χ^2_2. Por lo tanto el test de la RV rechaza, cuando la Expresión (5.2.3.2) sea mayor o igual que el valor $\chi^2_{2,\alpha}$.

5.2.4. Contrastes bayesianos

Cuando se tiene información inicial, acerca del parámetro θ, en la forma de una distribución de probabilidad sobre el espacio Θ, igual que en el caso de los estimadores por punto, véase la Sección 3.6.4, se recurre a calcular la probabilidad de la hipótesis nula y de la hipótesis alternativa a partir de la distribución final. Con notación de función de densidad la distribución final o a posteriori, cuando se observa el resultado muestral $\overrightarrow{x} = (x_1, ..., x_n)$ viene dada por

$$\pi(\theta|\overrightarrow{x}) = \frac{\pi(\theta) f(\overrightarrow{x}|\theta)}{\int_\Theta \pi(\theta) f(\overrightarrow{x}|\theta) d\theta} \tag{5.2.4.1}$$

se puede hablar entonces con toda propiedad de la probabilidad final de que la hipótesis nula y la alternativa sean verdaderas;

$$P(\Theta_0|\overrightarrow{x}) = \int_{\Theta_0} \pi(\theta|\overrightarrow{x}) d\theta \qquad P(\Theta_1|\overrightarrow{x}) = \int_{\Theta_1} \pi(\theta|\overrightarrow{x}) d\theta$$

por lo tanto la región crítica es

$$RC = \{\overrightarrow{x} = (x_1, ..., x_n) | P(\theta \in \Theta_1|\overrightarrow{x}) > P(\theta \in \Theta_0|\overrightarrow{x})\}. \tag{5.2.4.2}$$

De todas formas, en este caso el énfasis se hace sobre el espacio paramétrico Θ y no sobre el espacio muestral $\mathcal{X}$, por lo que no se va a estar interesado en todos los puntos de la región crítica, sino más bien en si para un punto muestral fijo, el que haya sido observado, se acepta o se rechaza la hipótesis nula. Obsérvese además que el comportamiento es del todo simétrico, en la hipótesis nula y en la alternativa. Se calculan las probabilidades de ambas hipótesis mediante (5.2.4.1) donde $P(\Theta_0|\overrightarrow{x}) + P(\Theta_1|\overrightarrow{x}) = 1$ y se acepta aquella hipótesis cuya probabilidad final sea mayor. Así pertenecen a la región crítica, aquellos puntos para los que la probabilidad final de la hipótesis alternativa sea mayor que la probabilidad final de la hipótesis nula.

Ejemplo 5.2.4.1
Para contrastar $H_0 : \theta = \theta_0$ frente a $H_1 : \theta = \theta_1$ con $\theta_1 > \theta_0$ para m.a.s. de tamaño n de una población $X \sim N(\theta, \sigma)$ con σ conocida, si la distribución inicial viene dada por $P\{\theta = \theta_0\} = p$, puede comprobarse que la región crítica está formada por

$$RC = \left\{ \overrightarrow{x} \,\middle|\, \overline{x} \geq \frac{\theta_1 + \theta_2}{2} + \frac{\sigma^2}{n(\theta_1 - \theta_0)} \ln \frac{p}{1-p} \right\}$$

■

Aunque en el contraste de hipótesis bayesiano las medidas de interés son las probabilidades finales de las dos hipótesis, es también interesante introducir el siguiente concepto.

Definición 5.2.4.1 (factor Bayes)
Para contrastar la hipótesis nula $H_0 : \theta \in \Theta_0$ frente a la hipótesis alternativa $H_1 : \theta \in \Theta_1$ se define el *factor Bayes* en favor de H_0, mediante el cociente

$$B(\overrightarrow{x}; H_0, H_1) = \frac{P(H_0|\overrightarrow{x})/\, P(H_1|\overrightarrow{x})}{P(H_0)/\, P(H_1)} \tag{5.2.4.3}$$

La idea es, que cuanto mayor sea el factor Bayes, mayor será la probabilidad final de la hipótesis nula frente a la probabilidad final de la hipótesis alternativa, y por tanto existe mayor evidencia a favor de la hipótesis nula.

Ejemplo 5.2.4.2
Para contrastar la hipótesis nula $H_0 : \theta = \frac{1}{2}$ frente a la hipótesis alternativa $H_1 : \theta = \frac{2}{3}$ para una m.a.s. de tamaño n de una población de Bernoulli de parámeto θ el factor Bayes a favor de H_0 viene dado por

$$\begin{aligned} B(\overrightarrow{x}; H_0, H_1) &= \left(\frac{\theta_0}{\theta_1}\right)^{\sum_1^n x_i} \left(\frac{1-\theta_0}{1-\theta_1}\right)^{n-\sum_1^n x_i} \\ &= 0'5^{\sum_1^n x_i} 1'5^n \end{aligned}$$

expresión que para $n = 10$ y cuando se obtienen 2 éxitos es igual a $14'42$ cuando se obtienen 5 éxitos es $1'80$ y cuando se producen 8 éxitos es igual a $0'23$. Es decir a medida que el número de éxitos aumenta, lógicamente disminuye la evidencia a favor de la hipótesis nula.
■

Se observa además que al factor Bayes, en el caso de que ambas hipótesis sean simples, no depende de las probabilidades iniciales, y vale $B(\overrightarrow{x}; H_0, H_1) = \frac{f(\overrightarrow{x}|\theta_0)}{f(\overrightarrow{x}|\theta_1)}$.

Se pueden entonces resumir los distintos contrastes bayesianos:

Contraste de hipótesis nula simple frente a alternativa simple:

Para contrastar $H_0 : \theta = \theta_0$ frente a $H_1 : \theta = \theta_1$, la probabilidad inicial vendrá dada por el par $(p, 1-p)$, con lo que las probabilidades finales serán

$$\begin{aligned} P(H_0|\overrightarrow{x}) &= \frac{pf(\overrightarrow{x}|\theta_0)}{pf(\overrightarrow{x}|\theta_0) + (1-p)f(\overrightarrow{x}|\theta_1)} \\ P(H_1|\overrightarrow{x}) &= \frac{(1-p)f(\overrightarrow{x}|\theta_1)}{pf(\overrightarrow{x}|\theta_0) + (1-p)f(\overrightarrow{x}|\theta_1)} \end{aligned}$$

De modo que se rechaza H_0 para aquellos puntos muestrales tales que

$$RC = \{\overrightarrow{x} | P(H_0|\overrightarrow{x}) < P(H_1|\overrightarrow{x})\}$$

o alternativamente

$$RC = \left\{ \overrightarrow{x} \left| \frac{f(\overrightarrow{x}|\theta_1)}{f(\overrightarrow{x}|\theta_0)} > \frac{p}{1-p} \right. \right\}. \tag{5.2.4.4}$$

Este contraste coincide, formalmente, con el que se obtendría mediante el Lema de Neyman-Pearson, véase el Teorema 5.3.1, sólo que en este caso la constante viene determinada por el cociente entre las probabilidades iniciales de ambas hipótesis.

El sentido de formalmente, es como ya ha sido señalado, que desde el punto de vista bayesiano, el énfasis está en θ y no en los puntos muestrales.

Contraste de hipótesis nula simple frente a alternativa compuesta:

En algunas situaciones prácticas, es interesante el contrastar $H_0 : \theta = \theta_0$ frente a $H_1 : \theta \neq \theta_0$, en este caso se utiliza como distribución inicial una que asigne probabilidad p a la hipótesis nula y reparta la restante $1-p$ con densidad $\pi(\theta)$ sobre $\theta \neq \theta_0$.

Por lo tanto

$$P(H_0|\overrightarrow{x}) = \frac{pf(\overrightarrow{x}|\theta_0)}{pf(\overrightarrow{x}|\theta_0) + (1-p)\int_\Theta f(\overrightarrow{x}|\theta)\pi(\theta)d\theta}$$

$$P(H_1|\overrightarrow{x}) = \frac{(1-p)\int_\Theta f(\overrightarrow{x}|\theta)\pi(\theta)d\theta}{pf(\overrightarrow{x}|\theta_0) + (1-p)\int_\Theta f(\overrightarrow{x}|\theta)\pi(\theta)d\theta}$$

donde por propiedades de la integral $\displaystyle\int_{\{\theta \neq \theta_0\}} f(\overrightarrow{x}|\theta)\pi(\theta)d\theta = \int_\Theta f(\overrightarrow{x}|\theta)\pi(\theta)d\theta$

con lo que se rechaza la hipótesis nula cuando

$$pf(\overrightarrow{x}|\theta) < (1-p)\int_\Theta \pi(\theta)f(\overrightarrow{x}|\theta)d\theta$$

o lo que es lo mismo

$$RC = \left\{ \overrightarrow{x} \left| \frac{f(\overrightarrow{x}|\theta_0)}{\int_\Theta \pi(\theta)f(\overrightarrow{x}|\theta)d\theta} < \frac{1-p}{p} \right. \right\} \tag{5.2.4.5}$$

Ejemplo 5.2.4.3

Dada una m.a.s. de tamaño n de una población $N(\theta, \sigma)$, con σ conocida, se pretende contrastar $H_0 : \theta = 0$ frente a $H_1 : \theta \neq 0$. Si se supone que la información inicial asigna probabilidad p a la hipótesis nula y reparte la restante, $1-p$, con densidad $N(0, \sigma_0)$ con σ_0 conocida.

En este caso es

$$f(\overrightarrow{x}|\theta) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2}$$

y

$$\int_\Theta \pi(\theta)f(\overrightarrow{x}|\theta)d\theta = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma_0}e^{-\frac{1}{2}\left[\frac{\theta^2}{\sigma_0^2}\right]} \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\theta)^2} d\theta$$

de acuerdo con (5.2.4.5) y mediante la expresión (3.2.3.2), se rechaza la hipótesis nula cuando

$$\sqrt{1+n\frac{\sigma_0^2}{\sigma^2}}e^{-\frac{1}{2}n^2\frac{\overline{x}^2}{\sigma^2}\frac{\sigma_0^2}{\sigma^2+n\sigma_0^2}} < \frac{1-p}{p}. \tag{5.2.4.6}$$

Si se compara este resultado con el contraste cuando no se tenía información inicial. Entonces la región crítica era la de la forma $|\overline{x}| \geq z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ (véase la Tabla 5.2.2.1). Si n es suficientemente grande y $|\overline{x}|$ es poco mayor que $z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$, lo suficiente para que el contraste clásico rechace y la aproximación de $|\overline{x}|$ por dicho valor sea aceptable, se tiene que el primer miembro de (5.2.4.6) es aproximadamente

$$\sqrt{1+n\frac{\sigma_0^2}{\sigma^2}}e^{-\frac{1}{2}z_{\frac{\alpha}{2}}^2}$$

expresión que será mayor que $\frac{1-p}{p}$. En resumen, mediante el contraste que da el teorema de Neymann-Pearson se rechazaría H_0 mientras que con el contraste bayesiano se aceptaría. Este resultado se conoce con el nombre de paradoja de Jeffreys o paradoja de Lindley.

■

Comparando el procedimiento sin información inicial con el procedimiento con información inicial, suele ocurrir que el p-valor, que es una medida de la evidencia en contra de la hipótesis nula, véase la Definición 5.1.4, es bastante menor que la probabilidad final sobre la citada hipótesis nula.

La relación entre la aproximación clásica y bayesiana en el contraste de hipótesis nula puntual, puede verse en Gómez Villegas y Gómez Sánchez-Manzano (1992), Gómez Villegas y Sanz (1998, 2000), y en Gómez Villegas, Maín y Sanz (2002).

Hay distintos aspectos que influyen en la forma de los tests bayesianos: uno de ellos es la forma del modelo, puede verse en Gómez Villegas y Maín (1992) y en Gómez Villegas, Maín, Navarro y Sanz (2004) ; otro es la robustez, puede verse en Gómez Villegas y Portela (2004).

Una crítica que se hace a los contrastes bayesianos, es que en algunas situaciones puede ser difícil determinar la distribución inicial, una solución en ciertos contextos puede verse en Gómez Sánchez-Manzano y Gómez Villegas (1990).

5.2.5. Contrastes invariantes

En esta Sección se utilizan consideraciones de invariancia para poder construir contrastes estadísticos, las ideas están íntimamente relacionadas con los estadísticos invariantes que han sido desarrollados en la Sección 2.4.

Como se va a ver la invariancia no permite construir de manera completa los tests invariantes, sino más bien, facilita la construcción de los mismos,

en el sentido de que si algunos puntos están contenidos en la región crítica, automáticamente se sigue la pertenencia a la región crítica de otros puntos muestrales función de los mismos; es decir, realmente lo que permiten estos procedimientos es caracterizar las regiones críticas de los tests invariantes en un menor número de puntos.

Definición 5.2.5.1 (test invariante)
Para contrastar $H_0 : \theta \in \Theta_0$ frente a $H_1 : \theta \in \Theta_1$ el test de región crítica RC es invariante, por un grupo de transformaciones $\mathcal{G} = \{g : \mathcal{X} \to \mathcal{X}\}$ del espacio muestral $\mathcal{X}$ en sí mismo, cuando el grupo de transformaciones deja invariante la hipótesis nula, la hipótesis alternativa y la región crítica del test. Es decir para cualquier $g \in \mathcal{G}$ la función $\overline{g}$ (véase la Definición 2.4.2) cumple $\overline{g}(\theta) \in \Theta_0$ si $\theta \in \Theta_0$ además $\overline{g}(\theta) \in \Theta_1$ si $\theta \in \Theta_1$ y $g(\overrightarrow{x}) \in RC$ si $\overrightarrow{x} \in RC$.

Ejemplo 5.2.5.1
Para m.a.s. de tamaño 10 de una población $X \sim Bernoulli(\theta)$, el test de región crítica $RC = \{0, 1, 9, 10\}$ es invariante para contrastar $H_0 : \theta = \frac{1}{2}$ frente a $H_1 : \theta \neq \frac{1}{2}$, mediante el grupo de transformaciones $\mathcal{G} = \{g : \mathcal{Y} \to \mathcal{Y}, g_1(y) = n - y \;\; g_2(y) = y\}$. Es claro que $\overline{g}_1(\theta) = 1 - \theta$ y $\overline{g}_2(\theta) = \theta$ por lo que al ser $\overline{g}_1(\frac{1}{2}) = \frac{1}{2}$ y $\overline{g}_2(\theta) = \theta$ dejan invariantes tanto a la hipótesis nula como a la alternativa. Por último si $y \in RC$ $g(y) = n - y$ también pertenece a RC, como puede comprobarse fácilmente.
■

Se observa que en este ejemplo no se ha utilizado el espacio muestral inicial $\mathcal{X}$ sino su simplificación mediante suficiencia $\mathcal{Y}$, el número de éxitos con n repeticiones, siempre que sea posible se empleará este mecanismo.

Ejemplo 5.2.5.2
En una población $X \sim N(\mu, \sigma)$ con ambos parámetros desconocidos, se quiere contrastar $H_0 : \mu \leq 0$ frente a $H_1 : \mu > 0$. Se van a emplear los estadísticos suficientes $\overline{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$ y $S^2 \sim \frac{\sigma^2}{n-1}\chi^2_{n-1}$. Si como grupo de transformaciones se utiliza el de escala, es decir $\mathcal{G} = \{g_c(\overline{x}, s^2) = (c\overline{x}, c^2 s^2) | c > 0\}$ resulta que $\overline{g}_c(\mu, \sigma^2) = (c\mu, c^2\sigma^2)$ es la transformación en los parámetros correspondientes a la transformación sobre los estadísticos suficientes. Esta transformación deja invariante $H_0 : \mu \leq 0$, ya que $c\mu \leq 0$ por ser c positiva.

$$c\overline{X} \sim N(c\mu, \frac{c\sigma}{\sqrt{n}}) \qquad c^2 S^2 \sim \frac{c^2\sigma^2}{n-1}\chi^2_{n-1},$$

por lo tanto el test de región crítica $RC = \{(\overline{x}, s^2) | \overline{x} \geq \frac{s}{\sqrt{n}} t_{n-1;\alpha}\}$ es invariante, ya que $c\overline{x} \geq c\frac{s}{\sqrt{n}} t_{n-1;\alpha}$ es equivalente, con $c > 0$, a $\overline{x} \geq \frac{s}{\sqrt{n}} t_{n-1;\alpha}$.
■

Como se ha visto en ambos ejemplos, la invariancia no es propiamente una manera de construir tests, pero sí permite una simplificación, en el primer ejemplo si se sabe qué hacer cuando se observan los valores 0 y 1 , automáticamente se sabe qué hacer cuando se observan los valores 10 y 9. Análogamente en el caso normal el rechazo y la aceptación de la hipótesis nula es solo función del cociente $\frac{\overline{x}}{s}$, en el sentido de que puntos muestrales que lleven al mismo valor del cociente, dan lugar a la misma norma de acción, de rechazo o aceptación de la hipótesis nula.

5.3. CRITERIOS DE COMPARACIÓN DE CONTRASTES

Como ya se ha dicho los contrastes que se buscan desde el punto de vista clásico, son aquellos que, manteniendo la probabilidad de error de tipo I menor o igual que un valor α, hacen mínima la probabilidad de cometer error de tipo II. Mediante el punto de vista bayesiano no hay más que un único contraste que es el que elige la hipótesis que más probabilidad final presenta. Por tanto todo lo que se diga en esta sección solo tiene aplicación para la aproximación clásica.

Definición 5.3.1 (contraste óptimo)
Si para contrastar $H_0 : \theta \in \Theta_0$ frente a $H_1 : \theta \in \Theta_1$ se considera la clase de contrastes $\mathcal{C}$, un contraste de esta clase dado por su región crítica RC se dice óptimo o uniformemente de máxima potencia, UMP, si para cualquier otro test de región crítica RC' de la misma clase $\mathcal{C}$ se tiene que $P\{RC|\theta\} \geq P\{RC'|\theta\}$ para todo $\theta \in \Theta_1$.
■

Es decir, el contraste óptimo de la clase $\mathcal{C}$ es el que tiene mayor función de potencia cuando θ se mueve en Θ_1. Normalmente la clase $\mathcal{C}$ será la de todos los contrastes que tienen de nivel α, es decir todos los que tienen acotada la probabilidad de cometer error de tipo I por α. Desafortunadamente en algunos problemas no existirá un test óptimo si no se restringe la clase $\mathcal{C}$ con la que se trabaja. En tales casos un método de continuar la búsqueda de un buen test va a ser considerar una subclase de la de los tests de nivel α, es la misma táctica que se utilizó para con los estimadores por punto en el Capítulo 3. En esta línea se introduce la siguiente definición.

Definición 5.3.2 (contraste insesgado)
Para contrastar $H_0 : \theta \in \Theta_0$ frente a $H_1 : \theta \in \Theta_1$ un test de región crítica RC se dice insesgado cuando $P\{RC|\theta\} \leq P\{RC|\theta'\}$ para todo $\theta \in \Theta_0$ y para todo $\theta' \in \Theta_1$. Es decir un contraste es insesgado si la probabilidad de rechazar la hipótesis nula cuando es falsa es mayor que cuando es verdadera.

Ejemplo 5.3.1
Para constrastar $H_0 : \mu = \mu_0$ frente a $H_1 : \mu = \mu_1$ con $\mu_1 > \mu_0$ con m.a.s. de tamaño n de una población $N(\mu, \sigma)$ con σ desconocida, el test de región crítica

$$RC = \left\{ \overrightarrow{x} \middle| \overline{x} \geq \mu_0 + t_{n-1;\alpha} \frac{s}{\sqrt{n}} \right\}$$

tiene de tamaño α y es insesgado, ya que para $\mu_1 > \mu_0$ es

$$P\left\{ \overline{X} \geq \mu_0 + t_{n-1;\alpha} \frac{S}{\sqrt{n}} \middle| \mu_1 \right\} = P\left\{ \frac{\overline{X} - \mu_1}{S} \sqrt{n} \geq \frac{\mu_0 - \mu_1}{S} \sqrt{n} + t_{n-1;\alpha} \middle| \mu_1 \right\}$$
$$= P\left\{ T_{n-1} \geq \frac{\mu_0 - \mu_1}{S} \sqrt{n} + t_{n-1\,\alpha} \right\} \geq \alpha$$

puesto que $\frac{\mu_0 - \mu_1}{S} \sqrt{n}$ es un valor negativo.

Teorema 5.3.1 (Neyman-Pearson)
Para contrastar la hipótesis nula simple $H_0 : \theta = \theta_0$ frente a $H_1 : \theta = \theta_1$, el test de región crítica

$$RC = \{ \overrightarrow{x} | f(\overrightarrow{x}|\theta_1) \geq k f(\overrightarrow{x}|\theta_0) \} \tag{5.3.1}$$

y de región de aceptación

$$RA = \{ \overrightarrow{x} | f(\overrightarrow{x}|\theta_1) < k f(\overrightarrow{x}|\theta_0) \} \tag{5.3.2}$$

con $k \geq 0$, que además tenga de tamaño α

$$\alpha = P\{RC|\theta_0\} \tag{5.3.3}$$

cumple que:

1) Es el test UMP en la clase de los de nivel α.

2) Si existe un test con región crítica RC^* y UMP en la clase de los de nivel α, ha de tener tamaño α es decir $\alpha = P\{RC^*|\theta_0\}$, además su región crítica ha de ser de la forma (5.3.1) y su región de aceptación ha de ser como en (5.3.2), salvo conjuntos de probabilidad nula bajo ambas hipótesis.

Demostración: 1) Sea RC' la región crítica de un test de nivel α, mediante las descomposiciones en conjuntos disjuntos

$$RC = (RC \cap RC') \cup (RC \cap RC'^c)$$
$$RC' = (RC' \cap RC) \cup (RC' \cap RC^c)$$

se sigue que

$$P\{RC|\theta_1\} - P\{RC'|\theta_1\} = P\{RC \cap RC'^c|\theta_1\} - P\{RC' \cap RC^c|\theta_1\} \tag{5.3.4}$$

por la construcción de RC y RA el segundo miembro de (5.3.4) cumple que

$$P\{RC \cap RC'^c|\theta_1\} - P\{RC' \cap RC^c|\theta_1\} \geq \geq k \int_{RC\cap RC'^c} f(\overrightarrow{x}|\theta_0)d\overrightarrow{x} - k \int_{RC'\cap RC^c} f(\overrightarrow{x}|\theta_0)d\overrightarrow{x} \tag{5.3.5}$$

donde se ha empleado la notación de función de densidad.

Si se vuelve a usar la igualdad (5.3.4), pero ahora en θ_0 se tiene

$$k\left[P\{RC \cap RC'^c|\theta_0\} - P\{RC' \cap RC^c|\theta_0\}\right] = k\left[P\{RC|\theta_0\} - P\{RC'|\theta_0\}\right] \tag{5.3.6}$$

por (5.3.3) y por ser el test de región crítica RC' de nivel α, el segundo miembro de (5.3.6) es no negativo, luego

$$P\{RC|\theta_1\} - P\{RC'|\theta_1\} \geq 0$$

y por lo tanto el test definido por (5.3.1) y (5.3.2) es UMP.

2) Si ahora se repite desde la expresión (5.3.4) hasta la (5.3.6) pero sustituyendo RC' por RC^* se obtiene

$$P\{RC|\theta_1\} - P\{RC^*|\theta_1\} \geq k\left[P\{RC|\theta_0\} - P\{RC^*|\theta_0\}\right] \tag{5.3.7}$$

por lo que ha de ser $P\{RC^*|\theta_0\} = \alpha$ o el test de región crítica RC^* no sería óptimo. Además de tener que ser

$$P\{RC|\theta_0\} = P\{RC^*|\theta_0\} \qquad \text{y} \qquad P\{RC|\theta_1\} = P\{RC^*|\theta_1\}$$

se sigue que RC^* ha de ser de la forma (5.3.1) y RA^* de la forma (5.3.2) salvo conjuntos de probabilidad nula bajo ambas hipótesis, ya que en caso contrario $P\{RC|\theta_0\} - P\{RC^*|\theta_0\}$ o $P\{RC|\theta_1\} - P\{RC^*|\theta_0\}$ serían estrictamente positivas; con lo que RC^* no se correspondería con un test óptimo. ■

De la demostración se deduce que los puntos para los que $f(\overrightarrow{x}|\theta_1) = = kf(\overrightarrow{x}|\theta_0)$ pueden ser colocados tanto en la región crítica como en la de aceptación, véase el paso de la desigualdad en la Expresión (5.3.5), pues estos puntos no contribuyen a aumentar el valor de la función de potencia.

Es importante señalar que el Teorema de Neyman-Pearson no dice que el test de la forma dada en las Expresiones (5.3.1) y (5.3.2) deba tener de tamaño α. Lo que puede ocurrir queda recogido en el siguiente enunciado.

Ejemplo 5.3.2

Para contrastar con m.a.s. de tamaño n de una población $X \sim Bernoulli(\theta)$ la hipótesis nula $H_0 : \theta = \theta_0$ frente a la alternativa $H_1 : \theta = \theta_1$ con $\theta_1 < \theta_0$ y ambos valores conocidos, si existe test óptimo, de acuerdo con (5.3.1), la región crítica ha de ser de la forma

$$RC = \left\{\frac{p(\overrightarrow{x}|\theta_1)}{p(\overrightarrow{x}|\theta_0)} \geq k\right\} = \left\{\overrightarrow{x} \;\middle|\; \left(\frac{\theta_1}{\theta_0}\frac{1-\theta_0}{1-\theta_1}\right)^{\sum_1^n x_i} \left(\frac{1-\theta_1}{1-\theta_0}\right)^n \geq k\right\}$$

que al ser $\frac{\theta_1}{\theta_0}\frac{1-\theta_0}{1-\theta_1} < 1$ puede ponerse $RC = \{\overrightarrow{x} | \sum_{i=1}^{n} x_i \leq c\}$ para un c conveniente. Las distintas regiones críticas que dan lugar a posibles tests óptimos se obtienen al ir variando c en el conjunto $\{0, 1, ..., n\}$, por lo que para todo valor $\alpha \in [0, 1]$ no será posible encontrar un valor de c de tal manera que $P\{RC|\theta_0\} = \alpha$. Por ejemplo si $n = 10, \theta_0 = 0'05$ y $\alpha = 0'05$ para $c = 0$ se obtienen $RC = \left\{\sum_{1}^{10} x_i \leq 0\right\}$ $P\{RC|0'05\} = 0'001$, para $c = 1$ con $RC = \left\{\sum_{1}^{10} x_i \leq 1\right\}$ es de la forma (5.3.1) y cumple que $P\{RC|0'05\} = 0'018$ luego es UMP en la clase de los de nivel $0'0108$.

Si a pesar de todo se quiere obtener el test óptimo de nivel $0'05$, puede recurrirse al artificio de construir el test de región crítica

$$RC = \left\{\sum_{1}^{10} x_i < 2 \text{ y el punto 2 con probabilidad } \gamma\right\}$$

y de región de aceptación

$$RA = \left\{\sum_{1}^{10} x_i > 2 \text{ y el punto 2 con probabilidad } 1 - \gamma\right\}.$$

Se observa que si $\gamma = 0'893$ entonces $P\{RC|0'5\} = 0'0108 + \gamma 0'0439 = 0'05$ con lo que el test tiene el tamaño deseado. Sin embargo la utilidad práctica de este test es más que dudosa; ya que si al realizar la observación se obtienen exactamente 2 éxitos debería efectuarse un sorteo a continuación que tuviera de probabilidad de rechazar H_0 $0'893$ y de probabilidad de aceptar H_0 $1 - 0'893 = 0'107$, este sorteo es independiente del valor de θ de la distribución de Bernoulli lo que no parece estar justificado.

De todas maneras puede admitirse, que en la situación del ejemplo es prácticamente igual actuar con una probabilidad de *error de tipo I* de $0'05$ o de $0'0547$, y con este segundo tamaño se evita el utilizar tests aleatorizados.

Ejemplo 5.3.3

Mediante el teorema de Neyman-Pearson y para muestras de tamaño uno de una población con función de densidad

$$f(x|\theta) = \frac{2}{\theta^2}(\theta - x)I_{(0,\theta)}(x) \quad \text{con} \quad \theta > 0$$

el test óptimo de tamaño α para contrastar $H_0: \theta = \theta_1$ frente a $H_1: \theta = \theta_1$ con $0 < \theta_1 < \theta_0$ será de la forma $RC = \left\{x \middle| \frac{f(x|\theta_1)}{f(x|\theta_0)} \geq k\right\}$, si k es compatible con que su tamaño sea α. De acuerdo con la Figura 5.3.1 la región crítica de los tests óptimos viene dada por $RC = \{x \leq c\}$ para algún valor de c. La determinación de la constante puede hacerse a partir de la expresión

$$\alpha = P\{X \leq c|\theta_0\} = \frac{2}{\theta_0}c - \frac{c^2}{\theta_0^2}$$

que siempre tiene solución en c cuando α recorre el intervalo $[0,1]$; $c = \theta_0 \pm \theta_0\sqrt{1-\alpha}$.

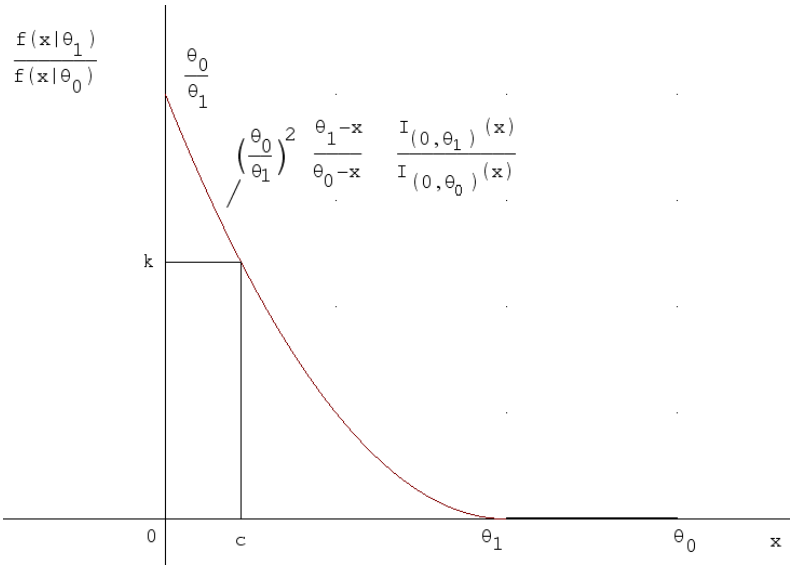


Figura 5.3.1. Cociente de verosimilitudes para el Ejemplo 5.3.3.

De donde se sigue que o bien el test rechaza siempre, y tiene de tamaño 1, o bien tiene de región crítica $RC = \{x \leq (1-\sqrt{1-\alpha})\theta_0\}$.

Ejemplo 5.3.4
Para m.a.s. de tamaño n de una población $N(\theta, \sigma)$ con σ conocida, el test óptimo para contrastar $H_0 : \theta = \theta_0$ frente a $H_1 : \theta = \theta_1$ con $\theta_1 > \theta_0$, viene dado por

$$RC = \{ f(\overrightarrow{x}|\theta_1)/ f(\overrightarrow{x}|\theta_0) \geq k\}$$

y si se tiene en cuenta que

$$\frac{f(\overrightarrow{x}|\theta_1)}{f(\overrightarrow{x}|\theta_0)} = e^{-\frac{n}{2\sigma^2}(\theta_1^2-\theta_0^2)} e^{\frac{\theta_1-\theta_0}{\sigma^2}\sum_1^n x_i}$$

se puede poner $RC = \{\overline{x} \geq c\}$. Además se obtiene el mismo contraste independientemente del valor θ_1 con tal que sea $\theta_1 > \theta_0$. Para la determinación de la constante se emplea que bajo θ_0, $\overline{X} \sim N(\theta_0, \frac{\sigma}{\sqrt{n}})$, de donde de acuerdo con el teorema de Neyman-Pearson el test óptimo tiene de región crítica $RC = \{\overline{X} \geq \theta_0 + z_\alpha \frac{\sigma}{\sqrt{n}}\}$.
■

Surge la pregunta de si lo ocurrido en este último ejemplo en el que se obtiene el mismo test mediante el teorema de Neyman-Pearson y la razón de verosimilitud ocurre siempre. Para comprender mejor la situación se considera el siguiente enunciado.

Ejemplo 5.3.5
Se toman muestras de tamaño 1 de una población $N(\theta, 1)$ y se quiere contrastar la hipótesis nula $H_0 : \theta = 0$ frente a $H_1 : \theta = 1$.

La razón de verosimilitud viene dada por

$$\lambda(x) = \begin{cases} 1 & \text{si} \quad x \leq \frac{1}{2} \\ e^{\frac{1}{2}-x} & \text{si} \quad x \geq \frac{1}{2} \end{cases}$$

con lo que la región crítica del test de la RV está formada por todos los puntos muestrales, si la constante que determina la región crítica vale 1, o bien la región crítica es $x \geq c$, si la constante c es mayor o igual que $0'5$. En cualquier caso, mediante el método de la RV sólo pueden obtenerse constantes de tamaño menor o igual que $P\{X \geq 0'5|\theta = 0\} = 0'30854$.

Por otro lado, mediante el teorema de Neyman-Pearson la región crítica es de la forma

$$RC = \{x|\ f(x|\theta = 1)/\ f(x|\theta = 0)\ \geq k\}$$

lo que lleva a que $RC = \{x|e^{-\frac{1}{2}+x} \geq k\}$ o bien $RC = \{x \geq c\}$. Por tanto mediante el Teorema de Neyman-Pearson se puede conseguir cualquier tamaño.

■

Se está en condiciones de probar la siguiente relación.

Proposición 5.3.1
Si para contrastar $H_0 : \theta = \theta_0$ frente a $H_1 : \theta = \theta_1$ existe un contraste construido a partir de la razón de verosimilitud que tiene el mismo tamaño que el test dado por el teorema de Neyman-Pearson, ambos coinciden.

Demostración: Dado α sea $RC_1 = \{\overrightarrow{x}|f(\overrightarrow{x}|\theta_1)/f(\overrightarrow{x}|\theta_0) \geq k_1\}$ la región crítica obtenida por el teorema de Neyman-Pearson con $P\{RC_1|\theta_0\} = \alpha$. Se puede poner $RC_1 = \{f(\overrightarrow{x}|\theta_0)/f(\overrightarrow{x}|\theta_1) \leq k_1'\}$ con $k_1' = k_1^{-1}$.

Por otro lado la razón de verosimilitud es

$$\lambda(\overrightarrow{x}) = \begin{cases} 1 & \text{si} \quad f(\overrightarrow{x}|\theta_0) \geq f(\overrightarrow{x}|\theta_1) \\ \frac{f(\overrightarrow{x}|\theta_0)}{f(\overrightarrow{x}|\theta_1)} & \text{si} \quad f(\overrightarrow{x}|\theta_0) \leq f(\overrightarrow{x}|\theta_1) \end{cases}$$

luego la región crítica del test de la RV es $RC = \{\overrightarrow{x}|f(\overrightarrow{x}|\theta_0)/f(\overrightarrow{x}|\theta_1) \leq k\}$. Si $k_2 < 1$, al ser el test de tamaño α, ha de ser $k_1' = k_2$. Si es $k_2 = 1$, la región crítica del test de la RV incluye a todos los puntos y $\alpha = 1$ con lo que los dos contrastes rechazan siempre y por tanto coinciden.

5.4. CONTRASTES DE HIPÓTESIS UNILATERALES Y BILATERALES

Como ya se había anticipado y si se contemplan los contrastes de hipótesis desde el punto de vista de la distribución en el muestreo, para la obtención de tests óptimos en el caso de hipótesis más complicadas, se necesita restringir la clase de los tests con los que se va a trabajar, para luego pasar a restringir la clase de distribuciones que sigue el modelo con el que se obtiene la muestra.

Como se ha visto en la sección anterior, la clase $\mathcal{C}$ de los contrastes a emplear, para hipótesis simples, era la de los tests de nivel α. Para encontrar contrastes óptimos en el caso de hipótesis unilaterales es necesario restringirse a modelos de distibuciones que tengan cociente de verosimilitud monótono, concepto que será definido a continuación. Por último, para el uso de hipótesis bilaterales se necesitan además restricciones a clases de contrastes que sean insesgados.

Este tipo de problemas no se presentan en la aproximación bayesiana, la hipótesis nula o la alternativa es aceptada en función de cuál sea su probabilidad final.

Definición 5.4.1. (distribuciones con cociente de verosimilitud monótono) La familia de funciones de densidad de la muestra $\mathcal{P} = \{f(\overrightarrow{x}|\theta)|\theta \in \Theta \subset \mathbb{R}\}$ tiene cociente de verosimilitud monótono (CVM), en un estadístico unidimensional $T(\overrightarrow{x})$, si cuando $\theta_1 < \theta_2$ el cociente $f(\overrightarrow{x}|\theta_2)/f(\overrightarrow{x}|\theta_1)$ es una función no decreciente de $T(\overrightarrow{x})$.
■

En esta definición se ha empleado notación de función de densidad, lo mismo es válido para funciones de masa. Los desarrollos posteriores también son válidos, con modificaciones obvias, si el cociente de verosimilitudes es una función no creciente de $T(\overrightarrow{x})$.

Ejemplo 5.4.1
Para una población $N(\theta, \sigma)$ con σ conocida y m.a.s. de tamaño n, la familia $\mathcal{P} = \left\{ \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum_1^n (x_i-\theta)^2} \,\middle|\, \theta \in \mathbb{R} \right\}$ tiene CVM en $T(\overrightarrow{x}) = \sum_{i=1}^n x_i$, ya que para $\theta_1 < \theta_2$ es

$$\frac{f(x_1, ..., x_n|\theta_2)}{f(x_1, ..., x_n|\theta_1)} = e^{\frac{1}{2}\frac{n}{\sigma^2}(\theta_1^2-\theta_2^2)} e^{\frac{\theta_2-\theta_1}{\sigma^2}\sum_1^n x_i}$$

que es una función monótona no decreciente en $T(\overrightarrow{x}) = \sum_1^n x_i$.

Ejemplo 5.4.2
Para una población $Uniforme(0,\theta)$ con m.a.s. de tamaño n, la familia de distribuciones $\mathcal{P} = \{\frac{1}{\theta^n} I_{(0,\theta)}(x_{(n)}) | \theta \in \mathbb{R}^+\}$ tiene CVM en $X_{(n)}$ ya que cuando es $\theta_1 < \theta_2$ el cociente

$$\frac{f(\overrightarrow{x}|\theta_2)}{f(\overrightarrow{x}|\theta_1)} = \left(\frac{\theta_1}{\theta_2}\right)^n \frac{I_{(0,\theta_2)}(x_{(n)})}{I_{(0,\theta_1)}(x_{(n)})}$$

es una función monótona no decreciente en $T(\overrightarrow{x}) = x_{(n)}$ el máximo de los valores muestrales.
■

Si la función de densidad o de masa de la muestra, es de la familia exponencial uniparamétrica, es decir, es de la forma

$$f(x_1, ..., x_n|\theta) = c(\theta)h(x_1, ..., x_n)e^{\pi(\theta)T(x_1,...,x_n)} \tag{5.4.1}$$

y la función $\pi(\theta)$ es monótona no decreciente en θ; entonces tiene CVM en el estadístico suficiente $T(x_1, ..., x_n)$. Por lo tanto poblaciones $Beta$ con un parámetro conocido, Normal con un parámetro desconocida, Binomial y Poisson, poseen CVM.

La distribución $Uniforme(0,\theta)$ aunque no pertenece a la familia exponencial uniparamétrica posee CVM en $X_{(n)}$.

La familia de distribuciones $\mathcal{P} = \{\frac{1}{\pi(1+(x-\theta)^2)} | \theta \in \mathbb{R}\}$ de Cauchy no posee CVM en x, ya que para $\theta_1 < \theta_2$ el cociente

$$\frac{f(x|\theta_2)}{f(x|\theta_1)} = \frac{1 + (x - \theta_1)^2}{1 + (x - \theta_2)^2}$$

tiende a 1 cuando x tiende a $\pm\infty$.

Ahora se está en condiciones de generalizar el teorema de Neyman-Pearson mediante el siguiente enunciado.

Teorema 5.4.1 (Karlin-Rubin)
Si θ es un parámetro real y la familia de distribuciones de la muestra tiene CVM en T, para contrastar $H_0 : \theta \leq \theta_0$ frente a $H_1 : \theta > \theta_0$ se sigue:

1) El test de región crítica

$$RC = \{\overrightarrow{x} | T(\overrightarrow{x}) \geq c\} \tag{5.4.2}$$

y de región de aceptación

$$RA = \{\overrightarrow{x} | T(\overrightarrow{x}) < c\} \tag{5.4.3}$$

que cumple

$$\alpha = P\{RC|\theta_0\} \tag{5.4.4}$$

es óptimo en la clase $\mathcal{C}$ de los tests de nivel α.

2) La función de potencia del test dado en 1) es no decreciente.

Demostración: 1) Puede verse que si se contrasta $H_0^* : \theta = \theta_0$ frente a $H_1^* : \theta = \theta_1$ con $\theta_1 > \theta_0$ ambos fijos, el test del enunciado es el óptimo según el Teorema 5.3.1 (Neyman-Pearson). En efecto $f(\overrightarrow{x}|\theta_1) \geq kf(\overrightarrow{x}|\theta_0)$ es equivalente a $T(\overrightarrow{x}) \geq c$ y al verificarse (5.4.4) el test tiene de tamaño α para la hipótesis nula $H_0^* : \theta = \theta_0$. Antes de continuar con la demostración de la parte 1) se puede ver 2). Bajo la hipótesis de que la familia de distribuciones de la muestra tenga *CVM*, es

$$P\{T \geq c|\theta_1\} \leq P\{T \geq c|\theta_2\}$$

siempre que $\theta_1 \leq \theta_2$.

Para todo $t \geq c$, por la hipótesis de CVM resulta

$$\frac{f(t|\theta_2)}{f(t|\theta_1)} \geq \frac{f(c|\theta_2)}{f(c|\theta_1)}$$

o alternativamente

$$f(t|\theta_1) \leq \frac{f(c|\theta_1)}{f(c|\theta_2)} f(t|\theta_2).$$

Integrando en ambos miembros se tiene

$$\int_c^{+\infty} f(t|\theta_1)dt \leq \frac{f(c|\theta_1)}{f(c|\theta_2)} \int_c^{+\infty} f(t|\theta_2)dt$$

de donde para cualquier c arbitrario

$$\frac{f(c|\theta_2)}{1 - F(c|\theta_2)} \leq \frac{f(c|\theta_1)}{1 - F(c|\theta_1)}.$$

Si se vuelve a integrar se tiene que

$$\int_{-\infty}^{t_0} \frac{f(t|\theta_2)}{1 - F(t|\,\theta_2)}dt \leq \int_{-\infty}^{t_0} \frac{f(t|\theta_1)}{1 - F(t|\theta_1)}dt$$

para cualquier valor t_0 arbitrario, de manera que

$$-\ln[1 - F(t_0|\theta_2)] \leq -\ln[1 - F(t_0|\theta_1)]$$

que equivale a $P\{T \geq t_0|\theta_1\} \leq P\{T \geq t_0|\theta_2\}$.

Por lo tanto, la función de potencia del test de región crítica (5.4.2) es monótona no decreciente en θ. Como la función de potencia del test cumple (5.4.4), esto sirve para determinar el valor de la constante c. Se tiene hasta ahora, que el test construido es óptimo para contrastar $H_0^* : \theta = \theta_0$ frente a $H_1 : \theta > \theta_0$, ya que el test al que se llegaría mediante el teorema de Neyman-Pearson no depende de cúal sea θ_1 siempre que sea mayor que θ_0.

El test del enunciado es óptimo, en la clase $\mathcal{C}'$ de los test de nivel α para contrastar $H_0^* : \theta = \theta_0$ frente a $H_1 : \theta > \theta_0$, pero esta clase es más amplia que la clase $\mathcal{C}$ de los tests de nivel α para contrastar $H_0 : \theta \leq \theta_0$ frente a $H_1 : \theta > \theta_0$ y como por 2) el test del enunciado está en la clase $\mathcal{C}$ y es óptimo en una clase más amplia, es óptimo también en $\mathcal{C}$.

Ejemplo 5.4.3
Para contrastar $H_0 : \theta \leq \theta_0$ frente a $H_1 : \theta > \theta_0$ con m.a.s. de tamaño n de una población $N(\theta, 1)$, si se quiere contrastar $H_0 : \theta \leq \theta_0$ frente a $H_1 : \theta > \theta_0$ existe test óptimo; ya que de acuerdo con el Ejemplo 5.4.1 la familia de las distribuciones $N(\theta, \sigma)$ con σ conocida tiene CVM en $\sum_1^n x_i$.

Por lo tanto el test de región crítica $RC = \left\{ \sum_{i=1}^{n} x_i \geq z_\alpha \sqrt{n} + n\theta_0 \right\}$, por el teorema de Karlin-Rubin, es óptimo.

Ejemplo 5.4.4
Para contrastar $H_0 : \sigma^2 \leq \sigma_0^2$ frente a $H_1 : \sigma^2 > \sigma_0^2$ con m.a.s. de tamaño n de una población $N(\mu, \sigma)$ de media conocida, dado que la distribución $N(\mu, \sigma)$ tiene CVM en $\sum_{i=1}^{n}(x_i - \mu)^2$ el test de región crítica $RC = \{\sum_1^n (x_i - \mu)^2 \geq \sigma_0^2 \chi^2_{n;\alpha}\}$ es óptimo.

■

El contraste de la hipótesis nula $H_0 : \theta \geq \theta_0$; frente a la hipótesis alternativa $H_1 : \theta < \theta_0$, con la utilización de familias de distribuciones con cociente de verosimilitud monótono no creciente en $T(\overrightarrow{x})$, no debería suponer ningún obstáculo al lector, para que éste enuncie una versión adecuada a esta situación del teorema de Karlin-Rubin.

Lo mismo que se dijo tras el enunciado del Teorema de Neyman-Pearson ocurre con el Teorema de Karlin-Rubin; el conjunto de puntos que verifican $f(\overrightarrow{x}|\theta_1) = kf(\overrightarrow{x}|\theta_0)$ pueden ser colocados tanto en la región crítica como en la región de aceptación, permaneciendo válidos ambos enunciados.

Para poner de manifiesto que sin introducir más restricciones no pueden obtenerse contrastes óptimos para las hipótesis bilaterales, supóngase que se quiere contrastar $H_0 : \theta = \theta_0$ frente a $H_1 : \theta \neq \theta_0$ mediante m.a.s. de tamaño n de una población $N(\theta, 1)$. Por tener la familia de distribuciónes CVM en $\overline{x}$, el test óptimo para contrastar $H_0' : \theta \leq \theta_0$ frente a $H_1' : \theta > \theta_0$ tiene de región crítica $RC_1 = \{\overline{x}|\overline{x} \geq \theta_0 + \frac{z_\alpha}{\sqrt{n}}\}$. De manera análoga el test óptimo para contrastar $H_0^{"} : \theta \geq \theta_0$ frente a $H_1^{"} : \theta < \theta_0$ tiene de región crítica $RC_2 = \{\overline{x}|\overline{x} \leq \theta_0 + \frac{z_{1-\alpha}}{\sqrt{n}}\}$, las funciones de potencia de ambos tests se representan en la Figura 5.4.1. Es claro que si existiera un test óptimo

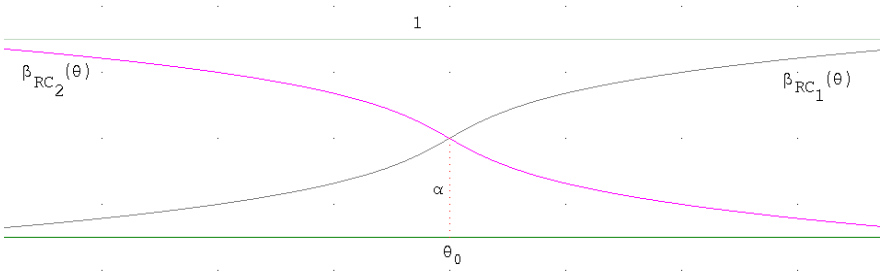


Figura 5.4.1. Función de potencia de los tests óptimos para hipótesis unilaterales en el caso normal.

para $H_0 : \theta = \theta_0$ frente a $H_1 : \theta \neq \theta_0$ de tamaño α, dado por una región crítica RC^*, su función de potencia $\beta_{RC^*}(\theta)$ debería coincidir con $\beta_{RC_1}(\theta)$, cuando $\theta > \theta_0$, y con $\beta_{RC_2}(\theta)$, cuando $\theta < \theta_0$. Pero esto es imposible, ya que por el Teorema de Karlin-Rubin en su parte 2) el test de región crítica RC^* ha de tener función de potencia no decreciente y como es $\beta_{RC^*}(\theta_0) = \alpha$ sería óptimo para contrastar $H_0 : \theta = \theta_0$ frente a $H_1 : \theta = \theta_1$ con $\theta_1 > \theta_0$.

De aquí se seguiría que RC^* ha de ser de la forma (5.4.1), por la parte de unicidad del Teorema de Neyman-Pearson, y por lo tanto $\beta_{RC^*}(\theta)$ a la izquierda de θ_0 no puede coincidir con $\beta_{RC_2}(\theta)$. Luego no puede haber un test cuya función de potencia coincida con $\beta_{RC_2}(\theta)$ a la izquierda de θ_0 y con $\beta_{RC_1}(\theta)$ a la derecha de θ_0.

Sin embargo, tanto el test de región crítica RC_1 como el de región crítica RC_2, deberían ser excluidos, cuando se pretenda contrastar $H_0 : \theta = \theta_0$ frente a $H_1 : \theta \neq \theta_0$, ya que tratan de manera asimétrica respecto a la función de potencia a los puntos que quedan por debajo de θ_0 de los que quedan por encima. Esto se logra restringiendo la clase de los contrastes a utilizar a la clase de los contrastes insesgados (véase la Definición 5.3.2), quedándose con los contrastes que rechazan la hipótesis nula con probabilidad mayor cuando ésta es falsa que cuando es cierta; es decir, solo se van a considerar aquellos contrastes tales que $P\{RC|\theta_0\} \leq \alpha$ y que $P\{RC|\theta\} \geq \alpha$ para cualquier $\theta \neq \theta_0$.

Obsérvese que el test dado en el Teorema de Karlin-Rubin, es insesgado para contrastar $H_0 : \theta \leq \theta_0$ frente a $H_1 : \theta > \theta_0$, pero no lo es para contrastar $H_0 : \theta = \theta_0$ frente a $H_1 : \theta \neq \theta_0$. En esta línea puede probarse el siguiente enunciado.

Teorema 5.4.2
Dado θ un parámetro real y cuando la familia de distribuciones de la muestra es la exponencial uniparamétrica $f(x_1, ..., x_n|\theta) = c(\theta)h(x_1, ..., x_n)e^{\pi(\theta)T(x_1,...,x_n)}$ con $\pi(\theta)$ función monótona no decreciente, existe contraste óptimo en la clase de los contrastes que son insesgados de nivel α para contrastar la

hipótesis nula $H_0 : \theta \in [\theta_1, \theta_2]$ frente a $H_1 : \theta \notin [\theta_1, \theta_2]$, donde eventualmente puede ser $\theta_1 = \theta_2 = \theta_0$. La región crítica del test óptimo es de la forma $RC = \{\overrightarrow{x} | T(\overrightarrow{x}) \leq c_1 \text{ o } T(\overrightarrow{x}) \geq c_2\}$ siempre que se tenga $P\{RC|\theta_1\} = P\{RC|\theta_2\} = \alpha$ en el caso en que $\theta_1 \neq \theta_2$, o $P\{RC|\theta_0\} = \alpha$ y $\frac{d}{d\theta} P\{RC|\theta\}|_{\theta=\theta_0} = 0$ en el caso $\theta_1 = \theta_2 = \theta_0$.

Demostración: La demostración puede verse en Lehmann(1986) pág. 135.

■

Para contrastar la hipótesis nula $H_0 : \theta \leq \theta_1$ o $\theta \geq \theta_2$ con $\theta_1 < \theta_2$ frente a la alternativa $H_1 : \theta \in [\theta_1, \theta_2]$ en la familia exponencial uniparamétrica no es necesario restingirse a la clase de los tests insesgados.

La hipótesis de que la distribución pertenezca a la familia exponencial uniparamétrica, con una función $\pi(\theta)$ monótona no decreciente, es necesaria para asegurar que la distribución tenga CVM en $T(\overrightarrow{x})$ y para que la derivada de la función de potencia exista y pueda anularse.

5.5. DUALIDAD ENTRE CONTRASTES DE HIPÓTESIS Y REGIONES DE CONFIANZA

Cuando en el Capítulo 4 se estudiaron las regiones de confianza, quedó pendiente el estudio de los criterios de óptimo para regiones de confianza. El motivo es que tales criterios se basan en el concepto de test óptimo, que ha sido introducido en este Capítulo. Más concretamente se va a ver en esta Sección, que existe un procedimiento natural de pasar de una familia de contrastes de hipótesis a una región de confianza y que este procedimiento permite trasladar las nociones de contrastes óptimos a definiciones de óptimo para su región de confianza correspondiente.

Conviene hacer notar, que esta dualidad solo tiene lugar en la aproximación clásica y no entre los contrastes y las regiones creíbles bayesianas. Con el fin de introducir la noción de dualidad considérese el siguiente Ejemplo.

Ejemplo 5.5.1

Para contrastar $H_0 : \theta \leq \theta_0$ frente a $H_1 : \theta > \theta_0$ con m.a.s. de tamaño n de una población $N(\theta, 1)$, el test óptimo de tamaño α tiene como región crítica $RC = \{\overline{X} \geq \theta_0 + z_\alpha \frac{1}{\sqrt{n}}\}$. Se rechaza la hipótesis nula para aquellos valores $\overline{X}$ que sean significativamente mayores de θ_0. Parece natural pensar que para cada $\overline{X}$ fijo los valores de θ compatibles con la media muestral serán aquellos que verifiquen $\overline{X} \leq \theta + z_\alpha \frac{1}{\sqrt{n}}$, es decir los valores de θ que hagan que $\overline{X}$ pertenezca a la región de aceptación del test. ¿Se puede entonces

considerar al conjunto

$$S(\overline{X}) = \left\{ \theta \left| \overline{X} - z_\alpha \frac{1}{\sqrt{n}} \leq \theta \right. \right\}$$

como una región de confianza? o más concretamente ¿es el intervalo $[\overline{X} - z_\alpha \frac{1}{\sqrt{n}}, +\infty)$, un intervalo de confianza para θ?

La respuesta es afirmativa, pues

$$\begin{aligned} P\{S(\overrightarrow{X}) \ni \theta|\theta\} &= P\{\overline{X} - z_\alpha \tfrac{1}{\sqrt{n}} \leq \theta|\theta\} = P\{\overline{X} \leq \theta + z_\alpha \tfrac{1}{\sqrt{n}}|\theta\} \\ &= P\{\overline{X} - \theta)\sqrt{n} \leq z_\alpha|\theta\} = 1 - \alpha. \end{aligned}$$

Por lo tanto $S(\overrightarrow{X})$ es una región de grado de confianza $1 - \alpha$.

■

El procedimiento es del todo general y ha ilustrado la correspondencia entre familias de contrastes hipótesis y regiones de confianza. Dada una familia de contrastes de hipótesis de tamaño α para contrastar $H_0 : \theta = \theta_0$ frente a $H_1 : \theta \in H(\theta_0)$ mediante las regiones críticas $RC(\theta_0)$ o bien mediante su complementario la región de aceptación $RA(\theta_0)$, se sigue que

$$S(\overrightarrow{X}) = \{\theta|\overrightarrow{X} \in RA(\theta)\}$$

es una región de grado de confianza $1 - \alpha$, ya que

$$P\{S(\overrightarrow{X}) \ni \theta|\theta\} = P\{\overrightarrow{X} \text{ tales que } \overrightarrow{X} \in RA(\theta)|\theta\} = 1 - P\{RC(\theta)|\theta\} \geq 1 - \alpha.$$

Se utiliza la notación $H_1 : \theta \in H(\theta_0)$ para hacer referencia a que la hipótesis alternativa dependerá de la hipótesis nula. Normalmente, para el caso en que el parámetro sea un valor real, a contrastes bilaterales les corresponderán regiones de confianza que sean intervalos con extremos finitos y a contrastes unilaterales regiones de confianza que sean intervalos con uno de sus extremos infinitos, como ha ocurrido en el Ejemplo 5.5.1.

Este método también puede ser invertido, así si $S(\overrightarrow{X})$ es una región de grado de confianza $1 - \alpha$, se puede construir una familia de contrastes de hipótesis, mediante el procedimiento de fijar arbitrariamente un punto como hipótesis nula $H_0 : \theta = \theta_0$ y definir

$$RC(\theta_0) = \{\overrightarrow{X}|\theta_0 \notin S(\overrightarrow{X})\}$$

entonces $RC(\theta_0)$ es la región crítica de un test de nivel α ; ya que

$$\begin{aligned} \beta_{RC(\theta_0)}(\theta_0) &= P\{\overrightarrow{X} \text{ tales que } \theta_0 \notin S(\overrightarrow{X})|\theta_0\} \\ &= 1 - P\{\overrightarrow{X} \text{ tales que } \theta_0 \in S(\overrightarrow{X})|\theta_0\} \\ &\leq \alpha, \end{aligned}$$

puesto que para cualquier θ es

$$P\{\overrightarrow{X} \text{ tales que } \theta \in S(\overrightarrow{X})|\theta\} \geq 1-\alpha.$$

A continuación se introducen las nociones de óptimo relativas a regiones de confianza.

Definición 5.5.1 (región de confianza uniformemente más ajustada)
Una región $S(X_1, ..., X_n)$ de grado de confianza $1-\alpha$ se dice uniformemente más ajustada (UMA) si para cualquier otra región $S'(X_1, ..., X_n)$ del mismo grado de confianza, se tiene que, para cualesquiera θ y θ'

$$P\{S(\overrightarrow{X}) \ni \theta'|\theta\} \leq P\{S'(\overrightarrow{X}) \ni \theta'|\theta\}$$

Es decir si $S(\overrightarrow{X})$ cubre valores incorrectos del parámetro con menor probabilidad que $S'(\overrightarrow{X})$.

También se emplea el término uniformemente más precisa o bien óptima.

Definición 5.5.2 (región de confianza insesgada)
Una región $S(X_1, ..., X_n)$ de grado de confianza $1-\alpha$ se dice insesgada, si para cualesquiera θ y θ' distintos

$$P\{S(\overrightarrow{X}) \ni \theta'|\theta\} \leq 1-\alpha$$

O lo que es lo mismo, si cubre valores correctos del parámetro con mayor probabilidad que valores incorrectos del mismo.
■

La relación entre estas definiciones y los conceptos de óptimo para contrastes, viene recogida en los dos teoremas siguientes

Teorema 5.5.1
A familias de contrastes de hipótesis de tamaño α óptimos, le corresponde una región de grado de confianza $1-\alpha$ UMA y recíprocamente.

Demostración: De acuerdo con lo lo visto en esta Sección, si se supone que la familia de contrastes para contrastar $H_0 : \theta = \theta_0$ frente a $H_1 : \theta \in H(\theta_0)$ viene dada por la colección de regiones críticas $RC(\theta_0)$ que son óptimas, donde se ha utilizado esta notación para poner de manifiesto que se tiene una colección de contrastes al ir variando θ_0. Su región de confianza asociada es $S(\overrightarrow{X}) = \{\theta|\overrightarrow{X} \notin RC(\theta)\}$, y para cualquier otra región $S'(\overrightarrow{X})$ de grado de confianza $1-\alpha$ se trata de ver si, cualesquiera que sean θ y θ', se cumple

$$P\{S(\overrightarrow{X}) \ni \theta'|\theta\} \leq P\{S'(\overrightarrow{X}) \ni \theta'|\theta\} \tag{5.5.2}$$

Sea $H_0 : \theta = \theta'$ y $H_1 : \theta \in H(\theta')$

$$\begin{aligned} P\{S(\overrightarrow{X}) \ni \theta'|\theta\} &= P\{\overrightarrow{X} \notin RC(\theta')|\theta\} \\ &= 1 - \beta_{RC(\theta')}(\theta) \end{aligned}$$

si se contruye la familia de contrastes, asociada a la región $S'(\overrightarrow{X})$, mediante $H_0 : \theta = \theta'$ frente a $H_1 : \theta \in H(\theta')$ será $RC'(\theta') = \{\overrightarrow{X}|S'(\overrightarrow{X}) \not\ni \theta'\}$. Es claro que

$$\begin{aligned} P\{S'(\overrightarrow{X}) \ni \theta'|\theta\} &= P\{\overrightarrow{X} \notin RC'(\theta')|\theta\} \\ &= 1 - \beta_{RC'(\theta')}(\theta) \end{aligned}$$

y, por la condición de óptimo de las regiones críticas $RC(\theta)$, ha de ser $\beta_{RC(\theta')}(\theta) \geq \beta_{RC'(\theta')}(\theta)$. Cambiando de signo esta expresión y sumando 1 en los dos miembros se obtiene (5.5.2).

Recíprocamente, para $S(\overrightarrow{X})$ una región de grado de confianza $1 - \alpha$ UMA, se construye su familia de contrastes correspondiente para $H_0 : \theta = \theta_0$ frente a $H_1 : \theta \in H_1(\theta)$ mediante $RC(\theta_0) = \{\overrightarrow{X}|\theta_0 \notin S(\overrightarrow{X})\}$. Si se tiene ahora una familia de contrastes de tamaño α $RC'(\theta_0)$, pasando a su región de confianza asociada $S'(\overrightarrow{X}) = \{\theta|\overrightarrow{X} \notin RC'(\theta)\}$ se tiene que

$$\beta_{RC'(\theta_0)}(\theta) = P\{S'(\overrightarrow{X}) \not\ni \theta_0|\theta\} = 1 - P\{S'(\overrightarrow{X}) \ni \theta_0|\theta\}$$

pero esta última expresión es menor o igual que $1 - P\{S(\overrightarrow{X}) \ni \theta_0|\theta\}$, por ser $S(\overrightarrow{X})$ UMA, por lo tanto

$$\beta_{RC'(\theta_0)}(\theta) \leq 1 - P\{S(\overrightarrow{X}) \ni \theta_0|\theta\} = P\{\overrightarrow{X} \text{ con } \theta_0 \notin S(\overrightarrow{X})|\theta\},$$

que coincide con $\beta_{RC(\theta_0)}(\theta)$ por construcción.

■

Teorema 5.5.2

A familia de contrastes de hipótesis de tamaño α insesgados, le corresponde una región de grado de confianza $1 - \alpha$ insesgada y recíprocamente.

Demostración: Sea $RC(\theta')$ la región crítica de un test insesgado de tamaño α para contrastar $H_0 : \theta = \theta'$ frente a $H_1 : \theta \in H(\theta')$, se trata de ver que con $S(\overrightarrow{X}) = \{\theta|\overrightarrow{X} \notin RC(\theta)\}$ es

$$P\{S(\overrightarrow{X}) \ni \theta'|\theta\} \geq P\{S(\overrightarrow{X}) \ni \theta|\theta\} \tag{5.5.3}$$

Por la Definición 5.5.2 es $P\{S(\overrightarrow{X}) \ni \theta'|\theta\} = 1 - \beta_{RC(\theta')}(\theta)$, y, de acuerdo con esto, $P\{S(\overrightarrow{X}) \ni \theta|\theta\} = 1 - \beta_{RC(\theta)}(\theta)$, pero al ser $RC(\theta')$ insesgado para todo θ' es

$$\beta_{RC(\theta')}(\theta) \leq \beta_{RC(\theta)}(\theta)$$

y por lo tanto se tiene (5.5.3).

Recíprocamente, si $S(\overrightarrow{X})$ es una región de grado de confianza $1-\alpha$ e insesgada, su familia de contrastes asociados para $H_0 : \theta = \theta_0$ frente a $H_1 : \theta \in H(\theta_0)$ tiene de regiones críticas $RC(\theta_0) = \{\overrightarrow{X} \text{ con } S(\overrightarrow{X}) \not\ni \theta_0\}$. Se trata de comprobar que

$$\beta_{RC(\theta_0)}(\theta) \geq \beta_{RC(\theta)}(\theta) \tag{5.5.4}$$

Ahora bien, $\beta_{RC(\theta_0)}(\theta) = P\{S(\overrightarrow{X}) \not\ni \theta_0|\theta\}$ y $\beta_{RC(\theta)}(\theta) = P\{S(\overrightarrow{X}) \not\ni \theta|\theta\}$ y por ser $S(\overrightarrow{X})$ insesgada (véase la Definición 5.5.2) es $P\{S(\overrightarrow{X}) \ni \theta_0|\theta\} \leq P\{S(\overrightarrow{X}) \ni \theta|\theta\}$ de donde se sigue (5.5.4).
■

Realmente, la parte de mayor interés de estos dos últimos enunciados, es la que permite pasar de los contrastes UMP a las regiones UMA y de los contrastes insesgados a las regiones insesgadas, ya que de esta forma se pueden construir regiones de confianza óptimas de manera automática, a partir del correspondiente test óptimo.

Ejemplo 5.5.2
Con muestras de tamaño uno, de una población con función de densidad

$$f(x|\theta) = \theta e^{-\theta x} I_{(0,+\infty)}(x) \quad \theta > 0,$$

para contrastar $H_0 : \theta \geq \theta_0$ frente a $H_1 : \theta < \theta_0$, puede comprobarse que el test óptimo e insesgado de tamaño α, tiene de región crítica $RC(\theta_0) = \{X \leq -\frac{\ln(1-\alpha)}{\theta_0}\}$. Por tanto la región de grado de confianza $1-\alpha$ UMA e insesgada viene dada por

$$S(X) = \left(-\frac{\ln(1-\alpha)}{X}, +\infty\right)$$

5.6. APROXIMACIÓN HISTÓRICA

La teoría de los contrastes estadísticos desde el punto de vista clásico es fruto de los trabajos de Egon Pearson y de Jerzy Neyman realizados entre los años 1926 y 1933. Egon Pearson, hijo de Karl Pearson, nació en Hampstead cerca de Londres en 1895 y estudió matemáticas en el *Trinity College* de Cambridge en 1914. Sus estudios se vieron perturbados por problemas de salud así como por el inicio de la segunda guerra mundial. En 1919 estudia astronomía para después ocupar una plaza en el Departamento de Estadística Aplicada del *University College* de Londres, donde se encuentra con Neyman en el 1925. Comenzó así un trabajo en equipo que llevaría a la elaboración de la teoría de los tests estadísticos y que duró hasta el año 1938, en que Neyman marcha a Estados Unidos, a la Universidad de Berkeley, California.

Los fundamentos de la aproximación bayesiana a la inferencia pueden fijarse en los trabajos de Bayes, Bernoulli y Laplace. Como ya se ha dicho en el Capítulo 3, la aproximación bayesiana tiene su origen en la Memoria de Bayes titulada *Ensayo hacia resolver un problema en la doctrina del azar*, publicada por Price en 1764 sobre la contribución a la probabilidad inversa. La traducción al castellano puede verse en Gómez Villegas y otros (2001). En la Memoria se obtiene por primera vez la versión continua del teorema de Bayes cuando la distribución inicial es uniforme, esto es,

$$\pi(\theta|x_1,...,x_n) = \frac{\pi(\theta)f(x_1,...,x_n|\theta)}{m(x_1,...,x_n)}$$

que cuando $\pi(\theta) = I_{(0,1)}(\theta)$ toma la forma

$$\pi(\theta|x_1,...,x_n) = \frac{\theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}}{\int_0^1 \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i} d\theta} \qquad \theta \in (0,1)$$

con $X_i \sim Bernoulli(\theta)$.

Bayes muere en 1761 y la Memoria contiene la aproximación que ha dado lugar al nacimiento de la inferencia bayesiana.

Continua siendo un misterio por qué el trabajo no fué publicado por el propio Bayes. Price en la presentación de la Memoria dice que los resultados en ella contenidos le han sido contados por Bayes y que la importancia de los mismos le decide a publicarlos.

El antecedente de la expresión continua del teorema de Bayes está en un trabajo de Jacob Bernoulli (1654-1705), quien en su libro *El Arte de la Conjetura*, publicado póstumamente por su sobrino Daniel Bernoulli en 1713, se plantea el problema de determinar la probabilidad θ de un suceso de Bernoulli. Afirma que si en n repeticiones el suceso se presenta m veces, se puede asegurar, con probabilidad todo lo próximo a 1 que se quiera, que

$$P\{|m/n - \theta| \leq \varepsilon\} > 1 - \delta \tag{5.6.1}$$

para ε y δ arbitrariamente pequeños.

Hasta ese momento el planteamiento había sido, dada una variable de $Bernoulli(\theta)$ con θ conocido, la variable aleatoria es m, que tiene la distribución $Binomial(n,\theta)$. Esta es la primera vez que se procede en sentido contrario, se pretende estimar el valor de θ a partir de m/n y además se calcula la probabilidad de que la probabilidad θ esté en el intervalo $\left(\frac{m}{n} - \varepsilon, \frac{m}{n} + \varepsilon\right)$. Por este motivo se dice que se trata de un enunciado de *probabilidad inversa*, fijados n grande y m se trata de hacer aseveraciones sobre θ; de invertir el camino que se venía siguiendo hasta entonces.

La Expresión (5.6.1) con m como v.a., es una ley de los grandes números y así es como realmente fué considerada por Jacob Bernoulli. Por el contrario la Expresión (5.6.1) empleada para determinar θ tiene sentido como fue interpretada por Bayes con θ como v.a. lo que lleva a introducir una distribución inicial, usar la verosimilitud y bien pasar a la distribución final, o bien tomarla como una determinación de un intervalo de confianza para θ, para lo que habría que esperar a la teoría de Egon Pearson y J. Neyman desarrollada alrededor de 1933, o a la teoría de la probabilidad fiducial introducida por Fisher sobre el 1930.

La idea de considerar a θ como una variable aleatoria fué retomada por Laplace (1749-1827) quien en su *Memoria sobre la probabilidad de las causas de los sucesos*, aparecida en 1774, recoge la demostración de la versión continua del teorema de Bayes, parece que sin conocer la obtenida previamente por Bayes. Más tarde vuelve a incluir esta expresión en su libro publicado en 1812 *Teoría Analítica de las Probabilidades*, donde la utiliza con distribuciones a priori distintas de la uniforme y donde incluye uno de los primeros problemas de contraste de hipótesis: el de contrastar si la probabilidad de que un recién nacido sea varón es $0'50$ y si esta probabilidad es la misma en Francia que en Inglaterra.

La aproximación bayesiana al contraste de hipótesis continuó desarrollándose hasta la aparición de los trabajos citados de Egon Pearson, J. Neyman y Fisher que interrumpieron la utilización de los métodos bayesianos, básicamente por depender de la distribución inicial, lo que hace que distintas personas con diferente información inicial y la misma información muestral lleguen a conclusiones distintas. Continuaron con la aproximación bayesiana, entre otros, Ramsey, quien en 1931 publica sus *Fundamentos de Matemáticas*, Jeffreys, con su *Teoría de la Probabilidad* en 1939; De Finetti, 1974; Leonard Savage en 1954 y Lindley en 1975, quienes han hecho que actualmente los métodos bayesianos, particularmente los contrastes de hipótesis, presenten un elevado desarrollo.

5.7. EJERCICIOS

5.7.1

Para m.a.s. de tamaño 12 de una población de $Poisson(\theta)$, con $\theta \in [0, 0'5]$. Si la región crítica para contrastar $H_0 : \theta = 0'5$ frente a $H_1 : \theta < 0'5$ viene dada por

$$RC = \{\overrightarrow{x} = (x_1, ..., x_{12}) | \sum_{i=1}^{12} x_i \leq 2\}$$

Hállese la función de potencia del test para $\theta = 0'5$, $\theta = 0'25$ y $\theta = 1/12$ y dibújese ésta de forma aproximada.

5.7.2
Para m.a.s. de tamaño 2 de una población con función de densidad

$$f(x|\theta) = \theta x^{\theta-1} I_{(0,1)}(x), \qquad \Theta = \{1, 2\}$$

se emplea la región crítica dada por

$$RC = \{(x_1, x_2)|0 < x_i < 1 \text{ i=1,2 con } 3/(4x_1) \leq x_2\}$$

para contrastar $H_0 : \theta = 1$ frente a $H_1 : \theta = 2$. Se pide:

a) Función de potencia.

b) Probabilidades de error de tipo I y de tipo II.

5.7.3
Para m.a.s. de tamaño 3 de una población de Bernoulli(θ) se quiere contrastar $H_0 : 0 \leq \theta \leq 0'5$ frente a $H_1 : 0'5 \leq \theta \leq 1$ y se aplican los tests de regiones críticas

$$RC_1 = \left\{ \sum_{i=1}^{3} x_i = 3 \right\} \qquad RC_2 = \left\{ \sum_{i=1}^{3} x_i \geq 2 \right\}$$

Compárense sus funciones de potencia.

5.7.4
Para m.a.s. de tamaño 16 de una población $N(\theta, 1)$ con $\theta \in \{0, 1\}$ se utiliza el test de región crítica

$$RC = \{\overline{x} \geq k\}$$

para contratar $H_0 : \theta = 0$ frente a $H_1 : \theta = 1$. Se pide:

a) Valor de k para que el test tenga de tamaño $0'01$.

b) Probabilidad de error de tipo I.

c) Probabilidad de error de tipo II.

d) Función de potencia.

5.7.5
En 1000 tiradas de una moneda se han obtenido 550 caras. ¿Es razonable suponer que la moneda es sesgada? Justifíquese la respuesta.

5.7.6
Calcúlese el p-valor en cada una de las siguientes situaciones:

a) Para contrastar $H_0 : \theta \leq 0'5$ frente a $H_1 : \theta > 0'5$ cuando se observan 7 éxitos en la repetición con independencia de 10 experimentos de Bernoulli.

b) Para contrastar $H_0 : \theta \leq 1$ frente a $H_1 : \theta > 1$ cuando se observa $x = 3$ con $X \sim Poisson(\theta)$.

c)Para contrastar $H_0 : \theta \leq 1$ frente a $H_1 : \theta > 1$ cuando se observa la m.a.s. (3,5,1) con $X \sim Poisson(\theta)$.

5.7.7
Para m.a.s. de tamaño n de una población $N(\theta, 1)$, determínese el test de la razón de verosimilitudes para contrastar $H_0 : \theta = 0$ frente a $H_1 : \theta > 0$.

a) Calcúlese su *función de potencia.*

b) Determínense n y c (constante del test), para que la función de potencia satisfaga que $\beta(0'1) = 0'95$ y $\beta(-0'1) = 0'05$.

5.7.8
Sea $X_1, ..., X_n$ una m.a.s. de tamaño n de una población $N(\mu, \sigma)$ con ambos parámetros desconocidos. Dedúzcase el test de la razón de verosimilitudes para contrastar $H_0 : \mu = \mu_0$ frente a $H_1 : \mu \neq \mu_0$.

5.7.9
Un examen consta de 100 preguntas, cada una de ellas admite la respuesta verdadero o falso. Se desea que los alumnos que superen la prueba sepan al menos el 50 % de la asignatura. Si se quiere que la probabilidad de suspender a un alumno que realmente se sabe la asignatura sea menor o igual que $0'01$ ¿cuál debe ser el número mínimo de respuestas correctas que se debe exigir para aprobar el examen? Supóngase que las preguntas abarcan uniformemente toda la asignatura y que para cada pregunta sólo una de las alternativas es la correcta.

5.7.10
En un examen tipo cuestionario se plantean 10 cuestiones, cada una de ellas admite la respuesta verdadero o falso. Aprobará el examen quien responda correctamente a 5 o más preguntas. ¿Cuál es la probabilidad de que apruebe el examen alguien que responde todo al azar? ¿Cuál es la probabilidad de que apruebe el examen una persona que sabe el 30 % de la asignatura? ¿Cuál debe ser el mínimo número de respuestas correctas que se debe exigir para aprobar, si se pretende que una persona que sepa al menos el 80 % de la asignatura apruebe con probabilidad $0'9$? En este último caso plantéese el contraste con todos sus elementos.

5.7.11
Para muestras de tamaño 1 de una población con distribución $Gamma(a = 0'5, p)$ hállese un test UMP de tamaño $0'05$ para contrastar $H_0 : p = 0'5$ frente a $H_1 : p = 1$

5.7.12
Sea $(X_1, ..., X_n)$ una m.a.s. de una población con distribución $Gamma(a, p = 1)$

a) Encuéntrese el test del cociente de verosimilitudes para contrastar $H_0 : a = a_0$ frente a $H_1 : a \neq a_0$.

b) Análogamente para contrastar $H_0 : a \geq a_0$ frente a $H_1 : a < a_0$.

5.7.13
Sea (X_1, X_2) una m.a.s. de una población de $Bernoulli(\theta)$ con $\theta \in \{0'25, 0'5\}$. Constrúyase un test óptimo de tamaño $\alpha = 3/16$ para contrastar $H_0 : \theta = 0'25$ frente a $H_1 : \theta = 0'5$.

5.7.14
Para una muestra de tamaño 1, determínese el test UMP de tamaño $0'05$ para contrastar H_0: procede de una población $N(0,1)$, frente a H_1 : procede de una población con función de densidad $Cauchy(0,1)$, es decir,

$$f_1(x) = \frac{1}{\pi}\frac{1}{1+x^2}, \quad x \in \mathbb{R}$$

5.7.15
Para una muestra de tamaño 1, determínese el test UMP de tamaño $0'01$ y su función de potencia, para contrastar H_0 :procede de una población $N(0,1)$, frente a H_1 : procede de una población con distribución $Gamma(1,1)$.

5.7.16
Para las siguientes familias de distribuciones de probabilidad, encuéntrense los tests óptimos de tamaño α para contrastar $H_0 : \theta \leq \theta_0$ frente a $H_1 : \theta > \theta_0$, basados en m.a.s. de n observaciones.

a) $f(x|\theta) = \theta e^{-\theta x}$, $x > 0, \theta > 0$.

b) $f(x|\theta) = \theta x^{\theta-1}$, $0 < x < 1, \theta > 0$.

5.7.17

a) De la producción de una fábrica se toma una m.a.s. de 50 elementos, resultando que $\overline{x} = 29$ y $\sum_1^{50}(x_i - \overline{x})^2 = 575$. Contrástese la hipótesis de que la media de la población es 30 mediante un test de tamaño $0'01$ suponiendo que la población tiene una distribución normal.

b) Se quiere saber si ha habido variación en la población para lo que se considera por una parte la muestra del apartado a) y una semana más tarde se toma otra m.a.s., independiente con la anterior, de tamaño 80 que arroja $\overline{y} = 22$ y $\sum_1^{80}(y_i - \overline{y})^2 = 650$. Contrástese la hipótesis de que el proceso no ha variado, mediante un test de tamaño $0'05$ bajo las mismas suposiciones de normalidad.

5.7.18
En un conjunto de 300 personas tomadas al azar de la ciudad A, hay 93 que ganan menos de 1,800 euros al mes. En la ciudad B, de un grupo de 200 personas tomadas al azar, hay 52 que ganan menos de 1,800 euros al mes. ¿Se podría rechazar la hipótesis de que la proporción de individuos que ganan menos de 1,800 euros al mes es la misma en ambas ciudades? Tómese para el test un tamaño de $0'05$.

5.7.19
Contrástese la hipótesis $H_0 : \theta = 1$ frente a $H_1 : \theta = 2$ para una m.a.s. de tamaño n de una población, con función de densidad

$$f(x|\theta) = \theta e^{-\theta x} \quad x > 0, \quad \theta > 0$$

suponiendo que la información inicial viene recogida por $P\{\theta = 1\} = \frac{2}{3}$.

5.7.20
A un estudiante se le somete a un test, con el fin de determinar su coeficiente de inteligencia θ que arroja como puntuación $x = 115$. Se supone que el resultado del test $X \sim N(\theta, 10)$ y que en el colectivo estudiado $\theta \sim N(100, 15)$.

Contrástese mediante un test bayesiano la hipótesis de que su coeficiente de inteligencia es menor o igual que 100 frente a su complementario. Calcúlese el p-valor correspondiente al test clásico para la misma situación.

5.7.21
Para contrastar la hipótesis nula $H_0 : \theta = 0$ frente a $H_1 : \theta \neq 0$ con m.a.s. de tamaño n de una población $N(\theta, \sigma)$ con σ conocida, se utiliza una distribución inicial que da $P\{\theta = 0\} = 0'5$ y si $\theta \neq 0$ vale $0'5\pi(\theta)$ donde $\pi(\theta)$ es la densidad de una $N(0, \tau_0)$ con τ_0 desviación típica conocida. Se pide:

a) Compruébese que, efectivamente $P\{-\infty < \theta < +\infty\} = 1$.

b) Calcúlese $P\{H_0|x_1, ..., x_n\}$.

c) Obténgase una expresión para el p-valor cuando se observe $\overline{x}$.

d) Para el caso $\sigma^2 = \tau_0^2 = 1$ y $n = 10$ calcúlense el p-valor y la probabilidad final de la hipótesis nula cuando los valores de la media muestral $\overline{x}$ han sido:

$$0, 0'1, 0'4, 0'6, 1, 2.$$

5.8. SOLUCIÓN A LOS EJERCICIOS

5.8.1
La función de potencia del contraste es

$$\beta_{RC}(\theta) = P\{RC|\theta\}$$

y como $\sum_{i=1}^{12} X_i$ tiene distribución de $Poisson(12\theta)$

$$\beta_{RC}(\theta) = P\left\{\sum_{i=1}^{12} X_i \leq 2|\theta\right\} = F(2|12\theta)$$

donde F es la función de distribución de la v.a. de $Poisson(12\theta)$. Se tiene que $F(2|6) = 0'062, F(2|3) = 0'4232$ y $F(2|1) = 0'9197$ por lo que la función de potencia es la recogida en la Figura 5.8.1, donde el valor correspondiente a $\theta = 0$ se determina por continuidad.

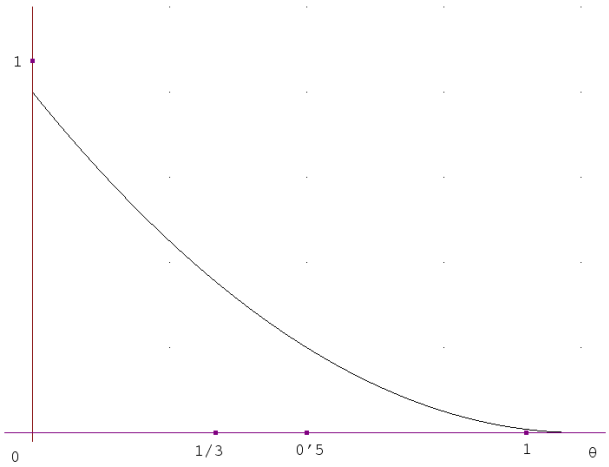


Figura 5.8.1. Función de potencia correspondiente al Ejercicio 5.7.2.

5.8.2

a) La función de potencia es

$$\beta_{RC}(\theta) = P\{(x_1, x_2) \ \text{ con } \ 0 < x_i < 1 \quad i = 1, 2 \ \text{ y } 3 \leq 4x_1x_2\}$$

se trata de calcular, véase la Figura 5.8.2,

$$\begin{aligned} P\{\text{zona rayada}|\theta\} &= \int_{3/4}^{1} \int_{\frac{3}{4x_1}}^{1} \theta^2 x_1^{\theta-1} x_2^{\theta-1} dx_2 dx_1 \\ &= \int_{3/4}^{1} \theta x_1^{\theta-1} \left[1 - \left(\frac{3}{4x_1}\right)^{\theta}\right] dx_1 \\ &= 1 - \left(\frac{3}{4}\right)^{\theta} (1 - \theta \ln \frac{3}{4}) \end{aligned}$$

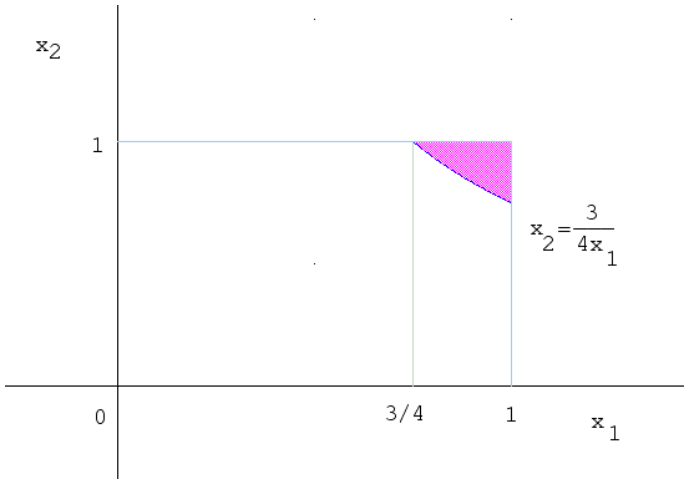


Figura 5.8.2. Representación de la región crítica del Ejercicio 5.7.2.

b) La probabilidad de cometer error de tipo I viene dada por

$$\beta_{RC}(1) = 1 - \frac{3}{4}\left(1 - \ln\frac{3}{4}\right) = 0'034.$$

La probabilidad de cometer error de tipo II es

$$1 - \beta_{RC}(2) = 1 - \left[1 - \frac{9}{16}(1 - 2\ln\frac{3}{4})\right] = 0'096.$$

5.8.3

Las funciones de potencia son

$$\beta_{RC_1}(\theta) = P\left\{\sum_{i=1}^{3} X_i = 3|\theta\right\} = \theta^3$$
$$\beta_{RC_2}(\theta) = P\left\{\sum_{i=1}^{3} X_i \geq 2|\theta\right\} = \theta^3 + 3\theta^2(1-\theta) = \theta^2(3-2\theta)$$

luego la función de potencia del primer test es menor que la función de potencia del segundo, y por tanto en la clase de los tests de nivel 0'5 es mejor el test de región crítica RC_1, véase la Figura 5.8.3.

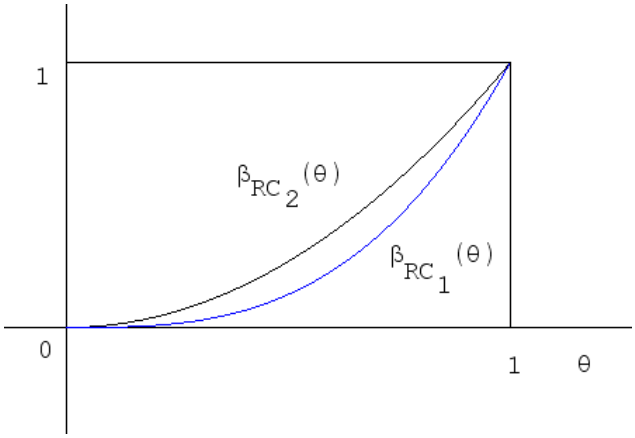


Figura 5.8.3. Funciones de potencia de los tests del Ejercicio 5.7.3.

5.8.4

Como bajo la suposición de normalidad $\overline{X} \sim N(\theta, \frac{1}{\sqrt{n}})$ se tiene

a) $P\{\overline{X} \geq k|\theta = 0\} = P\{Z \geq (k-\theta)\sqrt{n}\} = 0'01$

con $Z \sim N(0,1)$, de donde ha de ser $k4 = 2'326$ o lo que es lo mismo $k = 0'5815$.

b) La probabilidad de cometer error de tipo I es la probabilidad bajo $\theta = 0$ de rechazar la hipótesis nula, es decir

$$\begin{aligned} P\{RC|\theta = 0\} &= P\{\overline{X} \geq 0'5815|\theta = 0\} \\ &= P\{Z \geq 4 \cdot 0'5815\} \\ &= P\{Z \geq 2'326\} \\ &\simeq 0'01. \end{aligned}$$

c) La probabilidad de cometer error de tipo II es la probabilidad bajo $\theta = 1$ de aceptar la hipótesis nula

$$\begin{aligned} 1 - P\{RC|\theta = 1\} &= P\{\overline{X} < 0'5815|\theta = 1\} \\ &= P\left\{ \frac{\overline{X} - 1}{1/4} < \frac{0'5815 - 1}{1/4} \middle| \theta = 1 \right\} \\ &= P\{Z < 4(0'5815 - 1)\} \\ &= P\{Z \leq -1'674\} \\ &\simeq 0'047. \end{aligned}$$

d) De acuerdo con la Definición 5.1.3 la función de potencia es

$$\begin{aligned}\beta_{RC}(\theta) &= P\{RC|\theta\} = P\{\overline{X} \geq 0'5815|\theta\} \\ &= P\left\{\frac{\overline{X}-\theta}{1/4} \geq \frac{0'5815-\theta}{1/4}\middle|\,\theta\right\} \\ &= P\{Z \geq 4(0'5815-\theta)\} \\ &= 1-\Phi(4(0'5815-\theta))\end{aligned}$$

con Φ la función de distribución de la v.a. $N(0,1)$.

5.8.5

El test de la RV para contrastar $H_0 : \theta_0 = 0'5$ frente a $H_1 : \theta \neq 0'5$ tiene de región crítica

$$RC = \{\overrightarrow{x} = (x_1, ..., x_n)|\lambda(\overrightarrow{x}) \leq k\}$$

donde

$$\lambda(\overrightarrow{x}) = \frac{(0'5)^{n\overline{x}}(1-0'5)^{n(1-\overline{x})}}{\underset{\theta\in(0,1)}{\text{máx}}\ \theta^{n\overline{x}}(1-\theta)^{n(1-\overline{x})}}$$

el máximo del denominador se alcanza en $\widehat{\theta} = \overline{x}$ por lo que es

$$\lambda(\overrightarrow{x}) = \left(\frac{0'5}{\overline{x}}\right)^{n\overline{x}}\left(\frac{1-0'5}{1-\overline{x}}\right)^{n(1-\overline{x})}$$

Esta expresión es una función simétrica en $\overline{x}$ respecto al punto $0'5$, con lo que $RC = \{|\overline{x}-0'5| \geq c\}$. Para la determinación de c se emplea que la distribución aproximada de $\overline{X}$ es una normal $N(0'5, \sqrt{1000 \cdot 0'5^2})$ es decir $\overline{X} \simeq N(0'5, 15'8114)$.

Para la m.a.s. particular que arroja 550 caras en 1000 tiradas, el p-valor es

$$\begin{aligned}p-valor(0'55) &= P\{|\overline{X}-0'5| > |0'55-0'5|\ |\theta = 0'5\} \\ &\simeq P\left\{|Z| \geq \frac{0'05}{15'8114}\right\} \\ &= P\{|Z| \geq 3'16\} \\ &= 0'00079\end{aligned}$$

luego se debe rechazar el que la moneda sea equilibrada.

También hubiera sido posible contrastar $H_0 : \theta = 0'5$ frente a $H_1 : \theta > 0'5$, de todas maneras se ha preferido recoger la hipótesis alternativa más amplia.

5.8.6

a) Para contrastar $H_0 : \theta \leq 0'5$ frente a $H_1 : \theta > 0'5$ mediante el método de la RV se obtiene el test de $RC = \left\{ \sum_{i=1}^{n} x_i \geq c \right\}$ por lo que el p-valor con $\sum_{i=1}^{10} x_i = 7$ es

$$\sup_{\theta \in H_0} P\left\{ \sum_{i=1}^{10} X_i \geq 7 \middle| \theta \right\} = P\left\{ \sum_{i=1}^{10} X_i \geq 7 \middle| \theta = 0'05 \right\} =$$
$$= P\left\{ \sum_{i=1}^{10} X_i \leq 3 \middle| \theta = 0'5 \right\} =$$
$$= 0'1719$$

b) En el contraste $H_0 : \theta \leq 1$ frente a $H_1 : \theta > 1$ con $x = 3$ para una población $X \sim Poisson(\theta)$, el método de la RV conduce al test de $RC = \{x \geq c\}$, en este caso el p-valor correspondiente a $x = 3$ es

$$\sup_{\theta \in H_0} P\{X \geq 3|\theta\} = P\{X \geq 3|\theta = 1\} = 0'0803$$

c) Si se enfrenta $H_0 : \theta \leq 1$ frente a $H_1 : \theta > 1$ tras observar la m.a.s. (3,5,1), mediante el test de la RV la $RC = \left\{ \sum_{i=1}^{3} X_i \geq 0 \right\}$, por lo que el p-valor, viene dado por

$$\sup_{\theta \in H_0} P\left\{ \sum_{i=1}^{3} X_i \geq 3+5+1 \middle| \theta \right\} = P\{Y \geq 9|\theta = 3\} = 0'0038$$

donde $Y \sim Poisson(\theta = 3)$ en virtud de la reproductividad de la distribución de Poisson.

5.8.7

a) Para contrastar con m.a.s. de tamaño n de una población $N(\theta, 1)$ $H_0 : \theta = 0$ frente a $H_1 : \theta > 0$, el test de la RV viene dado por

$$RC = \{\lambda(x_1, ..., x_n) \leq c\}$$

donde

$$\lambda(x_1, ..., x_n) = \frac{f(x_1, ..., x_n|\theta)}{\max\limits_{\theta \geq 0} f(x_1, ..., x_n|\theta)}$$

como

$$\max_{\theta \geq 0} e^{-\frac{1}{2}\sum_{1}^{n}(x_i - \theta)^2} = \begin{cases} e^{-\frac{1}{2}\sum x_i^2} & \text{si} \quad \overline{x} \leq 0 \\ e^{-\frac{1}{2}\sum (x_i - \overline{x})^2} & \text{si} \quad \overline{x} > 0 \end{cases}$$

se sigue que

$$\lambda(\overline{x}) = \begin{cases} 1 & \text{si} \quad \overline{x} \leq 0 \\ e^{-\frac{1}{2}\left(\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n}(x_i - \overline{x})^2\right)} & \text{si} \quad \overline{x} > 0 \end{cases}$$

por lo que simplificando

$$\lambda(\overline{x}) = \begin{cases} 1 & \text{si} \quad \overline{x} \leq 0 \\ e^{-\frac{n}{2}\overline{x}^2} & \text{si} \quad \overline{x} > 0 \end{cases}$$

y por tanto, tras la representación de $\lambda(\overline{x})$, la región crítica es de la forma $RC = \{\overline{X} \geq k\}$.

b) La función de potencia es $\beta(\theta) = P\{\overline{X} \geq k|\theta\}$ y como $\overline{X} \sim N(\theta, \frac{1}{\sqrt{n}})$

$$\beta(\theta) = P\{Z \geq (k-\theta)\sqrt{n}\}$$

sin más que tipificar la variable $\overline{X}$, y donde $Z \sim N(0,1)$.

Se trata de resolver en n y k el sistema

$$P\{Z \geq (k-0'1)\sqrt{n}\} = 0'95$$
$$P\{Z \geq (k+0'1)\sqrt{n}\} = 0'05$$

o alternativamente

$$-1'6449 = (k-0'1)\sqrt{n}$$
$$1'6449 = (k+0'1)\sqrt{n}$$

de donde $n = 269$ y $k = 0$.

5.8.8

Con m.a.s. de tamaño n de una población $N(\mu, \sigma)$ el test de la RV para contrastar $H_0 : \theta = \theta_0$ frente a $H_1 : \theta \neq \theta_0$ con σ desconocida, viene dado por

$$RC = \{\lambda(x_1, ..., x_n) \leq k\}$$

con

$$\lambda(\overrightarrow{x}) = \frac{\max\limits_{\sigma} \left(\frac{1}{\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum(x_i-\theta_0)^2}}{\max\limits_{\theta,\sigma} \left(\frac{1}{\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum(x_i-\theta_0)^2}} \tag{5.8.1}$$

Para maximizar en θ y σ el denominador, se observa que para todo σ fijo el máximo en θ se obtiene minimizando en θ la expresión $\sum\limits_{1}^{n}(x_i - \theta)^2$ que como se sabe se consigue con $\widehat{\theta} = \overline{x}$. Por lo tanto bastará maximizar, ahora en σ, la expresión $\sigma^{-n}e^{-\frac{1}{2\sigma^2}\sum(x_i-\overline{x})^2}$. Tomando logaritmos se pretende maximizar $-n\ln\sigma - \frac{1}{2\sigma^2}\sum(x_i - \overline{x})^2$ cuya derivada igualada a cero es

$$-\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - \overline{x})^2 = 0$$

cuya única solución es $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$ la varianza muestral. Como la derivada segunda es $\frac{n}{\sigma^2} + \frac{3\sum_{i=1}^{n}(x_i-\overline{x})^2}{\sigma^4}$, que siempre es positiva, se sigue que $\widehat{\sigma}^2$ es un máximo y sustituyendo en el denominador de (5.8.1) se obtiene que éste vale $\widehat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \theta_0)^2$ y el numerador vale $\widehat{\sigma}_0^{-n}e^{-\frac{n}{2}}$, por lo que

$$\lambda(\overrightarrow{x}) = \left(\frac{\widehat{\sigma}}{\widehat{\sigma}_0}\right)^n = \left(\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{\sum_{i=1}^{n}(x_i - \theta_0)^2}\right)^n$$

sumando y restando en el denominador $\overline{x}$ se obtiene que la razón es $\lambda(t) = \left(1 + \frac{t^2}{n-1}\right)^{-n}$ con $t = \frac{\overline{x}-\theta_0}{S}\sqrt{n}$. Expresión que bajo la hipótesis nula se distribuye con distribución T_{n-1} de Student. En resumen, que al ser $\lambda(t)$ simétrica, la región crítica queda

$$RC = \{\lambda(t) \leq k\} = \{|t| \leq c\} = \left\{\left|\frac{\overline{x} - \theta_0}{S}\sqrt{n}\right| \leq c\right\}$$

Para la obtención de la constante hay que hacer que

$$\sup_{\sigma} P\{RC|\theta = \theta_0, \sigma\} = \alpha$$

con lo que basta que sea $c = t_{\frac{\alpha}{2}}$ y por lo tanto

$$RC = \left\{\left|\frac{\overline{x} - \theta_0}{S}\right|\sqrt{n} \leq t_{\frac{\alpha}{2}}\right\}.$$

5.8.9

Si se denota por $\theta \in [0,1]$ el porcentaje de asignatura que conoce un alumno, se pretende contrastar $H_0 : \theta \leq 0'5$ frente a $H_1 : \theta > 0'5$. Para aprobar, el alumno deberá contestar por lo menos a k preguntas de forma correcta, con k un entero menor que 100. Se puede suponer que el número de preguntas correctas contestadas es $\sum_{i=1}^{100} X_i$ una v.a. suma de v.a. de $Bernoulli(\theta)$ cuando se sabe un tanto por uno θ de la asignatura. Con este planteamiento un alumno aprobará si $\sum_{i=1}^{100} X_i \geq k$, y suspenderá si $\sum_{i=1}^{100} X_i < k$. La probabilidad de que un alumno conteste correctamente a una pregunta es

$$\begin{aligned} P\{X_i = 1|\theta\} = P\{&\text{se sepa la respuesta}\}\cdot \\ &\cdot P\{\text{la marque correctamente}|\text{se sabe la respuesta}\}+ \\ &+ P\{\text{no sepa la respuesta}\}\cdot \\ &\cdot P\{\text{la marque correctamente}|\text{no sabe la respuesta}\} \end{aligned}$$

que puede ponerse

$$p = P\{X_i = 1|\theta\} = \theta + (1-\theta)0'5 = 0'5 + 0'5\theta$$

por lo tanto $\sum_{i=1}^{100} X_i \sim Binomial(100, p)$.

La petición de acotar por $0'01$ la probabilidad de suspender cuando se sabe la asignatura, permite determinar el valor de k de la expresión

$$P\{\sum_{i=1}^{100} X_i < k|\theta = 0'5\} \leq 0'01$$

donde $\sum_{i=1}^{100} X_i \sim Binomial(100, 0'75)$. Utilizando la aproximación normal a la distribución Binomial $\overline{X} \sim N\left(0'75, \sqrt{\frac{0'75 \cdot 0'25}{100}}\right) = N(0'75, 0'043)$ se trata de obtener k de la expresión

$$P\left\{Z \leq \frac{k/100 - 0'75}{0'0433}\right\} = 0'01$$

donde $Z \sim N(0,1)$, lo que lleva a

$$\frac{k}{100} - 0'75 = 0'0433 \cdot (-2'3263)$$

y redondeando, ha de ser $k = 65$ el número mínimo de preguntas que un alumno ha de contestar correctamente para aprobar.

5.8.10

Entendiendo que el número de respuestas correctas viene modelizado por una variable aleatoria $\sum_{i=1}^{10} X_i$ donde $X_i = 1$ si la respuesta a la pregunta i-ésima es correcta. Una persona aprobará cuando $\sum_{i=1}^{10} X_i \geq 5$, mediante el *Teorema de la probabilidad total*, la probabilidad de que una persona que conoce un tanto por uno θ de la materia conteste correctamente es

$$\begin{aligned} p &= P\{X_i = 1|\theta\} \\ &= P\{\text{se sepa la respuesta}\}P\{\text{la marque correctamente}|\text{se sabe la respuesta}\} \\ &+ P\{\text{no sepa la respuesta}\}P\{\text{la marque correctamente}|\text{no sabe la respuesta}\} \\ &= 0'5 + 0'5\ \theta. \end{aligned}$$

Si la persona responde al azar es $p = 0'5$, valor correspondiente a $\theta = 0$; la probabilidad de que apruebe es

$$P\left\{\sum_{i=1}^{10} X_i \geq 5 \middle| \theta = 0\right\} = 1 - 0'377 = 0'623$$

La probabilidad de que apruebe el examen una persona que sabe el 30 % de la asignatura es

$$P\left\{\sum_{i=1}^{10} X_i \geq 5 \middle| \theta = 0'3\right\} = P\{Y \geq 5\} = 0'9052$$

con $Y \sim Binomial(10, 0'65)$ donde p es el valor correspondiente a $\theta = 0'3$.

Para calcular el número mínimo de preguntas correctas a responder para aprobar k, si se pretende que una persona que sepa al menos el 80 % de la asignatura apruebe con probabilidad $0'9$. Se trata de determinar k para que

$$P\{\sum_{i=1}^{10} X_i \geq k|\theta\} \geq 0'9$$

con $\theta \geq 0'8$, como al valor $\theta = 0'8$ le corresponde $p = 0'9$ se debe obtener k tal que

$$P\{Y \geq k\} \geq 0'9$$

con $Y \sim Binomial(10, 0'9)$, de donde $k = 8$.

En este último caso se ha pretendido contrastar la hipótesis H_0 : se sabe un porcentaje de asignatura $\theta \geq 0'8$ frente a H_1 : se sabe $\theta < 0'8$ y la región crítica es $RC = \left\{\sum_{i=1}^{10} X_i \leq 8\right\}$.

La probabilidad de cometer error de tipo I es

$$\sup_{\theta \geq 0'8} P\left\{\sum_{i=1}^{10} X_i \leq 8 \middle| \theta\right\} = P\left\{\sum_{i=1}^{10} X_i \leq 8 \middle| \theta = 0'8\right\} = P\{Y \leq 8\}$$

con $Y \sim Binomial(10, 0'9)$ y vale $0'0702$.

5.8.11

De acuerdo con el *Teorema de Neyman-Pearson*, para contrastar $H_0 : p = 0'5$ frente a $H_1 : p = 1$ el test óptimo tendrá como región crítica

$$RC = \{f(x|p = 1)/f(x|p = 0'5) > k\}$$

es decir, cuando

$$x^{0'5}\Gamma\left(\frac{0'5}{2}\right) > k$$

o equivalentemente

$$RC = \{x > c\}$$

Como bajo la hipótesis nula puede determinarse c para que

$$P\{X > c|a = 0'5, p = 0'5\} = P\{\chi_1^2 > c\} = 0'05$$

el test óptimo de tamaño $0'05$ viene dado por

$$RC = \{x > 3'842\}.$$

5.8.12

a) El test de la RV para contrastar $H_0 : a = a_0$ frente a $H_1 : a \neq a_0$ viene dado por

$$RC = \{\lambda(x_1, ..., x_n) \leq k\}$$

con

$$\lambda(x_1, ..., x_n) = \frac{a_0^n e^{-a_i \sum_1^n x_i}}{\max_a a^n e^{-a_i \sum_1^n x_i}}$$

Para calcular el máximo del denominador, tomando logaritmos se obtiene

$$n \ln a - a \sum X_i$$

e igualando a 0 la derivada ha de ser $a = \overline{x}^{-1}$, que al particularizar en la segunda derivada y como

$$-\frac{n}{a^2}\Big|_{a=\overline{x}^{-1}} < 0$$

se sigue que

$$\lambda(x_1, ..., x_n) = \frac{a_0^n e^{-a_0 n \overline{x}}}{\overline{x}^{-n} e^{-n}} = (a_0 \overline{x})^n e^{-n(a_o \overline{x} - 1)}$$

y de acuerdo con la Figura 5.8.4 se obtiene

$$RC = \{\overline{x} \leq c_1 \quad \text{ó} \quad \overline{x} \geq c_2\}$$

donde las constantes han de satisfacer

$$\left(\frac{c_1}{c_2}\right)^n = e^{-n(a_0 c_2 - a_0 c_1)}$$

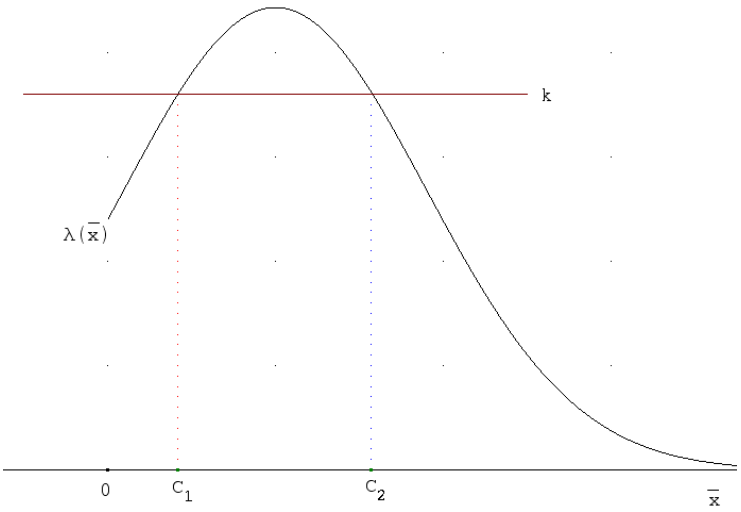


Figura 5.8.4. Representación de la RV correspondiente a la parte a) del Ejercicio 5.7.12.

b) El test de la *RV* para contrastar $H_0 : a \geq a_0$ frente a $H_1 : a < a_0$ vendrá dado por

$$RC = \{\lambda(x_1, ..., x_n) \leq k\}$$

con

$$\lambda(x_1, ..., x_n) = \frac{\underset{a \geq a_0}{\text{máx}}\ a^n e^{-n_n \overline{x}}}{\underset{a}{\text{máx}}\ a^n e^{-an\overline{x}}} = \frac{\widehat{a}_0^n e^{-\widehat{a}_0 n\overline{x}}}{\overline{x}^{-n} e^{-n}}$$
$$= \begin{cases} 1 & \text{si} \quad \overline{x}^{-1} \geq a_0 \\ \frac{a_0^n e^{-a_0 n\overline{x}}}{\overline{x}^{-n} e^{-n}} & \text{si} \quad \overline{x}^{-1} < a_0 \end{cases}$$

por lo que de acuerdo con la Figura 5.8.5, se obtiene

$$RC = \{\overline{x}^{-1} \leq c\}$$

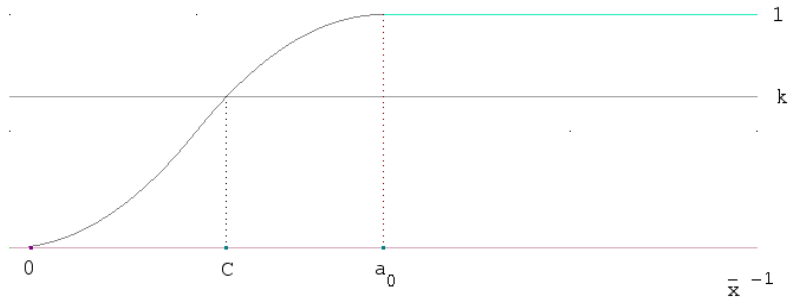


Figura 5.8.5. Representación de la RV correspondiente a la parte b) del Ejercicio 5.7.12.

5.8.13

Mediante el *Teorema 5.3.1 (Neyman-Pearson)* se debe calcular

$$\frac{P\{x_1, x_2|\theta = 0'5\}}{P\{x_1, x_2|\theta = 0'25\}} = 2^{x_1+x_2}\left(\frac{2}{3}\right)^{2-(x_1+x_2)}$$

Como la distribución de $Y = X_1 + X_2$ es una $Binomial(2, \frac{1}{4})$ bajo H_0 la región crítica del test es de la forma

$$RC = \{x_1 + x_2 \geq k\}$$

sin embargo es imposible hacer con ningún k que el tamaño del test sea $3/16$. Para lograrlo habría que recurrir a tests aleatorizados, es decir a tests que para algún valor de k, en este caso $k = 1$, se requiera de un sorteo que con probabilidad $1/3$ rechace la hipótesis nula y que con probabilidad $2/3$ acepte la hipótesis nula. Obsérvese que para tal test

$$\begin{aligned} P\{\text{rechazar } H_0|\theta = 0'25\} &= P\{X_1 + X_2 > 1|\theta = 0'25\} + \frac{1}{3}P\{X_1 + X_2 = 1|\theta = 0'25\} \\ &= \frac{1}{16} + \frac{1}{3}\frac{6}{16} \\ &= \frac{3}{16} \end{aligned}$$

No obstante no parece razonable el utilizar un contraste cuyo resultado depende de una aleatorización independiente del valor de θ, en este caso se emplearía de región crítica $\{x_1 + x_2 > 1\}$ que tiene de tamaño $\frac{1}{16}$ y no $\frac{3}{16}$.

5.8.14

Para contrastar $H_0 : X \sim N(0, 1)$ frente a $H_1 : X \sim Cauchy(0, 1)$, aplicando el *Teorema 5.3.1 (Neyman-Pearson*), el test óptimo tiene de región crítica

$$RC = \left\{ x \left| \frac{1}{\pi}\frac{1}{1+x^2} \geq k\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2} \right. \right\}$$

que puede ponerse alternativamente

$$RC = \{x| \ \ |x| \geq c\}$$

La constante se determina haciendo que $0'025 = P\{RC|H_0\}$ por lo que $c = 1'96$.

5.8.15

Se pretende contrastar $H_0 : X \sim N(0,1)$ frente a $H_1 : X \sim Gamma(1,1)$. Mediante el *Teorema 5.3.1 (Neyman-Pearson*), el test óptimo tiene la región crítica

$$RC = \left\{ x \left| e^{-x} \geq k \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \right. \right\}$$

o bien

$$RC = \{x|x \geq c \ \text{ con } \ c > 0\}$$

Para obtener la constante ha de ser $0'01 = P\{X \geq c|H_0\}$ por lo que $c = 2'33$.

La función de potencia es

$$\beta_{RC}(\theta) = P\{X \geq 2'33|H_1\} = \int_{2'33}^{+\infty} e^{-x} dx = e^{-2'33} = 0'097.$$

5.8.16

a) Si se fijan $\theta_1 > \theta_0$, el cociente

$$\frac{f(x_1, ..., x_n|\theta_1)}{f(x_1, ..., x_n|\theta_0)} = \left(\frac{\theta_1}{\theta_0}\right)^n e^{\sum_{i=1}^{n} x_i(\theta_1 - \theta_0)}$$

es una función monótona decreciente en $\sum_{i=1}^{n} x_i$, luego aplicando la versión adecuada del *Teorema 5.4.1 (Karlin-Rubin*), el test óptimo tiene de región crítica

$$RC = \left\{ \sum_{i=1}^{n} x_i \leq c \right\}$$

Para determinar la constante se utiliza que

$$0'01 = P\left\{ \sum_{i=1}^{n} x_i \leq c|H_0 \right\}$$

y por la reproductividad de la v.a. exponencial $\sum_{i=1}^{n} X_i \sim Gamma(a = \theta, p = n)$ de donde $2\theta \sum_{i=1}^{n} X_i \sim \chi^2_{2n}$ con lo que $c = \frac{1}{2\theta_0} \chi^2_{2n;1-\alpha}$.

b) Para θ_1 y θ_0 con $\theta_1 > \theta_0$, el cociente

$$\frac{f(x_1, ..., x_n|\theta_1)}{f(x_1, ..., x_n|\theta_0)} = \left(\frac{\theta_1}{\theta_0}\right)^n e^{\sum_{i=1}^{n} \ln x_i(\theta_1 - \theta_0)}$$

tiene RVM en $\sum_{i=1}^{n} \ln x_i$, luego mediante el *Teorema 5.4.1 (Karlin-Rubin)*, el test óptimo para contrastar $H_0 : \theta \leq \theta_0$ frente a $H_1 : \theta > \theta_0$ tiene de región crítica

$$RC = \left\{ \sum_{i=1}^{n} \ln x_i \geq c \right\}$$

Para determinar la constante se utiliza que la v.a. $-\ln X \sim Gamma(a = \theta, p = 1)$ luego $-\sum_{i=1}^{n} \ln X_i \sim Gamma(a = \theta, p = n)$ y por lo tanto $-2\theta \sum_{i=1}^{n} \ln X_i \sim \chi^2_{2n}$ con lo que $c = -\frac{1}{2\theta_0}\chi^2_{2n;1-\alpha}$.

5.8.17

a) Se pretende contrastar $H_0 : \mu = 30$ cm frente a $H_1 : \mu \neq 30$ cm para una población $N(\mu, \sigma)$ con σ desconocida. Mediante el método de la RV (véase la Tabla 5.2.2.1) ha de ser la región crítica del test

$$RC = \left\{ |\overline{x} - \mu_0| \geq \frac{S}{\sqrt{n}} t_{n-1;\frac{\alpha}{2}} \right\}$$

a partir de los datos $S^2 = \frac{1}{49}\sum_{i=1}^{50}(x_i - \overline{x})^2 = 11'7347$, $t_{49;0'005} = 2'6804$ con lo que la región crítica es $RC = \{|\overline{x} - 30| \geq 1'3117\}$. Como $|29 - 30| = 1$ no pertenece a la RC se acepta la hipótesis nula.

b) Se trata de contrastar $H_0 : \mu_1 - \mu_2 = 0$ frente a $H_1 : \mu_1 - \mu_2 \neq 0$. Como en este caso las dos varianzas son desconocidas se contrasta primero si ambas son iguales. Para contrastar $H_0 : \sigma_1 = \sigma_2$ frente a $H_1 : \sigma_1 \neq \sigma_2$, la región crítica es (véase la Tabla 5.2.2.4)

$$RC = \left\{ \frac{S_1^2}{S_2^2} \geq f_{m-1,n-1;\frac{\alpha}{2}} \right\}$$

ya que $S_1^2 \geq S_2^2, 11'7347 \geq 8'2278$. Cómo es $f_{49,79;0'025} = 1'6383$ y $S_1^2/S_2^2 = 1'4262$ no es mayor que $1'6383$ se acepta la igualdad de varianzas.

Para contrastar ahora la igualdad de medias, la región crítica es (véase la Tabla 5.2.2.3)

$$RC = \left\{ |\overline{x} - \overline{y}| \geq t_{m+n-2;\frac{\alpha}{2}} S_p \sqrt{\frac{1}{m} + \frac{1}{n}} \right\}$$

con la cuasivarianza ponderada

$$S_p^2 = \frac{(m-1)S_1^2 + (n-1)S_2^2}{m+n-2} = \frac{99 \cdot 11'7347 + 79 \cdot 8'2278}{178} = 10'1783$$

de donde $S_p = 3'1903$. Es $t_{178;0'025} = 1'9738$ el cuantil de orden $0'975$ de la distribución T_{178} y $\sqrt{\frac{1}{m}+\frac{1}{n}} = 0'15$. Por lo que la región crítica es

$$RC = \{|\overline{x} - \overline{y}| \geq 0'473\}$$

y como $|\overline{x} - \overline{y}| = 7$ se rechaza la hipótesis de igualdad de medias, luego el proceso ha variado.

5.8.18

Si se denota por $X_i = 1$ si la persona i-ésima de las obtenidas en la ciudad A gana menos de 1,800 euros al mes y análogamente $Y_j = 1$ respectivamente de las obtenidas en la ciudad B, se pueden considerar sendas variables:

$$\sum_{i=1}^{300} X_i \sim Binomial(m = 300, p_1)$$
$$\sum_{i=1}^{300} Y_j \sim Binomial(n = 200, p_2)$$

Se pretende contrastar la hipótesis nula $H_0 : p_1 = p_2 = p$ frente a $H_1 : p_1 \neq p_2$. Una posibilidad para realizar este contraste es emplear que $\overline{X} \sim N\left(p_1, \sqrt{\frac{p_1(1-p_1)}{m}}\right)$ $\overline{Y} \sim N\left(p_2, \sqrt{\frac{p_2(1-p_2)}{n}}\right)$ y sustituir en las expresiones de las desviaciones típicas p_1 por $\overline{x}$ y p_2 por $\overline{y}$. El problema con este procedimiento es que la normalidad es asintótica por lo que conceptualmente es más correcto utilizar el método de la RV para contrastar $H_0 : p_1 = p_2$ frente a $H_1 : p_1 \neq p_2$, lo que lleva a la región crítica

$$RC = \left\{ \frac{\max\limits_{p} p^{\sum_1^m x_i + \sum_1^n y_j} (1-p)^{m+n-\sum_1^m x_i - \sum_1^n y_j}}{\max\limits_{p_1,p_2} p_1^{\sum_1^m x_i} (1-p_1)^{m-\sum_1^m x_i} p_2^{\sum_1^n y_j} (1-p_2)^{n-\sum_1^n y_j}} \leq k \right\}$$

Es inmediato comprobar que

$$\widehat{p}_1 = \overline{x},\ \widehat{p}_2 = \overline{y},\ \widehat{p} = \frac{m\overline{x} + n\overline{y}}{m+n}$$

Mediante la distribución asintótica de $-2\ln\lambda\left(\sum_{i=1}^m x_i, \sum_{j=1}^n y_j\right)$ (véase el Teorema 5.2.3.2) se obtiene la convergencia a una distribución χ_1^2 por lo que la región crítica es

$$RC = \left\{ -2\ln\lambda\left(\sum_{i=1}^m x_i, \sum_{j=1}^n y_j\right) \geq c \right\}$$

donde $c = \chi^2_{1;\alpha}$, si se sustituye $\widehat{p}_1 = 0'31$, $\widehat{p}_2 = 0'26$ y $\widehat{p} = 0'29$, se tiene que la región crítica es

$$RC = \left\{ -2 \ln \lambda \left(\sum_{i=1}^{m} x_i, \sum_{j=1}^{n} y_j \right) \geq 3'842 \right\}$$

como $-2 \ln \lambda \left(\sum_{i=1}^{m} x_i, \sum_{j=1}^{n} y_j \right) = 1'4702$ se acepta la hipótesis nula de que la proporción de individuos que ganan menos de 1,800 euros al mes es la misma en ambas ciudades.

5.8.19

Dado un $\overrightarrow{x} = (x_1, ..., x_n)$ se rechaza la hipótesis nula H_0 cuando $P\{H_0|\overrightarrow{x}\} < P\{H_1|\overrightarrow{x}\}$. Como es

$$P\{\theta{=}1|\overrightarrow{x}\} = \frac{2/3 \ e^{-\sum x_i}}{2/3 \ e^{-\sum x_i} + 1/3 \ 2^n \ e^{-2\sum x_i}}$$

$$P\{\theta{=}2|\overrightarrow{x}\} = \frac{2/3 \ 2^n \ e^{-2\sum x_i}}{2/3 \ e^{-\sum x_i} + 1/3 \ 2^n \ e^{-2\sum x_i}}$$

$P\{\theta = 1|\overrightarrow{x}\} < P\{\theta = 2|\overrightarrow{x}\}$ es equivalente a

$$2/3 \ e^{-n\overline{x}} < 1/3 \ 2^n \ e^{-2n\overline{x}}.$$

Aunque mediante el método bayesiano se debe actuar condicionalmente a la muestra $\overrightarrow{x}$ observada, se mantiene la misma notación que en el caso clásico y se rechazará para todos aquellos puntos que cumplan

$$RC = \left\{ \overrightarrow{x} \left| \overline{x} < \frac{n-1}{n} \ln 2 \right. \right\}.$$

5.8.20

Se trata de constratar $H_0 : \theta \leq 100$ frente a $H_1 : \theta > 100$. Para el cálculo de la distribución final, se sabe que si la población se distribuye $N(\theta, \sigma)$ y la distribución inicial sobre θ es $N(\mu_0, \sigma_0)$, con μ_0, σ y σ_0 conocidas, la distribución final es $N(\mu_1, \sigma_1)$ con

$$\mu_1 = \frac{\frac{\mu_0}{\sigma_0^2} + \frac{\overline{x}}{\sigma^2/n}}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2/n}}$$

$$\sigma_1 = \frac{1}{\left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2/n} \right)^{1/2}}$$

(véase el Ejemplo 3.2.3.3).

Por lo tanto $\pi(\theta|115)$ es $N(110'38, 8'32)$ en este caso $n = 1 \quad \overline{x} = 115$, de manera que la probabilidad final sobre la hipótesis nula es

$$\begin{aligned} P\{H_0|x = 115\} &= P\{\theta \leq 100|x = 115\} \\ &= P\{Z \leq -1'248\} \\ &= 0'106 \end{aligned}$$

por lo que se rechaza la hipótesis nula.

Mediante el test clásico, si no se utiliza la información de que en el colectivo $\theta \sim N(100, 15)$, utilizando el test de la RV, para contrastar $H_0 : \theta \leq 100$ frente a $\mathrm{H}_1 : \theta > 100$, se tiene (véase la Tabla 5.2.2.1) que la región crítica es

$$RC = \left\{ \overline{x} \geq 100 + \frac{\sigma}{\sqrt{n}} z_\alpha \right\}$$

como se desea el p-valor correspondiente a $x = 115$

$$p - valor(115) = \sup_{\theta \leq 100} P\{X \geq 115|\theta\}$$

de donde

$$\begin{aligned} p - valor(115) &= P\{X \geq 115|100\} \\ &= P\left\{ Z \geq \frac{115 - 100}{10} \right\} \\ &= P\{Z \geq 1'5\} \\ &= 0'07 \end{aligned}$$

luego si el tamaño del test clásico es menor que 0'07 se aceptaría la hipótesis nula de que el estudiante tiene un coeficiente de inteligencia menor o igual que 100.

5.8.21

a)

$$\begin{aligned} P\{-\infty < \theta < \infty\} &= P\{\theta = 0\} + (1 - P\{\theta = 0\}) \int_{-\infty}^{+\infty} \pi(\theta) d\theta \\ &= 0'5 + 0'5 = 1. \end{aligned}$$

b) La probabilidad final de $H_0 : \theta = 0$ es

$$\begin{aligned} P\{\theta = 0|x_1, ..., x_n\} &= \frac{P\{\theta=0\} f(\overrightarrow{x}|0)}{P\{\theta=0\} f(\overrightarrow{x}|0) + P\{\theta \neq 0\} \int_{\theta \neq 0} \pi(\theta) f(\overrightarrow{x}|\theta) d\theta} \\ &= \frac{0'5 \ f(\overrightarrow{x}|0)}{0'5 f(\overrightarrow{x}|0) + 0'5 \int_{-\infty}^{+\infty} \pi(\theta) f(\overrightarrow{x}|\theta) d\theta} \\ &= \left(1 + \frac{1}{\sqrt{2\pi}\tau_0} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left[\left(\frac{1}{\tau_0^2} + \frac{1}{\sigma^2/n} \right) \theta^2 - 2 \frac{\overline{x}}{\sigma^2/n} \theta \right]} d\theta \right)^{-1} \end{aligned}$$

Esta integral se calcula obervando que es de la misma forma que la integral de la Expresión (3.2.3.2) en el Ejemplo 3.2.3.3, con

$$E = \left(\frac{n\overline{x}}{\sigma^2}\right)\left(\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}\right)^{-1}$$
$$V = \left(\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}\right)^{-1}$$

Así se obtiene que

$$P\{\theta = 0|x_1, ..., x_n\} = \left[1 + \left(1 + \frac{n}{\sigma^2}\tau_0^2\right)^{-1/2} \exp\frac{1}{2}\left(\frac{n\overline{x}}{\sigma^2}\right)^2 \left(\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}\right)^{-1}\right]^{-1}$$

c) Sin información inicial el test clásico mediante la RV para contrastar $H_0 : \theta = 0$ frente a $H_1 : \theta \neq 0$ viene dado por

$$RC = \{|\overline{X}| \geq k\}$$

(véase la Tabla 5.2.2.1), luego el $p\text{-}valor(\overrightarrow{x})$ es

$$\begin{aligned} p\text{-}valor(\overrightarrow{x}) &= P\{|\overline{X}| \geq |\overline{x}| \ \mid \ 0\} \\ &= 2(1 - \Phi(\sqrt{n}|\overline{x}|)) \end{aligned}$$

con Φ la función de distribución de la v.a. $N(0, 1)$.

d) Al particularizar $\sigma^2 = \tau_0^2 = 1$ y $n = 10$ en las expresiones de la probabilidad final y del p-valor, se obtiene

$$P\{\theta = 0|x_1, ..., x_{10}\} = [1 + (1 + n)^{-1/2} \exp(10\overline{x})^2(1 + n)^{-1}]^{-1}$$

lo que lleva a la Tabla 5.8.1, en la que se observa que la probabilidad final es siempre mayor que el p-valor, siendo esta diferencia bastante alta para valores moderados de $\overline{x}$.

$\overline{x}$	0	0'1	0'4	0'6	1	2
$p\text{-}valor(\overrightarrow{x})$	1	0'7518	0'2059	0'0578	0'0016	0
$P(\theta = 0\|\overrightarrow{x})$	0'7683	0'7602	0'6158	0'3924	0'034	0

Tabla 5.8.1. Valores del p-valor y la probabilidad final correspondiente al Ejercicio 5.7.21.

Capítulo 6

TEORÍA DE LA DECISIÓN

6.1. INTRODUCCIÓN

La estimación por punto, la estimación por intervalo y el contraste de hipótesis pueden considerarse como casos particulares de una teoría más general, la Teoría de la Decisión. En esta teoría se vuelven a estudiar los problemas de inferencia, en los que todos los elementos que están incluidos en el proceso de inferencia son formalmente definidos, y además se introducen y se desarrollan también los criterios de optimalidad deseables que deben cumplir los procedimientos que se utilizan y que incluyen, como casos particulares, a los que ya se han estudiado.

Desde este punto de vista, en la Teoría de la Decisión se obtienen conclusiones similares a las obtenidas en los capítulos anteriores pero, en algunas situaciones, se obtendrán, nuevas interpretaciones de los procedimientos ya vistos.

En la formulación de un modelo de decisión intervienen, los datos, descritos por un vector aleatorio $\overrightarrow{X} = \overrightarrow{x}$ que recorre el espacio muestral $\mathcal{X}$, el modelo, formado por todas las distribuciones de probabilidad sobre $\mathcal{X}$ y caracterizadas por un parámetro θ que recorre el espacio paramétrico Θ, las acciones a que son los elementos entre los que se puede escoger y que recorren un conjunto $\mathcal{A}$ y finalmente la función de pérdida $L(\theta, a)$ en que se incurre cuando el parámetro es θ y se elige la acción a, con $L(\theta, a) \in \mathbb{R}$.

Ya que el valor muestral $\overrightarrow{x}$ debe dar alguna información sobre el parámetro θ, se introducen las reglas de decisión $d(\overrightarrow{x})$ como las aplicaciones del espacio muestral en el conjunto de acciones, $\mathcal{D} = \{d : \mathcal{X} \to \mathcal{A}\}$.

Ejemplo 6.1.1
Una fábrica produce durante un periodo de tiempo N elementos, n de los cuales son muestreados aleatoriamente para determinar si son correctos o defectuosos, se pretende determinar θ la proporción de defectuosos que produce la fábrica. El lote de N elementos es vendido por un precio total de C unidades monetarias, con la garantia de que si la proporción de elementos

defectuosos excede a θ_0, la fábrica pagará al comprador una cantidad de G unidades monetarias por los inconvenientes causados. Para la colección de N elementos, el fabricante tiene dos acciones posibles: vender el lote con lo que recibe C unidades monetarias, o rectificarlo lo que tiene un coste de R unidades.

En este ejemplo $\mathcal{X} = \{0, 1, ..., n\}$ es el número de elementos defectuosos observados y el modelo viene dado por la distribución $Binomial(n, \theta)$

$$P\{x|\theta\} = \binom{n}{x}\theta^x(1-\theta)^{n-x} \quad \text{con} \quad x \in \mathcal{X}$$

donde $\Theta = [0, 1]$. El conjunto de acciones posibles es $\mathcal{A} = \{a_1, a_2\}$ siendo a_1 vender el lote y a_2 rectificarlo. La función de pérdida, en unidades monetarias, viene dada por la Tabla 6.1

$L(\theta, a)$	a_1	a_2
$\theta \leq \theta_0$	$-C$	R
$\theta > \theta_0$	$G - C$	R

Tabla 6.1. Función de pérdidas correspondiente al Ejemplo 6.1.1.

donde en el caso de tomar la acción a_1 cuando $\theta \leq \theta_0$ se produce una ganancia positiva de C u.m. por lo que se ha expresado el beneficio como una pérdida negativa.

Las reglas de decisión d son el conjunto $\mathcal{D}$ de aplicaciones del conjunto $\{0, 1, ..., n\}$ en $\{a_1, a_2\}$, y pueden ser expresadas formalmente como una colección de $n+1$ elementos en la que cada uno de los elementos es a_1 o a_2, donde se interpreta el valor i-ésimo de la colección como la acción a tomar cuando se observan x éxitos. Sin embargo parece natural considerar sólo las reglas de decisión que se aplican sobre a_1 cuando el número de elementos defectuosos queda menor o igual que x y sobre a_2 cuando dicho número es estrictamente mayor que x.

■

La Teoría de la Decisión es una teoría matemática para la toma de decisiones cuando existe incertidumbre. Así en su acepción más general, en la Teoría de la Decisión se estudian formas de determinar la función de pérdida, más bien su opuesta que es la función de utilidad, y de determinar *buenas* reglas de decisión. No obstante, este Capítulo, por ser una introducción, se va a restringir sólo al estudio de la Teoría de la Decisión Estadística, viendo como se pueden plantear en este contexto los problemas de estimación por punto, por regiones de confianza y los contrastes de hipótesis y estudiando los conceptos de *buenas* reglas de decisión, que quedan recogidos en los conceptos de reglas minimax y reglas Bayes. Un ejemplo de aplicación a un contexto distinto puede verse en Gómez Villegas (1979).

La calidad de una regla de decisión se cuantifica mediante la *función de riesgo* que es la esperanza de la función de pérdida respecto a la distribución

$f(\overrightarrow{x}|\theta)$.

$$R(\theta, d) = E[L(\theta, d(\overrightarrow{X}))] \tag{6.1.1}$$

Lo deseable es escoger una regla de decisión, cuya función de riesgo sea pequeña para todo valor de θ, en general esto no es posible y ocurre frecuentemente que para dos reglas de decisión d_1 y d_2 se tiene que

$$R(\theta_1, d_1) < R(\theta_1, d_2)$$

pero

$$R(\theta_2, d_1) > R(\theta_2, d_2)$$

con lo que d_1 es mejor que d_2 cuando el valor del parámetro es θ_1, pero d_2 es mejor que d_1 cuando el valor del parámetro es θ_2, con lo que no es posible escoger entre ambas.

Las dos maneras usuales de resolver esta dificultad son mediante la *regla minimax* y la *regla Bayes*. Una regla de decisión d se dice *minimax* si se cumple

$$\min_{d\in\mathcal{D}} \max_{\theta\in\Theta} R(\theta, d) = \max_{\theta\in\Theta} R(\theta, d^*) \tag{6.1.2}$$

es decir la regla d^* es minimax si con ella la máxima pérdida en que se puede incurrir es lo más pequeña posible.

Debe tenerse en cuenta que estas expresiones no son estrictamente correctas ya que el mínimo y el máximo, en general, pueden no ser alcanzables sin restringir $\mathcal{D}$ y Θ. De todas maneras se dejarán estas cuestiones para un tratamiento más avanzado de la Teoría de la Decisión.

El otro procedimiento, consiste en introducir, cuando esto sea posible, una distribución inicial mediante $\pi(\theta)$ y calcular el *riesgo Bayes* correspondiente a dicha distribución inicial, que viene definido mediante

$$B(\pi, d) = E[R(\theta, d)] \tag{6.1.3}$$

donde la esperanza está calculada respecto a la distribución dada por π. Si existe la regla d^π que minimice el *riesgo Bayes,* a tal regla se la denomina *regla Bayes.*

El procedimiento seguido para llegar a la regla minimax es demasiado conservador en el sentido de que asocia a cada regla de decisión lo peor que puede ocurrir al variar el valor del parámetro θ, es decir es como si el parámetro actuara en contra del que toma la decisión. En la regla de decisión Bayes se actua en promedio respecto a θ mediante la distribución dada por π, lo que parece un método más adecuado de tratar la función de riesgo.

También es cierto que desde una descripción matemáticamente más rigurosa, se deben hacer hipótesis para que exista el riesgo Bayes en (6.1.3) y sea alcanzable. En particular, también se deben considerar en $\mathcal{D}$ sólo aquellas funciones de $\mathcal{X}$ en $\mathcal{A}$ para las que la esperanza (6.1.1) exista.

6.2. MÉTODOS DE INFERENCIA EN LA TEORÍA DE LA DECISIÓN

En lo que sigue se aplican las ideas introducidas de la Teoría de la Decisión a los problemas ya estudiado de la inferencia estadística, es decir al contexto de la estimación por punto, por intervalo y al contraste de hipótesis.

6.2.1. Estimación por punto

Un problema de estimación por punto puede considerarse como un problema de la Teoría de la Decisión en que el espacio de acciones $\mathcal{A}$ coincide con el espacio paramétrico Θ. Las reglas de decisión de $\mathcal{D}$ son las aplicaciones del espacio muestral $\mathcal{X}$ en Θ por lo que una regla de decisión $d(x)$ es realmente un estimador $\widehat{\theta}(x)$ tal y como ha sido considerado en el Capítulo 3. Si ahora se considera como función de pérdida la cuadrática, $L(\theta, d(x)) = (\theta - d(x))^2$, se tiene que de acuerdo con (6.1.1) la función de riesgo viene dada por

$$\begin{aligned} R(\theta, d) &= E[(\theta - d(X))^2] = E[(d(X) - E[d(X)] + E[d(X)] - \theta)^2] \\ &= Var_\theta[d] + (Sesgo_\theta d)^2 \\ &= Var_\theta[d] + b_\theta^2[d] \\ &= ECM_d(\theta) \end{aligned}$$

con lo que la función de riesgo para la regla de decisión d coincide con el error cuadrático medio del estimador $\widehat{\theta}(x) = d(x)$ (véase la Sección 3.3.1). De esta manera, con pérdida cuadrática un buen estimador será el que minimice la varianza y el sesgo simultaneamente, con lo que el criterio de encontrar estimadores que minimicen el ECM toma un significado preciso en el contexto de la Teoría de la Decisión al utilizar la función de pérdida cuadrática.

6.2.2. Estimación por regiones de confianza

Un problema de estimación por regiones de confianza, puede considerarse como un problema de decisión en el que el espacio de acciones $\mathcal{A}$, son todos los subconjuntos del espacio paramétrico. Las reglas de decisión $\mathcal{D}$ son las aplicaciones del espacio muestral $\mathcal{X}$ en partes de Θ, por lo que una regla de decisión $d(\overrightarrow{x})$ es realmente una región de confianza $S(\overrightarrow{x}) \subset \Theta$. Si se considera como función de pérdida

$$L(\theta, S(\overrightarrow{x})) = 1 - I_{S(\overrightarrow{x})}(\theta) = \begin{cases} 1 & \text{si} \quad S(\overrightarrow{x}) \not\ni \theta \\ 0 & \text{si} \quad S(\overrightarrow{x}) \ni \theta \end{cases} \tag{6.2.2.1}$$

es decir, se pierde 1 si la región no contiene a θ y no se pierde nada si la región contiene a θ, se sigue, que la función de riesgo es, de acuerdo con (6.1.1),

$$R(\theta, d) = E[L(\theta, S(\overrightarrow{X}))] = P\{S(\overrightarrow{X}) \not\ni \theta | \theta\}.$$

En este caso, la función de riesgo coincide con la probabilidad de que la región de confianza no cubra el valor del parámetro. Por lo tanto una buena regla de decisión, desde el punto de vista de la Teoría de la Decisión, será aquella que haga mínima la probabilidad citada. Así se obtiene la región UMA (uniformemente más ajustada) introducida en la Sección 5.5. Un resultado sorprendente es que no aparecen en general las regiones de confianza de nivel $1-\alpha$, sino que se obtiene directamente la probabilidad de recubrir valores erróneos del parámetro. Para el modelo normal, en cierto sentido si puede lograrse (véase el Ejercicio 6.7.11).

En el caso en que $\theta = \mathbb{R}$ o un subconjunto propio y las acciones sean los intervalos de Θ, parece natural emplear como función de pérdida

$$L(\theta, S) = a \text{ longitud}(S) - I_S(\theta) \tag{6.2.2.2}$$

donde a es una constante positiva que recoge el peso relativo que se quiera dar al hecho de que el intervalo tenga una amplitud pequeña respecto a que debe tener probabilidad alta de recubrir valores correctos del parámetro. En este caso la función de riesgo viene dada por

$$R(\theta, S) = a\ E_\theta[longitud(S(\overrightarrow{X})] - P\{S(\overrightarrow{X}) \ni \theta|\theta\}$$

La poca utilización, no obstante, de la función de pérdida dada en (6.2.2.2) es debida a los problemas de cuantificación de la constante a.

6.2.3. Tests de hipótesis

Los tests de hipótesis se pueden considerar como un problema de decisión, en el que el espacio paramétrico queda dividido en dos conjuntos Θ_0 y Θ_1, el conjunto de acciones es $\mathcal{A} = \{a_0, a_1\}$ donde a_0 es la acción de aceptar la hipótesis nula de que $\theta \in \Theta_0$ y a_1 es la acción de rechazar la hipótesis nula; y en el que la función de pérdida viene dada por la Tabla 6.2.1, para dos valores

	a_0	a_1
$\theta \in \Theta_0$	0	c_I
$\theta \in \Theta_1$	c_{II}	0

Tabla 6.2.1. Función de pérdida de un test de hipótesis.

c_I y c_{II} convenientes. Una regla de decisión viene caracterizada por el conjunto de puntos de $\mathcal{X}$ para los que se toma la acción a_1, es decir para los que se rechaza la hipótesis nula. La función de riesgo viene dada por

$$R(\theta, d) = c_I P\{(d(\overrightarrow{x}) = a_1|\theta\} \quad \text{si} \quad \theta \in \Theta_0$$
$$R(\theta, d) = c_{II} P\{(d(\overrightarrow{x}) = a_0|\theta\} \quad \text{si} \quad \theta \in \Theta_1$$

que puede ponerse como

$$R(\theta, d) = c_I P\{\text{error de tipo I}|\theta\} \quad \text{si} \quad \theta \in \Theta_0$$
$$R(\theta, d) = c_{II} P\{\text{error de tipo II}|\theta\} \quad \text{si} \quad \theta \in \Theta_1$$

Por lo tanto en el caso en que $c_I = c_{II} = 1$, elegir la regla de decisión que minimice la función de riesgo supone minimizar también las probabilidades de cometer error de tipo I y error de tipo II, cuando se usa la función de pérdida recogida en la Tabla 6.2.1.

Alternativamente, la función de riesgo se puede poner en función de la potencia del test, de acuerdo con la Definición 5.1.3, como

$$R(\theta, d) = c_I \beta(\theta) \quad \text{si} \quad \theta \in \Theta_0$$
$$R(\theta, d) = c_{II} \beta(\theta) \quad \text{si} \quad \theta \in \Theta_1.$$

6.3. OBTENCIÓN DE REGLAS BAYES Y DE REGLAS MINIMAX

La obtención de reglas Bayes frente a una distribución inicial dada mediante π, pasa por el cálculo de la distribución final $\pi(\theta|x)$. En general, facilita la obtención de la regla de decisión que minimice el riesgo Bayes definido por la Expresión (6.1.3), el trabajar a partir de la distribución final y la distribución predictiva.

Teorema 6.3.1
Si para cada $\overrightarrow{x} \in \mathcal{X}$ y $a \in \mathcal{A}$ se denota por

$$r(a|\overrightarrow{x}) = \int_\Theta L(\theta, a)\pi(\theta|\overrightarrow{x})d\theta$$

y para cada $x \in \mathcal{X}$ existe una acción $a^* \in \mathcal{A}$ tal que

$$r(a^*|\overrightarrow{x}) = \inf_{a \in \mathcal{A}} r(a|\overrightarrow{x})$$

entonces $d^\pi(\overrightarrow{x}) = a^*$ es la regla Bayes con respecto a π.

Demostración: de acuerdo con la definición de riesgo Bayes, Expresión (6.1.3)

$$B(\pi, d) = \int_\Theta R(\theta, d)\pi(\theta)d\theta$$
$$= \int_\Theta \left[\int_{\mathcal{X}} L(\theta, d(\overrightarrow{x}))f(\overrightarrow{x}|\theta)d\overrightarrow{x} \right] \pi(\theta)d\theta$$

utilizando que

$$f(\overrightarrow{x}|\theta)\pi(\theta) = \pi(\theta|\overrightarrow{x})m(\overrightarrow{x})$$

e invirtiendo el orden de integración

$$\begin{aligned} B(\pi, d) &= \int_{\mathcal{X}} \left[\int_{\Theta} L(\theta, d(\overrightarrow{x}))\pi(\theta|\overrightarrow{x})d\theta \right] m(\overrightarrow{x})d\overrightarrow{x} \\ &= \int_{\mathcal{X}} r(a|\overrightarrow{x})m(\overrightarrow{x})d\overrightarrow{x} \end{aligned}$$

Por las hipótesis del teorema, el ínfimo para cada x se alcanza en a^*, luego el ínfimo en $B(\pi, d)$ se obtiene en $d^\pi(\overrightarrow{x}) = a^*$ ya que $m(\overrightarrow{x})$ es no negativa. Con lo que $d^\pi(\overrightarrow{x}) = a^*$ es la regla Bayes.

■

El teorema permite, para cada $\overrightarrow{x}$ fijo, obtener el ínfimo del riesgo Bayes en lugar de en el espacio de funciones $\mathcal{D}$ en el espacio de acciones $\mathcal{A}$, problema mucho más sencillo.

Si la función de pérdida es la cuadrática $r(a|x)$ coincide con la pérdida final esperada que se introdujo en la Sección 3.3.2, con lo que otra vez el modelo de la Teoría de la Decisión engloba al modelo de estimación.

Como consecuencia, para función de pérdida cuadrática la regla Bayes es $d^\pi(\overrightarrow{x}) = E[\theta|\overrightarrow{x}]$. Si la función de pérdida es la diferencia en valor absoluto, la regla Bayes es la mediana de la distribución final $\pi(\theta|x)$.

El Teorema 6.3.1 también se aplica al contraste de hipótesis, así en el Capítulo 5 se introdujeron los contrastes bayesianos en función del cálculo de las probabilidades finales de la hipótesis nula y de la alternativa. A la luz de la Teoría de la Decisión se puede enunciar el siguiente resultado.

Corolario 6.3.1
En un problema de decisión de contraste de hipótesis, donde la función de pérdida es la recogida en la Tabla 6.2.1, la regla Bayes frente a la distribución inicial $\pi(\theta)$ es

$$d^\pi(\overrightarrow{x}) = \begin{cases} a_0 & \text{si} \quad P\{\Theta_0|\overrightarrow{x}\} \geq c_{II}/(c_I + c_{II}) \\ a_1 & \text{si} \quad P\{\Theta_0|\overrightarrow{x}\} \leq c_{II}/(c_I + c_{II}) \end{cases} \tag{6.3.1}$$

Demostración: aplicando el Teorema 6.3.1

$$\begin{aligned} r(a_0|\overrightarrow{x}) &= c_{II}P\{\Theta_1|\overrightarrow{x}\} \\ r(a_1|\overrightarrow{x}) &= c_I P\{\Theta_0|\overrightarrow{x}\} \end{aligned}$$

la regla Bayes es por tanto la recogida en la Expresión (6.3.1).

■

Ejemplo 6.3.1
Con m.a.s. de tamaño n de una población $N(\theta, \sigma)$, para contrastar $H_0 : \theta \leq \theta_0$ frente a $H_1 : \theta > \theta_0$, la regla Bayes frente a la distribución inicial

$\pi(\theta)$ dada por una distribución $N(\mu,\tau)$, con σ,μ y τ conocidas, y cuando la función de pérdida es la recogida en la Tabla 6.2.1, viene dada por

$$(6.3.2)\qquad d^{\pi}(x_1,...,x_n)=\begin{cases} a_0 & \text{si} \quad (\theta_0-\eta\mu-(1-\eta)\overline{x})/(\tau\sqrt{\eta})\geq z \\ a_1 & \text{si} \quad (\theta_0-\eta\mu-(1-\eta)\overline{x})/(\tau\sqrt{\eta})\leq z \end{cases}$$

con $\eta=\sigma^2/(n\tau^2+\sigma^2)$ y z el cuantil de orden $c_{II}/(c_I+c_{II})$ de la distribución $N(0,1)$.

En efecto, como la distribución final $\pi(\theta|x_1,...,x_n)$ es la distribución normal con

$$E[\theta|\overrightarrow{x}]=\eta\mu+(1-\eta)\overline{x}$$
$$V[\theta|\overrightarrow{x}]=\eta\tau^2$$

(véase el Ejemplo 3.2.3.2). Utilizando el Corolario 6.3.1 se tiene la forma de la regla Bayes. Además como

$$\begin{aligned} P\{\Theta_0|\overrightarrow{x}\}&=P\{\theta\leq\theta_0|\overrightarrow{x}\} \\ &=P\left\{Z\leq\frac{\theta_0-\eta\mu-(1-\eta)\overline{x}}{\tau\sqrt{\eta}}\right\}\end{aligned}$$

con $Z\sim N(0,1)$, según (6.3.1), la regla Bayes es $d^{\pi}(\overrightarrow{x})=a_0$ cuando

$$[\theta_0-\eta\mu-(1-\eta)\overline{x}]/(\tau\sqrt{\eta})\geq z$$

con z el cuantil de orden $c_I/(c_I+c_{II})$ de la distribución $N(0,1)$.

En función de la muestra, la regla Bayes rechaza la hipótesis nula si

$$(6.3.3)\qquad \overline{x}\geq\frac{\theta_0-\eta\mu-z_{\alpha}\tau\sqrt{\eta}}{1-\eta}$$

El punto crítico, por encima del cuál se rechaza y por debajo se acepta, depende de las suposiciones iniciales y del valor α, que aquí es función de las constantes c_I y c_{II} , que determinan la función de pérdida. En el caso particular en que $c_I=c_{II}$ y $\mu=\theta_0$ se tiene que la Expresión (6.3.3) toma la forma $\overline{x}>\theta_0$, pues $\alpha=0'5$ y $z_{\alpha}=0$.

■

La obtención en casos generales de reglas de decisión minimax, es un problema bastante difícil, ya que, hay que calcular un máximo de un mínimo de una función de dos variables, no obstante se pueden encontrar, en determinados contextos, condiciones suficientes que aseguren que una regla es minimax.

Teorema 6.3.2
Si una regla de decisión d^{π} es Bayes respecto a la distribución inicial dada por $\pi(\theta)$ y tiene riesgo constante como función de θ (es una regla igualizadora), entonces es una regla minimax.

Demostración: si d^π no fuera minimax, existiría una regla $d' \in \mathcal{D}$ tal que

$$\sup_{\theta\in\Theta} R(\theta, d') < \sup_{\theta\in\Theta} R(\theta, d^\pi).$$

El riesgo Bayes frente a π correspondiente a d' cumple

$$\begin{aligned} B(\pi, d') &\leq \int_\Theta \sup_\theta R(\theta, d')\pi(\theta)d\theta \\ &< \int \sup_\theta R(\theta, d^\pi)\pi(\theta)d\theta \end{aligned}$$

por ser d^π igualizadora. Esta última expresión coincide con

$$\int_\theta R(\theta, d^\pi)\pi(\theta)d\theta = B(\pi, d^\pi)$$

lo que contradice el que d^π es Bayes frente a π.
■

Ejemplo 6.3.2
En una población de $Benoulli(\theta)$ con m.a.s. de tamaño n y función de pérdida

$$L(\theta, a) = \frac{(\theta - a)^2}{\theta(1-\theta)}$$

la regla de decisión $\overline{X}$ es igualizadora, ya que

$$R(\theta, \overline{X}) = E\left[\frac{(\overline{X} - \theta)^2}{\theta(1-\theta)}\right] = \frac{V[\overline{X}]}{\theta(1-\theta)} = \frac{\sigma^2}{n\theta(1-\theta)} = \frac{\theta(1-\theta)}{n\theta(1-\theta)} = \frac{1}{n}.$$

Además $\overline{X}$ es Bayes frente a $\pi(\theta) = U(0,1)$ ya que la distribución final viene dada por

$$\pi(\theta|\overrightarrow{x}) \sim Beta\left(\sum_1^n x_i + 1, n - \sum_1^n x_i + 1\right)$$

y por tanto minimizar en a la expresión

$$\frac{1}{B\left(\sum_1^n x_i + 1, n - \sum_1^n x_i + 1\right)} \int_0^1 \frac{(\theta - a)^2}{\theta(1-\theta)} \theta^{\sum_1^n x_i} (1-\theta)^{n - \sum_1^n x_i} d\theta$$

es equivalente a minimizar en a

$$\int_0^1 (\theta - a)^2 \theta^{\sum_1^n x_i - 1} (1-\theta)^{n - \sum_1^n x_i - 1} d\theta$$

expresión, que, a menos de una constante independiente de θ, es la desviación cuadrática a una constante, a, de una variable $\theta \sim Beta\left(\sum_1^n x_i, n - \sum_1^n x_i\right)$, que se hace mínima cuando es $a = \sum_1^n x_i \Big/ \left(\sum_1^n x_i + n - \sum_1^n x_i\right) = \overline{x}$, la media de las observaciones.

6.4. REGLAS ADMISIBLES Y CLASES COMPLETAS

En un problema de decisión, la clase de las reglas de decisión $\mathcal{D}$ a utilizar es demasiado amplia como para ser manejable. Lo que se hace es restringir la clase $\mathcal{D}$ a una clase más pequeña, en la esperanza de que sea más fácil elegir en ella una regla óptima.

En esta sección se introduce el concepto de *admisibilidad*, en la línea de poder desechar algunas reglas de decisión, y el concepto de *clase completa* de reglas de decisión, para asegurarse de que no se han desechado demasiadas reglas.

Definición 6.4.1 (regla de decisión admisible)
Una regla de decisión $d \in \mathcal{D}$ se dice *admisible* si no existe otra regla de decisión d' de $\mathcal{D}$ tal que

$$R(\theta, d') \leq R(\theta, d) \text{ para todo } \theta \in \Theta$$
$$R(\theta', d') < R(\theta', d) \text{ para al menos un } \theta' \in \Theta.$$

La idea intuitiva es que las reglas admisibles no son mejoradas de manera estricta, para algún valor del parámetro, por ninguna otra regla respecto a la función de riesgo.
■

Como ya se ha dicho en la introducción a este Capítulo, ocurre con frecuencia que para determinados valores del parámetro, la función de riesgo correspondiente a d' es menor o igual que la función de riesgo correspondiente a d, mientras que para otros valores el comportamiento se invierte.

Definición 6.4.2 (clase completa)
Una subclase C de reglas de decisión de $\mathcal{D}$ se dice *completa* si para cualquier regla de decisión $d' \notin C$ existe una regla $d \in C$ tal que

$$R(\theta, d) \leq R(\theta, d') \text{ para todo } \theta \in \Theta$$
$$R(\theta', d) < R(\theta', d') \text{ para al menos un } \theta' \in \Theta.$$

El siguiente enunciado recoge la relación entre reglas de decisión admisibles y clases completas de reglas de decisión.

Teorema 6.4.1
Si C es una clase completa de reglas de decisión, entonces C contiene a las reglas de decisión admisibles.

Demostración: Si d es una regla de decisión admisible y $d \notin C$ por ser C clase completa, existirá $d' \in C$ con $R(\theta, d') \leq R(\theta, d)$ para todo θ con algún valor θ' para el que la desigualdad será estricta, con lo que d no sería regla admisible.
■

El resultado siguiente establece condiciones bajo las cuales las reglas Bayes son admisibles, sin afinar demasiado puede decirse que las reglas Bayes son admisibles. Como por el Teorema 6.4.1 las reglas admisibles están contenidas en cualquier clase completa, resulta que las reglas Bayes están contenidas usualmente en cualquier clase completa, sería deseable que las reglas Bayes formaran clase completa, entre otros motivos porque de acuerdo con el Teorema 6.3.1 se dispone de un procedimiento constructivo relativamente sencillo para la obtención de tales reglas. De todas formas lo que falta por añadir al conjunto de reglas Bayes para obtener una clase completa son determinados límites de reglas Bayes. Resultados en esta línea pueden consultarse en Berger (1985) Capítulo 8.

Teorema 6.4.2
Se considera un problema de decisión en el que $\Theta \subset \mathbb{R}$ y, para cualquier regla de decisión, la función de riesgo, como función de θ, es continua. Si d^π es una regla Bayes frente a la distribución inicial de densidad $\pi(\theta)$ y para todo $\varepsilon > 0$ el intervalo $(\theta - \varepsilon, \theta + \varepsilon)$ tiene probabilidad positiva bajo π, entonces d^π es admisible.

Demostración: Si d^π no fuera admisible existiría una regla d tal que $R(\theta, d) \leq R(\theta, d^\pi)$ para todo $\theta \in \Theta$ y $R(\theta', d) < R(\theta', d^\pi)$ para al menos un θ'. Por la hipótesis de continuidad del enunciado y como la diferencia de funciones continuas es a su vez una función continua, se sigue que para la función $R(\theta, d^\pi) - R(\theta, d)$ si se denota por $\nu = R(\theta', d^\pi) - R(\theta', d) > 0$; dado $\frac{\nu}{2}$ existirá $\varepsilon > 0$ tal que si $|\theta - \theta'| \leq \varepsilon$ entonces

$$|R(\theta, d^\pi) - R(\theta, d) - (R(\theta', d^\pi) - R(\theta', d))| \leq \frac{\nu}{2}$$

Luego si $|\theta - \theta'| \leq \varepsilon$ se tiene que

$$\begin{aligned} R(\theta, d^\pi) - R(\theta, d) &\geq R(\theta', d^\pi) - R(\theta', d) - \tfrac{\nu}{2} \\ &= \tfrac{\nu}{2}. \end{aligned} \tag{6.4.1}$$

Integrando respecto a $\pi(\theta)$ se obtiene

$$B(\pi, d^\pi) - B(\pi, d) = \int_{|\theta-\theta'|\leq\varepsilon} (R(\theta, d^\pi) - R(\theta, d))\pi(\theta)d\theta + \int_{|\theta-\theta'|>\varepsilon} (R(\theta, d^\pi) - R(\theta, d))\pi(\theta)d\theta,$$

el segundo sumando es no negativo, por ser d^π no admisible, luego

$$B(\pi, d^\pi) - B(\pi, d) \geq \int_{|\theta-\theta'|\leq\varepsilon} (R(\theta, d^\pi) - R(\theta, d))\pi(\theta)d\theta$$

y por (6.4.1) si $|\theta - \theta'| \leq \varepsilon$

$$B(\pi, d^\pi) - B(\pi, d) \geq \frac{\nu}{2} \int_{|\theta-\theta'|\leq\varepsilon} \pi(\theta)d\theta$$

expresión, que es estrictamente positiva por hipótesis; luego se llega a contradicción con el hecho de que d^π era regla Bayes respecto a π. De aquí que d^π ha de ser admisible.

El mismo resultado se puede enunciar cuando Θ es discreto, en ese caso la regla de decisión debe ser Bayes con respecto a una distribución inicial, cuya función de masa $\pi(\theta)$ verifique $\pi(\theta) > 0$ para todo $\theta \in \Theta$.

A la distribución de probabilidad dada por π con esta característica, se dice que verifica la *regla de Cromwell*, esto es da probabilidad positiva a todo valor del parámetro, si éste es discreto, o a todo entorno de cualquier valor del parámetro, si éste es continuo.

■

La continuidad de la función $R(\theta, d)$ como función de θ, depende de la continuidad de la función de pérdida $L(\theta, a)$ y de la continuidad en θ de la función $f(x|\theta)$; condiciones precisas, pueden verse en Ferguson (1967) y Berger (1985). También puede estudiarse la relación entre reglas minimax y admisibilidad, como ejemplo se recoge el siguiente enunciado.

Teorema 6.4.3
Si d^*es una regla minimax y es única en el sentido de que si d es otra regla minimax es $R(\theta, d^*) = R(\theta, d)$ para todo $\theta \in \Theta$, entonces d^* es una regla de decisión admisible.

Demostración: Si d^* no fuera admisible existiría una regla d' tal que

$$R(\theta, d') \leq R(\theta, d^*) \text{ para todo } \theta$$
$$R(\theta', d') < R(\theta', d^*) \text{ para algún } \theta'$$

por lo tanto

$$\sup_\theta R(\theta, d') \leq \sup_\theta R(\theta, d^*)$$

si esta desigualdad fuera estricta, la regla d no sería minimax luego ha de cumplirse que

$$\sup_{\theta} R(\theta, d') = \sup_{\theta} R(\theta, d^*).$$

Pero entonces la regla d' también sería regla minimax y como existe al menos un valor de θ' tal que $R(\theta', d') < R(\theta', d")$ la regla minimax no sería única en el sentido establecido en el enunciado; luego la regla d^* tiene que ser admisible.

■

De todas formas el resultado no es de amplia aplicación, ya que suele ser difícil comprobar cuando se satisfacen las hipótesis del teorema.

6.5. PARADOJA DE STEIN

En esta sección se demuestra, que en el caso normal, para m.a.s. de tamaño n y función de pérdida cuadrática, el estimador $\overline{X}$ es admisible. Se prueba que en el caso en que se estimen varias medias de manera simultánea, con función de pérdida cuadrática, el estimador $\overline{X}$ es inadmisible.

En el Capítulo 3, se obtuvo que $\overline{X}$ es el $ECMV$ para θ cuando la población es $N(\theta, \sigma)$ con σ conocida. Además es el EMV y el estimador de mínimos cuadrados. Cuando se utiliza función de pérdida cuadrática no es estimador Bayes, pero es límite de reglas Bayes cuando la distribución inicial es $N(\mu, \tau)$ y τ tiende a $+\infty$.

Además $\overline{X}$ es una regla minimax, ya que si se toma como distribución inicial $\pi_n(\theta)$ la distribución $N(0, \tau_n)$ la regla Bayes d^{π_n} es la media de la distribución final que tiende a $\overline{X}$ cuando $\tau_n \to +\infty$, como se tiene que $B(\pi_n, d^{\pi_n}) \to \frac{\sigma^2}{n}$ cuando $\tau_n \to +\infty$ y $R(\theta, \overline{x}_n) = \frac{\sigma^2}{n}$ entonces $\overline{x}$ es minimax (véase el Ejercicio 6.7.19, con $p = \frac{\sigma^2}{n}$).

Tiene sentido preguntarse si, para $X_1, ..., X_n$ m.a.s. de una población $N(\theta, \sigma)$ con σ conocida y $L(\theta, d) = (\theta - d)^2$, es $\overline{X}$ admisible. La técnica de demostración que se va a seguir es general y se conoce como el método de Blyth pues fué introducida por este autor en 1951. La respuesta es afirmativa y su demostración se hace por reducción al absurdo de la siguiente manera.

Si $\overline{X}$ no fuera admisible exitiría una regla de decisión $d(X_1, ..., X_n)$ con

$$\begin{aligned} R(\theta, d) &\leq R(\theta, \overline{X}) \quad \text{para todo } \theta \\ R(\theta', d) &< R(\theta', \overline{X}) \quad \text{para al menos un valor } \theta' \end{aligned}$$

Por ser la función de densidad de la distribución $N(\theta, \sigma)$ y la función de pérdida cuadrática, funciones continuas en θ, la diferencia $R(\theta, \overline{X}) - R(\theta, d)$ es una función continua en θ; por lo tanto dada la constante estrictamente positiva $\frac{\nu}{2}$, con

$$R(\theta, \overline{X}) - R(\theta', d) = \nu$$

existirá ε positivo tal que si $|\theta - \theta'| \leq \varepsilon$, entonces

$$|R(\theta, \overline{X}) - R(\theta, d) - (R(\theta', \overline{X}) - R(\theta', d))| \leq \frac{\nu}{2}.$$

De manera que para todo θ en el intervalo $(\theta' - \varepsilon, \theta' + \varepsilon)$ se tiene que

$$\begin{aligned} R(\theta, \overline{X}) - R(\theta, d) &\geq R(\theta', \overline{X}) - R(\theta', d) - \tfrac{\nu}{2} \\ &= \tfrac{\nu}{2} \end{aligned} \tag{6.5.1}$$

si se emplea como distribución inicial para θ la distribución $N(0, \tau)$, con $\tau > 0$ conocida se sigue, integrando en los dos miembros de la Expresión (6.5.1) respecto a esta densidad, que

$$\begin{aligned} B(\pi, \overline{X}) - B(\pi, d) &= \int_{|\theta - \theta'| > \varepsilon} (R(\theta, \overline{X}) - R(\theta, d))\pi(\theta) d\theta \\ &\quad + \int_{\theta' - \varepsilon}^{\theta' + \varepsilon} (R(\theta, \overline{X}) - R(\theta, d))\pi(\theta) d\theta \\ &\geq \frac{\nu}{2} \int_{\theta' - \varepsilon}^{\theta' + \varepsilon} \frac{1}{\sqrt{2\pi}\tau} e^{-\frac{1}{2\tau^2}\theta^2} d\theta \end{aligned}$$

de donde se obtiene la acotación

$$\tau(B(\pi, \overline{X}) - B(\pi, d)) \geq \frac{\nu/2}{\sqrt{2\pi}} \int_{\theta' - \varepsilon}^{\theta' + \varepsilon} e^{-\frac{\theta^2}{2\tau^2}} d\theta. \tag{6.5.2}$$

Si, por otro lado, d^π es la regla Bayes respecto a la distribución $N(0, \tau)$, se tiene que

$$\begin{aligned} 0 \geq \tau(B(\pi, d^\pi) - B(\pi, d)) &= \tau(B(\pi, d^\pi) - B(\pi, \overline{X})) + \\ &\quad + \tau(B(\pi, \overline{X}) - B(\pi, d)) \end{aligned} \tag{6.5.3}$$

pero se comprueba que

$$\begin{aligned} B(\pi, \overline{X}) &= E_\pi E_{\overline{X}|\theta}(\overline{X} - \theta)^2 = E_\pi \left[\frac{\sigma^2}{n}\right] = \frac{\sigma^2}{n} \\ B(\pi, d^\pi) &= E_{m(\overrightarrow{x})} E_{\theta|\overrightarrow{x}}(d^\pi(\overrightarrow{x}) - \theta)^2 = E_{m(\overrightarrow{x})}[V[\theta|\overrightarrow{x}]] \end{aligned} \tag{6.5.4}$$

y como, de acuerdo con el Ejemplo 3.2.3.2, la distribución de $\theta|\overrightarrow{x}$ es $N((1-\eta)\overline{x}, \frac{1}{\sqrt{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}})$ con $\eta = \frac{\frac{1}{\tau^2}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}$, sustituyendo en la expresión (6.5.4) resulta

$$B(\pi, d^\pi) = E_{m(\overrightarrow{x})} \left(\frac{1}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}} \right) = \frac{1}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}} = \tau^2 \eta.$$

Volviendo a la Expresión (6.5.3)

$$\begin{aligned} 0 &\geq \tau(\tau^2\eta - \tfrac{\sigma^2}{n}) + \tfrac{\nu/2}{\sqrt{2\pi}} \textstyle\int_{\theta' - \varepsilon}^{\theta' + \varepsilon} e^{-\frac{\theta^2}{2\tau^2}} d\theta \\ &= -\tfrac{\frac{\sigma^2}{n}\frac{1}{\tau}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}} + \tfrac{\nu/2}{\sqrt{2\pi}} \textstyle\int_{\theta' - \varepsilon}^{\theta' + \varepsilon} e^{-\frac{\theta^2}{2\tau^2}} d\theta \end{aligned} \tag{6.5.5}$$

si se toman límites cuando $\tau \to \infty$, el primer sumando tiende a 0 y el integrando del segundo sumando tiende a 1, de donde se sigue que el segundo miembro de la Expresión (6.5.5) tiende a $\frac{\nu}{\sqrt{2\pi}}\varepsilon$ un número estrictamente positivo, con lo que se llega a una contradicción. Por lo tanto $\overline{X}$ ha de ser admisible.

La admisibilidad de $\overline{X}$ en el caso normal es cierta también aunque σ sea desconocida.

Como se ha visto la media muestral tiene buenas propiedades en el caso unidimensional. Se trata ahora de estimar de manera simultánea el vector de medias de varias poblaciones normales. Sean $X_i \sim N(\theta_i, 1) \quad i = 1, ..., p$ v.a.i. y se pretende estimar $\overrightarrow{\theta} = (\theta_1, ..., \theta_p)$ mediante una regla de decisión $\overrightarrow{d}(X_1, ..., X_p) = (d_1(X_1, ..., X_p), ..., d_p(X_1, ..., X_p))$ cuando la función de pérdida es

$$L(\overrightarrow{\theta}, \overrightarrow{d}(\overrightarrow{X})) = \sum_{i=1}^{p} (\theta_i - d_i(\overrightarrow{X}))^2. \tag{6.5.6}$$

Se tiene que la regla de decisión $d(\overrightarrow{X}) = (X_1, ..., X_p)$ es minimax y es inadmisible.

Una primera observación es que podría pensarse en la regla de decisión $d(\overrightarrow{X}) = (\overline{X}_1, ..., \overline{X}_p)$ donde $\overline{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$ con $X_{ij} \sim N(\theta_i, 1) \quad i = 1, ..., p$ y $j = 1, ..., n_i$ en lugar del caso más simple con $n_i = 1 \quad i = 1, ..., p$ que se ha enunciado. Con modificaciones obvias, el razonamiento que sigue es también válido en esa situación y únicamente se recoge con submuestras de tamaño 1 para simplificar los cálculos.

En efecto $d(\overrightarrow{X}) = (X_1, ..., X_p)$ es minimax. Para ello se va a utilizar que si se toman como distribuciones iniciales una sucesión $\pi_n(\overrightarrow{\theta})$ adecuada y se tiene que $B(\pi_n, d^{\pi_n}) \to p$, cuando $n \to +\infty$, entonces si una regla $d(\overrightarrow{X}) = (X_1, ..., X_p)$ cumple que $R(\theta, d) = p$ para todo $\theta = (\theta_1, ..., \theta_p)$ entonces es una regla minimax (véase el Ejercicio 6.7.19).

Se tiene así

$$\begin{aligned} R(\overrightarrow{\theta}, \overrightarrow{d}) &= E\left[\sum_{i=1}^{p} (\theta_i - d_i(\overrightarrow{X}))^2\right] \\ &= \int_{\mathcal{X}} ... \int_{\mathcal{X}} \sum_{i=1}^{p} (\theta_i - d_i(x_1, ..., x_p))^2 \prod_{i=1}^{p} f(x_i|\theta_i) dx_i \end{aligned} \tag{6.5.7}$$

donde $f(x_i|\theta_i)$ son las funciones de densidad correspondientes a las distribuciones $N(\theta_i, 1)$. Además, si la distribución inicial es $\pi(\overrightarrow{\theta}) = \prod_{i=1}^{p} \pi_i(\theta_i)$, con $\pi_i(\theta_i)$ funciones de densidad todas con distribución $N(0, \tau)$, el riesgo de la regla $\overrightarrow{d}$ frente a la distribución inicial π viene dado por

$$B(\pi, \overrightarrow{d}) = \int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} R(\overrightarrow{\theta}, \overrightarrow{d}) \pi_1(\theta_1) ... \pi_p(\theta_p) d\theta_1 ... d\theta_p \tag{6.5.8}$$

Dada la forma especial de las Expresiones (6.5.6) y (6.5.7) encontrar la regla Bayes $\overrightarrow{d}^{\pi}$ frente a π se reduce a encontrar, para estimar cada θ_i, con pérdida cuadrática de la forma $(\theta_i - d_i(x_1, ..., x_p))^2$, la regla Bayes frente a π_i. Esta regla es

$$d_i^{\pi}(X_1, ..., X_p) = d_i^{\pi_i}(x_i) = \frac{\tau^2}{\tau^2 + 1} X_i \quad i = 1, ..., p$$

y le corresponde el riesgo Bayes

$$B(\pi_i, d_i^{\pi_i}(X_i)) = \frac{\tau^2}{\tau^2 + 1} \qquad \text{i=1,...,p}$$

con lo que, mediante la Expresión (6.5.7) se obtiene que

$$B(\pi, d^{\pi}) = \sum_{i=1}^{p} \frac{\tau^2}{\tau^2 + 1} = \frac{p\tau^2}{\tau^2 + 1}. \tag{6.5.9}$$

Se tiene, entonces, que el riesgo Bayes correspondiente a d^{π} converge a p cuando $\tau \to +\infty$, pero como $R(\overrightarrow{\theta}, \overrightarrow{X}) = p$ se sigue, aplicando el Ejercicio 6.7.19, que la regla $\overrightarrow{X}$ es minimax.

Se va a demostrar ahora que, cuando $p \geq 3$, la regla de decisión

$$d_i^S(\overrightarrow{X}) = (1 - \frac{p-2}{\sum_{j=1}^{p} X_j^2}) X_i \qquad i = 1, ..., p \tag{6.5.10}$$

rebaja la función de riesgo correspondiente a la regla minimax $\overrightarrow{X}$. Esta regla d^S fué propuesta por James y Stein (1961).

Para la demostración se necesita el siguiente resultado.

Lema 6.5.1 (Stein)
Para una v.a. $X \sim N(\theta, \sigma)$ si g es una función derivable con $E|g'(X)| < +\infty$, entonces se tiene que

$$E\left[g(X)(X - \theta)\right] = \sigma^2 E\left[g'(X)\right]. \tag{6.5.11}$$

Demostración:

$$E\left[g(X)(X - \theta)\right] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} (x - \theta) g(x) e^{-\frac{1}{2}\frac{(x-\theta)^2}{\sigma^2}} dx$$

integrando por partes con

$$\begin{aligned} u &= g(x) & du &= g'(x) \\ dv &= (x - \theta) e^{-\frac{1}{2}\frac{(x-\theta)^2}{\sigma^2}} & v &= -\sigma^2 e^{-\frac{1}{2\sigma^2}(x-\theta)^2} \end{aligned}$$

se sigue que

(6.5.12)

$$\begin{aligned} E\left[g(X)(X-\theta)\right] &= \frac{1}{\sqrt{2\pi}\sigma}\left[-g(x)\sigma^2 e^{-\frac{1}{2}\frac{(x-\theta)^2}{\sigma^2}}\Big|_{-\infty}^{+\infty} + \right. \\ &\qquad \left. + \sigma^2 \int_{-\infty}^{+\infty} g'(x) e^{-\frac{1}{2}\frac{(x-\theta)^2}{\sigma^2}} dx\right] \\ &= \sigma^2 \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} g'(x) e^{-\frac{1}{2}\frac{(x-\theta)^2}{\sigma^2}} dx = \sigma^2 E\left[g'(X)\right]. \end{aligned}$$

La hipótesis $E|g'(X)| < +\infty$, es para asegurar que el primer sumando en la Expresión (6.5.12) es nulo.

■

La demostración de que la regla de decisión $d^s(X_1, ..., X_p)$ mejora a la regla minimax $\overrightarrow{X}$, pasa por el cálculo directo de la función de riesgo

$$R(\theta, d^S) = \sum_{i=1}^{p} E\left[\theta_i - (1 - \frac{p-2}{\sum_{j=1}^{p} X_j^2} X_i)\right]^2$$

$$= \sum_{i=1}^{p} E\left[\theta_i - X_i + \frac{(p-2)}{\sum_{j=1}^{p} X_j^2} X_i)\right]^2$$

que desarrollando el cuadrado puede escribirse como

$$\begin{aligned} R(\theta, d^S) &= \sum_{i=1}^{p}\left[E(\theta_i - X_i)^2 - 2E\left((\theta_i - X_i)\frac{(p-2)}{\sum_{j=1}^{p} X_j^2} X_i\right) + \right. \\ &\qquad \left. + E\left(\frac{(p-2)}{\sum_{j=1}^{p} X_j^2} X_i\right)^2\right] \\ &= \sum_{i=1}^{p} V[X_i|\theta_i] + 2\sum_{i=1}^{p} E\left((\theta_i - X_i)\frac{(p-2)}{\sum_{j=1}^{p} X_j^2} X_i\right) + \\ &\qquad + \sum_{i=1}^{p} E\left(\frac{(p-2)}{\sum_{j=1}^{p} X_j^2} X_i\right)^2 \end{aligned} \tag{6.5.13}$$

El primer sumando de esta expresión es igual a p, para el segundo suman-

do mediante el Lema de Stein se tiene que

$$(p-2)\sum_{i=1}^{p}\left(E(\theta_i - X_i)\frac{X_i}{\sum_{j=1}^{p} X_j^2}\right) = -(p-2)\sum_{i=1}^{p} E\left(\frac{\partial}{\partial X_i}\frac{X_i}{\sum_{j=1}^{p} X_j^2}\right)$$

el segundo miembro, puede entonces simplificarse mediante

$$\begin{aligned}
-(p-2)\sum_{i=1}^{p} E\left(\frac{\sum_{j=1}^{p} X_j^2 - 2X_i^2}{(\sum_{j=1}^{p} X_j^2)^2}\right) &= -(p-2)\sum_{i=1}^{p}\left\{E\left(\frac{\left(\sum_{j=1}^{p} X_j^2\right)}{\left(\sum_{j=1}^{p} X_j^2\right)^2}\right) - \right. \\
&\quad \left. -2E\left(\frac{X_i^2}{\left(\sum_{j=1}^{p} X_j^2\right)^2}\right)\right\} \\
&= -(p-2)\left\{pE\left(\frac{1}{\sum_{j=1}^{p} X_j^2}\right) - 2E\left(\frac{1}{\sum_{j=1}^{p} X_j^2}\right)\right\} \\
&= -(p-2)^2 E\left(\frac{1}{\sum_{j=1}^{p} X_j^2}\right)
\end{aligned}$$

Por último, el tercer sumando de la Expresión (6.5.13) puede ponerse como

$$(p-2)^2\sum_{i=1}^{p} E\left(\frac{X_i^2}{\left(\sum_{j=1}^{p} X_j^2\right)^2}\right) = (p-2)^2 E\sum_{i=1}^{p}\left(\frac{X_i^2}{\left(\sum_{j=1}^{p} X_j^2\right)^2}\right)$$

o bien como

$$(p-2)^2 E\left(\frac{1}{\sum_{j=1}^{p} X_j^2}\right)$$

Sustituyendo en la Expresión (6.5.13) estos tres sumandos, se sigue que

$$
\begin{aligned}
R(\theta, d^S) &= p - 2(p-2)^2 E\left(\frac{1}{\sum_{j=1}^{p} X_j^2}\right) + (p-2)^2 E\left(\frac{1}{\sum_{j=1}^{p} X_j^2}\right) \\
&= p - (p-2)^2 E\left(\frac{1}{\sum_{j=1}^{p} X_j^2}\right) < p = R(\theta, \overrightarrow{X})
\end{aligned} \tag{6.5.14}
$$

donde la última desigualdad solo es válida si $p \geq 3$, pues para $p = 1$ ó 2, es $E(1/\sum_{j=1}^{p} X_j^2) = +\infty$. Por lo tanto la regla $\overrightarrow{X}$ no es admisible.

No obstante para estimar valores de θ_i positivos, si la expresión $\sum_{j=1}^{p} X_j^2$ es próxima a cero el estimador de Stein, dado por la Expresión (6.5.10), tiene un comportamiento poco intuitivo, es muy negativo. Esto puede arreglarse introduciendo el estimador

$$
d_i^+(X_1, ..., X_p) = \left(1 - \frac{p-2}{\sum_{j=1}^{p} X_j^2}\right)^+ X_i \qquad \text{con} \quad i = 1, ..., p \tag{6.5.15}
$$

donde $x^+ = \text{máx}\{x, 0\}$. Brown (1971) ha demostrado que usando el estimador (6.5.15) se rebaja el riesgo correspondiente al estimador de Stein, Expresión (6.5.14), con lo que ni siquiera dicho estimador es admisible.

Como resumen se puede decir, con intuición geométrica, que si al estimador $\overrightarrow{X}$ se le encoge hacia el cero, que es lo que representa la Expresión (6.5.10), se consigue un estimador con riesgo menor, en el caso normal, con función de pérdida cuadrática y para dimensión mayor o igual que 3. Cuando se utiliza el planteamiento de un problema de decisión para abordar un problema con incertidumbre, es importante considerar lo que ocurre cuando se modifican tanto la distribución inicial cómo la función de pérdida. Algunos aspectos de estos problemas pueden verse en Gómez Villegas (1977), Gómez Villegas (1980) y Gómez Villegas (1983).

Como ya se ha dicho, el modelo de la Teoría de la Decisión puede aplicarse a situaciones muy generales, lo que puede obligar a generalizar distribuciones de probabilidad a nuevas situaciones, pueden verse ejemplos en Gómez y Gómez Villegas (1998) y Gómez, Gómez Villegas y Marín (1998). La aplicación a situaciones dependientes del tiempo puede verse en Gómez, Gómez Villegas y Marín (2002).

6.6. APROXIMACIÓN HISTÓRICA

La Teoría de la Decisión Estadística nace con los trabajos de Wald entre 1939 y 1952. Abraham Wald nació en 1902 en Cluj, Hungría, y murió en un accidente de aviación cuando iba a impartir unas conferencias a Travancore India, en 1950. De origen judío, se trasladó a Viena en 1927, donde obtuvo su doctorado en matemáticas en 1931. Inicialmente trabajó en el campo de la geometría y para aumentar sus ingresos fué tutor de un prominente banquero y economista vienés. Este trabajo le llevó a interesarse por la econometría, concretamente por las series temporales.

Cuando Austria fué ocupada por los nazis en 1738, Wald se trasladó a Estados Unidos, incorporándose a la *Corporación Carnegie*, y estudió estadística en la Universidad de Columbia con Harold Hotelling. En 1941 fué contratado como profesor de esta universidad a la que perteneció ya durante toda su vida. Sus primeros artículos en estadística fueron publicados en 1939. En ellos señala que los problemas de estimación y contraste de hipótesis no son sino casos particulares de la Teoría de Juegos, tal como ésta había sido desarrollada hasta el momento, principalmente por Von Neumann. La diferencia estaba en que en la Teoría de Juegos los oponentes tenían intereses encontrados, por lo que el criterio minimax era fundamental, mientras que en la Teoría de la Decisión el estadístico no parecía que debiera considerar como antagonista a la naturaleza, y por ello otro tipo de criterios parecían más adecuados. Wald definió la función de pérdida, la función de riesgo, las reglas minimax y demostró que una regla minimax tiene riesgo constante bajo ciertas condiciones de regularidad. Introdujo las reglas Bayes y las reglas de decisión admisibles, siendo paradójico que las introdujo para deducir teoremas de admisibilidad y para construir reglas minimax.

En 1950 recoge sus contribuciones a la Teoría de la Decisión en un libro titulado *Las funciones de Decisión Estadística*, de un alto nivel matemático. También en este año publica su otra gran contribución estadística, el análisis secuencial, obteniendo la distribución para muestras grandes del test secuencial y su propiedad de óptimo. Sus contribuciones no se agotan con esto, ya que también consiguió resultados en inferencia no paramétrica, procesos estocásticos y diseño de experimentos.

Wald era un excelente profesor, claro y preciso, y un maestro en obtener demostraciones elementales de resultados complicados.

Tras el fallecimiento de Wald en 1950, sus ideas fueron continuadas por Blackwell(1954), Raiffa(1968), DeGroot(1970), Lindley (1971), Savage(1972) y Morgenstern (1974) que extienden la Teoría de la Decisión a prácticamente todos los campos del saber en los que se trabaje con incertidumbre. Una perspectiva del desarrollo histórico de la Teoría de la Decisión puede verse en Fishburn (1989).

6.7. EJERCICIOS

6.7.1

En un problema de decisión, la función de pérdida y el resultado de la experimentación, vienen dados por las siguientes tablas:

$L(\theta, a)$	a_1	a_2	a_3
θ_1	2	0	4
θ_2	1	4	0
θ_3	5	6	2

$p(x_i\|\theta_j)$	x_1	x_2
θ_1	$0'4$	$0'6$
θ_2	$0'5$	$0'5$
θ_3	$0'3$	$0'7$

Se pide:

a) Establézcanse el conjunto de reglas de decisión y dígase cuántas hay.

b) Considérense las reglas de decisión $\{d_1, d_2, d_3\}$ descritas en la siguiente tabla y evalúense sus riesgos

	d_1	d_2	d_3
x_1	a_1	a_2	a_1
x_2	a_2	a_3	a_3

c) ¿Cuál es la *regla minimax* de las establecidas en la parte b)?

d) Suponiendo que la información sobre el parámetro $\Theta = \{\theta_1, \theta_2, \theta_3\}$ viene dada por la distribución uniforme discreta ¿cuál es la *regla Bayes* de las tres consideradas en la parte b)?

e) Si se observa x_1 ¿cuál será la *regla Bayes* entre todas las reglas de decisión?

6.7.2

Una moneda tiene una probabilidad desconocida θ de salir cara. Se tira una vez y se observa el resultado. A continuación se decide si jugar o no un juego en el que se predice el resultado de la siguiente tirada de la moneda. Si se predice correctamente se ganan 2 unidades monetarias y si se predice incorrectamente se pierde 1 u.m., si no se juega no se pierde ni se gana nada. Se pide:

a) Plantéese como un problema de decisión explicitando sus elementos.

b) ¿Cuál es la *regla minimax*?. ¿Cuál es la *regla Bayes* frente a una distribución inicial uniforme sobre θ?

6.7.3

Dada una observación $X = x$ se pretende saber si proviene de una distribución $N(0, 1)$ o de una $N(2, 1)$. Se supone que una de las dos distribuciones es escogida inicialmente con probabilidad $1/2$.

a) Bajo pérdida $0 - 1$ ¿Cuál es la *regla Bayes*?

b) Si se observa el valor $x = 0'5$ ¿a qué distribución asigna la *regla Bayes* esta observación?

6.7.4

Cada vez que se recibe una alarma de la llegada de un tornado en una cierta comunidad, se dispone de tres acciones posibles: a_1, no movilizarse; a_2, movilizarse parcialmente; a_3, movilización total. La severidad del tornado, es medida por un parámetro θ, que de menor a mayor toma los valores $0, 0'5$ y 1. La alerta especifica una predicción de severidad que queda recogida por una v.a. X con distribución

$$P\{X=0|\theta\}=1-\theta \qquad P\{X=0'5|\theta\} \qquad P\{X=1|\theta\}=\frac{\theta}{2}.$$

Se sabe que en el área ocupada por la comunidad, una severidad 0 ocurre con probabilidad $0'5$, una severidad de $0'5$ ocurre con probabilidad $0'25$ y una severidad de 1 ocurre con probabilidad $0'25$. ¿Cuál es la *acción Bayes* cuando la alerta específica $X=0$? Tómese como función de pérdida

$\mathrm{L}(\theta,a)$	a_1	a_2	a_3
$\theta=0$	0	1	2
$\theta=0'5$	5	0	1
$\theta=1$	0	5	0

6.7.5

Sea $X \sim Binomial(n,\theta)$ con n conocido.

a) Si la distribución inicial sobre θ es $Beta(\sqrt{n}/2, \sqrt{n}/2)$. ¿Cuál es la distribución final para θ? Calcúlese la regla de decisión $d(x)=E[\theta|x]$.

b) Con función de pérdida cuadrática, ¿es $d(x)$ una *regla minimax*?

c) Compárese la función de riesgo correspondiente a $d(x)=E[\theta|x]$ con la función de riesgo de $T(x)=\frac{x}{n}$ para $n=10, 50$ y 100. ¿Qué conclusión se extrae?

6.7.6

Un asesor de Bolsa recomienda comprar un paquete de acciones de una compañía porque éstas van a subir. Si el asesor tiene razón y se compra el paquete se obtiene una ganacia de 1.000 unidades monetarias (u.m.) pero si se equivoca y las acciones bajan se pierden 1.000 u.m. Una segunda opción es vender un paquete análogo, de manera que si el asesor acierta se pierden 1.000 u.m. y si el asesor se equivoca y las acciones bajan se gana, por haberlas vendido, 1.000 u.m. Una tercera opción es no hacer nada; con lo que ni se pierde ni se gana. En base a la experiencia que se tiene sobre las opiniones del asesor, se supone que tanto cuando las acciones suben como cuando bajan, él asesora correctamente con probabilidad $2/3$. La probabilidad inicial de que las acciones suban es $2/5$ y de que bajen $3/5$. Determínese la *regla Bayes.*

6.7.7

Demuéstrese que si existe regularidad suficiente, la *regla Bayes* para la función de pérdida $L(\theta,a)=c(\theta)(\theta-a)^2$ con $c(\theta)$ tal que

$$0<\int_{\Theta} c(\theta)\pi(\theta|\overrightarrow{x})d\theta<+\infty$$

viene dada por

$$d^*(\overrightarrow{x}) = \frac{\int_\Theta \theta c(\theta)\pi(\theta|\overrightarrow{x})d\theta}{\int_\Theta c(\theta)\pi(\theta|\overrightarrow{x})d\theta}.$$

6.7.8

Se supone que la distribución inicial de un parámetro θ es uniforme en un intervalo acotado (a, b). Pruébese que la distribución final es proporcional a la función de verosimilitud en dicho intervalo.

6.7.9

Se supone que la v.a. X tiene *Distribución Geométrica* de parámetro θ con $\theta \in (0, 1)$

$$P\{X = x|\theta\} = \theta(1-\theta)^{x-1} \qquad x = 1, 2, ...$$

y que la distribución inicial para θ es la $Uniforme(0, 1)$. Se pide:

a) ¿Cuál es la distribución final para θ?

b) ¿Cuál es la *regla Bayes* bajo pérdida cuadrática?

c) ¿Cuál es el EMV para θ?

6.7.10

Laplace afirmaba que si en n observaciones de un experimento de $Bernoulli(\theta)$ el suceso éxito se había observado las n veces, la probabilidad de que el suceso se presente en la observación $n+1$ es $(n+1)/(n+2)$. ¿Puede justificarse este resultado?

6.7.11

Sea $X \sim N(\theta, \sigma)$ con σ conocida. Para cada $k \geq 0$ se define la región de confianza para θ dada por $S(x) = [x - k\sigma, x + k\sigma]$ y se emplea la función de pérdida

$$L(\theta, S) = a \ \text{ longitud}(S) - I_S(\theta)$$

siendo a una constante positiva.

a) Compruébese que existen pares de valores de k y a para los que el estimador que minimiza el riesgo $R(\theta, S)$ como función de k, es el intervalo degenerado $[x, x]$.

b) Compruébese que el estimador que minimiza el riesgo $R(\theta, S)$ como función de k es el intervalo de grado de confianza $1 - \alpha$.

6.7.12

Sea $X \sim f(x|\theta)$ y supóngase que para estimar θ se utiliza el conjunto de reglas de decisión $D = \{S(x)|x \in \mathcal{X}\}$ donde $S(x)$ es un intervalo creíble de probabilidad final $1 - \alpha$. Se emplea la función de pérdida

$$L(\theta, S) = a \text{ longitud}(S) - I_S(\theta)$$

con a una constante positiva.

a) Si se denota la distribución inicial por $\pi(\theta)$, pruébese que la *regla Bayes* viene dada por

$$S^{\pi}(x) = \{\theta | \pi(\theta|x) \geq k\}$$

b) Si además se supone que $X \sim N(\theta, \sigma)$ con σ conocida, compruébese que existen elecciones de k y a tales que el intervalo de la forma $[x - k\sigma, x + k\sigma]$ para estimar θ que minimiza en k el riesgo final se convierte en el intervalo degenerado $[x, x] = \{x\}$. (Tómese una distribución inicial $\theta \sim N(\mu, \tau)$).

c) Repítase la parte b) viendo si es posible elegir k y a para que el intervalo que minimice el riesgo final sea el intervalo de gado de confianza $1 - \alpha$ para θ con $k = z_{\alpha/2}$ dado mediante $[x - z_{\alpha/2}\sigma, x + z_{\alpha/2}\sigma]$.

6.7.13

Del cargamento de un barco se extraen 5 elementos aleatoriamente y se observa si son defectuosos o no. Puede suponerse que la distribución del número de elementos en mal estado $X \sim Binomial(5, \theta)$ con θ la proporción desconocida de elementos defectuosos del cargamento. De anteriores envíos se sabe que $\theta \sim Beta(1, 9)$. Se denota por a_0 la acción de decidir que es $0 \leq \theta \leq 0'15$ y por a_1 la acción de decidir que es $\theta > 0'15$. Encuéntrese la *regla Bayes* en las siguientes situaciones:

a) Bajo pérdida $0 - 1$.

b) Cuando la función de pérdida viene dada por

$$L(\theta, a_0) = \begin{cases} 0 & \text{si} \quad \theta \leq 0'15 \\ 1 & \text{si} \quad \theta > 0'15 \end{cases} \qquad L(\theta, a_1) = \begin{cases} 29 & \text{si} \quad \theta \leq 0'15 \\ 0 & \text{si} \quad \theta > 0'15 \end{cases}$$

c) ¿Qué decisión se tomaría si se observan 4 elementos defectuosos con las funciones de pérdidas dadas en a) y en b)?

6.7.14

Para determinar el coeficiente de inteligencia θ de una persona, se le somete a un test cuyo resultado $X \sim N(\theta, \sigma = 10)$. Se supone que como consecuencia de investigaciones anteriores, en el colectivo al que pertenece la persona $\theta \sim N(100, \tau = 15)$. Además es importante detectar valores especialmente altos y bajos del coeficiente de inteligencia, por lo que se supone apropiada la función de pérdida

$$L(\theta, a) = (\theta - a)^2 e^{(\theta - 100)^2/900}.$$

Encuéntrese la regla *Bayes*.

6.7.15

(Berger 1985). Para clasificar la sangre en uno de los cuatro tipos A,B, AB y 0 se emplea un procedimiento que consiste en observar una v.a. X con función de densidad

$$f(x|\theta) = e^{-(x-\theta)} I_{(\theta, +\infty)}(x)$$

donde θ es un parámetro que varía en el intervalo $(0, +\infty)$ de tal modo que si $0 < \theta \leq 1$, la sangre es del tipo AB; si $1 < \theta \leq 2$, la sangre es del tipo A; si $2 < \theta \leq 3$, la sangre es del tipo B y si $\theta > 3$, la sangre es del tipo 0. En la población bajo estudio, θ se distribuye con densidad

$$\pi(\theta) = e^{-\theta} I_{(0,+\infty)}(\theta).$$

La función de pérdida viene dada por la siguiente tabla

$L(\theta, a)$	AB	A	A	0
AB	0	1	1	1
A	1	0	2	2
B	1	2	0	2
0	3	3	3	0

si se observa $x = 4$ como resultado del análisis ¿cuál es la *regla Bayes*?

6.7.16

Dada una observación x, se denota por $\alpha(x)$ la medida de confianza con que $\theta \in S(x)$, siendo $S(x)$ un intervalo de confianza para un parámetro θ. Demuéstrese que con función de pérdida cuadrática de la forma

$$L(\theta, \alpha(x)) = (I_{S(x)}(\theta) - \alpha(x))^2$$

la *regla Bayes* frente a $\pi(\theta)$ viene dada por

$$\alpha^{\pi}(x) = P\{\theta \in S(x)|x\}$$

6.7.17

(Ferguson (1969)) La función de pérdida para un problema de estimación viene dada por

$$L(\theta, a) = \begin{cases} k_0(\theta - a) & \text{si} \quad \theta \geq a \\ k_1(a - \theta) & \text{si} \quad \theta < a \end{cases}$$

Demuéstrese que la *regla Bayes*, respecto de $\pi(\theta)$, es cualquier cuantil de orden $k_0/(k_0 + k_1)$ de la distribución final.

6.7.18

(Berger (1985)) Un misil puede dispararse con trayectoria baja, acción a_1, o con trayectoria alta, acción a_2. La función de pérdida crece linealmente, desde 0 hasta 1, con la distancia lineal l entre el lugar de impacto y el objetivo, si ésta es mayor que 1 kilómetro se supone que la función de pérdida vale constantemente 1. Si se dispara con trayectoria baja el misil no puede ser contrarrestado por el fuego antimisil; en este caso su precisión depende de la amplitud θ de la zona cubierta de nubes que atraviesa el proyectil de modo que la distancia l a la que queda del blanco puede suponerse

distribuida uniformemente en el intervalo $(-\theta, \theta)$. La función de densidad de la variable aleatoria θ es una $Beta(2,2)$ de densidad

$$\pi_1(\theta) = 6\ \theta(1-\theta)I_{(0,1)}(\theta)$$

Si se dispara con trayectoria alta el misil alcanza exactamente el blanco, salvo que antes sea destruido por el fuego antimisil, lo que ocurre con probabilidad ξ. De lanzamientos anteriores se sabe que

$$\pi_2(\xi) = 2(1-\xi)I_{(0,1)}(\xi)$$

Se realiza un experimento para obtener información sobe ξ y se observa que tras tirar dos misiles con trayectoria alta, los dos dan en el blanco. Se pide:

a) Obténgase la función de pérdida para a_1 y a_2.
b) ¿Cuál de las dos es la *regla minimax*?
c) ¿Cuál de las dos es la *regla Bayes*?
d) ¿Son a_1 y a_2 reglas admisibles?

6.7.19
Si para una sucesión d^{π_n} de reglas Bayes frente a π_n su correspondiente sucesión de distribuciones iniciales, se tiene que

$$\lim_{n\to\infty} B(\pi_n, d^{\pi_n}) = p,$$

pruébese, que, cualquier regla de decisión d tal que $R(\theta, d) = p$ para todo $\theta \in \Theta$, es una *regla minimax.*

6.8. SOLUCIÓN A LOS EJERCICIOS

6.8.1

a) El conjunto de reglas de decisión, son las aplicaciones del espacio muestral $\mathcal{X} = \{x_1, x_2\}$ en el espacio de acciones $\mathcal{A} = \{a_1, a_2, a_3\}$. Cada regla de decisión se puede representar colocando los valores de x_1 y x_2 y debajo de cada uno la acción a tomar, por lo tanto tiene $3^2 = 9$ elementos.

b) La función de riesgo, de acuerdo con la Expresión (6.1.1), viene dada por

$$R(\theta, d) = E[L(\theta, d(X))]$$

por lo tanto

$$\begin{aligned} R(\theta_1, d_1) &= L(\theta_1 d_1(x_1))p(x_1|\theta_1) + L(\theta_1, d_1(x_2))p(x_2|\theta_1) \\ &= L(\theta_1, a_1)p(x_1|\theta_1) + L(\theta_1, a_2)p(x_2|\theta_1) \\ &= 2 \cdot 0'4 + 0 \cdot 0'06 = 0'8 \end{aligned}$$

procediendo de manera análoga se obtiene

$\mathrm{R}(\theta_i, d_j)$	d_1	d_2	d_3
θ_1	0'8	2'4	3'2
θ_2	2'5	2	0'5
θ_3	5'7	3'2	2'9

c) Mediante el *método minimax*, en virtud de la Expresión (6.1.2), si se escoge la regla de decisión d_1, lo peor que puede ocurrir es que se pierda $5'7$ si $\theta = \theta_3$, si se escoge d_2 lo peor que puede pasar es que se pierda $3'2$ y por último si se escoge d_3 lo peor que puede pasar es que se pierda $3'2$. Luego tanto la regla d_2 como la d_3 hacen perder como mucho $3'2$; cualquiera de ellas es minimax.

d) Con el *método bayesiano*, si la información inicial viene dada por la distribución discreta

$$\pi(\theta_1) = \pi(\theta_2) = \pi(\theta_3) = 1/3$$

el *riesgo Bayes* se calcula mediante

$$B(\pi, d) = E[R(\theta, d)]$$

(véase la Expresión (6.1.3)), con lo que se sigue que

$$\begin{aligned} B(\pi, d_1) &= R(\theta_1, d_1)\pi(\theta_1) + R(\theta_2, d_1)\pi(\theta_2) + R(\theta_3, d_1)\pi(\theta_3) \\ &= (0'8 + 2'5 + 5'7)\frac{1}{3} \\ &= 3 \end{aligned}$$

procediendo análogamente $B(\pi, d_2) = 2'53$ y $B(\pi, d_3) = 2'2$. Luego entre las tres reglas, la que da el método bayesiano como óptima es d_3.

e) Si se observa x_1, mediante la fórmula de Bayes, se puede comparar la pérdida en que se incurre si se escoge a_1 con respecto a la elección de a_2. Esto permite tratar en conjunto todas las reglas de decisión (véase el Teorema 6.3.1), por lo tanto si se calcula

$$\pi(\theta_i|x_1) = \frac{\pi(\theta_i)p(x_1|\theta_i)}{m(x_1)}$$

con

$$m(x_1) = \sum_{i=1}^{3} \pi(\theta_i)p(x_1|\theta_i)$$

se llega a que $m(x_1) = (0'4 + 0'5 + 0'3)\frac{1}{3} = 0'4$ y

$$\begin{aligned} \pi(\theta_1|x_1) &= 0'3333 \\ \pi(\theta_2|x_1) &= 0'4167 \\ \pi(\theta_3|x_1) &= 0'25 \end{aligned}$$

de donde

$$r(a_i|x_1) = \sum_{j=1}^{3} L(\theta_j, a_i)\pi(\theta_j|x_1)$$

Sustituyendo

$$r(a_1|x_1) = 2'3333$$
$$r(a_2|x_1) = 3'1668$$
$$r(a_3|x_1) = 1'8332$$

De manera que la *regla Bayes* consiste en escoger a_3 cuando se observa x_1, con lo que $d^{\pi}(x_1) = a_3$.

6.8.2

a) El espacio paramétrico es $\Theta = [0, 1]$ las distintas probabilidades de cara de la moneda. La experimentación $\mathcal{X} = \{1, 0\}$ entendiendo que $x = 1$ si el resultado de la primera tirada de la moneda es cara y $x = 0$ si el resultado es cruz. El modelo viene dado por la distribución de $Bernoulli(\theta)$. El espacio de acciones es $\mathcal{A} = \{a_1, a_2, a_3\}$ donde a_1 es jugar y decir que la siguiente tirada de la moneda saldrá cara, a_2 es jugar y decir que la siguiente tirada valdrá cruz y a_3 no jugar. La función de pérdida viene dada por

$$L(\theta, a_1) = -2\theta + 1 - \theta$$
$$L(\theta, a_2) = \theta + (1 - \theta)(-2)$$
$$L(\theta, a_3) = 0.$$

Las reglas de decisión son

x	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9
1	a_1	a_1	a_1	a_2	a_2	a_2	a_3	a_3	a_3
0	a_1	a_2	a_3	a_1	a_2	a_3	a_1	a_2	a_3

b) Para el cálculo de la *regla minimax* se tiene que

$$R(\theta, d_i) = E[L(\theta, d_i(X))] = L(\theta, d_i(1))\theta + L(\theta, d_i(0))(1 - \theta)$$

de donde

$$R(\theta, d_1) = -3\theta + 1 \qquad R(\theta, d_4) = 6\theta^2 - 6\theta + 1 \qquad R(\theta, d_7) = 3\theta^2 - 4\theta + 1$$
$$R(\theta, d_2) = -6\theta^2 + 6\theta - 2 \qquad R(\theta, d_5) = 3\theta - 2 \qquad R(\theta, d_8) = -3\theta^2 + 5\theta - 2$$
$$R(\theta, d_3) = -3\theta^2 + \theta \qquad R(\theta, d_6) = 3\theta^2 - 2\theta \qquad R(\theta, d_9) = 0$$

con lo que

$$\min_{d_i} \max_{\theta} R(\theta, d_i) = \min\left\{1, -\frac{1}{2}, \frac{1}{12}, 1, 1, 1, 1, \frac{5}{6}, 0\right\}$$
$$= -\frac{1}{2}$$

por lo que la *regla minimax* es d_2.

Para la obtención de la *regla Bayes*, hay que calcular para la densidad $\pi(\theta)$ inicial $Uniforme(0,1)$ los promedios

$$\begin{aligned} B(\theta, d_i) &= E[R(\theta, d_i)] \\ &= \int_0^1 R(\theta, d_i) d\theta \qquad i = 1, ..., 9 \end{aligned}$$

de donde

$$\begin{array}{lll} B(\pi, d_1) = -\frac{1}{2} & B(\pi, d_4) = 0 & B(\pi, d_7) = 0 \\ B(\pi, d_2) = -1 & B(\pi, d_5) = -\frac{1}{2} & B(\pi, d_8) = -\frac{1}{2} \\ B(\pi, d_3) = -\frac{1}{2} & B(\pi, d_6) = 0 & B(\pi.d_9) = 0 \end{array}$$

por lo que la regla que hace mínimo $B(\pi, d_i)$ es d_2 que es *Bayes* frente a la distribución $U(0,1)$.

En consecuencia la regla de decisión que consiste en jugar repitiendo el resultado que se haya visto en el primer lanzamiento es *minimax* y es *Bayes.*

6.8.3

a) Si se denota por θ_1 la distribución escogida es $N(0,1)$ y por θ_2 la distribución escogida es $N(2,1)$; y por a_1 la observación $X = x$ proviene de la distribución $N(0,1)$; y por a_2 la observación $X = x$ proviene de la distribución $N(2,1)$. La función de pérdida es

$L(\theta, a)$	a_1	a_2
θ_1	0	1
θ_2	1	0

y la *regla Bayes* frente a la distribución inicial $P(\theta_1) = P(\theta_2) = 0'5$ vendrá dada por

$$d^\pi(x) = \begin{cases} a_1 & \text{si} \quad r(a_1|x) < r(a_2|x) \\ a_1 \text{ ó } a_2 & \text{si} \quad r(a_1|x) = r(a_2|x) \\ a_2 & \text{si} \quad r(a_1|x) > r(a_2|x) \end{cases}$$

donde para el cálculo de estos riesgos finales se tiene

$$\begin{aligned} r(a_1|x) &= L(\theta_1, a_1)P(\theta_1|x) + L(\theta_1|a_1)P(\theta_2|x) \\ &= P(\theta_2|x) \end{aligned}$$

y análogamente

$$r(a_2|x) = P(\theta_1|x).$$

Por lo tanto la expresión $r(a_1|x) < r(a_2|x)$ es equivalente a $P(\theta_2|x) < P(\theta_1|x)$ y al ser la distribución inicial la uniforme discreta, esta expresión

es equivalente a que en el valor x, la función de densidad de la $N(2,1)$ sea menor que la función de densidad de la $N(0,1)$. Ya que

$$P(\theta_1|x) = \frac{0'5f(x|\theta_1)}{m(x)}$$
$$P(\theta_2|x) = \frac{0'5f(x|\theta_2)}{m(x)}$$

con $m(x) = 0'5f(x|\theta_1) + 0'5f(x|\theta_2)$ por lo tanto

$$d^*(x) = \begin{cases} a_1 & \text{si} \quad x < 1 \\ a_1 \text{ o } a_2 & \text{si} \quad x = 1 \\ a_2 & \text{si} \quad x > 1 \end{cases}$$

resultado intuitivo por otra parte.

b) De acuerdo con lo anterior la *regla Bayes* clasifica la observación de $x = 0'5$ como proviniente de la distribución $N(0,1)$.

6.8.4

El espacio muestral está formado por $\mathcal{X} = \{0, 0'5, 1\}$ por lo tanto una regla de decisión es cualquier aplicación $d : \mathcal{X} \to \mathcal{A}$ donde $\mathcal{A} = \{a_1, a_2, a_3\}$. Para buscar la *acción Bayes* frente a $x = 0$ hay que calcular el riesgo final $r(a_i|x)$ para cada acción $a_i \quad i = 1, 2, 3$.

$$\begin{aligned} r(a_i|x=0) = L(\theta=0, a_i)P\{\theta=0|x=0\} + \\ +L(\theta=0'5, a_i)P\{\theta=0'5|x=0\} \\ +L(\theta=1, a_i)P\{\theta=1|x=0\} \end{aligned} \tag{6.8.1}$$

donde

$$P\{\theta=0|x=0\} = \frac{P\{\theta=0\}P\{x=0|\theta=0\}}{P\{x=0\}}$$

la distribución inicial es $P\{\theta=0\} = 0'5$, $P\{\theta=0'5\} = 0'25$, $P\{\theta=1\} = 0'25$ y la distribución marginal de $x=0$ viene dada por

$$\begin{aligned} P\{x=0\} &= P\{\theta=0\}P\{x=0|\theta=0\} + P\{\theta=0'5\}P\{x=0|\theta=0'5\} + \\ &\quad + P\{\theta=1\}P\{x=0|\theta=1\} \\ &= 0'625 \end{aligned}$$

por lo tanto la distribución final es

$$P\{\theta=0|x=0\} = \frac{0'5 \cdot 1}{0'625} = 0'8$$
$$P\{\theta=0'5|x=0\} = \frac{0'25 \cdot 0'5}{0'625} = 0'2$$
$$P\{\theta=1|x=0\} = 0$$

sustituyendo en (6.8.1)

$$\begin{aligned} r(a_1|\text{x=0}) &= 1 \\ r(a_2|\text{x=0}) &= +0'8 \\ r(a_3|\text{x=0}) &= +1'8 \end{aligned}$$

luego la *acción Bayes*, cuando la alerta especifica una predicción de severidad de 0, es la a_2 movilizarse parcialmente.

6.8.5

a) La función de densidad final viene dada por

$$\begin{aligned} \pi(\theta|x) &= \frac{\frac{1}{B(\sqrt{n}/2,\sqrt{n}/2)}\theta^{\sqrt{n}/2-1}(1-\theta)^{\sqrt{n}/2-1}\binom{n}{x}\theta^x(1-\theta)^{n-x}}{\frac{1}{N(\sqrt{n}/2,\sqrt{n}/2)}\int_0^1 \theta^{\sqrt{n}/2-1}(1-\theta)^{\sqrt{n}/2-1}\binom{n}{x}\theta^x(1-\theta)^{n-x}d\theta} \\ &= \frac{\theta^{\sqrt{n}/2+x-1}(1-\theta)^{\sqrt{n}/2+n-x-1}}{B(\sqrt{n}/2+x,\sqrt{n}/2+n-x)} \end{aligned}$$

por lo tanto la distribución final es una distribución $Beta(\sqrt{n}/2+x, \sqrt{n}/2+n-x)$ y la regla de decisión $d(x)$ es

$$d(x) = \frac{\sqrt{n}/2+x}{\sqrt{n}+n}$$

b) $d(x)$ es una *regla minimax* ya que $R(\theta, d)$ es una función constante en θ, ya que

$$\begin{aligned} R(\theta,d) = E[L(\theta,d(X))] &= \sum_{x=0}^{n}\left(\theta - \frac{\sqrt{n}/2+x}{\sqrt{n}+n}\right)^2\binom{n}{x}\theta^x(1-\theta)^{n-x} \\ &= \frac{1}{(\sqrt{n}+n)^2}\sum_{x=0}^{n}\Big[\theta^2(\sqrt{n}+n)^2 + \frac{n}{4} + x^2 - \theta(\sqrt{n}+n)\sqrt{n} - \\ &\quad 2\theta(\sqrt{n}+n)x + \sqrt{n}x\Big]\cdot\binom{n}{x}\theta^x(1-\theta)^{n-x} \\ &= \frac{1}{(\sqrt{n}+n)^2}\Big[\theta^2(\sqrt{n}+n)^2 + \frac{n}{4} + E[X^2] - \theta(\sqrt{n}+n)\sqrt{n} - \\ &\quad 2\theta(\sqrt{n}+n)E[X] + \sqrt{n}E[X]\Big] \end{aligned}$$

y sustituyendo $E[X] = n\theta$ y $E[X^2] = n\theta + n(n-1)\theta^2$, se llega a que

$$R(\theta,d) = \frac{1}{4}(1+\sqrt{n})^{-2}.$$

Así pues la regla d es *Bayes* e *igualizadora*, luego es una *regla minimax*, véase el Teorema 6.3.2.

c) La función de riesgo correspondiente a $\frac{x}{n}$, el estimador de máxima verosimilitud para θ, es

$$R(\theta, \frac{x}{n}) = E\left[\left(\theta - \frac{X}{n}\right)^2\right] = \frac{1}{n^2} E\left[(X - n\theta)^2\right] = \frac{\theta(1-\theta)}{n}$$

de acuerdo con la siguiente tabla:

	$n = 10$	$n = 50$	$n = 100$
$R(\theta, d)$	$0'0144$	$0'0038$	$0'00207$
$R(\theta, \frac{x}{n})$	$\frac{\theta(1-\theta)}{10}$	$\frac{\theta(1-\theta)}{50}$	$\frac{\theta(1-\theta)}{100}$
$\widehat{R}$	$0'025$	$0'005$	$0'0025$

donde $\widehat{R}$ es el valor máximo de $R(\theta, \frac{x}{n})$ cuando $\theta \in [0, 1]$. Se observa, que para valores pequeños de n la regla d es sensiblemente mejor que el EMV, salvo cuando θ está en los extremos del intervalo, este efecto se vá perdiendo cuando n aumenta, en cuyo caso ambas reglas producen el mismo riesgo.

6.8.6

Si se denota por θ_1 las acciones de la compañia van a subir, y por θ_2 al suceso las acciones de la compañía van a bajar, se tiene explicitado el conjunto de estados $\Theta = \{\theta_1, \theta_2\}$. Las acciones posibles son: a_1 comprar el paquete de acciones, a_2 venderlo y a_3 no hacer nada. La función de pérdida viene dada por:

$L(\theta, a)$	a_1	a_2	a_3
θ_1	-1000	+1000	0
θ_2	+1000	-1000	0

donde $L(\theta_1, a_1) = -1000$ indica una pérdida de -1000 unidades monetarias es decir una ganancia de 1000, cuando las acciones suben y se compra el paquete de acciones.

El resultado de la experimentación viene dado por $X = 1$, cuando el asesor dice que las acciones van a subir, y $X = 0$, cuando dice que van a bajar, por lo que el modelo viene caracterizado por las distribuciones de probabilidad

$$P\{X = 1|\theta_1\} = \tfrac{2}{3} \quad P\{X = 0|\theta_1\} = \tfrac{1}{3}$$
$$P\{X = 1|\theta_2\} = \tfrac{1}{3} \quad P\{X = 0|\theta_2\} = \tfrac{2}{3}$$

mediante la fórmula de Bayes, la distribución final correspondiente a la inicial $P\{\theta_1\} = \frac{2}{5}$, $P\{\theta_2\} = \frac{3}{5}$ viene dada por

$$P\{\theta_1|\text{X=1}\} = \frac{P\{\theta_1\}P\{X = 1|\theta_1\}}{P\{\theta_1\}P\{X = 1|\theta_1\} + P\{\theta_2\}P\{X = 1|\theta_2\}}$$
$$= \frac{\frac{2}{5} \cdot \frac{2}{3}}{\frac{2}{5} \cdot \frac{2}{3} + \frac{3}{5} \cdot \frac{1}{3}} = \frac{4}{7}$$

$$P\{\theta_2|\text{X=1}\} = \frac{P\{\theta_2\}P\{X=1|\theta_2\}}{P\{\theta_1\}P\{X=1|\theta_1\} + P\{\theta_2\}P\{X=1|\theta_2\}}$$
$$= \frac{\frac{3}{5}\cdot\frac{1}{3}}{\frac{2}{5}\cdot\frac{2}{3} + \frac{3}{5}\cdot\frac{1}{3}} = \frac{3}{7}$$

de forma análoga se obtiene que

$$P\{\theta_1|\text{X=0}\} = \frac{1}{4}$$
$$P\{\theta_2|\text{X=0}\} = \frac{3}{4}$$

Para el cálculo de la regla Bayes se calculan los riesgos finales

$$r(a_1|X=1) = L(\theta_1,a_1)P\{\theta_1|X=1\} + L(\theta_2,a_1)P\{\theta_2|X=1\}$$
$$= -1000\cdot\frac{4}{7} + 1000\cdot\frac{3}{7} = -\frac{1000}{7}$$
$$r(a_2|X=1) = L(\theta_1,a_2)P\{\theta_1|X=1\} + L(\theta_2,a_2)P\{\theta_2|X=1\}$$
$$= 1000\cdot\frac{4}{7} - 1000\cdot\frac{3}{7} = \frac{1000}{7}$$
$$r(a_3|X=1) = 0$$

con lo que la *acción Bayes* cuando se observa $X=1$ es la a_1 que da mínima pérdida. Repitiendo los cálculos para el caso $X=0$

$$r(a_1|X=0) = L(\theta_1,a_1)P\{\theta_1|X=0\} + L(\theta_2,a_1)P\{\theta_2|X=0\}$$
$$= -1000\cdot\frac{1}{4} + 1000\cdot\frac{3}{3} = 500$$
$$r(a_2|X=0) = L(\theta_1,a_2)P\{\theta_1|X=0\} + L(\theta_2,a_2)P\{\theta_2|X=0\}$$
$$= 1000\cdot\frac{1}{4} - 1000\cdot\frac{3}{4} = -500$$
$$r(a_3|X=0) = 0$$

con lo que la *acción Bayes* cuando se observa $X=0$ es la a_2.

6.8.7

La *regla Bayes* respecto a la distribución inicial dada por $\pi(\theta)$ es la que minimiza el riesgo Bayes

$$B(\pi,d) = E[R(\theta,d)]$$

donde la anterior esperanza está calculada respecto a la distribución inicial $\pi(\theta)$, (véase la Expresión (6.1.3)). Alternativamente se puede minimizar en a el riesgo final

$$r(a|\overrightarrow{x}) = \int_\Theta L(\theta,a)\pi(\theta|\overrightarrow{x})d\theta$$
$$= \int_\Theta c(\theta)(\theta-a)^2\pi(\theta|\overrightarrow{x})d\theta$$

si se deriva la anterior expresión respecto de a y la integral y la derivada conmutan, la solución de la expresión

$$\frac{d}{da}\int_{\Theta} c(\theta)(\theta-a)^2\pi(\theta|\overrightarrow{x})d\theta = 0 \tag{6.8.1}$$

viene dada por

$$a^* = \frac{\int_{\Theta}\theta c(\theta)\pi(\theta|\overrightarrow{x})d\theta}{\int_{\Theta} c(\theta)\pi(\theta|\overrightarrow{x})d\theta} \tag{6.8.2}$$

y como volviendo a derivar respecto de a la Expresión (6.8.1) se obtiene

$$2\int_{\Theta} c(\theta)\pi(\theta|\overrightarrow{x})d\theta$$

una cantidad positiva, el valor de a^* dado por (6.8.2), se corresponde con un mínimo de $r(a|\overrightarrow{x})$ y permite construir la *regla Bayes* mediante $d^*(\overrightarrow{x}) = a^*$. Nótese que si $c(\theta) = 1$ se obtiene que $d^*(\overrightarrow{x}) = E[\theta|\overrightarrow{x}]$.

6.8.8
Mediante el teorema de Bayes

$$\begin{aligned}\pi(\theta|\overrightarrow{x}) &= \frac{I_{(a,b)}(\theta)\frac{1}{b-a}f(\overrightarrow{x}|\theta)}{\int_a^b I_{(a,b)}(\theta)\frac{1}{b-a}f(\overrightarrow{x}|\theta)d\theta}\\ &= I_{(a,b)}(\theta)\frac{f(\overrightarrow{x}|\theta)}{\int_a^b f(\overrightarrow{x}|\theta)d\theta}\end{aligned}$$

por lo tanto si en el intervalo (a,b), la función de verosimilitud tiene una integral finita no nula, la distribución final es proporcional a la verosimilitud.

6.8.9

a) La distribución final viene dada por

$$\begin{aligned}\pi(\theta|x) &= \frac{\theta(1-\theta)^{x-1}}{\int_0^1\theta(1-\theta)^{x-1}d\theta}\\ &= \frac{\theta(1-\theta)^{x-1}}{B(2,x)}\end{aligned}$$

es decir $\theta|x \sim Beta(2,x)$.

b) La *regla Bayes* es $d(x_1) = E[\theta|x] = \frac{2}{2+x}$.

c) El EMV para θ se obtiene maximizando en θ la función de verosimilitud o su logaritmo

$$\ln\theta + (x-1)\ln(1-\theta)$$

es fácil comprobar, mediante el estudio de la segunda derivada, que la raíz de la ecuación

$$\frac{1}{\theta} - \frac{x-1}{1-\theta} = 0$$

se corresponde con un máximo interior al intervalo, luego

$$EMV(\theta) = x^{-1}.$$

6.8.10

Si se supone la distribución inicial

$$\pi(\theta) = I_{(0,1)}(\theta)$$

la afirmación de Laplace se reduce a calcular

$$P\left\{X_{n+1} = 1 \middle| \sum_{i=1}^{n} X_i = n\right\} = \frac{P\left\{\sum_{i=1}^{n} X_i = n, X_{n+1} = 1\right\}}{P\left\{\sum_{i=1}^{n} X_i = n\right\}}$$

Mediante la expresión

$$\begin{aligned} P\left\{\sum_{i=1}^{n} X_i = n, X_{n+1} = 1\right\} &= \int_0^1 P\left\{\sum_{i=1}^{n} X_i = n, X_{n+1} = 1 \middle| \theta\right\} \pi(\theta) d\theta \\ &= \int_0^1 P\left\{\sum_{i=1}^{n} X_i = n \middle| \theta\right\} P\left\{X_{n+1} = 1 | \theta\right\} \pi(\theta) d\theta \\ &= \int_0^1 \theta^n \theta d\theta = \frac{1}{n+2} \end{aligned}$$

Análogamente

$$\begin{aligned} P\left\{\sum_{i=1}^{n} X_i = n\right\} &= \int_0^1 P\left\{\sum_{i=1}^{n} X_i = n \middle| \theta\right\} \pi(\theta) d\theta \\ &= \int_0^1 \theta^n d\theta = \frac{1}{n+1} \end{aligned}$$

Bajo estas suposiciones se obtiene la regla de sucesión de Laplace. Si un suceso ha ocurrido en n repeticiones, la probabilidad de que se presente en la repetición $n+1$ viene dada por la expresión $(n+1)/(n+2)$.

6.8.11

a)

$$\begin{aligned} R(\theta, S) &= \int_{\mathcal{X}} [a\ longitud(S(x)) - I_{S(x)}(\theta)] f(x|\theta) dx \\ &= a2k\sigma - P\{X - k\sigma < \theta < X + k\sigma | \theta\} \\ &= a2k\sigma - P\{\theta - k\sigma < X < \theta + k\sigma | \theta\} \\ &= a2k\sigma - 2\Phi(k) + 1 \end{aligned}$$

donde Φ es la función de distribución de la v.a. $N(0,1)$.

La función $R(\theta, S)$ para $k = 0$ vale 0 y su derivada respecto de k es

$$R'(\theta, S) = a2\sigma - 2\varphi(k)$$

con $\varphi(k)$ la función de densidad de la v.a. $N(0,1)$. Por lo tanto, si $a\sigma > \varphi(0)$ la derivada es positiva en todo k, por lo que el mínimo de $R(\theta, S)$ se alcanza para $k = 0$ en el intervalo degenerado $[x, x] = \{x\}$.

b) Repitiendo el razonamiento de la parte a) se tiene que la función de riesgo en $k = 0$ vale 0 y si $a\sigma < \varphi(0)$ la derivada en 0 es negativa. Cuando k tiende a infinito $R(\theta, S)$ tiende a $+\infty$ por lo que existe un valor de k para el que la función alcanza su mínimo que viene dado por la solución en k de la ecuación

$$a\sigma - \varphi(k) = 0.$$

En resumen si $a\sigma < 1/\sqrt{2\pi} = 0'0507$ y además es $k^* = \sqrt{-2ln(a\sigma 2'507)}$ el mínimo riesgo se alcanza para el intervalo $[x - k^*\sigma, x + k^*\sigma]$, si ahora se escoge el valor de a para que $k^*(a) = z_{\frac{\alpha}{2}}$ se obtiene el resultado deseado.

6.8.12

a) El riesgo final correspondiente al intervalo $S(x)$ viene dado por

$$r(S(x)|x) = a\ lng(S(x)) - \int_{S(x)} \pi(\theta|x)d\theta.$$

Verificar que $r(S^\pi(x)|x) \leq r(S(x)|x)$ supuesto que se cumple

$$P\{S^\pi(x)|x\} = P\{S(x)|x\} = 1 - \alpha, \tag{6.8.3}$$

es lo mismo que probar que $lng(S^\pi(x)) \leq lng(S(x))$. En efecto considérense las funciones

$$\varphi^*(\theta) = \begin{cases} 1 & \text{si} \quad \pi(\theta|x) \geq k \\ 0 & \text{si} \quad \pi(\theta|x) < k \end{cases} \qquad \varphi(\theta) = \begin{cases} 1 & \text{si} \quad \theta \in S \\ 0 & \text{si} \quad \theta \notin S \end{cases}$$

indicadores de los conjuntos $S^\pi(x)$ y $S(x)$ respectivamente.

Se tiene que

$$\int_\Theta (\varphi^*(\theta) - \varphi(\theta))(\pi(\theta|x) - k)d\theta \geq 0,$$

ya que en los puntos θ tales que $\pi(\theta|x) - k \geq 0$ es $\varphi^*(\theta) - \varphi(\theta) = 1 - \varphi(\theta)$ no negativo, y en los puntos θ en los que $\pi(\theta|x) - k \leq 0$ es $\varphi^*(\theta) - \varphi(\theta) = -\varphi(\theta)$ no positivo. Por lo tanto

$$\int_\Theta \varphi^*(\theta)\pi(\theta|x)d\theta - \int_\Theta \varphi(\theta)\pi(\theta|x)d\theta \geq k\left[\int_\Theta \varphi^*(\theta)d\theta - \int_\Theta \varphi(\theta)d\theta\right]$$

de manera que aplicando (6.8.3) se sigue que

$$0 \geq k\left[\int_{\Theta} \varphi^*(\theta)d\theta - \int_{\Theta} \varphi(\theta)d\theta\right]$$

luego el paréntesis de esta última expresión es no positivo con lo que

$$\int_{S^\pi(x)} d\theta \leq \int_{S(x)} d\theta$$

o lo que es lo mismo $lng(S^\pi(x)) \leq lng(S(x))$.

Es importante señalar dos cosas: la primera es que lo que se ha probado es que entre todos los intervalos que dan la misma probabilidad final, el de menor longitud es el intervalo creíble de más alta probabilidad final. La segunda es que el razonamiento que se ha seguido para hacerlo, es el mismo que se utiliza para demostrar el teorema de Neyman-Pearson en el caso de contrastar la hipótesis nula $H_0 : \pi(\theta|x)$ frente a la alternativa $H_1 : 1$ (la medida de Lebesgue), véase el Teorema 5.3.1.

b) Si la distribución inicial para θ es $N(\mu, \tau)$ la distribución final es $N(\mu_1, \tau_1)$ donde

$$\mu_1 = \frac{\frac{\mu}{\tau^2}+\frac{x}{\sigma^2}}{\frac{1}{\tau^2}+\frac{1}{\sigma^2}} \quad \tau_1^2 = \frac{1}{\frac{1}{\tau^2}+\frac{1}{\sigma^2}}$$

(véase el Ejemplo 3.2.3.2).

Para calcular el mínimo en $k \geq 0$ de $r(S|x)$ se va a denotar $r(S_k|x)$ en lugar de $R(S|x)$. Es decir

$$\begin{aligned} r(S_k|x) &= E[a\ lngS_k(x) - I_{S_k(x)}(\theta)|x] \\ &= a2k\sigma - P\{\theta \in S_k(x)|x\} \\ &= a2k\sigma - [F(x+k\sigma) - F(x-k\sigma)] \end{aligned}$$

con F la función de distribución de la v.a. $N(\mu_1, \sigma_1)$.

Se sigue que $r(S_0|x) = 0$ y que la derivada respecto de k de la función $r(S_k|x)$ es

$$r'(S_k|x) = a2\sigma - f(x+k\sigma)\sigma - f(x-k\sigma)\sigma$$

con lo que $r'(S_0|x) = a2\sigma - 2f(x)\sigma$, donde f es la función de densidad de la v.a. $N(\mu_1, \sigma_1)$. Por lo tanto si $f(x) > a$ la derivada en 0 es negativa, pero si se escoge un valor de a para la función de pérdida con $f(x) < a$ el mínimo de $r(S_k|x)$ se alcanza en $k = 0$, ya que $r'(S_k|x)$ es siempre positiva si $a > f(\mu_1)$, con lo que se obtiene el resultado deseado; el intervalo degenerado $\{x\}$ es la *regla Bayes*.

c) Por el apartado anterior, si se escogen valores de a menores que $f(x)$ entonces $r'(S_0|x)$ es negativa y como cuando k tiende hacia infinito $r(S_k|x)$ tiende también a infinito, se puede ver que el único mínimo existente

se corresponde con la solución en k de $r'(S_k|x) = 0$. Se trata de ver si pueden escogerse valores de a menores que $f(x)$ que cumplan

$$2a = f(x + z_{\frac{\alpha}{2}}\sigma) + f(x - z_{\frac{\alpha}{2}}\sigma) \tag{6.8.4}$$

De acuerdo con la Figura 6.8.1 esto en algunas situaciones es posible, basta llevar el valor de $f(x-k\sigma)$, el segmento AB, a continuación de $f(x+k\sigma)$ y dividir el segmento CB en dos partes iguales, cada una de las cuales es a.

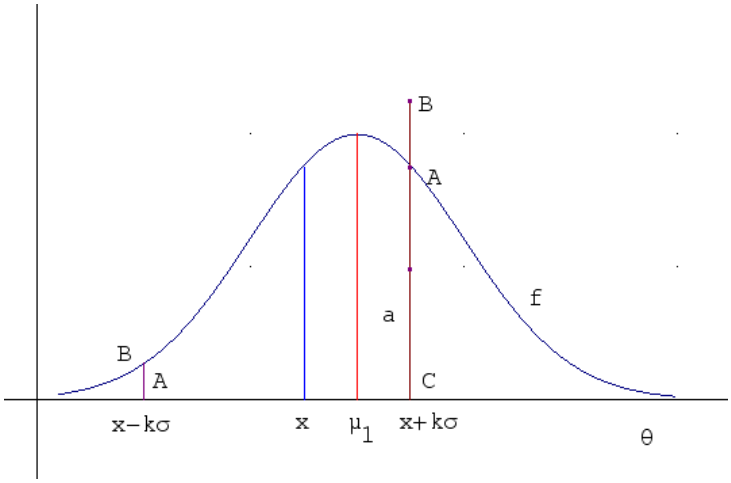


Figura 6.8.1. Elección de a para la parte c) del Ejercicio 6.8.12.

La conclusión es que en algunas situaciones, para valores de x, σ y α es posible c). Es importante considerar el caso en el que $x = \mu_1$ ya que se cumple que

$$f(x + k\sigma) = f(x - k\sigma) \tag{6.8.5}$$

En esta situación se puede hacer que $a < f(\mu_1)$ con

$$a = f(x - z_{\frac{\alpha}{2}}\sigma) = f(x + z_{\frac{\alpha}{2}}\sigma)$$

por lo que se satisfacen las expresiones (6.8.4) y (6.8.5), y el intervalo $[x - z_{\frac{\alpha}{2}}\sigma, x + z_{\frac{\alpha}{2}}\sigma]$ es una *regla Bayes* e intervalo de confianza de grado $1 - \alpha$, para el valor de a conveniente de acuerdo con la parte a).

Sin embargo para que $\mu_1 = x$ la distribución inicial tiene que tener $\tau = +\infty$, con lo que la distribución inicial es la medida de Lebesgue y el *riesgo Bayes* correspondiente es infinito, compruébese que vale $\int_{-\infty}^{+\infty}[2az_{\frac{\alpha}{2}}\sigma - (1 - \alpha)]d\theta$ es una *regla Bayes generalizada*, en el sentido de que minimiza el riesgo final aunque el *riesgo Bayes* es $+\infty$.

6.8.13

a) La *regla Bayes* viene dada por

$$d^\pi(x) = \begin{cases} a_0 & \text{si } P\{\Theta_0|x\} \geq 1/2 \\ a_1 & \text{si } P\{\Theta_0|x\} \leq 1/2 \end{cases}$$

donde $\Theta_0 = [0, 0'15]$ y $\Theta_1 = [0'15, 1]$, (véase el Corolario 6.3.1).

La distribución final queda caracterizada por la densidad

$$\pi(\theta|x) = \frac{\frac{1}{B(1,9)}(1-\theta)^8\binom{5}{x}\theta^x(1-\theta)^{5-x}}{\int_0^1 \frac{1}{B(1,9)}(1-\theta)^8\binom{5}{x}\theta^x(1-\theta)^{5-x}d\theta}$$
$$= \frac{1}{B(x+1, 13-x+1)}\theta^x(1-\theta)^{13-x}$$

Es decir, es la distribución *Beta* de parámetros $x+1$ y $14-x$, de donde

$$d^\pi(x) = \begin{cases} a_0 & \text{si } F(0'15|x) \geq 0'5 \\ a_1 & \text{si } F(0'15|x) \leq 0'5 \end{cases}$$

con $F(\theta|x)$ la función de distribución de la $Beta(x+1, 13-x+1)$.

b) Análogamente

$$d^\pi(x) = \begin{cases} a_0 & \text{si } P\{\Theta_0|x\} \geq 1/30 \\ a_1 & \text{si } P\{\Theta_0|x\} \leq 1/30 \end{cases}$$

y sustituyendo

$$d^\pi(x) = \begin{cases} a_0 & \text{si } F(0'15|x) \geq 1/30 \\ a_1 & \text{si } F(0'15|x) \leq 1/30 \end{cases}$$

donde otra vez $F(\theta|x)$ es la función de distribución de la $Beta(x+1, 14-$ $-x.)$.

c) La distribución final es $Beta(5, 10)$ por lo que $F(0'15|4) = 0'0467$, por lo tanto en el primer caso $F(0'15|4) \leq 0'5$ luego $d^\pi(4) = a_1$. En el segundo caso $F(0'15|4) = 0'0467 \geq 0'0333$ con lo que $d^\pi(4) = a_0$.

6.8.14

La distribución final $\theta|x \sim N(\mu_1, \sigma_1)$, donde

$$\mu_1 = \frac{100/225 + x/100}{1/225 + 1/100} \quad \sigma_1 = \frac{1}{\sqrt{1/225 + 1/100}}$$

(véase el Ejemplo 3.2.3.2). Para la obtención de la *regla Bayes* se calcula el riesgo final

$$r(a|x) = E[L(\theta, a)|x]$$
$$= \int_{-\infty}^{+\infty} (\theta - a)^2 e^{(\theta-100)^2/900}\pi(\theta|x)d\theta$$

desarrollando los cuadrados de los dos exponentes

$$r(a|x)=\frac{e^{\frac{100^2}{900}-\frac{\mu_1^2}{2\sigma_1^2}}}{\sqrt{2\pi}\sigma_1}\int_{-\infty}^{+\infty}(\theta-a)^2e^{\left(\frac{1}{900}-\frac{1}{2\sigma_1^2}\right)\theta^2+\left(\frac{\mu_1}{\sigma_1^2}-\frac{100}{450}\right)\theta}d\theta$$

derivando respecto de a e igualando a 0, se tiene que a debe cumplir

$$\int_{-\infty}^{+\infty}\theta e^{\left(\frac{1}{900}-\frac{1}{2\sigma^2}\right)\theta^2+\left(\frac{\mu_1}{\sigma_1^2}-\frac{1}{450}\right)\theta}d\theta=a\int_{-\infty}^{+\infty}e^{\left(\frac{1}{900}-\frac{1}{2\sigma^2}\right)\theta^2+\left(\frac{\mu_1}{\sigma_1^2}-\frac{1}{450}\right)\theta}d\theta$$

para el segundo miembro, mediante la Expresión (3.2.3.1) con

$$E=\frac{\mu_1/\sigma_1^2-\frac{100}{450}}{1/\sigma_1^2-\frac{1}{450}}\qquad V=\frac{1}{1/\sigma_1^2-\frac{1}{450}}$$

se sigue que puede ponerse como

$$e\sqrt{2\pi}\frac{1}{\sqrt{\frac{1}{\sigma_1^2}-\frac{1}{450}}}e^{-\frac{1}{2}\frac{(\mu_1/\sigma_1^2-100/450)^2}{1/\sigma_1^2-1/450}} \tag{6.8.6}$$

mientras que el primer miembro, si se usa que

$$\int_{-\infty}^{+\infty}\theta\frac{1}{\sqrt{2\pi}\sqrt{V}}e^{-\frac{1}{2V}(\theta-E)^2}d\theta=E$$

o alternativamente

$$\int_{-\infty}^{+\infty}\theta e^{-\frac{1}{2V}\theta^2+\frac{E}{V}\theta}d\theta=\sqrt{2\pi}E\sqrt{V}e^{-\frac{1}{2}E^2/V}$$

se tiene que es igual a

$$\sqrt{2\pi}\frac{1}{\sqrt{1/\sigma_1^2-1/450}}\frac{\mu_1/\sigma_1^2-100/450}{1/\sigma_1^2-1/450}e^{\frac{1}{2}(\mu_1/\sigma_1^2-\frac{100}{450})^2/(1/\sigma_1^2-1/450)^2} \tag{6.8.7}$$

e igualando (6.8.6) y (6.8.7) se puede despejar a, por lo que se comprueba que la *regla Bayes* viene dada por

$$d^\pi(x)=\frac{\mu_1/\sigma_1^2-100/450}{1/\sigma_1^2-1/450}.$$

6.8.15

La distribución final viene dada por

$$\begin{aligned}\pi(\theta|x)&=\frac{e^{-\theta}e^{-(x-\theta)}}{\int_0^x e^{-\theta}e^{-(x-\theta)}d\theta}\\&=x^{-1}I_{(0,x)}(\theta)\end{aligned}$$

por lo tanto el riesgo final de las distintas acciones, cuando se observa $x = 4$ viene dado por

$$\begin{aligned}
r(AB|\text{x=4}) &= \int_1^2 \frac{1}{4}d\theta + \int_2^3 \frac{1}{4}d\theta + 3\int_3^4 \frac{1}{4}d\theta = \frac{5}{4} \\
r(A|\text{x=4}) &= \int_0^1 \frac{1}{4}d\theta + 2\int_2^3 \frac{1}{4}d\theta + 3\int_3^4 \frac{1}{4}d\theta = \frac{6}{4} \\
r(B|\text{x=4}) &= \int_0^1 \frac{1}{4}d\theta + 2\int_1^2 \frac{1}{4}d\theta + 3\int_3^4 \frac{1}{4}d\theta = \frac{6}{4} \\
r(0|\text{x=4}) &= 2\int_0^1 \frac{1}{4}d\theta + 2\int_1^2 \frac{1}{4}d\theta + 2\int_2^3 \frac{1}{4}d\theta = \frac{6}{4}
\end{aligned}$$

luego la *regla Bayes* frente a $\pi(\theta)$ es $d^\pi(4) = AB$, valor para el que se alcanza el mínimo riesgo final.

6.8.16

El riesgo final correspondiente a la medida de confianza $\alpha(x)$ es

$$\begin{aligned}
r(\alpha(x)|x) &= E(L(\theta, \alpha(x)|x] \\
&= E[(I_{S(x)}(\theta) - \alpha(x))^2|x]
\end{aligned}$$

si se deriva respecto de α y se iguala a 0 , se obtiene

$$-2E[I_{S(x)}(\theta) - \alpha(x)|x] = 0$$

de donde

$$\begin{aligned}
\alpha(x) &= E[I_{S(x)}(\theta)|x] \\
&= \int_\Theta I_{S(x)}(\theta)\pi(\theta|x)d\theta \\
&= P\{\theta \in S(x)|x\}
\end{aligned}$$

como la derivada segunda de $r(\alpha(x)|x)$ es positiva, se sigue que el mínimo riesgo se alcanza para $\alpha^\pi(x)$ la probabilidad final de la región $S(x)$.

6.8.17

El riesgo final viene dado por

$$\begin{aligned}
r(a|x) &= \int_a^{+\infty} k_0(\theta - a)\pi(\theta|x)d\theta + \int_{-\infty}^a k_1(a - \theta)\pi(\theta|x)d\theta \\
&= k_0\left[\int_a^{+\infty} \theta\pi(\theta|x)d\theta - a(1 - F(a|x))\right] + k_1\left[aF(a|x) - \int_{-\infty}^a \theta\pi(\theta|x)d\theta\right]
\end{aligned}$$

donde $F(a|x)$ es la función de distribución de la distribución final. La derivada respecto de a es

$$\begin{aligned}\frac{\partial}{\partial a}r(a|x) &= -k_0 a\pi(a|x) - k_0(1 - F(a|x)) + k_0 a\pi(a|x) \\ &\quad + k_1 F(a|x) + k_1 a\pi(a|x) - k_1 a\pi(a|x) \\ &= k_1 F(a|x) - k_0 + k_0 F(a|x)\end{aligned}$$

que igualada a 0, tiene como solución cualquier valor de a tal que

$$F(a|x) = k_0/(k_0 + k_1)$$

como la derivada segunda respecto de a de $r(a|x)$ es

$$\frac{\partial^2}{(\partial a)^2}r(a|x) = k_1\pi(a|x) + k_0\pi(a|x)$$

que es positiva para todo a, el cuantil de orden $k_0|(k_0 + k_1)$ se corresponde con un mínimo.

Obsérvese que si $k_0 = k_1 = 1$ la función de pérdida es

$$L(\theta, a) = |\theta - a|$$

y se obtiene como caso particular de lo visto, que cualquier mediana de la distribución final es entonces la *regla Bayes* (véase el Ejercicio 3.7.14).

6.8.18

a) Para a_1 *trayectoria baja* la función de pérdida viene dada por

$$L((\theta, \xi), a_1) = |l| \qquad \text{con} \quad -\theta \leq l \leq \theta \ \text{ y } \ \theta \leq 1.$$

Donde θ es el porcentaje de zona cubierta de nubes y ξ la probabilidad de que el misil sea interceptado. La función de riesgo es

$$\begin{aligned}R((\theta, \theta), a_1) &= \int_{-\theta}^{+\theta} |l| \frac{1}{2\theta} dl \\ &= \frac{1}{2\theta} 2 \int_0^{\theta} l dl \\ &= \frac{\theta}{2} \qquad\qquad \text{con} \ \ \theta \leq 1\end{aligned}$$

y el *riesgo Bayes*

$$\begin{aligned}B(\pi_1, a_1) &= \int_0^1 R((\theta, \xi), a_1)\pi_1|\theta) d\theta \\ &= \int_0^1 \frac{\theta}{2} 6\theta(1 - \theta) d\theta \\ &= \frac{1}{4}.\end{aligned}$$

Para a_2 *trayectoria alta* la función de pérdida viene dada por

$$L((\theta,\xi),a_2)=\xi\ 1+(1-\xi)0 \qquad \text{con} \quad 0\le\xi\le 1$$

ya que, o bien el misil es alcanzado, lo que ocurre con probabilidad ξ, o no es alcanzado, en cuyo caso da exactamente en el blanco y no se pierde nada.

Por lo que la función de riesgo es

$$R((\theta,\xi),a_2)=\xi \quad \text{con} \quad 0\le\xi\le 1$$

y el riesgo Bayes

$$B(\pi_2,a_2)=\int_0^1 \xi\pi_2(\xi)d\xi.$$

Al realizar los dos lanzamientos se actualiza la densidad π_2 mediante

$$\begin{aligned}\pi_2(\xi|\text{dos éxitos}) &= \frac{\pi_2(\xi)P\{X_1=1,X_2=1|\xi\}}{\int_0^1\pi_2(\xi)P\{X_1=1,X_2=1|\xi\}d\xi}\\ &= \frac{2(1-\xi)(1-\xi)^2}{\int_0^1 2(1-\xi)(1-\xi)^2d\xi}\\ &= 4(1-\xi)^3 \qquad\qquad \text{con} \quad 0\le\xi\le 1,\end{aligned}$$

de donde

$$\begin{aligned}B(\pi_2,a_2) &= \int_0^1 \xi 4(1-\xi)^3d\xi\\ &= 4\ Beta(2,4)\\ &= \frac{1}{5}.\end{aligned}$$

b) Para la obtención de la *regla minimax*, el $\max\limits_{\theta} R((\theta,\xi),a_1)=\max\limits_{0\le\theta\le 1}\theta/2=\frac{1}{2}$ y con la regla a_2, $\max\limits_{\xi}R(\xi,a_2)=\max\limits_{0\le\xi\le 1}\xi=1$, luego la *regla minimax* es a_1 tirar con *trayectoria baja*.

c) La *regla Bayes* es la que minimiza el *riesgo Bayes*, por lo tanto a_1.

d) Las dos reglas son admisibles ya que dependiendo de los valores de θ y de ξ se puede hacer que $R((\theta_1,\xi_1),a_1)<R((\theta_1,\xi_1),a_2)$ mientras que para otra elección sea $R((\theta_2,\xi_2),a_1)>R((\theta_2,\xi_2),a_2)$.

6.8.19

Se trata de comprobar que dada cualquier otra regla $d'(\overrightarrow{x})$ es

$$\sup_{\theta\in\Theta} R(\theta,d)\le\sup_{\theta\in\Theta} R(\theta,d').$$

Por ser d^{π_n} *Bayes* frente a π_n es

$$B(\pi_n,d^{\pi_n})\le B(\pi_n,d')$$

y tomando límites

$$p = R(\theta, d) = \sup_{\theta \in \Theta} R(\theta, d) \leq \lim_{n \to r} B(\pi_n, d')$$

pero

$$\begin{aligned} \lim_{n \to \infty} B(\pi_n, d') &\leq \lim_{n \to \infty} \int_{\theta} \sup_{\theta} R(\theta, d') \pi_n(\theta) d\theta \\ &= \sup_{\theta} R(\theta, d') \int_{\Theta} \pi_n(\theta) d\theta \\ &= \sup_{\theta} R(\theta, d'). \end{aligned}$$

Obsérvese que este enunciado generaliza el Teorema 6.3.2.

Además el enunciado y la demostración son válidos también cuando θ y d son vectores.

Capítulo 7

ANÁLISIS DE LA VARIANZA Y REGRESIÓN LINEAL

7.1. INTRODUCCIÓN

En el Capítulo 5 se abordó el problema de comparar las medias de dos poblaciones, en la práctica es frecuente el problema más general de comparar las medias en un número mayor de dos poblaciones, así con frecuencia se trata de comparar distintos tratamientos médicos, diferentes procesos de fabricación, etc. Este Capítulo es una introducción al tratamiento de estos problemas y a los métodos que se emplean para su resolución que reciben el nombre de *análisis de varianza*. Contrariamente a lo que parece significar, se trata de comprobar si varias medias poblacionales son iguales. El Capítulo empieza estudiando el análisis de la varianza con uno y dos factores, mediante técnicas basadas en la distribución normal, dejándose su alternativa no paramétrica para el Capítulo siguiente, Sección 8.5.4.

La segunda parte del Capítulo se dedica al estudio de la regresión, sin duda la más popular herramienta estadística. El principal objetivo de la regresión es el estudio de la dependencia de una variable respecto de otra. Por último, y en homogeneidad con los distintos temas, se aborda la aproximación bayesiana a estos problemas.

7.2. ANÁLISIS DE LA VARIANZA CON UN SÓLO FACTOR

Ya se han obtenido contrastes de hipótesis para comparar las medias de dos poblaciones normales cuando las varianzas se suponen iguales (véase la Sección 5.2.2). En el análisis de la varianza se pretende contrastar $H_0 : \theta_1 =$

$\theta_2 = ... = \theta_k$ para k poblaciones normales, independientes, y con la misma varianza σ^2 desconocida. Tal situación es de gran interés práctico sobre todo en experimentos relacionados con la agricultura.

Se supone entonces el modelo

$$x_{ij} = \theta_i + \varepsilon_{ij} \qquad i = 1, ..., k \qquad j = 1, ..., n_i \tag{7.2.1}$$

donde $\varepsilon_{ij} \sim N(0, \sigma)$ e independientes. La introducción del modelo puede ser mejor comprendida mediante un ejemplo numérico. En la preparación de una sustancia se utilizan cuatro modificaciones de un procedimiento estándar, los datos están contenidos en la tabla 7.2.1 donde la primera columna se corresponde con el procedimiento estándar.

Métodos				
1	2	3	4	5
63	27	10	28	47
48	41	30	45	28
57	23	24	44	30
54	31	37	31	35
41	2	40	29	21
43	33	24	44	37

Tabla 7.2.1. Datos correspondientes a la preparación de una sustancia mediante 5 procedimientos de fabricación.

Su representación mediante un diagrama de cajas viene recogida en la Figura 7.2.1, donde se aprecia que el primer método es distinto de los cuatro restantes. Además la representación del segundo método no se ha cerrado inferiormente por incluir el valor 2, muy diferente de los restantes, en una aplicación debería revisarse este valor que aquí se va a admitir como correcto.

La función de verosimilitud, viene dada por

$$f(x_{11}, ..., x_{kn_k} | \theta_1, ..., \theta_k, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \theta_i)^2} \tag{7.2.2}$$

con $n = \sum_{j=1}^{k} n_i$ el número total de observaciones; de donde los estimadores de máxima verosimilitud son:

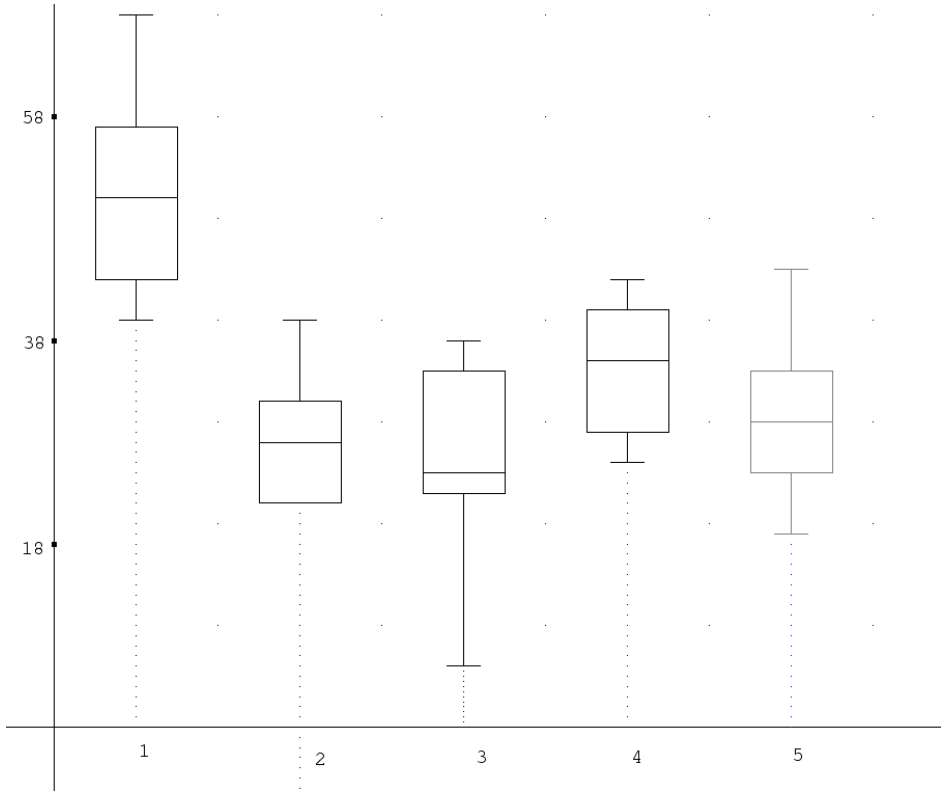


Figura 7.2.1. Diagrama de cajas correspondiente al experimento de la Tabla 7.2.1.

$$\widehat{\theta}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij} = \overline{x}_{i\cdot}$$

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \overline{x}_{i\cdot})^2 = \frac{1}{n}\sum_{i=1}^{k}(n_i - 1)s_i^2$$

siendo s_i^2 las cuasivarianzas muestrales. Obsérvese que esta última expresión no generaliza la varianza ponderada del caso $k = 2$ $\widehat{\sigma}_p^2$ (véase la Expresión (5.2.2.4)) ya que la fórmula análoga a la varianza ponderada es

$$\widehat{\sigma}_p^2 = \frac{\sum_{i=1}^{k}(n_i - 1)s_i^2}{n - k} \tag{7.2.4}$$

que aquí recibe el nombre de varianza residual (véase el Ejercicio 7.7.4).

Se llama residuo a la cantidad $x_{ij} - \widehat{\theta}_i = x_{ij} - \overline{x}_{i\cdot}$ que es el valor observado x_{ij} menos el valor estimado de θ_i. Se verifica que $\widehat{\sigma}_p^2$ es un estimador

centrado de σ^2, pues

$$E[\widehat{\sigma}_p^2] = \frac{1}{n-k}\sum_{i=1}^{k}(n_i-1)E[S_i^2]$$
$$= \frac{1}{n-k}\sum_{i=1}^{k}(n_i-1)\sigma^2 = \sigma^2.$$

Por último $n\frac{\widehat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-k}$; ya que $(n_i-1)s_i^2/\sigma^2 \sim \chi^2_{n_i-1}$ y por tanto

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{(x_{ij}-\overline{x}_{i\cdot})^2}{\sigma^2} = \sum_{i=1}^{k}(n_i-1)\frac{s_i^2}{\sigma^2} = \sum_{i=1}^{k}\chi^2_{n_i-1} \sim \chi^2_{n-k}.$$

La idea original del análisis de la varianza se debe a Fisher (1925) y está basada en la descomposición

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\overline{x}_{\cdot\cdot})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\overline{x}_{i\cdot})^2 + \sum_{i=1}^{k} n_i(\overline{x}_{i\cdot}-\overline{x}_{\cdot\cdot})^2 \tag{7.2.5}$$

donde $\overline{x}_{\cdot\cdot}$ es la media total

$$\overline{x}_{\cdot\cdot} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i}x_{ij} = \frac{1}{n}\sum_{i=1}^{k} n_i\overline{x}_{i\cdot}$$

La Expresión (7.2.5) se obtiene sin más que restar y sumar dentro del paréntesis del primer miembro de (7.2.5) $\overline{x}_{i\cdot}$ y desarrollando a continuación el cuadrado. Los tres términos que aparecen en la Expresión (7.2.5) reciben, respectivamente, el nombre de *suma de cuadrados total*, *suma de cuadrados residual* o *dentro* de los grupos y *suma de cuadrados explicada* o *entre* los grupos. Esto es

$$SC_T = SC_D + SC_E \tag{7.2.6}$$

con

(7.2.7)

$$SC_T = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\overline{x}_{\cdot\cdot})^2, \quad SC_D = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\overline{x}_{i\cdot})^2, \quad SC_E = \sum_{i=1}^{k} n_i(\overline{x}_{i\cdot}-\overline{x}_{\cdot\cdot})^2$$

La interpretación de la Expresión (7.2.5) es que la variación de cada elemento x_{ij} a la media total $\overline{x}_{\cdot\cdot}$ se puede explicar mediante la variación de cada elemento a la media de su grupo, más la debida a las medias en los distintos grupos.

Se pretende a continuación encontrar la distribución de las sumas de cuadrados. para ello se necesita el siguiente lema.

Lema 7.2.1
Si se tiene la colección Z_i de v.a.i. con $E[Z_i] = \theta_i$, $V[Z_i] = \sigma^2 \quad i = 1, ..., k$, entonces

$$E[(Z_i - \overline{Z})^2] = (\theta_i - \overline{\theta})^2 + \frac{k-1}{k}\sigma^2$$

donde

$$\overline{Z} = \frac{1}{k}\sum_{i=1}^{k} Z_i \qquad \overline{\theta} = \frac{1}{k}\sum_{i=1}^{k} \theta_i$$

Demostración: En efecto

$$E[(Z_i - \overline{Z})^2] = V[Z_i - \overline{Z}] + E^2[Z_i - \overline{Z}]$$

y claramente $E[Z_i - \overline{Z}] = \theta_i - \overline{\theta}$, mientras que

$$\begin{aligned} V[Z_i - \overline{Z}] &= V[Z_i] + V[\overline{Z}] - 2Cov(Z_i, \overline{Z}) \\ &= \sigma^2 + \frac{\sigma^2}{k} - 2\frac{\sigma^2}{k}. \end{aligned}$$

■

La esperanza de la *suma de cuadrados dentro* de los grupos, viene dada por

$$E[SC_E] = \sum_{i=1}^{k} n_i E[(\overline{X}_{i\cdot} - \overline{X}_{\cdot\cdot})^2],$$

pues, aplicando el Lema anterior con $Z_i = \overline{X}_{i\cdot}$, $\overline{Z} = \overline{X}_{\cdot\cdot}$, $E[Z_i] = \theta_i$ y $V[Z_i] = \sigma^2/n_i$, se tiene que

$$\begin{aligned} E[SC_E] &= \sum_{i=1}^{k} n_i \left[(\theta_i - \overline{\theta})^2 + \frac{k-1}{k}\frac{\sigma^2}{n_i}\right] \\ &= \sum_{i=1}^{k} n_i(\theta_i - \overline{\theta})^2 + (k-1)\sigma^2. \end{aligned}$$

Por lo tanto $E[SC_D/(n-k)] = \sigma^2$. Si se cumple la hipótesis $H_0 : \theta_1 = ... = \theta_k$ entonces también es $E[SC_E/(k-1)] = \sigma^2$ y se tienen dos estimadores insesgados de σ^2. Por el contrario; si alguno de los valores θ_i es diferente, la esperanza de $SC_E/(k-1)$ será estrictamente mayor que σ^2. Esto proporciona la intuición para realizar el contraste de que todos los valores $\theta_i \quad i = 1, ..., k$ son iguales, una vez que se disponga de la distribución de las dos sumas de cuadrados.

Teorema 7.2.1
Si $x_{ij} = \theta_i + \varepsilon_{ij} \quad i = 1, ..., k,\ j = 1, ..., n_i \quad n = \sum_{i=1}^{k} n_i$ con $\varepsilon_{ij} \sim N(0, \sigma)$ e independientes, entonces $SC_D/\sigma^2 \sim \chi^2_{n-k}$. Si además se cumple $H_0 : \theta_1 = ... = \theta_k$ entonces $SC_E/\sigma^2 \sim \chi^2_{k-1}$, con ambas distribuciones χ^2 independientes.

Demostración: la *suma de cuadrados dentro* o *residual* dividida por la varianza es

$$\frac{SC_D}{\sigma^2} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{(x_{ij} - \theta_{i.})^2}{\sigma^2} = \sum_{i=1}^{k} (n_i - 1) \frac{s_i^2}{\sigma^2}$$

y como, según el teorema de Fisher, véase el Teorema 1.2.5.1, cada sumando de la anterior expresión se distribuye $\chi^2_{n_i-1}$ y obviamente son independientes se sigue que $SC_D/\sigma^2 \sim \chi^2_{n-k}$; según la propiedad de aditividad de la distribución χ^2.

En el caso de que sea cierta la hipótesis $H_0 : \theta_1 = ... = \theta_k$, la *suma de cuadrados entre* se puede poner

$$SC_E = \sum_{i=1}^{k} n_i (\overline{x}_{i.} - \theta_i - (\overline{x}_{..} - \overline{\theta}_{..}))^2 \tag{7.2.8}$$

donde es $\overline{x}_{..} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij}/n$ y $\overline{\theta}_{..} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \theta_i/n$; por lo tanto la aplicación del Lema 1.5.2.1 a las variables $X_1, ..., X_n, Y_1$ con $X_i = \sqrt{n_i}(\overline{x}_{i.} - \theta_i)$ e

$$Y_1 = \sqrt{n}(\overline{x}_{..} - \overline{\theta}_{..}) = \left(\sqrt{\frac{n_1}{n}}, ..., \sqrt{\frac{n_k}{n}} \right) \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix}$$

(con lo que $X_i \sim N(0, \sigma)$), permite concluir que

$$\begin{aligned} Z &= \sum_{i=1}^{k} X_i^2 - Y_1^2 = \sum_{i=1}^{k} n_i (\overline{x}_{i.} - \theta_i)^2 - n(\overline{x}_{..} - \overline{\theta}_{..})^2 \\ &= \sum_{i=1}^{k} n_i (\overline{x}_{i.} - \theta_i - (\overline{x}_{..} - \overline{\theta}_{..}))^2 \end{aligned}$$

y por tanto $Z/\sigma^2 = SC_E/\sigma^2 \sim \chi^2_{k-1}$ e independiente de $\overline{x}_{..} - \overline{\theta}_{..}$. Por último las dos χ^2 son independientes, ya que la SC_E es función de $\overline{x}_{1.}, ..., \overline{x}_{k.}$ y SC_D depende de S_i^2 que es independiente de $\overline{x}_{i.}$ para $i' \neq i$ y cuando $i' = i$ S_i^2 es independiente de $\overline{x}_{i.}$, por el teorema de Fisher (véase el teorema 1.5.1). Como consecuencia, para contrastar la hipótesis nula $H_0 : \theta_1 = ... = \theta_k$ de que todas las variables tienen la misma media, frente a su complementaria; el test a utilizar tiene de región crítica

$$RC = \left\{ \frac{SC_E/\, k-1}{SC_D/\, n-k} > f_{k-1,n-k;\alpha} \right\} \tag{7.2.9}$$

ya que bajo H_0 el numerador y el denominador son estimadores de σ^2 y por tanto el cociente es uno. Por el contrario, si algún θ_i es diferente a los demás el numerador aumenta en la cantidad $(\theta_i - \overline{\theta}_{..})^2$. Además, en (7.2.9) $f_{k-1,n.k;\alpha}$ es el cuantil, de orden $1 - \alpha$, de la distribución de Snedecor de $k - 1$ y $n - k$ grados de libertad (véase la Definición 1.5.4.1).

Los datos acostumbran a presentarse en la forma de *tabla de análisis de la varianza* o tabla ADEVA (véase la Tabla 7.2.2).

Fuente de variación	Suma de cuadrados	Grados de libertad	Estadístico F
Entre grupos	$SC_E = \sum_{i=1}^{k} n_{i.}(\overline{x}_{i.} - \overline{x}_{..})^2$	$k-1$	$SC_E/k-1$
Dentro de los grupos o residual	$SC_D = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \overline{x}_{i.})^2$	$n-k$	$SC_D/n-k$
Total	$SC_T = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \overline{x}_{..})^2$	$n-1$	

Tabla 7.2.2. Tabla del análisis de la varianza (ADEVA) para un solo factor.

Ejemplo 7.2.1
Para los datos de la preparación de una sustancia química contenidos en la tabla 7.2.1, la tabla ADEVA viene dada por

Fuente de variación	Suma de cuadrados	Grados de libertad	Estadístico F
Entre grupos	2385′53	4	5′81
Dentro de los grupos	2567′17	25	
Total	4952′70	29	

La región crítica es $RC = \left\{ \frac{SC_E/k-1}{SC_D/n-k} > f_{4,25;0'05} = 2'76 \right\}$, donde el cuantil de la F se ha obtenido interpolando en la Tabla 6. Esto lleva a rechazar la hipótesis nula de igualdad de las medias.
■

El anterior ejemplo da pie al planteamiento de la siguiente pregunta: correcto, las medias no son todas iguales, pero ¿es posible que los métodos 2 y 5 sean iguales y todos los demás sean diferentes? Esto es, se puede contrastar la hipótesis nula, $H_0 : \theta_2 = \theta_5$ frente a $H_1 : \theta_2 \neq \theta_5$. Esta hipótesis es un caso particular de contrastes de la forma $H_0 : \sum_{i=1}^{k} a_i\theta_i = 0$ frente a $H_1 : \sum_{i=1}^{k} a_i\theta_i \neq 0$ siendo a_i un conjunto de constantes tales que $\sum_{i=1}^{k} a_i = 0$; basta tomar en este caso $a = (0, 1, 0, 0, -1)$.

Se tiene que $\sum_{i=1}^{k} a_i\overline{X}_{i_0}$ es una variable con distribución normal de media

$$E\left[\sum_{i=1}^{k} a_i\overline{X}_{i_0}\right] = \sum_{i=1}^{k} a_i\theta_i$$

y de varianza

$$V\left[\sum_{i=1}^{k} a_i \overline{X}_{i_0}\right] = \sigma^2 \sum_{i=1}^{k} \frac{a_i^2}{n_i}$$

con lo que

$$\frac{\sum_{i=1}^{k} a_i \overline{X}_{i_0} - \sum_{i=1}^{k} a_i \theta_i}{\sigma \sqrt{\sum_{i=1}^{k} a_i^2/n_i}} \sim N(0,1).$$

Como normalmente la varianza σ^2 será desconocida, se emplea su estimador centrado $\widehat{\sigma}_p^2$ (véase (7.2.4)) basado en la información suministrada por todos los grupos. Con ello se obtiene que

$$\frac{\sum_{i=1}^{k} a_i \overline{X}_{i_0} - \sum_{i=1}^{k} a_i \theta_i}{\widehat{\sigma}_p \sqrt{\sum_{i=1}^{k} a_i^2/n_i}} \sim T_{n-k} \tag{7.2.10}$$

es decir se distribuye según una distribución de Student con $n-k$ grados de libertad. Por tanto la región crítica del contraste es

$$RC = \left\{ \left| \frac{\sum_{i=1}^{k} a_i \overline{X}_{i_0}}{\widehat{\sigma}_p \sqrt{\sum_{i=1}^{k} a_i^2/n_i}} \right| \geq t_{n-k;\alpha/2} \right\} \tag{7.2.11}$$

Además (7.2.10) puede ser utilizada como variable pivotal para construir un intervalo de confianza para $\sum_{i=1}^{k} a_i \theta_i$

(7.2.12)

$$IC_{1-\alpha}\left(\sum_{i=1}^{k} a_i \theta_i\right) =$$
$$\left(\sum_{i=1}^{k} a_i \overline{X}_{i_0} - t_{n-k;\alpha/2} \widehat{\sigma}_p \sqrt{\sum_{i=1}^{k} a_i^2/n_i}, \sum_{i=1}^{k} a_i \overline{X}_{i_0} + t_{n-k;\alpha/2} \widehat{\sigma}_p \sqrt{\sum_{i=1}^{k} a_i^2/n_i}\right)$$

Ejemplo 7.2.2
Con los datos de la preparación de una sustancia química de la Tabla 7.2.1, para contrastar $H_0 : \theta_2 = \theta_5$ frente a $H_1 : \theta_2 \neq \theta_5$, la región crítica es

$$RC = \left\{ |\overline{x}_{2\cdot} - \overline{x}_{5\cdot}| \geq \widehat{\sigma}_p \sqrt{\frac{1}{n_2} + \frac{1}{n_5}} t_{n-k;\alpha/2} \right\}$$

que con $\widehat{\sigma}_p = 10'14, n_2 = n_5 = 6, t_{30-5;0'05} = 2'06$ y $\alpha = 0'1$ se convierte en

$$RC = \{|\overline{x}_{2\cdot} - \overline{x}_{5\cdot}| \geq 12'06\}.$$

Al ser $|\overline{x}_{2\cdot} - \overline{x}_{5\cdot}| = |26'17 - 33| = 6'83$, se acepta la igualdad de los dos métodos.

Como intervalo de confianza para $\theta_2 - \theta_5$ se obtiene

$$IC_{0'90}(\theta_2 - \theta_5) = (-5'23, 18'89).$$

■

Los tests de la forma $H_0 : \sum_{i=1}^{k} a_i\theta_i = 0$ frente a $H_1 : \sum_{i=1}^{k} a_i\theta_i \neq 0$ con $\sum_{i=1}^{k} a_i = 0$, permiten obtener conclusiones acerca de la ordenación de los valores θ_i, mediante la realización de varios contrastes simultáneos. Concretamente si se plantean los contrastes $H_0 : \theta_1 - \theta_2 = 0$ frente a $H_1 : \theta_1 > \theta_2$ y $H_0' : \theta_2 - \theta_3 = 0$ frente a $H_1' : \theta_2 > \theta_3$ y se rechaza la hipótesis nula en ambos casos, se obtiene que ha de ser $\theta_1 > \theta_2 > \theta_3$. Esto plantea el problema de la determinación del tamaño del test conjunto, un procedimiento sencillo de solucionarlo consiste en utilizar la desigualdad de Bonferroni, que establece que si $A_1, ..., A_n$ son sucesos cualesquiera entonces se verifica que

$$P\left(\bigcap_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} P(A_i) - (n-1) \tag{7.2.13}$$

en efecto; pasando al complementario

$$P\left(\bigcap_{i=1}^{n} A_i\right) = 1 - P\left(\bigcup_{i=1}^{n} A_i^c\right)$$

pero como

$$P\left(\bigcup_{i=1}^{n} A_i^c\right) \leq \sum_{i=1}^{n} P(A_i^c) = \sum_{i=1}^{n}[1 - P(A_i)] = n - \sum_{i=1}^{n} P(A_i)$$

multiplicando por -1 se sigue $-P\left(\bigcup_{i=1}^{n} A_i^c\right) \geq \sum_{i=1}^{n} P(A_i) - n$ de donde

$$1 - P\left(\bigcup_{i=1}^{n} A_i^c\right) \geq 1 - n + \sum_{i=1}^{n} P(A_i)$$

y como el primer miembro de esta desigualdad es $P\left(\bigcap_{i=1}^{n} A_i\right)$, se tiene demostrada la desigualdad.

Para realizar simultáneamente m contrastes, cuando cada uno de ellos tiene de tamaño $\alpha_i = \alpha/m$, si se contruyen

$$RA = \bigcap_{i=1}^{m} RA_i \qquad \text{y} \qquad RC = \bigcup_{i=1}^{m} RC_i$$

como la región de aceptación y la región crítica, respectivamente del test conjunto, es claro que en virtud de la Desigualdad (7.2.13)

$$\begin{aligned} P\{RC|H_0\} &= P\left\{\bigcup_{i=1}^{m} RC_i|H_0\right\} \\ &= 1 - P\left\{\bigcap_{i=1}^{m} RC_i^c|H_0\right\} \\ &\leq 1 - \sum_{i=1}^{m} P\{RC_i^c|H_0\} + m - 1 \end{aligned}$$

y como, al ser $P\{RC_i|H_0\} \leq \alpha_i$, es $-P\{RC_i^c|H_0\} \leq \alpha_i - 1$, por lo que sustituyendo se tiene

$$\begin{aligned} P\{RC|H_0\} &\leq \sum_{i=1}^{m} \left(\frac{\alpha}{m} - 1\right) + m \\ &= \alpha. \end{aligned}$$

Ejemplo 7.2.3
Para los datos de la preparación de la sustancia química contenidos en la Tabla 7.2.1, si se quiere contrastar simultáneamente $H_0 : \theta_1 = \theta_2$ frente a $H_1 : \theta_1 > \theta_2$ y $H_0' : \theta_2 = \theta_3$ frente a $H_1' : \theta_2 > \theta_3$, mediante un test de tamaño $\alpha = 0'05$. Basta tomar $\alpha_1 = \alpha_2 = \alpha/2 = 0'025$ con lo que las regiones críticas son

$$\begin{aligned} RC_1 &= \{\overline{x}_{1\cdot} - \overline{x}_{2\cdot} \geq \widehat{\sigma}_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \mathrm{t}_{n-k;0'025}\} \\ &= \{\overline{x}_{1\cdot} - \overline{x}_{2\cdot} \geq 4'136 \cdot 0'5773 \cdot 2'06\} \\ &= \{\overline{x}_{1\cdot} - \overline{x}_{2\cdot} \geq 4'919\} \end{aligned}$$

análogamente

$$RC_2 = \{\overline{x}_{2\cdot} - \overline{x}_{3\cdot} \geq 4'919\}$$

con lo que la región crítica conjunta es

$$RC = \{\overline{x}_{1\cdot} - \overline{x}_{2\cdot} \geq 4'919 \quad \text{o} \quad \overline{x}_{2\cdot} - \overline{x}_{3\cdot} \geq 4'919\}$$

por lo que al ser $\overline{x}_{1\cdot} - \overline{x}_{2\cdot} = 51 - 26'167 = 24'833$ se rechaza H_0 y como $\overline{x}_{2\cdot} - \overline{x}_{3\cdot} = 26'167 - 27'5 = -1'333$ se acepta H_0'; en consecuencia se tiene que $\theta_1 > \theta_2 = \theta_3$.
■

Existen otros procedimientos para modificar los tamaños del test conjunto, los más conocidos son, además del explicado de emplear la desigualdad de Bonferroni, el método de Scheffé, el de Tuckey y el de *la mínima diferencia significativa*; pueden verse en Stuart y Ord (1991), parágrafo 29.47 y siguientes.

Con el fin de extender el modelo del análisis de la varianza a situaciones prácticas de mayor complejidad, se necesita introducir primero un parámetro más en la modelización. En la introducción al análisis de la varianza con un sólo factor se ha supuesto el modelo de la forma (7.2.1), si bien existe un modelo alternativo expresado mediante

$$x_{ij} = \mu + \theta_i + \varepsilon_{ij} \quad i = 1, ..., k \quad j = 1, ..., n_i \tag{7.2.14}$$

donde de nuevo los errores $\varepsilon_{ij} \sim N(0, \sigma)$ son normales e independientes.

En este modelo, el parámetro μ representa una media general o un efecto global. Es decir, el efecto de cada tratamiento se descompone en un valor común μ y un efecto debido al tratamiento i que viene representado por θ_i. Como se ha introducido un parámetro más μ, se debe introducir también una relación más entre los parámetros con el fin de poder determinarlos todos, y una manera sencilla de hacerlo es mediante la expresión

$$\sum_{i=1}^{k} \theta_i = 0. \tag{7.2.15}$$

De no hacerse así, sería posible encontrar valores particulares de $\mu, \theta_1, ..., \theta_k$ tales que den lugar a la misma función $f(x_{11}, ..., x_{kn_k}|\mu_1, \theta_1, ..., \theta_k)$ para cualesquiera valores de $x_{11}, ..., x_{kn_k}$, se dice entonces que el modelo es no identificable. En concreto se tendría que siendo f la función de densidad explicitada en la Expresión (7.2.14) $f(x_{11}, ..., x_{kn_k}|0, 1, ..., 1) = f(x_{11}, ..., x_{kn_k}|1, 0, ..., 0)$. Es claro que al imponer la condición (7.2.15) el punto $\mu = 0 \quad \theta_1 = ... = \theta_k = 1$ queda eliminado.

Ahora, se está en condiciones de desarrollar las ideas correspondientes al análisis de la varianza con dos factores.

7.3. ANÁLISIS DE LA VARIANZA CON DOS FACTORES

En esta Sección se supone que el modelo es de la forma

$$x_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \tag{7.3.1}$$

con $i = 1, ..., r \quad j = 1, ..., s \quad k = 1, ..., t$ donde $\varepsilon_{ijk} \sim N(0, \sigma)$ e independientes. Cuando $t > 1$ se dice que el modelo es *balanceado* o con *replicación.*

Al igual que se hizo en el caso de un solo factor, se introducen las restricciones

$$\begin{aligned} &\sum_{i=1}^{r} \alpha_i = 0 \qquad \sum_{j=1}^{s} \beta_j = 0 \\ &\sum_{i=1}^{r} \gamma_{ij} = \sum_{j=1}^{s} \gamma_{ij} = 0 \quad i = 1, ..., r \quad j = 1, ..., s \end{aligned} \tag{7.3.2}$$

con el fin de que el modelo sea identificable.

La función de verosimilitud viene dada por

$$\begin{aligned} &f(x_{111}, ..., x_{rst} | \mu, \alpha_1, ..., \alpha_r, \beta_1, ..., \beta_s, \sigma) \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{rst}{2}} e^{-\frac{1}{2\sigma^2} \sum_i \sum_j \sum_k (x_{ijk} - \mu - \alpha_i - \beta_j - \gamma_{ij})^2} \end{aligned}$$

Es fácil verificar que los estimadores de máxima verosimilitud sujetos a las Restricciones (7.3.2) son

$$\begin{aligned} &\widehat{\mu} = \overline{x}_{...} \\ &\widehat{\alpha}_i = \overline{x}_{i..} - \overline{x}_{...} \quad i = 1, ..., r \\ &\widehat{\beta}_j = \overline{x}_{.j.} - \overline{x}_{...} \quad j = 1, ..., s \\ &\widehat{\gamma}_{ij} = \overline{x}_{ij.} - \overline{x}_{i..} - \overline{x}_{.j.} + \overline{x}_{...} \quad i = 1, ..., r \quad j = 1, ..., s \end{aligned} \tag{7.3.3}$$

Análogamente al caso de un único factor, las sumas de cuadrados debidos a los distintos efectos son

$$\begin{aligned} SC_A &= st \sum_{i=1}^{r} (\overline{x}_{i..} - \overline{x}_{...})^2 \\ SC_B &= rt \sum_{j=1}^{s} (\overline{x}_{.j.} - \overline{x}_{...})^2 \\ SC_{AB} &= t \sum_{i=1}^{r} \sum_{j=1}^{s} (\overline{x}_{ij.} - \overline{x}_{i..} - \overline{x}_{.j.} + \overline{x}_{...})^2 \\ SC_E &= \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} (x_{ijk} - \overline{x}_{ij.})^2 \\ SC_T &= \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} (x_{ijk} - \overline{x}_{...})^2 \end{aligned} \tag{7.3.4}$$

Puede comprobarse que

$$SC_T = SC_A + SC_B + SC_{AB} + SC_E$$

sin más que elevar el cuadrado y desarrollar la siguiente expresión

$$\begin{aligned} x_{ijk} - \overline{x}_{...} = &(x_{ijk} - \overline{x}_{ij.}) + (\overline{x}_{i..} - \overline{x}_{...}) \\ &+ (\overline{x}_{.j.} - \overline{x}_{...}) + (\overline{x}_{ij.} - \overline{x}_{i..} - \overline{x}_{.j.} + \overline{x}_{...}) \end{aligned}$$

El siguiente resultado, permite realizar los contrastes adecuados a esta situación

Teorema 7.3.1
Si $x_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \quad i = 1, ..., r \quad j = 1, ..., s \quad k = 1, ..., t$ con $n = rst$ el número de observaciones y donde los errores $\varepsilon_{ijk} \sim N(0, \sigma)$ son normales e independientes, entonces $SC_E/\sigma^2 \sim \chi^2_{rs(t-1)}$. Si se cumple $H_A : \alpha_1 = ... = \alpha_r = 0$, se sigue que $SC_A/\sigma^2 \sim \chi^2_{r-1}$ y ambas χ^2 son independientes. Si, además, se cumple $H_B : \beta_1 = ... = \beta_s = 0$ entonces $SC_B/\sigma^2 \sim \chi^2_{s-1}$ y es independiente de SC_R/σ^2. Por último si se cumple que $H_{AB} : \gamma_{ij} = 0$, $i = 1, ..., r$, $j = 1, ..., s$, entonces $SC_{AB}/\sigma^2 \sim \chi^2_{(r-1)(c-1)}$ e independiente de SC_E/σ^2.

La demostración puede verse en Stuart y Ord (1991), Parágrafo 29.15. Normalmente los cálculos se presentan en la llamada *tabla de la varianza con dos factores* o tabla ADEVA con dos factores, Tabla 7.3.1.

Fuentes de variación	Suma de cuadrados	Grados de libertad	Estadísticos F
Efecto fila	$SC_A = st \sum_{i=1}^{k} (\overline{x}_{i..} - \overline{x}_{...})^2$	r-1	$\frac{SC_A/r-1}{SC_E/rs(t-1)}$
Efecto columna	$SC_B = rt \sum_{j=1}^{s} (\overline{x}_{.j.} - \overline{x}_{...})^2$	s-1	$\frac{SC_B/s-1}{SC_E rs(t-1)}$
Interacción	$SC_{AB} = t \sum_{i=1}^{r} \sum_{j=1}^{s} (\overline{x}_{ij.} - \overline{x}_{i..} - \overline{x}_{.j.} + \overline{x}_{...})^2$	(r-1)(s-1)	$\frac{SC_{AB}/(r-1)(s-1)}{SC_E/rs(t-1)}$
Error	$SC_E = \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} (x_{ijk} - \overline{x}_{ij.})^2$	rs(t-1)	
Total	$SC_T = \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} (x_{ijk} - \overline{x}_{...})^2$	rst-1	

Tabla 7.3.1. Tabla del análisis de la varianza para dos factores (ADEVA).

Por lo tanto, para contrastar la hipótesis nula de que el efecto fila es nulo

$$H_A : \alpha_1 = ... = \alpha_r = 0 \text{ frente a } H'_A : \text{algún efecto fila es no nulo}$$

la región crítica es

$$RC = \left\{ \frac{SC_A/r - 1}{SC_E/rs(t-1)} \geq f_{r-1, rs(t-1); \alpha} \right\} \tag{7.3.5}$$

Análogamente, para contrastar que el efecto columna es nulo

$$H_B : \beta_1 = ... = \beta_r = 0 \text{ frente a } H'_B : \text{algún efecto fila es no nulo}$$

la región crítica es

$$RC = \left\{ \frac{SC_B/s - 1}{SC_E/rs(t-1)} \geq f_{s-1, rs(t-1); \alpha} \right\} \tag{7.3.6}$$

Por último, para contrastar que la interacción es nula

$H_{AB}: \gamma_{ij} = 0 \quad i = 1, ..., r \quad j = 1, ..., s$ frente a H'_{AB} : alguna interacción no es nula

la región crítica es

$$RC = \left\{ \frac{SC_{AB}/(r-1)(s-1)}{SC_E/rs(t-1)} \geq f_{(r-1)(s-1),rs(t-1);\alpha} \right\} \tag{7.3.7}$$

Ejemplo 7.3.1
Se experimentan en una clase tres métodos de explicación impartidos por dos profesores diferentes, tras hacer las correspondientes asignaciones aleatorias, se obtienen las calificaciones siguientes, al final del proceso de explicación

	profesores	
método	$\beta_1 = 1$	$\beta_2 = 1$
$\alpha_1 = 1$	$9'5, 8'5, 7'5, 7'5$	$6, 9, 8, 7$
$\alpha_2 = 2$	$9, 8, 9'5, 8'5$	$8'5, 9, 9, 8$
$\alpha_3 = 3$	$7, 8, 8'5, 8'5$	$6, 5, 7, 5, 7, 9$

Con el fin de estudiar el efecto debido al método de explicación y el efecto debido al profesor en los resultados obtenidos, se introduce el modelo

$$x_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \quad i = 1, 2, 3 \quad j = 1, 2 \quad k = 1, 2, 3, 4$$

donde μ es un valor medio general, α_i representa el efecto relativo al método de explicación, β_j el efecto debido a cada profesor, γ_{ij} el efecto relativo a la interacción profesor-método y ε_{ijk} el error aleatorio en cada observación con $\varepsilon_{ijk} \sim N(0, \sigma)$ e independientes.

Los diagramas de cajas correspondientes a las variables *método* y *notas*, así como *profesor* y *notas* vienen recogidos en la Figura 7.3.1, donde no parecen observarse demasiadas diferencias, únicamente los cuartiles correspondientes a los datos con el segundo profesor parecen más dispersos que los correspondientes al primero.

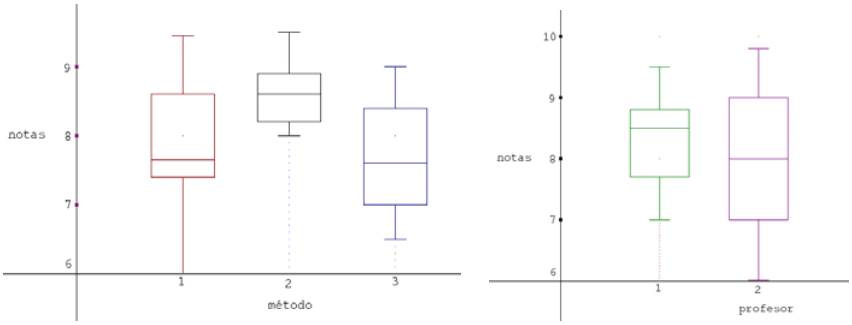


Figura 7.3.1. Diagramas de caja correspondiente al Ejemplo 7.3.1.

Los estimadores que se obtienen para los parámetros del modelo son

$$\overline{x}_{...} = 8'1042 \quad ,\widehat{\alpha}_1 = 7'875 \quad ,\widehat{\alpha}_2 = 8'687 \quad ,\widehat{\alpha}_3 = 7'75 \quad ,\widehat{\beta}_1 = 8'333 \quad ,\widehat{\beta}_2 = 7'875,$$
$$\widehat{\gamma}_{11} = 8'25 \quad ,\widehat{\gamma}_{12} = 7'5 \quad ,\widehat{\gamma}_{21} = 8'75 \quad ,\widehat{\gamma}_{22} = 8'625 \quad ,\widehat{\gamma}_{31} = 8 \quad \text{y} \quad \widehat{\gamma}_{32} = 7'5.$$

La tabla ADEVA

Fuentes de variación	Suma de cuadrados	Grados de libertad	Estadístico F
método	4'1458	2	2'54
profesor	1'2604	1	1'545
interacción	0'3958	2	0'243
error	14'6875	18	
Total	20'4896	23	

Para contrastar la hipótesis de que el efecto fila es nulo, es decir, que los tres métodos de explicación pueden ser considerados como equivalentes $H_A : \alpha_1 = \alpha_2 = \alpha_3 = 0$ con un tamaño $0'05$, la región crítica es

$$RC = \left\{ \frac{SC_A/r-1}{SC_E/rs(t-1)} \geq f_{2,18;0'05} = 3'55 \right\}$$

y como el valor obtenido para el cociente es $2'54$ se puede aceptar H_A.

Para contrastar la hipótesis de que el efecto columna es nulo, es decir, que los dos profesores consiguen los mismos resultados: $H_B : \beta_1 = \beta_2 = 0$,

$$RC = \left\{ \frac{SC_B/(s-1)}{SC_E/rs(t-1)} \geq f_{1,18;0'05} = 4'416 \right\}$$

como el cociente es $1'545$ se acepta H_B.

Por último, para contrastar si la interacción es nula: $H_{AB} : \gamma_{ij} = 0, i = 1, 2, 3,\ j = 1, 2$

$$RC = \left\{ \frac{SC_{AB}/(r-1)(s-1)}{SC_E/rs(t-1)} \geq f_{2,18;0'05} = 3'55 \right\}$$

y como el cociente es $0'243$ se acepta H_{AB}.

7.4. REGRESIÓN LINEAL

Con frecuencia, en estudios estadísticos se pretende poner una variable en función de otra u otras variables; bien porque éstas sean más sencillas de medir o bien porque se esté interesado en estudiar su relación. Así, se puede tratar de explicar la estatura de las personas en función de su edad, el área de las hojas de un árbol en función del peso de las mismas, o el peso de las personas en función de su altura y de su edad.

7.4.1. Regresión lineal simple

El caso más sencillo consiste en dada la colección de puntos (x_i, y_i) donde $i = 1, ..., n$ determinar la recta $y = \beta_0 + \beta_1 x$ que, con algún criterio, mejor se ajuste a la colección de puntos. A y se le da el nombre de *variable dependiente* o *variable respuesta* y a x se la denomina *variable independiente* o *variable predictora.*

El criterio usual que se emplea para la determinación de los parámetros β_0 y β_1 es el de los *mínimos cuadrados,* que consiste en minimizar en β_0 y β_1 la suma de cuadrados

$$SC(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \tag{7.4.1.1}$$

Geométricamente este método equivale a la minimización de la suma de los cuadrados de las distancias verticales, de acuerdo con la Figura 7.4.1.1. Aunque parece más intuitivo la minimización de la suma de las distancias ortogonales, este procedimiento no se utiliza por su mayor complejidad analítica.

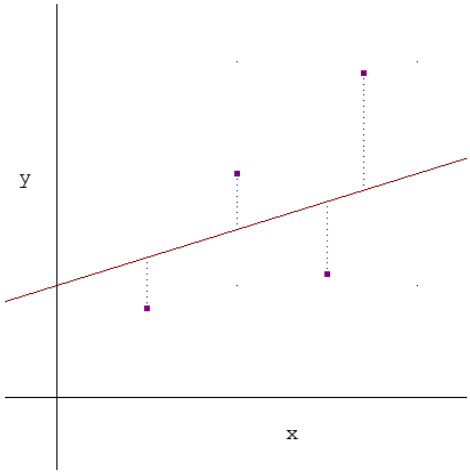


Figura 7.4.1.1. Representación de las distancias empleadas por el método de los mínimos cuadrados.

Para obtener β_0 y β_1 se debe resolver el sistema

$$\left.\begin{aligned} \frac{\partial}{\partial \beta_0} SC(\beta_0, \beta_1) &= -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = 0 \\ \frac{\partial}{\partial \beta_1} SC(\beta_0, \beta_1) &= -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) x_i = 0 \end{aligned}\right\}$$

que lleva a las llamadas *ecuaciones normales*

$$\left.\begin{aligned} \overline{y} &= \beta_0 + \overline{x}\beta_1 \\ \frac{1}{n} \sum_{i=1}^{n} x_i y_i &= \overline{x}\beta_0 + \frac{1}{n} \sum_{i=1}^{n} x_i^2 \beta_1 \end{aligned}\right\} \tag{7.4.1.2}$$

la primera ecuación, pone de manifiesto que la recta, en lo sucesivo llamada *recta de regresión,* pasa por el punto $(\overline{x}, \overline{y})$.

La solución en β_0 y β_1 del Sistema (7.4.1.2) permite construir los *estimadores de mínimos cuadrados* EMC

$$\begin{array}{l} \widehat{\beta}_0 = \overline{y} - \overline{x}\frac{b_{xy}}{b_x^2} \\ \widehat{\beta}_1 = \frac{b_{xy}}{b_x^2} \end{array} \tag{7.4.1.3}$$

donde $b_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$ y $b_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$ son los momentos respecto al centro correspondientes a la colección de puntos (x_i, y_i) $i = 1, ..., n$, (véase la Sección 1.2).

A la recta de regresión

$$y - \overline{y} = \frac{b_{xy}}{b_x^2}(x - \overline{x}) \tag{7.4.1.4}$$

se le da el nombre de *recta de regresión de y sobre x*. El mínimo de la expresión (7.4.1.1), $SC(\widehat{\beta}_0, \widehat{\beta}_1)$, es

$$\begin{array}{rl} SC(\widehat{\beta}_0, \widehat{\beta}_1) & = \sum_{i=1}^{n}(y_i - \overline{y} - \frac{b_{xy}}{b_x^2}(x_i - \overline{x}))^2 \\ & = nb_y^2 - n\frac{b_{xy}^2}{b_x^2} \\ & = nb_y^2(1 - r^2) \end{array} \tag{7.4.1.5}$$

donde r es el *coeficiente de correlación lineal*

$$r = \frac{b_{xy}}{b_x b_y} \tag{7.4.1.6}$$

(véase la Expresión (1.2.6)). La cantidad recogida en la Expresión (7.4.1.1) es siempre positiva, al ser una suma de cuadrados, por lo que el coeficiente de correlación es siempre $-1 \leq r \leq 1$. Además se interpreta r^2 como lo que queda de la variable y al restar su mejor ajuste lineal, en el sentido de los mínimos cuadrados, en función de la variable x. Valores de r próximos a $+1$ o a -1 indican una *buena* explicación de la variable y por la variable x en sentido lineal; valores próximos a cero del *coeficiente de correlación lineal* indican una *mala* explicación de la segunda variable como una función lineal de la primera.

Ejemplo 7.4.1.1
Un agricultor hace una estimación a ojo, del peso real de las naranjas contenidas en 10 árboles de un huerto, obteniendo los siguientes resultados medidos en kilogramos

Peso estimado	59	47	52	60	67	48	44	58	76	58
Peso real	61	42	50	58	67	45	39	57	71	53

con el fin de ver si tiene sentido un ajuste lineal, que explique el peso real en función del peso estimado, se recurre inicialmente a representar la colección de puntos,

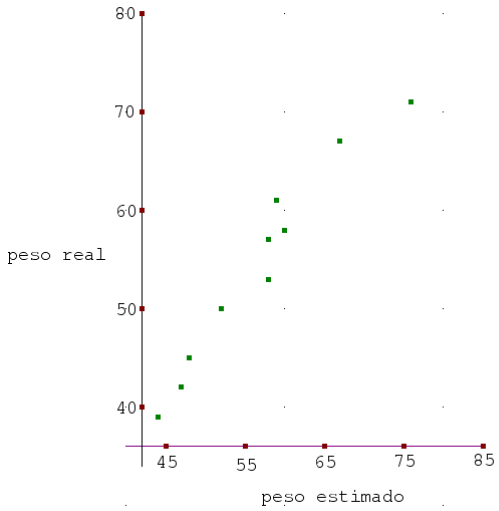


Figura 7.4.1.2. Representación de la nube de puntos del Ejemplo 7.4.1.1.

De la Figura 7.4.1.2 se desprende que el ajuste lineal puede ser correcto. Para realizar el mismo, las ecuaciones normales son

$$\left.\begin{array}{r} 54'3 = \beta_0 + 56'9\beta_1 \\ 3179'4 = 56'9\beta_0 + 3322'7\beta_1 \end{array}\right\}$$

que conducen a los estimadores

$$\widehat{\beta}_0 = 54'3 - 56'9\frac{89'73}{85'09} = -5'7028$$
$$\widehat{\beta}_1 = \frac{89'73}{85'09} = 1'0545$$

por lo que la recta de regresión es

$$y - 54'3 = 1'0545(x - 56'9)$$

o bien

$$y = -5'7028 + 1'0545x.$$

El coeficiente de correlación lineal es

$$r = \frac{89'73}{9'224 \cdot 9'99} = 0'974.$$

Lo que indica que el ajuste lineal es bastante aceptable.

■

Para encontrar las propiedades de los estimadores en el caso de la regresión lineal, se deben hacer hipótesis probabilísticas sobre el modelo, en lo que sigue se va a suponer

$$y_i = \beta_0 + \beta_1 x_i + e_i, \qquad i = 1, ..., n$$

donde las variables aleatorias e_i se interpretan como errores y se admite que son independientes con $E[e_i] = 0$ y $V[e_i] = \sigma^2$. Se hace notar que ambas variables no desempeñan un papel simétrico, la y es una v.a. por serlo e, los valores de x son conocidos independientemente de que puedan provenir de otra variable aleatoria o no. Así en este modelo se actua condicionado a estos valores de x.

Teorema 7.4.1.1
Para el modelo $y_i = \beta_0 + \beta_1 x_i + e_i$, $i = 1, ..., n$ con $E[e_i] = 0, V[e_i] = \sigma^2$ se tiene que los EMC $\widehat{\beta}_0$ y $\widehat{\beta}_1$ de β_0 y β_1 dados por

$$\widehat{\beta}_0 = \overline{y} - \overline{x}\frac{b_{xy}}{b_x^2} \qquad \widehat{\beta}_1 = \frac{b_{xy}}{b_x^2}$$

son insesgados, y se verifica que

$$\begin{aligned} V[\widehat{\beta}_0] &= \frac{\sigma^2 \frac{1}{n}\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}; \\ V[\widehat{\beta}_1] &= \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}; \\ Cov(\widehat{\beta}_0, \widehat{\beta}_1) &= -\frac{\sigma^2 \overline{x}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2} \end{aligned} \tag{7.4.1.7}$$

Demostración: De la Expresión (7.4.1.3)

$$E[\widehat{\beta}_0] = E[\overline{Y}] - \frac{\overline{x}}{b_x^2} E[b_{xy}]$$

como $b_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})y_i$. Al ser $\sum_{i=1}^{n}(x_i - \overline{x})\overline{y} = 0$, se puede poner

$$\begin{aligned} E[\widehat{\beta}_0] &= \beta_0 + \beta_1\overline{x} - \frac{\overline{x}}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sum_{i=1}^{n}(x_i - \overline{x})E[Y_i] \\ &= \beta_0 + \beta_1\overline{x} - \frac{\overline{x}}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sum_{i=1}^{n}(x_i - \overline{x})(\beta_0 + \beta_1 x_i) \\ &= \beta_0 + \beta_1\overline{x} - \frac{\overline{x}}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\beta_0\sum_{i=1}^{n}(x_i - \overline{x}) - \frac{\overline{x}}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\beta_1\sum_{i=1}^{n}(x_i - \overline{x})x_i \end{aligned}$$

pero como $\sum_{i=1}^{n}(x_i-\overline{x})=0$ y $\sum_{i=1}^{n}(x_i-\overline{x})x_i=\sum_{i=1}^{n}(x_i-\overline{x})^2$ se tiene que

$$\begin{aligned}E[\widehat{\beta}_0] &= \beta_0+\beta_1\overline{x}-\frac{\beta_1\overline{x}}{\sum_{i=1}^{n}(x_i-\overline{x})^2}\sum_{i=1}^{n}(x_i-\overline{x})^2\\ &= \beta_0.\end{aligned}$$

Análogamente

$$\begin{aligned}E[\widehat{\beta}_1] &= \sum_{i=1}^{n}\frac{(x_i-\overline{x})}{\sum_{i=1}^{n}(x_i-\overline{x})^2}E[Y_i]\\ &= \beta_0\sum_{i=1}^{n}\frac{(x_i-\overline{x})}{\sum_{i=1}^{n}(x_i-\overline{x})^2}+\beta_1\sum_{i=1}^{n}\frac{(x_i-\overline{x})x_i}{\sum_{i=1}^{n}(x_i-\overline{x})^2}\\ &= \beta_1\end{aligned}$$

por ser $\sum_{i=1}^{n}(x_i-\overline{x})=0$ y $\sum_{i=1}^{n}(x_i-\overline{x})x_i=\sum_{i=1}^{n}(x_i-\overline{x})^2$.

Como $\widehat{\beta}_1=\sum_{i=1}^{n}\frac{(x_i-\overline{x})Y_i}{\sum_{i=1}^{n}(x_i-\overline{x})^2}$ se tiene que

$$V[\widehat{\beta}_1]=\sum_{i=1}^{n}\frac{(x_i-\overline{x})^2}{\left(\sum_{i=1}^{n}(x_i-\overline{x})^2\right)^2}V[Y_i]$$

y al sustituir $V[Y_i]=V[e_i]=\sigma^2$, se obtiene

$$\begin{aligned}V[\widehat{\beta}_1] &= \sigma^2\sum_{i=1}^{n}\frac{(x_i-\overline{x})^2}{\left(\sum_{i=1}^{n}(x_i-\overline{x})^2\right)^2}\\ &= \sigma^2\frac{1}{\sum_{i=1}^{n}(x_i-\overline{x})^2}.\end{aligned}$$

Las expresiones para $V[\widehat{\beta}_0]$ y $Cov(\widehat{\beta}_0,\widehat{\beta}_1)$ pueden obtenerse como caso particular del planteamiento matricial de la regresión que se hará más adelante, Expresión (7.4.2.5).

■

Como se desprende de (7.4.1.7), para el cálculo de la varianza de los estimadores $\widehat{\beta}_0$ y $\widehat{\beta}_1$ se necesita conocer σ^2. En la Sección 7.4.2. se demostrará

que lo que se ha notado por $SC(\widehat{\beta}_0, \widehat{\beta}_1)$ en la Expresión (7.4.1.4) si se divide por $n-2$, es un estimador insesgado de σ^2, es decir, que notando por

$$SC_R = SC(\widehat{\beta}_0, \widehat{\beta}_1) = \sum_{i=1}^{n} (y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i)^2$$

expresión a la que se le dá el nombre de *suma de cuadrados residual*, se sigue que $E\left[\frac{SC_R}{n-2}\right] = \sigma^2$.

En la práctica, lo que se hace es utilizar este estimador para estimar las varianzas de $\widehat{\beta}_0$ y de $\widehat{\beta}_1$. Además, en el Teorema no se utiliza la hipótesis de normalidad, bajo la suposición de que los errores $e_i \sim N(0, \sigma)$ se obtiene la distribución normal de ambos estadísticos, ya que son combinaciones lineales de v.a. normales. En cualquier caso, esta suposición de normalidad no es demasiado fuerte si el tamaño muestral es grande en virtud del teorema central del límite.

A los valores $(\widehat{V}[\widehat{\beta}_i])^{1/2}$ se les da el nombre de *errores estandar* de los estimadores, donde se sustituye en la expresión de la varianza, σ^2 por $SC_R/(n-2)$ y se sigue que

$$\frac{\widehat{\beta}_i - \beta_i}{\sqrt{\widehat{V}[\widehat{\beta}_i]}} \sim T_{n-2} \tag{7.4.1.8}$$

tiene distribución de Student con $n-2$ grados de libertad. Esta última expresión permite realizar constrastes de hipótesis y obtener regiones de confianza para los parámetros.

Ejemplo 7.4.1.2
Si se aplican los resultados anteriores a los datos obtenidos por el agricultor en el Ejemplo 7.4.1.1, se obtienen los valores siguientes

Coeficientes	Estimadores	Errores estándar	Estadísticos T
β_0	$-5'7028$	$5'0317$	$-1'1334$
β_1	$1'5045$	$0'0873$	$12'0805$

y $SC_R = 51'8698$. Por lo tanto $\widehat{\sigma}^2 = 51'8698/8 = 6'4837$ de donde

$$\widehat{V}(\widehat{\beta}_0) = \tfrac{6'4837 \cdot 3322'7}{850'9} = 25'3184 \quad \sqrt{\widehat{V}(\widehat{\beta}_0)} = 5'0317$$
$$\widehat{V}(\widehat{\beta}_1) = \tfrac{6'4837}{850'9} = 0'00762 \quad \sqrt{\widehat{V}(\widehat{\beta}_1)} = 0'0873$$

Para contrastar $H_0 : \beta_1 = 0$ frente a $H_1 : \beta_1 \neq 0$ se emplea que

$$\frac{\widehat{\beta}_1 - \beta_1}{\sqrt{\widehat{V}(\widehat{\beta}_1)}} \sim T_8$$

y por tanto

$$\begin{aligned} RC &= \{|\widehat{\beta}_1| \geq \sqrt{\widehat{V}(\widehat{\beta}_1)}t_{8;0'025} = 0'0873 \cdot 2'306\} \\ &= \{|\widehat{\beta}_1| \geq 0'2013\} \end{aligned}$$

como $1'0545$ pertenece a RC, se rechaza H_0.

Un intervalo de grado de confianza $0'95$ para β_1 se obtiene mediante

$$IC_{0'95}(\beta_1) = \left(\widehat{\beta}_1 - \sqrt{\widehat{V}(\widehat{\beta}_1)}t_{n-2;\alpha/2}, \widehat{\beta}_1 + \sqrt{\widehat{V}(\widehat{\beta}_1)}t_{n-2;\alpha/2}\right)$$

que se convierte en

$$\begin{aligned} IC_{0'95}(\beta_1) &= (1'0545 - 0'0873, 1'0545 + 0'0873 \cdot 2'306) \\ &= (0'8532, 1'2558) \end{aligned}$$

■

Para evitar aplicaciones incorrectas del modelo de regresión simple, debe comprobarse siempre, que se cumplen todas las hipótesis supuestas para la elaboración del modelo. Inicialmente es conveniente representar los puntos (x_i, y_i) para ver si tiene sentido un ajuste lineal. Una buena herramienta que se puede utilizar para la validación del modelo son los residuos $\widehat{e}_i = y_i - \widehat{y}_i = y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i$, la representación de $(x_i, \widehat{e}_i)$ debe ser del tipo de la Figura 7.4.1.3. Por ejemplo, una representación como la de la Figura 7.4.1.4 desaconsejaría la suposición de que la varianza σ^2 no depende de x_i (homocedasticidad)

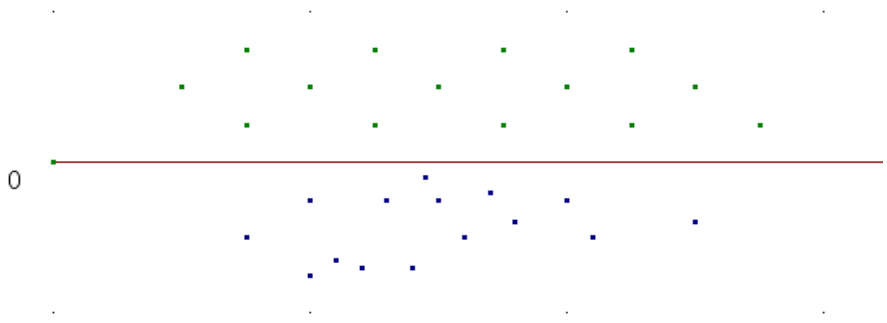


Figura 7.4.1.3. Representación compatible con las suposiciones del modelo de regresión simple de los puntos $(x_i, \widehat{e}_i)$.

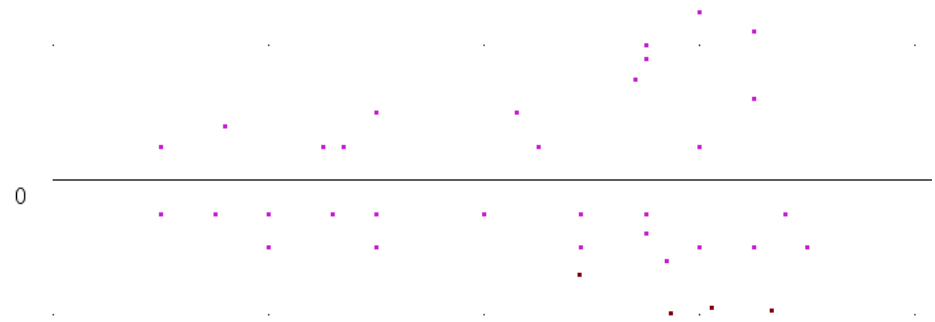


Figura 7.4.1.4. Representación correspondiente a falta de homocedasticidad de los puntos $(x_i, \widehat{e}_i)$ en un modelo de regresión simple.

Con vistas a la mayor o menor exactitud de los contrastes de hipótesis y las regiones de confianza (la determinación de las constantes que intervienen en ellos), es interesante verificar también la hipótesis de normalidad para los puntos $(x_i, \widehat{e}_i)$, puede verse en la Sección 8.4.3.

7.4.2. Regresión lineal múltiple: aproximación matricial

Al trabajar con problemas reales, es necesario recurrir a modelos más complejos, aunque todavía lineales, en los que una variable dependiente se explica por dos o más variables independientes. En esta Sección se va a recurrir a ajustar a los datos

$$y_i, x_{i1}, x_{i2}, ..., x_{i,p-1}, \qquad i = 1, ..., n$$

un modelo de la forma

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_{p-1} x_{p-1}$$

con $p \leq n$, es decir, un hiperplano.

En forma matricial se puede escribir como

$$Y = X\beta + \varepsilon$$
$$n \times 1 = n \times p \cdot p \times 1 + n \times 1$$

donde los vectores se suponen vectores columnas y la matriz X viene dada por

$$\begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1,p-1} \\ 1 & x_{21} & x_{22} & \cdots & x_{2,p-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{n,p-1} \end{pmatrix}$$

El criterio que se emplea para la estimación de los parámetros $\beta_0, ..., \beta_{p-1}$ es el de *mínimos cuadrados*, como en el caso de la regresión lineal simple, y consiste en minimizar la siguiente expresión

(7.4.2.1)

$$\begin{aligned} SC(\beta_0, \beta_1, ..., \beta_{p-1}) &= \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1\, x_{i,1} - ... - \beta_{p-1} x_{i,p-1})^2 \\ &= \|Y - X\beta\|^2 = \left\| Y - \widehat{Y} \right\|^2 \end{aligned}$$

donde $\|Z\|^2 = \sum_{i=1}^{n} Z_i^2$ es el cuadrado de la norma del vector Z.

Derivando respecto a $\beta_j, j = 0, ..., p-1$ la expresión $SC(\beta_0, \beta_1, ..., \beta_{p-1})$ e igualando a cero, se obtienen las *ecuaciones normales*

$$\left.\begin{array}{c} n\beta_0 + \sum_{i=1}^{n} x_{i1}\beta_1 + ... + \sum_{i=1}^{n} x_{i,p-1}\beta_{p-1} = \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{i1}\beta_0 + \sum_{i=1}^{n} x_{i1}^2\beta_1 + ... + \sum_{i=1}^{n} x_{i,p-1}x_{i1}\beta_{p-1} = \sum_{i=1}^{n} x_{i1}y_i \\ \vdots \\ \sum_{i=1}^{n} x_{i,p-1}\beta_0 + \sum_{i=1}^{n} x_{i1}x_{i,p-1}\beta_1 + ... + \sum_{i=1}^{n} x_{i,p-1}^2\beta_{p-1} = \sum_{i=1}^{n} x_{i,p-1}y_i \end{array}\right\}$$

que matricialmente pueden escribirse como

$$X'X\widehat{\beta} = X'Y \tag{7.4.2.2}$$

Si es $|X'X| \neq 0$ se pueden resolver en $\widehat{\beta}$, obteniéndose

$$\widehat{\beta} = (X'X)^{-1}X'Y \tag{7.4.2.3}$$

desde el punto de vista computacional, la inversión de la matriz $X'X$ puede dar lugar a errores, por lo que se advierte de la existencia de métodos alternativos para la resolución de la Ecuación (7.4.2.2), (véase Strang (1980)).

En los dos ejemplos siguientes, se expresan matricialmente resultados ya estudiados.

Ejemplo 7.4.2.1

En el caso de la regresión lineal simple $y = \beta_0 + \beta_1 x$, con (x_i, y_i) $i = 1, ..., n$, se comprueba que

$$X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad X'X = \begin{pmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{pmatrix}, \quad X'Y = \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{pmatrix},$$

$$(X'X)^{-1} = \frac{1}{n\sum_{i=1}^{n} x_i^2 - n^2\overline{x}^2} \begin{pmatrix} \sum_{i=1}^{n} x_i^2 & -\sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & n \end{pmatrix}$$

$$\widehat{\beta} = (X'X)^{-1}X'Y = \begin{pmatrix} \dfrac{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i - \sum_{i=1}^n x_i y_i \sum_{i=1}^n x_i}{n\sum_{i=1}^n x_i^2 - n^2\overline{x}^2} \\ \dfrac{n\sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n\sum_{i=1}^n x_i^2 - n^2\overline{x}^2} \end{pmatrix} = \begin{pmatrix} \dfrac{\frac{1}{n}\sum_{i=1}^n x_i^2\overline{y} - \frac{1}{n}\sum_{i=1}^n x_i y_i\overline{x}}{\frac{1}{n}\sum_{i=1}^n x_i^2 - \overline{x}^2} \\ \dfrac{\frac{1}{n}\sum_{i=1}^n x_i y_i - \overline{xy}}{\frac{1}{n}\sum_{i=1}^n x_i^2 - \overline{x}^2} \end{pmatrix}$$

$$= \begin{pmatrix} \dfrac{\frac{1}{n}\sum_{i=1}^n (x_i^2 - \overline{x}^2 + \overline{x}^2)\overline{y} - \frac{1}{n}\sum_{i=1}^n x_i y_i\overline{x}}{b_x^2} \\ \dfrac{\frac{1}{n}\sum_{i=1}^n x_i y_i - \overline{xy}}{b_x^2} \end{pmatrix} = \begin{pmatrix} \dfrac{b_x^2\overline{y} + \overline{x}^2\overline{y} - \frac{1}{n}\sum_{i=1}^n x_i y_i\overline{x}}{b_x^2} \\ \dfrac{b_{xy}}{b_x^2} \end{pmatrix}$$

y el primer elemento de este vector es

$$\overline{y} + \frac{\overline{x}^2\overline{y} - \frac{1}{n}\sum_{i=1}^n x_i y_i\overline{x}}{b_x^2} = \overline{y} - \overline{x}\frac{b_{xy}}{b_x^2}$$

que coincide con la Expresión (7.4.1.3).

Ejemplo 7.4.2.2

La Expresión $\sum_{i=1}^n (x_i - \overline{x})^2$ puede escribirse matricialmente como

$$\sum_{i=1}^n (x_i - \overline{x})^2 = X'(I - \frac{1}{n}M)X$$

donde M es la matriz $n \times n$ que tiene todos sus elementos iguales a la unidad.

Basta comprobar que

$$\begin{aligned} \sum_{i=1}^n (x_i - \overline{x})^2 &= \sum_{i=1}^n x_i^2 - n\overline{x}^2 = X'X - \frac{1}{n}(x_1 + ... + x_n)^2 = \\ &= X'X - \frac{1}{n}\left(\sum_{i=1}^n x_i^2 + 2\sum_{i=1}^n \sum_{j>i} x_i x_j\right) \\ &= X'X - \frac{1}{n}(X'X + X'\begin{pmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 1 & 1 & \cdots & 0 \end{pmatrix} X) \\ &= X'X - \frac{1}{n}(X'MX) \end{aligned}$$

La matriz M y el número n pueden expresarse como $M = \mathbf{1}\mathbf{1}'$ y $n = \mathbf{1}'\mathbf{1}$, siendo $\mathbf{1}$ el vector columna n dimensional con todas sus componentes iguales a uno.

Otras expresiones matriciales útiles, están contenidas en los dos resultados siguientes.

Teorema 7.4.2.1
Si X es un vector columna aleatorio de dimensión n, de media μ y de matriz de covarianzas Σ_X, entonces la esperanza de la forma cuadrática $X'AX$, para cualquier matriz $A \;\; n \times n$, viene dada por

$$E(X'AX) = traza(A\Sigma_X) + \mu' A\mu \tag{7.4.2.4}$$

Demostración: la traza de una matriz es la suma de los elementos de su diagonal principal. El resultado se sigue de considerar que $X'AX = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j$ y por tanto

$$\begin{aligned} E(X'AX) &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}E(X_iX_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}(\sigma_{ij} + \mu_i\mu_j) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\sigma_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\mu_i\mu_j \\ &= traza(A\Sigma_X) + \mu' A\mu. \end{aligned}$$

■

Teorema 7.4.2.2
Si X es un vector columna aleatorio de dimensión n con matriz de covarianzas Σ_X y se construyen

$$\underset{p\times 1}{Y} = \underset{p\times n}{A}\underset{n\times 1}{X} \qquad \underset{m\times 1}{Z} = \underset{m\times n}{B}\underset{n\times 1}{X}$$

con A y B matrices de constantes, entonces

$$\Sigma_{YZ} = A\Sigma_X B'.$$

Demostración:

$$\begin{aligned} Cov(Y_i, Z_k) &= Cov(\sum_{j=1}^{n} a_{ij}X_j, \sum_{l=1}^{n} b_{kl}X_l) \\ &= \sum_{j=1}^{n} a_{ij}\sum_{l=1}^{n} b_{kl}Cov(X_j, X_l) \\ &= \sum_{j=1}^{n}\sum_{l=1}^{n} a_{ij}\sigma_{jl}b_{kl} \end{aligned}$$

que constituye el elemento i,k de la matriz $A\Sigma_X B'$.
■

Como consecuencia del Ejemplo 7.4.2.2 y del Teorema 7.4.2.1, se puede obtener otra demostración de que si $X_i \sim N(\mu, \sigma)$ $i = 1, ..., n$ e independientes, entonces

$$E[(X_i - \overline{X})^2] = (n-1)\sigma^2.$$

En efecto: $\sum_{i=1}^{n}(X_i - \overline{X})^2 = X'AX$ con $A = I - \frac{1}{n}M$ y M una matriz cuyos elementos son todos 1, luego

$$E\left[\sum_{i=1}^{n}(X_i - \overline{X})^2\right] = traza(A\Sigma_X) + \mu' A\mu$$

pero $\Sigma_X = \sigma^2 I$, por ser independientes, la matriz

$$(I - \frac{1}{n}M) = \begin{pmatrix} (1-\frac{1}{n}) & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & (1-\frac{1}{n}) & \cdots & -\frac{1}{n} \\ \vdots & \vdots & & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \cdots & (1-\frac{1}{n}) \end{pmatrix}$$

y haciendo $\boldsymbol{\mu} = \mu \mathbf{1}_\sigma$, se obtiene

$$\begin{aligned} \boldsymbol{\mu}'(I - \frac{1}{n}\mathbf{1}\mathbf{1}')\boldsymbol{\mu} &= \mu^2(\mathbf{1}'\mathbf{1} - \frac{1}{n}\mathbf{1}'\mathbf{1}\mathbf{1}'\mathbf{1}) \\ &= \mu^2(n - \frac{1}{n}nn) \\ &= 0 \end{aligned}$$

luego

$$\begin{aligned} E\left[\sum_{i=1}^{n}(X_i - \overline{X})^2\right] &= traza\left[\left(I - \frac{1}{n}M\right)\sigma^2 I\right] \\ &= (n-1)\sigma^2. \end{aligned}$$

■

Se está ahora en condiciones de obtener la media y la matriz de covarianzas de los EMC en el caso de la regresión múltiple.

Teorema 7.4.2.3
Para el modelo

$$Y_i = \beta_0 + \sum_{j=1}^{p-1}\beta_j x_{ij} + e_i, \qquad i = 1, ..., n$$

donde $E[e_i] = 0$ y $Cov(e_i, e_j) = \delta_{ij}\sigma^2$, que en notación matricial puede ponerse como

$$\begin{array}{ccccc} Y & = & X & \beta & +e \\ n \times 1 & & n \times p & p \times 1 & n \times 1 \end{array}$$

con $E[e] = 0 \quad \Sigma_e = \sigma^2 I$, el vector de estimadores de mínimos cuadrados (EMC) de los valores β_i es

$$\widehat{\beta} = (X'X)^{-1}X'Y$$

y cumple que

1) $E[\widehat{\beta}] = \beta$ 2) $\Sigma_{\widehat{\beta}} = \sigma^2(X'X)^{-1}$.

Demostración: 1) $E[\widehat{\beta}] = (X'X)^{-1}X'E[Y] = (X'X)^{-1}X'X\beta$.

2) Como $\widehat{\beta} = AY$ con $A = (X'X)^{-1}X'$, por el Teorema 7.4.2.2 con $B = A$ se tiene que

$$\begin{aligned}\Sigma_{\widehat{\beta}} &= A\Sigma_Y A' = A\sigma^2 I A' = (X'X)^{-1}X'X(X'X)^{-1}\sigma^2 \\ &= \sigma^2(X'X)^{-1}.\end{aligned}$$

Ejemplo 7.4.2.3

Volviendo al caso de la regresión lineal simple, se puede utilizar el Teorema 7.4.2.3 para demostrar que

$$V(\widehat{\beta}_0) = \frac{\sigma^2 \frac{1}{n}\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2} \qquad Cov(\widehat{\beta}_0, \widehat{\beta}_1) = -\frac{\sigma^2\overline{x}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}$$

Del Ejemplo 7.4.2.1 se tiene que

$$(X'X)^{-1} = \frac{1}{n\sum_{i=1}^{n} x_i^2 - n^2\overline{x}^2}\begin{pmatrix} \sum_{i=1}^{n} x_i^2 & -\sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & n \end{pmatrix}$$

por lo que sustituyendo

$$V[\widehat{\beta}_0] = \sigma^2\frac{\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n} x_i^2 - n^2\overline{x}^2} = \frac{\sigma^2\frac{1}{n}\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}$$

y

$$Cov(\widehat{\beta}_0, \widehat{\beta}_1) = -\frac{\sigma^2\sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} x_i^2 - n^2\overline{x}^2} = \frac{-\sigma^2\overline{x}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}.$$

■

Con el fin de obtener un estimador insesgado de la varianza σ^2 en el modelo de regresión lineal múltiple, se introduce el siguiente resultado.

Lema 7.4.2.1
La matriz de dimensión $n \times p$ $\quad P = X(X'X)^{-1}X'$ cumple que

$$P = P' = P^2$$
$$I - P = (I - P') = (I - P)^2$$

Demostración: $P' = X[(X'X)^{-1}]'X' = X[(X'X)']^{-1}X'$ ya que por ser $X'(X^{-1})' = (X^{-1}X)' = I' = I$ es $(X^{-1})' = (X')^{-1}$ por lo que $P' = X(X'X)^{-1}X' = P$.

Además $P^2 = X(X'X)^{-1}X'X(X'X)^{-1}X' = X(X'X)^{-1}X' = P$.

De manera análoga se procede con $I - P$.

■

Utilizando la matriz P, se puede calcular la suma de los cuadrados de los residuos y demostrar que

$$E\left[\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2\right] = \sigma^2(n-p). \tag{7.4.2.5}$$

En efecto,

$$\begin{aligned}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2 &= \left\|Y - \widehat{Y}\right\|^2 = \left\|Y - X\widehat{\beta}\right\|^2 = \left\|Y - X(X'X)^{-1}X'Y\right\|^2 \\ &= \|Y - PY\|^2 = \|(I-P)Y\|^2 = Y'(I-P)'(I-P)Y \\ &= Y'(I-P)Y\end{aligned}$$

y, mediante la aplicación del Teorema 7.4.2.1,

$$\begin{aligned}E\left[\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2\right] &= E\left[Y'(I-P)Y\right] \\ &= traza\left[(I-P)\Sigma_Y\right] + (E[Y])'(I-P)E[Y]\end{aligned}$$

y por ser $E[Y] = X\beta$ y $\Sigma_Y = \sigma^2 I$

$$\begin{aligned}E\left[\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2\right] &= traza\left[\sigma^2(I-P)I\right] + \beta'X'(I-P)X\beta \\ &= \sigma^2 traza(I-P) + \beta'X'(X\beta - PX\beta) \\ &= \sigma^2(n - traza\ P)\end{aligned}$$

ya que $\beta'X'(X\beta - X(X'X)^{-1}X'X\beta) = \beta'X'(X\beta - X\beta) = 0$.

Por último, dado que $traza(AB) = traza(BA)$ se tiene que $traza\ P = traza[X'X(X'X)^{-1}] = n - traza\ I$, donde I tiene dimensión $p \times p$, con lo que se obtiene la Expresión (7.4.2.5). La suma de los cuadrados de los residuos dividida por $n-p$ es un estimador insesgado de σ^2. $E[SC_R/(n-p)] = \sigma^2$.

■

Una propiedad de óptimo que tienen los EMC es el contenido del siguiente resultado.

Teorema 7.4.2.4 (Gauss-Markov)
Los estimadores $\widehat{\beta}_i$ de $\widehat{\beta} = (X'X)^{-1}X'Y$ tienen varianza mínima, en la clase de los estimadores centrados que son funciones lineales de las observaciones $Y_i \quad i = 1, ..., n$.

Demostración: Se tiene que $\widehat{\beta} = BY$ con $B = (X'X)^{-1}X'$. Si $T(Y) = QY$ con Q de dimensión $p \times n$, es un estimador cualquiera que sea función lineal de las observaciones, y centrado para β, ha de cumplir

$$E[T(Y)] = QE[Y] = QX\beta = \beta$$

con lo que ha de ser $QX = I$, con I de dimensión $p \times p$.

Por el Teorema 7.4.2.2

$$\Sigma_T = Q\Sigma_Y Q' = \sigma^2 QQ'$$

tomando $A = Q - B$ se tiene que

$$I = QX = (A + B)X = AX + (X'X)^{-1}X'X = AX + I$$

luego $AX = \mathbf{O}$ (matriz de ceros de dimensión $p \times p$).

Por otro lado, dado que $Q = A + B$,

$$\Sigma_T = \sigma^2(A + B)(A + B)' = \sigma^2(AA' + AB' + BA' + BB')$$

pero

$$\begin{aligned} BB' &= (X'X)^{-1}X'X(X'X)^{-1} = (X'X)^{-1} \\ AB' &= AX(X'X)^{-1} = \mathbf{O} \\ BA' &= (X'X)^{-1}X'A' = (X'X)^{-1}(AX)'\mathbf{O} \end{aligned}$$

luego

$$\Sigma_T = \sigma^2(AA' + (X'X)^{-1}).$$

Los elementos de la diagonal principal de AA' son la suma de los cuadrados de las filas de la matriz A y son siempre no negativos; se tiene que la varianza de los elementos del vector $T(Y)$ cumplirá

$$V[T_i] \geq V[\widehat{\beta}_i] \quad i = 1, ..., p.$$

■

Una vez que se sabe estimar σ^2 mediante la Expresión (7.4.2.5), para hacer inferencias respecto de β es preciso hacer alguna hipótesis acerca de la distribución de los errores e_i. Si se supone que $e_i \sim N(0, \sigma) \quad i = 1, ..., n$ e independientes, como $\widehat{\beta} = (X'X)^{-1}X'Y$, se tiene que $Y \sim N(X\beta, \sigma^2 I)$, de donde $\widehat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$. Por lo tanto $\widehat{\beta}_i \sim N(\beta_i, \sigma\sqrt{c_{ii}})$, siendo

$C = (X'X)^{-1}$. Al ser σ desconocida se estima por $s = \sqrt{\frac{SC_R}{n-p}}$ y se puede demostrar que $SC_R \sim \chi^2_{n-p}$, con lo que

$$\frac{\widehat{\beta}_i - \beta_i}{s\sqrt{c_{ii}}} \sim T_{n-p} \tag{7.4.2.6}$$

Esto permite realizar contrastes respecto a valores de β_i, del tipo $H_0 : \beta_i = \beta_{i0}$ conocido.

Ejemplo 7.4.2.4
En un estudio sobre el hacinamiento en una gran ciudad, se estudian 10 manzanas de casas escogidas al azar y se obtienen los siguientes datos

Manzana	1	2	3	4	5	6	7	8	9	10
Habitaciones	60	52	58	56	62	51	72	48	71	58
Personas	115	80	82	93	105	109	130	93	109	95

se pretende estimar el número de personas y en función del número de habitaciones x y de su cuadrado, mediante una función del tipo

$$y = \beta_0 + \beta_1 x + \beta_2 x^2$$

obteniéndose $\widehat{\beta}_0 = 263'733$, $\widehat{\beta}_1 = -6'685$ y $\widehat{\beta}_2 = 0'065$.

Para realizar el contraste $H_0 : \beta_2 = 0$ se emplea la Expresión (7.4.2.6) con lo que $\widehat{\beta}_2 / s\sqrt{c_{33}} \sim T_{n-p}$ con $n - p = 8$ grados de libertad.

Si con el fin de realizar los cálculos se obtienen las filas

x_i^2 :	3600	2704	3364	3136	3844	2601	5184	2304
	5041	3364						
$\widehat{y}_i$:	96'633	91'873	94'663	93'213	99'123	91'363	119'373	92'613
	116'763	94'663						

lo que permite calcular

$$s^2 = \frac{SC_R}{n-p} = \frac{\Sigma(y_i - \widehat{y}_i)^2}{n-p} = \frac{1140'387}{8} = 142'548.$$

Se puede verificar que

$$X'X = \begin{pmatrix} 10 & 588 & 1011 \\ 588 & 35142 & 60166 \\ 1011 & 60166 & 104359 \end{pmatrix}$$

con lo que

$$C = (X'X)^{-1} = \begin{pmatrix} 6'76376 & -0'076321 & -0'021524 \\ -0'076321 & 0'0030612 & -0'0010255 \\ -0'021524 & -0'0010255 & 0'00080932 \end{pmatrix}$$

de donde resulta que $c_{33} = 0'00080932$.

La región crítica viene dada por

$$RC = \left\{ \frac{\left|\widehat{\beta}_2 - \beta_2\right|}{s\sqrt{c_{33}}} \geq k \right\}$$

Para la determinación de k, se tiene en cuenta que

$$P\{RC|\beta_2 = 0\} = P\{|T_8| \geq k\} = \alpha$$

si se escoge $\alpha = 0'05$ se obtiene $k = 2'306$. Como $\widehat{\beta}_2/s\sqrt{c_{33}} = 0'065/11'939 \cdot 0'028 = 0'194$, se obtiene, que los datos son compatibles con $\beta_2 = 0$.

■

Dado un modelo de regresión múltiple, también es interesante hacer predicciones para valores conocidos de la variable predictora $x = x_0$. Como el modelo supuesto es de la forma

$$y(x_0) = \beta_0 + \beta_1 x_{01} + ... + \beta_{p-1} x_{0,p-1} + e \tag{7.4.2.7}$$

el estimador obvio es $\widehat{y}(x_0) = x_0'\widehat{\beta}$, que es centrado y que tendrá como varianza

$$\begin{aligned} V[\widehat{y}(x_0)] &= x_0' \Sigma_{\widehat{\beta}} x_o \\ &= \sigma^2 x_0'(X'X)^{-1} x_0 \end{aligned}$$

que, a su vez puede ser estimada sustituyendo s^2 por σ^2, mediante la Expresión (7.4.2.5), con lo que la suposición de normalidad sobre los errores, o el teorema central del límite, permiten determinar un intervalo de grado de confianza α para $y(x_0)$, mediante

$$(x_0'\widehat{\beta} - t_{n-p;\alpha/2}\sqrt{\widehat{V}[\widehat{y}(x_0)]}, x_0'\widehat{\beta} + t_{n-p;\alpha/2}\sqrt{\widehat{V}[\widehat{y}(x_0)]}) \tag{7.4.2.8}$$

Ejemplo 7.4.2.5

Para los datos sobre el hacinamiento en una gran ciudad del Ejemplo 7.4.2.4 y supuesto el modelo

$$y = \beta_0 + \beta_1 x + \beta_2 x^2$$

el intervalo de grado de confianza $0'95$, para el número de personas que corresponderían a 65 habitaciones, viene dado, de acuerdo con la expresión (7.4.2.8), por

$$x_0'\widehat{\beta} \pm t_{n-p;\alpha/2}\sqrt{\widehat{V}[\widehat{y}(x_0)]}$$

con $x_0' = (1, 65, 4225)$ $\quad x_0'\widehat{\beta} = 103'833 \quad t_{8;0'025} = 2'306$

$$\begin{aligned} \widehat{V}[\widehat{y}(x_0)] &= s^2 x_0'(x'x)^{-1} x_0 = 142'54 \cdot 13711'476 \\ &= 1954433'731 \end{aligned}$$

con lo que se obtiene

$$IC_{0'95}(\widehat{y}(65)) = (-3119'980, 3327'646).$$

■

Como se ha visto en el Ejemplo 7.4.2.4, existe la posibilidad de variar el número de variables explicativas a incluir en el ajuste. En el caso lineal, el mayor o menor ajuste venía caracterizado por lo próximo al valor uno del cuadrado del coeficiente de correlación lineal r^2, siendo

$$r = \frac{b_{xy}}{b_x b_y}$$

(véanse las Expresiones (7.4.1.4) y (7.4.1.5)). En el caso de la regresión múltiple, esta última expresión no es generalizable pero sí lo es la (7.4.1.4), por lo que a partir de ésta se define el *coeficiente de determinación* o *cuadrado del coeficiente de correlación múltiple* mediante el valor R^2 tal que

$$\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2 = nb_y^2(1 - R^2)$$

esto es

$$R^2 = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{b_y^2}$$

Para su cálculo se emplea

$$\begin{aligned} R^2 &= \frac{\frac{1}{n}\sum_{i=1}^{n}(y_i-\overline{y})^2 - \frac{1}{n}\sum_{i=1}^{n}(y_i-\widehat{y}_i)^2}{\frac{1}{n}\sum_{i=1}^{n}(y_i-\overline{y})^2} \\ &= \frac{b_y - \frac{1}{n}SC_R}{b_y} \end{aligned} \tag{7.4.2.9}$$

y se verifica que $0 \leq R^2 \leq 1$. Valores próximos a uno significan un mejor ajuste o lo que es lo mismo hay un mayor porcentaje de b_y, explicado por el mejor ajuste en el sentido de los mínimos cuadrados.

Ejemplo 7.4.2.6
Se pretende determinar la relación lineal existente, entre la producción en un cierto área y el esfuerzo en trabajo y capital. Se dispone de los datos de un periodo de tiempo contenidos en la Tabla 7.4.2.1. Las colecciones de puntos relacionando (Trabajo, Producción) y (Capital, Producción) se muestran en la Figura 7.4.2.1, y ponen de manifiesto la conveniencia de un ajuste lineal.

Año	Trabajo	Capital	Producción
1990	7'60	7'45	8'90
1991	7'66	7'26	8'32
1992	7'18	6'30	7'50
1993	7'30	6'55	8'00
1994	7'20	6'74	7'70
1995	7'12	6'50	7'62
1996	7'25	7'39	8'23
1997	7'45	7'42	8'50
1999	7'38	7'07	8'20
2000	7'45	7'12	8'72

Tabla 7.4.2.1. Datos correspondientes al Ejemplo 7.4.2.6.

La colección de datos es muy pequeña por lo que las conclusiones sólo son válidas a nivel pedagógico.

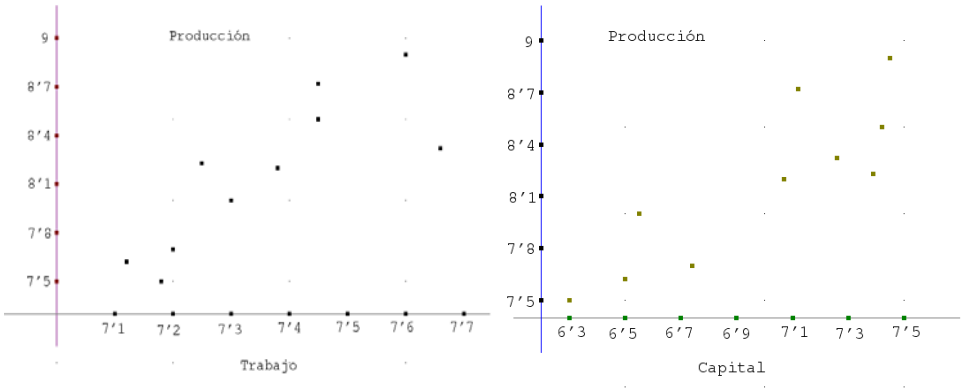


Figura 7.4.2.1. Colecciones de puntos correspondientes a (trabajo, producción) y (capital, producción) del Ejemplo 7.4.2.6.

Ejemplo 7.4.2.7
Si se considera primero, la predicción de la producción por el trabajo y la predicción de la producción por el capital, se obtiene

	Trabajo	Capital
$\widehat{\beta}_0$	-7'55	1'55
$\widehat{\beta}_1$	2'137	0'948
r	0'827	0'860
SC_R	0'623	0'512
R^2	0'684	0'740

Se consigue un mejor ajuste lineal de la producción en función del capital, pues con él se explica el 74 % de la varianza $b_y^2 = 0'197$, que en función del

trabajo, que explica sólo el $68'4\,\%$ de la misma. Si se recurre a explicar la producción en función del trabajo y el capital conjuntamente, se obtiene

$$producción = -4'128 + 1'096\ trabajo + 0'606\ capital$$

el coeficiente de determinación R^2 calculado mediante la Expresión (7.4.2.9) es ahora de $0'774$, ligeramente superior, al obtenido con cada una de las variables por separado, ya que con las dos variables se explica el $77'4\,\%$ de la varianza muestral.

7.5. APROXIMACIÓN BAYESIANA DEL ANÁLISIS DE LA VARIANZA

Para llevar a cabo el análisis bayesiano del modelo lineal estudiado en el Teorema 7.4.2.3, de la forma

$$\begin{array}{ccccccc} Y & = & X & \beta & + & e \\ n\times 1 & & n\times p & p\times 1 & & n\times 1 \end{array} \tag{7.5.1}$$

hay que definir una distribución de probabilidad inicial $\pi(\beta,\sigma^2)$ sobre los parámetros $(\beta_1,...,\beta_p,\sigma^2)$ y calcular la distribución de probabilidad final mediante la densidad

$$\pi(\beta,\sigma^2|y) = \frac{\pi(\beta,\sigma^2)f(y|\beta,\sigma^2)}{\int_\beta\int_{\sigma^2}\pi(\beta,\sigma^2)f(y|\beta,\sigma^2)d\sigma^2 d\beta} \tag{7.5.2}$$

Para obtener expresiones análogas a las del Teorema 7.4.2.3 citado, se va a utilizar como distribución inicial

$$\pi(\beta,\sigma^2) = \frac{1}{\sigma^2} \quad \text{con} \quad \sigma^2 > 0 \tag{7.5.3}$$

y bajo la suposición de que los errores tienen $e \sim N(0,\sigma^2 I)$ distribución normal será

$$f(y|\beta,\sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\sum_{i=1}^{p}x_{ij}\beta_j)^2} \tag{7.5.4}$$

Se hace notar que la Expresión (7.5.3) no es una función de densidad, ya que la integral sobre todo el espacio paramétrico es infinita, sin embargo la utilización de esta función como densidad inicial, impropia por tanto, permite obtener los resultados clásicos en el contexto bayesiano, como se verá a continuación.

Si se denota, en uniformidad con la Expresión (7.4.2.1), la suma de cuadrados por

$$SC(\beta) = \sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2 = (y - X\beta)'(y - X\beta) \tag{7.5.5}$$
$$= y'y + 2\beta' X'y + \beta' X'X\beta$$

y se hace

$$X'X\widehat{\beta} = X'Y$$

se tiene

$$\widehat{\beta} = (X'X)^{-1}X'y \qquad \beta' X'X\widehat{\beta} = \beta' X'y$$

Entonces (7.5.5) toma la forma

$$SC(\overrightarrow{\beta}) = (\beta - \widehat{\beta})'X'X(\beta - \widehat{\beta}) + y'y - \widehat{\theta}' X'X\widehat{\beta}$$

Si se emplea la notación

$$SC_T(\overrightarrow{\beta}) = (\beta - \widehat{\beta})'X'X(\beta - \widehat{\beta})$$
$$SC_R(\overrightarrow{\beta}) = y'y - \widehat{\beta}' X'X\widehat{\beta}$$

la función de verosimilitud en (7.5.4) puede escribirse como

$$f(\overrightarrow{y}|\beta, \sigma^2) = (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}\{SC_T(\beta) + SC_R(\beta)\}} \tag{7.5.6}$$

La expresión $SC_R(\beta)$ es la suma de cuadrados de los errores, ya que

$$(y - \widehat{y})'(y - \widehat{y}) = y'y - \widehat{\beta}' X'X\widehat{\beta}$$

sin más que operar y utilizar el Lema 7.4.2.1, con lo que

$$(I - P)'(I - P) = I - P,$$

y

$$X'Y = X'X\widehat{\beta}.$$

Recordando que el estimador de σ^2 era $SC_R/(n-p)$ (véase la Expresión (7.4.2.5)) se tiene que la función de verosimilitud se expresa como

$$f(\overrightarrow{y}\,|\,\overrightarrow{\beta}, \sigma^2) = (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}\{SC_T(\beta) + (n-p)s^2\}} \tag{7.5.7}$$

donde

$$s^2 = \frac{1}{n-p} SC_R = \frac{1}{n-p}(y - \widehat{y})'(y - \widehat{y})$$
$$= \frac{1}{n-p} \sum_{i=1}^{n} (y_i - \widehat{y}_i).$$

El cálculo de la distribución final de los parámetros se reduce a

(7.5.8)
$$\pi(\overrightarrow{\beta},\sigma^2|\overrightarrow{y}) = \frac{\sigma^{-2}(2\pi\sigma^2)^{-n/2}e^{-\frac{1}{2\sigma^2}\{SC_T(\beta)+(n-p)s^2\}}}{\int_{\overrightarrow{\beta}}\int_0^{+\infty}\sigma^{-2}(2\pi\sigma^2)^{-n/2}e^{-\frac{1}{2\sigma^2}\{SC_T(\beta)+(n-p)s^2\}}d\sigma^2 d\overrightarrow{\beta}}$$

Para hacer inferencias acerca de los parámetros β o σ^2 se necesita calcular la distribución marginal correspondiente. Para β se tiene

(7.5.9)
$$\begin{aligned}\pi(\overrightarrow{\beta}\,|\,\overrightarrow{y}) &= \int_0^{+\infty}\pi(\beta,\sigma^2|y)d\sigma^2 \\ &= \frac{\int_0^{+\infty}(\sigma^2)^{-(\frac{n}{2}+1)}e^{-\frac{1}{2\sigma^2}\{SC_T(\beta)+(n-p)s^2\}\sigma^{-2}}d\sigma^2}{\int_\beta\int_0^{+\infty}(\sigma^2)^{-(\frac{n}{2}+1)}e^{-\frac{1}{2\sigma^2}\{SC_T(\beta)+(n-p)s^2\}\sigma^{-2}}d\sigma^2 d\beta}\end{aligned}$$

pero por propiedades de la función gamma

(7.5.10)
$$\int_0^{+\infty} x^{-(k+1)}e^{-ax^{-1}}dx = a^{-k}\Gamma(k)$$

para $k=\frac{n}{2}$ y $a=\frac{1}{2}\{SC_T(\beta)+(n-p)S^2\}$, (véase el Ejercicio 7.17) se sigue que

$$\begin{aligned}\pi(\beta|y) &= \frac{\left[SC_T(\beta)+(n-p)s^2\right]^{-n/2}}{\int_\beta\left[SC_T(\beta)+(n-p)s^2\right]^{-n/2}d\beta} \\ &= \frac{\left[1+\frac{SC_T(\beta)}{(n-p)s^2}\right]^{-n/2}}{\int_\beta\left[1+\frac{SC_T(\beta)}{(n-p)s^2}\right]^{-n/2}d\beta}\end{aligned}$$

y como $SC_T(\beta)=(\beta-\widehat{\beta})'X'X(\beta-\widehat{\beta})$, el denominador es, por propiedades de la distribución T de Student p dimensional $T(n-p,\widehat{\beta},s^2(X'X)^{-1})$, igual a

$$\frac{\Gamma^p\left(\frac{1}{2}\right)\Gamma\left(\frac{n-p}{2}\right)(\sqrt{n-p})^{-p}}{\Gamma\left(\frac{n}{2}\right)|X'X|^{1/2}s^{-p}}$$

(véase el Apéndice sobre distribuciones), con lo que

(7.5.11)
$$\pi(\overrightarrow{\beta}\,|\,\overrightarrow{y}) = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma^p\left(\frac{1}{2}\right)\Gamma\left(\frac{n-p}{2}\right)}\frac{s^{-p}|X'X|^{1/2}}{(\sqrt{n-p})^p}\left[1+\frac{(\beta-\widehat{\beta})'X'X(\beta-\widehat{\beta})}{(n-p)s^2}\right]^{-n/2}$$

la distribución final de $\overrightarrow{\beta}$ es una distribución de Student p-multivariante

$$\beta|y \sim T(n-p,\widehat{\beta},s^2(X'X)^{-1})$$

Pueden hacerse inferencias sobre β mediante la distribución dada por (7.5.11), sin embargo se puede demostrar que si β sigue una distribución

T de Student p dimensional de parámetros $(n-p, \widehat{\beta}, s^2(X'X)^{-1})$ (véase la distribución en el Apéndice), entonces el siguiente cociente se distribuye según una F de Snedecor

$$\frac{(\beta-\widehat{\beta})'X'X(\beta-\widehat{\beta})}{ps^2} \sim F_{p,n-p} \tag{7.5.12}$$

(la demostración puede verse en Box y Tiao (1973), pág. 117).

Es importante señalar la identidad formal de este resultado con el análisis de la varianza clásica, tal y como ha sido recogido, por ejemplo, en el enunciado del Teorema 7.2.1, ya que permite dar una interpretación diferente de la tabla ADEVA recogida en la Tabla 7.2.2. En efecto, el primer miembro de (7.5.12) puede verse alternativamente como $\frac{SC_T/p}{SC_R/n-p}$, es decir se puede considerar, desde el punto de vista clásico, como el cociente de dos distribuciones independientes χ^2 divididas por sus grados de libertad respectivos; por lo tanto su distribución es una $F_{p,n-p}$. La novedad radica en la diferente interpretación de lo que se considera v.a., así en (7.5.12) las v.a. son las coordenadas del parámeto β mientras que en la aproximación clásica las v.a. son $\widehat{\beta}$ y s^2.

Si se está interesado en realizar contrastes respecto a un subconjunto de elementos del vector β, se puede utilizar que la distribución de Student p dimensional cumple, que un subconjunto de elementos de dimensión l, con $l \leq p$, tiene una distribución T de Student l dimensional (véase la distribución en el Apéndice). Por lo tanto $\beta_1, ..., \beta_l | \overrightarrow{y} \sim T(n-p, \widehat{\beta}_1, s^2 C_{11})$ donde

(7.5.13)

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \begin{matrix} l \\ p-l \end{matrix} \qquad \widehat{\beta} = \begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{pmatrix} \begin{matrix} l \\ p-l \end{matrix}$$

$$X'X = \left(\begin{array}{c|c} \overset{l}{X_1'X_1} & \overset{p-l}{X_1'X_2} \\ \hline X_2'X_1 & X_2'X_2 \end{array}\right) \begin{matrix} l \\ p-l \end{matrix} \qquad (X'X)^{-1} = \begin{pmatrix} \overset{l}{C_{11}} & \overset{p-l}{C_{12}} \\ C_{21} & C_{22} \end{pmatrix} \begin{matrix} l \\ p-l \end{matrix}$$

Esto permite realizar contrastes de hipótesis bayesianos sobre subcojuntos de valores de β.

Antes de aplicar estas ideas a un ejemplo, se va a estimar el otro valor desconocido del modelo. Para hacer inferencias respecto de la varianza σ^2, al igual que se hizo para β, se calcula su distribución marginal mediante

$$\begin{aligned} \pi(\sigma^2|\overrightarrow{y}) &= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \pi(\overrightarrow{\beta}, \sigma^2|\overrightarrow{y}) d\beta_1 ... d\beta_p \\ &= \frac{\int_{\overrightarrow{\beta}} (\sigma^2)^{-(\frac{n}{2}+1)} e^{-\frac{1}{2}\{SC_T(\beta)+(n-p)s^2\}\sigma^{-2}} d\overrightarrow{\beta}}{\int_0^{\infty} \int_{\overrightarrow{\beta}} (\sigma^2)^{-(\frac{n}{2}+1)} e^{-\frac{1}{2}\{SC_T(\beta)+(n-p)s^2\}\sigma^{-2}} d\overrightarrow{\beta}\, d\sigma^2} \\ &= \frac{(\sigma^2)^{-(\frac{n}{2}+1)} e^{-\frac{1}{2\sigma^2}(n-p)s^2} \int_{\overrightarrow{\beta}} e^{-\frac{1}{2\sigma^2}SC_T(\beta)} d\overrightarrow{\beta}}{\int_0^{\infty} (\sigma^2)^{-(\frac{n}{2}+1)} e^{-\frac{1}{2\sigma^2}(n-p)s^2} \int_{\overrightarrow{\beta}} e^{-\frac{1}{2\sigma^2}SC_T(\beta)} d\overrightarrow{\beta}\, d\sigma^2} \end{aligned} \tag{7.5.14}$$

Por propiedades de la distribución normal multivariante, (véase el Apéndice

con las distribuciones) se obtiene que

$$\int_{\overrightarrow{\beta}} e^{-\frac{1}{2\sigma^2}SC_T(\beta)}d\overrightarrow{\beta} = \int_{\overrightarrow{\beta}} e^{-\frac{1}{2\sigma^2}(\beta-\widehat{\beta})'\frac{X'X}{\sigma^2}(\beta-\widehat{\beta})}d\overrightarrow{\beta}$$
$$= \left(\sqrt{2\pi}\right)^p \left|\frac{X'X}{\sigma^2}\right|^{-\frac{1}{2}} = \left(\sqrt{2\pi}\right)^p \left|X'X\right|^{-\frac{1}{2}} (\sigma^2)^{p/2}$$

luego

$$\pi(\sigma^2|\overrightarrow{y}) = \frac{(\sigma^2)^{-\left(\frac{n}{2}+1\right)}e^{-\frac{1}{2\sigma^2}(n-p)s^2}\left(\sqrt{2\pi}\right)^p|X'X|^{-\frac{1}{2}}(\sigma^2)^{p/2}}{\int_0^{+\infty}(\sigma^2)^{-\left(\frac{n}{2}+1\right)}e^{-\frac{1}{2\sigma^2}(n-p)s^2}\left(\sqrt{2\pi}\right)^p|X'X|^{-\frac{1}{2}}(\sigma^2)^{p/2}d\sigma^2}$$
$$= \frac{(\sigma^2)^{-\left(\frac{n-p}{2}+1\right)}e^{-\frac{1}{2\sigma^2}(n-p)s^2}}{\int_0^{+\infty}(\sigma^2)^{-\left(\frac{n-p}{2}+1\right)}e^{-\frac{1}{2\sigma^2}(n-p)s^2}d\sigma^2}$$

aplicando la Expresión (7.5.10) con $k = \frac{n-p}{2}$ y $a = \frac{(n-p)s^2}{2}$, se obtiene

$$\pi(\sigma^2|y) = \frac{\left[(n-p)s^2\right]^{\frac{n-p}{2}}e^{-\frac{(n-p)s^2}{2\sigma^2}}}{2^{\frac{n-p}{2}}\Gamma\left(\frac{n-p}{2}\right)}(\sigma^2)^{-\left(\frac{n-p}{2}+1\right)} \tag{7.5.15}$$

que es la densidad de la distribución *Gamma-Inversa* de parámetros $a = \frac{(n-p)s^2}{2}$ y $\frac{n-p}{2}$ (véase la distribución en el Apéndice).

Otra vez se ha obtenido una identidad formal de este resultado con el análisis de la varianza clásico. En el Teorema 7.2.1, se obtuvo que $SC_R/\sigma^2 \sim \chi^2_{n-k}$, donde las v.a. eran las observaciones x_{ij} y por lo tanto SC_R, actuando σ^2 como parámetro. Aquí, desde el punto de vista bayesiano, si la distribución inicial es $\pi(\beta_1, ..., \beta_k, \sigma^2) = \sigma^{-2}$, la distribución final de SC_R/σ^2 es una χ^2_{n-k} donde la v.a. es σ^2 ; o bien, la distribución final de σ^2 es la *Gamma-Inversa* de parámetros $a = SC_R$ y $p = n - k$.

Los resultados de esta sección pueden resumirse en el siguiente enunciado.

Teorema 7.5.1
Para el modelo lineal

$$y_i = \sum_{j=1}^{p} x_{ij}\beta_j + e_i \qquad i = 1, ..., n$$

con $e_i \sim N(0, \sigma)$ e independientes, si la distribución inicial viene dada por

$$\pi(\beta_1, ..., \beta_p, \sigma^2) = \frac{1}{\sigma^2}, \quad \sigma^2 > 0$$

entonces la distribución final

$$\beta_1, ..., \beta_p|\overrightarrow{y} \sim T_p(n-p, \widehat{\beta}, s^2(X'X)^{-1})$$

además condicionado a $\overrightarrow{y}$

$$\frac{(\beta-\widehat{\beta})'X'X(\beta-\widehat{\beta})}{ps^2}=\frac{(\beta-\widehat{\beta})'X'X(\beta-\widehat{\beta})/r}{(y-\widehat{y})'(y-\widehat{y})/n-r}\sim F_{r,n-r}$$

$$\widehat{\beta}=(X'X)^{-1}X'Y,\qquad \widehat{Y}=X\widehat{\beta},\qquad s^2=\frac{SC_R}{n-p}=\frac{1}{n-p}(Y-\widehat{Y})'(Y-\widehat{Y}).$$

La distribución final de σ^2 es la *Gamma-Inversa*

$$\sigma^2|y\sim GI(s^2,n-p).$$

■

El análisis bayesiano del modelo lineal, con la distribución inicial que se ha empleado, reproduce el modelo clásico aunque, como se ha dicho, en el modelo clásico las v.a. son las observaciones y_i mientras que en el modelo bayesiano se actúa condicionado por las observaciones y_i y las v.a. son $\overrightarrow{\beta}$ y σ^2. Un desarrollo más general, con otras distribuciones iniciales, puede verse en Press (1989), pág. 139.

Ejemplo 7.5.1
Para los datos de la preparación de una sustancia química contenidos en la Tabla 7.2.1 el modelo es

$$y=X\beta+e$$

donde

$$y'=(63,...,37)\qquad X'X=\begin{pmatrix}6&0&0&0&0\\0&6&0&0&0\\0&0&6&0&0\\0&0&0&6&0\\0&0&0&0&6\end{pmatrix}\qquad X'y=\begin{pmatrix}306\\157\\165\\281\\198\end{pmatrix}$$

$$(X'X)^{-1}=\begin{pmatrix}1/6&0&0&0&0\\0&1/6&0&0&0\\0&0&1/6&0&0\\0&0&0&1/6&0\\0&0&0&0&1/6\end{pmatrix}\qquad \widehat{\beta}=(X'X)^{-1}X'y=\begin{pmatrix}51\\26'1667\\27'5\\46'8333\\33\end{pmatrix}$$

$$s^2=\frac{1}{30-5}(y-\widehat{y})'(y-\widehat{y})=\frac{4627'1667}{25}$$

por lo tanto, si

$$\pi(\beta_1,...,\beta_5,\sigma^2)=\sigma^{-2}$$

entonces

$$\beta_1,...,\beta_5|\overrightarrow{y}\sim T(5,\widehat{\beta},30'84778I,25)$$

siendo I la matriz identidad de dimensión 5×5.

7.6. APROXIMACIÓN HISTÓRICA

Las primeras ideas relacionadas con el análisis de la varianza, subdividiendo una suma de cuadrados en dos componentes, pueden remontarse a Gauss (1777-1855) y Laplace (1749-1827) cuando introducen el método de los mínimos cuadrados. No obstante, su estudio sistemático y la utilización de la ortogonalización, se deben al genio de Fisher alrededor de 1923.

La paternidad de la correlación ha dado lugar a un interesante debate. Se ha dicho que Gauss puede introducirla relacionada con su estudio de la distribución normal. Karl Pearson en 1895, considera a Augusto Brevais (1811-1863) como el precursor por unos comentarios que dirige al trabajo de Gauss, no obstante parece más ajustado señalar a Francis Galton como su auténtico originador en 1877, cuando recalca la necesidad de introducir una medida de la relación existente entre variables bidimensionales. A él se debe la acuñación del término *reversión* como la tendencia de un tipo filial medio a desviarse del tipo de los padres para volver al cabo de alguna generación a *regresar* al tipo medio ideal. Así, es Galton quién expresa la necesidad de construir un *coeficiente de reversión* que indique la reducción de la variabilidad, el término reversión se transformará en *regresión* en un trabajo posterior, de 1889. En 1896 Karl Pearson retomará los conceptos de Galton, para darles la formulación con la que son conocidos actualmente.

Francis Galton (1822-1911) nace cerca de Birmingham, en una familia de banqueros y armadores y es primo de Charles Darwin, estudia matemáticas en Cambridge y medicina en Londres. A los 72 años muere su padre dejándole una gran fortuna, lo que le permite viajar y completar su formación, sus descubrimientos geográficos le valieron la medalla de oro de la *Geographical Society*. Ha sido considerado el último científico de la era victoriana, con un saber erudito en diferentes campos, como psicología, antropología, sociología, educación y huellas dactilares. A él se debe el término anticiclón y el término Eugenesia, para la ciencia que tiene por objeto la mejora de las especies. Con su interés por el estudio intrínseco de las cantidades, contribuyó al desarrollo de los métodos estadísticos y su aplicación a otros campos de las ciencias aplicadas. Fundó la revista *Biometrika* para acoger publicaciones relacionadas con la medida de magnitudes relacionadas con la Biología y dotó la cátedra Galton, que sería ocupada por Karl Pearson.

Como estadístico Galton toma el relevo de Quetelet, sin embargo rompe con él en el sentido de que no trabaja con el hombre medio y prefiere realizar las comparaciones mediante diagramas de puntos. Encuentra que las desviaciones de los individuos, siguen una curva para cada una de las especies y cada propiedad, lo que le permite dar una característica de cada especie, así como detectar si un individuo pertenece o no a la especie en estudio.

En 1875 considera el aspecto de un fruto como debido a una combinación de efectos como cantidad de sol, terreno donde fué sembrado, riego, etc y observa que el aspecto del fruto se desvia de un ideal de acuerdo con la dis-

tribución normal. A continuación clasifica los frutos en distintas categorías del tipo: los que recibieron mucho sol, los que recibieron una cantidad media y los que recibieron poco sol, para observar que cada uno de éstos también se desvía de un ideal medio con arrego a una normal. Se plantea la pregunta de cómo de subpoblaciones normales se llega a una población normal y anticipa así la idea de la reproductividad de la distribución normal.

Los trabajos de Galton fueron continuados por Karl Pearson y por sus colaboradores entre los que cabe citar a Edgeworth (1845-1926) y Yule (1871-1951), lo que ha dado lugar a la introducción de otras medidas de dependencia más generales, que serán estudiadas en el capítulo siguiente.

7.7. EJERCICIOS

7.7.1

Para determinar la calidad de una alimentación, quince ratas fueron alimentadas asignándolas al azar a una de tres dietas con distintos niveles de proteínas, obteniéndose las siguientes ganancias en peso:

nivel de proteínas en la dieta		
bajo	medio	alto
3'91	8'56	22'35
3'85	8'87	23'31
3'25	9'20	32'82
2'60	9'42	24'15
3'70	10'40	20'29

mediante un análisis de la varianza, contrástese si la ganancia media en peso es significativamente distinta con los distintos niveles de proteínas. A continuación contrástese si la ganancia media en peso es mayor con el nivel de proteínas alto que con el nivel medio y con éste mayor que con el nivel bajo.

7.7.2

Cuatro atletas de triple salto saltan las siguientes distancias, en metros, en diferentes reuniones de atletismo. ¿Puede suponerse que los cuatro saltan la misma distancia?

atletas			
1	2	3	4
8,25	7,31	7	6'69
6'97	8,26	7,25	7'32
7'65	6'50	8	
7'86	9'42		
	10'40		

7.7.3

La siguiente tabla recoge el rendimiento, en unidades convenientes, de 48 parcelas en las que se han experimentado 3 tipos de abonos y 4 tratamientos distintos, asignando las combinaciones al azar. Realícese un análisis de la varianza de dos factores, para ver si existen diferencias entre los tipos de abonos, entre los distintos tratamientos y si existe interacción entre abonos y tratamientos.

tipos de abonos	tratamientos							
	1		2		3		4	
1	5'2	6'6	10'1	12'9	6'5	6'7	6'5	9'1
	6'7	6'4	10'7	9'1	8'5	9'8	8'6	8'2
2	5'7	5	11'1	8	6'6	5'7	7'6	12'2
	6'1	4'4	6'8	14'3	5'3	6'2	9'1	5'8
3	4'3	4'2	4'9	5'6	4'5	4'7	5	5'6
	3'9	4'4	5'7	4'8	4'6	4'4	5'1	5'3

7.7.4

Verifíquese que, en el análisis de la varianza con un sólo factor, si el número de muestras es 2, la expresión dada por (7.2.4)

$$\widehat{\sigma}_p^2 = \frac{1}{n-k}\sum_{i=1}^{k}(n_i-1)s_i^2$$

se reduce a la varianza muestral ponderada, utilizada en la comparación de las medias para dos distribuciones normales recogida en la Expresión (5.2.2.4)

$$S_p^2 = \frac{(m-1)S_1^2 + (n-1)S_2^2}{m+n-2}$$

7.7.5

Compruébese que en un análisis de la varianza con un solo factor y dos muestras, el estadístico F coincide con T^2, donde T es el estadístico de Student que aparece en el contraste de igualdad de medias con distribuciones normales de varianzas desconocidas pero iguales.

7.7.6

La siguiente tabla, recoge el ph de la placa dental a los 20, 30 y 50 minutos de haber comido azúcar en 8 voluntarios. Se sabe por otras consideraciones que el pH es no decreciente. ¿Puede asegurarse que el pH es constante en

los 3 instantes de tiempo mediante el análisis de la varianza?

	tiempo		
sujetos	20'	30'	50'
1	4'6	4'4	4'1
2	4'2	4'7	4'7
3	3'9	5'0	5'7
4	4'6	5'2	5'6
5	4'6	5'1	5'7
6	4'7	5'3	5'3
7	4'9	5'6	5'9
8	5'0	6'0	5'9

7.7.7

Para estudiar la duración de los generadores solares de una determinada marca, se considera el número de horas que son utilizados por semana, que se supone a dos niveles, y el tamaño del colector, que se supone tiene tres niveles. Se realizan 3 observaciones por cada combinación del nivel de horas por semana con el nivel correspondiente al tamaño del colector. La tabla siguiente recoge los resultados, en miles de horas, del experimento hasta que el generador solar se estropea.

horas por	tamaño del colector		
semana	1	2	3
1	3'4	7'6	6'1
	5'1	12'4	2'1
	4'2	8'9	7'5
2	2'9	5'6	8'1
	4'8	11'2	12
	10	9'5	5'9

Se pide:

a) Realícese un análisis de la varianza para contrastar los efectos de la interacción.

b) Represéntese en abcisas los niveles del factor tamaño del colector y en ordenadas los totales correspondientes a cada combinación de nivel de horas por semana y tamaño del colector, e interprétese gráficamente la interacción observada en la parte a).

7.7.8

Un especie de árbol ha sido plantada y el diámetro del tronco en 10 años sucesivos ha sido:

año	1	2	3	4	5	6	7	8	9	10
diámetro (cm.)	1'8	3	4'2	5'8	6'9	7'7	7'9	8'3	9'1	9'7

a) Represéntese gráficamente el diámetro en función del año y determínese la recta de regresión de y sobre x.

b) Calcúlese la suma de cuadrados de los errores cometidos por la aproximación lineal.

c) La recta de regresión para un estudio de una especie de árbol similar ha sido $y = 0'8 + 0'8x$. Contrástese si la recta de regresión obtenida es compatible con ésta.

7.7.9

La tabla siguiente, tomada de Rice (1995), recoge la presión en función de la temperatura a la que se produce una determinada reacción química:

Temperatura (en grados)	Presión (en barios)
20'9	25.526
21	25.216
22'1	25.187
22'5	25.177
25	25.088
34	24.711
34'1	24.716
42'7	24.394
49'9	24.067
50'1	24.057

Se pide:

a) Represéntese la colección de puntos (temperatura, presión) y ajústese una relación lineal de la presión en función de la temperatura.

b) Contrástese si la pendiente de la recta de regresión es 0.

c) Represéntense los residuos.

7.7.10

a) Demuéstrese que si para una regresión lineal se estima el valor correspondiente a x_0 por $\widehat{y}(x_0) = \widehat{\beta}_0 + \widehat{\beta}_1 x_0$ la distribución de

$$\frac{\widehat{\beta}_0 + \widehat{\beta}_1 x_0 - (\beta_0 + \beta_1 x_0)}{s\sqrt{\frac{1}{n}\left[1 + \frac{(x_0 - \overline{x})^2}{b_x^2}\right]}}$$

donde $s^2 = \frac{1}{n-2}\sum_{j=1}^{n}(Y_j - \widehat{Y}_j)^2$ y $b_x^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j - \overline{x})^2$, es una T de Student con $n-2$ grados de libertad.

b) Utilícese el apartado anterior, para determinar un intervalo de grado de confianza $1-\alpha$ para el valor

$$y(x_0) = \beta_0 + \beta_1 x_0.$$

c) Dados los datos de la siguiente tabla, correspondientes a alturas en metros de 8 individuos:

A los 2 años	0'76	0'77	0'81	0'86	0'87	0'9	0'91	0'99
A los 14 años	1'63	1'60	1'61	1'70	1'72	1'73	1'78	1'80

¿Cuál es el intervalo de grado de confianza 0′95, para la altura a los 14 años de un individuo que a los 2 años medía 0'95 metros?

7.7.11

En un estudio para relacionar la proporción de unidades defectuosas y la producción realizada, se dispone de los siguientes datos:

Producción en unidades por meses	1000	1500	2000	2500	3000	3500
Proporción de unidades defectuosas	5	5'2	5'7	6'5	6'8	7'1

4000	4500	5000	5500	6000	6500	7000	7500
8'1	10	10'2	10'3	13	15'3	20	25

Un analista sugiere utilizar un modelo exponencial de la forma $y = ab^x$ mientras que otro sugiere emplear un polinomio de segundo grado. Determínese qué modelo ofrece mejor ajuste. Utilícense las unidades producidas por mes como variable no aleatoria.

7.7.12

En un estudio para relacionar los gastos realizados en alojamiento con los gastos hechos en actividades de esparcimiento, se dispone de los datos en tanto por ciento correspondientes a 15 casas tomadas al azar:

Gastos en alojamiento	12	13	13'5	14	14'5	15	16'5	17	17'5
Gastos en esparcimiento	20'3	20	20'5	19'5	18'5	18	17'5	17	16

18	19	21	22'5	24	25
18	16	15'5	15'2	15	15

a) Se sugiere utilizar un modelo de la forma

$$y = a + \frac{b}{x}$$

donde x es el tanto por ciento de gastos en alojamiento.

b) ¿Sería mejor utilizar un ajuste lineal?

7.7.13

a) Para los datos del Ejercicio 7.7.11, ajústese un modelo potencial de la forma

$$y = ax^b$$

b) ¿Es mejor el modelo potencial o el modelo exponencial?

7.7.14

Para estudiar la relación de las pérdidas de cosechas con la lluvia, el periodo de tiempo lloviendo y la temperatura, en una determinada área, se dispone de los siguientes datos:

Lluvia	Periodo	Temperatura	Pérdidas
4'3	59	70	12'3
4'5	61	71	12'5
4'3	74	72	15'0
6'1	60	66	13'1
5'6	78	65	18'7
5'6	85	64	17'9
4'1	65	63	13'5
5'5	76	64	18'1
5'0	83	66	16'3
5'6	71	66	14'0
5'2	77	66	14'0
4'8	84	69	19'5
3'8	63	77	12'7
3'4	65	76	14'5
3'6	77	75	16'1
3'9	71	73	15'0
5'1	75	71	15'5
5'9	77	70	17'3
4'6	70	68	15'9
4'8	77	66	14'6
4'9	75	66	16'5
5'1	69	65	14'5
5'4	81	63	18'0
6'5	80	62	18'5
6'8	87	60	19'0
6'2	87	60	19'5

a) Represéntense las colecciones de puntos correspondientes a los pares (periodo, pérdidas), (temperatura, pérdidas) y (lluvia, pérdidas).

b) Ajústense linealmente las pérdidas en función de las variables lluvia, periodo y temperatura.

c) Ajústense linealmente las pérdidas en función de la lluvia y la temperatura.

7.7.15

Demuéstrese que si X es un vector aleatorio de n filas y una columna con esperanza $E[X] = \mu\mathbf{1}$ y matriz de covarianzas $\Sigma_X = \sigma^2 I$, entonces la v.a. $\overline{X}$ está incorrelada con cada una de las diferencias $X_i - \overline{X}$, $i = 1, ..., n$.

7.7.16

Véase, que el modelo del análisis de la varianza con un sólo factor

$$y_{ij} = \theta_i + \varepsilon_{ij} \qquad i = 1, ..., k \qquad j = 1, ..., n_i$$

en el que se contrasta $H_0 : \theta_1 = ... = \theta_k$, puede expresarse matricialmente en la forma

$$Y = X\theta + \varepsilon$$

con X matriz $n \times k$, donde $n = \sum_{i=1}^{k} n_i$, θ es un vector columna de k elementos y $H_0 : H\theta = \mathbf{0}$ para alguna matriz H de dimensión $(k-1) \times k$.

7.7.17

Compruébese que para $a > 0$ y $k > 0$ se tiene

$$\int_0^{+\infty} x^{-(k+1)} e^{-ax^{-1}} dx = a^{-k}\Gamma(k).$$

7.7.18

Box y Tiao (1973), pág. 126, recogen un experimento para determinar los pesos (β_1, β_2) de dos objetos extremadamente pequeños A y B, a partir de 18 pesadas en microgramos (1 microgramo=10^{-3} gramos) realizadas de la forma siguiente:

únicamente A	únicamente B	ambos A y B
109,85	114,121,140,122,189 129,98,134,133	217,203,243,229, 233,221,221

si se supone válido el modelo

$$y_{ij} = \beta_i + \varepsilon_{ij} \quad i = 1, 2 \quad j = 1, ..., n_i$$

donde los errores ε_{ij} tienen distribución $N(0, \sigma)$ con σ desconocida y son independientes. Se pide:

a) Determínese la distribución final $\pi(\beta_1, \beta_2|\overrightarrow{y})$.

b) Contrástese $H_0 : \beta_1 = 100, \beta_2 = 125$ frente a H_2: el complementario de H_0.

7.7.19

Para los datos de la Tabla 7.2.1, que se reproduce a continuación

Métodos				
1	2	3	4	5
63	27	10	28	47
48	41	30	45	28
57	23	24	44	30
54	31	37	31	35
41	2	40	29	21
43	33	24	44	37

y supuesto válido el modelo normal con

$$y_{ij} = \beta_i + \varepsilon_{ij} \quad i = 1, ..., 5 \quad j = 1, ..., 6$$

donde los errores $\varepsilon_{ij} \sim N(0, \sigma)$ con σ desconocida e independientes. Se pide:

a) Determínese la distribución final $\pi(\beta_1, ..., \beta_5 | \overrightarrow{y})$.

b) Contrástese $H_0 : \beta_2 = 30, \beta_5 = 30$ frente a $H_1 : \beta_2 \neq 30, \beta_5 \neq 30$.

7.8. SOLUCIÓN A LOS EJERCICIOS

7.8.1

El modelo es

$$x_{ij} = \theta_i + \varepsilon_{ij} \quad i = 1, 2, 3 \quad j = 1, ..., 5$$

y se trata de constrastar $H_0 : \theta_1 = \theta_2 = \theta_3$ frente a su complementario. El test tiene de región crítica

$$RC = \left\{ \frac{SC_E/k-1}{SC_D(n-k} \geq f_{k-1,n-k;\alpha} \right\}$$

con $SC_E = \sum_{i=1}^{3} n_i(\overline{x}_{i\cdot} - \overline{x}_{\cdot\cdot})^2$ *suma de cuadrados entre* grupos y $SC_D = \sum_{i=1}^{3} \sum_{j=1}^{5} (x_{ij} - \overline{x}_{i\cdot})^2$ *suma de cuadrados dentro* de los grupos o residual (véase la Expresión (7.2.9)), poniendo $k - 1 = 2$, $n - k = 12$ y con $\alpha = 0'05$

$$RC = \left\{ \frac{SC_E/2}{SC_D/12} \geq 3'89 \right\}$$

como el valor del estadístico es $\frac{1165'87/2}{80'56/12} = 86'83$ se rechaza la hipótesis nula, con lo que la ganancia en peso media es significativamente distinta a los distintos niveles de proteinas.

Para contrastar si es $H_0 : \theta_3 > \theta_2 > \theta_1$, se puede recurrir a contrastar simultáneamente las hipótesis $H_0 : \theta_1 - \theta_2 = 0$ frente a $H_1 : \theta_1 - \theta_2 < 0$ y $H_0' : \theta_2 - \theta_3$ frente a $H_1' : \theta_2 - \theta_3 < 0$. Es decir, el primer contraste es con $\alpha_1 = 0'05/2 = 0'025$ $a_1 = (1, -1, 0)$ y el segundo con $\alpha_1 = 0'025$, $a_2 = (0, 1, -1)$. Para el primero la región crítica viene dada por

$$RC_1 = \left\{ \overline{x}_{1\cdot} - \overline{x}_{2\cdot} \leq \widehat{\sigma}_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} t_{n-k;\alpha_1} \right\}$$

(véase la Expresión (7.2.10)). Mientras que para el segundo

$$RC_2 = \left\{ \overline{x}_{2\cdot} - \overline{x}_{3\cdot} \leq \widehat{\sigma}_p \sqrt{\frac{1}{n_2} + \frac{1}{n_3}} t_{n-k;\alpha_2} \right\}$$

como $\overline{x}_{1\cdot} = 3'46$, $\overline{x}_{2\cdot} = 9'29$, $\overline{x}_{3\cdot} = 24'38$, $n_1 = n_2 = n_3 = 5$, $\alpha_1 = \alpha_2 = 0'025$, $t_{12;0'025} = 23'3$ y $\widehat{\sigma}_p^2 = 6'97$ se sigue que

$$RC_1 = \{\overline{x}_{1\cdot} - \overline{x}_{2\cdot} \leq 38'92\}$$

y como $\overline{x}_{1\cdot} - \overline{x}_{2\cdot} = -4'83$ es $\theta_1 - \theta_2 < 0$. De manera análoga

$$RC_2 = \{\overline{x}_{2\cdot} - \overline{x}_{3\cdot} \leq 38'92\}$$

y como $\overline{x}_{2\cdot} - \overline{x}_{3\cdot} = -15'09$ es $\theta_2 - \theta_3 < 0$; con lo que efectivamente es $\theta_1 < \theta_2 < \theta_3$.

7.8.2

Se supone que la distancia a la que salta cada uno, x_{ij}, sigue el modelo

$$x_{ij} = \theta_i + \varepsilon_{ij} \qquad i = 1, ..., 4 \qquad j = 1, ..., n_i$$

con $n_1 = 4$, $n_2 = 5$, $n_3 = 3$ y $n_4 = 2$. Se pretende contrastar

$$H_0 : \theta_1 = \theta_2 = \theta_3 = \theta_4 \quad \text{frente a } H_1 : \text{al menos dos son diferentes.}$$

La región crítica viene dada por

$$RC = \left\{ \frac{SC_E/k-1}{SC_D/n-k} \geq f_{k-1,n-k;\alpha} \right\}$$

(véase (7.2.9)), sustituyendo $SC_E = 2'7519$, $SC_D = 8'1879$, $k-1 = 3$, $n-k-1 = 10$, $\alpha = 0'05$ y $f_{3,10;0'05} = 3'71$ queda

$$RC = \left\{ \frac{SC_E/k-1}{SC_D/n-k} \geq 3'71 \right\}$$

como el valor del estadístico es $1'12$ se acepta H_0; los 4 atletas saltan lo mismo.

7.8.3

El modelo es de la forma

$$x_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \qquad i = 1, ..., 3 \quad j = 1, ..., 4 \quad k = 1, ..., 4$$

donde μ es el valor medio, α_i representa el efecto debido al tipo de abono, β_j el efecto debido a cada tratamiento, γ_{ij} el efecto debido conjuntamente a abono y tratamiento y ε_{ijk} el error aleatorio de cada observación.

Los diagramas de caja correspondientes a las variables abono y rendimiento, así como tratamiento y rendimiento, vienen recogidos en la Figura 7.8.1, donde se observan diferencias en el rendimiento en función de los distintos tipos de abono y menos marcadas en el caso del rendimiento en función de los tratamientos.

Las estimaciones de los parámetros del modelo son:

$$\overline{x}_{...} = 7'84 \quad \widehat{\alpha}_1 = 8'225 \quad \widehat{\alpha}_2 = 7'494 \quad \widehat{\alpha}_3 = 4'812 \quad \widehat{\beta}_1 = 5'242 \quad \widehat{\beta}_2 = 8'667$$
$$\widehat{\beta}_3 = 6'125 \quad \widehat{\beta}_4 = 7'342 \quad \widehat{\gamma}_{11} = 6'225 \quad \widehat{\gamma}_{12} = 10'7 \quad \widehat{\gamma}_{13} = 7'875 \quad \widehat{\gamma}_{14} = 8'1$$

$$\widehat{\gamma}_{21} = 4'3 \quad \widehat{\gamma}_{22} = 10'05 \quad \widehat{\gamma}_{23} = 5'95 \quad \widehat{\gamma}_{24} = 8'675$$
$$\widehat{\gamma}_{31} = 4'2 \quad \widehat{\gamma}_{32} = 5'25 \quad \widehat{\gamma}_{33} = 4'55 \quad \widehat{\gamma}_{34} = 5'25.$$

La tabla ADEVA es de la forma

Fuentes de variación	Suma de cuadrados	Grados de libertad	Estadísticos F
abono	1003'3012	2	23'222
tratamiento	79'8506	3	11'967
interacción	25'0137	6	1'874
error	80'0725	36	
total	288'2381	47	

Para contrastar la hipótesis de que los tres tipos de abono producen el mismo efecto, es decir, $H_A : \alpha_1 = \alpha_2 = \alpha_3 = 0$ frente a su complementario, con $\alpha = 0'05$ la región crítica es

$$RC = \left\{ \frac{SC_A/p-1}{SC_E/rs(t-1)} \geq f_{2,36;0'05} = 3'26 \right\}$$

y como el valor obtenido del estadístico es $23'32$ se rechaza H_A significativamente, los tipos de abono producen efectos diferentes.

Para contrastar la hipótesis de que los 4 tratamientos son equivalentes

$$H_B : \beta_1 = \beta_1 = \beta_3 = \beta_4 = 0$$

la región crítica es de la forma

$$RC = \left\{ \frac{SC_B/s-1}{SC_E/rs(t-1)} \geq f_{2,36;0'05} = 2'87 \right\}$$

como el estadístico es $11'967$ se rechaza H_B.

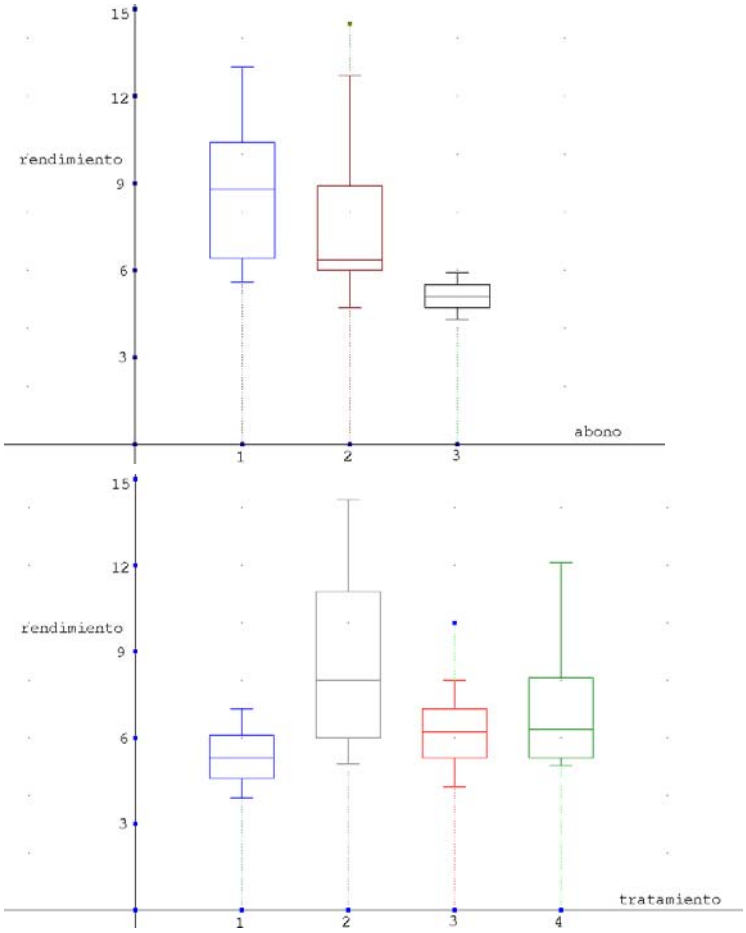


Figura 7.8.1. Diagramas de caja correspondiente al Ejercicio 7.7.3 de las variables (abono, rendimiento) y (tratamiento, rendimiento).

Por último, para contrastar si la interacción es nula,

$$H_{AB} : \gamma_{ij} = 0 \qquad i = 1, 2, 3 \qquad j = 1, 2, 3, 4$$

la región crítica es

$$RC = \left\{ \frac{SC_{AB}/(r-1)(s-1)}{SC_E/rs(t-1)} \geq f_{6,36;0'05} = 2'36 \right\}$$

y como el estadístico vale $1'875$ se acepta que no hay interacción.

7.8.4

Para $k = 2$ con $n_1 = m$, $n_2 = n$ se tiene que $n = n_1 + n_2 = m + n$, por

tanto

$$\widehat{\sigma}_p^2 = \frac{1}{n-2}\sum_{i=1}^{2}(n_i-1)s_i^2 = \frac{1}{m+n-2}\sum_{i=1}^{2}(n_i-1)s_i^2$$
$$= \frac{(m-1)s_1^2+(n-1)s_2^2}{m+n-2}$$

que es la expresión de la varianza muestral ponderada.

7.8.5
El estadístico F vale, (véase la Tabla 7.2.2)

$$F = \frac{SC_E/k-1}{SC_D/n-k} = \frac{\sum_{i=1}^{k} n_i(\overline{x}_{i\cdot}-\overline{x}_{\cdot\cdot})^2/(n-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\overline{x}_{i\cdot})^2/(n-k)}$$

que para $k=2$ y como el denominador es $\frac{1}{m+n-2}\left[\sum_{j=1}^{m}(x_{1j}-\overline{x})^2+\sum_{j=1}^{n}(x_{2j}-\overline{y})^2\right]$ puede escribirse como

$$F = \frac{m(\overline{x}-\overline{x}_{\cdot\cdot})+n(\overline{y}-\overline{x})^2}{S_p^2}$$

si se sustituyen $\overline{x}-\overline{x}_{\cdot\cdot} = n(\overline{x}-\overline{y})/(m+n)$ y $\overline{y}-\overline{y}_{\cdot\cdot} = m(\overline{y}-\overline{x})/(m+n)$ se puede poner

$$F = \frac{(\overline{x}-\overline{y})^2(mn^2+nm^2)/(m+n)^2}{S_p^2}$$
$$= \frac{(\overline{x}-\overline{y})^2 mn/((m+n)}{S_p^2}$$

que coincide con T^2, ya que por la Expresión (5.2.2.3)

$$T = \frac{[\overline{x}-\overline{y}-(\theta_1-\theta_2)]/\sqrt{\frac{1}{m}+\frac{1}{n}}}{S_p}$$

y si se cumple $H_0 : \theta_1 = \theta_2$ es

$$T = \frac{(\overline{x}-\overline{y})\sqrt{\frac{mn}{m+n}}}{S_p}.$$

En este sentido, el análisis de la varianza generaliza la comparación de dos medias para distribuciones normales.

7.8.6

Se supone el modelo

$$x_{ij} = \mu + \theta_i + \varepsilon_{ij} \quad \text{con} \quad i = 1,2,3 \quad j = 1,...,8$$

y se pretende contrastar la hipótesis nula $H_0 : \theta_1 = \theta_2 = \theta_3 = 0$ frente a la alternativa $H_1 : \theta_1 < \theta_2 < \theta_3$. Si se realizan los dos contrastes $H_0 : \theta_1 = \theta_2$ frente a $H_1 : \theta_1 < \theta_2$ y $H_0' : \theta_2 = \theta_3$ frente a $H_1' : \theta_2 < \theta_3$, mediante un test conjunto de tamaño $\alpha = 0'05$, caso de que se rechacen ambas hipótesis se tendría que es $\theta_1 < \theta_2 < \theta_3$.

Para conseguir que el test conjunto sea de tamaño $0'05$, basta imponer que $\alpha_i = \alpha/2 = 0'025$ para cada una de las regiones críticas de los tests. Con lo que para el contraste $H_0 : \theta_1 = \theta_2$ frente a $H_1 : \theta_1 < \theta_2$

$$RC_1 = \left\{ \overline{x}_{1\cdot} - \overline{x}_{2\cdot} \leq \widehat{\sigma}_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} t_{n-k;1-\frac{\alpha}{2}} \right\}$$

lo que lleva a $\widehat{\sigma}_p = 2'882$, $\sqrt{\frac{1}{8} + \frac{1}{8}} = 0'5$ y $t_{21;0'975} = -2'08$

$$RC_1 = \{\overline{x}_{1\cdot} - \overline{x}_{2\cdot} \leq -2'997\}$$

el valor del estadístico es $\overline{x}_{1\cdot} - \overline{x}_{2\cdot} = 4'562 - 5'162 = -0'6$, luego se mantiene H_0.

Para el contraste $H_0' : \theta_2 = \theta_3$ frente a $H_1 : \theta_2 < \theta_3$

$$RC_2 = \left\{ \overline{x}_{2\cdot} - \overline{x}_{3\cdot} \leq \widehat{\sigma}_p \sqrt{\frac{1}{n_2} + \frac{1}{n_3}} t_{n-k;1-\frac{\alpha}{2}} \right\}$$

y como

$$RC_2 = \{\overline{x}_{2\cdot} - \overline{x}_{3\cdot} \leq -2'997\}$$

al ser el valor del estadístico $\overline{x}_{2\cdot} - \overline{x}_{3\cdot} = 5'162 - 5'362 = -0'2$, se mantiene H_0'. Por lo tanto se concluye que el valor del pH es constante.

7.8.7

El modelo es de la forma

$$x_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \qquad i = 1,2 \quad j = 1,2,3 \quad k = 1,2,3$$

con μ el valor medio, α_i el efecto debido al número de horas por semana empleado en el experimento, β_j el efecto debido al tamaño del colector y γ_{ij} el efecto debido a la interacción.

La representación mediante el diagrama de caja queda recogida en la Figura 7.8.2, donde se observan discrepancias en la representación de las duraciones en función del tamaño del colector.

a) La tabla ADEVA viene dada por

Fuentes de variación	Suma de cuadrados	Grados de libertad	Estadísticos F
horas por semana	8'96	1	1'168
tamaño del colector	51'388	2	3'349
interacción	14'014	2	0'913
error	92'073	12	
total	166'436	17	

Para contrastar si la interacción es nula, es decir, $H_{AB} : \gamma_{ij} = 0 \;\; i = 1, 2$ $j = 1, 2, 3$ frente a su complementario, con tamaño $0'05$, la región crítica es

$$RC = \left\{ \frac{SC_{AB}/(r-1)(s-1)}{SC_E/rs(t-1)} \geq f_{2,12;0'05} = 3'89 \right\}$$

el valor del estadístico es $0'913$ luego se acepta H_0, los datos no son significativos de que haya interacción.

b) Los totales para cada celda, combinación del nivel de horas por semana y tamaño del colector son:

horas por semana	tamaño del colector		
	1	2	3
1	12'7	28'9	15'7
2	17'7	26'3	26

La representación gráfica de la Figura 7.8.3 pone de manifiesto la pequeña discrepancia que existe entre el factor fila, horas por semana, al nivel 1 y nivel 2, en función del factor columna, tamaño del colector. Lo que explica que el test de la parte a) saliera no significativo, ya que cuanto más alejadas estén de ser paralelas ambas rectas, más alta será la interacción entre ambos factores.

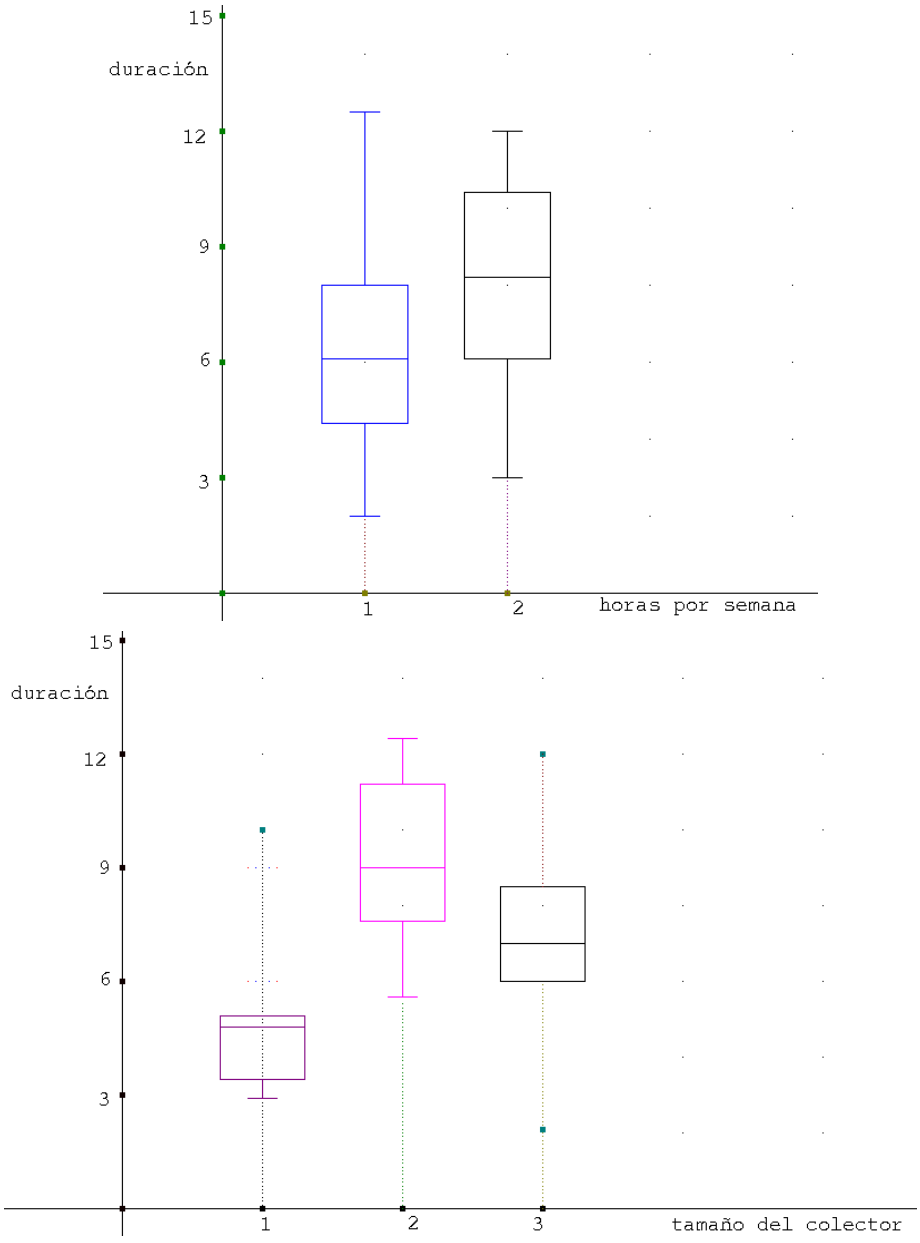


Figura 7.8.2. Diagramas de caja, correspondientes al Ejercicio 7.8.7, de las variables (horas por semana, duración) y (tamaño del colector, duración).

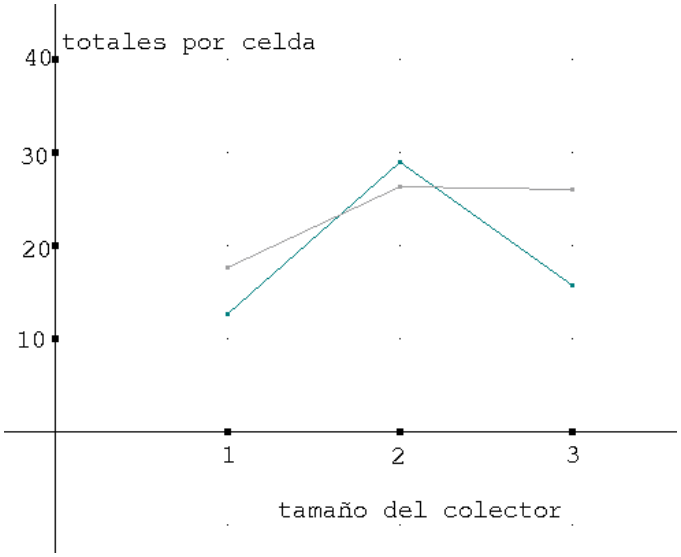


Figura 7.8.3. Totales por celda del factor fila en función del factor columna para el Ejercicio 7.8.7.

7.8.8

a) La representación gráfica de la Figura 7.8.4, parece aconsejar la conveniencia de la aproximación lineal, en ella se ha incluido también la representación de la recta de regresión de y sobre x.

Para el cálculo de la recta de regresión de y sobre x

$$y - \overline{y} = \frac{b_{xy}}{b_x^2}(x - \overline{x})$$

(véase (7.4.1.4)) se necesitan los siguientes valores

$$\overline{x} = 5'5 \qquad \overline{y} = 6'44 \qquad \sum_{i=1}^{10} x_i^2 = 385 \qquad \sum_{i=1}^{10} y_i^2 = 478'62 \qquad \sum_{i=1}^{10} x_i y_i = 424'9$$

$$b_{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{xy} = 42'49 - 35'42 = 7'07$$

$$b_x^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2 = 38'5 - 30'25 = 8'25$$

de donde

$$y - 6'44 = \frac{7'07}{8'25}(x - \overline{x})$$

o alternativamente

$$y = 1'7265 + 0'857x$$

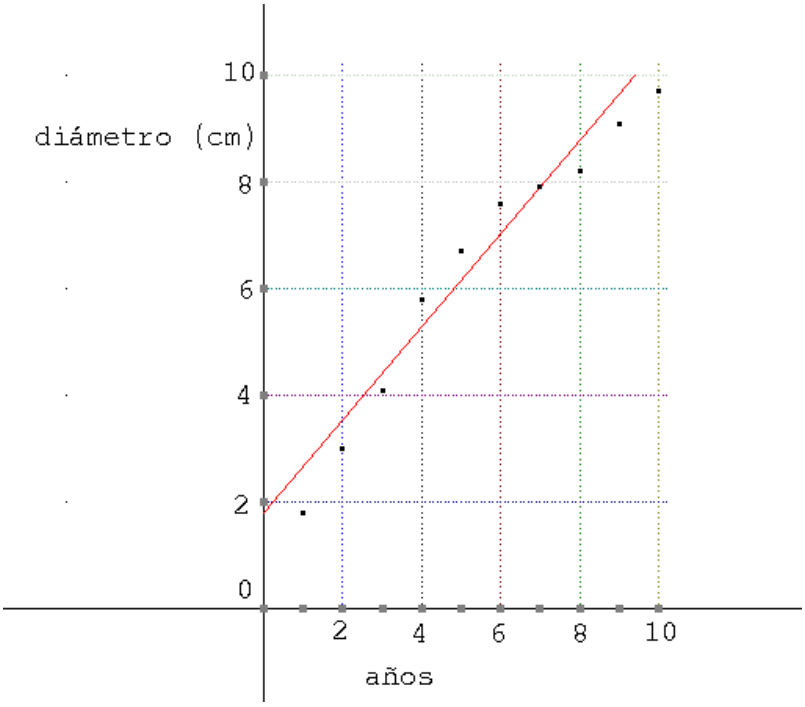


Figura 7.8.4. Representación y recta de regresión de la variable diámetro en función de la variable año, del Ejercicio 7.8.

b) La suma de los cuadrados de los errores cometidos es

$$\begin{aligned} SC(\widehat{\beta}_0, \widehat{\beta}_1) &= \sum_{i=1}^{n} (y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i)^2 \\ &= n b_y^2 (1 - r^2) \end{aligned}$$

con

$$b_y^2 = \frac{1}{n} \sum_{i=1}^{n} y_i^2 - \overline{y}^2 = 47'862 - 41'474 = 6'388$$

$$r = \frac{b_{xy}}{b_x b_y} = \frac{7'07}{2'872 \cdot 2'527} = 0'974$$

(véanse las Expresiones (7.4.1.4) y (7.4.1.5)). Por lo que

$$SC(\widehat{\beta}_0, \widehat{\beta}_1) = 3'296$$

c) Para contrastar si las dos rectas son compatibles, se utiliza que

$$\frac{\widehat{\beta}_i - \beta_i}{\sqrt{\widehat{V}[\widehat{\beta}_i]}} \sim T_{n-2}$$

(véase el comentario a continuación del Teorema 7.4.1.1), por lo que para contrastar $H_0 : \beta_0 = 0'8$ frente a $H_1 : \beta_0 \neq 0'8$, la región crítica es

$$RC = \left\{ \frac{|\widehat{\beta}_0 - 0'8|}{\sqrt{\widehat{V}[\widehat{\beta}_0]}} \geq t_{n-2;\frac{\alpha}{2}} \right\}$$

en este caso con $\alpha = 0'05$ y $t_{8;0'025} = 2'306$ como

$$V[\widehat{\beta}_0] = \frac{\sigma^2 \frac{1}{n} \sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}$$

si se sustituye σ^2 por su estimador insesgado. $SC(\widehat{\beta}_0, \widehat{\beta}_1)/(n-2) = 0'412$ se tiene que $\widehat{V}[\widehat{\beta}_0] = 0'412 \cdot 0'466 = 0'192$, por lo que el valor del estadístico es $\left| \frac{1'727 - 0'8}{0'438} \right| = 2'116$, luego se acepta la hipótesis nula.

Para contrastar ahora si son iguales las pendientes de las rectas, se contrasta $H_0 : \beta_1 = 0'85$ frente a $H_1 : \beta_1 \neq 0'85$, la región crítica es

$$RC = \left\{ \frac{\left|\widehat{\beta}_1 - 0'85\right|}{\sqrt{\widehat{V}[\widehat{\beta}_1]}} \geq t_{n-2;\frac{\alpha}{2}} = 2'306 \right\}$$

como

$$\widehat{V}[\widehat{\beta}_1] = \widehat{\sigma}^2 \frac{1}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2} = 0'412 \cdot 0'012 = 0'005$$

y el valor del estadístico es $\left| \frac{0'857 - 0'85}{0'0707} \right| = 1'315$, se acepta también la hipótesis nula.

7.8.9

a) La representación gráfica de la Figura 7.8.5, permite utilizar la aproximación lineal de la presión en función de la temperatura. Se ha incluido también la recta de regresión de y sobre x.

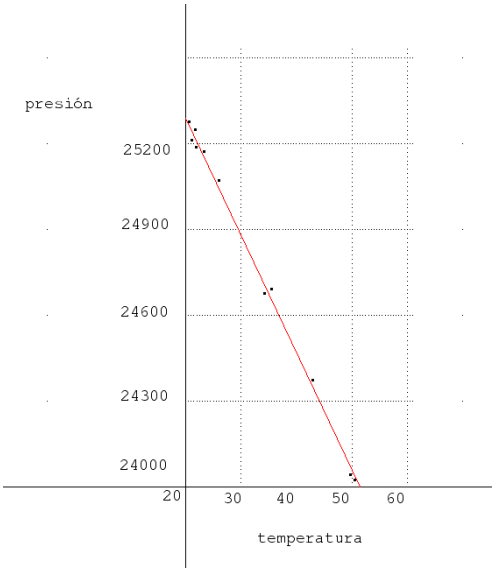


Figura 7.8.5. Representación y recta de regresión de presión sobre temperatura para el Ejercicio 7.8.9.

Mediante un paquete estándar, el programa de regresión permite obtener

Coeficientes	Estimadores	Errores estandar	Estadísticos T
β_0	26082	14'84	1757'55
β_1	-40'183	0'435	-92'381

b) Para el contraste de la pendiente de la recta de regresión como nula; $H_0 : \beta_1 = 0$ frente a $H_1 : \beta_1 \neq 0$ se emplea que

$$\frac{\widehat{\beta}_1 - \beta_1}{\sqrt{\widehat{V}[\widehat{\beta}_1]}} \sim T_{n-2}$$

(véase la expresión (7.4.1.7)), de donde la región crítica es

$$RC = \left\{ \left| \widehat{\beta}_1 \right| \geq \sqrt{\widehat{V}[\widehat{\beta}_1]} t_{n-1;\frac{\alpha}{2}} \right\}$$

sustituyendo con $\alpha = 0'05$, el error estandar por $0'435$ y $t_{8;0'025} = 2'306$ queda como región crítica

$$RC = \left\{ \left| \widehat{\beta}_1 \right| \geq 1'003 \right\}$$

y como el valor de estadístico es $40'183$, se rechaza la hipótesis nula.

c) La representación (temperatura, residuo), donde el residuo es $y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i$, de la Figura 7.8.6 no pone de manifiesto ninguna anomalía

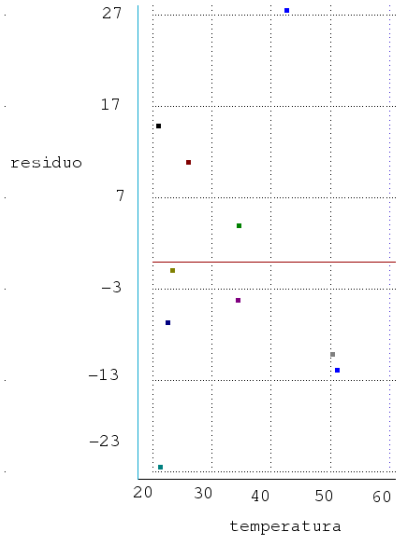


Figura 7.8.6. Residuos en función de la temperatura para el Ejercicio 7.8.9.

7.8.10

a) De acuerdo con el modelo de regresión lineal

$$Y = \beta_0 + \beta_1 x + e$$

por lo tanto, si se estima el valor correspondiente a x_0 , parece razonable utilizar

$$\widehat{Y}(x_0) = \widehat{\beta}_0 + \widehat{\beta}_1 x_0$$

Se sigue que,

$$E[\widehat{Y}(x_0)] = \beta_0 + \beta_1 x_0$$
$$V[\widehat{Y}(x_0)] = V[\widehat{\beta}_0] + x_0^2 V[\widehat{\beta}_1] + 2x_0 Cov(\widehat{\beta}_0, \widehat{\beta}_1)$$

(véase el Teorema 7.4.1.1), de donde

$$\begin{aligned} V[\widehat{Y}(x_0)] &= \frac{\sigma^2 \frac{1}{n}\sum_{j=1}^{n} x_j^2}{\sum_{j=1}^{n} x_i^2 - n\overline{x}^2} + x_0^2 \frac{\sigma^2}{\sum_{j=1}^{n} x_j^2 - n\overline{x}^2} - 2x_0 \frac{\sigma^2 \overline{x}}{\sum_{j=1}^{n} x_j^2 - n\overline{x}^2} \\ &= \frac{\sigma^2}{nb_x^2}\left[\frac{1}{n}\sum_{j=1}^{n} x_j^2 + x_0^2 - 2x_0\overline{x}\right] \\ &= \frac{\sigma^2}{nb_x^2}\left[\frac{1}{n}\sum_{j=1}^{n} x_j^2 + (x_0 - \overline{x})^2 - \overline{x}\right] \\ &= \frac{\sigma^2}{nb_x^2}\left[b_x^2 + (x_0 - \overline{x})\right]. \end{aligned}$$

Además, como $\widehat{\beta}_0 + \widehat{\beta}_1 x_0$ es combinación lineal de las v.a. $Y_1, ..., Y_n$ que son normales, se sigue que

$$\widehat{\beta}_0 + \widehat{\beta}_1 x_0 \sim N\left(\beta_0 + \beta_1 x_0,\ \sigma\sqrt{\frac{1}{n}\left[1 + \frac{(x_0 - \overline{x})^2}{b_x^2}\right]}\right).$$

Como σ es desconocida, se puede utilizar que $(n-2)\frac{s^2}{\sigma^2} \sim \chi^2_{n-2}$ con lo que si se demuestra que $(\widehat{\beta}_0, \widehat{\beta}_1)$ es independiente de S^2 se puede utilizar la definición de la distribución de Student, (véase la Definición 1.5.3.1) para tener demostrada la parte a) del enunciado.

Para ver la independencia de S^2, es suficiente ver que $(\widehat{\beta}_0, \widehat{\beta}_1)$ es independiente de cada residuo $\widehat{\varepsilon}_i = Y_i - \widehat{Y}_i \quad i = 1, ..., n$. Pero se puede escribir

$$\begin{aligned} \widehat{\varepsilon}_i &= Y_i - \overline{Y} - \frac{b_{xy}}{b_x^2}(x_i - \overline{x}) \\ &= \sum_{j=1}^{n}\left[\delta_{ij} - \frac{1}{n} - \frac{(x_i - \overline{x})x_j}{nb_x^2} + \frac{(x_i - \overline{x})\overline{x}}{nb_x^2}\right] Y_j \end{aligned}$$

donde

$$\delta_{ij} = \begin{cases} 1 & \text{si} \quad j = i \\ 0 & \text{si} \quad j \neq i \end{cases}$$

por lo que alternativamente puede ponerse

$$\widehat{\varepsilon}_i = \sum_{j=1}^{n} \left[\delta_{ij} - (c_i + d_i x_j)\right] Y_j$$

con

$$c_i = \frac{1}{n} - \frac{\overline{x}(x_j - \overline{x})}{nb_x^2} \qquad \text{y} \qquad d_i = \frac{x_i - \overline{x}}{nb_x^2}.$$

Por otro lado,

$$\widehat{\beta}_0 = \sum_{j=1}^{n}\left[\frac{1}{n} - \frac{\overline{x}(x_j-\overline{x})}{nb_x^2}\right] \quad \widehat{\beta}_1 = \sum_{j=1}^{n}\left[\frac{1}{n} - \frac{(x_j-\overline{x})}{nb_x^2}Y_j\right]$$
$$= \sum_{j=1}^{n} c_j Y_j \qquad\qquad = \sum_{j=1}^{n} d_j Y_j$$

Es sencillo ver que

$$Cov(\widehat{\varepsilon}_i, \widehat{\beta}_0) = Cov(\widehat{\varepsilon}_i, \widehat{\beta}_1) = 0 \quad i = 1, ..., n$$

sin más que verificar las igualdades

$$\sum_{j=1}^{n} [\delta_{ij} - (c_i + d_i x_j)]\, c_j = \sum_{j=1}^{n} [\delta_{ij} - (c_i + d_i x_j)]\, d_j = 0.$$

Así para la primera basta comprobar que

$$\sum_{j=1}^{n} c_j = 1 \qquad \text{y} \qquad \sum_{j=1}^{n} c_j x_j = 0$$

y para la segunda, basta comprobar que

$$\sum_{j=1}^{n} d_j = 0 \qquad \text{y} \qquad \sum_{j=1}^{n} d_j x_j = 1.$$

b) Mediante el apartado anterior

$$P\left\{\left|\frac{\widehat{\beta}_0 + \widehat{\beta}_1 x_0 - (\beta_0 + \beta_1 x_0)}{s\sqrt{\frac{1}{n}\left[1 + \frac{(x_0-\overline{x})^2}{b_x^2}\right]}}\right| \leq t_{n-2;\frac{\alpha}{2}}\right\} = 1 - \alpha$$

si se denota por

$$s_0 = s\sqrt{\frac{1}{n}\left[1 + \frac{(x_0-\overline{x})^2}{b_x^2}\right]}$$

la expresión se puede poner en la forma

$$P\{-s_0 t_{n-2;\frac{\alpha}{2}} \leq \widehat{\beta}_0 + \widehat{\beta}_1 x_0 - (\beta_0 + \beta_1 x_0) \leq s_0 t_{n-2;\frac{\alpha}{2}}\} = 1 - \alpha$$

si se deja en el centro de la cadena de desigualdades $\beta_0 + \beta_1 x_0$, haciendo operaciones es inmediato comprobar que un intervalo de grado de confianza $1 - \alpha$ viene dado por

$$IC_{1-\alpha}(\beta_0 + \beta_1 x_0) = (\widehat{\beta}_0 + \widehat{\beta}_1 x_0 - t_{n-2;\frac{\alpha}{2}} s_0, \widehat{\beta}_0 + \widehat{\beta}_1 x_0 + t_{n-2;\frac{\alpha}{2}} s_0)$$

es interesante señalar que la amplitud del intervalo varía con el valor de x_0 y es mínima para $x_0 = \overline{x}$.

c) Por el apartado anterior, se debe calcular

$$\widehat{\beta}_0 = 0'896, \quad \widehat{\beta}_1 = 0'932, \quad t_{8;0'025} = 2'306, \quad s = 0'027, \quad \overline{x} = 0'859$$

$$b_x^2 = \frac{1}{n}\sum_{j=1}^{n} x_j^2 - \overline{x}^2 = 0'743 - -0'738 = 0'00478$$

$$s_0 = s\sqrt{\frac{1}{n}\left[1 + \frac{(x_0 - \overline{x})^2}{b_x^2}\right]} = 0'0157$$

de donde

$$IC_{0'95}(\beta_0 + \beta_1 0'95) = (1'744, 1'819).$$

En la Figura 7.8.7 se incluyen la recta de regresión y los residuos.

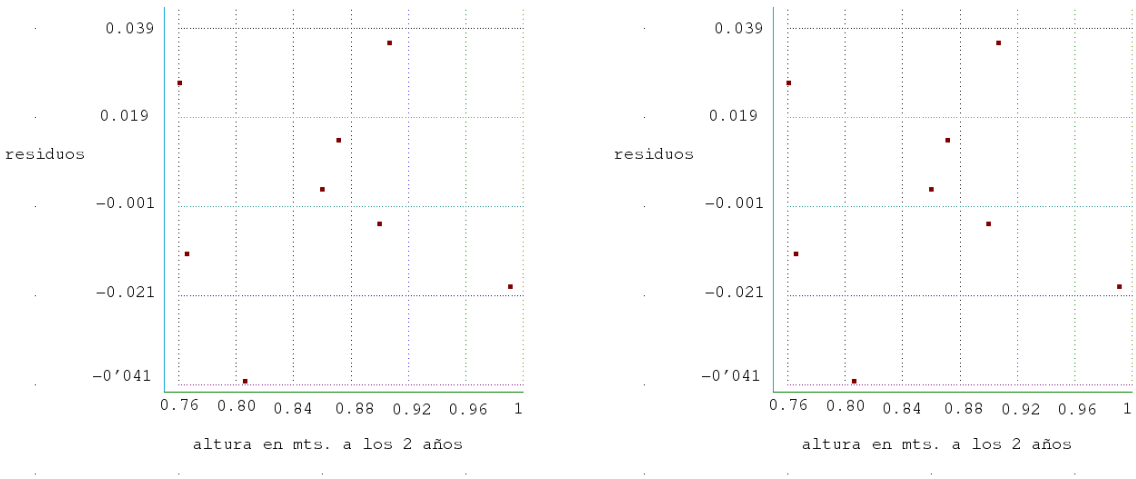


Figura 7.8.7. Recta de regresión de y sobre x, con región de grado de confianza 0'95 *y representación de los residuos en función de x, para el Ejercicio 7.8.10.*

7.8.12

La representación de la Figura 7.8.8, de la variable *tanto por ciento de unidades defectuosas* en función de la variable *unidades producidas al mes*, sugiere la utilización de una aproximación exponencial de la forma

$$y = ab^x$$

o bien

$$y = \beta_0 + \beta_1 x + \beta_2 x^2$$

como dice el enunciado.

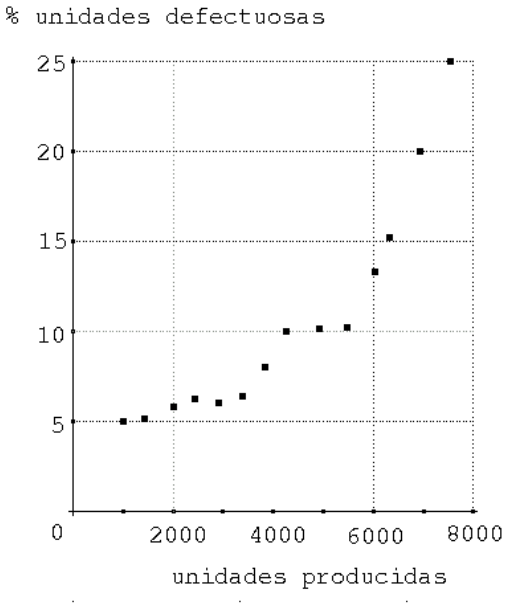


Figura 7.8.8. Representación de los pares de puntos (producción, % defectuosos) para el Ejercicio 7.8.11.

Se observa que tomando logaritmos en la expresión de la función exponencial

$$\ln y = \ln a + (\ln b)x$$

las variables x y $\ln y$ están relacionadas linealmente.

El cuadro de datos para la realización de los cálculos queda recogido a continuación

Producción x	% y	$\ln y$
1.000	5	1'609
1.500	5'2	1'649
2.000	5'7	1'740
2.500	6'5	1'872
3.000	6'8	1'917
3.500	7'1	1'96
4.000	8'1	2'092
4.500	10	2'303
5.000	10'2	2'322
5.550	10'3	2'332
6.000	13	2'565
6.500	15'3	2'728
7.000	20	2'996
7.500	25	3'219

Con el ajuste lineal se obtiene $\ln a = 1'249 \quad \ln b = 2'333 \cdot 10^{-4}$, por lo que el modelo exponencial viene dado por

$$y = 3'487 \cdot 1'0002^x$$

y le corresponde un *coeficiente de determinación* de $R^2 = 0'951$.

El ajuste mediante el polinomio de segundo grado es

$$y = 8'072 - 0'00253x + 5'989 \cdot 10^{-7}$$

y le corresponde un *coeficiente de determinación* de $R^2 = 0'958$.

Es, por lo tanto, ligeramente mejor el ajuste mediante el polinomio de segundo grado.

7.8.13

a) La representación de los puntos (gastos en alojamiento, gastos en esparcimiento) de la Figura 7.8.9 permite ajustar una curva con una asíntota horizontal sobre la ordenada 15.

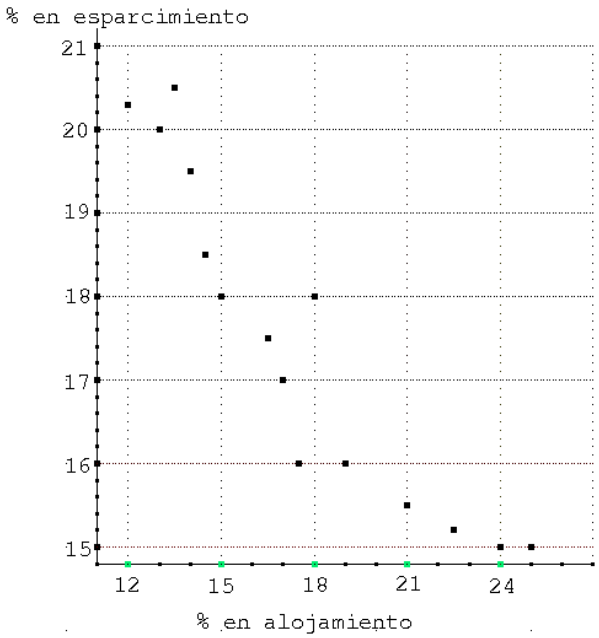


Figura 7.8.9. Nube de puntos (% de gastos en alojamiento, % de gastos en esparcimiento) correspondientes al Ejercicio 7.8.12.

Por lo tanto parece razonable ajustar un modelo de la forma

$$y = a + b/x$$

si se introduce la nueva variable auxiliar $u = 1/x$, se puede realizar un ajuste lineal a la nube de puntos (u, y), con lo que se obtienen los parámetros

$$a = 8'947 \qquad b = 142'042$$

Como medida del ajuste el *coeficiente de correlación lineal* vale $r = 0'758$ o bien interpretado éste como caso particular del *coeficiente de determinación* $R^2 = 0'915$.

b) Si se utiliza un ajuste lineal en las variables x= *% de gastos en alojamiento* e y = *% de gastos en entretenimiento*

$$y = a' + b'x$$

se llega a los valores

$$a' = 25'3473 \qquad b' = -0'450$$

obteniéndose que $r = -0'927$ y $R^2 = 0'859$, lo que quiere decir que el ajuste es ligeramente peor en este segundo caso.

7.8.14

a) Vistos los datos del Ejercicio 7.8.11, parece que el modelo potencial no va a dar un buen ajuste, ya que la representación gráfica de la función potencial es la recogida en la Figura 7.8.10 que pasa por el origen.

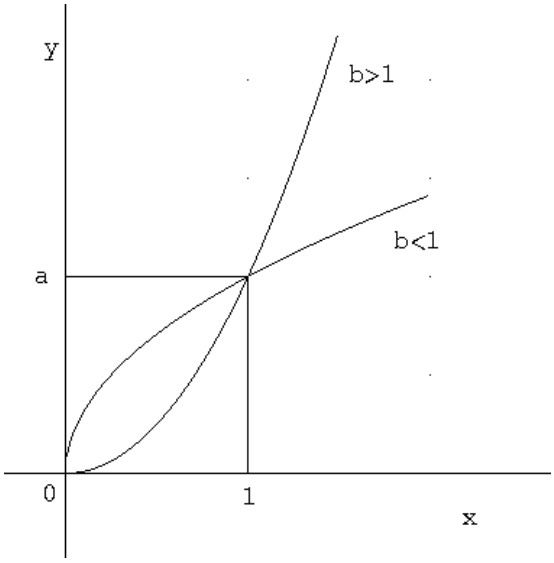


Figura 7.8.10. Representación de la función potencial $y = ax^b$ *en función de los signos de los parámetros* a y b.

Se observa que al tomar logaritmos en la función potencial se obtiene

$$\ln y = \ln a + b \ln x$$

por lo que se puede realizar un ajuste lineal en las nuevas variables $(\ln x, \ln y)$. Las variables auxiliares para realizar los cálculos vienen recogidas a continuación

$\ln x$	$\ln y$
6'9077	1'6094
7'3132	1'6487
7'6009	1'7405
7'8240	1'8718
8'0064	1'9169
8'1605	1'9601
8'2940	2'0919
8'4118	2'3026
8'5172	2'3224
8'6125	2'3321
8'6995	2'5649
8'7796	2'7279
8'8537	2'9957
8'9227	3'2189

obteniéndose los valores

$$\ln a = -3'7845 \qquad b = 0'7335$$

de donde el modelo potencial ajustado es

$$y = 0'0227x^{0'7335}$$

b) Para comparar ambos modelos el potencial con el exponencial se utiliza el *coeficiente de determinación*, que para el caso del ajuste del modelo potencial vale $0'809$ que es ligeramente inferior, como era de esperar, al que se obtuvo con el ajuste exponencial que era $0'951$.

Por consiguiente es mejor el modelo potencial.

7.8.15

a) Con el fin de estudiar las posibles influencias entre las variables, en la Figura 7.8.11 se recogen las distintas representaciones.

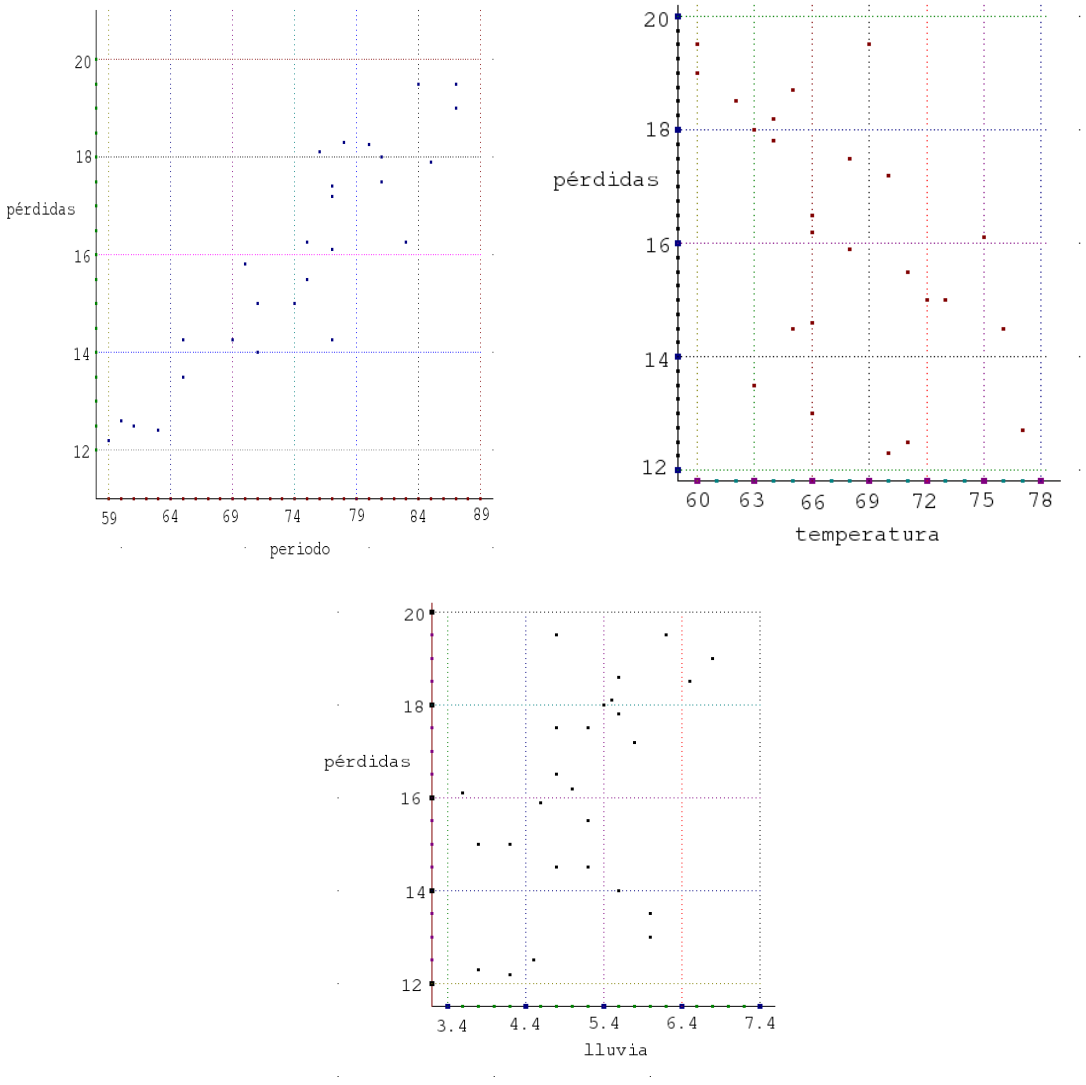


Figura 7.8.11. Representación de las variables (periodo, pérdidas), (temperaturas, pérdidas) y (lluvia, pérdidas) correspondientes al Ejercicio 7.8.14.

b) El ajuste lineal de las pérdidas en función de la lluvia, periodo y temperatura viene dado mediante

$$\textit{pérdidas} = -3'9349 + 0'3326\ \ \textit{lluvia} + 0'2346\ \textit{periodo} + 0'0126\ \textit{temperatura}$$

a este ajuste le corresponde un *coeficiente de determinación* $R^2 = 0'8054$.

c) Mientras que el ajuste lineal de las pérdidas en función única-

mente de la lluvia y de la temperatura, viene dado por

$$\textit{pérdidas} = 32'1847 + 0'0685\ \textit{lluvia} - 0'244\ \textit{temperatura}$$

con un *coeficiente de determinación* $R^2 = 0'2125$. Por lo tanto este segundo ajuste es peor.

7.8.16

Se va a calcular $\Sigma_{X-\overline{X},\overline{X}}$, donde $X-\overline{X}$ es un vector de n filas y una columna y $\overline{X}$ es una v.a., mediante la utilización del Teorema 7.4.2.2

$$X - \overline{X} = X - \mathbf{1}\mathbf{1}'X = (I - \frac{1}{n}\mathbf{1}\mathbf{1}')X$$

donde $\mathbf{1}$ es un vector columna de n valores iguales a uno.

Como

$$\overline{X} = \frac{1}{n}\mathbf{1}'X$$

se sigue que

$$\begin{aligned}\Sigma_{X-\overline{X},\overline{X}} &= (I - \frac{1}{n}\mathbf{1}\mathbf{1}')\Sigma_X(\frac{1}{n}\mathbf{1}')' \\ &= (I - \frac{1}{n}\mathbf{1}\mathbf{1}')\sigma^2 I\frac{1}{n}\mathbf{1} \\ &= \sigma^2\frac{1}{n}\mathbf{1} - \frac{1}{n}\mathbf{1}\mathbf{1}'\sigma^2 I\frac{1}{n}\mathbf{1} \\ &= \frac{\sigma^2}{n}\mathbf{1} - \frac{\sigma^2}{n^2}n\mathbf{1}\end{aligned}$$

que es un vector columna de n valores iguales a cero. Luego cada diferencia $X_i - \overline{X}$ está incorrelada con $\overline{X}$.

7.8.17

Basta poner

$$Y = (y_{11}, ..., y_{1n_1}, ..., y_{k1}, ..., y_{kn_k})'$$

$$X = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{1}_{n_k} \end{pmatrix}$$

con $\mathbf{1}_{n_i}$ un vector columna de n_i elementos iguales a la unidad

$$\theta = (\theta_1, ..., \theta_1, ..., \theta_k, ..., \theta_k)'$$

Por último una posibilidad es considerar para H la matriz

$$H = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ 1 & 0 & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & -1 \end{pmatrix}$$

de dimensión $(k-1)\times k$. Con lo que $H_0 : \theta_0 = ... = \theta_k$ puede escribirse como

$$H\theta = \mathbf{0}.$$

7.8.18

Mediante el cambio $y = x^{-1}$ en la integral, se obtiene

$$\int_0^{+\infty} x^{-(k+1)} e^{-ax^{-1}} dx = \int_0^{+\infty} y^{k-1} e^{-ay} dy$$

y como la función gamma cumple

$$\int_0^{+\infty} e^{-ax} x^{p-1} dx = a^{-p}\Gamma(p)$$

se sigue que la integral pedida es igual a $a^{-k}\Gamma(k)$.

7.8.19

a) Matricialmente, el modelo es de la forma

$$y = X\beta + e$$

donde

$$y = (109, 85, 114, ..., 133, 217, ..., 221)'$$

$$x = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}'$$

$$X'X = \begin{pmatrix} 9 & 7 \\ 7 & 16 \end{pmatrix} \quad (X'X)^{-1} = \begin{pmatrix} 0'1684 & -0'0737 \\ -0'0737 & 0'0947 \end{pmatrix} \quad X'y = \begin{pmatrix} 1761 \\ 2683 \end{pmatrix}$$

$$\widehat{\beta} = (X'X)^{-1}X'y = \begin{pmatrix} 98'895 \\ 124'421 \end{pmatrix} \qquad s^2 = \frac{1}{18-2}(y-\widehat{y})'(y-\widehat{y}) = \frac{2{,}525}{16}.$$

Si se toma como distribución inicial

$$\pi(\beta_1, \beta_2, \sigma^2) = \sigma^{-2}$$

entonces

$$\beta_1, \beta_2 | \overrightarrow{y} \sim T(2, \widehat{\beta}, 157'812(X'X)^{-1}, 16)$$

(véase la Expresión (7.5.12)).

b) Para contrastar $H_0 : \overrightarrow{\beta} = (100, 125)$ frente a $H_1 : \overrightarrow{\beta} \neq (100, 125)$ se emplea que

$$\frac{(\beta - \widehat{\beta})'X'X(\beta - \widehat{\beta})}{2S^2} \sim F_{2,n-2}$$

(véase el Teorema 7.5.1). Es decir, se determina una región de probabilidad final calculada respecto a $\pi(\beta_1, \beta_2|\overrightarrow{y})$ igual a $1-\alpha$, y se rechaza si el punto no pertenece a la región. En analogía a la notación clásica, se toma como

$$RC = \left\{ \overrightarrow{\beta} \left| \frac{SC_T(\overrightarrow{\beta})}{2S^2} > f_{2,n-2;\alpha} \right. \right\}$$

donde

$$SC_T(\overrightarrow{\beta}) = (\beta - \widehat{\beta})' X' X (\beta - \widehat{\beta})$$
$$S^2 = \frac{1}{n-p} \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2$$

En este caso es $f_{2,16;0'05} = 3'63$ y como

$$SC_T(\overrightarrow{\beta}) = (100 - 98'895, 125 - 124'421) \begin{pmatrix} 9 & 7 \\ 7 & 16 \end{pmatrix} \begin{pmatrix} 100 - 98'895 \\ 125 - 124'421 \end{pmatrix}$$
$$= 25'3102$$
$$S^2 = \frac{1}{16} 0'158$$

por lo tanto $\frac{SC_T(\overrightarrow{\beta}_0)}{2S^2} = 80'1971 > 3'63$ luego se rechaza H_0.

Obsérvese que se debería de haber usado el contraste de hipótesis nula puntual bayesiana y utilizado una distribución inicial de tipo mixto (véase la Sección 5.2.4), sin embargo por facilidad de cálculo, se suele preferir determinar una *región creíble* y ver si $\overrightarrow{\beta}_0$ pertenece a ella.

7.8.20

a) Matricialmente el modelo puede escribirse

$$y = X\beta + e$$

donde

$$y = (63, 48, ..., 47, ..., 77)'$$
$$X = \begin{pmatrix} \mathbf{1}_6 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_6 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{1}_6 \end{pmatrix}$$

con $\mathbf{1}_6$ un vector columna de unos y $\mathbf{0}$ un vector columna de seis ceros

$$X'X = \begin{pmatrix} 6 & 0 & 0 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 6 \end{pmatrix} \qquad (X'X)^{-1} = \begin{pmatrix} 1/6 & 0 & 0 & 0 & 0 \\ 0 & 1/6 & 0 & 0 & 0 \\ 0 & 0 & 1/6 & 0 & 0 \\ 0 & 0 & 0 & 1/6 & 0 \\ 0 & 0 & 0 & 0 & 1/6 \end{pmatrix}$$

$$X'Y = \begin{pmatrix} 306 \\ 132 \\ 165 \\ 221 \\ 198 \end{pmatrix} \qquad \widehat{\beta} = (X'X)^{-1}X'Y = \begin{pmatrix} 51 \\ 22 \\ 27'5 \\ 36'833 \\ 66 \end{pmatrix}$$

$$S^2 = \frac{1}{30-5}(y-\widehat{y})'(y-\widehat{y}) = \frac{2567'17}{25}$$

para la distribución inicial

$$\pi(\beta_1, \beta_2, ..., \beta_5, \sigma^2) = \sigma^{-2}$$

la distribución final es

$$\beta_1, ..., \beta_5 | y \sim T(5, \widehat{\beta}, 102'6868 I_5, 25)$$

(véase la Expresión (7.5.12)).

b) Para contrastar $H_0 : \beta_2 = 30, \beta_5 = 30$ frente a $H_1 : \beta_2 \neq 30, \beta_5 \neq 30$ se calcula la distribución marginal $\beta_2, \beta_5 | y \sim T(2, (\widehat{\beta}_2, \widehat{\beta}_5), S^2 c_{11}^{-1}, p = n-2)$ donde

$$c_{11} = \begin{pmatrix} c_{22} & c_{25} \\ c_{52} & c_{55} \end{pmatrix} = \begin{pmatrix} 1/6 & 0 \\ 0 & 1/6 \end{pmatrix}.$$

La región crítica es de la forma

$$RC = \left\{ \beta \text{ tales que } \frac{(\beta - \widehat{\beta})' C_{11}^{-1} (\beta - \widehat{\beta})}{2 \cdot 5^2} \geq f_{k,\nu;\alpha} \right\}$$

en este caso $f_{2,25;0'05} = 3'39$ (véase la Tabla 6). Todo se reduce a comprobar si $\overrightarrow{\beta}_0 = (\beta_2 = 30, \beta_5 = 30)$ pertenece o no a la región crítica

$$(30 - 36'1667, 30 - 33) \begin{pmatrix} 6 & 0 \\ 0 & 6 \end{pmatrix} \begin{pmatrix} 30 - 26'1667 \\ 3' - 23 \end{pmatrix} = 382'1743$$

que dividido por $2S^2 = 205'3736$ hace que el estadístico sea $1'8609$, que no pertenece a la región crítica, luego se acepta la hipótesis nula y los datos son compatibles con $\beta_0 = 30 \;\; \beta_5 = 30$.

Ha de tenerse en cuenta el comentario realizado al final del Ejercicio 7.8.18, se ha empleado una *región creíble* en lugar de realizar el contraste de hipótesis nula puntual (véase la Sección 5.2.4).

Capítulo 8

ESTADÍSTICA NO PARAMÉTRICA

8.1. INTRODUCCIÓN

Al aplicar los métodos estadísticos ocurre con frecuencia que las suposiciones que parecían razonables no se cumplen, así la muestra puede ser demasiado pequeña, la suposición de normalidad poco convincente, etc. En estas situaciones es conveniente disponer de procedimientos que sean válidos en contextos suficientemente amplios.

Existen un conjunto de técnicas estadísticas, a cuyo estudio se dedica este capítulo, que reciben el nombre genérico de *procedimientos no paramétricos* y que se basan en funciones de las observaciones muestrales, que no dependen de la distribución específica de la población de la cuál la muestra ha sido extraida, sino de una propiedad de la distribución, como la continuidad o la simetría.

De manera estricta no se va a poder decir que exista un parámetro θ que caracterice a la función de distribución de la población, como se ha venido suponiendo hasta ahora. Se dice de manera estricta porque si se consideran, por ejemplo, los tests de *bondad de ajuste*, la no existencia de parámetros no es del todo correcta, de hecho se va a trabajar con la distribución multinomial, pero es cierto que los procedimientos que se emplean están basados en características de la distribución poblacional que no son hipótesis sobre el valor de sus parámetros.

En esta línea se comenzará contrastando la adecuación entre la muestra y el modelo, en lo que se conoce como *problema de la bondad del ajuste*, para a continuación contrastar si varias muestras provienen de una misma población en el *problema de homogeneidad*. Se estudiará el problema de los *valores discordantes* y la *estimación no paramétrica de densidades*, para acabar tratando la independencia entre dos variables aleatorias.

8.2. AJUSTES RELATIVOS A LA DISTRIBUCIÓN MULTINOMIAL

Este grupo de problemas se basan en trasladar la hipótesis no paramétrica a la distribución multinomial y utilizar el estadístico de la razón de verosimilitudes, cuya distribución asintótica fué obtenida por Wilks (1938), o el estadístico de la distancia al cuadrado introducido por K. Pearson (1900) que compara las frecuencias observadas experimentalmente con las frecuencias esperadas bajo la hipótesis nula. Una aproximación bayesiana a estos problemas puede verse en Gómez Villegas (1988).

8.2.1. Bondad del ajuste

Se trata de contrastar si la m.a.s. $(x_1, ..., x_n)$ se distribuye de acuerdo con una función de distribución determinada F. Se van a distinguir dos casos, según que F sea perfectamente conocida, por ejemplo viene dada por la normal de media 0 y de desviación típica 1, o bien que F sea conocida a menos de k parámetros, por ejemplo el caso más realista en que F venga dada por la normal de media μ y de desviación típica σ, en esta situación F es conocida a menos de 2 parámetros.

En el primer caso, se divide el recorrido de la población X en k intervalos disjuntos, en principio de forma arbitraria, $I_1, ..., I_k$ con lo que es posible calcular bajo F las probabilidades $p_i^0 = P\{x \in I_i\}$ $i = 1, ..., k$.

Si la muestra es compatible con F se puede suponer la hipótesis nula H_0 : consistente en que el vector $(n_1, ..., n_k)$, con n_i número de veces que en la muestra aparece un elemento del intervalo I_i, tiene una distribución $(N_1 = n_1, ..., N_k = n_k) \sim Multinomial(n, p_1^0, ..., p_k^0)$, frente a la hipótesis alternativa $H_1 : (N_1 = n_1, ..., N_k = n_k) \sim Multinomial(n, p_1, ..., p_k)$. La única diferencia está en que bajo la hipótesis nula los parámetros de la distribución multinomial son $p^0 = (p_1^0, ..., p_k^0)$ conocidos, mientras que bajo la hipótesis alternativa son los parámetros $p = (p_1, ..., p_k)$ desconocidos.

Si se emplea el contraste de la razón de verosimilitudes, de acuerdo con la Sección 5.2.1, se tiene que

$$\lambda(n_1, ..., n_k) = \frac{\frac{n!}{n_1!...n_k!}(p_1^0)^{n_1}...(p_k^0)^{n_k}}{\sup\limits_{p_1,...,p_k} \frac{n!}{n_1!...n_k!}p_1^{n_1}...p_k^{n_k}}$$

El supremo del denominador, usando que $\sum\limits_{i=1}^{k} p_i = 1$, se alcanza en $\widehat{p}_i = \frac{n_i}{n}$. Por tanto la región crítica será de la forma

$$RC = \left\{ \lambda(n_1, ..., n_k) = \prod_{i=1}^{k} \left(\frac{p_i^0}{\widehat{p}_i} \right)^{n_i} \leq k \right\}.$$

Bajo H_0 la distribución asintótica de $-2\ln\lambda(n_1,...,n_k)$ es una χ^2 con $k-1$ grados de libertad, ya que, $H_0 \cup H_1$ tiene de dimensión $k-1$ y la hipótesis nula tiene de dimensión 0, (véase el Teorema 5.2.3.2). Para un tamaño α, se obtiene la región crítica

$$RC = \left\{ -2\sum_{i=1}^{k} n_i \left[\ln p_i^0 - \ln \widehat{p}_i\right] \geq \chi^2_{k-1;\alpha} \right\}$$

K. Pearson introdujo un procedimiento alternativo de resolver el problema, mediante una distancia ponderada entre el EMV del vector multinomial $\widehat{p}$ y el valor p^0 supuesto bajo la hipótesis nula, en concreto la cantidad

$$D^2(p^0, \widehat{p}) = \sum_{i=1}^{k} \lambda_i (\widehat{p}_i - p_i^0)^2$$

con

$$\lambda_i = \frac{n}{p_i^0}$$

y rechazar la hipótesis nula para valores grandes de la distancia; por lo que la región crítica es

$$RC = \left\{ \sum_{i=1}^{k} \frac{(n_i - np_i^0)^2}{np_i^0} \geq c \right\}$$

Se observa, que las n_i son las frecuencias observadas de los intervalos I_i en la muestra, mientras que np_i^0 son las frecuencias esperadas de los mismos bajo la hipótesis nula, por lo que se acostumbra a notar alternativamente por

$$D^2(p_0, \widehat{p}) = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{8.2.2.1}$$

Con el fin de poder determinar la constante c para que el contraste tenga de tamaño α, es necesario conocer la distribución de la cantidad en la Expresión (8.2.2.1) bajo la hipótesis nula, resultando que asintóticamente tiene la misma distribución que $-2\ln\lambda(n_1,...,n_k)$. Una justificación de este hecho radica en observar que

$$-2\ln\lambda(n_1,...,n_k) = -2\sum_{i=1}^{k} n_i \ln \frac{p_i^0}{\widehat{p}_i}$$

y si se emplea que, aproximadamente, $\ln x \simeq (x-1) - \frac{1}{2}(x-1)^2$, se tiene que

$$\begin{aligned} -2\ln\lambda(n_1,...,n_k) &\simeq -2\sum_{i=1}^{k} n_i \left[\left(\frac{p_i^0}{\widehat{p}_i} - 1 \right) - \frac{1}{2}\left(\frac{p_i^0}{\widehat{p}_i} - 1 \right)^2 \right] = \\ &= \sum_{i=1}^{k} n_i \left(\frac{p_i^0}{\widehat{p}_i} - 1 \right)^2 = \sum_{i=1}^{k} \frac{(n_i - np_i^0)^2}{n_i} \simeq \sum_{i=1}^{k} \frac{(n_i - np_i^0)^2}{np_i^0} \end{aligned}$$

donde la última aproximación se debe a que $n_i \simeq np_i^0$.

Una demostración del resultado que se ha justificado, es el contenido del siguiente enunciado.

Teorema 8.2.1.1
Si $(N_1, ..., N_k) \sim Multinomial(n, p_1, ..., p_k)$ entonces

$$D^2 = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i} \xrightarrow{\mathcal{L}} \chi^2_{k-1}.$$

Demostración: El vector $(N_1, ..., N_{k-1})$ tiene de media $(np_1, ..., np_{k-1})$ y de matriz de covarianzas

$$\sigma_{ij} = \begin{cases} -np_ip_j & \text{si} \quad i \neq j \\ np_i(1-p_i) & \text{si} \quad i = j \end{cases}$$

o bien

$$\sigma_{ij} = (p_i\delta_{ij} - p_ip_j)n$$

con lo que $Y = \left(\frac{N_1 - np_1}{\sqrt{n}}, ..., \frac{N_{k-1} - np_{k-1}}{\sqrt{n}}\right)$ converge en ley a una v.a. $N_{k-1}(0, \Sigma)$ con $\Sigma = (\sigma_{ij}) = (p_i\delta_{ij} - p_ip_j)$ por lo que se tiene, (véase Rao (1973), pág. 384, Expresión (6.1.1.6)) que

$$Y'\Sigma^{-1}Y \xrightarrow{\mathcal{L}} \chi^2_{k-1}$$

y como $\Sigma^{-1} = \left(\frac{\delta_{ij}}{p_i} + \frac{1}{p_k}\right)_{i,j=1,...,k-1}$ se sigue que

$$\begin{aligned} Y'\Sigma^{-1}Y &= \sum_{1}^{k-1} \left(\frac{1}{p_i} + \frac{1}{p_k}\right) \frac{(N_i - np_i)^2}{n} + \sum_{i=1}^{k-1}\sum_{j\neq i}^{k-1} \frac{(N_i - np_i)(N_j - np_j)}{n} = \\ &= \sum_{1}^{k-1} \frac{(N_i - np_i)^2}{np_i} + \frac{1}{np_k}\left\{\sum_{1}^{k-1}(N_i - np_i)^2 + \sum_{i\neq j}(N_i - np_i)(n_j - np_j)\right\} \\ &= \sum_{1}^{k} \frac{(N_i - np_i)^2}{np_i}. \end{aligned}$$

■

Se observa que aunque ambos métodos llegan a la misma distribución asintótica, los valores de los dos estadísticos pueden ser diferentes para un tamaño muestral fijo.

Para aplicar estos métodos es fundamental apoyarse en la distribución multinomial. Si la variable X es discreta, en algunas ocasiones, puede emplearse $(X_1, ..., X_n)$ como $(N_1, ..., N_k)$ y entonces $k = n$. Así en el caso de las tiradas de un dado, cada intervalo I_i es equivalente a cada una de las caras. Es por este motivo que algunos autores llaman a estos problemas estudios de variables cualitativas, aquí se ha preferido hacer el énfasis sobre la distribución multinomial subyacente que se usa como verosimilitud.

Ejemplo 8.2.1.1
Mendel postuló como probabilidades de aparición de las distintas clases, los valores que aparecen en la primera fila de la siguiente tabla:

LA	RA	LV	RV
$\frac{9}{16}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{16}$
315	101	108	32

donde la última fila son los valores observados para una m.a.s. de 556 elementos. Para contrastar la adecuación de la hipótesis, con un tamaño $\alpha = 0'1$, se tiene que mediante el método de la RV la región crítica es

$$RC = \{-2 \ln \lambda(n_1, ..., n_4) \geq \chi^2_{3;0'1}\}$$

con

$$-2 \ln \lambda(n_1, ..., n_4) = -2 \sum_{i=1}^{k} n_i \ln \frac{p_i^0}{n_i/n} = 0'48$$

y como $\chi^2_{3;0'1} = 6'251$ se mantiene la hipótesis nula.

Alternativamente mediante el método de la distancia de K. Pearson, la región crítica es

$$RC = \{D^2 \geq \chi^2_{3;0'1}\}$$

con

$$D^2 = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i} = \frac{(315 - 556 \cdot \frac{9}{16})^2}{556 \cdot \frac{9}{16}} + ... + \frac{(31 - 556 \cdot \frac{1}{16})^2}{556 \cdot \frac{1}{16}}$$
$$= 0'0162 + 0'1013 + 0'1349 + 0'2176 = 0'47$$

y como $0'47$ no pertenece a la RC el contraste resulta también no significativo.
■

Se comprueba que

$$\sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = \sum_{i=1}^{k} \frac{O_i^2}{E_i} - n$$

No obstante, es recomendable hacer la determinación a partir del cuadrado de cada diferencia, para detectar cuál es la causante del rechazo, caso de producirse éste.

Como se dijo al principio de esta Sección, se aborda ahora el problema de ajuste a una distribución parcialmente especificada, en la práctica una situación que se presenta con mayor frecuencia. Se empleará la notación $p_i(\theta_1, ..., \theta_r)$ para las probabilidades de los intervalos I_i con el fin de resaltar esta diferencia. En uniformidad con el caso anterior, todo se reduce a contrastar, a partir de los datos n_i $i = 1, ..., k$, la hipótesis nula H_0: provienen de una multinomial de parámetros $n, p_i(\theta_1, ..., \theta_r)$ $i = 1, ..., k$ con $k \geq r$, frente

a la hipótesis alternativa H_1: provienen de una multinomial de parámetros n, $p_i \quad i = 1, ..., k$.

Si se emplea el contraste de la RV se debe utilizar

$$\lambda(n_1, ..., n_k) = \frac{\sup_{\theta_1,...,\theta_r} \frac{n!}{n_1!...n_k!} p_1^{n_1}(\theta_1, ..., \theta_r)...p_k^{n_k}(\theta_1, ..., \theta_r)}{\sup_{p_1,...,p_k} \frac{n!}{n_1!...n_k!} p_1^{n_1}...p_k^{n_k}}$$

Para el cálculo del denominador, se sabe que el supremo se alcanza para $\widehat{p}_i = \frac{n_i}{n} \quad i = 1, ...k$ y para el numerador, hay que resolver en $\theta_1, ..., \theta_r$ el sistema

$$\left.\begin{array}{c} \sum_{i=1}^{k} \frac{n_i}{p_i(\theta_1,...,\theta_r)} \frac{\partial}{\partial \theta_j} p_i(\theta_1, ..., \theta_r) = 0 \\ j = 1, ..., r. \end{array}\right\}$$

En el caso de que exista regularidad suficiente, las soluciones de este sistema $p(\widehat{\theta})$ maximizarán el numerador de la RV, con lo que la región crítica ha de ser

$$RC = \left\{ \prod_{i=1}^{k} \left[\frac{p_i(\widehat{\theta})}{n_i/n} \right]^{n_i} \leq k \right\}$$

y por los resultados asintóticos de la RV, bajo H_0, se sigue que

$$-2 \ln \lambda(n_1, ..., n_k) \xrightarrow{\mathcal{L}} \chi^2_{k-1-r}$$

ya que la hipótesis nula tiene de dimensión r, mientras que $H_0 \cup H_1$ tiene de dimensión $k - 1$, (véase el Teorema 5.2.3.2).

Por lo tanto la región crítica viene dada por

$$RC = \left\{ \lambda(n_1, ..., n_k) = \prod_{i=1}^{k} \left[\frac{p_i(\widehat{\theta})}{n_i/n} \right]^{n_i} \leq k \right\}$$

Alternativamente, se puede aplicar el método de la distancia de K. Pearson para lo que se introduce la cantidad

$$D^2(p(\widehat{\theta}), \widehat{p}) = \sum_{i=1}^{k} \lambda_i \left(\frac{n_i}{n} - p_i(\widehat{\theta}) \right)^2$$

y si se pone $\lambda_i = \frac{n}{p_i(\widehat{\theta})}$ queda

$$D^2(p(\widehat{\theta}), \widehat{p}) = \sum_{i=1}^{k} \frac{(n_i - np_i(\widehat{\theta}))^2}{np_i(\widehat{\theta})}$$

para ver que la analogía con el Teorema 8.2.1.1, se mantiene también en esta situación se necesita el siguiente resultado asintótico, cuya demostración puede verse en Cramér (1963) pág. 489. También ocurre como en el caso

en que la distribución está totalmente especificada, los valores $np_i(\theta)$ son las frecuencias esperadas de los intervalos I_i bajo la hipótesis nula, que se estiman de manera natural por $np_i(\widehat{\theta})$ por lo que el estadístico de contraste es el mismo que el de la Expresión (8.2.2.1).

Teorema 8.2.1.2
Si $(N_1, ..., N_k) \sim Multinomial(n, p_1, ..., p_k)$, A es un intervalo de $\mathbb{R}^r$ y $p_i = p_i(\theta_1, ..., \theta_r)$ $i = 1, ..., k$ son funciones positivas y se cumple:

1) $\sum_{i=1}^{k} p_i(\overrightarrow{\theta}) = 1 \quad p_i(\overrightarrow{\theta}) > c > 0 \quad i = 1, ..., k$
2) las funciones $p_i(\overrightarrow{\theta})$ tienen derivadas parciales continuas
3) la matriz $\left(\frac{\partial}{\partial\theta_j} p_i(\overrightarrow{\theta})\right)_{\substack{i=1,...,k \\ j=1,...,r}}$ es de rango r

entonces, se tiene asegurada la existencia de una raíz consistente del sistema

$$\left.\begin{array}{c} \sum_{i=1}^{k} \frac{N_i}{p_i(\theta)} \frac{\partial}{\partial\theta_j} p_i(\theta) = 0 \\ j = 1, ..., r \end{array}\right\}$$

a la que se notará por $\widehat{\theta}_n$, y se cumple que

$$D^2(p(\widehat{\theta}_n), \widehat{p}_n) = \sum_{i=1}^{k} \frac{(N_i - np_i(\widehat{\theta}_n))^2}{np_i(\widehat{\theta}_n)} \xrightarrow{\mathcal{L}} \chi^2_{k-1-r}.$$

Ejemplo 8.2.1.2
Mediante un modelo genético, se supone que las probabilidades de 4 sucesos son $\left(\frac{1}{4}(2+\theta)\right), \frac{1}{4}(1-\theta), \frac{1}{4}(1-\theta), \frac{1}{4}\theta)$ con $0 < \theta < \frac{1}{4}$. Para contrastar la adecuación del modelo a los datos (n_1, n_2, n_3, n_4) con $n = \sum_{1}^{4} n_i$, puede comprobarse que hay que resolver la ecuación de segundo grado en θ

$$n\theta^2 + (n_4 - n_1 + 2n_2 + 2n_3)\theta - 2n_4 = 0$$

y obtener su raíz entre 0 y $\frac{1}{4}$.

Los dos estadísticos tienen asintóticamente distribución χ^2_2.

■

En la práctica aunque se ha dicho que la partición en los intervalos I_i puede ser arbitraria, es conveniente evitar valores muy pequeños en las frecuencias de los mismos. Así en las aplicaciones suelen tomarse los intervalos A_i para que las frecuencias esperadas sean mayores o iguales que 5, caso contrario se procede a refundir intervalos adyacentes hasta conseguirlo. Esto tanto para el problema de ajuste a una distribución totalmente especificada como a una parcialmente especificada.

8.2.2. Homogeneidad entre varias muestras

Se trata de contrastar si m muestras $(X_{i1},...,X_{in_i})\ \ i=1,...,m$ provienen todas de la misma población. Mediante la determinación de k intervalos y de forma análoga al problema de ajuste, se construye la siguiente tabla:

I_1	$\cdots$	I_k	
n_{11}	$\cdots$	n_{1k}	$n_{1\cdot}$
$\vdots$		$\vdots$	$\vdots$
n_{m1}	$\cdots$	n_{mk}	$n_{m\cdot}$
$n_{\cdot 1}$	$\cdots$	$n_{\cdot k}$	n

donde $n_{i\cdot}=\sum_1^k n_{ij}$, $n_{\cdot j}=\sum_1^m n_{ij}$, $n=\sum_1^m n_{i\cdot}=\sum_1^k n_{\cdot j}$, $i=1,...,m\ \ j=1,...,k$.

Lo que se contrasta realmente es la hipótesis nula H_0 : cada $(n_{i1},...,n_{ik})$ se distribuye mediante $Multinomial(n_{i\cdot},p_{\cdot 1},...,p_{\cdot k})\ \ i=1,...,m$ frente a la hipótesis alternativa H_1 : cada $(n_{i1},...,n_{ik})$ se distribuye mediante la $Multinomial(n_{i\cdot},p_{i1},...,p_{ik})$.

Mediante el Teorema 5.2.3.2, adecuado a esta situación, se tiene que

$$\lambda(n_{11},...,n_{mk})=\frac{\sup\limits_{p_{\cdot 1},...,p_{\cdot k}}\left[\prod\limits_{i=1}^m \frac{n_i!}{n_{ik}!...n_{ik}!}\right]p_{\cdot 1}^{n_{\cdot 1}},...,p_{\cdot k}^{n_{\cdot k}}}{\sup\limits_{p_{11},...,p_{mk}}\left[\prod\limits_{i=1}^m \frac{n_{i\cdot}!}{n_{i1}!...n_{ik}!}\right]p_{11}^{n_{11}},...,p_{mk}^{n_{mk}}}$$

por lo que la región crítica del contraste de la RV es

$$RC=\left\{\lambda(n_{11},...,n_{mk})=\frac{\prod\limits_{j=1}^k p_{\cdot j}^{n_{\cdot j}}}{\prod\limits_{i=1}^m\prod\limits_{j=1}^k p_{ij}^{n_{ij}}}\le k\right\}$$

es fácil comprobar que $\widehat{p}_{ij}=\frac{n_{ij}}{n_{i\cdot}}$ es el EMV bajo H_0 y que $\widehat{p}_{\cdot j}=\frac{n_{ij}}{n_{i\cdot}}$ es el EMV bajo H_1. Alternativamente, la distancia de K. Pearson introduce la cantidad

$$D^2(\widehat{p}_{\cdot j},\widehat{p}_{ij})=\sum_{i=1}^m\sum_{j=1}^k \lambda_{ij}(\widehat{p}_{ij}-\widehat{p}_{\cdot j})^2$$

con $\lambda_{ij}=\frac{n_{i\cdot}}{\widehat{p}_{\cdot j}}$, lo que puede escribirse como

$$D^2(\widehat{p}_{\cdot j},\widehat{p}_{ij})=\sum_{i=1}^m\sum_{j=1}^k \frac{(n_{ij}-n_{i\cdot}n_{\cdot j}/n)^2}{n_{i\cdot}n_{\cdot j}/n}.$$

En este caso la hipótesis $H_0\cup H_1$ tiene de dimensión $m(k-1)$ mientras que la hipótesis nula tiene de dimensión $k-1$. Aplicando el Teorema 8.2.1.2 se puede poner la región crítica del test de la razón de verosimilitudes como

$$RC=\{-2\ln\lambda(n_{11},...,n_{mk})\ge\chi^2_{(m-1)(k-1);\alpha}\}.$$

Igual que en el caso del problema de la bondad del ajuste, se pueden interpretar los valores $n_{i.}n_{.j}/n$ como las frecuencias esperadas de la muestra de tamaño $n_{i.}$ en el intervalo I_j bajo la hipótesis de homogeneidad, ya que al ser la muestra de tamaño $n_{i.}$ dicha frecuencia esperada debería de ser $n_{i.}p_{.j}$, valor que puede ser estimado por $n_{i.}\widehat{p}_{.j}$ con $\widehat{p}_{.j}=\frac{n_{.j}}{n}$ el valor del estimador de máxima verosimilitud bajo la suposición de verosimilitud multinomial que implica la hipótesis nula.

La región crítica bajo el método de la distancia de K. Pearson puede escribirse en la forma

$$RC = \left\{ \sum_{i=1}^{m} \sum_{j=1}^{k} \frac{(n_{ij} - n_{i.}n_{.j}/n)^2}{n_{i.}n_{.j}/n} \geq \chi^2_{(m-1)(k-1);\alpha} \right\}$$

El enunciado que establece la distribución asintótica de este test puede verse en Cristóbal (1995), pág. 627.

Por el mismo motivo que en el caso del problema del ajuste a una distribución, siempre que sea posible se refundirán los intervalos I_j, para conseguir que las frecuencias sean mayores o iguales que 5.

Si la población es una variable aleatoria discreta, se pueden tomar los intervalos I_i como los valores que toma la variable.

Ejercicio 8.2.2.1

El estudio del grupo sanguíneo de 353 individuos de la comunidad C_1 y de 364 de la comunidad C_2, dió los siguientes resultados:

	grupos				
	0	A	B	AB	
C_1	121	120	79	33	353
C_2	118	95	121	30	364
	239	215	200	63	717

Para contrastar la homogeneidad respecto a los grupos sanguíneos de las dos comunidades, con un nivel de significación $\alpha = 0'05$, se tiene que los estimadores de MV, con distribución multinomial, bajo la hipótesis nula son $\widehat{p}_{.j} = \frac{n_{.j}}{n}$ y bajo la hipótesis alternativa los estimadores, con distribución multinomial, de las probabilidades de los grupos son diferentes en ambas comunidades y valen $\widehat{p}_{ij} = \frac{n_{ij}}{n_{i.}}$. Por lo que se tiene

$$H_0 : 0'333, 0'2999, 0'2789, 0'0879$$
$$H_1 : 0'3428, 0'3399, 0'2238, 0'0824$$
$$0'3242, 0'2610, 0'3324, 0'0824$$

Mediante el método de la RV

$$\lambda(n_{11}, ..., n_{24}) = \prod_{j=1}^{n} \widehat{p}_{.j}^{\,n_{.j}} \Big/ \prod_{i=1}^{2} \prod_{j=1}^{4} \widehat{p}_{ij}^{\,n_{ij}}$$

y la región crítica es

$$RC = \{-2 \ln \lambda(n_{11}, ..., n_{24}) \geq \chi^2_{3;0'05}\}$$

el valor del estadístico es $2(930'0864 - 924'1808) = 11'8112$ y como $\chi^2_{3;0'05} = 7'815$ se rechaza la hipótesis nula.

Mediante el método de la distancia de K. Pearson, el valor del estadístico es

$$D^2 = \frac{(121 - 353 \cdot 239/717)^2}{353 \cdot 239/717} + ... + \frac{(30 - 364 \cdot 63/717)^2}{364 \cdot 63/717} = 11'74$$

por lo que también se rechaza la hipótesis nula al ser $11'74 > 7'815$.

8.2.3. Tablas de contingencia

Se supone que se tiene una m.a.s. correspondiente a una v.a. bidimensional $(x_1, y_1), ..., (x_n, y_n)$ y se trata de contrastar la hipótesis nula H_0: las dos variables son independientes, frente a la hipótesis alternativa H_1 : las dos variables no son independientes. Dividiendo en intervalos el recorrido de las v.a. respectivas, se obtienen los sucesos $A_i \;\; i = 1, ..., m$ y $B_j \;\; j = 1, ..., k$, con lo que se construye la siguiente tabla:

	B_1	$\cdots$	B_k	
A_1	n_{11}	$\cdots$	n_{1k}	$n_{1\cdot}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$
A_m	n_{m1}	$\cdots$	n_{mk}	$n_{m\cdot}$
	$n_{\cdot 1}$	$\cdots$	$n_{\cdot k}$	n

donde los valores n_{ij} son el número de elementos en la muestra contenidos en el suceso $A_i \times B_j$, y los restantes elementos son

$$n_{i\cdot} = \sum_{j=1}^{k} n_{ij} \qquad n_{\cdot j} = \sum_{i=1}^{m} n_{ij} \qquad n = \sum_{i=1}^{m} n_{i\cdot} = \sum_{j=1}^{k} n_{\cdot j}$$

La hipótesis nula de independencia, viene caracterizada por $H_0 : (n_{11}, ..., m_{mk})$ se distribuye como una multinomial $Multinomial(n, p_{1\cdot}p_{\cdot 1}, ..., p_{m\cdot}p_{\cdot k})$ frente a la hipótesis alternativa $H_1 : (n_{11}, ..., n_{mk})$ tiene distribución $Multinomial(n,\, p_{11}, ..., p_{mk})$.

La hipótesis nula tiene dimensión $m-1+k-1$ mientras que la hipótesis alternativa tiene de dimensión $mk - 1$. La RV viene dada por

$$\lambda(n_{11}, ..., n_{mk}) = \frac{\sup\limits_{p_{1\cdot}, ..., p_{m\cdot}, p_{\cdot 1}, ..., p_{\cdot k}} \frac{n!}{n_{11}! ... n_{mk}!} (p_{1\cdot}p_{\cdot 1})^{n_{11}} ... (p_{m\cdot}p_{\cdot k})^{n_{mk}}}{\sup\limits_{p_{11}, ..., p_{mk}} \frac{n!}{n_{11}! ... n_{mk}!} p_{11}^{n_{11}} ... p_{mk}^{n_{mk}}}$$

Es fácil ver que

$$\widehat{p}_{ij} = \frac{n_{ij}}{n} \qquad \widehat{p}_{i\cdot} = \frac{n_{i\cdot}}{n} \qquad \widehat{p}_{\cdot j} = \frac{n_{\cdot j}}{n}, \quad i = 1, ..., m, \ j = 1, ..., k$$

con lo que, de acuerdo con el Teorema 5.2.3.2 la región crítica del test de la RV es

$$RC = \left\{ -2 \sum_{i=1}^{m} \sum_{j=1}^{k} n_{ij} \ln \frac{\widehat{p}_{i\cdot}\widehat{p}_{\cdot j}}{\widehat{p}_{ij}} \geq \chi^2_{(m-1)(k-1);\alpha} \right\}$$

Alternativamente, se puede utilizar el método de la distancia de K. Pearson mediante la cantidad

$$D^2(\widehat{p_{i\cdot}\widehat{p_{\cdot j}}}, \widehat{p}_{ij}) = \sum_{i=1}^{m} \sum_{j=1}^{k} \lambda_{ij} (\widehat{p_{i\cdot}\widehat{p_{\cdot j}}} - \widehat{p}_{ij})^2$$

con $\lambda_{\iota j} = \frac{n}{\widehat{p_{i\cdot}\widehat{p_{\cdot j}}}}$ con lo que queda en la forma

$$D^2(\widehat{p_{i\cdot}\widehat{p_{\cdot j}}}, \widehat{p}_{ij}) = \sum_{i=1}^{m} \sum_{j=1}^{k} \frac{(n_{ij} - n_{i\cdot}n_{\cdot j}/n)^2}{n_{i\cdot}n_{\cdot j}/n}$$

lo que lleva, de acuerdo con el Teorema 8.2.1.2 adecuado a esta situación, a la región crítica

$$RC = \left\{ \sum_{i=1}^{m} \sum_{j=1}^{k} \frac{(n_{ij} - n_{i\cdot}n_{\cdot j}/n)^2}{n_{i\cdot}n_{\cdot j}/n} \geq \chi^2_{(m-1)(k-1);\alpha} \right\}$$

Los valores $n_{i\cdot}n_{\cdot j}/n$ son las frecuencias esperadas bajo la hipótesis nula, ya que el número esperado de observaciones del suceso $A_i \times B_j$ en la muestra es $np_{i\cdot}p_{\cdot j}$, los valores desconocidos se sustituyen por sus estimadores de MV $\widehat{p}_{i\cdot}$ y $\widehat{p}_{\cdot j}$ respectivamente.

Ejemplo 8.2.3.1
Para contrastar la independencia entre los turnos y las líneas de producción de una factoría, se dispone de los siguientes datos

	Turnos		
Líneas	1	2	3
1	10	12	13
2	14	9	12

Si se quiere contrastar la independencia entre ambas variables con un nivel de significación $\alpha = 0'05$, se tiene que mediante el método de la razón de verosimi- litudes, la región crítica es

$$RC = \left\{ -2 \sum_{i=1}^{2} \sum_{j=1}^{3} n_{ij} \ln \frac{\widehat{p}_{i\cdot}\widehat{p}_{\cdot j}}{\widehat{p}_{ij}} \geq \chi^2_{2;0'05} = 5'991 \right\}$$

y como

$$-2\sum_{i=1}^{2}\sum_{j=1}^{3} n_{ij}\ln\frac{\widehat{p}_{i\cdot}\widehat{p}_{\cdot j}}{\widehat{p}_{ij}} = -2\left\{10\left(\ln\frac{35}{70}\cdot\frac{24}{70}-\ln\frac{10}{70}\right)+....\right.$$
$$\left.... + 12\left(\ln\frac{35}{70}\cdot\frac{25}{70}-\ln\frac{12}{70}\right)\right\} = 1'14$$

es menor que $5'991$, se mantiene la hipótesis nula.

Por otro lado mediante el método de la distancia de K. Pearson, la región crítica es de la forma

$$RC = \left\{-2\sum_{i=1}^{2}\sum_{j=1}^{3} n_{ij}\ln\frac{\widehat{p}_{i\cdot}\widehat{p}_{\cdot j}}{\widehat{p}_{ij}} \geq \chi^2_{2;0'5} = 5'991\right\}$$

donde

$$\frac{\left(10-\frac{35\cdot 24}{70}\right)^2}{\frac{35\cdot 24}{70}} + ... + \frac{\left(12-\frac{35\cdot 25}{70}\right)^2}{\frac{35\cdot 25}{70}} = 1'14$$

con lo cual se acepta también la hipótesis nula.

■

Es importante tener en cuenta a la hora de trabajar con tablas de contingencia, el siguiente hecho, tomado de DeGroot (1988) pág. 521, que recibe el nombre de Paradoja de Simpson. Consiste en que en el estudio de tablas de contingencia, puede ocurrir que no se obtenga la misma conclusión de una tabla que se disgrega en varias tablas de la misma dimensión o recíprocamente, que de varias tablas de las que se saca una conclusión al agregarlas en una sola se obtenga la conclusión contraria.

Ejemplo 8.2.3.2
Se realiza un experimento, para comparar un nuevo tratamiento de cura de cierta enfermedad con el tratamiento estándar, y se obtienen los siguientes resultados:

Con todos los pacientes	Mejoran	No mejoran
Nuevo tratamiento	20	20
Tratamiento estándar	24	16

De los datos parece desprenderse que las dos poblaciones no son homogéneas y que el tratamiento estándar era mejor.

Si se realiza el mismo análisis, pero por separado para hombres y mujeres se obtienen los resultados

Hombres	Mejoran	No mejoran
Nuevo tratamiento	12	18
Tratamiento estándar	3	7

Mujeres	Mejoran	No mejoran
Nuevo tratamiento	8	2
Tratamiento estándar	21	9

aquí se observa que el porcentaje de mejoras con el nuevo tratamiento es más alto en ambos casos.

La explicación de la paradoja es que las mujeres se recuperan, con el nuevo tratamiento o sin él, en un porcentaje más alto que los hombres. Aunque se han analizado tantos hombres como mujeres, 40 de cada, una alta proporción de mujeres y una baja proporcion de hombres recibieron el tratamiento estándar con lo que el efecto del tratamiento es enmascarado por el efecto debido al sexo, entre los hombres a 30 se les dió el tratamiento y a 10 no; mientras que en las mujeres es justo al contrario.
■

Para evitar este tipo de situaciones hay que recurrir a utilizar no el mismo número de elementos sino el mismo porcentaje.

8.3. APLICACIONES DE LOS ESTADÍSTICOS ORDENADOS

En la línea de lo dicho en la introducción de este Capítulo, se estudia el comportamiento de algún método de *distribución libre*, se dice que un estadístico tiene distribución libre cuando su distribución, no depende de la distribución de la población. La única suposición que se hace es que la función de distribución de la población sea una función continua. A continuación se estudian los *intervalos de tolerancia,* cómo determinar intervalos de confianza para los cuantiles de la población y la manera de contrastar hipótesis sobre los cuantiles de la población.

8.3.1. Intervalos de tolerancia

Con frecuencia en procedimientos industriales se está interesado en estimar un amplio porcentaje de la distribución que sigue el proceso de fabricación. Para fijar ideas se supone que se están fabricando una clase de tuercas por una máquina. Las tuercas han de tener una longitud de 2 cm pero puede permitirse, pues son utilizables todavía, que la longitud de las mismas esté entre 1'99 y 2'01 cm. Tiene interés el determinar mediante una muestra de tamaño n dos funciones $L_I(X_1, ..., X_n)$ y $L_S(X_1, ..., X_n)$ tales que la probabilidad de que la v.a. X, longitud de las tuercas producidas, esté entre dichas funciones sea suficientemente alta. Más precisamente se trata de, fijados dos valores próximos a 1, p y $1-\alpha$ determinar las funciones para que se cumpla

$$P\{P\{L_I(X_1, ..., X_n) < X < L_S(X_1, ..., X_n)\} \geq p\} \geq 1 - \alpha. \qquad (8.3.1.1)$$

Si el intervalo $(L_I(X_1, ..., X_n), L_S(X_1, ..., X_n))$ estuviera contenido en el intervalo $(1'99, 2'01)$ se aceptaría como correcto el proceso de fabricación

y caso contrario se rechazaría y se haría revisar la máquina. El problema es que para satisfacer (8.3.1.1) se necesita conocer la distribución de la población X. Una solución es tratar de determinar $L_I(X_1, ..., X_n) = X_{(r)}$ $L_S(X_1, ..., X_n) = X_{(s)}$ con r menor que s tales que

$$P\{P\{X_{(r)} < X < X_{(s)}\} \geq p\} \geq 1 - \alpha$$

o lo que es lo mismo

$$P\{F(X_{(s)}) - F(X_{(r)}) \geq p\} \geq 1 - \alpha.$$

Como $F(X_{(s)}) - F(X_{(r)}) = C_{r+1} + ... + C_s$ donde C_i es el recubrimiento de orden i (véase la Sección 1.6.3), por ser la distribución de una suma de recubrimientos función sólo del número de sumandos, se trata de determinar r y s para que

$$P\{F(X_{(r-s)}) \geq p\} \geq 1 - \alpha$$

o lo que es lo mismo

$$P\{F(X_{(r-s)}) < p\} \leq \alpha$$

para lo que ha de cumplirse

$$\sum_{i=r-s}^{n} p_i \leq \alpha \tag{8.3.1.2}$$

con p_i las probabilidades correspondientes a una v.a. $Binomial(n, p)$.

En resumen un *intervalo de tolerancia* de coeficiente $1-\alpha$ es $(X_{(r)}, X_{(s)})$ si se verifica que contiene, entre sus extremos aleatorios al menos un porcentaje del 100p % de la distribución de probabilidad de la población con la que se haya obtenido la muestra. Para su determinación basta despejar $r - s$ de la Ecuación (8.3.1.2) o, alternativamente de la Expresión

$$\sum_{i=0}^{s-r-1} p_i \geq 1 - \alpha.$$

Ejemplo 8.3.1.1
Para determinar un intervalo de tolerancia de coeficiente $1 - \alpha = 0'8$, que cubra el 50 % de la probabilidad de la v.a. X, con m.a.s. de tamaño $n = 5$, se emplea la distribución $Binomial(5, 0'5)$ con lo que se tiene:

x	p_i	$F(x)$
0	0'0312	0'0312
1	0'1562	0'1874
2	0'3125	0'4999
3	0'3125	0'8124
4	0'1562	0'9686
5	0'0312	0'9998

y como

$$\sum_{i=0}^{3} p_i = 0'8124 \geq 0'8$$

ha de ser $s - r - 1 = 3$ cuya única solución compatible con el *tamaño muestral* $n = 5$ es $r = 1$ y $s = r + 4 = 5$ por lo que el intervalo de tolerancia pedido es $(X_{(1)}, X_{(5)})$.

Se observa, que si se toma de coeficiente $1 - \alpha = 0'9$ y $p = 0'5$ no se puede determinar un intervalo de tolerancia con muestras de tamaño 5 ya que entonces $s - r - 1 = 4$ y $s = r + 5$ que para $r = 1$ da como $s = 6$, un valor superior a 5.

■

En la práctica, suele ser más interesante fijar p y α y determinar el mínimo valor n para el cual $(X_{(1)}, X_{(n)})$ sea un *intervalo de tolerancia* de nivel $1 - \alpha$ para un porcentaje p de la distribución. Se trata entonces de despejar n de la Expresión

$$p_{n-1} + p_n \leq \alpha.$$

Así, en el caso del ejemplo de las tuercas que ha servido de introducción, se puede concretar la petición, en determinar el menor valor n a partir del cual se tenga

$$P\{P\{X_{(1)} < X < X_{(n)}\} \geq 0'8\} \geq 0'99$$

con lo que se trata de obtener el menor valor de n para el cual

$$p_{n-1} + p_n \leq 0'01.$$

Se observa que con

$$n = 30,\ p_{n-1} + p_n = 0'0012 + 0'0093 = 0'0105 \not\leq 0'01$$
$$n = 31,\ p_{n-1} + p_n = 0'0010 + 0'0077 = 0'0087 \leq 0'01$$

luego se deberán observar 31 tuercas y aceptar el lote si el intervalo $(X_{(1)}, X_{(31)})$ está contenido en el intervalo $(1'99, 2'01)$.

Ejemplo 8.3.1.2

Si se quiere determinar el tamaño muestral n que ha de tomarse, para que con probabilidad $0'90$ se cubra un porcentaje del $99\,\%$ de la distribución de la v.a., basta obtener el mínimo n para que sea $p_0 + p_1 + ... + p_{n-2} \geq 0'9$ con p_i las probabilidades correspondientes a una v.a. $Binomial(n, 0'99)$. Ha de verificarse que $P\{X \leq n - 2\} \geq 0'9$ o equivalentemente

$$P\left\{\frac{X - np}{\sqrt{npq}} \leq \frac{n - 2 - np}{\sqrt{npq}}\right\} = P\left\{Z \leq \frac{n - 2 - np}{\sqrt{npq}}\right\}$$

siendo Z una v.a. $N(0, 1)$. Así de las tablas de la v.a. $N(0, 1)$ (véase la Tabla 3) se obtiene $\frac{n-2-np}{\sqrt{npq}} = 1'29$ con lo que la resolución de la ecuación de segundo grado conveniente, permite obtener de manera aproximada que ha de ser $n \geq 185$.

8.3.2. Intervalos de confianza para los cuantiles

La siguiente aplicación a estudiar dentro de los métodos no paramétricos es la determinación de intervalos de confianza para los cuantiles de la población. Si se supone que la población X se distribuye con una función de distribución continua F, el cuantil de orden p es el mínimo valor k_p tal que $F(k_p) = p$. Se pretende encontrar r y s tales que $s > r$ verificando

$$P\{X_{(r)} \leq k_p \leq X_{(s)}\} \geq 1 - \alpha$$

si se aplica F a las desigualdades, se tiene que

$$P\{F(X_{(r)}) \leq p \leq F(X_{(s)})\} \geq 1 - \alpha.$$

Ahora bien, como la variable $F(x)$ tiene distribución uniforme, esto equivale a que

$$P\{U_{(r)} \leq p \leq U_{(s)}\} \geq 1 - \alpha \tag{8.3.2.1}$$

siendo $U_{(r)}$ y $U_{(s)}$ los correspondientes estadísticos ordenados de la distribución $Uniforme(0,1)$. Esta expresión es independiente de cual sea la función de distribución de la población.

Alternativamente, la Expresión (8.3.2.1) se puede escribir como

$$P\{U_{(r)} \leq p\} - P\{U_{(s)} \leq p\} = \sum_{i=r}^{n} p_i - \sum_{j=s}^{n} p_j$$

donde p_i son las probabilidades correspondientes a los valores que toma la v.a. $Binomial(n,p)$. Basta entonces, dado $1-\alpha$ próximo a 1, determinar r y s para que se cumpla

$$\sum_{i=r}^{s-1} p_i \geq 1 - \alpha$$

como es más sencillo, normalmente se determinan resolviendo el sistema

$$\left.\begin{array}{c} p_0 + ... + p_{r-1} \leq \frac{\alpha}{2} \\ p_s + ... + p_n \leq \frac{\alpha}{2} \end{array}\right\} \tag{8.3.2.2}$$

También tiene sentido, como en el caso de los intervalos de tolerancia, preguntarse por el menor tamaño muestral para el cual $(X_{(1)}, X_{(n)})$ es un intervalo de grado de confianza $1-\alpha$, la particularización del Sistema (8.3.2.2) a este caso, viene dada por $p_0 \leq \frac{\alpha}{2}$ y $p_n \leq \frac{\alpha}{2}$.

Ejemplo 8.3.2.1
Para m.a.s. de tamaño 10, un intervalo de grado de confianza $0'9$ para la mediana de la población, que se corresponde con $p = 0'5$, es el $(X_{(2)}, X_{(9)})$

ya que de acuerdo con las tablas de la distribución *Binomial* si $X \sim Binomial(10, 0'5)$ se tiene

x	$P\{X = x\}$	$F(x)$
0	0'0010	0'0010
1	0'0098	0'0108
2	0'0439	0'0457

luego mediante el Sistema (8.3.2.2), al ser $F(1) \leq 0'05$ ha de cumplirse que $r - 1 = 1$, de donde $r = 2$, y, por la simetría de la *Binomial* para $p = 0'5$ es $s = 9$.

Ejemplo 8.3.2.2
El menor valor n tal que el intervalo $(X_{(1)}, X_{(n)})$ contenga a la mediana de la población con probabilidad mayor o igual que $0'9$ es 5, ya que, de las tablas de la $Binomial(n, 0'5)$, la primera vez que se satisfacen a la vez $p_0 \leq 0'05$ y $p_n \leq 0'05$ es con $n = 5$.

Se puede repetir el enunciado y determinar el menor valor de n, tal que $(X_{(2)}, X_{(n-1)})$ sea un intervalo de grado de confianza $0'9$ para la mediana. Esta vez las Ecuaciones (8.3.2.2) se reducen a

$$\left.\begin{array}{c} p_0 + p_1 \leq 0'05 \\ p_{n-1} + p_n \leq 0'05 \end{array}\right\}$$

y debido a la simetría, la primera ecuación es suficiente y es satisfecha por primera vez para $n = 8$, con $p_0 + p_1 = 0'0351$, por lo que el intervalo es el $(X_{(2)}, X_{(7)})$.

8.3.3. Contrastes de hipótesis sobre los cuantiles

Suponiendo que la población tiene función de distribución continua, se trata de contrastar la hipótesis nula $H_0 : k_p = k$ de que el cuantil de orden p es k, es decir $F(k) = p$. Las posibles hipótesis alternativas y las regiones críticas, quedan recogidas en la Tabla 8.3.3.1.

H_0	H_1	Región Crítica
$k_p = k$	$k_p > k$	$X_{(r)} > k$
	$k_p < k$	$X_{(s)} < k$
	$k_p \neq k$	$X_{(r)} > k$ ó $X_{(s)} < k$ con $r > s$

Tabla 8.3.3.1. Regiones críticas para contrastar que el cuantil de orden p de una población es k.

Si la muestra es compatible con la hipótesis nula el $100p\,\%$ de las observaciónes deberían ser menores que k y las restantes $100(1-p)\,\%$ mayores

que k. Por el contrario si hay muchas observaciones por encima de k, presumiblemente será $k_p > k$; el que haya muchas observaciones por encima de k se detecta mediante $X_{(r)} > k$ con r razonablemente pequeño. La determinación de r, se hace a partir del tamaño α deseado para el contraste, así si se quiere que

$$P\{X_{(r)} > k|H_0\} \leq \alpha$$

aplicando F, ha de cumplirse que

$$P\{F(X_{(r)}) > F(k) = p|H_0\} \leq \alpha$$

o, lo que es lo mismo,

$$P\{U_{(r)} > p\} = 1 - P\{U_{(r)} \leq p\} = \sum_{i=0}^{r-1} p_i \leq \alpha,$$

donde, $U_{(r)}$ es el estadístico ordenado de orden r de una muestra de una población con distribución $Uniforme(0,1)$ y p_i las probabilidades correspondientes a una v.a. $Binomial(n,p)$.

Una manera alternativa de realizar este contraste consiste en construir la muestra $X_1 - k, ..., X_n - k$ y contabilizar ahora el número de diferencias positivas existentes. Bajo la hipótesis nula, la probabilidad de obtener una diferencia positiva es

$$P\{X_i > k|H_0\} = 1 - F(k) = 1 - p$$

y el número de diferencias positivas, bajo la hipótesis nula, es una v.a. $Binomial(n, 1-p)$. Se puede entonces, trabajar con la variable número de diferencias positivas en lugar de con $X_{(r)}$. Este test será desarrollado con más detalle más adelante en la Sección 8.4.2.

Ejemplo 8.3.3.1
Se pretende contrastar, a un tamaño de $0'1$, que la mediana de una distribución continua es 2 frente a que es mayor. Se obtiene la muestra

$$-3, -6, 1, 9, 4, 10, 12.$$

La región crítica es de la forma

$$RC = \{X_{(r)} > 2\}$$

donde r se obtiene de

$$\sum_{i=0}^{r-1} p_i \leq 0'1$$

siendo p_i las probabilidades de la v.a. $Binomial(7, 0'5)$ (véase la Tabla 8.3.3).

Esto lleva a que $r = 2$ ya que $p_0 = 0'0078$, $p_0 + p_1 = 0'0625$ y $p_0 + p_1 + p_2 = 0'2266$. Para la muestra dada es $X_{(2)} = -3$, que no es mayor que 2, luego $X_{(2)}$ no pertenece a la región crítica y se mantiene la hipótesis nula.

Alternativamente, el número de signos positivos en la muestra $X_i - 2$, es una v.a. $Binomial(7, p)$, en la que $p = P\{X > 2\} = 0'5$. Se rechaza la hipótesis nula hacia la hipótesis alternativa, cuando el número de signos positivos sea muy grande, es decir, la región crítica es de la forma

$$RC = \{\text{número de signos positivos } \geq k\}$$

y para la determinación de k se emplea

$$0'1 = P\{\text{número de signos positivos } \geq k\} = \sum_{i=k}^{n} q_i$$

con q_i las probabilidades correspondientes a la $Binomial(7, 0'5)$; luego ha de ser $k = 6$. Como el número de signos positivos es 4, se mantiene H_0.

Obsérvese que es equivalente decir que el número de signos positivos es superior a 6 a que $X_{(2)}$ es mayor que 2.

8.4. PROBLEMAS NO PARAMÉTRICOS RELATIVOS A UNA MUESTRA

Estos problemas se dividen en dos grupos claramente diferenciados: los relativos a la *bondad del ajuste* y los *problemas de localización*.

Los primeros son análogos a los problemas estudiados en la Sección 8.2 relativos a la distribución multinomial, tratados mediante los tests de la χ^2, pero difieren de ellos en que se trabaja con funciones de distribución continuas. Este primer grupo de problemas está basado en la introducción de los *estadísticos de Kolmogorov-Smirnov*.

En el segundo grupo, los *problemas de localización*, se trata de realizar tests sobre los cuantiles de la población y básicamente se supone, además de la hipótesis de continuidad de la función de distribución, la hipótesis de simetría de la misma. En concreto se van a estudiar el *contraste de los signos* y el *contraste de los rangos signados* o de *Wilcoxon*.

8.4.1. Contrastes de Kolmogorov-Smirnov para una muestra

Dada una m.a.s. de tamaño n de una población X con función de distribución $F(x)$ continua se ha definido (véase la Definición 1.4.1) la función de distribución empírica o muestral mediante $F_n^*(x) = \frac{i}{n}$ si $x_{(i)} \leq x \leq x_{(i+1)}$ con $i = 0, 1, ..., n$ donde aquí se añade por convenio $x_{(0)} = -\infty$ y $x_{(n+1)} = +\infty$. Se introducen los siguientes estadísticos.

Definición 8.4.1.1 (Estadísticos de Kolmogorov-Smirnov)
Se llaman *estadísticos unilaterales de Kolmogorov-Smirnov* a

$$D_n^+ = \sup_x (F_n^*(x) - F(x))$$
$$D_n^- = \sup_x (F(x) - F_n^*(x))$$

y *estadístico bilateral de Kolmogorov-Smirnov* a

$$D_n = \sup_x |F_n^*(x) - F(x)|.$$

■

El *teorema de Glivenko-Cantelli* (véase el Teorema 1.4.7), asegura que $F_n^*(x)$ converge a $F(x)$ uniformemente con probabilidad uno. Por lo tanto la magnitud de las diferencias entre $F_n^*(x)$ y $F(x)$, proporciona información de la compatibilidad entre la muestra y $F(x)$.

Proposición 8.4.1.1
Para m.a.s. de tamaño n de una población X con función de distribución F continua, la función de distribución de los estadísticos D_n^+, D_n^- y de D_n no depende de F.

Demostración: Recordando que $x_{(0)} = -\infty$ y $x_{(n+1)} = +\infty$, se puede poner

$$D_n^+ = \max_{0\le i\le n} \sup_{x_{(i)}\le x\le x_{(i+1)}} \left(\frac{i}{n} - F(x)\right)$$

Como en el intervalo $[x_{(n)}, x_{(n+1)})$ la función $1-F(x)$ es positiva o nula y $\sup_{x_{(i)}\le x<x_{(i+1)}} \left(\frac{i}{n} - F(x)\right) = \frac{i}{n} - F(x_{(i)})$ se tiene que

$$D_n^+ = \max_{1\le i\le n} \left(\frac{i}{n} - F(x_{(i)})\right) \tag{8.4.1.1}$$

como $F(X_{(i)}) = U_{(i)}$ es el estadístico ordenado de orden i de una muestra de la v.a. $Uniforme(0,1)$; es claro que esta expresión no depende de la distribución F.

Análogamente

$$\begin{aligned} D_n^- &= \max_{0\le i\le n} \sup_{x_{(i)}\le x\le x_{(i+1)}} \left(F(x) - \tfrac{i}{n}\right) \\ &= \max_{0\le i\le n} \left(F(x_{(i+1)}) - \tfrac{i}{n}\right) \\ &= \max_{1\le i\le n} \left(F(x_{(i)}) - \tfrac{i-1}{n}\right) \end{aligned} \tag{8.4.1.2}$$

sin más que observar que en el intervalo $[x_{(0)}, x_{(1)})$ la diferencia $F(x_{(i+1)}) - \frac{i}{n}$ es positiva o nula.

Por último al ser $D_n = \max\{D_n^+, D_n^-\}$ se sigue que la distribución de los tres estadísticos es independiente de la función de distribución F.

Se puede entonces enunciar esta proposición diciendo, que los *estadísticos de Kolmogorov-Smirnov* tienen *distribución libre* (véase la Sección 8.3).

■

Se ha dado esta demostración porque permite obtener la expresión que se utiliza para el cálculo de estos estadísticos en las aplicaciones. Existe un procedimiento más sencillo de prueba mediante el paso de la variable X a la variable $U = F(X)$ que tiene distribución uniforme, en concreto para D_n se tiene

$$\begin{aligned} D_n &= \sup_x |F_n^*(x) - F(x)| = \sup_{u\in(0,1)} |F_n^*(F^{-1}(u)) - F(F^{-1}(u))| \\ &= \sup_{u\in(0,1)} |F_n^*(F^{-1}(u)) - u| \end{aligned}$$

y por ser $F_n^*(k) = \frac{\text{n}^o \text{ de } X_i \leq k}{n}$ se sigue que

$$F_n^*(F^{-1}(u)) = \frac{\text{n}^o \text{ de } X_i \leq F^{-1}(u)}{n} = \frac{\text{n}^o \text{ de } F(X_i) \leq u}{n}$$

con lo que

$$D_n = \sup_{u\in(0,1)} |G_n^*(u) - u|$$

donde $G_n^*(u)$ es la distribución empírica correspondiente a una m.a.s. de tamaño n de una población uniforme. Por tanto D_n tiene distribución independiente de F.

El mismo tipo de razonamiento es aplicable para D_n^+ y D_n^-.

■

La distribución exacta de estos estadísticos queda recogida en el siguiente teorema.

Teorema 8.4.1.1
Para una población, con función de distribución continua y para m.a.s. de tamaño n es

$$P\{D_n^+ \leq z\} = \begin{cases} 0 & \text{si } z \leq 0 \\ n! \int_{1-z}^{1} \int_{\max\{\frac{n-1}{n}-z,0\}}^{u_n} \cdots & \\ \quad \cdots \int_{\max\{\frac{2}{n}-z,0\}}^{u_3} \int_{\max\{\frac{1}{n}-z,0\}}^{u_2} du_1 du_2 ... du_{n-1} du_n & \text{si } 0 \leq z \leq 1 \\ 1 & \text{si } z \geq 1 \end{cases}$$

Demostración: En virtud de la Expresión (8.4.1.1)

$$P\{D_n^+ \leq z\} = P\left\{\max_{1\leq i\leq n}\left(\frac{i}{n} - U_{(i)}\right) \leq z\right\} = P\left\{U_{(i)} \geq \frac{i}{n} - z, i = 1, 2, ..., n\right\}$$

Dado que la función de densidad del estadístico ordenado de una población uniforme es

$$f(u_1,...,u_n) = \begin{cases} n! & \text{si } 0 \le u_1 \le u_2 \le ...u_n \le 1 \\ 0 & \text{resto} \end{cases}$$

(véase el Teorema 1.6.1.1), se obtiene inmediatamente la expresión del enunciado.

■

Para D_n^-, usando un razonamiento análogo y a partir de la Expresión (8.4.1.2), se puede ver que éste tiene la misma distribución que D_n^+. Encontrar ahora la distribución del máximo, al menos desde un punto de vista teórico, es viable y por lo tanto se conoce la distribución de D_n.

Las distribuciones asintóticas de estos estadísticos fueron obtenidas por Kolmogorov (1933, 1941) y por Smirnov (1939, 1948), quien dió una demostración más sencilla y generalizó el problema al caso de dos muestras.

La utilización de estos estadísticos, para contrastar si una m.a.s. proviene de una distribución totalmente especificada $F_0(x)$, queda recogida en la Tabla 8.4.1.1

H_0	H_1	Región Crítica
$F = F_0$	$F \neq F_0$	$D_n > d_{n,\alpha}$
	$F > F_0$	$D_n^+ > d_{n,\alpha}^+$
	$F < F_0$	$D_n^- > d_{n,\alpha}^-$

Tabla 8.4.1.1. Regiones críticas del test de Kolmogorov-Smirnov para una muestra.

Si por ejemplo la hipótesis alternativa es que, $F(x) \ge F_0(x)$ existiendo algún valor para el que la desigualdad sea estricta, resultará que cuando H_0 sea cierta $F_n^*(x)$ será aproximadamente igual a $F_0(x)$, para todo x, y D_n^+ será pequeño. Por el contrario bajo la hipótesis alternativa $F_n^*(x)$ se aproximará a la verdadera función de distribución de la población F que al ser mayor que F_0 hará que aparezcan valores más altos de D_n^+, por lo tanto la región crítica habrá de ser de la forma

$$RC = \{D_n^+ > k\}.$$

La determinación de la constante se obtiene de

$$P\{RC|H_0\} \le \alpha$$

con lo que ha de ser $k = d_{n,\alpha}^+$. De manera análoga se justifican los restantes valores.

Los *estadísticos de Kolmogorov-Smirnov* se emplean también para determinar una banda de confianza $1-\alpha$ para la función de distribución de la población.

Dados $1-\alpha$ y n, se puede obtener, véase la Tabla 7 en el Apéndice al final del libro, el valor $d_{n,\alpha}$ para el que se cumple que

$$P\{D_n \leq d_{n\,,\alpha}\} \geq 1-\alpha. \tag{8.4.1.3}$$

Normalmente, se fija un valor ε suficientemente pequeño y se toma el menor tamaño muestral para el que se tiene (8.4.1.3) con $\varepsilon = d_{n,\alpha}$. En cualquier caso

$$\begin{aligned} P\{\sup_x |F_n^*(x) - F_0(x)| \leq d_{n,\alpha}\} & \\ &= P\{x : |F_n^*(x) - F_0(x)| \leq d_{n\,\alpha}\} \\ &= P\{x : F_n^*(x) - d_{n,\alpha} \leq F_0(x) \leq F_n^*(x) + d_{n,\alpha}\} \end{aligned}$$

con lo que al ir variando x, la banda

$$\left[\text{máx}\left\{F_n^*(x) - d_{n,\alpha}, 0\right\}, \text{mín}\left\{F_n^*(x) + d_{n,\alpha}, 1\right\}\right]$$

tiene una probabilidad mayor o igual que $1-\alpha$ de contener, para cualquier x, al valor $F_0(x)$.

Ejemplo 8.4.1.1
Se desea contrastar si la siguiente colección de números

$$20,\ 88,\ 89,\ 69,\ 50,\ 77,\ 98,\ 56,\ 6,\ 41$$

es una m.a.s. de la *Uniforme*$(0, 100)$, con un tamaño de $\alpha = 0'1$. Se puede construir la siguiente tabla:

$x_{(i)}$	$\frac{i}{n}$	$F_0(x_{(i)})$	$\frac{i}{n} - F_0(x_{(i)})$	$\frac{i-1}{n}$	$F_0(x_{(i)}) - \frac{i-1}{n}$
6	0'1	0'06	0'04	0	0'06
20	0'2	0'20	0	0'1	0'10
41	0'3	0'41	-0'11	0'2	0'21
50	0'4	0'50	-0'10	0'3	0'20
56	0'5	0'56	-0'06	0'4	0'16
69	0'6	0'69	-0'09	0'5	0'19
79	0'7	0'79	-0'09	0'6	0'19
88	0'8	0'88	-0'08	0'7	0'18
89	0'9	0'89	0'01	0'8	0'09
98	1	0'98	0'09	0'9	0'08

Aplicando la Expresión (8.4.1.1) se obtiene para $D_n^+ = 0'04$ y mediante (8.4.2.2) es $D_n^- = 0'21$, de donde $D_n = 0'21$. La región crítica será

$$RC = \{D_n > d_{n,0'1}\}$$

y de la Tabla 7 se tiene que $d_{10,0'1} = 0'369$. Como el valor del estadístico es $D_n = 0'21$, éste no pertenece a la región crítica y, por tanto, se mantiene

la hipótesis de que los números tienen distribución uniforme. De hecho el p-valor es $0'769$.

Para determinar una banda de confianza con $1-\alpha = 0'9$ para F_0, basta, de acuerdo con la Tabla 7 del Apéndice, con sumar y restar el valor $0'369$ a la función de distribución empírica, con lo que se obtiene la Figura 8.4.1.1.

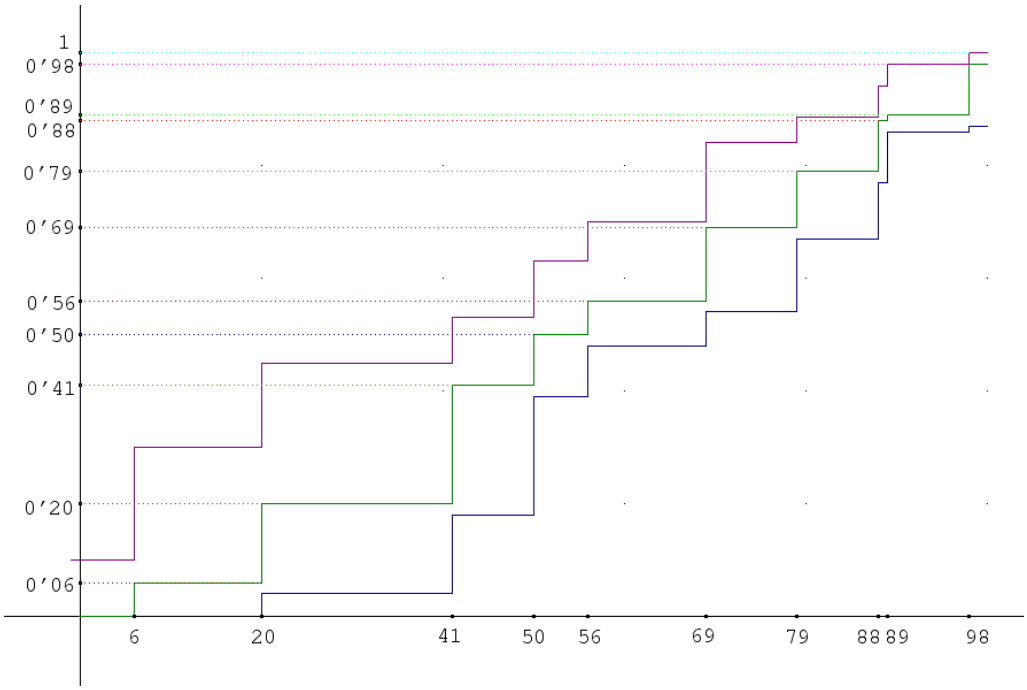


Figura 8.4.1.1. Banda de confianza de probabilidad $0'9$ *para el Ejemplo 8.4.1.1*

■

Una generalización natural de este contraste, es al caso en que se estimen parámetros de la función de distribución F_0 a partir de la muestra, en este contexto, se debe trabajar con un nivel de significación más bajo que el utilizado cuando los parámetros no se estiman a partir de la muestra. Así para la distribución normal, cuando se estiman los parámetros μ y σ^2, a partir de la muestra mediante $\overline{x}$ y s^2, el problema está resuelto. Lilliefors (1967) investigó las propiedades del test de Komogorov-Smirnov modificando, mediante simulación, los niveles críticos del contraste, que pueden verse en la Tabla 8.

Ejemplo 8.4.1.2

Para verificar el ajuste de los 20 datos siguientes a una distribución normal,

con un nivel de significación de 0′05.

86	93'7	106	101'3	88	103'1	96'7	91'2	79	107'9
109'8	101'9	100'7	99'1	104'6	88'3	93	106'9	112'2	117'2

se estiman los parámetros $\widehat{\mu} = \overline{x} = 99'345$ $\widehat{\sigma} = s = 9'76$ y se construye la tabla siguiente:

$x_{(i)}$	$\frac{i}{n}$	$F_0(x_{(i)})$	$\frac{i}{n} - F_0(x_{(i)})$	$\frac{i-1}{n}$	$F_0(x_{(i)}) - \frac{i-1}{n}$
79	0'05	0'019	0'031	0	0'019
86	0'10	0'086	0'014	0'05	0'036
88	0'15	0'128	0'027	0'10	0'023
88'3	0'20	0'129	0'071	0'15	-0'021
91'2	0'25	0'202	0'048	0'20	0'002
93	0'30	0'258	0'042	0'25	0'008
93'7	0'35	0'282	0'068	0'30	-0'018
96'7	0'40	0'393	0'007	0'35	0'043
99'1	0'45	0'490	-0'040	0'40	0'090
100'7	0'50	0'555	-0'055	0'45	0'105
101'3	0'55	0'579	-0'029	0'50	0'079
101'9	0'60	0'603	-0'003	0'55	0'053
103'1	0'65	0'650	0	0'60	0'050
104'6	0'70	0'705	-0'005	0'65	0'055
106	0'75	0'762	-0'012	0'70	0'062
106'9	0'80	0'781	0'019	0'75	0'031
107'9	0'85	0'810	0'040	0'80	0'010
109'8	0'90	0'858	0'042	0'85	0'008
112'2	0'95	0'906	0'044	0'90	0'006
117'2	1	0'966	0'034	0'95	0'016

El valor del *estadístico de Kolmogorov-Smirnov* es $D_{20} = 0'105$, ya que $D_{20}^{+} = 0'071$ y $D_{20}^{-} = 0'105$. De la Tabla 7 es $d_{20;0'05} = 0'294$, que pasa a ser $d_{20;0'05} = 0'192$ si se tiene en cuenta la modificación de Lilliefors, ver Tabla 8. Como $0'105$ no pertence a la región crítica, se acepta la hipótesis de normalidad.

■

Como ya se ha estudiado, el contraste de normalidad, puede hacerse también mediante el *test de la chi-cuadrado*, un argumento a favor del *test de Kolmogorov-Smirnov*, bajo normalidad, puede verse en Kac, Kiefer y Wolfowitz (1955), y consiste en que a tamaños muestrales grandes, el *test de Kolmogorov-Smirnov* tiene mayor potencia.

8.4.2. Contrastes de localización

En la sección anterior se ha tratado el problema del ajuste a una distribución totalmente especificada, frecuentemente no es necesario conocer toda la

función de distribución sino que basta con alguna característica de ésta, por ejemplo, si se puede suponer que el fenómeno bajo estudio sufre únicamente una translación en la distribución pero su forma se mantiene exactamente igual. En estos casos, bastará con hacer hipótesis sobre el cuantil de orden p para tener asegurada la adecuación entre la muestra y la función de distribución.

Se van a estudiar dos tipos de contrastes adecuados a esta situación: el *test de los signos*, válido para cualquier cuantil, y el *test de Wilcoxon* o de los *rangos signados*, que a costa de utilizar una mayor cantidad de la información contenida en la muestra, sólo es válido para contrastar hipótesis sobre el cuantil de orden $0'5$, la mediana.

Test de los signos.

Se supone que la función de distribución de una v.a. es continua, y se pretende contrastar si su cuantil de orden p es un número conocido k, es decir si es $F(k) = p$. Si se resta a los valores de la muestra el valor conocido a contrastar k, la v.a. Y número de diferencias positivas (o de signos positivos) en $x_1-k, ..., x_n-k$ bajo la hipótesis nula, tiene distribución $Binomial(n, 1-p)$, ya que la probabilidad de éxito es

$$P\{X - k > 0\} = 1 - P\{X \leq k\} = 1 - p.$$

Un número de diferencias positivas bajo, indicará que el cuantil será menor que el valor a contrastar, por el contrario, si el número de diferencias positivas es alto, el cuantil será mayor que el valor k que se ha supuesto. Los diferentes contrastes que se pueden realizar quedan recogidas en la Tabla 8.4.2.1.

H_0	H_1	Región Crítica
$k_p = k$	$k_p > k$	$Y \geq c_1$
	$k_p < k$	$Y \leq c_2$
	$k_p \neq k$	$Y \geq c_1$ o $Y \leq c_2$ (con $c_1 > c_2$)

Tabla 8.4.2.1. Regiones críticas del test de los signos.

La existencia de valores iguales a k, que aunque es un suceso de probabilidad cero puede darse en la práctica, se tratan eliminando este valor de la muestra si es único. Si ocurre pocas veces, se asignan arbitrariamente signos positivos y negativos, en la esperanza de que esto no modifique sustancialmente las características del test. La presencia en la muestra de varios valores iguales a k invalidaría la hipótesis hecha, de que la función de distribución es continua, al menos en el valor k.

Para la determinación de las constantes c_1 y c_2 se utiliza la distribución Binomial, así en el primer caso, c_1 ha de ser tal que

$$\alpha \geq P\{Y \geq c_1 | H_0\} = \sum_{i \geq c_1}^{n} p_i$$

donde $p_i = P\{Y = i\}$ con $Y \sim Binomial(n, 1-p)$.

En el contraste bilateral, las constantes c_1 y c_2 se determinan a partir de las expresiones

$$P\{Y \geq c_1 | H_0\} \leq \frac{\alpha}{2} \qquad P\{Y \leq c_2 | H_0\} \leq \frac{\alpha}{2}$$

Si el tamaño muestral es grande, se pueden utilizar las aproximaciones usuales para la binomial; por ejemplo mediante la aproximación normal se puede emplear para Y la distribución $N(n(1-p), \sqrt{n(1-p)p})$.

Ejemplo 8.4.2.1
Se postula que los alumnos de una determinada materia, de un curso en un colegio, han de obtener una calificación de manera que el cuartil de orden 1 sea igual a $3'5$. Si éste es menor, se concluye que el aprovechamiento ha sido poco satisfactorio. Se dispone de las calificaciones de 12 alumnos tomados al azar y éstas son

$$4'5,\ 1,\ 3,8'5,\ 4,\ 0'5,\ 5'5,\ 8,\ 2,\ 2'5,\ 4'5,\ 6.$$

Se trata de contrastar la hipótesis nula $H_0 : k_{0'25} = 3'5$ frente a la hipótesis alternativa $H_1 : k_{0'25} < 3'5$, con un nivel de significación $\alpha = 0'1$, la región crítica es de la forma $\{Y \leq c_2\}$.

Dado que Y tiene distribución aproximada $N(12 \cdot 0'75, \sqrt{12 \cdot 0'75 \cdot 0'25})$, al tipificar la v.a. queda

$$P\{Y \leq c_2\} = P\{Z \leq \frac{c_2 - 9}{1'5}\}$$

siendo Z una v.a. $N(0,1)$, de donde $\frac{c_2-9}{1'5} = -1'2815$ y por tanto $c_2 = 7'0777$.

La región crítica es $RC = \{Y \leq 7'0777\}$ y como Y, número de diferencias positivas, vale 7, se rechaza la hipótesis nula.

Test de Wilcoxon o de los rangos signados

El test de los signos no utiliza toda la información suministrada por la muestra, ya que no tiene en cuenta la magnitud de las desviaciones al valor del cuantil. El test que se va a estudiar ahora considera estas magnitudes, pero a cambio, sólo se pueden hacer contrastes de hipótesis respecto a la mediana y no sobre cualquier otro cuantil.

Para contrastar la hipótesis nula H_0 : la muestra proviene de una distribución simétrica con mediana M_0, frente a la hipótesis alternativa H_1: la mediana M es mayor que M_0, se construyen las variables $D_i = X_i - M_0$, y a sus valores absolutos, $|D_i|$, se les asigna un número de 1 a n de menor a mayor, llamado *rango*. Se llama *estadístico de Wilcoxon* y se denota por T_n^+, a la suma de los rangos correspondientes a valores de las variables D_i positivas.

Analíticamente

$$T_n^+ = \sum_{i=1}^{n} Z_i r(|D_i|) \tag{8.4.2.1}$$

donde

$$Z_i = \begin{cases} 1 & \text{si } D_i > 0 \\ 0 & \text{si } D_i < 0 \end{cases}$$

Si la población tiene distribución continua, la probabilidad de que aparezcan valores tanto con el mismo rango, como iguales a la mediana es cero. Se adoptará el convenio de ignorar los valores iguales a M_0, disminuyendo el tamaño muestral, y se asignará el promedio de los rangos, a los valores que presenten empates. Por ejemplo si la muestra es $X_i : 5, 2, -2, 3, 3, 4, 8$ y la hipótesis es la mediana $M_0 = 4$, se emplea la nueva muestra $X_i : 5, 2, -2, 3, 3, 8$ a la que corresponden los valores $D_i = X_i - M_0 : 1, -2, 2, -1, -1, 4$ con lo que $|D_i| : 1, 2, 2, 1, 1, 4$ y $R_i : 2, 4'5, 4'5, 2, 2, 6$ obtenidos de los rangos $1, 4, 5, 2, 3, 6$ al promediar $1, 2$ y 3 por un lado y 4 y 5 por otro. Algunos autores proponen diferentes soluciones para este problema, que básicamente suponen ser conservadores con la hipótesis nula. Un estudio detallado del tratamiento de los empates puede verse en Lehman (1975) pág. 130.

El contraste, también hubiera podido hacerse, utilizando la suma de los rangos correspondientes a valores negativos, es decir a partir del estadístico

$$T_n^- = \sum_{i=1}^{n} (1 - Z_i) r(|D_i|)$$

ya que si la distribución es continua

$$T_n^+ - T_n^- = \sum_{i=1}^{n} r(|D_i|) = \frac{(n+1)n}{2}. \tag{8.4.2.2}$$

Los distintos contrastes basados en el *estadístico de Wilcoxon* pueden verse en la Tabla 8.4.2.2.

H_0	H_1	Región Crítica
$M = M_0$	$M > M_0$	$T_n^+ > c_1$
	$M < M_0$	$T_n^+ < c_2$
	$M \neq M_0$	$T_n^+ > c_1$ o $T_n^+ < c_2$ con $c_1 > c_2$

Tabla 8.4.2.2. Regiones críticas del test de Wilcoxon.

Para poder determinar las constantes, se necesita conocer la distribución del estadístico T_n^+, bajo la hipótesis nula, para lo que se necesita el siguiente enunciado.

Proposición 8.4.2.1
Dada una v.a. D con distribución simétrica y de mediana 0, el signo y el valor absoluto de D son v.a. independientes.

Demostración: Basta ver que

$$P\{D > 0, |D| < d\} = P\{D > 0, -d < D < d\} = P\{0 < D < +d\}$$

pero, por la hipótesis de simetría

$$P\{0 < D < +d\} = \frac{1}{2}P\{-d < D < d\} = \frac{1}{2}P\{|D| < d\}$$

y por ser la mediana 0

$$P\{D > 0\} = \frac{1}{2}$$

con lo que

$$P\{D > 0, |D| < d\} = P\{D > 0\}P\{|D| < d\}.$$

Como consecuencia, las variables aleatorias signo y rango son independientes, ya que ésta última variable solo es función de $|D|$.

■

Mediante este resultado, con $D = X - M_0$, se puede obtener la distribución de T_n^+, al menos para tamaños muestrales pequeños.

Para ello, bajo la hipótesis nula, se puede considerar la muestra ordenada de menor a mayor de manera que los rangos correspondientes sean 1,2,...,n respectivamente y a continuación, con independencia, se asigna a cada rango un signo + o un signo − con probabilidad $\frac{1}{2}$. De esta forma se puede suponer cada muestra, como una sucesión de n signos, cada uno + o − , asociados a los n primeros números naturales y que tendrá de probabilidad $\left(\frac{1}{2}\right)^n$, ya que

$$P\{+|H_0\} = P\{X_i > M_0|H_0\} = \frac{1}{2}.$$

Por lo tanto, se puede tabular la distribución de T_n^+, al menos para

valores pequeños de n. Así por ejemplo, para $n = 3$, se tiene

(1,	2,	3)	$T_3^+ = t$	$P_{H_0}\{T_3^+ = t\}$
+	+	+	6	1/8
-	+	+	5	1/8
+	-	+	4	1/8
+	+	-	3	1/8
+	-	-	1	1/8
-	+	-	2	1/8
-	-	+	3	1/8
-	-	-	0	1/8

lo que lleva a la distribución

$T_3^+ = t$	0	1	2	3	4	5	6
$P_{H_0}\{T_3^+ = t\}$	1/8	1/8	1/8	2/8	1/8	1/8	1/8

Se observa que la distribución de T_n^+ es simétrica alrededor de $n(n+1)/4$, ya que si se sustituyen los signos + por − y los signos − por +, cualquier muestra compatible con t se corresponde con un valor t' donde

$$t = \sum_{i=1}^{n} Z_i r(|D_i|) \qquad t' = \sum_{i=1}^{n} (1 - Z_i) r(|D_i|)$$

y por tanto

$$t + t' = \sum_{i=1}^{n} i = \frac{n+1}{2} n$$

y de ahí la simetría que se observa respecto al valor $(n + 1)n/4$.

Con el fin de encontrar la distribución de T_n^+ para tamaños muestrales mayores, se emplea que bajo H_0 es

$$P\{T_n^+ = t\} = P\{D_n > 0\} P\{T_n^+ = t | D_n > 0\} + P\{D_n < 0\} P\{T_n^+ = t | D_n < 0\}$$

expresión que se convierte en

$$P\{T_n^+ = t\} = \frac{1}{2} P\{T_{n-1}^+ = t - n\} + \frac{1}{2} P\{T_{n-1}^+ = t\}.$$

Esta ecuación permite calcular la distribución del *estadístico de Wilcoxon* para una muestra de tamaño n, a partir de su distribución para una muestra de tamaño $n - 1$.

Con el fin de obtener la distribución asintótica del *estadístico de Wilcoxon* bajo la hipótesis nula, se puede poner la Expresión (8.4.2.1) en la forma

$$T_n^+ = \sum_{i=1}^{n} i W_i \tag{8.4.2.3}$$

donde W_i vale 1 si la observación que tiene de rango i proviene de una observación X_j mayor que M_0 y vale 0 en caso contrario.

Es claro que bajo la hipótesis nula

$$E[W_i] = P\{X_j > M_0\} = \frac{1}{2}$$
$$V[W_i] = E[W_i^2] - E^2[W_i] = E[W_i] - \frac{1}{4} = \frac{1}{4}$$

y por lo tanto

$$E[T_n^+] = \frac{1}{2}\sum_{i=1}^{n} i = \frac{n(n+1)}{4}$$
$$V[T_n^+] = \frac{1}{4}\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{24}$$

donde es importante tener en cuenta que las variables iW_i son independien-
pero no igualmente distribuidas.

El teorema central del límite permite asegurar que bajo H_0 se tiene la convergencia

$$\frac{T_n^+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \xrightarrow{\mathcal{L}} Z$$

donde Z se distribuye $N(0,1)$.

Debe advertirse que, cuando como resultado del contraste se rechaza la hipótesis nula, puede ser bien porque la mediana no sea el valor supuesto o quizás también porque la distribución no sea simétrica.

Ejemplo 8.4.2.2
Dada la siguiente m.a.s. de una v.a. $N(0,1)$

-0'465 0'120 -0'238 -0'869 -1'016 0'417 0'056 0'561

para contrastar la hipótesis nula $H_0 : M_0 = -1$, frente a la hipótesis alternativa $H_1 : M_0 > -1$, con un nivel de significación $\alpha = 0'01$, se pasa a los valores $|D_i| = |X_i - M_0|$

0'535 1'120 0'762 0'131 |-0'016| 1'417 1'056 1'561

a los que les corresponden los rangos $3, 6, 4, 2, 1, 7, 5, 8$ respectivamente. Se tiene que

$$T_8^+ = \frac{(8+1)8}{2} - 1 = 35$$

y de acuerdo con la Tabla 8.4.2.2, la región crítica es de la forma

$$RC = \{T_8^+ > c_1\}$$

con $c_1 = T_{8,0'01} = 34$. Por lo que, como era de esperar, se rechaza la hipótesis nula.

Si se contrasta ahora el verdadero valor, es decir $M_0 = 0$, se obtienen los rangos $5, 2, 3, 7, 8, 4, 1, 6$ y como el valor del estadístico es $T_8^+ = 13$, se acepta la hipótesis nula.

Ejemplo 8.4.2.3
Dada la m.a.s. de tamaño 10

0'435	0'143	0'889	0'697	0'648
0'482	0'152	0'477	0'741	0'885

Se pretende contrastar, mediante el *test de los signos* y el *de Wilcoxon*, si proviene de una población $Uniforme(0, 1)$, con un nivel de significación $\alpha = 0'05$.

Mediante el *test de los signos*, se trata de contrastar $H_0 : M = k_{\frac{1}{2}} = 0'5$ frente a $H_1 : M \neq 0'5$. La región crítica es de la forma $RC = \{Y < c_2 \text{ o } Y > c_1\}$ con $Y \sim Binomial(10, 0'5)$ de donde si se quiere que

$$P\{Y > c_1 | H_0\} \leq 0'025,$$

y dado que $p_1 + p_{10} = 0'001 + 0'0098 = 0'0108$ y $p_8 + p_9 + p_{10} > 0'025$, se obtiene $c_1 = 8$ y por simetría $c_2 = 2$. Como el número de signos positivos al hacer las diferencias de la muestra a M_0 es 6, se acepta la hipótesis nula.

Aplicando el *test de Wilcoxon* se obtiene

$D_i = X_i - M_0$:	-0'065	-0'357	0'389	0'197	0'148
	-0'018	-0'348	-0'023	0'241	0'385
$\lvert D_i \rvert$:	0'065	0'357	0'389	0'197	0'148
	0'018	0'348	0'023	0'241	0'385
R_i :	3	8	10	5	4
	1	7	2	6	9

de acuerdo con la Tabla 8.4.2.2, la región crítica es de la forma $RC = \{T_n^+ > c_1 \text{ o } T_n^+ < c_2\}$ y por la restricción del tamaño, ha de ser

$$P\{T_{10}^+ > c_1 | H_0\} \leq 0'025$$

con lo que $c_1 = T^+_{10;0'025} = 46$ (véase la Tabla 9), y por simetría $c_2 = \frac{n(n+1)}{2} - 46 = 9$.

Como la suma de los rangos correspondientes a diferencias positivas vale 34, se acepta la hipótesis nula.

■

Con frecuencia el test de Wilcoxon se aplica para ver la homogeneidad de dos poblaciones X e Y, a partir de una m.a.s. de la forma $(X_1, Y_1), ..., (X_n, Y_n)$.

Para ello, se construye la variable diferencia $Z_i = X_i - Y_i$ y se contrasta la hipótesis nula H_0 :es 0 la mediana de la población Z. En general no es cierto que $M_Z = M_X - M_Y$ y en todo caso, ha de suponerse también la hipótesis de simetría para la variable diferencia.

Ejemplo 8.4.2.4
Se escogen aleatoriamente 10 unidades de un determinado proceso de fabricación y se toman dos medidas de cada unidad con un intervalo de tiempo, obteniéndose los siguientes valores:

nº de unidad	1	2	3	4	5	6	7	8	9	10
1^a medida	71	108	72	140	61	97	90	127	101	114
2^a medida	77	105	71	148	78	117	92	131	112	105
diferencia	-6	3	1	-8	-17	-20	-2	-4	-11	9

Para contrastar la hipótesis de que las dos medidas se refieren a la misma magnitud, con un nivel de significación $\alpha = 0'05$, la región crítica viene dada por

$$RC = \{T_{10}^+ > c_1 \ \text{ o } \ T_{10}^+ < c_2, \ \ c_1 > c_2\}.$$

mediante las tablas de la distribución del *estadístico de Wilcoxon* (véase la Tabla 9), cómo ha de ser

$$P\{T_{10}^+ > c_1 | H_0\} \leq 0'025$$
$$P\{T_{10}^+ > c_2 | H_0\} \leq 0'025$$

se obtiene que $c_1 = T^+_{10,0'025} = 46$ y por ser la distribución del estadístico simétrica ha de ser $c_2 = \frac{n(n+1)}{2} - c_1 = 9$. Los rangos correspondientes al valor absoluto de la variable diferencia son

$$5, 3, 1, 6, 9, 10, 2, 4, 8, 7,$$

con lo que se obtiene que $T_n^+ = 11$; luego se acepta la hipótesis de homogeneidad.

8.4.3. Contrastes de normalidad

Por la importancia y el papel central en estadística de la distribución normal, se estudia a continuación el contraste para comprobar si una m.a.s. proviene de una población normal. Como se ha dicho para realizar este contraste son aplicables tanto el modelo multinomial desarrollado en la *Bondad del ajuste* (véase la Sección 8.2.1) como el *contraste de Kolmogorov-Smirnov* (véase la Sección 8.4.1).

El siguiente procedimiento gráfico, basado en la transformación mediante la función de distribución, constituye un método bastante intuitivo aunque no muy preciso, que permite resolver el problema. Basta observar que dada

la m.a.s. $(x_1, ..., x_n)$, la nueva muestra $(F(x_1), ..., F(x_n))$, con F la función de distribución de la v.a. $N(\mu, \sigma)$, tiene distribución uniforme (véase la Sección 1.6.3); por lo tanto la imagen de la muestra ordenada puede considerarse como una muestra ordenada $(u_{(1)}, ..., u_{(n)})$ de la distribución $U(0, 1)$, y de acuerdo con la propiedad de los recubrimientos (véase la Sección 1.5.3, Expresión (1.5.3.1)) la esperanza de los recubrimientos es $\frac{1}{n+1}$. Esto implica, que los puntos $(x_{(i)}, F^{-1}(\frac{i}{n+1}))$ deberán estar aproximadamente sobre la bisectriz del primer cuadrante, ya que se ha sustituido el valor de una variable por su esperanza, y ésto para toda F función de distribución continua. En particular, los puntos $\left(\frac{x_{(i)}-\mu}{\sigma}, \Phi^{-1}\left(\frac{i}{n+1}\right)\right)$, siendo Φ la función de distribución de la v.a. $N(0, 1)$, estarán aproximadamente sobre la bisectriz del primer cuadrante, es decir

$$\Phi^{-1}\left(\frac{i}{n+1}\right) \simeq \frac{x_{(i)} - \mu}{\sigma}$$

o lo que es lo mismo

$$x_{(i)} \simeq \mu + \sigma\Phi^{-1}\left(\frac{i}{n+1}\right).$$

En resumen, si la m.a.s. $(x_1, ..., x_n)$ proviene de una población normal, los puntos $(x_{(i)}, \Phi^{-1}(\frac{i}{n+1}))$ han de estar alineados. La ordenada en el origen sería μ y la pendiente de la recta σ. Esto permite además, al menos de una forma rudimentaria, estimar μ y σ, ya que en general estos parámetros no serán conocidos. Un ejemplo ayudará a entender la forma de proceder.

Ejemplo 8.4.3.1
Los siguientes datos están tomados de Stigler (1977) y se refieren a los experimentos realizados por Michelson, entre el 5 de Junio de 1879 y el 2 de Julio de 2879, para determinar de manera experimental, la velocidad de la luz. Se ha restado 299.000 a cada dato

890	900	810	1070	820
720	800	770	950	760
980	740	920	1000	910
850	650	880	830	840
790	860	960	780	870

Con el fin de contrastar su ajuste a una distribución normal de paráme-

tros convenientes, se construye la siguiente Tabla:

i	1	2	3	4	5	6	7
$x_{(i)}$	650	720	740	760	770	780	790
$i/(n+1)$	0'0385	0'0769	0'1154	0'1538	0'1923	0'2308	0'2692
$\Phi^{-1}(i/(n+1))$	-1'7684	-1'4262	-1'1983	-1'0203	-0'8694	-0'7362	-0'6152
	8	9	10	11	12	13	14
	800	810	820	830	840	850	860
	0'3077	0'3461	0'3846	0'4231	0'4615	0'5000	0'5385
	-0'5024	-0'3959	-0'2934	-0'1940	-0'0966	0	0'0966
	15	16	17	18	19	20	21
	870	880	890	900	910	920	950
	0'5769	0'6154	0'6538	0'6923	0'7308	0'7692	0'8077
	0'1940	0'2934	0'3959	0'5024	0'6152	0'7362	0'8694
	22	23	24	25			
	960	980	1000	1070			
	0'861	0'8846	0'9231	0'9615			
	1'0203	1'1983	1'4263	1'7484			

Por comodidad se acostumbra a representar $(x_{(i)}, 100\Phi^{-1}(i/(n+1))$, lo que permite construir el diagrama contenido en la Figura 8.4.2.

El problema es cuantificar numéricamente el ajuste, para ver si debe ser aceptada la suposición de normalidad. Esto se puede hacer calculando el *coeficiente de correlación muestral* que en este caso vale 0'994 (véase la Expresión (1.2.6)), lo que pone de manifiesto un buen ajuste. Además, mediante la determinación de la recta de regresion de 100 $\Phi^{-1}(i/(n+1))$ sobre $x_{(i)}$, que es

$$100\ \Phi^{-1}(i/(n+1)) = -802'438 + 0'9396\, x_{(i)}$$

(véase la Expresión (7.4.1.4)), se puede determinar, sin recurrir al método gráfico, que $\widehat{\mu} = 85'402$ y $\widehat{\sigma} = 1006'428$.

En resumen, el ajuste a la normal es bueno y se puede afirmar que a partir de los experimentos de Michelson, se hubiera dado como estimador de la velocidad de la luz el valor 299,085'402, obtenido sumando 299,000 a la media estimada.

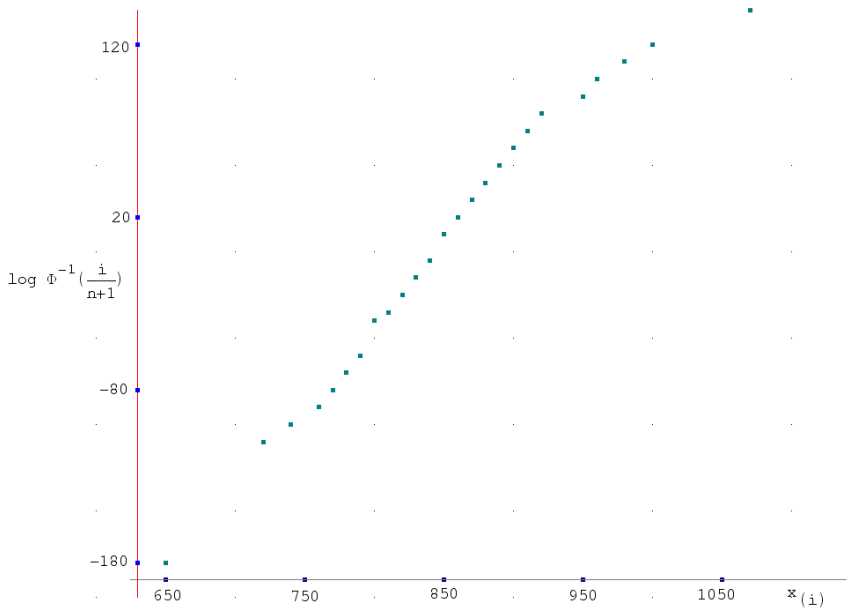


Figura 8.4.2. Diagrama de ajuste a la distribución normal para los datos del Ejemplo 8.4.3.2.

8.4.4. Valores discordantes

Con frecuencia al estudiar una muestra aparecen observaciones que no son homogéneas con las restantes en su orden de magnitud. A tales observaciones se les da el nombre de *discordantes* o *atípicas* y el primer problema es decidir si una observación puede ser clasificada como tal. Una solución puede ser utilizar la desigualdad de Tchebycheff y suponer que una observación es *discordante* si dista de la media muestral en valor absoluto, más de tres veces la desviación típica muestral. Pero la media muestral depende fuertemente de la observación *discordante* por lo que se prefiere utilizar una medida de centralización que no dependa tanto de las observaciones extremas, tal como la mediana. Un test simple y razonable para clasificar una observación muestral x_i como *discordante* es ver si cumple

$$|x_i - M|/med(|x_i - M|) > 5 \qquad (8.4.4.1)$$

donde M es la mediana muestral. El denominador de ésta expresión, la mediana de las desviaciones en valor absoluto a la mediana muestral, es una medida de dispersión. El siguiente ejemplo explica la manera de proceder

Ejemplo 8.4.4.1
Los siguientes datos

3'7 6'9 6'2 8'9 7'2 5'4 2'8 17'2 13'7

son analizados para ver si puede ser considerado como observación discordante el valor 17'2. En efecto, la mediana muestral es $M = 6'9$, con lo que se tiene la siguiente tabla:

x_{ij}	2'8	3'7	5'4	6'2	6'9	7'2	6'9	13'7	17'2
$d_i = \lvert x_{(i)} - M \rvert$	4'1	3'2	1'5	0'7	0	0'3	2	6'8	10'3
$d_{(i)}$	0	0'3	0'7	1'5	2	3'2	4'1	6'8	10'3

la $med(d_i) = 2$ y el único valor que cumple que $|x_i - 6'9| > 5'2$ es 17'2, por lo que puede ser considerado discordante.
■

Una vez detectada la existencia de un valor *atípico* surge el problema de qué se debe hacer con él se le podría eliminar de la muestra, otra posibilidad es medir la influencia que tiene en el procedimiento estadístico que se esté empleando. Esta segunda idea llevó a Huber (1972) a la introducción de los *m-estimadores*. Un estudio de los *m-estimadores* así como de las *funciones de influencia* puede verse en (Randle y Wolfe (1979), Sección 7.4).

8.4.5. Estimación no paramétrica de densidades

Lo mismo que la función de distribución empírica aproxima a la función de distribución de la población, surge de manera natural la pregunta ¿cómo se puede estimar la función de densidad de la población? un candidato natural es el histograma, obtenido dividiendo el recorrido de la variabla aleatoria en un número adecuado de clases y poniendo como alturas números proporcionales a las frecuencias observadas en la muestra. Si la amplitud de cada clase no es la misma, el histograma puede dar formas bastantes diferentes, para evitar esto y también para hacer que el área del histograma valga la unidad, se supondrán todas las amplitudes de las clases iguales a h y las alturas iguales a $\frac{n_i}{nh}$, con n_i las frecuencias observadas en cada clase y n el tamaño de la m.a.s. observada. Dos problemas principales aparecen en este procedimiento: la función que se obtiene no es derivable y además depende fuertemente de la amplitud h con que se escojan las clases.

Ejemplo 8.4.5.1 (Silverman (1996))
Los datos de la Tabla 8.4.5.1 recogen el número de días que duró el tratamien-

to psicológico suministrado a 86 pacientes.

1	1	1	5	7	8	8	13	14	14	17	18	21
21	22	25	27	27	30	30	31	31	32	34	35	36
37	38	39	39	40	49	49	54	56	56	62	63	65
65	67	75	76	79	82	83	84	84	84	90	91	92
93	93	103	103	111	112	119	122	123	126	129	134	144
147	153	163	167	175	228	231	235	242	256	256	257	311
314	322	369	415	573	609	640	737					

Tabla 8.5.4.1. Días de tratamiento dado a 86 pacientes en un estudio psicológico.

Como se puede ver existe una notable discrepancia entre la representación con la amplitud del intervalo $h = 50$, Figura 8.4.5.1 (a), y la amplitud $h = 75$, Figura 8.4.5.2 (b).

Como norma general, a medida que la amplitud de banda aumenta, la forma que presentan los datos representados tiende a difuminarse.

■

Como alternativa al histograma de rectángulos, puede construirse un estimador $\widehat{f}(x)$ de la función de densidad, que sea una función continua, observando que la función de densidad verifica

$$f(x) = \lim_{h \to 0} \frac{1}{2h} P\{x - h < X \leq x + h\}$$

para un valor de h fijo, parece lógico estimar $P\{x - h < X_i \leq x + h\}$ por la proporción de puntos muestrales que caen dentro del intervalo $(x - h, x + h]$. Entonces un estimador natural $\widehat{f}$ de la densidad puede venir dado, si se fija un valor pequeño para h, por

$$\widehat{f}(x) = \frac{1}{2hn} \sum_{i=1}^{n} I_{(x-h,x+h]}(x_i) \tag{8.4.5.1}$$

o alternativamente mediante

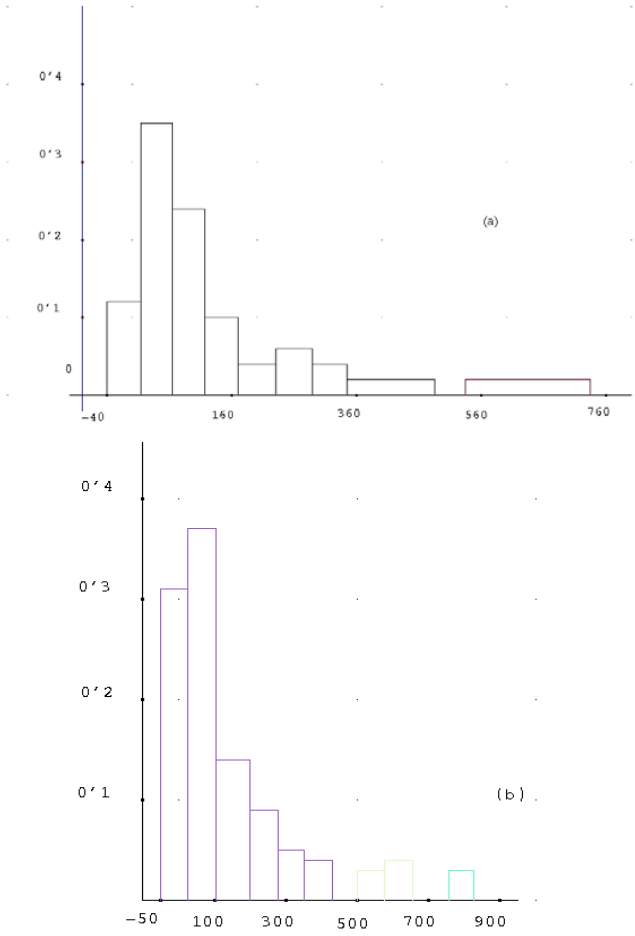


Figura 8.4.5.1. Representación del histograma de frecuencias correspondiente al Ejemplo 8.4.5.1 para (a) una amplitud de banda $h = 50$ *(b) una amplitud de banda* $h = 75$.

$$\widehat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2h}I_{(x-h,x+h]}(x_i) \tag{8.4.5.2}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}w\left(\frac{x-x_i}{h}\right) \tag{8.4.5.3}$$

donde

$$w(y) = \begin{cases} \frac{1}{2} & \text{si } |y| < 1 \\ 0 & \text{resto} \end{cases} \tag{8.4.5.4}$$

La Expresión (8.4.5.2) proporciona la clave para la obtención de una expresión continua, ya que obedece a colocar sobre cada punto muestral x_i una caja de base $2h$ y de altura $(2h)^{-1}$ y después promediar en todas las observaciones. Se puede tomar como $w(y)$ una función no negativa, simétrica, centrada en cero y que integre 1. De esta forma

$$w_h(y) = \frac{1}{h}\, w\left(\frac{y}{h}\right)$$

con lo que definitivamente es

$$\widehat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} w_h(x - x_i) \tag{8.4.5.5}$$

A este estimador se le llama *estimador núcleo* de la función de densidad, a la función w *función núcleo* y al parámetro h *amplitud de banda* o *factor de suavizado.*

Una posibilidad, ampliamente usada en la práctica, es tomar como función w la densidad de la distribución $N(0,1)$, con lo que $w_h(x - x_i)$ es la función de densidad de la distribución $N(x_i, h)$.

En las aplicaciones es conveniente realizar la estimación con diferentes valores de *amplitud de banda*, en general si el valor de h es demasiado pequeño la estimación será muy abrupta, mientras que si es demasiado grande la función será muy suave.

Un estudio de la estimación no paramétrica de densidades puede verse en el libro de Silverman (1986). Una revisión de problemas en relación con la elección del *factor de suavizado* puede verse en Devroye (1997).

Ejemplo 8.4.5.2
Para los datos del Ejemplo 8.4.5.1, la Figura 8.4.5.2 recoge los *estimadores núcleo* con *amplitudes de banda* de 20 y 60 días, respectivamente. Se observa que la primera amplitud no suaviza lo suficiente a partir del valor correspondiente a los 200 días, mientras que la segunda suaviza demasiado esta parte de la densidad y además acorta el efecto del cuerpo principal de la densidad.

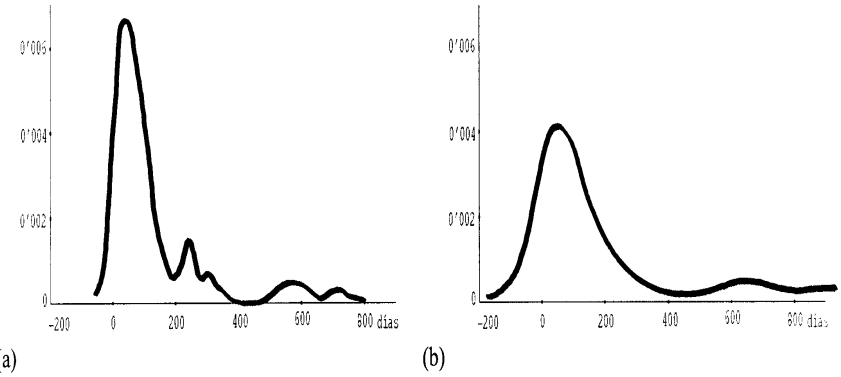


Figura 8.4.5.2. Estimación de la función de densidad para los datos del Ejemplo 8.4.5.1 para (a) una amplitud de banda de 20 días (b) una amplitud de banda de 60 días.

8.5. PROBLEMAS NO PARAMÉTRICOS RELATIVOS A DOS MUESTRAS

Los problemas que son tratados dentro de este Epígrafe, tienen como hipótesis nula que dos m.a.s. e independientes entre sí, provienen de una misma distribución, que se supone es de la clase de las funciones de distribución continuas. La hipótesis alternativa es que o bien ambas funciones de distribución son distintas o bien que una de las variables *domina estocásticamente* a la otra. Se dice que la v.a. Y domina estocásticamente a la v.a. X si $F_X(x) \geq F_Y(x)$.

8.5.1. Contrastes de Kolmogorov-Smirnov para dos muestras

Dados los valores de dos m.a.s. $(x_1, ..., x_m)$ e $(y_1, ..., y_n)$ independientes de dos poblaciones con funciones de distribución $F_X(x)$ y $F_Y(x)$ respectivamente, se define el *estadístico de Kolmogorov-Smirnov bilateral*, mediante

$$D_{m,n} = \sup_x |F_m^*(x) - G_n^*(x)| \tag{8.5.1.1}$$

y los *estadísticos de Kolmogorov-Smirnov unilaterales*, mediante

$$D_{m,n}^+ = \sup_x (F_m^*(x) - G_n^*(x)) \tag{8.5.1.2}$$

$$D_{m,n}^- = \sup_x (G_n^*(x) - F_m^*(x)) \tag{8.5.1.3}$$

donde

$$F_m^*(x) = \frac{1}{m}\sum_{i=1}^{m} I_{(-\infty,x]}(x_i)$$
$$G_n^*(x) = \frac{1}{n}\sum_{j=1}^{n} I_{(-\infty,x]}(y_j)$$

De forma análoga al caso de una muestra única, $D_{m,n}$, $D_{m,n}^+$ y $D_{m,n}^-$ tienen distribución libre. Ya que mediante el cambio $u = F(x)$ se tiene que

$$D_{m,n} = \sup_{u\in[0,1]} |F_m^*(F^{-1}(u)) - G_n^*(F^{-1}(u))| \tag{8.5.1.4}$$

y

$$F_m^*(F^{-1}(u)) = \frac{\text{n}^o \text{ de valores } x_i \le F^{-1}(u)}{m} = \frac{\text{n}^o \text{ de valores } F(x_i) \le u}{m}$$

expresión que es la función de distribución empírica correspondiente a una m.a.s. de tamaño m, de una población *Uniforme*$(0,1)$. Análogamente $G_n^*(F^{-1}(u))$ es la función de distribución empírica correspondiente a una m.a.s. de tamaño n de una población *Uniforme*$(0,1)$, por lo tanto el estadístico de la Expresión (8.5.1.1) tiene distribución independiente de F_X y F_Y y lo mismo es válido para los restantes estadísticos.

Los distintos contrastes quedan recogidos en la Tabla 8.5.1.1, donde el valor de las constantes está expresado con la notación usual.

H_0	H_1	Región Crítica
$F_X = F_Y$	$F_X \neq F_Y$	$D_{m,n} > d_{m,n;\alpha}$
	$F_X > F_Y$	$D_{m,n}^+ > d_{m,n;\alpha}^+$
	$F_X < F_Y$	$D_{m,n}^- > d_{m,n;\alpha}^-$

Tabla 8.5.1.1. Regiones críticas del test de Kolmogorov-Smirnov para dos muestras.

La idea es que, si por ejemplo, la variable Y domina estocásticamente a la X, es decir, $F_Y(x) \le F_X(x)$, cabe esperar en ambas muestras, que los valores de la v.a. X tengan tendencia a aparecer a la derecha de los valores de la v.a. Y, por lo que de acuerdo con el *teorema de Glivenko-Cantelli* cabrá esperar valores de $D_{m,n}^+$ grandes. Por el contrario, si la hipótesis nula es cierta $D_{m,n}^+$ ha de ser muy próximo a cero.

Realmente no es necesario utilizar el estadístico $D_{m,n}^-$ ya que si se quiere como hipótesis alternativa $F_X < F_Y$ intercambiando la muestra X con la muestra Y se puede trabajar con $D_{m,n}^+$.

La determinación de la distribución exacta de $D_{m,n}$, bajo la hipótesis nula se atribuye a Gnedenko (1954); y está basada en técnicas de recorridos aleatorios, puede verse en Hájek y Sidak (1967), pág. 141. La distribución asintótica cuando m y n crecen, mientras el cociente se mantiene constante, se debe a Smirnov (1939), una demostración puede verse en el mencionado libro de Hájek y Sidak (1967), pág. 189.

Para valores pequeños, puede emplearse el siguiente método recursivo para determinar la distribución exacta bajo la hipótesis nula

$$P\left\{D_{m,n} \leq \tfrac{k}{mn} | H_0\right\} = P\left\{\sup_x |F_m^*(x) - G_n^*(x)| \leq \tfrac{k}{mn} \middle| H_0\right\} =$$
$$P\left\{(X_1, ..., X_m) \text{ e } (Y_1, ..., Y_n) \text{ con } \sup_x |F_m^*(x) - G_n^*(x)| \leq \tfrac{k}{mn} \middle| H_0\right\}$$

Cada muestra conjunta ordenada se puede considerar como una poligonal Γ con un segmento por cada observación, que tiene su vértice inicial en el punto (0,0) y en la que cada vértice genérico (x, y) se enlaza; bien con el punto $(x+1, y)$, si la observación corresponde a la primera muestra, o bien con el punto $(x, y+1)$, si la observación se corresponde con la segunda; así hasta terminar en el vértice (m, n). Por ejemplo la muestra conjunta $x_{(1)} < y_{(1)} < y_{(2)} < x_{(2)} < x_{(3)} < y_{(3)} < y_{(4)}$ se corresponde con la poligonal de la Figura 8.5.1.1.

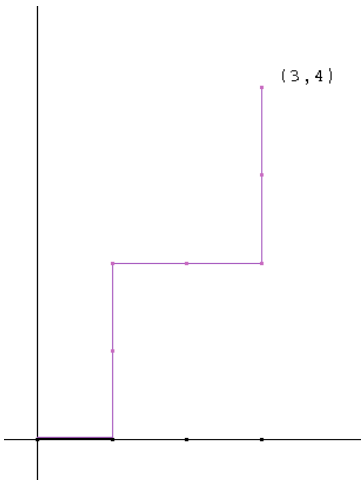


Figura 8.5.1.1. Poligonal asociada a la muestra conjunta $x_{(1)} < y_{(1)} < y_{(2)} < x_{(2)} < x_{(3)} < y_{(3)} < y_{(4)}$.

A medida que x recorre la recta real, $(mF_m^0(x), nG_n^*(x))$ va recorriendo los valores (u, v) correspondientes a las coordenadas de los vértices de la

poligonal. Por lo tanto

$$P\left\{ D_{m,n} \leq \frac{k}{mn} \middle| H_0 \right\} = P\left\{ \text{máx} \left| \frac{u}{m} - \frac{v}{n} \right| \leq \frac{k}{mn} \middle| H_0 \right\} =$$
$$= P\left\{ \underset{(u,v)}{\text{máx}} \frac{1}{n} \left| \frac{n}{m} u - v \right| \leq \frac{k}{mn} \middle| H_0 \right\}$$

donde el último máximo está calculado cuando los valores (u, v) se van moviendo por los vértices de la poligonal. Simplificando $\frac{1}{n}$ y cómo la diferencia $\frac{n}{m}u - v$ representa la distancia en vertical del vértice (u, v) a la recta que une el origen y el punto (m, n), se tiene que

$$P\left\{ D_{m,n} \leq \frac{b}{mn} \middle| H_0 \right\} = \frac{A_k(m,n)}{\binom{m+n}{m}}$$

con $A_k(m, n)$ el número de poligonales que unen el origen con el punto (m, n) y que están a distancia vertical de dicha recta menor o igual que $\frac{k}{m}$. El número combinatorio del denominador se obtiene de considerar que bajo la hipótesis nula ambas muestras tienen la misma distribución, por lo que una configuración de m valores de la primera muestra y n de la segunda tiene de probabilidad $\binom{m+n}{m}^{-1}$.

Es inmediato comprobar la siguiente expresión recursiva

$$A_k(u, v) = A_k(u - 1, v) + A_k(u, v - 1)$$

donde se supone que $A_k(u, v) = 0$ para los puntos (u, v) cuya distancia vertical a la recta que une el origen con (m, n) sea mayor que $\frac{k}{m}$. Este razonamiento se traslada fácilmente para calcular la distribución de $D^+_{m,n}$.

Ejemplo 8.5.1.1
Para $m = 3$ $n = 4$ puede verificarse que:

x	$\frac{3}{14}$	$\frac{4}{12}$	$\frac{5}{12}$	$\frac{6}{12}$	$\frac{8}{12}$	$\frac{9}{12}$	$\frac{12}{12}$
$P\{D_{3,4} \leq x\}$	$\frac{1}{35}$	$\frac{4}{35}$	$\frac{12}{35}$	$\frac{21}{35}$	$\frac{27}{35}$	$\frac{33}{35}$	$\frac{35}{35}$

donde para el cálculo de la distribución de $D_{3,4}$ se tiene

$$P\left\{ D_{3,4} \leq \frac{6}{12} \middle| H_0 \right\} = P\left\{ \underset{(u,v)}{\text{máx}} \frac{1}{4} \left| \frac{4}{3} u - v \right| \leq \frac{6}{3 \cdot 4} \middle| H_0 \right\}$$
$$= P\left\{ \underset{(u,v)}{\text{máx}} \left| \frac{4}{3} u - v \right| \leq 2 \middle| H_0 \right\}$$

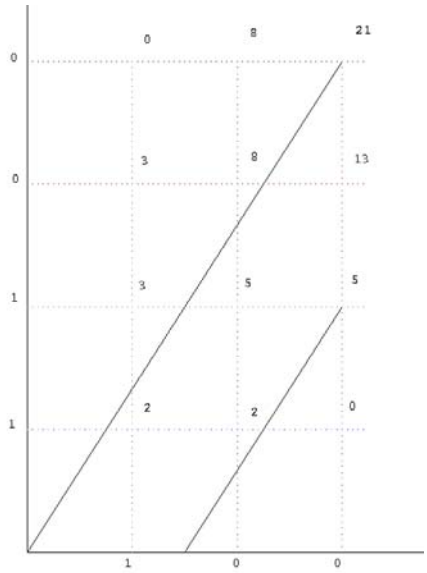


Figura 8.5.1.2. Cálculo de $A_k(3,4)$.

se construye ahora una banda de una altura de 2 unidades por encima y por debajo de la diagonal y se asigna 0 a todos los elementos que quedan fuera de la banda. El número que se asigna a cada vértice dentro de la banda es la suma de los dos números que quedan inmediatamente a su izquierda y por debajo, véase la Figura 8.5.1.2.

■

Para tamaños pequeños y con muestras de igual dimensión, es posible calcular de forma explícita la distribución de $D_{n,n}$ y de $D^+_{n,n}$ bajo la hipótesis nula. Basta observar que

$$P\left\{ D^+_{n,n} \leq \frac{k}{n} \middle| H_0 \right\} = P\left\{ \text{máx}\left(\frac{u}{n} - \frac{v}{n}\right) \leq \frac{k}{n} \middle| H_0 \right\}$$
$$= P\{\text{máx}(u-v) \leq k | H_0\}.$$

Si se asocia a cada muestra conjunta ordenada, una poligonal de lados paralelos a las diagonales del primer y del cuarto cuadrante; según que el valor se corresponda con la primera o segunda muestra respectivamente, es posible calcular, de forma explícita, la distribución de probabilidad.

A título de ejemplo la muestra conjunta $x_{(1)} < y_{(1)} < y_{(2)} < x_{(2)} < x_{(3)} < y_{(3)}$ se corresponde con la poligonal de la Figura 8.5.1.3.

Si se emplea la notación $D_{n,n} = \text{máx}(\frac{u}{n} - \frac{v}{n})$, como $u - v$ es precisamente la ordenada correspondiente a cada vértide de la poligonal, se tiene

$$P\left\{ D_{n,n} \leq \frac{k}{n} \middle| H_0 \right\} = P\left\{ \text{máx}\left(\frac{u}{n} - \frac{v}{n}\right) \leq \frac{k}{n} \middle| H_0 \right\}$$

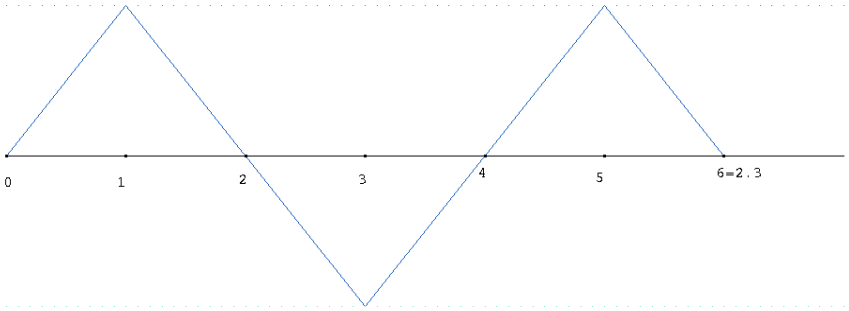


Figura 8.5.1.3. Poligonal asociada a la muestra conjunta $x_{(1)} < y_{(1)} < y_{(2)} < x_{(2)} < x_{(3)} < y_{(3)}$.

que puede ponerse como

$$P\{\text{máx}(u-v) \leq k | H_0\} = \frac{B_k(n)}{\binom{2n}{n}}$$

siendo $B_k(n)$, el número de poligonales de longitud $2n$ que unen el origen con el punto $(2n, 0)$ sin subir más arriba de la ordenada k.

El número de poligonales que unen el origen con el punto $(2n, 0)$ sin alcanzar la ordenada $k+1$, coincide con el número total de poligonales entre los citados puntos, menos las que los unen alcanzando el valor $k+1$. Estas últimas poligonales se pueden contar, ya que hay el mismo número de las que unen el origen con el punto $(2n, 0)$ tocando $k+1$, que las que unen el punto $(0, 2k+2)$ con el $(2n, 0)$. Para ver esto último basta reflejar, respecto a la recta $k+1$, la parte de poligonal que va del punto $(0, 0)$ hasta tocar a la ordenada $k+1$. Como las poligonales que unen $(0, 2k+2)$ con $(2n, 0)$ tienen i tramos ascendentes y j tramos descendentes, ha de cumplirse

$$\left.\begin{array}{c} 2k+2+i-j=0 \\ i+j=2n \end{array}\right\}$$

esto lleva a que $j = n+k+1$ y por lo tanto

$$P\left\{ D_{n,n}^{+} \leq \frac{k}{n} \middle| H_0 \right\} = 1 - \frac{\binom{2n}{n+k+1}}{\binom{2n}{n}}$$

ya que, por la hipótesis nula, las configuraciones posibles de longitud $2n$ de la muestra conjunta son $\binom{2n}{n}$.

Ejemplo 8.5.1.2
Para $n = 3$ las configuraciones posibles son:

	$u - v$	$u - v$
X X X Y Y Y	3	0
X X Y X Y Y	2	0
X X Y Y X Y	2	0
X X Y Y Y X	2	1
X Y X X Y Y	2	0
X Y X Y X Y	1	0
X Y X Y Y X	1	1
X Y Y X X Y	1	1
X Y Y X Y X	1	1
X Y Y Y X X	1	2

La segunda columna que lleva por encabezamiento $u - v$ se obtiene cambiando los valores de X por los de Y.

La distribución

d_{33}^+	0	$\frac{1}{3}$	$\frac{2}{3}$	1
$P\{D_{33}^+ = d_{33}^+\}$	$\frac{5}{20}$	$\frac{9}{20}$	$\frac{5}{20}$	$\frac{1}{20}$

Se puede obtener directamente observando que

$$P\left\{ D_{33}^+ \leq \frac{k}{3} \middle| H_0 \right\} = 1 - \frac{\binom{6}{4+k}}{\binom{6}{3}}$$

Las distribuciones asintóticas bajo H_0 de los *estadísticos de Kolmogorov-Smirnov* vienen dadas por

$$\lim_{n,m\to\infty} P\left\{ \sqrt{\frac{m+n}{mn}} D_{mn} < x \right\} = 1 - 2\sum_{i=1}^{\infty} (-1)^{i+1} e^{-2i^2x^2} \qquad x > 0$$

$$\lim_{n,m\to\infty} P\left\{ \sqrt{\frac{m+n}{mn}} D_{mn}^+ < x \right\} = 1 - e^{-2x^2} \qquad x > 0$$

puede verse en Hájek & Sidak (1967) pág 190.

La distribución exacta de los estadísticos $D_{m,n}^+$ y $D_{m,n}^-$ es la misma (véase Hájek (1969), pág. 80).

Ejemplo 8.5.1.3
Para comparar la duración de dos marcas de neumáticos, se usan hasta que se gastan obteniéndose los siguientes recorridos en miles de kilómetros:

A	17'8	19'6	19'9	20'6	21'4	28'4	30'1	32'1
B	16'9	17'9	18'7	19'8	27'6	27'9	30'8	34'1

Si se quiere contrastar la hipótesis $H_0 : F_X = F_Y$ frente a la hipótesis alternativa $H_1 : F_X \neq F_Y$ con un nivel de significación $\alpha = 0'05$ mediante el *test de Kolmogorov-Smirnov*, se construye la siguiente tabla:

	$n\lvert F_n^*(x) - G_n^*(x)\rvert$		$n\lvert F_n^*(x) - G_n^*(x)$
$y_{(1)} = 16'9$	1	$x_{(5)} = 21'4$	1
$x_{(1)} = 17'8$	0	$y_{(5)} = 27'6$	0
$y_{(2)} = 17'9$	1	$y_{(6)} = 27'9$	1
$y_{(3)} = 18'7$	2	$x_{(6)} = 28'4$	0
$x_{(2)} = 19'6$	1	$x_{(7)} = 30'8$	1
$y_{(4)} = 19'8$	2	$y_{(7)} = 30'8$	0
$x_{(3)} = 19'9$	1	$\mathrm{x}_{(8)} = 32'1$	1
$x_{(4)} = 90'6$	0	$y_{(8)} = 34'1$	0

La región crítica es de la forma $RC = \{D_{8,8} > d_{8,8;0'05} = \frac{5}{8}\}$ (véase la Tabla 11 al final del libro). Al ser el valor del estadístico $D_{8,8} = \frac{2}{8}$, que no pertenece a la RC se mantiene H_0.

8.5.2. Test de la mediana

El *test de Kolmogorov-Smirnov* está indicado para contrastar hipótesis alternativas muy generales. Si se tiene evidencia de que existe un parámetro de localización, es decir $F_Y(x) = F_X(x-\theta)$, con θ desconocido, se sigue que las distribuciones de las dos poblaciones son trasladadas la una de otra y entonces se puede utilizar el *test de la mediana* para contrastar si ambas poblaciones son en realidad la misma.

Si se consideran dos m.a.s. $(X_1, ..., X_m)$ e $(Y_1, ..., Y_n)$, cada una de una población y se ordenan conjuntamente, se obtiene $(Z_{(1)}, Z_{(2)}, ..., Z_{(n+m)})$, la mediana muestral M es el valor $\frac{1}{2}(Z_{(k)} + Z_{(k+1)})$ si N es par o el valor $Z_{(k+1)}$ si N es impar, siendo $N = m + n$ y $k = \left[\frac{N}{2}\right]$ (véase la Definición 1.3.2.1). Esto se concreta en que hay $k = \left[\frac{N}{2}\right]$ observaciones menores o iguales a M tanto si N es par como impar. Si se introduce el estadístico T, número de observaciones de la primera muestra que son menores o iguales que M, es intuitivo el contenido de la Tabla 8.5.2.1.

H_0	H_1	Región Crítica
$M_X = M_Y$	$M_X < M_Y (\theta > 0)$	$T > c_1$
	$M_X > M_Y (\theta < 0)$	$T < c_2$
	$M_X \neq M_Y (\theta \neq 0)$	$T > c_1$ o $T < c_2$ (con $c_1 > c_2$)

Tabla 8.5.2.1. Regiones críticas del test de la Mediana para dos muestras.

En concreto, si T es un valor alto, próximo a M, esto apoyaría una configuración parecida, al ordenar las muestras de menor a mayor, a la

$XX...XYY...Y$ donde X significa un valor de la primera muestra e Y de la segunda, y, por tanto, tal que $M_X < M_Y$. De manera análoga se justifican los otros contrastes.

Para la determinación de las constantes, se utiliza que bajo H_0, T se distribuye con la ley $Hipergeométrica(m+n, m, k)$, así por debajo de la mediana ha de haber t observaciones de la primera muestra y $k-t$ de la segunda, por lo que

$$P\{T=t|H_0\} = \frac{\binom{m}{t}\binom{n}{k-t}}{\binom{m+n}{k}}$$

con

$$t = \text{máx}\{0, k-n\}, ..., \text{mín}\{m, k\}$$

Como aproximaciones, se pueden emplear las correspondientes a la disribución *Hipergeométrica*, es decir, si $m, n \to \infty$ con $\frac{m}{n} \to \theta$ se utiliza la v.a. $Binomial(k, \frac{m}{n})$. También se puede utilizar la aproximación normal

$$\frac{T - \frac{mk}{n}}{\sqrt{mnk(N-k)}} N\sqrt{N-1} \xrightarrow{\mathcal{L}} Z \sim N(0,1).$$

Ejemplo 8.5.2.1
Los datos siguientes representan la duración en horas de dos tipos diferentes de baterías:

Tipo 1:	40'1	30'5	40'2	45	55'5	30'6
Tipo 2:	50'5	50'7	45'5	55'7	60	40'8

Para contrastar su diferencia mediante un test de tamaño $\alpha = 0'05$, se puede utilizar el contraste de la mediana con $H_0 : M_X = M_Y$ ó $F_X = F_Y$ frente a $H_1 : M_X < M_Y$ o bien $F_X > F_Y$, ya que cabe suponer la existencia de un parámetro de localización. La región crítica es de la forma $RC = \{T > c_1\}$. La muestra conjunta reordenada es

$$\underline{30'5} < \underline{30'6} < \underline{40'1} < \underline{40'2} < 40'8 < \underline{45} < 45'5 < 50'1 < 50'7 < \underline{55'5} < 55'7 < 60$$

donde los valores subrayados pertenecen a la primera muestra.

Como $m+n = 12$ la mediana muestral es $M = (45 + 45'5)/2 = 45'25$ y el estadístico T toma el valor 5. Dado que

$$P\{T{=}6|H_0\} = \frac{\binom{6}{6}\binom{6}{0}}{\binom{12}{6}} = 0'0011$$

$$P\{T{=}5|H_0\} = \frac{\binom{6}{5}\binom{6}{1}}{\binom{12}{6}} = 0'0389$$

$$P\{T{=}4|H_0\} = \frac{\binom{6}{4}\binom{6}{2}}{\binom{12}{6}} = 0'2435$$

se tiene que $P\{T > 4|H_0\} = 0'039 < 0'05$, por lo que la región crítica es $RC = \{T > 4\}$ y 5>4. Luego se rechaza la hipótesis nula.

8.5.3. Test de Mann-Whitney-Wilcoxon

Este test fue obtenido por Mann-Whitney (1947) y por Wilcoxon (1947) mediante una expresión equivalente por un cambio lineal. Se trata de un test válido para contrastar la homogeneidad entre dos poblaciones continuas. Dadas dos m.a.s., la primera de tamaño m de una población con función de distribución F_X, y la segunda de tamaño n de una población con función de distribución F_Y, se define el *estadístico de Mann-Whitney-Wilcoxon* como

$$U = \sum_{i=1}^{m} \sum_{j=1}^{n} D_{ij} \quad \text{siendo} \quad D_{ij} = \begin{cases} 1 & \text{si} \quad X_i < Y_j \\ 0 & \text{resto} \end{cases}$$

Para cada i fijo, $\sum_{j=1}^{n} D_{ij}$ es el número de veces que las observaciones de la segunda muestra están por encima del valor $X_i, i = 1, ..., m$; por lo tanto U es el número de veces que las observaciones de la segunda muestra están por encima de las observaciones de la primera. Es claro que $U \in \{0, 1, ..., mn\}$, que a la configuración $YY...YXX...X$ le corresponde $U = 0$ y a la configuración $XX...XYY...Y$ le corresponde $U = mn$.

Valores grandes de U apoyarán el que $F_X > F_Y$ -que Y domina estocástiamente a X- mientras que valores pequeños de U apoyarán el que $F_X < F_Y$, con lo que se obtiene la Tabla 8.5.3.1

H_0	H_1	Región Crítica
$F_X = F_Y$	$F_X > F_Y$	$U > c_1$
	$F_X < F_Y$	$U < c_2$
	$F_X \neq F_Y$	$U > c_1$ o $U < c_2$ $(c_1 > c_2)$

Tabla 8.5.3.1. Regiones críticas del test de Mann-Whitney-Wilcoxon.

Para la determinación de las constantes se emplea que bajo la hipótesis nula, las $\binom{m+n}{m}$ ordenaciones conjuntas son equiprobables y por tanto

$$P\{U = u | H_0\} = \frac{r_{m,n}(u)}{\binom{m+n}{m}} \quad 0 \leq u \leq m \cdot n \tag{8.5.3.1}$$

donde $r_{m,n}(u)$ es el número de ordenaciones conjuntas compatibles con el valor $U = u$.

Para tamaños pequeños, por enumeración, se puede determinar de forma exacta la distribución de U bajo H_0. Por ejemplo si $m = 2$ y $n = 3$, y se emplea la notación $(yxxyy)$ para simbolizar la muestra conjunta ordenada

$(y_{(1)}, x_{(1)}, x_{(2)}, y_{(2)}, y_{(3)})$, se obtiene la Tabla 8.5.3.2

muestras	U	muestras	U
$xxyyy$	6	$yxyxy$	3
$xyxyy$	5	$yyyxx$	0
$xyyxy$	4	$yyxyx$	1
$yxxyy$	4	$yxyyx$	2
$xyyyx$	3	$yyxxy$	2

Tabla 8.5.3.2. Valores del estadístico de Mann-Whitney-Wilcoxon para $m = 2$ y $n = 3$.

que lleva a la distribución de probabilidad de la Tabla 8.5.3.3

u	0	1	2	3	4	5	6
$P\{U = u\|H_0\}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{2}{10}$	$\frac{2}{10}$	$\frac{1}{10}$	$\frac{1}{10}$

Tabla 8.5.3.3. Distribución del estadístico de Mann-Whitney-Wilcoxon para $m = 2$ y $n = 3$.

Las 4 últimas configuraciones de la Tabla 8.5.3.2 se convierten en las cuatro primeras cuando se lee la muestra de derecha a izquierda, las ordenaciones 5 y 6 son invariantes mediante este mecanismo. Luego si llamamos u' a los valores correspondientes de U para ésta segunda parte de la tabla se tiene que $u+u' = \sum_{i=1}^{m}\sum_{j=1}^{n} 1 = mn$ con lo que se demuestra que la distribución de U es simétrica alrededor del valor $\frac{mn}{2}$.

Para el cálculo de los numeradores de la Expresión (8.5.3.1) es válida la siguiente relación recurrente

$$r_{m,m}(u) = r_{m-1,n}(u) + r_{m,n-1}(u-m)$$

ya que dada una configuración conjunta de m observaciones de X y n observaciones de Y, si acaba en X y se suprime esta última X, queda una sucesión de $m-1$ observaciones de X y n observaciones de Y, que ha de ser compatible con $U = u$, ya que la X suprimida no contribuye al valor de U. Por el contrario si acaba en Y y se la suprime, queda una configuración de m observaciones de X y $n-1$ observaciones de Y que ha de corresponderse con $U = u - m$, ya que la Y suprimida está por encima de m observaciones de la v.a. X. Por lo tanto se tiene que bajo H_0

$$\begin{aligned} P_{m,n}\{U = u|H_0\} &= P\{\text{acabar en } X|H_0\}P_{m,n}\{U = u|\text{acabar en } X,\, \mathrm{H}_0\}+ \\ &\quad +P\{\text{acabar en } Y|H_0\}P_{m,n}\{U = u|\text{acabar en } Y,\ H_0\} \\ &= \tfrac{m}{n+m}P_{m-1,n}\{U = u|H_0\} + \tfrac{n}{m+n}P_{m,n-1}\{U = u - m|H_0\}. \end{aligned}$$

Obviamente, el test no depende del orden en que se obtengan las dos muestras, por lo que se toma como X la variable a la que corresponda el menor tamaño muestral; es decir se supone que es $m \leq n$.

Para obtener la distribución asintótica del *estadístico de Mann-Whitney-Wilcoxon,* como U es suma de v.a. dependientes, se construye una sucesión de v.a. independientes que converjan a lo mismo y cuya media y varianza se conserven (véase Mann y Whitney (1947)). Además

$$E\left[U|H_0\right] = \sum_{i=1}^{m}\sum_{j=1}^{n} E\left[D_{ij}|H_0\right] = \sum_{i=1}^{m}\sum_{j=1}^{n} P\{X_i < Y_j|H_0\} = \frac{mn}{2}$$

ya que bajo H_0 X e Y tienen la misma distribución y,

$$P\{X_i < Y_j|H_0\} = 1/2.$$

Para la varianza se tiene

$$V\left[U|H_0\right] = \sum_{i=1}^{m}\sum_{j=1}^{n} V\left[D_{ij}|H_0\right] + \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k\neq j} Cov\left[(D_{ij}, D_{ik})|H_0\right] +$$
$$+ \sum_{i=1}^{m}\sum_{j\neq i}\sum_{k=1}^{n} Cov\left[(D_{ik}, D_{jk})|H_0\right] + \sum_{i=1}^{m}\sum_{j\neq i}\sum_{k=1}^{n}\sum_{l\neq k} Cov[(D_{ik}, D_{jl})|H_0].$$

Es fácil ver que

$$V[D_{ij}|H_0] = \frac{1}{4}, Cov[(D_{ik}, D_{jl})|H_0] = 0$$

y

$$Cov[(D_{ij}, D_{ik})|H_0] = E[D_{ij}D_{ik}|H_0] - E[D_{ij}|H_0]E[D_{ik}|H_0] = \frac{1}{3} - \frac{1}{2}\frac{1}{2}.$$

Por lo tanto

$$V[U|H_0] = \tfrac{mn}{4} + \tfrac{mn(n-1)}{12} + \tfrac{mn(m-1)}{12} = mn\tfrac{(m+n+1)}{12}.$$

Aplicando el teorema central del límite, se tiene la convergencia

$$\left(U - \frac{mn}{2}\right) \Big/ \sqrt{\frac{mn(m+n+1)}{12}} \xrightarrow{\mathcal{L}} Z \sim N(0,1)$$

Ejemplo 8.5.3.1
Dadas las dos muestras de tamaños $m = 8$ y $n = 9$ siguientes:

$$X : 1, 2, 3, 5, 7, 9, 11, 18$$
$$Y : 4, 6, 8, 10, 12, 13, 14, 15, 19$$

se puede contrastar mediante el test de Mann-Whitney-Wilcoxon de tamaño $0'1$ si ambas provienen de la misma población. Para ello

$$U = \sum_{i=1}^{8}\sum_{j=1}^{9} D_{ij} \quad \text{con} \quad D_{ij} = \begin{cases} 1 & \text{si } X_i < Y_j \\ 0 & resto \end{cases}$$

Con lo que el estadístico U toma el valor $U = 9+9+9+8+7+6+5+1 = 54$. Si la hipótesis alternativa es $H_1 : F_X \neq F_Y$ la región crítica es de la forma $RC = \{U = c_1 \text{ o } U < c_2 \text{ con } c_1 > c_2\}$ y dado que ha de ser $P\{U > c_1|H_0\} \leq 0'05$, de la Tabla 12, se obtiene que $c_1 = 53$; por simetría el valor $c_2 = 8 \cdot 9 - 53 = 19$ es tal que

$$P\{U < c_2|H_0\} \leq 0'05.$$

Por último como $54 > c_1$ pertenece a la RC, se rechaza la hipótesis de homogeneidad de ambas poblaciones.

■

El test de MWW puede también utilizarse cuando ambas poblaciones sean iguales salvo un parámetro de localización, es decir en el mismo contexto que el *test de la mediana*, de hecho muchos estadísticos consideran que el test de MWW es el más indicado para realizar este tipo de contraste.

Wilcoxon (1947) propuso definir, para la muestra conjunta a la que se le asignan sus rangos, el estadístico $V = \sum_{i=1}^{m} R_i$ siendo R_i los rangos correspondientes a la primera muestra en la muestra conjunta, es decir $V = \sum_{i=1}^{m+n} iZ_i$ con $Z_i = 1$ si el valor i-ésimo es de la primera muestra y $Z_i = 0$ si es de la segunda.

Si R_i es el rango de la i-ésima observación de la primera muestra, detrás de ella hay $m + n - R_i$ observaciones de la muestra conjunta, de las cuales $m+n-R_i-(m-i)$ son de la segunda muestra. Sumando en i desde 1 hasta m se obtiene que

$$\sum_{i=1}^{m} [m + n - R_i(m - i)] = m(m + n) - V - \frac{m(m-1)}{2} = U$$

con lo que el *estadístico de Wilcoxon* V y el *estadístico de Mann-Whitney* U están relacionados linealmente, por lo tanto con ambos se realiza el mismo contraste.

Ejemplo 8.5.3.2

Para los datos del Ejemplo 8.5.1.3, se puede contrastar mediante el test de MWW la hipótesis nula $H_0 : F_X = F_Y$ con un tamaño $\alpha = 0'05$.

En este caso es $U = \sum_{i=1}^{8} \sum_{j=1}^{8} D_{ij} = 7 + 5 + 4 + 4 + 4 + 2 + 2 = 29$, y la región crítica es de la forma

$$RC = \{U > c_1 \text{ o } U < c_2, \text{ con } c_1 > c_2\}.$$

Como ha de ser $P\{U > c_1\} \leq 0'25$, se sigue de la Tabla 19 al final del libro, que $c_1 = 50$ y por simetría $c_2 = mn - 50 = 64 - 50 = 14$. Luego como 29 no pertenece a la región crítica, se mantiene la hipótesis de homogeneidad de ambas poblaciones.

8.5.4. Test de Kruskal-Wallis

Como se dijo al tratar del análisis de la varianza (véase la Sección 7.1), algunas veces las hipótesis de normalidad del modelo no se cumplen. En ese caso se puede suponer que las observaciones son de la forma

$$x_{ij} = \theta_i + \varepsilon_{ij} \quad i = 1, ..., k \quad j = 1, ..., n_i$$

Si se quiere contrastar $H_o : \theta_1 = \theta_2 = ... = \theta_k$ frente a H_1: al menos dos valores θ_i son diferentes, sin hacer la hipótesis de que los valores de ε_{ij} tienen distribución normal, se pueden utilizar los rangos R_{ij} de las observaciones x_{ij} en la ordenación de todas las observaciones $n = \sum_{i=1}^{k} n_i$ y construir

$$\overline{R}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} R_{ij}$$

$$\overline{R}_{\cdot\cdot} = \frac{1}{k} \sum_{i=1}^{k} \overline{R}_{i\cdot} = \frac{1}{n} \sum_{i=1}^{n} i = \frac{n+1}{2}$$

Si hubiera uniformidad en la asignación de los rangos, lo que se sigue de H_0, a la muestra i-ésima de tamaño n_i le corresponderían proporcionalmente $n_i(n+1)/2$ rangos. Por otro lado, la suma de los rangos que de hecho corresponden a la muestra i-ésima es

$$R_{i\cdot} = \sum_{j=1}^{n_i} R_{ij}$$

por lo que un test razonable consiste en medir la distancia

$$S = \sum_{i=1}^{k} \left(R_{i\cdot} - \frac{n_i(n+1)}{2} \right)^2 \tag{8.5.4.1}$$

y rechazar cuando S sea grande.

Aunque se puede calcular la distribución bajo H_0 del estadístico S, se prefiere utilizar el estadístico H introducido por Kruskal y Wallis

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \left(R_{i\cdot} - \frac{n_i(n+1)}{2} \right)^2 \frac{1}{n_i} \tag{8.5.4.2}$$

una ligera modificación del anterior, que se introduce por ser conocido su comportamiento asintótico. En efecto, puede demostrarse que asintóticamente, bajo H_0, la distribución de H es una χ^2_{k-1} (véase Lehmann (1975), pág. 396). Para facilitar el cálculo de H se utiliza la expresión alternativa

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} n_i \overline{R}_{i\cdot}^2 - 3(n+1) \tag{8.5.4.3}$$

que se obtiene desarrollando el cuadrado de (8.5.4.2) y teniendo en cuenta que

$$\sum_{i=1}^{k} R_{i\cdot} = \frac{(n+1)n}{2}$$

Ejemplo 8.5.4.1
Para los datos de la preparación de una sustancia química, contenidos en la Tabla 7.2.1 y recogidos en la Tabla 8.5.4.1 para comodidad del lector.

métodos				
1	2	3	4	5
63	27	10	28	47
48	41	30	45	28
57	23	24	44	30
54	31	37	31	35
41	2	40	29	21
43	33	24	44	37

Tabla 8.5.4.1. Datos correspondientes a la preparación de una sustancia mediante 5 procedimientos de fabricación.

El análisis de la varianza sin la suposición de normalidad lleva al cálculo de los rangos contenidos en la Tabla 8.5.4.2. Se observa que cuando aparecen valores repetidos se les asigna el rango medio.

métodos	1	2	3	4	5
	30	7	2	8'5	26
	27	20'5	11'5	25	8'5
	28	4	5'5	23'5	11'5
	29	13'5	17'5	13'5	16
	20'5	1	19	10	3
	22	15	5'5	23'5	17'5
$\overline{R}_{i\cdot}$	26'083	10'167	10'167	17'33	13'75

Tabla 8.5.4.2. Rangos y rangos medios correspondientes a cada método de preparación del Ejemplo 8.5.4.1.

Con este convenio las expresiones de H deberían ser modificadas ligeramente, cosa que no se va a hacer al no ser grande, en este caso, el error que se comete. La región crítica es de la forma

$$RC = \{H \geq \chi^2_{k-1;\alpha}\}$$

y el valor que toma el estadístico es $H = 13'573$, con $\chi^2_{4;0'05} = 9'49$, por lo que se rechaza la hipótesis nula de igualdad de los métodos de preparación. ■

El *test de Kruskal-Wallis* es una generalización a k muestras del *test de Mann-Whitney-Wilcoxon* para dos muestras (véase la Sección 8.5.3), en él no se hacen suposiciones de normalidad por lo que es una alternativa al análisis de la varianza, y además, al ser función de los rangos, no se ve tan afectado por posibles errores en la toma de los datos. Por último se conoce la distribución exacta del H, bajo H_0 para valores pequeños de k y de n_i $i = 1, ..., k$ (véase Lehmann(1975), expresión (5.4), Tabla I).

8.6. CONTRASTES NO PARAMÉTRICOS DE INDEPENDENCIA

En esta Sección, se estudian contrastes para verificar la posible dependencia o independencia entre las dos componentes de una variable aleatoria bidimensional (X, Y). Dada la realización de una m.a.s. de tamaño n $(x_1, y_1), ..., (x_n, y_n)$ se trata de contrastar la hipótesis nula $H_0 : F(x, y) = F_1(x)F_2(y)$, donde F_1 y F_2 son las funciones de distribución marginales. Como es usual se supone que todas las distribuciones son continuas.

Una posible medida de la dependencia entre las dos variables, podría venir recogida por la covarianza entre ambas. Cuando las dos variables son independientes la covarianza es nula; cuando ésta es positiva al crecimiento o decrecimiento de una de las variables les corresponde, respectivamente, el crecimiento o decrecimiento de la otra variable; y, por el contrario, cuando la covarianza es negativa al crecimiento o decrecimiento de una de las variables les corresponde el decrecimiento o crecimiento de la otra variable. Estos hechos son fáciles de comprobar mediante las ecuaciones de las rectas de regresión.

El estimador natural de la covarianza poblacional, es la covarianza muestral, sin embargo la covarianza, tiene el inconveniente de depender de las unidades de medida. Una posible alternativa podría ser utilizar el coeficiente de correlación que, al menos por transformaciones lineales, es invariante, por contra, tiene el inconveniente de que para obtener su distribución hay que recurrir a hipótesis paramétricas sobre la variable bidimensional.

Por estos motivos se van a introducir medidas basadas en los estadísticos ordenados y los rangos.

8.6.1. Test de Kendall

Con el fin de introducir una medida en el intervalo $[-1, 1]$ que valga cero cuando X e Y sean independientes, tome el valor -1 cuando al aumentar una variable la otra disminuya y el valor 1 cuando las dos presenten el mismo comportamiento, se introducen las definiciones siguientes.

Definición 8.6.1.1 (pares concordantes)
El par (x_i, y_i) es *concordante* con el (x_j, y_j) cuando se verifica que si $x_i < x_j$

entonces es $y_i < y_j$ o bien si es $x_i > x_j$ entonces es $y_i > y_j$. En resumen, cuando $(x_j - x_i)(y_j - y_i) > 0$.

Definición 8.6.1.2 (pares discordantes)
El par (x_i, y_i) es *discordante* con el (x_j, y_j) cuando se verifica que si $x_i < x_j$ entonces es $y_i > y_j$ o bien si es $x_i > x_j$ entonces es $y_i < y_j$. En resumen, cuando $(x_j - x_i)(y_j - y_i) < 0$.
■

Parece intuitivo que una medida de la dependencia o independencia existente entre las dos variables de (X, Y) pueda venir dada por la diferencia entre las probabilidades de concordancia π_c y de discordancia π_d, definidas por

$$\pi_c = P\{(X_j - X_i)(Y_j - Y_i) > 0\} \qquad \pi_d = P\{(X_j - X_i)(Y_j - Y_i) < 0\}$$

la medida viene recogida en la siguiente definición.

Definición 8.6.1.3 (coeficiente de Kendall)
Se define el *coeficiente de Kendall* como la diferencia entre la probabilidad de obtener un par de *pares concordantes* menos la probabilidad de obtener un par de *pares discordantes*. Es decir

$$\tau = \pi_c - \pi_d \tag{8.6.1.1}$$

Si las distribuciones son continuas $P\{X_j = X_i\} = P\{Y_j = Y_i\} = 0$ por lo que se tiene que $\pi_c + \pi_d = 1$ y por lo tanto $\tau = 2\pi_c - 1 = 1 - 2\pi_d$.
■

Bajo la hipótesis de independencia de las componentes de la muestra $(X_1, ..., X_n)$, se tiene que $P\{X_i \leq X_j\} = \frac{1}{2}$, ya que

$$P\{X_i \leq X_j\} = \int_{-\infty}^{+\infty} \left[\int_{x_1}^{+\infty} f(x_1, x_2) dx_2 \right] dx_1$$

siendo $f(x_1, x_2)$ es la función de densidad conjunta de las observaciones X_i y X_j, por lo que $f(x_1, x_2) = f_1(x_1) f_1(x_2)$ con f_1 la función de densidad marginal de la variable X. Se tiene entonces que

$$\begin{aligned} P\{X_i \leq X_j\} &= \int_{-\infty}^{+\infty} f_1(x_1) \left[\int_{x_1}^{+\infty} f_1(x_2) dx_2 \right] dx_1 \\ &= \int_{-\infty}^{+\infty} \left[1 - F_1(x_1)\right] f_1(x_1) dx_1 \\ &= 1 - \int_0^1 u du = \frac{1}{2} \end{aligned}$$

la última integral se obtiene haciendo el cambio $F_1(X) = U \sim Uniforme(0, 1)$.

A este resultado se puede llegar de manera más directa, puesto que al intercambiar X_i con X_j se tiene que

$$P\{X_i \leq X_j\} = P\{X_j \leq X_i\}$$

y ambas probabilidades suman 1.

En el caso de que sea verdadera la hipótesis de que X e Y son v.a. independientes se verifica que $\tau = 0$, ya que

$$\pi_c = P\{X_i \leq X_j \cap Y_i \leq Y_j\} + P\{X_i \geq X_j \cap Y_i \geq Y_j\}$$

por la hipótesis de independencia puede ponerse

$$\pi_c = P\{X_i \leq X_j\}P\{Y_i \leq Y_j\} + P\{X_i \geq X_j\}P\{Y_i \geq Y_j\}$$

por ser $P\{X_i \leq X_j\} = P\{X_i \geq X_j\}$

$$\begin{aligned}\pi_c &= P\{X_i \geq X_j\}P\{Y_i \leq Y_j\} + P\{X_i \leq X_j\}P\{Y_i \geq Y_j\} \\ &= P\{X_i \geq X_j \cap Y_i \leq Y_j\} + P\{X_i \leq X_j \cap Y_i \geq Y_j\} \\ &= \pi_d.\end{aligned}$$

Dada la realización particular de una m.a.s. bidimensional $(x_1, y_1), ..., (x_n, y_n)$ se pueden estimar π_c y π_d y mediante la diferencia de ambos estimadores estimar π.

Sea

$$a_{ij} = \begin{cases} 1 & \text{si } (x_j - x_i)(y_j - y_i) > 0 \\ 0 & \text{si } (x_j - x_i)(y_j - y_i) = 0 \\ -1 & \text{si } (x_j - x_i)(y_j - y_i) < 0 \end{cases}$$

es claro que a_{ij} toma el valor 1 con probabilidad π_c, el valor -1 con probabilidad π_d y el valor 0 con probabilidad $1 - (\pi_c + \pi_d)$, por lo tanto al variar la m.a.s. $E[a_{ij}] = \tau$.

Se introduce el *estadístico muestral de Kendall* mediante

$$T = \frac{1}{\binom{n}{2}} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j>i}}^{n} a_{ij} \tag{8.6.1.2}$$

donde sólo se consideran $\binom{n}{2}$ valores a_{ij} ya que a_{ij} y a_{ji} dan la misma información respecto a concordancias o discordancias.

Para cualquier distribución bidimensional, T es un estimador insesgado de τ, ya que

$$E[T] = \frac{1}{\binom{n}{2}} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j>i}}^{n} (\pi_c - \pi_d) = \pi_c - \pi_d = \tau.$$

Nótese que T se puede expresar alternativamente como

$$T = \frac{P - N}{\binom{n}{2}} \tag{8.6.1.3}$$

siendo P el número de valores de a_{ij} positivos en la muestra y N el número de valores a_{ij} negativos.

Esta última expresión es la más rápida de calcular, ya que se puede reordenar la muestra para que quede en la forma (i, S_i) $i = 1, ..., n$ con S_i los rangos correspondientes de la variable Y, lo que facilita los cálculos. Es decir se ha cambiado la muestra (x_i, y_i) por la muestra de los rangos (R_i, S_i) ordenados de menor a mayor por la primera variable, ya que los valores de los estadísticos no varían.

Cuando haya independencia cabe esperar un valor de T próximo a 0, si la dependencia es positiva (en el sentido de que cuando una variable aumenta también lo hace la otra) habrá mayor número de pares concordantes que discordantes, con lo que T será grande, mientras que si la dependencia es negativa (cuando una variable aumenta la otra disminuye) el número de pares discordantes será mayor.

A la muestra $(1, 1), ..., (n, n)$ le corresponde un valor de $P = (n-1) + ... + 1 = \frac{n(n-1)}{2}$ con lo que si no hay repeticiones es $T = 1$. Los diferentes contrastes que pueden hacerse quedan recogidos en la Tabla 8.6.1.1

H_0	H_1	Región Crítica
$F(x,y) = F_1(x)F_2(y)$	$F(x,y) \neq F_1(x)F_2(y)$	$\|T\| > c$
	dependencia positiva	$T < c_1$
	dependencia negativa	$T < c_2$

Tabla 8.6.1.1. Regiones críticas para el test de Kendall.

Para la determinación de las constantes bajo H_0, se puede utilizar que

$$P\{P = p|H_0\} = \frac{r(n,p)}{n!}$$

donde $r(n, p)$ es el número de permutaciones compatibles con el valor p para las ordenaciones de n elementos.

Con valores pequeños de n se puede construir directamente la distribución exacta del estadístico T, como indica la Tabla 8.6.1.2, donde se ha

s_1	s_2	s_3	P	T	s_1	s_2	s_3	P	T
1	2	3	3	1	3	2	1	0	-1
1	3	2	2	1/3	2	3	1	1	-1/3
2	1	3	2	1/3	3	1	2	1	-1/3

Tabla 8.6.1.2. Valores de P número de coincidencias y T estadístico de Kendall con $n = 3$.

empleado que $T = \frac{2}{3}P - 1$ y que la permutación ordenada de atrás hacia adelante se corresponde con un valor del estadístico simétrico respecto al 0.

De esta manera se obtiene la distribución contenida en la Tabla 8.6.1.3

$T = t$	-1	-1/3	1/3	1
$P\{T = t\|H_0\}$	1/6	1/6	1/6	1/6

Tabla 8.6.1.3. Distribución del estadístico de Kendall para $n = 3$.

Es posible disponer de una expresión recursiva en la forma

$$P_n\{P = p|H_0\} = \tfrac{1}{n}(P_{n-1}\{P = p|H_0\} + P_{n-1}\{P = p-1|H_0\} + ...+ \\ +P_{n-1}\{P = p-(n-1)|H_0\})$$

ya que de acuerdo con el teorema de la probabilidad total, la probabilidad del conjunto de muestras de tamaño n compatibles con el valor $P = p$, es la suma en i de las probabilidaes de las muestras de tamaño n que tienen el valor n en la i-ésima posición, con $i = 1, 2, ..., n$. Bajo H_0, la probabilidad de que una muestra tenga el valor n en la posición i es $1/n$, y la probabilidad de que $P = p$, condicionada porque tenga un n en la i-ésima posición, es $P_{n-1}\{P = p-(i-1)|H_0\}$, ya que el valor n está por encima de los valores $1, 2, ..., i-1$.

Este razonamiento, permite calcular la distribución exacta de T para valores no muy grandes de n.

La distribución asintótica de T es

$$3\sqrt{\frac{n(n-1)}{2(2n+5)}}T \xrightarrow{\mathcal{L}} Z \sim N(0,1)$$

(véase Gibbons y Chakraborti (1992), pág. 349).

Debe notarse que cuando existen observaciones repetidas, tanto en la primera componente, como en la segunda, como en ambas; a_{ij} toma el valor cero, por lo tanto el valor del estadístico no se ve afectado, aunque deja de ser cierto que

$$P + N = \binom{n}{2}$$

En la práctica cuando hay repeticiones, lo que se hace es sustituir el rango de los valores repetidos por el promedio de los rangos que tienen estas observaciones repetidas.

El test de Kendall puede utilizarse también, para ver si la sucesión $y_1, ..., y_n$ es aleatoria, para ello se contrasta la independencia de la variable bidimensional mediante la muestra $(i, y_i) \quad i = 1, ..., n$.

Ejemplo 8.6.1.1
Para contrastar la independencia entre las notas de Matemáticas e Inglés en un colectivo, se escogen aleatoriamente 6 alumnos y se observa que han obtenido las siguientes calificaciones:

Matemáticas	5	6,50	7	7,50	9	8'75
Inglés	7	6	6'50	6'75	8'50	9'50

Al reordenar la muestra por la nota de Matemáticas y empleando los rangos, se obtiene la muestra equivalente

$$(1,4)\ \ (2,1)\ \ (3,2)\ \ (4,3)\ \ (5,6)\ \ (6,5)$$

con lo que el valor del *estadístico de Kendall* es

$$T = \frac{2P}{\binom{6}{2}} - 1 = \frac{2 \cdot 11}{\binom{6}{2}} - 1 = 7/15.$$

La forma de la región crítica, es de acuerdo con la Tabla 8.6.1.1

$$RC = \{|T| > c\}$$

con lo que a un tamaño de $0'05$, se obtiene para $c = 9/15$ (véase la Tabla 13 al final del libro). Por lo tanto como $7/15$ no pertenece a la RC se mantiene la hipótesis de independencia.

8.6.2. Test del coeficiente de correlación de los rangos de Spearman

Dada una m.a.s. de tamaño n de una población bivariante, se puede contrastar la independencia entre las variables, mediante el coeficiente de correlación calculado a partir de los rangos de la muestra. Por comodidad de cálculo, se supone la muestra reordenada en la forma $(1, S_1), ..., (n, S_n)$ donde S_i es el rango correspondiente a la observación de la segunda componente que se corresponde con la de rango i de la primera componente.

El *coeficiente de correlación de los rangos de Spearman* viene dado por

$$R = \frac{\sum_{i=1}^{n}(i - \overline{i})(S_i - \overline{S})}{\sqrt{\sum_{i=1}^{n}(i - \overline{i})^2 \sum_{i=1}^{n}(S_i - \overline{S})^2}} \tag{8.6.2.1}$$

expresión que puede simplificarse observando que

$$\overline{i} = \overline{S} = \frac{n+1}{2}$$

$$\sum_{i=1}^{n}(i - \overline{i})^2 = \sum_{i=1}^{n}(S_i - \overline{S})^2 = \frac{n(n+1)(n-1)}{12}$$

pues

$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$$

con lo que se puede poner alternativamente

$$R = \frac{12}{n(n^2-1)} \sum_{i=1}^{n} iS_i - 3\frac{n+1}{n-1} \tag{8.6.2.2}$$

Por ser R un coeficiente de correlación, está contenido en el intervalo $(-1, 1)$ y si las variables X e Y son independientes $E[R|H_0] = 0$, ya que

$$E[S_i|H_0] = \sum_{i=1}^{n} \frac{1}{n} i = \frac{n+1}{2}$$

y, por tanto

$$E[R|H_0] = \frac{12}{n(n^2-1)} \sum_{i=1}^{n} i\frac{(n+1)}{2} - 3\frac{n+1}{n-1} = 0.$$

Bajo la hipótesis nula, la distribución de R es simétrica alrededor de cero, ya que los valores de los rangos S_i pueden ser considerados como una permutación de los n primeros números naturales, que si es compatible con un valor $R = r$, la permutación ordenada de atrás hacia adelante asigna el valor $n + 1 - i$ al que antes era i, lo que lleva al valor $R = r'$ con

$$\begin{aligned} r' &= \frac{12}{n(n^2-1)} \sum_{i=1}^{n} (n+1-i)s_i - 3\frac{n+1}{n-1} \\ &= 3\frac{n+1}{n-1} - \frac{12}{n(n^2-1)} \sum_{i=1}^{n} iS_i \\ &= -r. \end{aligned}$$

La Tabla 8.6.2.1 recoge la forma de las regiones críticas basadas en el *estadístico de Spearman*.

H_0	H_1	Región Crítica
$F(x,y) = F_1(x)F_2(y)$	$F(x,y) \neq F_1(x)F_2(y)$	$\lvert R\rvert > c$
	dependencia positiva	$R > c_1$
	dependencia negativa	$R < c_2$

Tabla 8.6.2.1. Regiones críticas para el test de Spearman.

El cálculo de la distribución exacta bajo la hipótesis nula, para valores pequeños de n, puede efectuarse de forma directa. Así para $n = 3$, la

Expresión (8.6.2.2) lleva a la Tabla 8.6.2.2

S_1	S_2	S_3	$\sum_{i=1}^{3} iS_i$	R	S_1	S_2	S_3	$\sum_{i=1}^{3} iS_i$	R
1	2	3	14	1	3	2	1	10	-1
1	3	2	13	0'5	2	3	1	11	-0'5
2	1	3	13	0'5	3	1	2	11	-0'5

Tabla 8.6.2.2. Valores del estadístico de Spearman para $n = 3$.

con lo que se obtiene la distribución contenida en la Tabla 8.6.2.3

$R = r$	-1	-0'5	0'5	1
$P\{R = r\|H_0\}$	1/6	1/3	1/3	1/6

Tabla 8.6.2.3. Distribución del estadístico de Spearman para $n = 3$.

La distribución asintótica del *estadístico de Spearman* viene dada por

$$\sqrt{n-1}\, R \to Z \sim N(0,1)$$

(puede consultarse en Fraser (1957), pág. 247).

El tratamiento de los rangos repetidos consiste en sustituirlos por su rango promedio, esto implica que ha de calcularse R mediante la expresión directa, Expresión (8.6.2.1), y no mediante su expresión simplificada, Fórmula (8.6.2.2). De todas formas si n es grande, el error que se comete al emplear la expresión simplificada es mínimo.

El test de Spearman puede utilizarse igual que el test de Kendall, para verificar si la sucesión $y_1, ..., y_n$ es aleatoria, mediante el contraste de independencia basado en la muestra $(i, y_i) \quad i = 1, ..., n$.

Ejemplo 8.6.2.1
A 8 cocineros que preparan un mismo plato que se presenta a un concurso, dos jueces les otorgan la siguiente puntuación:

primer juez:	2	1	3	5	4	8	7	6
segundo juez:	1	2	4	5	7	6	8	3

donde 1 es la mejor calificación y 8 la peor. Para contrastar la hipótesis nula de que los dos jueces califican con independencia, la región crítica, con un test de tamaño $0'05$, es

$$RC = \{|R| > 0'714\}$$

(véase la Tabla 14 del Apéndice al final del libro). Si se reordena la muestra por la calificación del primer juez se obtiene

$$(1,2)\ \ (2,1)\ \ (3,4)\ \ (4,7)\ \ (5,5)\ \ (6,3)\ \ (7,8)\ \ (8,6)$$

con lo que el valor del *estadístico de Spearman* es $R = \frac{12}{8\cdot 63}191 - 3\frac{9}{7} = 0'69$, luego se mantiene la hipótesis de independencia.

8.7. APROXIMACIÓN HISTÓRICA

Puede decirse que la primera consideración de un estudio no paramétrico, fué la del test de los signos realizado por Arbuthnot (1667-1735) en 1710, año en que se publica *Un argumento para la divina providencia, tomado de la ley de regularidad observada en el nacimiento de ambos sexos.*

El problema de la existencia de valores discordantes es tratado inicialmente por Boscovich (1711-1787), un astrónomo que en 1770 se pregunta por la conveniencia de calcular la media de un conjunto de observaciones, una vez suprimidas las observaciones más grandes y más pequeñas, y que parece haber preconizado el rechazo de los valores *discordantes*. Esta idea es recogida por Legendre (1752-1833) en 1805, pero el problema no es tratado desde el punto de vista probabilístico hasta 1852, en que Benjamin Pierce, astrónomo y matemático de la Universidad de Harvard, enuncia el primer criterio basado en la probabilidad de rechazo de observaciones *discordantes*, criterio que por cierto fué inmediatamente combatido por otros astrónomos. A partir de este momento las investigaciones avanzan por dos caminos diferentes. Por un lado aparece una corriente que estudia la ponderación de los datos muestrales de manera diferentes, dentro de ésta se puede citar a Francis Edgeworth (1845-1926) en 1885. Por otro lado se van introduciendo criterios de rechazo; así en 1884 Wright propone rechazar cualquier observación, tal que $|X_i - \overline{X}|$ sea mayor que 5 veces el *error probable*, donde esta última cantidad es la mitad del intervalo intercuartílico de la distribución $N(\mu, \sigma)$, otras reglas han sido introducidas por Student en 1927, por Thompson en 1935 y, más recientemente, por Dixon en 1950 y 1951.

Estos desarrollos dieron lugar al estudio del comportamiento de los estimadores, cuando las suposiciones usuales de normalidad en la población no se cumplen, lo que llevó a la introducción de la mediana y de los estadísticos ordenados por Galton en 1875, quien preconizó la utilización de la mediana muestral para distribuciones con colas más altas que la distribución normal, y que le llevaron a él y a Edgeworth, en 1889, a estimar, en esos casos, mediante funciones de los estadísticos ordenados. Esto originó el que Simon Newcomb, probablemente el más celebre astrónomo americano del siglo XIX, estudiara, en 1882, la distribución de mixturas de distribuciones normales con varianzas desiguales. Así en 1886 Newcomb obtuvo un estimador de centralización que es más *robusto* que la media, dando origen al

término *robusto* que fué usado por George Box en 1953 y que ha permitido el desarrollo de una fructífera rama de investigación, la *teoría de la robustez*.

Otro hito importante en los métodos no paramétricos, fué el célebre test de la χ^2 introducido por Karl Pearson(1857-1936) y publicado en 1900. Algo más tarde, Harald Cramér en 1928 y Richard von Mises en 1931, contruyeron un test de ajuste basado en la distancia entre la función de distribución empírica y la que se pretende contrastar bajo la hipótesis nula, calculada la distancia mediante $sup|F_n^*(x) - F_0(x)|^2$. Esto dio pie a Kolmogorov a la introducción del estadístico que lleva su nombre y a que obtuviera su distribución asintótica. En 1939 Smirnov emplea por primera vez el estadístico de Kolmogorov para un contraste unilateral y generaliza el mismo al caso de dos muestras; alrededor de 1960 extiende las mismas ideas a más de dos muestras.

Los tests de independencia nacen en el campo de la Psicología junto con las medidas de asociación entre caracteres; así el coeficiente de correlación de los rangos fué introducido por Spearman en 1904 en un artículo titulado *La demostración y medida del coeficiente de asociación entre dos cosas.* En la misma línea Kendall, en 1938 introduce su medida de independencia con un trabajo titulado *Una nueva medida de correlación de rangos.*

El estudio posterior de la homogeneidad entre poblaciones, dio lugar a la introducción de los rangos, con los trabajos de Hotelling en 1936, que culminan con el estudio de Wilcoxon (1892-1965) publicado en 1945 y que tiene por título *Comparaciones individuales por los métodos de los rangos.* En este estudio Wilcoxon introduce los tests de rangos, que son continuados en 1947 por Mann y Whitney.

Sin duda, si se debe fijar más la atención en algun autor dentro de la inferencia no paramétrica este es Frank Wilcoxon, irlandés que pasó a los Estados Unidos donde consiguió su doctorado en Química. Su interés en la estadística comenzó en 1925, con el estudio del libro de Fisher *Métodos Estadísticos para Investigadores.* Su contribución más importante fué el ya citado artículo de 1945 en el que introdujo el test de los rangos signados para dos muestras y el mismo estadístico para muestras *emparejadas.* Hodges y Lehmann en 1961 muestran la gran eficacia de estos tests, para prácticamente cualquier distribución poblacional.

La construcción de los intervalos de tolerancia y de los contrastes para los cuantiles de una población, son debidos a Wilcoxon en 1947.

8.8. EJERCICIOS

8.8.1

Contrástese si los datos siguientes provienen de una distribución $N(100, 10)$ con un test de tamaño $0'05$

107'9	96'7	91'2	79	103'1	88	101'3	106	93'7	86
107'7	99'4	104'6	117'2	112'2	106'9	93	88'3	101'9	109'8

8.8.2
Se supone que el número de accidentes diarios en una red viaria sigue una distribución de Poisson de parámetro 2. Se recogen observaciones durante 200 días, obteniéndose los siguientes resultados

nº de accidentes	0	1	2	3	4	5	6	7 ó más
nº de días	22	53	58	39	20	5	2	1

Contrástese el ajuste a la distribución indicada, mediante un test de tamaño 0′05.

8.8.3
Un programa de ordenador genera números distribuidos aleatoriamente con distribución *Uniforme* en el intervalo $(0, 10)$. Tras obtener 250 números se observan los siguientes resultados:

intervalos	$[0, 2]$	$(2, 4]$	$(4, 6]$	$(6, 8]$	$(8, 10]$
observaciones	38	55	54	41	62

¿Son estos resultados compatibles con las hipótesis hechas? Tómese un tamaño de 0′05.

8.8.4
Al lanzar un dado 300 veces, se han obtenido las siguientes frecuencias

valores	1	2	3	4	5	6
frecuencias	43	49	56	45	66	11

Contrástese si el dado es equilibrado mediante un test de tamaño 0′05.

8.8.5
La observación de defectos congénitos en una muestra de 100 individuos de una población, dió lugar a la siguiente tabla:

nº de defectos	0	1	2	3	4	5
frecuencias	84	9	3	2	1	1

¿Puede decirse que se ajustan a una distribución de Poisson? Tómese un test de tamaño 0′05.

8.8.6
Se está estudiando la distribución de los grupos sanguíneos en dos comunidades. Los resultados obtenidos fueron:

	0	A	B	AB
Comunidad 1	121	120	79	33
Comunidad 2	118	95	121	30

a) ¿Se puede considerar que ambas comunidades tienen la misma distribución de grupos sanguíneos?

b) ¿Se puede afirmar que el modelo teórico que asigna probabilidades a los distintos grupos sanguíneos es el contenido en la siguiente tabla?

grupos	O	A	B	AB
probabilidades	r^2	p^2+2pr	q^2+2qr	$2pq$

con $p+q+r=1$. Supóngase inicialmente $\widehat{p}=0'25 \quad \widehat{q}=017$. Tómese un test de tamaño $0'05$.

8.8.7

La siguiente tabla recoge los resultados de una m.a.s. de 400 individuos empleados en una misma compañía

	sueldos en euros		
años de permanencia	≤ 480	$(480-900]$	≥ 900
≤ 5	75	75	25
$(5-10]$	25	50	25
>10	25	75	50

Mediante un test de tamaño $0'05$, contrástese la independencia entre las variables, años de permanencia y sueldos.

8.8.8

Dadas cuatro muestras de tamaños respectivos $n_1=24, n_2=20, n_3=25$ y $n_4=20$, se quiere contrastar si provienen todas de una población con distribución $N(0,1)$. Para ello se divide la recta real en los intervalos $A_1=(-\infty,-0'75], A_2=(-0'75,0], A_3=(0,0'75]$ y $A_4=(0'75,+\infty)$ llegándose a la siguiente tabla:

	A_1	A_2	A_3	A_4
muestra 1	3	7	7	7
muestra 2	3	5	5	7
muestra 3	9	5	5	6
muestra 4	4	6	8	2

Contrástese la hipótesis de homogeneidad mediante un test de tamaño $0'05$.

8.8.9

La siguiente tabla recoge la distribución observada de dos características, que presentan ciertas plantas de guisantes

Forma	Color		
del polen	Morado	Rojo	
alargado	6	1	7
redondo	1	2	3
	7	3	10

Se pide:

a) Contrástese el ajuste de las distribuciones empíricas marginales a la proporción 3:1 mediante el método de la *distancia de K. Pearson*.

b) Contrástese la independencia entre la forma y el color mediante el test de la razón de verosimilitud.

Empléese en ambos casos un tamaño de $0'05$.

8.8.10

Las ventas en dos establecimientos E_1 y E_2 de una misma cadena, fueron controladas aleatoriamente en 10 y 15 días respectivamente, arrojando los resultados siguientes en unidades convenientes:

E_1	143'5	125'6	142'8	150'2	128'5	147'9	139'8	140'6	147'3	145'9
E_2	150'7	145'8	147'3	148'9	153'5	160'1	140'4	128'9	155'6	120'1
	152'3	151'3	124'2	139'6	110'2					

Contrástese mediante el *test de Kolmogorov-Smirnov* si cabe suponer que las ventas son iguales. Tómese un tamaño de $0'05$.

8.8.11

Los siguientes 20 datos han sido obtenidos por simulación en un ordenador a partir de una distribución $N(1,3)$. Contrástese su adecuación a una $N(0,1)$ y a una $N(\mu,\sigma)$ mediante contrastes de tamaño $0'05$.

-0'09	1'41	3'99	7'62	-1'56	1'52	-'057	1'09	3'56	0'78
1'60	2'96	-0'14	-1'81	-1'14	3'34	8'16	0'12	1'05	5'46

8.8.12

Para contrastar la habilidad de los parasicólogos para inducir influencia paranormal, se les pide a 9 de ellos que induzcan roturas en las visitas a una tienda de porcelanas. El número de roturas entre 500 visitantes; el día que estaba el parasicólogo y el mismo día de la semana anterior, sin la presencia del mismo, queda recogido en la siguiente tabla:

	1	2	3	4	5	6	7	8	9
con parasicólogo	2'5	3'5	1'0	4'6	3'8	5'9	6'1	15'1	2'4
sin parasicólogo	0	3'7	2'4	3'8	4'1	1'2	1'1	1'3	0

Realícense contrastes de localización, para ver si el fenómeno paranormal se ha presentado. Utilícese un tamaño de $0'05$.

8.8.13

En una estantería hay 28 libros, 12 de *Estadística* y los restantes clasificados como *generales*. Mediante el *test de Mann-Whitney-Wilcoxon* contrástese si puede suponerse que tienen el mismo número de páginas. Repítase el contraste mediante el test de normalidad. Utilícese en ambos casos un tamaño de 0'05.

Estadística	126,142,156,228,245,246,370,419,433,454,478,503
Generales	29,39,60,71,82,112,125,170,192,224,263,275,276,286,369,756

8.8.14

Dadas las siguientes muestras correspondientes a dos poblaciones

X	13	53	59	75	76	79	129	138
Y	96	102	104	107	110	130	133	141

Se pide:

a) Contrástese, mediante la primera muestra, si el cuantil de la población de que proviene es $k_{0'75} = 75'5$ frente a que es distinto de ese valor.

b) Mediante el *test de Kolmogorov-Smirnov*, contrástese si las dos muestras provienen de la misma población frente a que es $F_X > F_Y$.

c) Repítase la parte b) para el *test de la mediana*.

d) Repítase la parte b) para el *test de Mann-Whitney-Wilcoxon*.

Utilícese en todos los casos un tamaño de $0'1$.

8.8.15

Se han tomado dos muestras de tamaño 12 de las calificaciones dadas por dos profesores que corrigen dos grupos distintos de una misma asignatura. Los resultados obtenidos han sido

P_1	1'9	2'2	3	3'7	5'6	6'6	6'7	7'1	7'5	7'8	8'2	9'5
P_2	1'3	2'6	3'3	5'9	7'2	7'6	7'7	8	8'1	8'5	8'6	9

Mediante el *test de Mann-Whitney-Wilcoxon*, contrástese si ambas muestras proceden de la misma población, frente a la hipótesis de que las poblaciones difieren en un parámetro de localización. Utilícese un tamaño de $0'05$.

8.8.16

Diez niños y diez niñas de 4 años de edad, fueron observados durante dos sesiones de juegos de 15 minutos cada una. Durante este tiempo se les midió el grado de agresión mediante un test conveniente, a calificación más alta como resultado del test, mayor grado de agresión. Se obtuvieron las siguientes observaciones:

niñas	15	26	16	9	7	16	58	22	40	55
niños	104	41	50	45	65	113	65	72	69	86

Contrástese si existe diferencia de agresividad significativa mediante el *test de Kolmogorov-Smirnov*, el *de la Mediana* y el *de Mann-Whitney-Wilcoxon*. Tómese un tamaño de $0'05$.

8.8.17

Se quiere averiguar si existe independencia entre las calificaciones en Estadística y Análisis del tercer curso de la Licenciatura en Matemáticas. Se escogen al azar las notas de 15 alumnos y se observan los valores siguientes:

Estadística	5'8	2'4	1'5	7'3	4'6	7'9	8'2	9'3	4'5	3'5
	1	4'7	3'2	9'5	6'5					
Análisis	5'5	7	6'1	5	4'2	1'7	5'2	6'2	5'7	7'2
	6	2	3'1	5'4	7'1					

Contrástese la hipótesis nula de que hay independencia mediante un test de tamaño $0'05$.

8.9. SOLUCIÓN A LOS EJERCICIOS

8.9.1

Mediante el test del cuadrado de la *distancia de K. Pearson* (véase la Sección 8.2.1), se construye la siguiente tabla

A_i	$\leq 93'26$	$(93'26, 100]$	$(100, 106'745]$	$> 106'745$	
$P(A_i\|H_0)$	0'25	0'25	0'25	0'25	1
O_i	6	3	5	6	20
$E_i = 20P(A_i\|H_0)$	5	5	5	5	20

como la región crítica es de la forma

$$RC = \left\{ \sum_{i=1}^{4} (O_i - E_i)^2 / E_i > \chi^2_{k-1;\alpha} \right\}$$

y el valor del estadístico es $1'2$, al ser $\chi^2_{3;0'05} = 7'81$ (véase la Tabla 4 al final del libro), se acepta la hipótesis nula. Al valor $1'2$ le corresponde un p-valor de $0'753$.

Mediante el *test de la RV* la región crítica es

$$RC = \left\{ \lambda(n_1, ..., n_4) = \prod_{i=1}^{4} \left(\frac{p_i}{n_i/n} \right)^{n_i} \leq c \right\}$$

para la determinación de la constante se emplea que $-2 \ln \lambda(N_1, ..., N_k)$ se distribuye aproximadamente con distribución χ^2_{k-1}, luego

$$RC = \left\{ -2 \ln \lambda(n_1, ..., n_4) = -2 \sum_{i=1}^{4} n_i \ln \frac{p_i}{n_i/n} \geq -\ln c \right\}$$

el valor del estadístico es $-2\,[-1'0939 + 1'5325 - 1'0939] = 1'3106$, y como $-\ln c = \chi^2_{k-1;0'05} = 7'81$ se acepta H_0 que la muestra proviene de la distribución $N(100, 10)$.

Por último, mediante el *estadístico de Kolmogorov-Smirnov* (véase la Sección 8.4.1), se tiene que $D^+_{20} = 0'0857$ $D^-_{20} = 0'1017$ luego $D_{20} = 0'1017$, como la región crítica es de la forma

$$RC = \{D_n > d_{n;\alpha}\}$$

(veáse la Tabla 8.4.1), y como $d_{20;0'05} = 0'294$ (véase la Tabla 7 al final del libro), se acepta también H_0, al valor del estadístico $0'1017$ le corresponde un p-valor de $0'98$. De todas formas, al ser tratados los datos de forma agrupada se pierde información, por lo que la utilización del *test de Kolmogorov-Smirnov* no es demasiado adecuada.

8.9.2

Se trata de contrastar el ajuste a un distribución de $Poisson(2)$. Mediante el método de la *distancia de K. Pearson* (véase la Sección 8.2.1), se contruye la tabla

A_i	0	1	2	3	4	5
$P(A_i\|H_0)$	0'1353	0'2707	0'2707	0'1404	0'0902	0'0361
O_i	22	53	58	39	20	5
$E_i = 200P(A_i\|H_0)$	27'06	54'14	54'14	36'08	18'04	7'22

A_i	6	≥ 7	
$P(A_i\|H_0)$	0'0120	0'0043	0'9597
O_i	2	1	200
$E_i = 200P(A_i\|H_0)$	2'4	0'86	199'94

se reagrupan las tres últimas clases, para que las frecuencias esperadas E_i sean siempre mayores o iguales que 5.

A_i	0	1	2	3	4	≥ 5	
$P(A_i\|H_0)$	0'1353	0'2707	0'2707	0'1404	0'0902	0'0524	0'9597
O_i	22	53	58	39	20	8	200
$E_i = 200P(A_i\|H_0)$	27'06	54'14	54'14	36'08	18'04	10'48	199'94

La región crítica es de la forma

$$RC = \left\{ \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \geq \chi^2_{5;0'05} = 11'1 \right\}$$

y el valor del estadístico es $3'881$, luego se acepta H_0. Al valor $3'881$ le corresponde un p-valor de $0'57$.

Si se aplica ahora el *test de la RV* la región crítica es

$$RC = \left\{ -2 \sum_{i=1}^{6} O_i \ln \frac{P(A_i|H_0)}{O_i/n} \geq \chi^2_{5;0'05} = 11'1 \right\}$$

y el valor del estadístico

$$-2[4'554 + 1'279 - 3'9944 - 3'0351 - 2'0628 + 2'1602] = 2'5004$$

luego se acepta H_0.

Si se utiliza ahora el *estadístico de Kolmogorov-Smirnov* (véase la Sección 8.4.1), se obtiene $D_6 = 0'36$ lo que lleva a un p-valor de $0'4183$ por lo que se acepta H_0, de la Tabla 7 en el Apéndice es $d_{6;0'05} = 0'519$, la región crítica es de la forma

$$RC = \{D_n > d_{n;\alpha} = 0'519\}$$

(véase la Tabla 8.4.1.1).

8.9.3

Se trata de contrastar $H_0 : f(x) = \frac{1}{10} \quad x \in (0,10)$ frente a $H_1 : f(x)$ cualquiera. Mediante el método de la *distancia de K. Pearson* se construye la tabla:

A_i	(0,2]	(2,4]	(4,6]	(6,8]	(8,10]	
$\mathrm{P}(\mathrm{A}_i\|H_0)$	0'2	0'2	0'2	0'2	2	1
O_i	38	55	54	41	62	250
$E_i = nP(A_i\|H_0)$	50	5	05	05	50	250

La región crítica es de la forma

$$RC = \left\{ \sum_{i=1}^{5} \frac{(O_i - E_i)^2}{E_i} \geq \chi^2_{4;0'05} = 9'49 \right\}$$

el valor del estadístico es $8'2$, luego se acepta H_0. Al valor del estadístico le corresponde un p-valor de $0'085$.

Mediante el *test de la RV* se tiene la región crítica

$$RC = \left\{ -2 \sum_{i=1}^{5} O_i \ln \frac{P(A_i|H_0)}{O_i/250} \geq \chi^2_{4;0'05} = 9'49 \right\}$$

el valor del estadístico es $8'36 \ngeq 9'49$ luego se acepta la hipótesis nula. El *estadístico de Kolmogorov-Smirnov* no debe emplearse en este caso por venir las frecuencias agrupadas.

8.9.4

Se pretende contrastar $H_0 : Multinomial(300, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$ frente a $H_1 : Multinomial(300, p_1, ..., p_5)$ con $\sum_{i=1}^{6} p_i = 1$. Para utilizar el método de la *distancia de K. Pearson* se tiene la tabla:

A_i	1	2	3	4	5	6	
$P(A_i\|H_0)$	1/6	1/6	1/6	1/6	1/6	1/6	1
O_i	43	49	56	45	66	41	300
$E_i = nP(A_i\|H_0)$	50	50	50	50	50	50	300

La región crítica es de la forma

$$RC = \left\{ \sum_{i=1}^{6} \frac{(O_i - E_i)^2}{E_i} \geq \chi^2_{5;0'05} = 11'1 \right\}$$

Mediante el *test de la RV* la región crítica es

$$RC = \left\{ -2 \sum_{i=1}^{6} O_i \ln \frac{P(A_i|H_0)}{O_i/n} \geq \chi^2_{5;0'05} = 11'1 \right\}$$

el estadístico da $8'63$ luego se acepta también la hipótesis nula.

Por venir los datos agrupados, no está indicado el *test de Kolmogorov-Smirnov.*

8.9.5

Se trata de contrastar $H_0 : Multinomial(100, p_1(\lambda), ..., p_6(\lambda))$ frente a $H_1 : Multinomial(100, p_1, ..., p_6)$ con $\sum_{i=1}^{6} p_i = 1$, donde $p_i(\lambda) = e^{-\lambda}\frac{\lambda^i}{i!}$. Para determinar λ, se debe resolver

$$\max_{\lambda} \prod_{i=1}^{6} [p_i(\lambda)]^{n_i}$$

Tomando logaritmos y derivando respecto de λ, se resuelve en λ la ecuación

$$\sum_{i=1}^{6} n_i \frac{\frac{\partial}{\partial\lambda} p_i(\lambda)}{p_i(\lambda)} = 0$$

expresión que toma la forma

$$\sum_{i=1}^{6} n_i \left[-e^{-\lambda}/i! + e^{-\lambda}\frac{\lambda^{i-1}}{i-1!} \right] \Big/ e^{-\lambda}\frac{\lambda^i}{i!} = 0$$

de donde

$$\widehat{\lambda} = \sum_{i=1}^{6} i n_i \Big/ \sum_{i=1}^{6} n_i = \frac{30}{100} = 0'3$$

Para realizar el contraste de la *distancia de K. Pearson*, se construye la tabla:

A_i	1	2	3	4	5	≥ 5	
$p_i(\widehat{\lambda})$	0'7408	0'2222	0'0333	0'0033	0'0002	0'0000	1
O_i	84	9	3	2	1	1	100
$E_i = 100 p_i(\widehat{\lambda})$	74'08	82'22	3'33	0'33	0'02	0	100

con el fin de lograr que las frecuencias esperadas sean próxima a 5, se agrupan las tres últimas clases, lo que lleva a la nueva tabla:

A_i	1	2	≥ 2	
$p_i(\widehat{\lambda})$	0'7408	0'2222	0'0368	1
O_i	84	9	7	100
$E_i = 100\ p_i(\widehat{\lambda})$	74'08	82'22	3'68	100

La región crítica es de la forma

$$RC = \left\{ \sum_{i=1}^{3} \frac{(O_i - E_i)^2}{E_i} \geq \chi^2_{1;0'05} = 3'84 \right\}$$

el estadístico vale $12'188$, luego se rechaza la hipótesis nula. El p-valor correspondiente a $3'84$ es de $0'0002$.

Mediante el *método de la RV* la región crítica viene dada por

$$RC = \left\{ -2 \sum_{i=1}^{3} O_i \ln \frac{p_i(\widehat{\lambda})}{O_i/100} \geq \chi^2_{1;0'05} = 3'84 \right\}$$

el valor estadístico es $13'846$; luego también se rechaza la hipótesis nula.

8.9.6

a) Es un *test de homogeneidad* (véase la Sección 8.2.2) donde las probabilidades de las multinomiales son respectivamente

$$H_0 : p_1, p_2, p_3, p_4 \ \text{ con } \ \sum_{i=1}^{4} p_i = 1 \qquad H_1 : p_{11} ... p_{14} \ \text{ con } \ \sum_{j=1}^{4} p_{1j} = 1$$

$$\widehat{p}_1 = \tfrac{239}{717} \ \ \widehat{p}_2 = \tfrac{215}{717} \ \ \widehat{p}_3 = \tfrac{200}{717} \ \ \widehat{p}_4 = \tfrac{63}{717} \qquad p_{21} ... p_{24} \ \text{ con } \ \sum_{j=1}^{4} p_{2j} = 1$$

$$\widehat{p}_{ij} = \tfrac{n_{ij}}{n_{i\cdot}}$$

se tienen los siguientes resultados

121	120	79	33	353
118	95	121	30	364
239	215	200	63	717

por lo tanto se trata de contrastar

$$\mathrm{H}_0 : 0'333 \ \ 0'300 \ \ 0'279 \ \ 0'088 \qquad H_1 : 0'343 \ \ 0'340 \ \ 0'224 \ \ 0'093 \ \ 0'324 \ \ 0'261 \ \ 0'332 \ \ 0'082$$

Mediante el *método de K. Pearson*

$$RC = \left\{ \sum_{\substack{i=1,2 \\ j=1,...,4}} \frac{(n_{ij} - \frac{n_{i\cdot} - n_{\cdot j}}{n})^2}{\frac{n_{i\cdot} - n_{\cdot j}}{n}} \geq \chi^2_{(r-1)(k-1);\alpha} = \chi^2_{3;0'05} \right\}$$

el valor del estadístico es $11'74$ y el cuantil $\chi^2_{3;0'05} = 7'81$, luego se rechaza la hipótesis, el p-valor es $0'0083$.

Mediante el *método de la RV*

$$RC = \left\{ -2 \sum_{i,j} n_{ij} \ln \frac{\widehat{p}_i}{\widehat{p}_{ij}} \geq 7'81 \right\}$$

el valor del estadístico es $11'682$; luego se rechaza H_0.

b) Como por la parte a) hay que rechazar la homogeneidad entre ambas muestras, se va a proceder a contrastar el modelo muestra a muestra. Se pretende contrastar la hipótesis nula

$$H_0 : \mathrm{r}^2 \quad p^2 + 2pr \quad q^2 + 2qr \quad 2pq$$

donde sustituyendo por

$$\widehat{p} = 0'25 \quad \widehat{q} = 0'17 \quad \widehat{r} = 0'58$$

quedan realmente las probabilidades

$$H_0 : 0'3364 \quad 0'3525 \quad 0'2261 \quad 0'085$$

frente a

$$H_1 : 0'343 \quad 0'340 \quad 0'224 \quad 0'093.$$

Para realizar el contraste mediante la *distancia de K. Pearson* se construye la siguiente tabla:

A_i	0	A	B	AB	
$P(A_i\|H_0)$	0'3364	0'3525	0'2261	0'085	1
O_i	121	120	79	33	353
E_i	118'75	124'43	79'81	30	353

la región crítica es de la forma

$$RC = \left\{ \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i} \geq \chi^2_{3;0'05} = 7'81 \right\}$$

el valor del estadístico es $0'5086$ luego se acepta H_0, y el modelo es correcto.

Aunque no es necesario hacerlo para la segunda muestra, ya que ambas no eran homogéneas, se puede repetir el procedimiento. Es decir, si se contrasta la hipótesis nula

$$H_0 : \quad 0'3364 \quad 0'3525 \quad 0'2261 \quad 0'085$$

se tiene la siguiente tabla:

A_i	0	A	B	AB	
$P(A_i\|H_0)$	0'3364	0'3525	0'2261	0'085	1
O_i	118	95	121	30	364
E_i	122'45	128'31	82'3	30'94	364

La región crítica es de la forma

$$RC = \left\{ \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i} \geq 7'81 \right\}$$

el valor del estadístico es $27'0357$, que se corresponde con un p-valor de $0'122 \times 10^{-4}$, luego como era de esperar se rechaza la hipótesis nula.

El lector puede repetir el ejercicio utilizando el *método de la RV*.

8.9.7

Es una tabla de contingencia (véase la Sección 8.2.3). Mediante el *método de K. Pearson* la región crítica es de la forma

$$RC = \left\{ \sum_{\substack{i=1,2,3 \\ j=1,2,3}} \frac{(n_{ij} - \frac{n_{i.}n_{.j}}{n})^2}{\frac{n_{i.}n_{.j}}{n}} \geq \chi^2_{2\times 2;0'05} = 9'49 \right\}$$

el valor del estadístico es $33'3929$, luego se rechaza la hipótesis de independencia. Mediante el *test de la* RV la región crítica es de la forma

$$RC = \left\{ -2 \sum_{i,j} n_{ij} \ln \frac{\frac{n_{i.}}{n}\frac{n_{.j}}{n}}{\frac{n_{ij}}{n}} \geq \chi^2_{4;0'05} = 9'49 \right\}$$

el valor del estadístico es $33'8986$ luego se rechaza también la hipótesis nula.

8.9.8

Es un contraste de homogeneidad (véase la Sección 8.2.2). Mediante la tabla siguiente:

muestras	A_1	A_2	A_3	A_4	
1	3	7	7	7	24
2	3	5	5	7	20
3	9	5	5	6	25
4	4	6	8	2	20
	19	23	25	22	89

se pretende contrastar la hipótesis nula

$$H_0 : \ 0'2266 \ \ 0'2734 \ \ 0'2734 \ \ 0'2266$$

donde estas probabilidades p_i han sido calculadas bajo la hipótesis de normalidad, frente a la hipótesis alternativa $H_1 : p_{11}...p_{14}$ con $\sum_{j=1}^{4} p_{1j} = 1$, $p_{21}, ..., p_{24}$ con $\sum_{j=1}^{4} p_{2j} = 1$, $p_{31}, ..., p_{34}$ con $\sum_{j=1}^{4} p_{3j} = 1$, $p_{41}, ..., p_{44}$ con $\sum_{j=1}^{4} p_{4j} = 1$. La

$$RC = \left\{ \sum_i \sum_j \frac{(n_{ij} - \frac{n_{i.}n_{.j}}{n})^2}{\frac{n_{i.}n_{.j}}{n}} \geq \chi^2_{12;0'05} = 21 \right\}$$

el estadístico vale $9'21$; luego se acepta la hipótesis de homogeneidad, el p-valor es de $0'9555$. Obsérvese que al estar la hipótesis nula perfectamente especificada, el número de grados de libertad de la χ^2 es 12.

El *test de la RV* tiene de región crítica

$$RC = \left\{ -2 \sum_{\substack{i=1 \\ j=1}}^{4,4} n_{ij} \ln \frac{p_i}{n_{ij}/n_{i\cdot}} \geq 21 \right\}$$

el valor del estadístico es 15′7856, luego también se acepta la hipótesis nula.

8.9.9

a) Hay que contrastar

$$H_0 : \ 0'75, \ 0'25 \quad \text{frente a} \quad H_1 : p_1, p_2 \ \text{ con } \ p_1 + p_2 = 1$$

mediante el método de la *distancia de K. Pearson* (véase la Sección 8.2.1).

Se tiene la tabla

A_i	M	R	
$P(A_i\|H_0)$	0'75	0'26	1
O_i	7	3	10
$E_i = 10P(A_i\|H_0)$	7'5	2'5	1

La región crítica es de la forma

$$RC = \left\{ \sum_{i=1}^{2} \frac{(O_i - E_i)^2}{E_i} \geq \chi^2_{1;0'05} = 3'84 \right\}$$

el valor del estadístico es 0′133; luego se acepta el ajuste a la distribución 3:1, el p-valor es 0′715.

b) Por el resultado de la parte a) se toman como distribuciones marginales las dadas por (0′75, 0′25) y se trata de contrastar

$$H_0 : p_{ij} = p_{i\cdot} p_{j\cdot} \ \text{ i=1,2 } \ \text{j=1,2} \quad \text{con } \ p_{1\cdot} = p_{\cdot 1} = 0'75 \quad p_{2\cdot} = p_{\cdot 2} = 0'25$$

frente a

$$H_1 : p_{11}, p_{12}, p_{21}, p_{22} \quad \text{con } \ \sum_{i,j} p_{ij} = 1,$$

mediante el *método de la RV* como los EMV bajo H_1 son $\widehat{p}_{ij} = n_{ij}/n$, la región crítica es de la forma

$$RC = \left\{ -2 \sum_{\substack{i=1,2 \\ j=1,2}} n_{ij} \ln \frac{p_{i\cdot} p_{\cdot j}}{n_{ij}/n} \geq \chi^2_{3;0'05} = 7'81 \right\}$$

y el valor del estadístico es 4′5344, luego se acepta la hipótesis de independencia.

8.9.10

Se pretende contrastar $H_0 : F_X = F_Y$ frente a $H_1 : F_X \neq F_Y$ mediante el *test de Kolmogorov-Smirnov* para dos muestras (véase la Sección 8.5.1). La región crítica viene dada, véase la Tabla 8.5.1, por

$$RC = \{D_{mn} > d_{mn;\alpha q}\}$$

con

$$D_{mn} = \sup_x |F_m^*(x) - G_n^*(x)|$$

y donde $m = 10, n = 15$ y $\alpha = 0'05$.

Los cálculos pueden llevarse a cabo de la manera siguiente:

muestra	$\lvert F_{10}^*(x) - G_{15}^*(x)\rvert$	muestra	$\lvert F_{10}^*(x) - G_{15}^*(x)\rvert$
$y_{(1)} = 110'2$	0'0667	$x_{(7)} = 145'9$	0'233
$y_{(2)} = 120'1$	0'133	$y_{(8)} = 147'3*$	0'167
$y_{(3)} = 124'8$	0'200	$y_{(8)} = 147'3*$	0'267
$x_{(1)} = 125'6$	+0'100	$x_{(9)} = 147'9$	0'367
$x_{(2)} = 128'5$	0	$y_{(9)} = 148'9$	0'300
$y_{(4)} = 128'9$	0'067	$x_{(10)} = 150'2$	0'400
$y_{(5)} = 139'6$	0'133	$y_{(10)} = 150'7$	0'323
$x_{(5)} = 139'8$	0'033	$y_{(11)} = 151'3$	0'267
$y_{(6)} = 140'4$	0'100	$y_{(12)} = 152'3$	0'200
$x_{(4)} = 140'6$	0	$y_{(13)} = 153'5$	0'133
$x_{(5)} = 142'8$	0'100	$y_{(14)} = 155'6$	0'067
$x_{(6)} = 143'5$	0'200	$y_{(15)} = 160'1$	0
$y_{(7)} = 145'8$	0'133		

Se observa, que al estar repetido el valor $147'3$ ha sido asignado a la segunda muestra, si se hubiera asignado a la primera, hubiera quedado la secuencia primero $x_{(8)} = 147'3$ y luego $y_{(8)} = 147'3$ con lo que los números $0'167$ y $0'267$ se hubiera sustituido por $0'333$ y $0'267$ respectivamente que no afectan al valor del máximo que es $D_{10,15} = 0'4$. Por lo tanto la región crítica es

$$RC = \{D_{10,15} > d_{10,15;0'05} = 0'5\}$$

(véase la Tabla 10 al final del libro), luego se acepta la hipótesis de igualdad de las dos distribuciones.

8.9.11

Mediante el *test de Kolmogorov-Smirnov* para una muestra (véase la Sección 8.4.1), se trata de contrastar $H_0 : X \sim Normal(0, 1)$ frente a H_1 : cualquier otra distribución continua. La región crítica viene dada (véase la Tabla 8.4.1.1) por

$$RC = \{D_n > d_{n;\alpha}\}$$

con $n = 20$ y $\alpha = 0'05$.

Los cálculos pueden disponerse de la siguiente forma:

$x_{(i)}$	$\frac{i}{n}$	$F_0(x_{(i)})$	$\frac{i}{n} - F_0(x_i)$	$\frac{i-1}{n}$	$F_0(x_{(i)}) - \frac{i-1}{n}$
-1'81	0'05	0'035	0'015	0	0'035
-1'56	0'10	0'059	0'041	0'05	0'009
-1'14	0'15	0'127	0'023	0'10	0'027
-0'57	0'20	0'284	-0'084	0'15	0'134
-0'14	0'25	0'444	-0'194	0'20	0'244
-0'09	0'30	0'464	-0'164	0'25	0'214
0'12	0'35	0'548	-0'198	0'30	0'248
0'78	0'40	0'782	.0'382	0'35	0'432
1'05	0'45	0'853	-0'403	0'40	0'453
1'09	0'50	0'862	-0'362	0'45	0'412
1'41	0'55	0'921	-0'371	0'50	0'421
1'52	0'60	0'936	-0'336	0'55	0'386
1'60	0'65	0'945	-0'295	0'60	0'345
2'96	0'70	0'998	-0'298	0'65	0'348
3'34	0'75	0'999	-0'249	0'70	0'299
3'56	0'80	0'999	-0'199	0'75	0'249
3'99	0'85	0'999	-0'149	0'80	0'199
5'46	0'90	1	-0'1	0'85	0'150
7'62	0'95	1	-0'05	0'90	0'1
8'16	1	1	0	0'95	0'05

Por tanto $D_{20}^{+} = 0'041$, $D_{20}^{-} = 0'453$ y $D_{20} = 0'453$. De las tablas se obtiene $d_{20;0'05} = 0'294$ (véase la Tabla 7 al final del libro), luego se rechaza el ajuste a la distribución $N(0,1)$.

Para realizar el ajuste a una distribución normal, se calculan $\overline{x} = 1'8675$ y $s = 2'81308$, se trata ahora de contrastar

$$H_0 : X \sim N(1'87, 2'81) \quad \text{frente a} \quad H_1 : \text{ cualquier otra distribución continua.}$$

La región crítica es la misma de antes, pero donde el valor de $d_{n,\alpha}$ se toma de la tabla del *test de Kolmogorov-Sminov* cuando los parámetros se ajustan a partir de la muestra, $d_{20;0'05} = 0'192$ (véase la Tabla 8 del

Apéndice). Los cálculos pueden disponerse de la siguiente forma:

$x_{(i)}$	$\frac{i}{n}$	$F_0^*(x_{(i)})$	$\frac{i}{n} - F_0^*(x_i)$	$\frac{i-1}{n}$	$F_0^*(x_{(i)}) - \frac{i-1}{n}$
-1'81	0'05	0'095	-0'045	0	0'095
-1'56	0'10	0'111	-0'011	0'05	0'061
-1'14	0'15	0'142	0'008	0'10	0'042
-0'57	0'20	0'193	0'007	0'15	0'043
-0'14	0'25	0'237	0'013	0'20	0'037
-0'09	0'30	0'243	0'057	0'25	-0'007
0'12	0'35	0'267	0'083	0'30	-0'033
0'78	0'40	0'349	0'051	0'35	-0'001
1'05	0'45	0'385	0'065	0'40	-0'015
1'09	0'50	0'391	0'109	0'45	-0'059
1'41	0'55	0'435	0'115	0'50	-0'065
1'52	0'60	0'450	0'150	0'55	-0'100
1'60	0'65	0'462	0'188	0'60	-0'138
2'96	0'70	0'651	0'049	0'65	0'001
3'34	0'75	0'699	0'051	0'70	-0'001
3'56	0'80	0'726	0'074	0'75	-0'024
3'99	0'85	0'775	0'075	0'80	-0'025
5'46	0'90	0'899	0'001	0'85	0'049
7'62	0'95	0'980	-0'030	0'90	0'080
8'16	1	0'987	0'013	0'95	0'037

donde ahora los valores de $F_0^*(x)$ se corresponden con los de la distribución $N(1'87, 2'81)$. Se tiene que $D_{20}^+ = 0'188$, $D_{20}^- = 0'095$ y $D_{20} = 0'188$ luego la región crítica es (véase la Tabla 8.4.1.1)

$$RC = \{D_{20} > 0'192\}$$

y se acepta el ajuste, de hecho el p-valor es aproximadamente $0'477$.

8.9.12

Si se denota por x_i al resultado obtenido con el parasicólogo presente, e y_i al resultado obtenido la semana anterior, la sucesión $d_i = x_i - y_i$ debe contener aproximadamente el mismo número de diferencias positivas que negativas, si no hay predicción paranormal. Por lo tanto se trata de ver, mediante el test de los signos, si la sucesión

$$+ \; - \; - \; + \; - \; + \; + \; + \; +$$

es compatible con $H_0 : k_{\frac{1}{2}} = 0$ frente a $H_1 : k_{\frac{1}{2}} > 0$, ya que en este último caso el número de diferencias positivas sería mayor. Si se denota por Y al número de diferencias positivas es $Y = 6$ y la región crítica viene dada por $RC = \{Y \geq c_1\}$ donde la distribución de Y bajo H_0 es $Binomial(9, 0'5)$ (véase la Sección 8.4.2).

Con $\alpha = 0'05$, para que $P\{Y \geq c_1|H_0\} \leq 0'05$ ha de ser $c_1 = 8$ puesto que $P\{Y \geq 8|H_0\} = 0'0196$, $P\{Y \geq 9|H_0\} = 0'0899$ (véase la Tabla 1 del Apéndice), luego se acepta la hipótesis nula y por lo tanto se concluye que no hay inducción paranormal.

También es aplicable el *test de Wilcoxon de los rangos signados* (véase la Sección 8.4.2). Se trata de calcular el *estadístico de Wilcoxon* $T^+ = \sum_{i=1}^{9} Z_i r(|D_i|)$ con $Z_i = 1$ para los rangos correspondientes a diferencias positivas (véase la Expresión (8.4.2.1)).

d_i :	2'5	-0'2	-1'4	0'8	-0'3	4'7	5	13'8	2'4
$\|d_i\|$:	2'5	0'2	1'4	0'8	0'3	4'7	5	13'8	2'4
$r(\|d_i\|)$:	6	1	4	3	2	7	8	9	5

Por lo tanto $T_9^+ = 6 + 3 + 7 + 8 + 9 + 5 = 38$. De acuerdo con la Tabla 8.4.2.2 la región crítica es $RC = \{T_9^+ > c_1\}$, de la restricción de tamaño, ha de ser $P\{T_9^+ > c_1\} \geq 0'05$ de donde se sigue que $c_1 = 36$, luego se rechaza la hipótesis nula.

Como se ve, con los dos tests se han obtenido resultados distintos; en general el *test de los rangos signados de Wilcoxon* utiliza de manera más matizada la información —ya que utiliza los rangos— que el de los signos —que solo utiliza las diferencias—; de todas formas, al no haber demasiada diferencia, se debería recurrir a un estudio posterior con un tamaño muestral mayor.

8.9.13

Mediante el *test de Mann-Whitney-Wilcoxon* se trata de contrastar

$$H_0 : F_X = F_Y \quad \text{frente a } H_1 : F_X \neq F_y$$

la región crítica es $RC = \{U > c_1 \text{ o } U < c_2\}$ con $U =$n^o de veces que la primera muestra está por encima de la segunda (véase la Sección 8.5.3). El valor del estadístico es

$$U = 7 + 7 + 7 + 10 + 10 + 10 + 15 + 15 + 15 + 15 + 15 + 15 = 141.$$

Para determinar las constantes, se utiliza la restricción de tamaño mediante

$$P\{U > c_1|H_0\} \leq 0'025$$

y empleando que

$$\frac{U - \frac{mn}{2}}{\sqrt{\frac{1}{12}(m+n+1)mn}} \xrightarrow{\mathcal{L}} Z \sim N(0,1)$$

con $m = 12$ y $n = 16$ queda

$$P\{Z > \frac{c_1 - 96}{21'54}\} = 0'025$$

con lo que $c_1 - \frac{96}{21'54} = 1'96$ y por lo tanto $c_1 = 138'22$ y $c_2 = 96 - 1'96 \cdot 21'54 = 53'78$. La región crítica es de la forma

$$RC = \{U > 138'22 \ \text{ ó } \ U < 53'78\}$$

(véase la Tabla 8.5.3.1), luego se rechaza la hipótesis nula; los *libros de Estadística* y *generales* no tienen el mismo número de páginas.

Mediante la suposición de que las muestras provienen de sendas poblaciones normales $X \sim N(\mu, \sigma_1)$ e $Y \sim N(\mu_2, \sigma_2)$ se empieza contrastando la igualdad de varianzas

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1 \quad \text{frente a } H_1 : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$$

con lo que la región crítica es de la forma (véase, en la Sección 5.2.2, la Tabla 5.2.2.4)

$$RC = \left\{ \frac{S_1^2}{S_2^2} \geq f_{m-1,n-1;\frac{\alpha}{2}} \quad \text{si} \quad S_1^2 \geq S_2^2 \quad \text{ó} \quad \frac{S_2^2}{S_1^2} \geq f_{n-1,m-1;\frac{\alpha}{2}} \quad \text{si} \quad S_1^2 < S_2^2 \right\}$$

se tiene que $m = 12$, $n = 16$, $S_1^2 = 19738'8$, $S_2^2 = 31977'3$, $\frac{S_2^2}{S_1^2} \geq f_{n-1,m-1;\frac{\alpha}{2}} = f_{15,11;0'025} = 3'33$. Como $1'62 < 3'33$, se acepta la igualdad de las varianzas.

Si se contrasta ahora la igualdad de medias

$$H_0 : \mu_1 = \mu_2 \ \text{ frente a } H_1 : \mu_1 \neq \mu_2$$

la región crítica es (véase la Tabla 5.2.2.3)

$$RC = \left\{ |\overline{x} - \overline{y}| \geq t_{m+n-2;\frac{\alpha}{2}} S_p \sqrt{\frac{1}{m} + \frac{1}{n}} \right\}$$

con $S_p^2 = \frac{(m-1)S_1^2 + (n-1)S_2^2}{m+n-2} = \frac{11 \cdot 19738'8 + 15 \cdot 31977'3}{26} = 26{,}799'4$, $S_p = 163'705$ y $\sqrt{\frac{1}{12} + \frac{1}{16}} = 0'382$, luego la región crítica es

$$RC = \{|\overline{x} - \overline{y}| \geq 2'068 \cdot 163'705 \cdot 0'382 = 129'323\}$$

y como el valor del estadístico es $|\overline{x} - \overline{y}| = 316'67 - 208'063 = 108'607$ se acepta la hipótesis de igualdad de medias y por lo tanto, bajo la suposición de normalidad, ambas poblaciones son iguales.

No se ha contrastado sin embargo la hipótesis de normalidad de muestra, que se deja como ejercicio para el lector (véase la Sección 8.4.3).

8.9.14

a) Se trata de contrastar mediante el *test de los signos*

$$H_0 : k_{0'75} = 75'5 \ \text{ frente a } \quad H_1 : k_{0'75} \neq 0'75$$

La región crítica es de la forma

$$RC = \{Y \geq c_1 \ \text{ ó } \ \mathrm{Y} \leq c_2 \ \text{ con } \ \mathrm{c}_1 > c_2\}$$

con $Y = \mathrm{n}^o$ de diferencias positivas en la muestra $x_i - 75'5$, que en este caso es de 4 (véase la Sección 8.4.2 y la Tabla 8.4.2.1).

Bajo H_0 la distribución de $Y \sim Binomial(8, 0'25)$, luego para que sea

$$P\{Y \geq c_1 | H_0\} \leq 0'05 \quad P\{Y \leq c_2 | H_0\} \leq 0'05$$

de la primera ecuación, mediante las tablas de la *distribución Binomial*, se obtiene $P\{Y \geq 4|H_0\} = 0'1138$ y $P\{Y \geq 5|H_0\} = 0'0273$ luego $c_1 = 5$ y la región crítica es

$$RC = \{Y \geq 5 \ \text{ ó } \ Y < 0\},$$

como $4 \ngeq 5$ se acepta H_0.

b) Mediante el *test de Kolmogorov-Smirnov* para dos muestras y para contrastar

$$H_0 : F_X = F_Y \quad \text{frente a} \quad H_1 : F_X > F_Y$$

La región crítica es

$$RC = \{D^+_{m,n} > d^+_{m,n;\alpha}\}$$

con $D^+_{m,n} = \sup_x(F^*_m(x) - G^*_n(x))$ $m = n = 8$ (véase la Sección 8.5.1 y la Tabla 8.5.1). Para el cálculo del estadístico se utiliza la siguiente tabla:

datos ordenados	$n(F^*_n(x) - G^*_n(x))$
$x_{(1)} = 13$	1
$x_{(2)} = 53$	2
$x_{(3)} = 59$	3
$x_{(4)} = 75$	4
$x_{(5)} = 76$	5
$x_{(6)} = 79$	6
$y_{(1)} = 96$	5
$y_{(2)} = 102$	4
$y_{(3)} = 104$	3
$y_{(4)} = 107$	2
$y_{(5)} = 110$	1
$x_{(7)} = 129$	2
$y_{(6)} = 130$	1
$y_{(7)} = 133$	0
$x_{(9)} = 138$	1
$y_{(8)} = 141$	0

De la restricción de tamaño ha de ser

$$P\{D_{88}^{+} < d_{88;0'1}^{+}\} \leq 0'1$$

y de la Tabla 11 del Apéndice

$$d_{88;0'1}^{+} = \frac{4}{8}$$

Como el valor del estadístico es $D_{88}^{+} = \frac{6}{8}$ y $\frac{6}{8} > \frac{4}{8}$ se rechaza la hipótesis de homogeneidad.

c) Para contrastar la igualdad de ambas poblaciones mediante el *test de la mediana* $H_0 : M_X = M_Y$ frente a $H_1 : M_X \neq M_Y$, la región crítica correspondiente viene dada por

$$RC = \{T > c_1 \ \text{ó} \ T < c_2 \ \text{ con } \ c_1 > c_2\}$$

siendo T el número de observaciones de la primera muestra que quedan por debajo de la mediana muestral (véase la Sección 8.5.2 y la Tabla 8.5.2).

Como $m = n = 8$, $N = m + n = 16$ y $\left[\frac{N}{2}\right] = 8$ la mediana muestral es $Z_{([\frac{N}{2}]+1)} = Z_{(9)} = y_{(3)} = 104$. La distribución de T es

$$P\{T = t|H_0\} = \frac{\binom{m}{t}\binom{n}{k-t}}{\binom{m+n}{k}}$$

con $k = [\frac{N}{2}] = 8$.

Debido a que $T = 6$, y $P\{T > 6|H_0\} = 0'00497 + 0'0000777 = 0'0050477$, se tiene que el p-valor es menor que $0'1$, luego se rechaza que las dos poblaciones sean la misma.

d) Mediante el *test de MWW* se pretende contrastar

$$H_0 : F_X = F_Y \ \text{ frente a } H_1 : F_X > F_Y.$$

Se introduce el estadístico $U = \text{n}^o$ de veces que la primera muestra queda por debajo estrictamente de la segunda muestra (véase la Sección 8.5.3) y la región crítica es de la forma (véase la Tabla 8.5.3.1)

$$RC = \{U > c_1\},$$

mediante la restricción de tamaño ha de ser

$$P\{U > c_1|H_0\} \leq 0'1$$

de donde se obtiene $c_1 = 44$ (véase la Tabla 12 al final del libro). Como el valor del estadístico es

$$U = 8 + 8 + 8 + 8 + 8 + 8 + 3 + 1 = 52$$

se rechaza H_0 y las dos muestras no son homogéneas.

8.9.15
Para contrastar mediante el *test de MWW*.

$$H_0 : F_X = F_Y \quad \text{frente a} \quad H_1 : F_X \neq F_Y$$

la forma de la región crítica es

$$RC = \{U > c_1 \quad \text{ó} \quad U < c_2 \quad \text{con} \quad c_1 > c_2\}$$

(véase la Tabla 8.5.3.1 de la Sección 8.5.3). Donde U es el número de veces que la segunda muestra está por encima estrictamente de la primera, es decir

$$\begin{aligned} U &= 11 + 11 + 10 + +9 + 9 + 8 + 8 + 8 + 7 + 5 + 3 \\ &= 89 \end{aligned}$$

Para satisfacer la restricción de tamaño, ha de cumplirse

$$P\{U > c_1\} \leq 0'025$$

como

$$\frac{U - \frac{mn}{2}}{\sqrt{\frac{mn(m+n+1)}{12}}} \xrightarrow{\mathcal{L}} Z \sim N(0,1)$$

ha de ser

$$P\{Z > \frac{c_1 - 72}{17'3205}\} = 0'025$$

de donde $c_1 = 72 + 1'96 \cdot 17'3205 = 105'948$ y por simetría $c_2 = 72 - 1'96 \cdot 17'3205 = 38'052$. Por lo tanto se acepta que ambas poblaciones son iguales y que los dos grupos son homogéneos.

8.9.16
Como gran parte de los valores de la muestra de los niños quedan por encima, vamos a contrastar, inicialmente mediante el *test de Kolmogorov-Smirnov*

$$H_0 : F_X = F_Y \quad \text{frente a} \quad H_1 : F_X > F_y$$

con lo que la región crítica es de la forma

$$RC = \{D^+_{m,n} > d^+_{m,n;\alpha}\}$$

(véase la Tabla 8.5.1.1). De la Tabla 11 del Apéndice, $d^+_{10,190;0'05} = \frac{5}{10}$ y

construyendo la tabla:

datos ordenados	$n(F_n^*(x) - G_n^*(x))$	datos ordenados	$n(F_n^*(x) - G_n^*(x))$
$x_{(1)} = 7$	1	$y_{(3)} = 50$	5
$x_{(2)} = 9$	2	$x_{(9)} = 55$	6
$x_{(3)} = 16$	3	$x_{(10)} = 58$	7
$x_{(4)} = 16*$	4	$y_{(4)} = 65*$	6
$x_{(5)} = 16*$	5	$y_{(5)} = 65*$	5
$x_{(6)} = 22$	6	$y_{(6)} = 69$	4
$x_{(7)} = 26$	7	$y_{(7)} = 72$	3
$x_{(8)} = 40$	8	$y_{(8)} = 86$	2
$y_{(1)} = 41$	7	$y_{(9)} = 104$	1
$y_{(2)} = 45$	6	$y_{10} = 113$	0

se obtiene $D_{m,n}^{+} = \frac{8}{10}$, donde se observa que los valores repetidos, que aparecen con un asterisco detrás, no afectan al valor de estadístico. Como $0'8$ pertenece a la RC se rechaza la hipótesis nula.

Mediante el *test de la mediana* se trata de contrastar

$$H_0 : M_X = M_Y \text{ frente a } H_1 : M_X < M_Y$$

la región crítica es de la forma

$$RC = \{T > c_1\}$$

siendo $T = \text{n}^o$ de observaciones de la primera muestra que quedan por debajo de la mediana muestral (véase la Sección 8.5.2). Como $m = n = 10, N = m + n = 20$ y $K = \left[\frac{N}{2}\right] = 10$, la mediana muestral es $Z_{\left(\left[\frac{N}{2}+1\right]\right)} = Z_{(11)} = y_{(3)} = 50$. Bajo H_0 sustituyendo en

$$P\{T = t|H_0\} = \frac{\binom{m}{t}\binom{n}{k-t}}{\binom{m+n}{k}}$$

como $T = 8 \;\; P\{T > 8|H_0\} = 0'0005713 + 0'0000057 = 0'00057$ y el p-valor es menor que $0'05$, luego se rechaza H_0.

Por último, mediante el *test de MWW* (véase la Sección 8.5.3), se trata de contrastar

$$H_0 : F_X = F_Y \quad \text{frente a} \quad H_1 : F_X > F_Y$$

con lo que se rechaza cuando

$$RC = \{U > c_1\}$$

donde U es el número de veces que la primera muestra queda estrictamente por debajo de la segunda (véase la Tabla 8.5.3.1). De la Tabla 12 al final del libro se obtiene que como

$$P\{U > c_1|H_0\} \leq 0'05$$

ha de ser $c_1 = 72$ y como el valor del estadístico es $U = 6\cdot 10+7+2\cdot 10+7 = 94$ que queda en la región crítica, luego se rechaza H_0.

8.9.17

Mediante el *test de Kendall* (véase la Sección 8.6.1), la forma de la región crítica para contrastar $H_0 : F(x, y) = F_1(x)F_2(y)$ frente a $H_1 : F(x, y) \neq F_1(x)F_2(y)$ es

$$RC = \{|T| > c\}$$

(véase la Tabla 8.6.1.1). El valor del *estadístico de Kendall* es $T = \dfrac{2P}{\binom{n}{2}} - 1 = \dfrac{2 \cdot 44}{\binom{15}{2}} - 1 = -0'162$, ya que el número de concordancias es 44 y el de discordancias es 61.

Para la determinación de las constantes, como ha de cumplirse que

$$P\{T > c|H_0\} \leq 0'025,$$

mediante la distribución asintótica $3\sqrt{\frac{n(n-1)}{2(2n+5)}}T \to Z \sim N(0, 1)$ ha de verificarse que

$$P\left\{ Z > c\ \sqrt{\frac{n(n-1)}{2(2n+5)}} \middle| H_0 \right\} = P\{Z > c\ 5'196\} \leq 0'025$$

de donde $z_{0'025} = c\ 5'196 = 1'96$, por lo que $c = 0'377$.

Como $-0'162$ no pertenece a la RC se acepta la hipótesis de independencia, el p-valor es $0'4$.

Mediante el *coeficiente de correlación de Spearman* , la forma de la región crítica es

$$RC = \{|R| > c\}$$

donde

$$R = \frac{12}{n(n^2-1)} \left\{ \sum_{i=1}^{15} iS_i - \frac{n(n+1)^2}{4} \right\}$$

(véase la Sección 8.6.2 y la Tabla 8.6.2.1). Como ha de ser

$$P\{|R| > c|H_0\} \leq 0'05$$

empleando la distribución asintótica

$$\sqrt{n-1}\ R \xrightarrow{\mathcal{L}} Z \sim N(0,1)$$

se sigue

$$P\{R > c|H_0\} \simeq P\{Z > c\sqrt{n-1}\} \leq 0'025$$

por lo que $c\ 3'742 = 1'96$, luego $c = 0'524$.

Como el valor del estadístico es $-0'2643$ se acepta la hipótesis de independencia, el p-valor es $0'3227$.

DISTRIBUCIONES USUALES

$Bernoulli(\theta)$	$\theta \in [0,1]$
función de masa	$P\{X = x\|\theta\} = \theta^x, \;\; x \in \{0,1\}$
media	$E[X] = \theta$
varianza	$V[X] = \theta(1-\theta)$
función característica	$\varphi_X(t) = 1 + \theta(e^{it} - 1)$

$Binomial(n,\theta)$	$n \in \{1,2,...\}$, $\theta \in [0,1]$
función de masa	$P\{X = x\|\theta\} = \binom{n}{x}\theta^x(1-\theta)^{n-x}$, $x \in \{0,1,...,n\}$
media	$E[X] = \theta$
varianza	$V[X] = n\theta(1-\theta)$
función característica	$\varphi_X(t) = [1 + \theta(e^{it} - 1)]^n$

$Geométrica(\theta)$	$\theta \in [0,1]$
función de masa	$P\{X = x\|\theta\} = \theta(1-\theta)^{x-1}, \quad x = 1,2,...$
media	$E[X] = \frac{1}{\theta}$
varianza	$V[X] = \frac{1-\theta}{\theta^2}$
función característica	$\varphi_X(t) = (\theta e^{it})/[1 - (1-\theta)e^{it}]$

$Hipergeométrica(N,N,n)$	$N > 0, 0 \leq M \leq N, n \in \{1,2,...,N\}$, M y N *enteros*
función de masa	$P\{X = x\|N,M,n\} = \dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$, $x \in \{0,1,...,n\}$
media	$E[X] = \frac{nM}{N}$
varianza	$V[X] = \frac{nM(N-M)(N-n)}{N^2(N-1)}$

$Binomial\ Negativa(r,\theta)$	$r \geq 1$ entero, $\theta \in [0,1]$
función de masa	$P\{X=x\|r,\theta\} = \binom{r+x-1}{x}\theta^r(1-\theta)^x,$ $x=0,1,...$
media	$E[X] = \frac{r(1-\theta)}{\theta}$
varianza	$V[X] = \frac{r(1-\theta)}{\theta^2}$
función característica	$\varphi_X(t) = [\theta/(1-(1-\theta)e^{it})]^r$

$Poisson(\theta)$	$\theta \in (0,+\infty)$
función de masa	$P\{X=x\|\theta\} = e^{-\theta}\frac{\theta^x}{x!}, \quad x=0,1,...$
media	$E[X] = \theta$
varianza	$V[X] = \theta$
función característica	$\varphi_X(t) = e^{\theta(e^{it}-1)}$

$Uniforme(\theta_1,\theta_2)$	$-\infty < \theta_1 < \theta_2 < +\infty$
función de densidad	$f(x\|\theta_1,\theta_2) = \frac{1}{\theta_2-\theta_1}I_{(\theta_1,\theta_2)}(x)$
media	$E[X] = \frac{\theta_1+\theta_2}{2}$
varianza	$V[X] = \frac{(\theta_2-\theta_1)^2}{12}$
función característica	$\varphi_X(t) = (e^{it\theta_2} - e^{it\theta_1})/(it(\theta_2-\theta_1))$

$Normal(\mu,\sigma)$	$\mu \in \mathbb{R} \quad \sigma \in (0,+\infty)$
función de densidad	$f(x\|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad x \in \mathbb{R}$
media	$E[X] = \mu$
varianza	$V[X] = \sigma^2$
función característica	$\varphi_X(t) = e^{it\mu-\frac{1}{2}\sigma^2t^2}$

$Exponencial(\theta)$	$\theta \in (0,+\infty)$
función de densidad	$f(x\|\theta) = \theta e^{-\theta x}I_{(0,+\infty)}(x)$
media	$E[X] = \frac{1}{\theta}$
varianza	$V[X] = \frac{1}{\theta^2}$
función característica	$\varphi_X(t) = [1-(it)/a]^{-p}$

$Doble\ Exponencial(\mu,\sigma)$	$\mu \in \mathbb{R} \quad \sigma \in (0,+\infty)$
función de densidad	$f(x\|\mu,\sigma) = \frac{1}{2\sigma}e^{-\frac{\|x-\mu\|}{\sigma}}, \quad x \in \mathbb{R}$
media	$E[X] = \mu$
varianza	$V[X] = 2\sigma^2$
función característica	$\varphi_X(t) = e^{i\mu t}/(1+\sigma^2t^2)$

$Gamma(a,p)$	$a \in (0,+\infty) \quad p \in (0,+\infty)$
función de densidad	$f(x\|a,p) = \frac{a^p}{\Gamma(p)} e^{-ax} x^{p-1} I_{(0,+\infty)}(x)$
media	$E[X] = \frac{p}{a}$
varianza	$V[X] = \frac{p}{a^2}$
función característica	$\varphi_X(t) = [1-(it)/a]^{-p}$

χ^2_p	$p \in (0,+\infty)$
función de densidad	$f(x\|p) = \frac{1}{2^{p/2}\Gamma(\frac{p}{2})} e^{-\frac{1}{2}x} x^{p-1} I_{(0,+\infty)}(x)$
media	$E[X] = p$
varianza	$V[X] = 2p$
función característica	$\varphi_X(t) = (1-2it)^{-\frac{p}{2}}$

$Beta(p,q)$	$p \in (0,+\infty) \quad q \in (0,+\infty)$
función de densidad	$f(x\|p,q) = \frac{1}{B(p,q)} x^{p-1}(1-x)^{q-1} I_{(0,1)}(x)$
media	$E[X] = \frac{p}{p+q}$
varianza	$V[X] = \frac{pq}{(p+q)^2(p+q+1)}$

$Cauchy(\theta,\sigma)$	$\theta \in \mathbb{R} \quad \sigma \in (0,+\infty)$
función de densidad	$f(x\|\theta,\sigma) = \frac{1}{\pi\sigma} \frac{1}{1+\left(\frac{x-\theta}{\sigma}\right)^2}, \quad x \in \mathbb{R}$
media	no existe
varianza	no existe
función característica	$\varphi_X(t) = e^{it\theta-\sigma\|t\|}$

T_p *de Student*	$p \in (0,+\infty)$
función de densidad	$f(x\|p) = \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{1}{2}\right)} \frac{1}{\sqrt{p}} \left(1+\frac{x^2}{p}\right)^{-\frac{p+1}{2}}, \quad x \in \mathbb{R}$
media	$E[X] = 0 \quad \text{si} \quad p > 1$
varianza	$V[X] = \frac{p}{p-2}, \text{ si} \quad p > 2$

$F_{p,q}$ *de Snedecor*	$p \in (0,+\infty), q \in (0,+\infty)$
función de densidad	$f(x\|p,q) =$
	$= \frac{\Gamma\left(\frac{p+q}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{q}{2}\right)} \left(\frac{p}{q}\right)^{\frac{p}{2}} x^{\frac{p-2}{2}} \left(1+\frac{p}{q}x\right)^{-\frac{p+q}{2}} I_{(0,+\infty)}(x)$
media	$E[X] = \frac{p}{p-2}, \text{ si} \quad p > 2$
varianza	$V[X] = 2\left(\frac{q}{q-2}\right)^2 \frac{p+q-2}{p(q-4)}, \quad \text{si} \quad q > 4$

Gamma Inversa(a,p)	$a \in (0,+\infty), p \in (0,+\infty)$
función de densidad	$f(x\|a,p) = \frac{a^p}{\Gamma(p)} e^{-\frac{a}{x}} x^{-(p+1)} I_{(0,+\infty)}(x)$
media	$E[X] = \frac{a}{p-1}$, si $p > 1$
varianza	$V[X] = \frac{a^2}{(p-1)^2(p-2)}$ si $p > 2$

Normal $k-$*dimensional*$(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$\boldsymbol{\mu} \in \mathbb{R}^k, \boldsymbol{\Sigma}$ matriz $k \times k$ simétrica definida positiva
función de densidad	$f(x\|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{k}{2}} \|\Sigma\|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})}$, $\mathrm{x} \in \mathbb{R}^k$
media	$E[\mathbf{X}] = \boldsymbol{\mu}$
matriz de covarianzas	$\boldsymbol{\Sigma}$
función característica	$\varphi_X(t_1, ..., t_k) = e^{it' \boldsymbol{\mu} - \frac{1}{2} t' \boldsymbol{\Sigma} t}$

Multinomial$(n, \theta_1, ..., \theta_k)$	$n > 0$ entero, $\theta_i \in (0,1)$, $i = 1, ..., k$, $\sum_{i=1}^{k} \theta_i = 1$
función de masa	$P\{X_1 = x_1, ..., X_k = x_k \vert \overrightarrow{\theta}\}$
	$= \frac{n!}{\prod_{i=1}^{k} x_i!} \prod_{i=1}^{k} \theta_i^{x_i}$, $x_i = 0, 1, ..., n, i = 1, ..., k, \sum_{i=1}^{k} x_i = n$
media	$E[X_i] = n\theta_i$
varianza	$V[X_i] = n\theta_i(1-\theta_i)$
covarianzas	$Cov(X_i, X_j) = -n\theta_i\theta_j$
función característica	$\varphi_X(t_1, ..., t_k) = (\theta_1 e^{it_1} + \theta_2 e^{it_2} + ... + \theta_k e^{it_k})^n$

T de *Student* $k-$*dimensional*$(p, \boldsymbol{\mu}, \boldsymbol{\Sigma})$	$p > 0, \boldsymbol{\mu} \in \mathbb{R}^k, \boldsymbol{\Sigma}$ matriz $k \times k$ simétrica definida positiva
función de densidad	$f(x\|p, \boldsymbol{\mu}, \boldsymbol{\Sigma}) =$
	$= \frac{\Gamma\left(\frac{k+p}{2}\right)}{(p\pi)^{\frac{k}{2}} \Gamma(\frac{k}{2}) \|\Sigma\|^{\frac{1}{2}}} \left[1 + \frac{(\mathbf{x}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})}{p}\right]^{-\frac{k+p}{2}}$
	$\mathbf{x} \in \mathbb{R}^k$
media	$E[\mathbf{X}] = \boldsymbol{\mu}$, si $p > 1$
matriz de covarianzas	$p\boldsymbol{\Sigma}/(p-2)$, si $p > 2$

TABLAS

TABLA 0

Números aleatorios

03731	59936	58116	39874	93896	42327	31159	72287	66279	26356
18090	49516	58621	78150	21823	02400	34007	42346	59754	71470
52234	17091	56495	74247	94785	37236	92526	24496	96456	11684
09936	48962	26323	69783	57220	30463	87227	24368	44320	74139
08773	18567	11844	49743	12023	98081	29493	94179	92819	54374
56563	66101	62681	41375	85594	47660	03878	81855	14889	78637
66718	75178	88893	05298	80729	32281	84366	43146	46277	53036
32876	18793	39096	58810	12739	59161	40034	63282	93654	91372
49826	61427	97495	25215	11243	28205	59066	94455	56348	04036
71051	15390	10730	90834	76780	52007	38345	09287	46355	42874
65517	68222	51977	54559	41626	84072	68417	99378	98840	98135
56558	23322	78797	89739	17783	08111	85513	50612	10740	55402
71144	85714	49586	49505	86744	35167	32337	51161	08045	55793
19714	11208	02202	43267	22846	52008	99408	71784	26445	18016
48755	77145	24301	94688	84466	04791	54548	80117	77558	09621
27370	84604	32283	24361	24342	28098	90886	30371	27943	13315
18567	53691	91653	46671	29586	49136	31216	57727	48889	87194
70766	97857	20486	14835	61349	36502	58310	19503	85001	82627
71699	88864	23062	35265	85636	99498	19581	90749	27163	27239
59565	44695	15960	00501	90406	78752	12224	20661	33351	13568
71866	37343	20163	36808	54129	46029	42793	28470	95834	31839
54111	25499	83727	38030	51360	15434	24935	98258	94905	20242
51843	63171	08509	70777	16463	14003	85017	27355	83598	23494
25849	19785	14483	56563	95565	64385	81611	56672	42055	75771
53203	34049	19923	61044	71986	21351	37304	95454	21119	93850
93842	67539	46391	69979	69835	49199	82403	99682	59341	39001
07830	37887	04196	35970	32637	93412	50504	58470	47718	98525
98888	82866	35038	49706	32410	80637	28218	20287	01202	17155
20834	30485	95188	09270	17256	72806	69249	69620	13563	28823
60882	65583	34445	11147	26057	59819	70534	04496	53864	34269

16906	97573	44574	30028	65405	90348	57805	64883	44410	53796
44664	15801	66124	95049	63615	95067	43491	02686	00875	54906
67898	72763	62437	94103	81477	00874	64676	55592	12902	08992
61914	33089	79549	04573	00569	91255	24138	37883	55766	35267
89086	35035	95056	73584	79702	35398	99431	01685	74790	49195

48817	67578	17668	63797	50544	86813	73963	76120	01504	72763
73008	91396	48775	42795	98540	78504	07933	57598	86875	36260
58983	93406	69977	58860	62419	17339	45154	56550	15193	50463
42035	79372	16781	75021	86335	17428	08197	47392	57448	36871
93250	18519	49350	79561	44977	25366	78441	17901	29572	99810

21035	45555	32458	10074	23234	01836	38688	05104	02756	85452
39370	54207	28591	24715	09289	62178	44244	91678	36633	02256
94356	58771	40787	00603	23165	28637	20704	38786	00417	16653
54054	80201	11919	42317	18801	28801	86575	88994	25073	99888
93797	51409	35901	24259	04516	04516	70437	70437	10088	28203

56358	39689	90343	26717	92995	07617	38946	47500	51326	12283
16700	55656	06408	44361	72858	24768	50568	43046	50951	50875
84568	24527	14502	87238	34922	77819	60395	83548	43300	57738
38380	34342	43044	12909	45233	25839	68127	96675	47099	10557
57492	93825	54974	70866	60923	20115	14044	17011	95523	38618

71751	17175	52210	48425	44245	89159	43718	14814	63810	55545
33728	61666	01980	81066	26249	00336	80284	82283	54691	48444
77537	25939	83182	78936	81098	61950	44529	35795	05013	03379
78884	58147	49099	78252	16807	94604	00053	21029	62437	99665
10923	92431	15806	87224	91665	18996	96505	40259	94288	72236

85389	88642	61135	87472	16355	82284	61555	39269	07019	75607
47819	20731	05178	51186	92979	53899	60733	94935	76106	40610
01456	32067	28091	88240	99237	94240	51556	86121	13655	47356
77116	51260	31172	55746	89696	06490	14108	51894	15770	26669
92174	28636	20037	25215	25222	88754	21522	83252	58422	41380

74984	87712	62719	02682	10201	92109	27856	25661	14344	89097
64348	83806	28773	76463	55750	82511	78111	99265	84426	60215
97753	41035	02886	25076	06552	75500	27781	51597	55188	41520
49777	82984	10835	54075	35452	33115	16944	01325	48700	49526
16945	13245	16191	46311	55895	74107	94525	78469	06300	99924

96775	22368	34924	55104	38302	30872	81828	44426	99356	29539
12096	40082	72392	40501	44143	68603	98721	83030	91607	65620
10812	66958	93886	12080	23447	15277	11060	00087	54974	83563
18838	63294	42021	33734	75687	67508	92871	63282	48928	57164
16728	87890	58794	73370	77844	16632	86356	83481	48185	68137

59018	70951	56109	45946	35837	50485	92141	67001	57870	80116
22173	89400	29758	22532	16594	02574	21841	89595	37099	09814
63334	55543	43358	10390	95411	66037	94403	94403	26503	47418
60049	19647	34278	29364	84476	90925	31518	31518	51548	71003
49195	23241	14430	34797	96085	84084	26133	26133	79370	54447

21331	25442	41125	33737	07963	09059	33459	14156	80115	00959
17608	14940	89081	22745	40099	59954	62155	26265	02189	24789
47789	11834	00504	91544	49996	47579	49165	48306	03305	30593
92104	02180	79405	85389	87941	67983	24990	93124	60298	81640
00557	51395	66954	42342	27685	99855	22622	44136	37489	30013

64397	11830	08081	51312	15377	79516	13860	95025	79116	43855
36423	38185	28256	04511	56711	39955	59388	85275	14749	83563
73386	80056	02085	28812	48829	26309	35127	25673	99728	47273
05468	02671	50314	44471	74079	89479	91760	91687	31633	30680
15587	01921	42513	36709	62067	84737	13908	92272	59455	46109

82874	95632	28276	75931	76170	79358	51907	16019	66036	67777
67739	95300	37587	80249	03114	79818	06707	34073	75839	36184
96389	22799	38524	45320	49416	08192	26820	72832	31844	50383
20355	86729	82586	98687	39451	98536	64621	01169	04240	99461
22219	56520	33861	37342	63848	10932	00839	57839	65959	51730

TABLA 1

Función de masa de la v.a. $Binomial(n, p)$

$$P\{X = x\} = \binom{n}{x} p^x (1-p)^{n-x} \quad n = 1, 2, ..., 10$$

n	x \ θ	.01	.05	.10	.15	.20	.25	.30	$\frac{1}{3}$	.35	.40	.45	.49	.50
2	0	.9801	.9025	.8100	.7225	.6400	.5625	.4900	.4444	.4225	.3600	.3025	.2601	.2500
	1	.0198	.0950	.1800	.2550	.3200	.3750	.4200	.4444	.4550	.4800	.4950	.4998	.5000
	2	.0001	.0025	.0100	.0225	.0400	.0625	.0900	.1111	.1225	.1600	.2050	.2401	.2500
3	0	.9703	.8574	.7290	.6141	.5120	.4219	.3430	.2963	.2746	.2160	.1644	.1327	.1250
	1	.0294	.1354	.2430	.3251	.3840	.4219	.4410	.4444	.4436	.4320	.4084	.3823	.3750
	2	.0003	.0071	.0270	.0574	.0960	.1406	.1890	.2222	.2389	.2880	.3341	.3674	.3750
	3	.0000	.0001	.0010	.0034	.0080	.0156	.0270	.0370	.0429	.0640	.0911	.1176	.1250
4	0	.9606	.8145	.6561	.5220	.4096	.3164	.2401	.1975	.1785	.1296	.0915	.0677	.0625
	1	.0388	.1715	.2916	.3685	.4096	.4219	.4116	.3951	.3845	.3456	.2995	.2600	.2500
	2	.0006	.0135	.0486	.0975	.1536	.2109	.2646	.2963	.3105	.3456	.3675	.3747	.3750
	3	.0000	.0005	.0036	.0115	.0256	.0469	.0756	.0988	.1115	.1536	.2005	.2400	.2500
	4	.0000	.0000	.0001	.0005	.0016	.0039	.0081	.0123	.0150	.0256	.0410	.0576	.0625
5	0	.9510	.7738	.5905	.4437	.3277	.2373	.1681	.1317	.1160	.0778	.0503	.0345	.0312
	1	.0480	.2036	.3280	.3915	.4096	.3955	.3602	.3292	.3124	.2592	.2059	.1657	.1562
	2	.0010	.0214	.0729	.1582	.2043	.2637	.3087	.3292	.3364	.3456	.3369	.3185	.3125
	3	.0000	.0011	.0081	.0244	.0512	.0879	.1323	.1646	.1811	.2304	.2757	.3060	.3125
	4	.0000	.0000	.0004	.0022	.0064	.0145	.0284	.0412	.0488	.0768	.1128	.1470	.1562
	5	.0000	.0000	.0000	.0001	.0003	.0010	.0024	.0041	.0053	.0102	.0185	.0283	.0312
6	0	.9415	.7351	.5314	.3771	.2621	.1780	.1176	.0872	.0754	.0467	.0277	.0176	.0156
	1	.0571	.2321	.3543	.3993	.3932	.3560	.3025	.2634	.2437	.1866	.1359	.1014	.0938
	2	.0014	.0305	.0984	.1762	.2458	.2966	.3241	.3292	.3280	.3110	.2780	.2437	.2344
	3	.0000	.0021	.0146	.0415	.0819	.1312	.1852	.2195	.2355	.2765	.3032	.3121	.3125
	4	.0000	.0001	.0012	.0055	.0154	.0330	.0592	.0823	.0951	.1382	.1861	.2249	.2344
	5	.0000	.0000	.0001	.0004	.0015	.0044	.0102	.0165	.0205	.0369	.0609	.0864	.0938
	6	.0000	.0000	.0000	.0000	.0002	.0002	.0007	.0014	.0018	.0041	.0083	.0139	.0156
7	0	.9321	.6983	.4783	.3206	.2097	.1335	.0824	.0585	.0490	.0280	.0152	.0090	.0078
	1	.0659	.2573	.3720	.3960	.3670	.3115	.2471	.2048	.1848	.1306	.0872	.0603	.0547
	2	.0020	.0406	.1240	.2097	.2753	.3115	.3177	.3073	.2985	.2613	.2140	.1740	.1641
	3	.0000	.0036	.0230	.0617	.1147	.1730	.2269	.2561	.2679	.2903	.1918	.2786	.2734
	4	.0000	.0002	.0026	.0109	.0287	.0577	.0972	.1280	.1442	.1935	.2388	.2676	.2734
	5	.0000	.0000	.0002	.0012	.0043	.0115	.0250	.0384	.0466	.0774	.1172	.1543	.1641
	6	.0000	.0000	.0000	.0001	.0004	.0013	.0036	.0064	.0084	.0172	.0320	.0494	.0547
	7	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0005	.0006	.0016	.0037	.0068	.0078
8	0	.9227	.6634	.4305	.2725	.1678	.1001	.0576	.0390	.0319	.0168	.0084	.0046	.0039
	1	.0746	.2793	.3826	.3847	.3355	.2670	.1977	.1373	.1373	.0896	.0548	.0352	.0312
	2	.0001	.0515	.1488	.2376	.2936	.3115	.2965	.2587	.2587	.2090	.1569	.1183	.1094
	3	.0000	.0054	.0351	.0839	.1468	.2076	.2541	.2786	.2786	.2787	.2568	.2273	.2188
	4	.0000	.0004	.0046	.0185	.0459	.0865	.1361	.1875	.1875	.2322	.2627	.2730	.2734
	5	.0000	.0000	.0004	.0026	.0092	.0251	.0467	.0808	.0808	.1239	.1719	.2098	.2188
	6	.0000	.0000	.0000	.0002	.0011	.0038	.0100	.0217	.0217	.0413	.0703	.1008	.1094
	7	.0000	.0000	.0000	.0000	.0001	.0004	.0012	.0033	.0033	.0079	.0164	.0277	.0312
	8	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0002	.0007	.0017	.0033	.0039
9	0	.9135	.6302	.3874	.2316	.1342	.0751	.0404	.0260	.0207	.0101	.0046	.0023	.0020
	1	.0830	.2985	.3874	.3679	.3020	.2253	.1556	.1171	.1004	.0605	.0339	.0202	.0176
	2	.0034	.0629	.1722	.2597	.3020	.3003	.2668	.2341	.2162	.1612	.1110	.0776	.0705
	3	.0001	.0077	.0446	.1069	.1762	.2336	.2668	.2731	.2716	.2508	.2119	.1739	.1641
	4	.0000	.0006	.0074	.0283	.0661	.1168	.1715	.2048	.2194	.2508	.2600	.2506	.2461
	5	.0000	.0000	.0008	.0050	.0165	.0389	.0735	.1024	.1181	.1672	.2128	.2408	.2461
	6	.0000	.0000	.0001	.0006	.0028	.0087	.0210	.0341	.0424	.0743	.1160	.1542	.1641
	7	.0000	.0000	.0000	.0000	.0003	.0012	.0039	.0073	.0098	.0212	.0407	.0636	.0703
	8	.0000	.0000	.0000	.0000	.0000	.0001	.0004	.0009	.0013	.0035	.0083	.0153	.0176
	9	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0003	.0008	.0016	.0020
10	0	.9044	.5987	.3487	.1969	.1074	.0563	.0282	.0173	.0135	.0060	.0025	.0012	.0010
	1	.0914	.3151	.3874	.3474	.2684	.1877	.1211	.0867	.0725	.0403	.0207	.0114	.0098
	2	.0042	.0746	.1937	.2759	.3020	.2816	.2335	.1951	.1757	.1209	.0763	.0495	.0439
	3	.0001	.0105	.0574	.1298	.2013	.2503	.2668	.2601	.2522	.2150	.1665	.1267	.1172
	4	.0000	.0010	.0112	.0401	.0881	.1460	.2001	.2276	.2377	.2508	.2384	.2130	.2051
	5	.0000	.0001	.0015	.0085	.0264	.0584	.1029	.1366	.1536	.2007	.2340	.2456	.2461
	6	.0000	.0000	.0001	.0012	.0055	.0162	.0368	.0569	.0689	.1115	.1596	.1966	.2051
	7	.0000	.0000	.0000	.0001	.0008	.0031	.0090	.0163	.0212	.0425	.0746	.1080	.1172
	8	.0000	.0000	.0000	.0000	.0001	.0004	.0014	.0030	.0043	.0106	.0229	.0389	.0439
	9	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0005	.0016	.0042	.0083	.0098
	10	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0008	.0010

TABLA 2

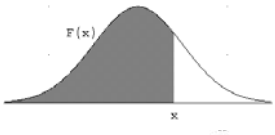
Función de masa de la v.a. de $Poisson(\theta)$
$P\{X = x\} = e^{-\theta}\theta^x/x!$ $\quad x = 0, 1, ...$ $\theta = 1, 2, 3, ..., 100$

$\theta \backslash x$	0	1	2	3	4	5	6	7	8	9	10	11	12
0.1	.9048	.0905	.0045	.0002	.0000								
0.2	.8187	.1637	.0164	.0011	.0001	.0000							
0.3	.7408	.2222	.0333	.0033	.0002	.0000							
0.4	.6703	.2681	.0536	.0072	.0007	.0001	.0000						
0.5	.6065	.3033	.0758	.0126	.0016	.0002	.0000						
0.6	.5468	.3293	.0968	.0199	.0030	.0004	.0000						
0.7	.4960	.3476	.1217	.0867	.0050	.0007	.0001	.0000					
0.8	.4493	.3595	.1438	.0998	.0077	.0012	.0002	.0000					
0.9	.4066	.3659	.1647	.1128	.0111	.0020	.0003	.0000					
1.0	.3679	.3679	.1839	.1255	.0153	.0031	.0005	.0001	.0000				
1.1	.3329	.3662	.2014	.0736	.0203	.0045	.0008	.0001	.0000				
1.2	.3012	.3614	.2169	.0867	.0260	.0062	.0012	.0002	.0000				
1.3	.2726	.3543	.2303	.0998	.0324	.0084	.0018	.0003	.0001	.0000			
1.4	.2466	.3452	.2417	.1128	.0395	.0111	.0026	.0005	.0001	.0000			
1.5	.2231	.3347	.2510	.1255	.0471	.0141	.0035	.0008	.0001	.0000			
1.6	.2019	.3230	.2584	.1378	.0551	.176	.0047	.0011	.0002	.0000			
1.7	.1827	.3106	.2540	.1496	.0636	.0216	.0061	.0015	.0003	.0001	.0000		
1.8	.1653	.2975	.2678	.1607	.0723	.0260	.0078	.0020	.0005	.0001	.0000		
1.9	.1490	.2842	.2700	.1710	.0812	.0309	.0098	.0027	.0006	.0001	.0000		
2.0	.1353	.2707	.2707	.1804	.0902	.0361	.0120	.0034	.0009	.0002	.0000		
2.2	.1108	.2438	.2781	.1966	.1082	.0476	.0174	.0055	.0015	.0004	.0001	.0000	
2.3	.0907	.2177	.2613	.2090	.1254	.0602	.0241	.0083	.0025	.0007	.0002	.0000	
2.6	.0743	.1931	.2510	.2176	.1414	.0735	.0319	.0118	.0038	.0011	.0003	.0001	.0000
2.8	.0608	.1703	.2384	.2225	.1557	.0872	.0407	.0163	.0057	.0018	.0005	.0001	.0000
3.0	.0498	.1494	.2240	.2240	.1680	.1008	.0504	.0216	.0081	.0027	.0008	.0002	.0001
3.2	.0408	.1304	.2087	.2226	.1781	.1140	.0608	.0278	.0111	.0040	.0013	.0004	.0001
3.4	.0334	.1135	.1929	.2186	.1868	.1264	.0716	.0348	.0148	.0055	.0019	.0006	.0002
3.6	.0273	.0984	.1771	.2126	.1812	.1277	.0826	.0425	.0191	.0076	.0028	.0009	.0003
3.8	.0224	.850	.1515	.2046	.1944	.1477	.0836	.0608	.0241	.0102	.0039	.0013	.0004
4.0	.0183	.733	.1465	.1954	.1954	.1563	.1042	.0695	.0298	.0132	.0063	.0018	.0006
5.0	.0067	.0337	.0842	.1404	.1755	.1755	.1462	.1044	.0653	.0363	.0181	.0082	.0034
6.0	.0025	.0149	.0446	.0892	.1339	.1606	.1606	.1377	.1033	.0688	.0413	.0225	.0113
7.0	.0009	.0064	.0223	.0521	.0912	.1277	.1490	.1490	.1304	.1014	.0710	.0452	.0264
8.0	.0003	.0077	.0107	.0286	.0673	.0916	.1221	.1395	.1396	.1241	.0993	.0722	.0481
9.0	.0001	.0011	.0050	.0150	.0337	.0607	.0911	.1171	.1318	.1318	.1186	.0970	.0728
10.0	.0000	.0005	.0023	.0076	.0189	.0378	.0631	.0901	.1126	.1251	.1251	.1137	.0948

$\theta \backslash x$	13	14	15	16	17	18	19	20	21	22	23	24	25
5.0	.0013	.0005	.0002	.0000	.0000								
6.0	.0052	.0022	.0009	.0003	.0001	.0000							
7.0	.0142	.0071	.0033	.0014	.0006	.0002	.0001	.0000					
8.0	.0296	.0169	.0090	.0045	.0021	.0009	.0004	.0002	.0001	.0000			
9.0	.0504	.0324	.0193	.0109	.0058	.0029	.0014	.0006	.0003	.0001	.0000	.0000	
10.0	.0729	.0521	.0347	.0217	.0128	.0071	.0037	.0019	.0009	.0004	.0002	.0001	.0000

TABLA 3

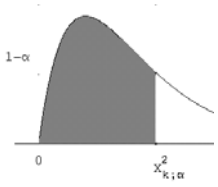
Función de distribución de la v.a. $Normal(0,1)$.
La tabla proporciona los valores de $F(x)$ tales que $F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$



x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70914	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84850	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92786	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	097725	0.97778	0.97831	097882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99085	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99586	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99859	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99897	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99026	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99957	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997

TABLA 4

Distribución de la v.a. χ^2_k.
La tabla proporciona los valores de $\chi^2_{k;\alpha}$ tales que $P\{\chi^2_k \leq \chi^2_{k;\alpha}\} = 1 - \alpha$.



1-α	0'995	0'99	0'975	0'95	0'9	0'75	0'5	0'25	0'1	0'05	0'025	0'01	0'005
k													
1	7.88	6.63	5.02	3.84	2.71	1.32	0.455	0.102	0.0158	0.0039	0.0010	0.0002	0.0000
2	10.6	9.21	7.38	5.99	4.61	2.77	1.39	0.575	0.211	0.103	0.0506	0.0201	0.0100
3	12.8	11.3	9.35	7.81	6.25	4.11	2.37	1.21	0.584	0.352	0.216	0.115	0.072
4	14.9	13.3	11.1	9.49	7.78	5.39	3.36	1.92	1.06	0.711	0.484	0.297	0.207
5	16.7	15.1	12.8	11.1	9.24	6.63	4.35	2.67	1.61	1.15	0.831	0.554	0.412
6	18.5	16.8	14.4	12.6	10.6	7.84	5.35	3.45	2.20	1.64	1.24	0.872	0.676
7	20.3	18.5	16.0	14.1	12.0	9.04	6.35	4.25	2.83	2.17	1.69	1.24	0.989
8	22.0	20.1	17.5	15.5	13.4	10.2	7.34	5.07	3.49	2.73	2.18	1.65	1.34
9	23.6	21.7	19.0	16.9	14.7	11.4	8.34	5.90	4.17	3.33	2.70	2.09	1.73
10	25.2	23.2	20.5	18.3	16.0	12.5	9.34	6.74	4.87	3.94	3.25	2.56	2.16
11	26.8	24.7	21.9	19.7	17.3	13.7	1.03	7.58	5.58	4.57	3.82	3.05	2.60
12	28.3	26.2	23.3	21.0	18.5	14.8	11.3	8.44	6.30	5.23	4.40	3.57	3.07
13	29.8	27.7	24.7	22.4	19.8	16.0	12.13	9.30	7.04	5.89	5.01	4.11	3.57
14	31.3	29.1	26.1	23.7	21.1	17.1	13.3	10.2	7.79	6.57	5.63	4.66	4.07
15	32.8	30.6	27.5	25.0	22.3	18.2	14.3	11.0	8.55	7.26	6.26	5.23	4.60
16	34.3	32.0	28.8	26.3	23.5	19.4	15.3	11.9	9.31	9.86	6.91	5.81	5.14
17	35.7	33.4	30.2	27.6	24.8	20.5	16.3	12.8	10.1	8.67	7.56	6.41	5.70
18	37.2	34.8	31.5	28.9	26.0	21.6	17.3	13.7	10.9	9.39	8.23	7.01	6.26
19	38.6	36.2	32.9	30.1	27.2	22.7	18.3	14.6	11.7	10.1	8.91	7.63	6.84
20	40.0	37.6	34.2	31.4	28.4	23.8	19.3	15.5	12.4	10.9	9.59	8.26	7.43
21	41.4	38.9	35.5	32.7	29.6	24.9	20.3	16.3	13.2	11.6	10.3	8.90	8.03
22	42.8	40.3	36.8	33.9	30.8	26.0	21.3	17.2	14.0	12.3	11.0	9.54	8.64
23	44.2	41.6	38.1	35.2	32.0	27.1	22.3	18.1	14.8	13.1	11.7	10.2	9.26
24	45.6	43.0	39.4	36.4	33.2	28.2	23.3	19.0	15.7	13.8	12.4	10.9	9.89
25	46.9	44.3	40.6	37.7	34.4	29.3	24.3	19.9	16.5	14.6	13.1	11.5	10.5
26	48.3	45.6	41.9	38.9	35.6	30.4	25.3	20.8	17.3	15.4	13.8	12.2	11.2
27	49.6	47.0	43.2	40.1	36.7	31.5	26.3	21.7	18.1	16.2	14.6	12.9	11.8
28	51.0	48.3	44.5	41.3	37.9	32.6	27.3	22.7	18.9	16.9	15.3	13.6	12.5
29	52.3	49.6	45.7	42.6	39.1	33.7	28.3	33.6	19,8	17.7	16.0	14.3	13.1
30	53.7	50.9	47.0	43.8	40.3	34.8	29.3	24.5	20.6	18.5	16.8	15.0	13.8
40	66.8	63.7	59.3	55.8	51.8	45.6	39.3	33.7	29.1	26.5	24.2	22.2	20.7
50	79.5	76.2	71.4	67.5	63.2	56.3	49.3	42.9	37.7	34.8	32.4	29.7	28.0
60	92.0	88.4	83.3	79.1	74.4	67.0	59.3	52.3	46.5	43.2	40.5	37.5	35.5
70	104.2	100.4	95.0	90.5	85.5	77.6	69.3	61.7	55.3	51.7	48.8	45.4	43.3
80	116.3	112.3	106.6	101.9	96.6	88.1	79.3	71.1	64.3	60.4	57.2	53.5	51.2
90	128.3	124.1	118.1	113.1	107.6	98.6	89.3	80.6	73.3	69.1	65.6	61.8	59.2
100	140.2	135.8	129.6	124.3	118.5	109.1	99.3	90.1	82.4	77.9	74.2	70.1	67.3

Para $k > 100$ utilícese que $\sqrt{2\chi^2_k} - \sqrt{2k-1} \simeq N(0,1)$.

TABLA 5

Distribución de la v.a. T_n *de Student*
La tabla proporciona los valores de t_α tales que $P\{T_n \leq t_\alpha\} = 1 - \alpha$

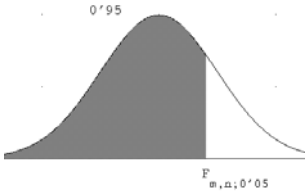


$1-\alpha$	0.750	0.900	0.950	0.975	0.990	0.995	0.999	0.9995
n								
1	1.000	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	0.816	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	0.765	1.638	2.363	3.182	4.541	5.841	10.213	12.924
4	0.741	1.633	2.132	2.776	3.747	4.604	7.173	8.610
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893	8.669
6	0.718	1.440	1.943	2.447	3.143	3.707	5.208	5.969
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.700	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.697	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.695	1.366	1.782	2.179	2.681	3.055	3.930	4.318
13	0.694	1.360	1.771	2.160	2.650	3.012	3.852	4.221
14	0.692	1.345	1.761	2.145	2.621	2.977	3.787	4.140
15	0.691	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.690	1.337	1.746	2.120	2.583	2.921	3.688	4.015
17	0.689	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.688	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.688	1.328	1.729	2.093	2.539	2.861	3.679	3.883
20	0.687	1.321	1.725	2.088	2.528	2.845	3.552	3.850
21	0.686	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.685	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.685	1.319	1.714	2.069	2.500	2.807	3.485	3.767
24	0.685	1.318	1.711	2.064	2.492	2.797	3.487	3.745
25	0.684	1.316	1.798	2.060	2.485	2.787	3.450	3.725
26	0.684	1.315	1.706	2.068	2.479	2.779	3.435	3.707
27	0.684	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.683	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.683	1.311	1.699	2.045	2.462	2.756	3.395	3.659
30	0.683	1.310	1.697	2.042	2.467	2.750	3.385	3.646
40	0.681	1.383	1.684	2.021	2.423	2.704	3.307	3.551
60	0.679	1.298	1.671	2.000	2.390	2.660	3.232	3.480
120	0.677	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	0.674	1.282	1.645	1.960	2.328	2.578	3.090	3.291

Para valores grandes de n utilícese que $T_n\sqrt{\frac{n-2}{n}} \simeq N(0,1)$.

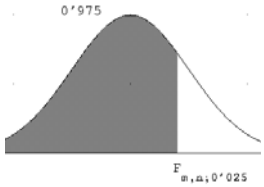
TABLA 6

Distribución de la v.a. $F_{m,n}$ *de Snedecor*.
La tabla proporciona los valores de $F_{m,n;0'05}$ tales que $P\{F_{m,n} \leq F_{m,n;0'05}\} = 0'95$



m	1	2	3	4	5	6	7	8	9	10	15	20	30	40	60	120	∞
n																	
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	245.5	241.9	245.9	248.0	250.1	151.1	252.2	25.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.49	19.43	19.45	19.46	19.47	19.48	19.40	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.70	8.66	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.86	5.80	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.62	4.56	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.94	3.87	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.51	3.44	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.7	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.22	3.15	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.02	2.94	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.85	2.77	2.70	2.66	.62	2.58	2.54
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.40	2.33	2.25	2.20	2.16	2.11	2.07
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.20	2.12	2.04	1.99	1.95	2.90	1.84
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.01	1.93	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	1.92	1.84	1.7	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.84	1.75	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.75	1.66	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.67	1.57	1.46	1.39	1.32	1.22	1.00

TABLA 6 (continuación)



Distribución de la v.a. $F_{m,n}$ *de Snedecor.*
La tabla proporciona los valores de $F_{m,n;0'05}$ tales que $P\{F_{m,n} \leq F_{m,n;0'025}\} = 0'975$

m	1	2	3	4	5	6	7	8	9	10	15	20	30	40	60	120	∞
n																	
1	647.8	799.5	864.2	899.6	912.8	937.1	948.2	956.7	963.3	968.6	984.9	993.1	1001	1006	1010	1014	1018
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.43	39.45	39.46	39.47	39.48	39.49	39.50
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.25	14.17	14.08	14.04	13.99	13.95	13.90
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.66	8.56	8.46	8.41	8.36	8.31	8.26
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.43	6.33	6.26	6.18	6.12	6.07	6.02
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.27	5.17	5.07	5.01	4.96	4.90	4.85
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.57	4.47	4.36	4.31	4.25	4.20	4.14
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.10	4.00	3.89	3.84	3.78	3.73	3.67
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.77	3.67	3.56	3.51	3.45	3.39	3.33
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3,85	3.78	3.72	3.52	3.42	3.31	3.26	3.20	3.14	3-08
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.86	2.76	2.64	2.59	2.52	2.46	2.40
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.57	2.46	2.35	2.29	2.22	2.16	2.09
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.31	2.20	2.07	2.01	1.94	2.87	1.79
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.18	2.07	1.94	1.88	1.80	1.72	1.64
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.06	1.94	1.82	1.74	1.67	1.58	1.48
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	1.94	1.82	1.69	1.61	1.53	1.43	1.31
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.83	1.71	1.57	1.48	1.39	1.27	1.00

TABLA 7

Distribución del *estadístico de Kolmogorov-Smirnov de una muestra.* Esta tabla contiene los valores de $d^+_{n\alpha}$ y $d_{n,\alpha}$ tales que $P\{D^+_n > d^+_{n,\alpha}\} \leq \alpha$ y $P\{D_n > d_{n,\alpha}\} \leq \alpha$.

Test unilateral $\alpha =$	.10	.05	.025	.01	.005	$\alpha =$	.10	.05	.025	.01	.005
Test bilateral $\alpha =$	.20	.10	.05	.02	.01	$\alpha =$	.20	.10	.05	.02	.01
n =1	.900	.950	.975	.990	.995	n=21	.226	.259	.287	.321	.344
2	.681	.776	.842	.900	.929	22	.221	.253	.281	.314	.337
3	.565	.636	.708	.785	.829	23	.216	.247	.275	.307	.330
4	.493	.565	.624	.689	.734	24	.212	.242	.269	.301	.323
5	.447	.509	.563	.627	.669	25	.208	.238	.264	.295	.317
6	.410	.468	.519	.577	.617	26	.204	.233	.259	.290	.311
7	.381	.436	.483	.538	.576	27	.200	.229	.254	.284	.305
8	.358	.410	.454	.507	.542	28	.197	.225	.250	.279	.300
9	.339	.387	.430	.480	.513	29	.193	.221	.246	.275	.295
10	.323	.369	.409	.457	.489	30	.190	.218	.242	.270	.290
11	.308	.352	.391	.437	.468	31	.187	.214	.238	.266	.285
12	.296	.338	.375	.419	.449	32	.184	.211	.234	.262	.281
13	.285	.325	.361	.404	.432	33	.182	.208	.231	.258	.277
14	.275	.314	.349	.390	.418	34	.179	.205	.227	.254	.273
15	.266	.304	.338	.377	.404	35	.177	.202	.224	.251	.269
16	.258	.295	.327	.366	.392	36	.174	.199	.221	.247	.265
17	.250	.286	.318	.355	.381	37	.172	.196	.218	.244	.262
18	.244	.279	.309	.346	.371	38	.170	.194	.215	.241	.258
19	.237	.271	.301	.337	.361	39	.168	.191	.213	.238	.255
20	.232	.265	.294	.329	.352	40	.165	.189	.210	.235	.252
	Aproximación para $n > 40$						$\frac{1{,}07}{\sqrt{n}}$	$\frac{1{,}22}{\sqrt{n}}$	$\frac{1{,}36}{\sqrt{n}}$	$\frac{1{,}52}{\sqrt{n}}$	$\frac{1{,}63}{\sqrt{n}}$

TABLA 8

Distribución del *estadístico de Kolmogorov-Smirnov* para el ajuste a una v.a. Normal con los parámetros estimados a partir de la muestra.
La tabla contiene los valores de $d^+_{n\,\alpha}$ tales que $P\{D^+_n > d^+_{n\,\alpha}\} \leq \alpha$.

n	α	0'20	0'15	0'10	0'05	0'01	0'001
4		.303	.321	.346	.376	.413	-433
5		.289	.303	.319	.343	.397	.439
6		.269	.281	.297	.323	.371	.424
7		.252	.264	.280	.304	.351	.402
8		.239	.250	.265	.288	.333	.384
9		.227	.238	.252	.274	.317	.365
10		.217	.228	.241	.262	.304	.352
11		.208	.218	.231	.251	.291	.338
12		.200	.210	.222	.242	.281	.325
13		.193	.202	.215	.234	.271	.314
14		.187	.196	.208	.226	.262	.305
15		.181	.190	.201	.219	.254	.296
16		.176	.184	.195	.213	.247	.287
17		.171	.179	.190	.207	.240	.279
18		.167	.175	.185	.202	.234	.273
19		.163	.170	.181	.197	.228	.266
20		.159	.166	.176	.192	.223	.260
25		.143	.150	.159	.173	.201	.236
30		.131	.138	.146	.159	.185	.217
40		.115	.120	.128	.139	.162	.189
100		.074	.077	.082	.089	.104	.122
400		.037	.039	.041	.045	.052	.061
900		.025	.026	.028	.030	.035	.042

TABLA 9

Distribución del *estadístico de los rangos signados de Wilcoxon.*
Esta tabla contiene los valores de t_α tales que $P\{T^+ > t_\alpha\} \leq \alpha$.
Los valores correspondientes a $t_{1-\alpha}$ se obtienen de
$t_{1-\alpha} = m(n+1)/2 - t_\alpha$.

n	α .01	.025	.05	.10
3	6	6	6	6
4	10	10	10	9
5	15	15	14	12
6	21	20	18	17
7	27	25	24	22
8	34	32	30	27
9	41	39	36	34
10	49	46	44	40
11	58	55	52	48
12	67	64	60	56
13	78	73	69	64
14	89	84	79	73
15	100	94	89	83
16	112	106	100	93
17	125	118	111	104
18	138	130	123	115
19	152	143	136	127
20	166	157	149	140

La distribución asintótica del *estadístico de Wilcoxon* viene dada por

$$\frac{T_n^+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \simeq N(0,1).$$

TABLA 10

Distribución del *estadístico de Kolmogorov-Smirnov para dos muestras de distinto tamaño.*

Esta tabla contiene los valores de $d^+_{mn,\alpha}$ y $d_{mn,\alpha}$ tales que $P\{D^+_{m,n} > d^+_{m,n,\alpha}\} \leq \alpha$ y $P\{D_{m,n} > d_{m,n\,\alpha}\} \leq \alpha$ con $N_1 = \text{mín}\{m, n\}$ y $N_2 = \text{máx}\{m, n\}$

Test unilateral α = / Test bilateral α =		0.10 / .20	0.05 / .10	0.025 / .05	0.01 / .02	0.005 / .01
N_1=1	N_2=9	17/18				
	10	9/10				
N_1=2	N_2=3	5/6				
	4	3/4				
	5	4/5	4/5			
	6	5/6	5/6			
	7	5/7	6/7			
	8	3/4	7/8	7/8		
	9	7/9	8/9	8/9		
	10	7/10	4/5	9/10		
N_1=3	N_2=4	3/4	3/4			
	5	2/3	4/5	4/5		
	6	2/3	2/3	5/6		
	7	2/3	5/7	6/7	6/7	
	8	5/8	3/4	3/4	7/8	
	9	2/3	2/3	7/9	8/9	8/9
	10	3/5	7/10	4/5	9/10	8/10
	12	7/12	2/3	3/4	5/6	11/12
N_1=4	N_2=5	3/5	3/4	4/5	4/5	
	6	7/12	2/3	3/4	5/6	5/6
	7	17/28	5/7	3/4	6/7	6/7
	8	5/8	5/8	3/4	7/8	7/8
	9	5/9	2/3	3/4	7/9	8/9
	10	11/20	13/20	7/10	4/5	4/5
	12	7/12	2/3	2/3	3/4	5/6
	16	9/16	5/8	11/16	3/4	13/16
N_1=5	N_2=6	3/5	2/3	2/3	5/6	5/6
	7	4/7	23/35	5/7	29/35	6/7
	8	11/20	5/8	27/40	4/5	4/5
	9	5/9	3/5	31/45	7/9	4/5
	10	1/2	3/5	7/10	7/10	4/5
	15	8/15	3/5	2/3	11/15	11/15
	20	1/2	11/20	3/5	7/10	3/4
N_1=6	N_2=7	23/42	4/7	29/42	5/7	5/6
	8	1/2	7/12	2/3	3/4	3/4
	9	1/2	5/9	2/3	13/18	7/9
	10	1/2	17/30	19/30	7/10	11/15
	12	1/2	7/12	7/12	2/3	3/4
	18	4/9	5/9	11/18	2/3	13/18
	24	11/24	1/2	7/12	5/8	2/3

TABLA 10 (continuación)

Test unilateral $\alpha =$ / Test bilateral $\alpha =$		0.10 / 0.20	0.05 / 0.10	0.025 / 0.05	0.01 / 0.02	0.005 / 0.01
N_1=7	N_2=8	27/56	33/56	5/8	41/56	3/4
	9	31/63	5/9	40/63	5/7	47/63
	10	33/70	39/70	43/70	7/10	5/7
	14	3/7	1/2	4/7	9/14	5/7
	28	3/7	13/28	15/28	17/28	9/14
N_1=8	N_2=9	4/9	13/24	5/8	2/3	3/4
	10	19/40	21/40	23/40	27/40	7/10
	12	11/24	1/2	7/12	5/8	7/3
	16	7/16	1/2	9/16	5/8	5/8
	32	13/32	7/16	1/2	9/16	19/32
N_1=9	N_2=10	7/15	1/2	26/45	2/3	31/45
	12	4/9	1/2	5/9	11/28	2/3
	15	19/45	22/45	8/15	3/5	29/45
	18	7/18	4/9	1/2	5/9	11/18
	36	13/36	5/12	17/36	19/36	5/9
N_1=10	N_2=15	2/5	7/15	1/2	17/30	19/30
	20	2/5	9/20	1/2	11/20	3/5
	40	7/20	2/5	9/20	1/2	
N_1=12	N_2=15	23/60	9/20	1/2	11/20	7/12
	16	3/8	7/16	23/48	13/24	7/12
	18	13/36	5/12	17/36	19/36	5/9
	20	11/30	5/12	7/15	31/60	17/30
N_1=15	N_2=20	7/20	2/5	13/30	29/60	31/60
N_1=16	N_2=20	27/80	31/80	17/40	19/40	41/80
Aproximación para m y n grandes		$10\sqrt{\frac{m+n}{mn}}$	$1.22\sqrt{\frac{m+n}{mn}}$	$1.36\sqrt{\frac{m+n}{mn}}$	$1.52\sqrt{\frac{m+n}{mn}}$	$1.63\sqrt{\frac{m+n}{mn}}$

TABLA 11

Distribución del estadístico de *Kolmogorov-Smirnov para dos muestras de igual tamaño.*
Esta tabla contiene los valores de $d^{+}_{nn;\alpha}$ y $d_{nn;\alpha}$ tales que

$$P\{D^{+}_{nn} > d_{nn;\alpha}\} \leq \alpha \quad \text{y} \quad P\{D_{nn} > d_{nn;\alpha}\} \leq \alpha.$$

Test unilateral $\alpha =$	.10	.05	.025	.01	.005	$\alpha =$	.10	.05	.025	.01	.005
Test bilateral $\alpha =$	.20	.10	.05	.02	.01	$\alpha =$	.20	.10	.05	.02	.01
n=3	2/3	2/3				n=20	6/20	7/20	8/20	9/20	10/20
4	3/4	3/4	3/4			21	6/21	7/21	8/21	9/21	10/21
5	3/5	3/5	4/5	4/5	4/5	22	7/22	8/22	8/22	10/22	10/22
6	3/6	4/6	4/6	5/6	5/6	23	7/23	8/23	9/23	10/23	10/23
7	4/7	4/7	5/7	5/7	5/7	24	7/24	8/24	9/24	10/24	11/24
8	4/8	4/8	5/8	5/8	6/8	25	7/25	8/25	9/25	10/25	11/25
9	4/9	5/9	5/9	6/9	6/9	26	7/26	8/26	9/26	10/26	11/26
10	4/10	5/10	6/10	6/10	7/10	27	7/27	8/27	9/27	11/27	11/27
11	5/11	5/11	7/11	7/11	7/11	28	8/28	9/28	10/28	11/28	12/28
12	5/12	5/12	7/12	7/12	7/12	29	8/29	9/29	10/29	11/29	12/29
13	5/13	6/13	7/13	7/13	8/13	30	8/30	9/30	10/30	11/30	12/30
14	5/14	6/14	7/14	7/14	8/14	31	8/31	9/31	10/31	11/31	12/31
15	5/15	6/15	7/15	8/15	8/15	32	8/32	9/32	10/32	12/32	12/32
16	6/16	6/16	7/16	8/16	9/16	34	8/34	10/34	11/34	12/34	12/34
17	6/17	7/17	7/17	8/17	9/17	36	9/36	10/36	11/36	12/36	12/36
18	6/18	7/18	8/18	9/18	9/18	38	9/38	10/38	11/38	13/38	14/38
19	6/19	7/19	8/19	9/19	9/19	40	9/40	10/40	12/40	13/40	14/40
Aproximación para $n > 40$:							$\frac{1,52}{\sqrt{n}}$	$\frac{1,73}{\sqrt{n}}$	$\frac{1,92}{\sqrt{n}}$	$\frac{2,15}{\sqrt{n}}$	$\frac{2,30}{\sqrt{n}}$

TABLA 12

Distribución del *estadístico de Mann-Whitney-Wilcoxon.*
Esta tabla contiene los valores de U_α tales que $P\{U > u_\alpha\} \leq \alpha$.
Por simetría $u_{1-\alpha} = mn - u_\alpha$.

m	α	n	2	3	4	5	6	7	8	9	10
2	.01		4	6	8	10	12	14	16	18	20
	.025		4	6	8	10	12	14	15	17	19
	.05		4	6	8	9	11	13	14	16	18
	.10		4	5	7	8	10	12	13	15	16
3	.01			9	12	15	18	20	20	25	28
	.025			9	12	14	16	19	21	24	26
	.05			8	11	13	15	18	20	22	25
	.10			7	10	12	14	16	18	21	23
4	.01				16	19	22	26	29	32	36
	.025				15	18	21	24	27	31	34
	.05				14	17	20	23	26	29	32
	.10				12	15	18	21	24	26	29
5	.01					23	27	31	35	39	43
	.025					22	26	29	33	37	41
	.05					20	24	28	31	35	38
	.10					19	22	26	29	32	36
6	.01						32	37	41	46	51
	.025						30	35	39	43	48
	.05						28	33	37	41	45
	.10						26	30	34	38	42
7	.01							42	48	53	58
	.025							40	45	50	55
	.05							37	42	47	52
	.10							35	39	44	48
8	.01								54	60	66
	.025								50	56	62
	.05								48	53	59
	.10								44	49	55
9	.01									66	73
	.025									63	69
	.05									59	65
	.10									55	61
10	.01										80
	.025										76
	.05										72
	.10										67

La distribución asintótica del *estadístico de Mann-Whitney-Wilcoxon* viene dada por

$$\left(U - \frac{mn}{2}\right) \Big/ \sqrt{\frac{mn(m+n+1)}{12}} \simeq N(0,1)$$

TABLA 13

Distribución del *estadístico de Kendall.*
Esta tabla contiene los valores de s_α para los que $P\{T > t_\alpha\} \leq \alpha$ donde $S = \binom{n}{2} T$ es el *estadístico de Kendall.* Por simetría $s_{1-\alpha} = -s_\alpha$.

n	α .100	.050	.025	.01
3	3	3	3	3
4	4	4	6	6
5	6	6	8	8
6	7	9	11	11
7	9	11	13	15
8	10	14	16	18
9	12	16	18	22
10	15	19	21	25

La distribución asintótica del *estadístico de Kendall* viene dada por

$$3\sqrt{[n(n-1)]/[2(2n+5)]}\; T \simeq N(0,1).$$

TABLA 14

Distribución del estadístico de los *rangos de Spearman.*
Esta tabla contiene los valores de r_α tales que $P\{R > r_\alpha\} \leq \alpha$.
Por simetría $r_{1-\alpha} = -r_\alpha$.

n	α .01	.025	.05	.10
3	1.000	1.000	1.000	1.000
4	1.000	1.000	.800	.800
5	.900	.900	.800	.700
6	.886	.829	.771	.600
7	.857	.750	.679	.536
8	.810	.714	.619	.500
9	.767	.667	.583	.467
10	.721	.636	.552	.442

La distribución asintótica del *estadístico de Spearman* viene dada por

$$R\sqrt{n-1} \simeq N(0,1).$$

REFERENCIAS BIBLIOGRÁFICAS

Arbuthnot, J. (1710), An argument for Divine Providence, taken from the constant regularity observed in the births of both sexes. *Philosophical Transactions of the Royal Society of London*, **27**, 186-190. Puede verse en Kendall, M. y Plackett, R.L. (1977) Studies in the History of Statistics and Probability Vol II, London: Charles Griffin, pág. 30-34.

Basu, D. (1955), On statistics independent of a complete sufficient statistic. *Sankhya*, **15**, 377-380.

Bayes, T. (1763) An essay towards solving a problem in the doctrine of chances. *Phil. Trans. Roy. Soc. Lond* **53**, 370-418 y **54**, 296-325. Reeditado en *Biometrika* 45 (1958), 293-315, con una nota biográfica por G.A. Barnard. Reproducido en S.J. Press (1989) *Bayesian Statistics: Principles, Models, and Applications*. Nueva York: Wiley. Traducido al castellano por Girón, F.J., Gómez Villegas, M.A., Martínez, M.L. y Ríos, D. en *Rev. R. Acad. Cienc. Exact. Fis. Nat. (Esp)* **95** (2001), **1-2**, 63-80.

Bennet, J.H. (1990) *Statistical Methods, Experimental Design and Scientific Inference*. Reedición de los libros de R.A. Fisher. Oxford: Oxford University Press.

Berger, J. (1985) *Statistical Decision Theory and Bayesian Analysis*. Second Edition. Nueva York: Springer-Verlag.

Bernoulli, J. (1713) *Ars Conjectandi*. Basilea: Thurnisiorum.

Bickel, P.J. y Doksum, K.A. (1977) *Mathematical Statistics: Basic Ideas and Selected Topics*. San Francisco: Holden-Day.

Birnbaum, A. (1962) On the foundations of statistical inference. *J. Amer. Statist. Assoc.* **57**, 269-306.

Blackwell, D. y Girshick, M.A. (1954) *Theory of Games and Statistical Decisions*. Nueva York: Wiley.

Blyth, C.R. (1951) On minimax statistical decision procedures and their admissibility. *Annals of Mathematical Statistics* **22**, 22-42.

Boscovich, R.J. (1770) *Voyage Astronomique et Geographique, dans L'État de L'Église*. Paris: N.M. Tilliar.

Box, G.E.P. (1953) Non-normality and tests on variances. *Biometrika* **40**, 318-335.

Box, G.E.R. y Tiao, G.C. (1873) *Bayesian Inference in Statistical Analysis.* Londres: Addison-Wesley.

Brown, L.D. (1971) Admissible estimators, recurrent diffusions, and indisoluble boundary-value problems. *Annals of Mathematical Statistics* **42**, 855-903.

Cramér, H. (1968, 1973) *Métodos Matemáticos de Estadística*. Madrid: Aguilar.

Cristóbal, J.A. (1995) *Inferencia Estadística.* Zaragoza: Prensas Universitarias de Zaragoza.

de Finetti, B. (1974) *Theory of Probability*, 1,2. Chichester: Wiley.

DeGroot, M.H. (1970) *Optimal Statistical Decisions*. N. York: McGraw-Hill.

DeGroot, M. (1988) *Probabilidad y Estadística*. México: Addison-Wesley Iberoameriaca S.A.

Devroye, L. (1997) Universal smoothing factor selection in density estimation: theory and practice. *Test* 6, 2, 223-320.

Dudewicz, E.J.Z. y Mishra, S.N. (1988) *Modern Mathematical Statistics*. Nueva York: Wiley.

Edgeworth, F.Y. (1885) Observations and statistics: an essay on the theory of errors of observation and the first principles of statistics. *Transactions of the Cambridge Philosophical Society* **14**, 138-169.

Edgeworth, F.Y. (1889) Review of Natural Inheritance by F. Galton. *Nature* **39**, 603-604.

Ferguson, T.S. (1967) *Mathematical Statistics: A Decision Theoretic Approach*. New York: Academic Press.

Fishburn, P.C. (1989) Foundations of decision analysis: along the way. *Management Science* Vol. 35, **4**, 387-405.

Fisher, R.A. (1912) On an absolute criterion for fitting frequencey curves. *Messenger Math.* **41**, 155-60. Reeditado en Stat. Sci. **12**, 39-41.

Fisher, R.A. (1915) Frecuency distribution of the values of the correlation coefficiente in samples from an indefinitely large population. *Biometrika* **10**, 507-521.

Fisher, R.A. (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinburgh* **52**, 399-433.

Fisher, R.A. (1920). A mathematical examination of the methods of determining the accuracy of an observation by the mean error, and by the mean square error. *Mon. Not. R. Astron. Soc.* **80**, 758-770.

Fisher, R.A. (1923) Statistical tests of agreement between observation and hypothesis. *Economica* **3**, 139-147.

Fisz, M. (1963) *Probability Theory and Mathematical Statistics.* Nueva York: John Wiley.

Fraser, D.A.S. (1957) *Nonparametric Methods in Statistics.* Nueva York: John Wiley.

Galton, F. (1875) Statistic by intercomparison, with remarks on the law of frequency of error. *Philosophical Magazine* **49**, 33-46.

Galton, F. (1877) Typical laws of heredity. *Nature* **15**, 492-495, 512-514, 532-533.

Galton, F. (1889) *Natural Inheritance.* Londres: Macmillan.

Galton, F. (1889) Correlations and their measurement, chiefly from anthropometric data. *Proc. R. Soc. London* **45**, 135-145.

García Nogales, A. (1998) *Estadística Matemática.* Cáceres: Servicio de Publicaciones de la Universidad de Extremadura.

Gibbons, J.D. y Chakraborti, S. (1992) *Nonparametric Statistical Inference.* Nueva York: Marcel Dekker.

Giri, N.C. (1993) *Introduction to Probability and Statistics.* Nueva York: Marcel Dekker.

Girón, F.J. y Gómez Villegas, M.A. (1977) *Teoría de Juegos*. Madrid: UNED.

Girón, F.J. y Gómez Villegas, M.A. (1998): R.A. Fisher: su contribución a la Ciencia Estadística. *Historia de la Matemática en el siglo XX.* Madrid: Real Academia de Ciencias Exactas, Físicas y Naturales, 43-61.

Gnedenko, B.V. (1954) Tests of homogeneity of probability distributions in two independent samples. *Math. Nachrichten* **19**, 29-66.

Gómez, V. y Gómez Villegas, M.A. (1998) Simplifying the initialization of the Kalman filter for nonstationary state space models. *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales* **92, 4**, 377-385.

Gómez, E., Gómez Villegas, M.A. y Marín J.M. (1998) A multivariate generalization of the power exponential family of distributions. *Commun. Satist. Theory Meth.* **27, 3**, 589-600.

Gómez, E., Gómez Villegas, M.A. y Marín, J.M. (2002) Continuous elliptical and exponential power linear dynamic models. *Journal of Multivariate Analysis* **83**, 22-36.

Gómez Sánchez-Manzano, E. y Gómez Villegas, M.A. (1990) Three Methods for constructing reference prior distributions. *Revista Matemática de la Universidad Complutense de Madrid* **3, 2** y **3,** 153-162.

Gómez Villegas, M.A. (1977) Sensibilidad del criterio de decisión de Bayes. *Trabajos de Estadística y de Investigación Operativa* **28**, **2** y **3**, 63-84.

Gómez Villegas, M.A. (1979) Determinación de la política óptima de explotación de un embalse. *Trabajos de Estadística y de Investigación Operativa* **30**, **2**, 97-102.

Gómez Villegas, M.A. (1980) Théorie de la décision statistique. Sensibilité et lemme de Neyman-Pearson. *C.R. Acad. Sc. Paris* **290**, A, 1141-1143.

Gómez Villegas, M.A. (1983) Sensibilidad respecto a la función de pérdida en la teoría de la decisión. *Rev. R. Acad. Cienc. Exact. Fís. Nat.* (Esp) **77, 2,** 463-465.

Gómez Villegas, M.A. (1988) Ajuste bayesiano a una distribución mediante la familia exponencial potencial. *Estadística Española* **30**, **110**, 203-213.

Gómez Villegas, M.A. (2004) Estadísticos significativos *Historia de la Probabilidad y la Estadística II* (Por Santos, J. y García, M., eds) Madrid: Delta Publicaciones, 271-286.

Gómez Villegas, M.A. y Gómez Sánchez-Manzano, E. (1992) Bayes factor in testing precise hypotheses. *Commun. Statist. Theory Meth.* **21**, **6**, 1707-1715.

Gómez Villegas, M.A. y Maín, P. (1992) The influence of prior and likelihood tail behaviour on the posterior distribution. *Bayesian Statistics 4*. (Bernardo M. , Berger J.O. David, A.P. y Smith, A.F.M., eds) Oxford: University Press, 661-667.

Gómez Villegas, M.A., Maín, P., Navarro, H. y Sanz, L. (2004) Asymptotic relationships between posterior probabilities and p-values using the hazard rate. *Statistics & Probability Letters*. **66**, **1**, 59-66.

Gómez Villegas, M.A. y Portela, J. (2004) Implementation of a robust Bayesian method. *Journal of Statistical Computation and Simulation*. **74**, **4**, 235-248.

Gómez Villegas, M.A. y Sanz, L. (1998) Reconciling Bayesian and frequentist evidence in the point null testing problem. *Test* **4**, **1**, 207-216.

Gómez Villegas, M.A. y Sanz, L. (2000) ε-contaminated prior in testing point null hypothesis: a procedure to determine the prior probability. *Statistics & Probability Letters* **47**, 53-60.

Hájek, J. (1969) *A Course in Nonparametric Statistics*. San Francisco: Holden-Day.

Hájek, J.Z. y Šidák, Z. (1967) *Theory of Rank Tests*. Nueva York: Academic Press.

Hartigan, J.A. (1983) *Bayes Theory*. Nueva York: Springer-Verlag.

Hodges, J.L. y Lehmann, E.L. (1961) Comparison of the normal scores and Wilcoxon tests. *Proceedings of the fourth Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: California Press, **1**, 307-317.

Hotelling, H. y Palost, M. (1936) Rank correlation and test of significance involving no assumption of normality. *Ann. Math. Statist.* **7**, 28-43.

Huber, P.J. (1972) Robust statistics: a review. A*nn. Math. Statist.* **43**, 1041-1067.

James, W. y Stein, C. (1961) Estimation with quadratic loss. *Proceedings of the fourth Berkeley symposium on mathematics statistics and probability* 361-380, Berkeley (California): University of California Press.

Jeffreys, H. (1939, 1961) *Theory of Probability.* Oxford: Oxford University Press.

Kac, M. Kiefer, J. y Wolfowitz, J. (1955) On tests of normality and other tests of fitdness of fit bases on distance methods. *Ann. Math. Stat.* **26**, 189-211.

Kendall, M.G. (1938) A new measure of rank correlation. *Biometrika* **30**, 81-93.

Kolmogorov, A.N. (1933) Sulla determinazione empirica di una legge di distribuzione. *Giornale Inst. Ital. Attuari* **4**, 83-91.

Kolmogorov, A.N. (1941) Confidence limits for an unknown distribution function. *Ann. Math. Statist.* **12,** 461-463.

Laplace, P.S. (1774) Mémoire sur la probabilité des causes par las évènements. *Mémoires de l'Académie royale des sciences presentés par divers savans* **6**, 621-656.

Laplace, P.S. (1812, 1814, 1820, 1825) *Théorie Analytique des Probabilités.* Paris: Courcier.

Legendre, A.M. (1805) *Nouvelles méthodes pour la détermination des orbits des comètes.* París: Courcier. Puede verse en Smith, D.E. (1952) A Course Book in Mathematics. Nueva York: McGraw-Hill.

Lehmann, E.L. (1975) *Nonparametric: Statistical Methods Based on Ranks.* San Francisco: Holden Day.

Lehmann, E.L. (1986): *Testing Statistical Hypotheises.* Nueva York: Wiley.

Lehmann, E. y Casella, G. (1998) *Theory of Point Estimation (segunda edición).* Londres: Springer.

Lindley, D. (1969) *Introduction to Probability an Statistics From a Bayesian Viewpoint. Part 1. Probability.* Cambridge University Press.

Lindley, D. (1970) *Introduction to Probability and Statistics From a Bayesian Viewpoint. Part 2. Inference.* Cambridge University Press.

Lilliefors, H.W. (1967) On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *J. Amer. Statist. Assoc.* **62**, 399-402.

Loève, M. (1963) *Probability Theory.* Nueva York: Van Nostrand.

Mann, H.B. and Whitney, D.R. (1947) On a test of whether one of two randon variables is stochastically larger than the other. *Ann. Math. Statist.* **18**, 50-60.

Mises, R. von (1931) Über einige abschätzungen von erwartungswerten. J. Reine Angew. math. **165**, 184-193. Reeditado en 1964 *Selected Papers of Richard von Mises*, vol. 2, 135-148. Nueva York: American Mathematical Society.

Newcomb, S. (1903) *The Reminiscences of an Astronomer*. Boston: Houghton Hifflin.

Pearson, K. (1892, 1900, 1911) *The Grammar of Science*. Londres: Walter Scott.

Pearson, K. (1896) Mathematical contributions to the theory of evolution III: regresión, heredity and panmisia. *Philosophical Transactions fo the Royal Society of London* A. **187,** 253-318.

Pearson, K. (1900) On the criterion that a given system of derivations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine* **50**, 157-175. Reeditado en Pearson 1956, 339-357.

Pearson, K. (1904) Mathematical contributions to the theory of evolution XIII. On the theory of contingency and its relation to association and normal correlation. *Biometric* **2**, Reeditada en K. Pearson (1948).

Pearson, K. (1931) *Tables for statisticians and Biometricians*. Cambridge: Cambridge University Press.

Pearson, K. (1948) *Karl Pearson's Early Statistical Papers*. Editado por E.S. Pearson. Cambridge: Cambridge University Press.

Pearson, K. (1956) *Karl Pearson's Early Statistical Papers*. Cambridge: Cambridge University Press.

Pearson, K. (1978) *The History of Statistics in the 17th and 18th Centuries.* Lectures by Karl Pearson given at University College London during the academic sessions 1921-1933. Editado por E.S. Pearson. Londres. Griffin.

Pearson, E.S. y Kendall, M.G. (1970) *Studies in the History of Statistics and Probability.* Londres: Griffin.

Peña, D. (2001) *Fundamentos de Estadística*. Madrid: Alianza Editorial.

Press, S.J. (1989) *Bayesian Statistics: Principles, Models and Applications.* Nueva York: John Wiley & Sons.

Quetelet, A. (1846) *Lettres à S.A.R. le Duc Régnant de Saxe-Lobourg te Gotha, sur la Théorie des Probabilités, Appliqué aux Sciences Morales et Politiques*. Bruselas: Hayesz.

Randles, R.H. y Wolfe, D.A. (1979) *Introduction to the Theory of Nonparametric Statistics*. Nueva York: John Wiley & Sons.

Rao, C.R. (1973) *Linear Statistical Inference and its Applications*. Nueva York: Wiley.

Rudemo, H. (1982) Empirical choice of histograms and kernel density estimators. *Scand. J. Stat.* **9**, 65-78.

Rice, J.A. (1995) *Mathematical Statistics and Data Analysis*. California: Duxbury Press.

Rios, S. (1975) *Métodos Estadísticos.* Madrid: Ed. Castillo.

Savage, L.J. (1997) *Theory of Statistics*. Nueva York: Springer.

Sheynin, O.B. (1973) *Arch. History Exact Sci.* **9**, 306-324. (Contiene una revisión de los trabajos estadísticos de Boscovich).

Silverman, B.W. (1996) *Density Estimation for Statistics and Data Analysis.* Londres: Chapman & Hall.

Simpson, T. (1755) A letter to the right Honorable George Earl of Macclesfield, President of the Royal Society, on the advantage of taking the mean of a number of observations in practical astronomy. *Philosophical Transactions of the Royal Society of London* **49**, 82-93.

Smirnov, N.V. (1939) Estimate of deviation between empirical distribution functions in two independent samples. *Bulletin of Moscow University* **2**, 3-16.

Smirnov, N.V. (1948) Tables for estimating the goodness of fit of empirical distributions. *Ann. Math. Statist.* **19**, 279-81.

Smirnov, N.V. (1960) Tables of the Normal Probability Integral, Normal Density and its Normalized Derivative. *Isdatel'stvo Akad. Nauk SSSR*. Moscow.

Spearman, C. (1904) The proof and measuremente of association between two things. *Amer. J. Psychol.* **15**, 72-101.

Stigler, S.M. (1986) *The History of Statistics: The Measurement of Uncertainty before 1900*. Cambridge: Harvard University Press.

Strang, G. (1980) *Linear Algebra and Its Applications*. Nueva York: Academic Press.

Stuart, A. y Ord, J.K. (1991) *Kendall's advanced theory of Statistics.* Vol. 2. Classical Inference and Relationship. Londres. Edward Arnold.

Student (1927) Errors in routine analysis. *Biometrika* **19,** 151-64.

Sudderth, W.D. (1976) Finitely additive priors, coherence and the marginalization parados. *J. Roy. Statist. Soc.* B **42**, 339-341.

Vélez, R. y Garcia, A. (1993) *Principios de Inferencia Estadística*. Madrid: UNED.

Wald, A. (1939) Contributions to the theory of statistical estimation and testing hypothesis. *Ann. Math. Statist.* **10**, 299-326.

Wald, A. (1950) *Statistical Decision Functions.* Nueva York: Wiley.

Wilcoxon, F. (1945) Individual comparinson by ranking methods. *Biometrics* **1**, 80-83.

Wilcoson, F. (1947) Probability tables for individual comparisons by ranking methods. *Biometrics* **3**, 119-122.

Wilks, S.S. (1938) The large-sample distributions of the likelihood ratio for testing composite hypotehses. *Ann. Math. Statist.* **9**, 60-62.

Índice alfabético

acción Bayes, 292, 300, 301, 303
admisibilidad, 280, 285
amplitud de banda, 428
análisis de datos, 9
análisis de la varianza, 315, 325
aproximación
    bayesiana, 207
    clásica, 207
Arago, 141
Arbuthnot, 452

banda de confianza, 410
Bayes, 246
Bennet, 187
Berger, 140, 281, 282, 295
Bernoulli, Daniel, 246
Bernoulli, Jacob, 140, 246
Bickel, 120
Birnbaum, 79
Blackwell, 290
Blyth, 283
bondad de ajuste, 389
Borel, 140, 187
Boscovich, 43, 452
Bouvard, 141
Box, 453
Breiman, 2
Brevais, 355
Brown, 289

Cantor, 140
Casella, 137, 140
Chakraborti, 448
clase completa, 280, 281
cociente de verosimilitudes (RV), 234
coeficiente de
    correlación, 121
    correlación de Spearman, 449, 475
    correlación lineal, 331, 332, 347, 381
    correlación muestral, 78, 423
    determinación, 347, 380–383
    Kendall, 445
completitud, 67, 93, 115
completo, 157
con reemplazamiento, 4
conjugada, 143, 200
consistencia, 99, 135
contraste
    aleatorizado, 208, 233
    bayesiano, 225, 251
    de hipótesis, 209, 275
    de hipótesis unilaterales, 236
    de homogeneidad, 462
    de Kendall, 444, 447, 448, 475
    de Kolmogorov-Smirnov, 456–458, 461, 466, 473
    de Kolmogorov-Smirnov para dos muestras, 429, 430, 471
    de Kolmogorov-Smirnov para una muestra, 407, 410, 466
    de Kruskal-Wallis, 442, 444
    de la mediana, 436, 457, 472, 474
    de la razón de verosimilitudes (RV), 211, 212, 224, 249, 458–461, 464
    de localización, 413
    de los rangos signados, 407

de los rangos signados para dos muestras, 453
de los signos, 407, 414, 420, 468, 470
de Mann-Whitney-Wilcoxon, 438, 440, 444, 457, 469
de normalidad, 421
de Spearman, 450, 451
de Wilcoxon, 414, 415, 420
de Wilcoxon de los rangos signados, 469
del coeficiente de correlación de los rangos de Spearman, 449
estadístico, 208
insesgado, 230, 240, 245
invariante, 228, 229
no paramétrico de independencia, 444
óptimo, 230, 233, 234, 245
relativo a la diferencia de medias, 217
relativo a la media, 213
relativo a la media de la distribución normal, 213
relativo a la varianza, 215
relativo al cociente de varianzas, 218
uniformemente de máxima potencia (UMP), 231, 245
convergencia en ley, 17
correlación, 355
cota de Fréchet-Cramér-Rao (FCR), 120–127, 145, 146, 157, 161
Cournot, 141
covarianza, 120
Cramér, 1, 17, 394, 453
criterio
de comparación de contrastes, 230
de la pérdida final esperada, 107
del error cuadrático medio, 105, 144
cuantil
muestral, 13
poblacional, 13
cuasivarianza
muestral, 6, 8, 12, 13, 45, 118, 143
muestral ponderada, 28, 179

D'Alembert, 43
Darwin, 78, 355
De Finetti, 42, 247
DeGroot, 290, 400
desigualdad de
Bonferroni, 323, 325
Cauchy-Schwarz, 114
Devroye, 428
diagrama de
caja, 14, 15, 50, 328, 364, 366, 368, 370
rectángulos, 9
sectores, 9
tallos y hojas, 14, 44, 49
distancia de K. Pearson, 393, 394, 396, 397, 399, 400, 456, 458, 460, 463, 465
distribución
χ^2_n, 23
asintótica, 138, 140
de la distribución final, 139
de la razón de verosimilitudes (RV), 221, 223
del estimador de máxima verosimilitud (EMV), 136
Beta, 156, 200, 301
Binomial, 246, 272, 405
Binomial Negativa, 145, 156, 158
conjunta de los recubrimientos, 38
de Behrens-Fisher, 179
de Bernoulli, 156, 199, 298
de la media muestral, 21, 45
de Poisson, 191, 251, 454

de Snedecor, 30, 180, 199, 219, 320
de Snedecor no centrada, 32
de Student, 25, 27, 28, 52, 179, 198, 204, 322, 335, 376
de Student multivariante, 351
de Student no centrada, 28
empírica, 20
final, 100, 102, 103, 108, 109, 139, 143, 144, 151, 152, 154–156, 167, 191, 200, 201, 225, 276, 300
Gamma, 191, 249, 250
Gamma-Inversa, 143, 154, 353
Geométrica, 293
impropia, 202
inicial, 100, 102–104, 143–145, 152, 184, 300
libre, 401, 409
marginal, 300
predictiva, 100, 276
Triangular, 193
Uniforme, 454
distribuciones con cociente de verosimilitud monótono, 236
Dixon, 452
Doksum, 120
Dualidad, 241
Dudewicz, 179

ecuación de verosimilitud, 135, 136, 166, 182, 222
ecuaciones de verosimilitud, 131
ecuaciones normales, 330, 338
Edgeworth, 43, 356, 452
Egon Pearson, 187, 188, 245, 247
error
cuadrático medio (ECM), 105–110, 118, 144, 153, 154
de tipo I, 209, 210, 230, 233, 248, 253
de tipo II, 209, 210, 230, 253
estándar, 335
probable, 452
espacio muestral, 271
Estadística Descriptiva, 9
estadístico, 95, 142, 157
ancilario, 65, 80, 81
completo, 66, 80, 81, 91, 93, 112, 115, 157
de Kendall, 449, 475
de Kolmogorov-Smirnov, 458–460
de Kolmogorov-Smirnov bilateral, 429
de Mann-Whitney, 441
de Mann-Whitney-Wilcoxon, 438–440
de Spearman, 450–452
de Student, 357
de Wilcoxon, 416, 418, 421, 441, 469
invariante, 70, 71
minimal suficiente, 63, 68, 81, 82, 93
muestral, 4, 6
de Kendall, 446
suficiente, 59–61, 74, 79, 81–83, 92, 115–117, 133, 157, 196
suficiente bayesiano, 101
estadísticos de Kolmogorov-Smirnov, 407, 409, 410, 435
estadísticos ordenados, 33, 401
estimación
no paramétrica de densidades, 389, 425
por punto, 274
por regiones de confianza, 274
estimador, 81
Bayes, 108, 138
bayesiano, 100, 108, 109, 138–140, 144, 145, 154–156
centrado, 96, 97, 146, 148, 149, 158, 159
centrado de mínima varianza (ECMV), 109, 111–113, 115–117, 122, 124–126, 145, 150,

158–160, 283
consistente, 98, 142, 150
de máxima verosimilitud (EMV), 129–131, 133–135, 138, 140, 142, 146, 147, 151, 162–167, 182, 189, 191, 194, 283, 304, 316
de máxima verosimilitud (EMV) en sentido débil, 131
de mínimos cuadrados (EMC), 331, 341, 342
de Stein, 289
eficiente, 126, 127, 146, 161, 162
inadmisible, 107
insesgado, 81, 90, 115, 116, 142, 146, 148
núcleo, 428
óptimo, 112, 115, 116, 118
por punto, 95
evidencia estadística, 73

factor
Bayes, 226
de suavizado, 428
familia
conjugada, 103, 104, 151
de contrastes de hipótesis, 242, 244
de distribuciones completa, 65
de distribuciones invariante, 70
exponencial biparamétrica, 118
exponencial k-paramétrica, 64
exponencial uniparamétrica, 118, 122, 125, 126, 140, 156, 157, 160, 237, 241
Ferguson, 282, 295
Fishburn, 290
Fisher, 2, 25, 43, 44, 78, 79, 129, 141, 180, 186–188, 247, 318, 355, 453
Fisz, 21, 129
frecuencia relativa, 12
función de
influencia, 425
pérdida, 108, 109, 144, 271, 272, 274, 290, 291
pérdida cuadrática, 283, 292
potencia, 209, 230, 247, 249, 251
riesgo, 272, 274, 276, 280, 290, 292, 296
utilidad, 272
verosimilitud, 72, 130, 132, 133, 136, 163–166
inducida, 134
funcion núcleo, 428

Gómez Sánchez-Manzano, 228, 289
Gómez Villegas, 142, 228, 246, 272, 289, 390
Galton, 43, 78, 355, 452
Gauss, 355
Gibbons, 448
Giri, 28, 32
Glivenko-Cantelli, 21
Gnedenko, 431
Gosset, 43
grado de confianza, 172, 175, 190
grados de libertad, 22, 179, 224
grupo de transformaciones, 229

Hájek, 431, 435
hipótesis
alternativa, 207
bilaterales, 209
compuesta, 208
de homogeneidad, 455
nula, 207
simple, 208
unilaterales, 209
histograma de frecuencias, 44
Hodges, 453
homocedasticidad, 336
homogeneidad, 453
entre varias muestras, 396
Hotelling, 290, 453
Hsu, 179

Huber, 425

información, 121, 122
de Fisher, 119–121, 138, 140
final, 155, 156
inicial, 144, 151, 156, 184, 191, 247, 251
muestral, 156, 247
intercambiabilidad, 40–42, 47, 58
intervalo
asintótico, 190
creíble, 184, 185, 191, 201, 205, 293
de confianza, 171, 173, 176, 177, 179, 182, 183, 189, 192, 196–198
de confianza aproximado, 200
de confianza asintótico, 199
de confianza de amplitud mínima, 196
de confianza para el cociente de varianzas, 179
de confianza para la diferencia de medias, 178
de confianza para la media, 177, 198
de grado de confianza, 175, 177, 178, 180, 182, 183, 189
de tolerancia, 401–403
fiducial, 187
invariancia, 133

James, 286
Jeffreys, 247

Karl Pearson, 43, 44, 78, 127, 141, 186, 187, 355, 453
Kendall, 2, 453
Kintchine, 16
Kolmogorov, 453

Laplace, 43, 78, 141, 142, 246, 247, 293, 355
Lebesgue, 187
Legendre, 452
Lehmann, 2, 101, 125, 137, 140, 241, 416, 442, 453
lema de
Neyman-Pearson, 227
Stein, 288
Lindley, 2, 290
Loève, 16, 24

m-estimadores, 425
mínima diferencia significativa, 325
mínimos cuadrados, 330, 338
método
de estimación bayesiano, 138, 297
de K. Pearson, 462, 464
de la máxima verosimilitud, 129, 134, 146
de la variable pivotal, 195
de los mínimos cuadrados, 355
de los momentos, 127–129, 146, 163
de Neyman, 174, 175, 189, 196
de Scheffé, 325
minimax, 297
Maín, 228
Mann, 440, 453
Marín, 289
matriz de covarianzas, 340, 341
media, 44
muestral, 6
truncada, 12
mediana, 13, 44
de las desviaciones absolutas (MDA), 13
muestral, 12, 13, 163, 474
medidas de
centralización, 11
dispersión, 12
Memoria de Bayes, 246
Mendel, 393
Michelson, 422
Mishra, 179
modelos log-lineales, 224
momento muestral

respecto al centro, 6
respecto al origen, 6
momentos
cruzados, 8
marginales respecto al origen, 8
muestrales respecto al centro, 8, 12
muestrales respecto al origen, 11
poblacionales, 128, 129
respecto al origen, 128
Morgenstern, 290
muestra aleatoria simple (m.a.s.), 4

Navarro, 228
Newcomb, 452
Neyman, 187, 188, 245, 247
nivel de significación, 209

observaciones atípicas, 425
observaciones discordantes, 452
Ord, 325, 327

p-valor, 210, 211, 248, 251, 255, 256, 268, 269
Péano, 140
pérdida
0-1, 294
cuadrática, 108, 139, 144, 145, 155
final esperada (PFE), 105, 107, 108, 138, 139, 144, 154, 277
parámetro de no centralidad, 24
paradoja de
Jeffreys, 228
Lindley, 228
Simpson, 400
Stein, 283
pares
concordantes, 444
discordantes, 445
Pierce, 452
población, 4
Press, 354
Price, 246
primer cuartil, 13
principio de
condicionalidad, 74, 76
suficiencia, 74, 76
verosimilitud, 72, 73, 76, 77, 80, 145, 156
probabilidad
final, 225, 251
de la hipótesis alternativa, 225
de la hipótesis nula, 225
inversa, 246
problema
de homogeneidad, 389
problemas
de localización, 407
no paramétricos relativos a dos muestras, 429

Quetelet, 43, 141, 142, 355

Raiffa, 290
Ramsey, 247
Randle, 425
rango, 37
medio, 443
rangos, 443, 449
Rao, 114, 223, 392
razón de verosimilitudes (RV), 211, 214
recorrido intercuartílico muestral (RIM), 13, 44
recorrido intercuartílico poblacional, 13
recorrido muestral, 33, 37, 89, 90
recta de regresión, 331, 359, 371, 373, 378
recubrimientos, 37, 38, 402
reducción de datos, 72
región
crítica, 208
creíble, 184, 185, 191, 387

de más alta distribución final, 185
de más alta probabilidad, 191, 204
de más alta probabilidad final, 185, 204
de aceptación, 208
de confianza, 138, 169, 171, 181, 188, 189, 192
de confianza insesgada, 243
de confianza para la media y la varianza, 180
de confianza uniformemente más ajustada (UMA), 243, 245, 275
de grado de confianza, 181
insesgada, 245
regla
admisible, 280
Bayes, 273, 276–278, 281–283, 286, 309–313
Bayes generalizada, 308
de Bayes, 284
de Cromwell, 282
de decisión, 271, 272, 275, 296
de decisión admisible, 280, 282
de decisión Bayes, 291, 294, 299, 303, 304
de decisión minimax, 278, 282, 290, 291
de sucesión de Laplace, 305
igualizadora, 278
minimax, 276, 278, 283, 285, 313
regresión, 315
lineal, 333
lineal múltiple, 337, 342
lineal simple, 342
múltiple, 341, 346
replicación, 325
representación de variables intercambiables, 41
residuo, 317, 343, 359, 374, 376
Riemann, 140
riesgo
a posteriori, 108
Bayes, 273, 276, 286, 297, 312, 313
final, 300
robusta, 13
robusto, 452

Sanz, 228
Savage, 247, 290
segundo cuartil, 13
sesgo, 96, 106, 115, 143, 153
Sheynin, 43
Sidak, 431
Silverman, 425, 428
Simpson, 43
sin reemplazamiento, 5
Slutsky, 17, 137
Smirnov, 431, 453
Snedecor, 43
soporte, 93, 135, 156
soporte de la distribución, 119
Stein, 286
Stigler, 422
Strang, 338
Stuart, 2, 325, 327
Student, 43, 177, 187, 452
sucesión consistente, 136
Sudderth, 41
suma de cuadrados
total, 318

tabla
de análisis de la varianza (ADEVA), 321, 327, 365, 369
de contingencia, 398
de frecuencias, 10, 11, 14
tallo, 14, 49
tamaño, 209
Tchebycheff, 48, 183, 198
teoría de
la robustez, 453
teoría de
Juegos, 290

la Decisión, 271, 273, 289, 290
teorema de
Basu, 68, 69, 86
Bayes, 246, 247
Birnbaum, 75, 77
caracterización de estadísticos minimal suficientes, 91
existencia de estadísticos completos, 67, 92
factorización, 60, 92, 133
Fisher, 51, 320
Gauss-Markov, 344
Glivenko-Cantelli, 408, 430
Karlin-Rubin, 237, 239, 240, 265
la probabilidad total, 259
Lehmann-Scheffé, 115, 150, 157, 158, 160
Neyman, 197
Neyman-Pearson, 231–235, 237, 260, 263
Rao-Blackwell, 113, 115
Slutsky, 18, 19, 150, 151, 223
tercer cuartil, 13
Thomas Bayes, 140
Thompson, 452
traza, 340
Tuckey, 325
Tycho Brahe, 43, 141
universo, 4
valores discordantes, 389, 424, 452
variable
aleatoria de Snedecor, 56
aleatoria de Student, 55
dependiente, 330
estadística, 11
independiente, 330
pivotal, 172, 173, 175, 177–179, 188, 189, 192, 195, 196, 322
predictora, 330, 346
respuesta, 330
variables aleatorias intercambiables, 39, 40
varianza
muestral, 7
residual, 317
verosimilitud, 134, 135, 201
inducida, 133, 134
von Mises, 453
Von Neumann, 290
Wald, 290
Whitney, 440, 453
Wilcoxon, 407, 438, 441, 453
Wilks, 390
Wolfe, 425
Wright, 452
Yule, 356