

LUIS RUIZ-MAYA PÉREZ
JAVIER MARTÍN PLIEGO

*Fundamentos
de
Inferencia
Estadística*

Editorial AC

PRÓLOGO

En la investigación empírica existen conjuntos de elementos sometidos a análisis, denominados poblaciones, que estadísticamente se sintetizan en modelos caracterizados por expresiones matemáticas conteniendo uno o varios parámetros, con valores que permiten diferenciar unas poblaciones de otras de la misma familia.

El conocimiento completo del modelo hace posible el tratamiento estadístico poblacional. Sin embargo, en la realidad, la situación no es ésta, pues lo frecuente es hallarnos ante el desconocimiento de la expresión matemática y de los parámetros, conduciendo a la inutilidad práctica del modelo, siendo preciso obtener aproximaciones.

El tamaño de la población que, a efectos prácticos, puede suponerse infinito o inabarcable, impide o dificulta su análisis completo, por lo que sólo será factible estudiar una parte denominada muestra, generalizando a la población las conclusiones obtenidas a partir de la muestra.

El proceso de paso de lo particular (muestra) a lo general (población), inferencia estadística (inferencia inductiva), tiene como característica inherente que sus resultados no son exactos sino probables, frente a las conclusiones de la inferencia deductiva exactas y válidas en todas circunstancias.

Este manual, Fundamentos de Inferencia Estadística, expone en los dos primeros capítulos el concepto de estadísticos, como funciones de los valores muestrales, sus distribuciones de probabilidad, y los importantes temas de suficiencia e información, directamente relacionados con los estadísticos y con las muestras.

Basándose en la información proporcionada por la muestra, el proceso de la inferencia estadística contempla dos ámbitos: estimación de parámetros, procedimientos que proporcionan valores aproximados de los parámetros desconocidos, y contrastación de hipótesis: métodos que permiten optar por una de dos hipótesis establecidas sobre el valor de un parámetro o sobre el tipo de modelo matemático supuesto.

VI ■ FUNDAMENTOS DE INFERENCIA ESTADÍSTICA

La estimación de parámetros se desarrolla a lo largo de los capítulos tercero al quinto, estudiando los estimadores, sus propiedades, métodos de obtención y estimación por intervalos. La contrastación de hipótesis comprende los capítulos seis al nueve, tratando las contrastaciones paramétrica y no paramétrica, así como los contrastes de significación.

En el último capítulo se presenta, de forma somera, una introducción a la teoría de la decisión y a la Inferencia Bayesiana.

Madrid, Julio de 1999

LOS AUTORES

CONTENIDO

Capítulo 1

Muestreo: estadísticos y sus distribuciones.....	1
1.1 Población y muestra	1
1.2 Tipos de muestreo.....	2
1.2.1. Distribución conjunta de la muestra	5
1.3 Distribución de la población, distribución de la muestra y distribución en el muestreo. Estadísticos	6
1.4 Características de las distribuciones en el muestreo	14
1.4.1. Momentos muestrales con respecto al origen	14
1.4.2. Momentos muestrales con respecto a la media.....	16
1.4.3. Momentos muestrales de distribuciones bidimensionales	18
1.5 Estadísticos ordenados.....	20
1.5.1. Distribución del estadístico ordenado u_i	21
1.5.2. Distribución del menos valor de la muestra, u_1	22
1.5.3. Distribución del mayor valor de la muestra, u_n	23
1.6 Muestreo en poblaciones normales	24
1.7 Varianza poblacional conocida	24
1.7.1. Distribución de la media muestral	24
1.7.2. Distribución de la diferencia de medias muestrales	25
1.8 Distribución conjunta de la media y la varianza muestrales o Lema de Fisher-Cochram	26
1.9 Varianza poblacional desconocida	26
1.9.1. Distribución de la media muestral.....	26
1.9.2. Distribución de la diferencia de medias muestrales	27
1.10 Distribución en el muestreo del coeficiente de correlación lineal.....	29
1.11 Distribución de los coeficientes de regresión lineal muestrales	30
Ejercicios resueltos	31

Capítulo 2

Suficiencia e información	43
2.1 Parámetros desconocidos	43
2.2 Función de verosimilitud.....	44
2.3 Suficiencia.....	47
2.4 Criterio de factorización de Fisher-Neyman	52
2.5 Información	54
2.5.1. Información de Fisher.....	56
2.5.2. Estadístico suficiente e información de Fisher	60
Ejercicios resueltos	62

Capítulo 3

Estimación puntual: propiedades de los estimadores	67
3.1 Introducción.....	67
3.2 Concepto de estimador, estimación y error de estimación	68
3.2.1. Error cuadrático medio del estimador	69
3.2.2. Propiedades de los estimadores.....	71
3.3 Estimador insesgado	72
3.3.1. Propiedades de los estimadores insesgados	74
3.4. Estimador eficiente	77
3.4.1. Cota de Cramér-Rao.....	78
3.4.2. Unicidad del estimador eficiente	83
3.4.3. Eficiencia relativa.....	84
3.4.4. Estimadores UMVUE.....	85
3.5 Estimador consistente.....	86
3.5.1. Propiedades de los estimadores consistentes	88
3.6 Suficiencia.....	90
3.6.1. Criterio de factorización de Fisher-Neyman.....	91
3.7 Estimador invariante.....	92
3.8 Estimador robusto.....	93
Ejercicios resueltos	95

Capítulo 4

Métodos de estimación	109
4.1 Metodo de estimación de máxima verosimilitud.....	109
4.2 Propiedades de los estimadores máximo-verosímiles	116
4.3 Estimación por el método de los momentos	117
4.4 Estimación por el método de los mínimos cuadrados	119
Ejercicios resueltos	122

Capítulo 5

Estimación por intervalos	131
5.1 Introducción.....	131
5.2 Métodos de construcción de intervalos de confianza	133
5.2.1. Método de la cantidad pivotal.....	133
5.2.2. Metodo general.....	136
5.2.3. Intervalos de confianza de longitud mínima	139
5.3 Intervalos de confianza en poblaciones normales	141
5.3.1. Intervalos de confianza para la media μ de una distribución normal siendo σ conocida.....	142
5.3.2. Intervalo de confianza para la media μ de una distribución normal siendo σ desconocida.....	145
5.3.3. Intervalo de confianza para la varianza σ^2 de una población normal	146
5.4 Intervalos de confianza en poblaciones no normales.....	148
5.4.1. Caso general.....	148
5.4.2. Intervalos de confianza basados en grandes muestras.....	150
5.5 Determinación del tamaño de la muestra	153

Capítulo 6

Contrastación de hipótesis	157
6.1 Conceptos generales	157
6.2 Contraste de hipótesis	158
6.2.1. Hipótesis simples	160
6.2.2. Lema de Neyman-Pearson	163
6.3 Hipótesis compuestas	171
6.3.1. Función de potencia	172
6.3.2. Contrastes uniformemente más potentes (CUMP).....	174
Ejercicios resueltos	177

Capítulo 7

Contrastes de significación.....	183
7.1. Principios básicos	183
7.1.1. Contraste de la media de una población normal con σ conocida	186
7.1.2. Contraste de la media de una población normal con σ desconocida ...	188
7.2 Contraste razón de verosimilitud	190
7.3 Contrastes asintóticos basados en la función de verosimilitud.....	197
7.3.1. Test de Wald	197
7.3.2. Test de los multiplicadores de Lagrange.....	200
7.3.3. Comparación de los tests razón de verosimilitud, Wald y multiplicadores de Lagrange	204
7.4 Relación entre los contrastes de hipótesis y los intervalos de confianza.....	205

Capítulo 8

Contrastes paramétricos..... 209

8.1	Contrastación de hipótesis sobre la media de una población normal	209
8.1.1.	Varianza poblacional conocida.....	209
8.1.1.1.	Hipótesis simples: $H_0 [\mu = \mu_0]$, $H_1 [\mu = \mu_1]$	209
8.1.1.2.	Hipótesis compuestas: $H_0 [\mu = \mu_0]$, $H_1 [\mu > \mu_0]$ o $H_1 [\mu < \mu_0]$...	215
8.1.1.3.	Hipótesis compuesta: $H_0 [\mu = \mu_0]$, $H_1 [\mu \neq \mu_0]$	216
8.1.2.	Varianza poblacional desconocida	219
8.1.2.1.	Hipótesis simples: $H_0 [\mu = \mu_0]$, $H_1 [\mu = \mu_1]$	219
8.1.2.2.	Hipótesis compuestas: $H_0 [\mu = \mu_0]$, $H_1 [\mu > \mu_0]$ o $H_1 [\mu < \mu_0]$...	221
8.1.2.3.	Hipotesis compuesta: $H_0 [\mu = \mu_0]$, $H_1 [\mu \neq \mu_0]$	222
8.1.3.	Cuadro resumen de los contrastes de medias en población normales .	224
8.2	Contrastación de hipótesis sobre la varianza de una población normal	226
8.2.1.	Media poblacional conocida	226
8.2.1.1.	Hipótesis simples: $H_0 [\sigma^2 = \sigma_0^2]$, $H_1 [\sigma^2 = \sigma_1^2]$	226
8.2.1.2.	Hipótesis compuestas: $H_0 [\sigma^2 = \sigma_0^2]$, $H_1 [\sigma^2 > \sigma_0^2]$ o $H_1 [\sigma^2 < \sigma_0^2]$	231
8.2.1.3.	Hipótesis compuesta: $H_0 [\sigma^2 = \sigma_0^2]$, $H_1 [\sigma^2 \neq \sigma_0^2]$	231
8.2.2.	Media poblacional desconocida	235
8.2.2.1.	Hipótesis simple: $H_0 [\sigma^2 = \sigma_0^2]$, $H_1 [\sigma^2 = \sigma_1^2]$	235
8.2.2.2.	Hipótesis compuestas: $H_0 [\sigma^2 = \sigma_0^2]$, $H_1 [\sigma^2 > \sigma_0^2]$ o $H_1 [\sigma^2 < \sigma_0^2]$	238
8.2.2.3.	Hipótesis compuestas: $H_0 [\sigma^2 = \sigma_0^2]$, $H_1 [\sigma^2 \neq \sigma_0^2]$	238
8.2.3.	Cuadro resumen de los contrastes de varianza en poblaciones normales.....	238
8.3	Contrastación de hipótesis sobre el coeficiente de correlación lineal de una población normal bidimensional.....	240
8.3.1.	Contraste de incorrelación lineal: $H_0[\rho = 0]$, $H_1[\rho \neq 0]$	240
8.3.2.	Contraste de asociación lineal: $H_0[\rho = \rho_0]$, $H_1[\rho \neq \rho_0]$	241
8.4	Contrastación de hipótesis sobre los coeficientes de regresión de una población normal	242
8.5	Contrastes entre dos poblaciones	244
8.5.1.	Contrastes de igualdad de medias	244
8.5.1.1.	Varianzas poblacionales conocidas	244
8.5.1.2.	Varianza poblacional común desconocida	249
8.5.2.	Contraste de igualdad de varianzas	253
8.5.2.1.	Medias poblacionales conocidas.....	253
8.5.2.2.	Medias poblacionales desconocidas.....	258
8.6	Contrastación de hipótesis en muestras grandes	259
8.6.1.	Planteamiento general.....	259
8.6.2.	Contraste de proporciones.....	260

Capítulo 9

Contrastes no paramétricos	271
9.1 Introducción.....	271
9.2 Contrastes de bondad del ajuste.....	273
9.2.1. Test χ^2	274
9.2.1.1. Test χ^2 cuando no se estiman parámetros.....	274
9.2.1.2. Test χ^2 cuando se estiman parámetros	278
9.2.2. Test G^2 de la razón de verosimilitud	281
9.2.3. Test de Kolmogorov-Smirnov	284
9.2.4. Comparación de los tests χ^2 , G^2 y de Kolmogorov-Smirnov.....	286
9.3 Test de rachas para el contraste de aleatoriedad.....	287
9.4 Test de Wilcoxon: contraste de localización	290
9.5 Test de correlación por rangos de Spearman para asociación	293
9.6 Tablas de contingencia	298
9.6.1. Contraste de independencia	299
9.6.2. Contraste de homogeneidad	304

Capítulo 10

Introducción a la teoría de la decisión e inferencia bayesiana	309
10.1 Teoría de la decisión.....	309
10.1.1. Tipología de la decisión	311
10.1.2. Decisión sin experimentación.....	311
10.1.2.1. Decisión en ambiente de incertidumbre: Criterio minimax	312
10.1.2.2. Decisión con riesgo: Solución de Bayes	312
10.1.3. Decisión con experimentación.....	313
10.1.3.1. Función de decisión	313
10.1.3.2. Función de pérdida	313
10.1.3.3. Función de riesgo.....	314
10.1.3.4. Decisión en ambiente de incertidumbre: Criterio minimax	314
10.1.3.5. Decisión con riesgo: Decisión de Bayes.....	317
10.2 Inferencia bayesiana: distribuciones a priori y a posteriori.....	318
10.3 Estimación de Bayes.....	319
10.4 Contrastación bayesiana de hipótesis.....	324
Ejercicios resueltos	327
Bibliografía	333
Tablas estadísticas	337
Índice analítico	349

CAPÍTULO 1

Muestreo: estadísticos y sus distribuciones

1.1 Población y muestra

El conjunto de todos los posibles resultados de un fenómeno, o experimento aleatorio, recibe el nombre de **población**¹, denominándose **elemento** a cada uno de sus componentes, pudiendo ser su número finito o, en teoría, infinito. Cabe, también, contemplar una población como un conjunto de elementos con una o más características comunes. La población se caracteriza probabilísticamente mediante variables aleatorias y éstas por sus campos de variación y distribuciones de probabilidad, que especifican el comportamiento aleatorio de la población.

Una **muestra** es un subconjunto² de una población. Si el número de elementos del subconjunto es n , diremos que la muestra es de tamaño n . También cabe definir la muestra como la observación de n elementos de una población, o el resultado de repetir n veces un experimento aleatorio. Designaremos la muestra genérica por X ,

¹ También universo o colectivo.

² En la práctica, finito.

y al referirnos a sus elementos por (x_1, \dots, x_n) si es de tamaño n , de tal forma que x_i es el elemento que corresponde a la i -ésima observación.

Si tomáramos las posibles muestras de tamaño n de una población, el conjunto de todas ellas constituye el **espacio muestral**, \mathfrak{X} , integrado por tantos elementos como muestras distintas sea posible obtener, y dado que cada muestra viene definida por un conjunto de n valores, (x_1, \dots, x_n) podemos pensar, también, en el espacio muestral como incluido en un espacio n -dimensional, en el que las coordenadas de cada punto son los correspondientes elementos muestrales, (x_1, \dots, x_n) , de cada una de ellas.

Según lo expuesto, una muestra concreta puede ser considerada como la realización de un experimento consistente en la extracción de una muestra de entre todas las integrantes del espacio muestral, por lo cual es posible contemplar la muestra genérica \mathbf{X} como una variable aleatoria n -dimensional, cuya distribución de probabilidad depende de la distribución de probabilidad de la variable aleatoria poblacional, $F(x)$, del procedimiento de selección y del tamaño n de la muestra.

La finalidad «ideal» perseguida al tomar una muestra es la **representación** a escala de la población, por lo que el procedimiento seguido para obtener la muestra debe garantizar esta representación «perfecta».

En una bolsa tenemos mil bolas del mismo tamaño: 900 de color blanco y 100 negro. Para que una muestra sea *exactamente representativa* deberá contener bolas con los colores en la misma proporción que en la población: 90% blancas y 10% negras. Si el tamaño muestral es 100 debería haber en tal muestra 90 bolas de color blanco y 10 de color negro. En la mayoría de los casos esto no sucede, apartándose el resultado muestral de la situación ideal deseada (por ejemplo, la muestra obtenida puede estar constituida por 87 bolas blancas y 13 negras), atribuyéndose las desviaciones al proceso de selección de la muestra, desviaciones que no invalidan los resultados del proceso inferencial, siempre que las diferencias no sean sistemáticas y tengan origen aleatorio, es decir, sean debidas al azar.

1.2 Tipos de muestreo

El procedimiento utilizado para la obtención de cada uno de los elementos muestrales puede ser aleatorio o no, disyuntiva que conduce a que la selección de la muestra sea probabilística o no.

Para garantizar la necesaria **representatividad de la muestra**, los elementos deben ser extraídos de la manera más objetiva posible, a fin de no influir en la selección de cada uno de ellos, lográndose la representatividad cuando la extracción se realiza al azar. Si ésto es así, diremos que la muestra es probabilística o aleato-

ria. Por otra parte, el procedimiento probabilístico permitirá, como veremos en capítulos posteriores, conocer en términos de probabilidad el error que se comete al utilizar la muestra como reflejo de la población.

Como ejemplo de selección no probabilística tenemos la **muestra opinática** donde el investigador elige, con criterios no aleatorios, los elementos que va a estudiar. Esa manera de proceder impide medir el posible error que pueda cometerse en el análisis que se esté realizando.

En el muestreo probabilístico se distinguen dos tipos dependiendo de cual sea el procedimiento aleatorio de extracción utilizado, pues influye en la probabilidad de obtención de los elementos de la muestra. Podemos extraer los elementos muestrales de dos maneras: con reemplazamiento y sin reemplazamiento.

El muestreo se lleva a cabo **con reemplazamiento** cuando una vez extraído un elemento, y hechas sobre él las observaciones oportunas, se devuelve al colectivo, lo que supone que la población no es modificada y el elemento puede ser elegido de nuevo. Esta forma de actuar implica que la probabilidad de extracción de cada elemento es la misma para todos ellos, pues la población no cambia.

El muestreo tiene lugar **sin reemplazamiento** cuando el elemento elegido se excluye del colectivo y, por consiguiente, éste se modifica de tal manera que la probabilidad de extracción de un elemento depende de cuántos y cuáles hayan salido con anterioridad³.

El muestreo con reemplazamiento conduce a que los elementos de la muestra sean probabilísticamente independientes, mientras que en el muestreo sin reemplazamiento no lo son.

En todo el estudio estadístico que realizaremos en este manual las muestras serán aleatorias con reemplazamiento, denominándolas **muestras aleatorias simples**.⁴

Si en el ejemplo anterior la muestra de tamaño 100 se obtiene *con reemplazamiento*, la probabilidad de que la primera bola sea blanca es $900/1000$, al devolver la bola a la bolsa la probabilidad de sacar la segunda bola blanca es, también, $900/1000$ y así sucesivamente. En el muestreo *sin reemplazamiento*, la bola extraída se aparta, la probabilidad de extraer la primera bola blanca es $900/1000$, igual que en el caso anterior, pero la probabilidad de la segunda blanca es $899/999$, pues la primera no fue devuelta a la bolsa, produciendo tal manera de proceder una modificación de la estructura de la bolsa al pasar de un total de 1000 bolas a 999 y de 900 bolas blancas a 899.

³ Se supone que, junto con el procedimiento aleatorio, se mantienen fijas las condiciones bajo las que se realiza el experimento. Esta condición es relativamente factible en la experimentación en las ciencias físicas, químicas, etc., no es así en las sociales donde el control es, en el mejor de los casos, difícil.

⁴ Algunos autores lo denominan muestreo aleatorio.

EJEMPLO 1

Tenemos 100 bolas de las mismas características numeradas con 1, 2 y 3 en la siguiente cuantía: 20 con el número 1, 30 con el 2 y 50 con el 3. Se toman muestras aleatorias simples de tamaño dos y construimos su espacio muestral bidimensional \mathfrak{X} . En el plano, el espacio muestral está constituido por los puntos cuyas coordenadas son las posibles muestras distintas de tamaño dos

3	(1, 3)	(2, 3)	(3, 3)
2	(1, 2)	(2, 2)	(3, 2)
1	(1, 1)	(2, 1)	(3, 1)
	1	2	3

La muestra (1, 3), por ejemplo, está integrada por una bola con el número 1 y otra con el 3.

Las probabilidades que tienen de ser elegidas estas muestras son

3	0,10	0,15	0,25
2	0,06	0,09	0,15
1	0,04	0,06	0,10
	1	2	3

El cálculo de las probabilidades se ha hecho de la siguiente forma, por ejemplo, para la muestra (1, 3),

$$P(1, 3) = P[(\text{bola } 1) \cap (\text{bola } 3)] = P(1 \cap 3)$$

como la muestra es aleatoria simple sus elementos son independientes y la probabilidad de su intersección igual al producto de las probabilidades, quedando

$$P(1, 3) = P(1) \cdot P(3) = 0,2 \cdot 0,5 = 0,10.$$

La probabilidad de extraer una muestra con un 1 y un 3, sin importar el orden de obtención se calcula del siguiente modo: Esta combinación de valores de la variable puede proceder de la muestra (1, 3) o de la muestra (3, 1), por tanto,

$$P(\text{muestra con un 1 y un 3}) = P[\text{muestra } (1, 3) \cup \text{muestra } (3, 1)]$$

como las dos muestras son sucesos disjuntos, la probabilidad es igual a

$$P[\text{muestra } (1, 3)] + P[\text{muestra } (3, 1)] = 0,10 + 0,10 = 0,20.$$

EJEMPLO 2

En el ejemplo 1 obtuvimos para las muestras aleatorias simples de tamaño 2 las probabilidades correspondientes en su espacio muestral \mathfrak{X} . A continuación hacemos lo mismo pero siendo las muestras aleatorias extraídas sin reemplazamiento, lo que implica que sus elementos no son independientes. El ejemplo se desarrolla en dos cuadros. En el primero tenemos el proceso de cálculo de las probabilidades y, en el segundo, sus valores.

3	$\frac{20}{100} \frac{50}{99}$	$\frac{30}{100} \frac{50}{99}$	$\frac{50}{100} \frac{49}{99}$
2	$\frac{20}{100} \frac{30}{99}$	$\frac{30}{100} \frac{29}{99}$	$\frac{50}{100} \frac{30}{99}$
1	$\frac{20}{100} \frac{19}{99}$	$\frac{30}{100} \frac{20}{99}$	$\frac{50}{100} \frac{20}{99}$
	1	2	3

3	0,1010	0,1515	0,2475
2	0,0606	0,0879	0,1515
1	0,0384	0,0606	0,1010
	1	2	3

Se aprecia, con claridad, que los valores de las probabilidades para cada clase de muestra difieren de un tipo de muestreo a otro.

1.2.1. DISTRIBUCIÓN CONJUNTA DE LA MUESTRA

La probabilidad de extracción de una muestra aleatoria simple concreta (x_1, \dots, x_n) , si la variable poblacional es discreta con función de cuantía $P(\xi = x)$, se calcula de la siguiente forma.

El suceso final es $\{\xi_1 = x_1\} \cap \{\xi_2 = x_2\} \cap \dots \cap \{\xi_n = x_n\}$ y como la muestra es aleatoria simple sus elementos son independientes, por lo cual,

$$P(x_1, x_2, \dots, x_n) = P[\{\xi_1 = x_1\} \cap \{\xi_2 = x_2\} \cap \dots \cap \{\xi_n = x_n\}] = P(\xi_1 = x_1) \cdot P(\xi_2 = x_2) \dots P(\xi_n = x_n)$$

siendo $P(\xi = x_i)$ la probabilidad de obtener en la población un elemento cuyo valor sea x_i y $P(x_1, \dots, x_n)$ la función de cuantía conjunta de la muestra.

En el caso que la variable aleatoria poblacional sea continua, con función de densidad $f(x)$, la probabilidad elemental de obtener el resultado muestral concreto (x_1, \dots, x_n) , por ser la muestra aleatoria simple, es

$$f(x_1, \dots, x_n) dx_1 \dots dx_n,$$

donde $f(x_1, \dots, x_n)$ es la función de densidad conjunta de la muestra, verificándose

$$f(x_1, \dots, x_n) = f(x_1) \dots f(x_n).$$

por ser independientes sus elementos.

En una **muestra aleatoria simple** (x_1, \dots, x_n) se verifican las siguientes relaciones entre sus elementos:

$$\begin{array}{l} \text{I) } F(x_1) = F(x_2) = \dots = F(x_n) = F(x) \\ \text{II) } F(x_1, x_2, \dots, x_n) = F(x_1) \cdot F(x_2) \dots F(x_n) \end{array}$$

es decir, «*las variables aleatorias x_i son independientes e idénticamente distribuidas con la misma distribución de probabilidad que tenga la población*».

La manera práctica de obtener una muestra aleatoria simple es la de realizar el muestreo con reemplazamiento.

Si la muestra no fuera aleatoria simple (es decir, si la extracción hubiera tenido lugar sin reemplazamiento y, por consiguiente, sus elementos no son independientes), la modificación de la estructura poblacional conduce a que la probabilidad de una extracción venga condicionada por las extracciones anteriores, y la función de cuantía conjunta es

$$P(x_1, \dots, x_n) = \prod_{i=1}^n P(\xi_i = x_i / \xi_1 = x_1, \dots, \xi_{i-1} = x_{i-1})$$

y la función de densidad conjunta

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

1.3 Distribución de la población, distribución de la muestra y distribución en el muestreo. Estadísticos

Un concepto de gran importancia en el planteamiento de la Inferencia Estadística es el de **función de distribución empírica de la muestra**, definida en una muestra de tamaño n como

$$F_n(x_i) = \frac{N_i}{n},$$

siendo N_i el número de observaciones muestrales menores o iguales que x_i , es decir, la **frecuencia acumulada**. Esta función presenta tantos saltos como valores muestrales distintos haya, siendo la cuantía del salto $\frac{1}{n}$ cuando no se repite el valor

x_i , y $\frac{n_i}{n}$ cuando x_i se repite n_i veces, lo que indica que la función de distribución empírica es siempre discreta.

En la función de distribución empírica de la muestra podemos calcular todos sus momentos, uni o k -dimensionales, con respecto al origen o a la media, como en una población cualquiera y, para distinguirlos de los poblacionales, se les denominará momentos **de la muestra** o **muestrales**, representándolos por a_r o m_r (según sean respecto al origen o respecto a la media).

Una bolsa contiene 1.000 bolas, todas de igual tamaño, y marcadas con cuatro números distintos en la siguiente cuantía: 400 con el número 1, 100 con el 2, 300 con el 3 y las 200 restantes con el 4. El campo de variación de la variable aleatoria ξ , numeración de las bolas, está integrado por los cuatro números enteros 1, 2, 3 y 4, y la **distribución de probabilidad de la población** es

$$\begin{aligned}
 P(\xi = 1) &= 0,4; & P(\xi = 2) &= 0,1; \\
 P(\xi = 3) &= 0,3; & P(\xi = 4) &= 0,2.
 \end{aligned}$$

Se expresa esta distribución mediante el diagrama de barras de la función de cuantía que aparece en la figura 1.1(a).

Tomamos una muestra aleatoria simple de tamaño 100, siendo el resultado 43 bolas con el número 1, 6 con el 2, 28 con el 3 y 23 con el número 4. Los elementos muestrales tienen el mismo campo de variación que en la población. La **distribución de frecuencias de la muestra** obtenida es la siguiente:

$$\begin{aligned}
 \frac{n_1}{n} &= 0,43; & \frac{n_2}{n} &= 0,06; \\
 \frac{n_3}{n} &= 0,28; & \frac{n_4}{n} &= 0,23.
 \end{aligned}$$

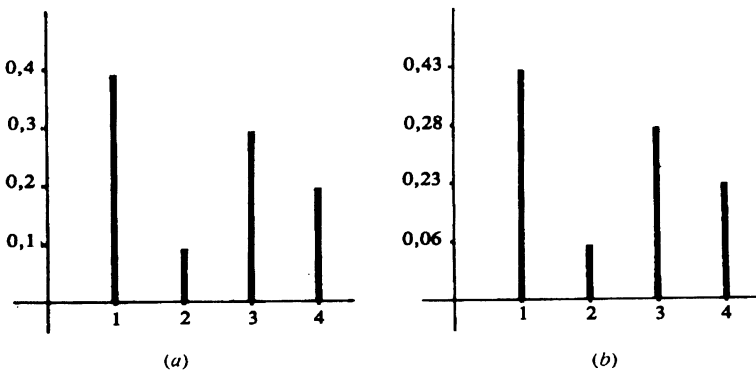


FIGURA 1.1

- (a) Distribución de la población.
- (b) Distribución de la muestra de tamaño 100.

Si comparamos el diagrama de barras muestral, Fig. 1.1(b), con el poblacional apreciamos que pese a ser muy parecidos no coinciden, pues la muestra no reproduce exactamente la estructura de la población, debiéndose esta diferencia a la variabilidad introducida en la estricta aleatoriedad teórica de la muestra. Si tomamos otras muestras, cada una de ellas tendrá su propia distribución, que se aproximará tanto más a la de la población cuanto «más aleatorio»⁵ haya sido el proceso de selección, es decir, «más objetivo».

En general, en una muestra concreta, sus características (momentos, etc.) no tienen por qué coincidir exactamente con las correspondientes de la población a causa de la aleatoriedad del procedimiento de extracción de los elementos (el azar, no mecanicista, no asegura la «objetividad absoluta»), pero si la muestra ha sido tomada con las máximas garantías de aleatoriedad, con la máxima objetividad, es de esperar que los valores de las características muestrales no se alejen demasiado de los poblacionales, lo que proporciona a la muestra sus posibilidades inductivas.

EJEMPLO 3

Consideremos una característica cualquiera de la población y la correspondiente de la muestra, por ejemplo, la esperanza matemática, la media. En la población anterior su valor es

$$\begin{aligned}\mu &= 1 \cdot P(\xi = 1) + 2 \cdot P(\xi = 2) + 3 \cdot P(\xi = 3) + 4 \cdot P(\xi = 4) = \\ &= 1 \cdot 0,4 + 2 \cdot 0,1 + 3 \cdot 0,3 + 4 \cdot 0,2 = \\ &= 2,3\end{aligned}$$

y en la muestra

$$\begin{aligned}\bar{x} &= 1 \cdot \frac{n_1}{n} + 2 \cdot \frac{n_2}{n} + 3 \cdot \frac{n_3}{n} + 4 \cdot \frac{n_4}{n} = \\ &= 1 \cdot 0,43 + 2 \cdot 0,06 + 3 \cdot 0,28 + 4 \cdot 0,23 = \\ &= 2,31\end{aligned}$$

valores que, aunque muy próximos, no coinciden.

Ahora bien, lo que realmente confiere a las muestras aleatorias toda su potencialidad inductiva, es que la función de distribución empírica converge en probabilidad a la función de distribución poblacional (teorema de Glivenko-Cantelli,⁶ denomina-

⁵ Si fuera posible emplear esta expresión, científicamente inadecuada.

⁶ Si $D_n = \sup_{-\infty < x < \infty} |F_n^*(x) - F(x)|$ y $\lim_{n \rightarrow \infty} P(D_n \geq \varepsilon) = 0$. MARTÍN PLIEGO y RUIZ-MAYA: *Fundamentos de Probabilidad*, Capítulo 9.

do **teorema fundamental de la Estadística**) implicando, como veremos, que las características muestrales, bajo condiciones generales, convergen en probabilidad a las correspondientes poblacionales⁷.

Estadístico

Cualquier función de los elementos muestrales recibe el nombre de **estadístico**, siempre que no contenga parámetros desconocidos, designándose por $T(\mathbf{X}) = T(x_1, \dots, x_n)$, por ejemplo,

$$T_1(\mathbf{X}) = x_1 + \dots + x_n = \sum_{i=1}^n x_i$$

$$T_2(\mathbf{X}) = x_1^2 + \dots + x_n^2 = \sum_{i=1}^n x_i^2$$

$$T_3(\mathbf{X}) = \frac{\sum_{i=1}^n x_i}{n} = \bar{x}$$

$$T_4(\mathbf{X}) = u_1 = \text{mín}(x_1, \dots, x_n)$$

$$T_5(\mathbf{X}) = \sum_{i=1}^n \ln x_i$$

En particular, trataremos con estadísticos muy concretos: los momentos muestrales (concretamente, media, varianza, covarianza), el valor mínimo o máximo de la muestra, etc.

Los elementos que integran la muestra son variables aleatorias, por lo que cualquier función de estos elementos, el estadístico, también será variable aleatoria.⁸ Como tal variable aleatoria el estadístico tendrá su propio campo de variación y su distribución de probabilidad determinados, a su vez, unívocamente por el campo de variación y la distribución de la población.

El campo de variación del estadístico es el conjunto de valores que toma para cada uno de los elementos del espacio muestral correspondiente. Si consideramos un estadístico cualquiera, tomamos todas las posibles muestras y en cada una de ellas calculamos su valor, habremos obtenido todos sus posibles valores, su campo de variación.

⁷ De ahí el acierto de CRAMÉR al designar la muestra como *imagen estadística de la población*.

⁸ MARTÍN PLIEGO y RUIZ-MAYA: *Fundamentos...* Capítulo 2.

Dado que un estadístico se genera en el proceso de muestreo, su distribución de probabilidad recibe el nombre de **distribución de probabilidad en el muestreo**. El concepto de distribución en el muestreo y el teorema de Glivenko-Cantelli constituyen las piezas clave de la Inferencia Estadística, al ser la base sobre la que se construye todo el edificio de estimación y contrastación de hipótesis que se estudiarán más adelante.

Por otra parte, según se estudiará con detalle en el capítulo 2, el estadístico induce una **partición** del espacio muestral \mathfrak{X} .

Para analizar con más detalle el concepto de distribución en el muestreo, tomaremos una variable discreta, calculando las medias y varianzas de las correspondientes distribuciones en el muestreo.

EJEMPLO 4

Una variable aleatoria ξ presenta los valores 1, 2 y 3, con probabilidades 0,1, 0,2 y 0,7. Tomamos muestras aleatorias simples de tamaño 3 y consideramos como estadístico la media muestral. En el cuadro de la página siguiente tenemos todas las posibles muestras distintas (clases de muestras), el valor de la media de cada muestra y la probabilidad de obtención de estas muestras.

El cálculo de las medias y de las probabilidades se ha realizado de la siguiente forma:

El número de valores de la media de cada muestra depende del tipo de muestra, pues hay valores medios que sólo aparecen en una única clase de muestra así, por ejemplo, en la muestra (1, 1, 1) la media vale 1, $\frac{1+1+1}{3} = 1$, y no se repite.

En otros casos, el valor de la media corresponde a un único tipo de muestra, aunque sus elementos aparezcan en distinto orden, por ejemplo, la muestra (1, 1, 2) puede aparecer con tres ordenaciones distintas (1, 1, 2), (1, 2, 1) y (2, 1, 1) y, en los tres casos, la media es la misma, 4/3.

Por último, la media puede tomar un valor procedente de clases de muestras distintas: si la media es igual a 7/3, este valor puede proceder de la muestra tipo (2, 2, 3) o de la muestra tipo (3, 3, 1), en ambos casos tenemos tres posibles muestras.

Partiendo de lo expuesto, calculamos las probabilidades de obtención de los valores concretos de la media en estos tres ejemplos:

■ *Media igual a 1*

$$\begin{aligned} P(\bar{x} = 1) &= P[\text{obtención de la muestra } (1, 1, 1)] = \\ &= P[(\text{obtener 1 en la primera extracción}) \cap \\ &\quad \cap (\text{obtener 1 en la segunda extracción}) \cap \\ &\quad \cap (\text{obtener 1 en la tercera extracción})]. \end{aligned}$$

<i>Clases de muestras</i>	<i>Muestras distintas</i>	<i>Medias muestrales</i>	<i>Probabilidades de obtención de las muestras</i>
(1, 1, 1)	(1, 1, 1)	1	$0,1 \cdot 0,1 \cdot 0,1 = 0,001$
(1, 1, 2)	(1, 1, 2)	4/3	$0,1 \cdot 0,1 \cdot 0,2 = 0,002$
	(1, 2, 1)	4/3	$0,1 \cdot 0,2 \cdot 0,1 = 0,002$
	(2, 1, 1)	4/3	$0,2 \cdot 0,1 \cdot 0,1 = 0,002$
		4/3	0,006
(1, 1, 3)	(1, 1, 3)	5/3	$0,1^2 \cdot 0,7 = 0,007$
	(1, 3, 1)	5/3	0,007
	(3, 1, 1)	5/3	0,007
		5/3	0,021
(1, 2, 3)	(1, 2, 3)	2	$0,1 \cdot 0,2 \cdot 0,7 = 0,014$
	(1, 3, 2)	2	0,014
	(2, 1, 3)	2	0,014
	(2, 3, 1)	2	0,014
	(3, 1, 2)	2	0,014
	(3, 2, 1)	2	0,014
		2	0,084
(2, 2, 1)	(1, 2, 2)	5/3	$0,1 \cdot 0,2^2 = 0,004$
	(2, 1, 2)	5/3	0,004
	(2, 2, 1)	5/3	0,004
		5/3	0,012
(2, 2, 2)	(2, 2, 2)	2	$0,2^3 = 0,008$
(2, 2, 3)	(2, 2, 3)	7/3	$0,2^2 \cdot 0,7 = 0,028$
	(2, 3, 2)	7/3	0,028
	(3, 2, 2)	7/3	0,028
		7/3	0,084
(3, 3, 1)	(3, 3, 1)	7/3	$0,7^2 \cdot 0,1 = 0,049$
	(3, 1, 3)	7/3	0,049
	(1, 3, 3)	7/3	0,049
		7/3	0,147
(3, 3, 2)	(3, 3, 2)	8/3	$0,2 \cdot 0,7^2 = 0,098$
	(3, 2, 3)	8/3	0,098
	(2, 3, 3)	8/3	0,098
		8/3	0,294
(3, 3, 3)	(3, 3, 3)	3	$0,7^3 = 0,343$

Como la muestra es aleatoria simple sus elementos son independientes⁹ y la probabilidad de la intersección de sucesos independientes es igual al producto de las probabilidades de cada uno de los sucesos, siendo la expresión igual a:

$$\begin{aligned}
 & P(\text{obtener 1 en la primera extracción}) \cdot \\
 & \quad \cdot P(\text{obtener 1 en la segunda extracción}) \cdot \\
 & \quad \cdot P(\text{obtener 1 en la tercera extracción}) = \\
 & = P(1) \cdot P(1) \cdot P(1) = [P(1)]^3 = 0,1^3 = \\
 & = 0,001.
 \end{aligned}$$

■ Media igual a 4/3

$$P(\bar{x} = 4/3) = P[\text{muestra } (1, 1, 2) \cup \text{muestra } (1, 2, 1) \cup \text{muestra } (2, 1, 1)].$$

Las muestras son aleatorias simples y disjuntas, por lo que, en cada una los elementos son independientes siendo la probabilidad igual a

$$\begin{aligned}
 P(\bar{x} = 4/3) &= P(1) \cdot P(1) \cdot P(2) + P(1) \cdot P(2) \cdot P(1) + P(2) \cdot P(1) \cdot P(1) = \\
 &= 3 P(1) \cdot P(1) \cdot P(2) = 3 \cdot 0,1^2 \cdot 0,2 = \\
 &= 0,006.
 \end{aligned}$$

■ Media igual a 7/3

$$\begin{aligned}
 P(\bar{x} = 7/3) &= P[\text{muestra tipo } (2, 2, 3) \cup \text{muestra tipo } (3, 3, 1)] = \\
 &= P[\text{muestra tipo } (2, 2, 3)] + P[\text{muestra tipo } (3, 3, 1)]
 \end{aligned}$$

$$\begin{aligned}
 P[\text{clase de muestra } (2, 2, 3)] &= P[\text{muestra } (2, 2, 3) \cup \text{muestra } (2, 3, 2) \cup \text{muestra } (3, 2, 2)] = \\
 &= P[\text{muestra } (2, 2, 3)] + P[\text{muestra } (2, 3, 2)] + P[\text{muestra } (3, 2, 2)] = \\
 &= P(2) P(2) P(3) + P(2) P(3) P(2) + P(3) P(2) P(2) = \\
 &= 3 [P(2)]^2 P(3) = 3 \cdot 0,2^2 \cdot 0,7 = \\
 &= 0,084
 \end{aligned}$$

$$P[\text{clase de muestra } (3, 3, 1)] = 3 [P(3)]^2 P(1) = 3 \cdot 0,7^2 \cdot 0,1 = 0,147$$

$$\begin{aligned}
 P(\bar{x} = 7/3) &= P[\text{clase de muestra } (2, 2, 3)] + P[\text{clase de muestra } (3, 3, 1)] = \\
 &= 0,084 + 0,147 = 0,231.
 \end{aligned}$$

⁹ El muestreo se ha realizado con reemplazamiento.

La distribución de probabilidad *en* el muestreo de la media muestral resulta:

Distribución *en* el muestreo de la media muestral

\bar{x}_i	$P(\bar{x} = \bar{x}_i)$
1	0,001
4/3	0,006
5/3	0,033
2	0,092
7/3	0,231
8/3	0,294
3	0,343

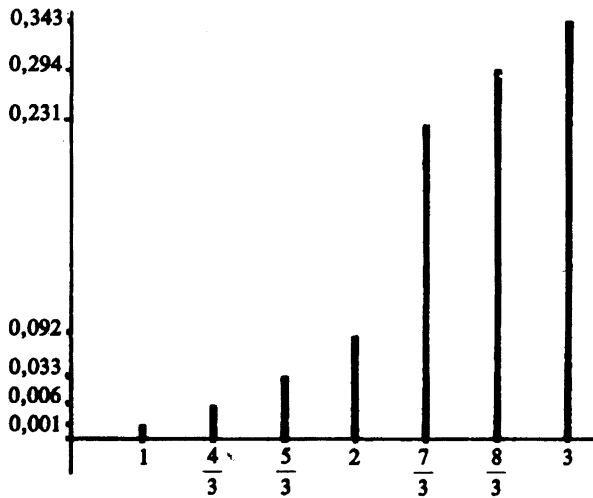


FIGURA 1.2

Distribución *en* el muestreo de la media de muestras aleatorias simples de tamaño 3.

En el campo de variación de la media muestral aparecen valores que se generan en el proceso de muestreo y que no están contemplados en el campo de la variable poblacional.

En la figura 1.2 representamos la función de cuantía muestral de la media, pudiéndose comprobar al compararla con la de la población y la de la muestra concreta de la figura 1.1, que se trata de una variable distinta aunque ligada a la primera.

La anterior distribución de probabilidad *en* el muestreo es una distribución de probabilidad como cualquier otra, y el cálculo de sus características se lleva a cabo mediante los procedimientos habituales. Calcularemos la esperanza y la varianza de la media muestral.

■ Esperanza

$$\begin{aligned}
 E(\bar{x}) &= \bar{x}_1 P(\bar{x} = \bar{x}_1) + \bar{x}_2 P(\bar{x} = \bar{x}_2) + \bar{x}_3 P(\bar{x} = \bar{x}_3) + \bar{x}_4 P(\bar{x} = \bar{x}_4) + \\
 &\quad + \bar{x}_5 P(\bar{x} = \bar{x}_5) + \bar{x}_6 P(\bar{x} = \bar{x}_6) + \bar{x}_7 P(\bar{x} = \bar{x}_7) = \\
 &= 1 \cdot 0,001 + \frac{4}{3} \cdot 0,006 + \frac{5}{3} \cdot 0,033 + 2 \cdot 0,092 + \\
 &\quad + \frac{7}{3} \cdot 0,231 + \frac{8}{3} \cdot 0,294 + 3 \cdot 0,343 = \\
 &= 2,6
 \end{aligned}$$

■ Varianza

$$\begin{aligned}
 E(\bar{x}^2) &= \bar{x}_1^2 P(\bar{x} = \bar{x}_1) + \bar{x}_2^2 P(\bar{x} = \bar{x}_2) + \bar{x}_3^2 P(\bar{x} = \bar{x}_3) + \bar{x}_4^2 P(\bar{x} = \bar{x}_4) + \\
 &\quad + \bar{x}_5^2 P(\bar{x} = \bar{x}_5) + \bar{x}_6^2 P(\bar{x} = \bar{x}_6) + \bar{x}_7^2 P(\bar{x} = \bar{x}_7) = \\
 &= 1^2 \cdot 0,001 + \left(\frac{4}{3}\right)^2 \cdot 0,006 + \left(\frac{5}{3}\right)^2 \cdot 0,033 + 2^2 \cdot 0,092 + \\
 &\quad + \left(\frac{7}{3}\right)^2 \cdot 0,231 + \left(\frac{8}{3}\right)^2 \cdot 0,294 + 3^2 \cdot 0,343 = \\
 &= 6,9067
 \end{aligned}$$

$$V(\bar{x}) = E(\bar{x}^2) - [E(\bar{x})]^2 = 6,9067 - 2,6^2 = 0,1467.$$

1.4 Características de las distribuciones en el muestreo

Estudiaremos como características de las distribuciones *en* el muestreo las esperanzas y varianzas de los momentos respecto al origen y a la media. En general, supondremos que las muestras han sido extraídas de poblaciones con media μ y varianza σ^2 .

1.4.1. MOMENTOS MUESTRALES CON RESPECTO AL ORIGEN

En una muestra aleatoria X de tamaño n (x_1, \dots, x_n), de una variable aleatoria ξ , se definen los momentos muestrales de orden r con respecto al origen como

$$a_r = \frac{1}{n} \sum_{i=1}^n x_i^r.$$

Calculamos la esperanza matemática de a_r , variable aleatoria por ser función de los elementos muestrales,

$$E(a_r) = E\left[\frac{1}{n} \sum_{i=1}^n x_i^r\right] = \frac{1}{n} \sum_{i=1}^n E(x_i^r) = \frac{1}{n} \sum_{i=1}^n E(\xi^r) = \frac{1}{n} n \alpha_r = \alpha_r.$$

El resultado dice que, *sea cual sea la distribución que tenga la variable aleatoria poblacional*, la esperanza matemática de cada momento muestral respecto al origen siempre es igual al correspondiente momento poblacional. Además, esto sucede aunque la muestra no sea aleatoria simple, es decir, aunque sus elementos no sean independientes.

Calculamos, a continuación, las varianzas de las distribuciones en el muestreo de estos momentos. Si la muestra aleatoria es simple, sus elementos son variables independientes y la varianza de su suma es igual a la suma de las varianzas¹⁰

$$\begin{aligned} V(a_r) &= V\left[\frac{1}{n} \sum_{i=1}^n x_i^r\right] = \frac{1}{n^2} \sum_{i=1}^n V(x_i^r) = \frac{1}{n^2} \sum_{i=1}^n V(\xi^r) = \\ &= \frac{1}{n^2} n V(\xi^r) = \frac{1}{n} E[\xi^r - E(\xi^r)]^2 = \frac{1}{n} E[\xi^r - \alpha_r]^2 = \\ &= \frac{1}{n} E[\xi^{2r} + \alpha_r^2 - 2\alpha_r \xi^r] = \frac{1}{n} (\alpha_{2r} - \alpha_r^2). \end{aligned}$$

Frente a lo apuntado para la esperanza matemática de los momentos, el resultado a que hemos llegado en la varianza exige que las variables aleatorias integrantes de la muestra sean independientes, es decir, la muestra tiene que ser aleatoria simple.

Media muestral

Centrándonos, en particular, en la media de la muestra, momento de orden uno, tenemos que

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = a_1$$

$$E(\bar{x}) = E(a_1) = \alpha_1 = \mu$$

$$V(\bar{x}) = V(a_1) = \frac{1}{n} (\alpha_2 - \alpha_1^2) = \frac{\sigma^2}{n}$$

¹⁰ MARTÍN PLIEGO y RUIZ-MAYA: *Fundamentos...* Capítulo 4.

En el ejemplo 4, el valor de la varianza poblacional es 0,44 y el de la varianza de la media muestral

$$V(\bar{x}) = \frac{0,44}{3} = 0,1467$$

que coincide con el obtenido a partir de la distribución en el muestreo de la media muestral.

1.4.2. MOMENTOS MUESTRALES CON RESPECTO A LA MEDIA

Definimos los momentos muestrales de orden r con respecto a la media muestral \bar{x} , m_r , como

$$m_r = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^r$$

demostrándose que

$$E(m_r) = \mu_r + O\left(\frac{1}{n}\right)$$

lo que revela el hecho que el valor esperado de los momentos muestrales respecto a la media no coincide con sus respectivos momentos poblacionales, diferenciándose en la cantidad $O\left(\frac{1}{n}\right)$, que será tanto menor cuanto mayor sea el tamaño de la muestra.

Varianza muestral

Como caso particular estudiaremos la varianza muestral cuya expresión es

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Para calcular su esperanza matemática hacemos, previamente, algunos cálculos sumando y restando la esperanza de la variable aleatoria poblacional ξ , μ , en el paréntesis del segundo miembro,

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x} + \mu - \mu)^2}{n} = \frac{1}{n} \sum_{i=1}^n [(x_i - \mu) - (\bar{x} - \mu)]^2.$$

Desarrollamos el cuadrado

$$\begin{aligned}
 s^2 &= \frac{1}{n} \sum_{i=1}^n [(x_i - \mu)^2 + (\bar{x} - \mu)^2 - 2(\bar{x} - \mu)(x_i - \mu)] = \\
 &= \frac{1}{n} \left[\sum_{i=1}^n (x_i - \mu)^2 + n(\bar{x} - \mu)^2 - 2(\bar{x} - \mu) \sum_{i=1}^n (x_i - \mu) \right] = \\
 &= \frac{1}{n} \left[\sum_{i=1}^n (x_i - \mu)^2 + n(\bar{x} - \mu)^2 - 2(\bar{x} - \mu)(n\bar{x} - n\mu) \right] = \\
 &= \frac{1}{n} \left[\sum_{i=1}^n (x_i - \mu)^2 - n(\bar{x} - \mu)^2 \right].
 \end{aligned}$$

Calculamos la esperanza matemática de la varianza muestral, s^2 , mediante la expresión anterior

$$\begin{aligned}
 E(s^2) &= \frac{1}{n} E \left[\sum_{i=1}^n (x_i - \mu)^2 - n(\bar{x} - \mu)^2 \right] = \\
 &= \frac{1}{n} \left[\sum_{i=1}^n E(x_i - \mu)^2 - n E(\bar{x} - \mu)^2 \right].
 \end{aligned}$$

En el segundo miembro aparecen primera dos esperanzas, la $E(x_i - \mu)^2$ coincide con la varianza poblacional, σ^2 , al ser la muestra aleatoria simple y la segunda $E(\bar{x} - \mu)^2$ es la varianza de la media muestral, $\frac{\sigma^2}{n}$, llegando por último a

$$E(s^2) = \frac{1}{n} \left[n\sigma^2 - n \frac{\sigma^2}{n} \right] = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n} \sigma^2.$$

La varianza de la varianza muestral, $V(s^2)$, de cálculo complicado, es

$$V(s^2) = \frac{\mu_4 - \mu_2^2}{n} - \frac{2(\mu_4 - 2\mu_2^2)}{n^2} + \frac{\mu_4 - 3\mu_2^2}{n^3}.$$

Adquiere especial importancia en la Inferencia Estadística la **cuasivarianza muestral**¹¹, s_1^2 , definida como

$$s_1^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}.$$

¹¹ MARTÍN PLIEGO: Capítulo 4.

La relación entre la cuasivarianza y la varianza es inmediata, pues

$$s_1^2 = \frac{n}{n-1} s^2.$$

La esperanza de la cuasivarianza es igual a σ^2

$$E(s_1^2) = E\left[\frac{n}{n-1} s^2\right] = \frac{n}{n-1} \cdot \frac{n-1}{n} \sigma^2 = \sigma^2$$

y la varianza¹²

$$\begin{aligned} V(s_1^2) &= V\left[\frac{n}{n-1} s^2\right] = \frac{n^2}{(n-1)^2} V(s^2) = \\ &= \frac{n^2}{(n-1)^2} \left[\frac{\mu_4 - \mu_2^2}{n} - \frac{2(\mu_4 - 2\mu_2^2)}{n^2} + \frac{\mu_4 - 3\mu_2^2}{n^3} \right] = \\ &= \frac{1}{(n-1)^2} \left[(n-2)\mu_4 - (n-4)\mu_2^2 + \frac{\mu_4 - 3\mu_2^2}{n} \right]. \end{aligned}$$

1.4.3. MOMENTOS MUESTRALES DE DISTRIBUCIONES BIDIMENSIONALES

Al generalizar lo expuesto respecto a los momentos vamos a centrarnos, únicamente, en el caso de poblaciones bidimensionales. Se parte de una muestra aleatoria de tamaño n $[(x_1; y_1), \dots, (x_n; y_n)]$ de la población $(\xi; \eta)$.

Las expresiones de los momentos mixtos muestrales son

$$a_{rs} = \frac{1}{n} \sum_{i=1}^n x_i^r y_i^s$$

$$m_{rs} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^r (y_i - \bar{y})^s$$

llegándose a que

$$E(a_{rs}) = \alpha_{rs}.$$

¹² En el caso de una población con distribución normal

$$V(s^2) = \frac{2(n-1)\sigma^4}{n^2} \quad \text{y} \quad V(s_1^2) = \frac{2\sigma^4}{n-1}$$

dado que $\mu_4 = 3\sigma^4$ y $\mu_2 = \sigma^2$.

De los momentos con respecto a la media, el más importante es el de orden (uno, uno), la **covarianza muestral**

$$m_{11} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

cuya esperanza es

$$E(m_{11}) = \mu_{11} - \frac{\mu_{11}}{n}.$$

De manera análoga a lo estudiado en los momentos de las distribuciones bidimensionales poblacionales pasamos en la muestra¹³ a los momentos de las distribuciones unidimensionales marginales. Si en la expresión general de los momentos respecto al origen hacemos $r = 1, s = 0$ y $r = 0, s = 1$, llegamos a que

$$a_{10} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}; \quad a_{01} = \frac{1}{n} \sum_{i=1}^n y_i = \bar{y}$$

son las medias muestrales de las dos distribuciones marginales. Si, ahora, particularizamos en las expresiones de los momentos respecto a las medias para $r = 2, s = 0$ y $r = 0, s = 2$, resultan las varianzas de las dos distribuciones marginales

$$m_{20} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = s_x^2; \quad m_{02} = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 = s_y^2.$$

El **coeficiente de correlación muestral**, r , es igual a

$$\begin{aligned} r &= \frac{m_{11}}{\sqrt{m_{20} m_{02}}} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}} = \\ &= \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sqrt{\left[\sum_{i=1}^n x_i^2 - n \bar{x}^2 \right] \left[\sum_{i=1}^n y_i^2 - n \bar{y}^2 \right]}}. \end{aligned}$$

El coeficiente de correlación muestral r , como el poblacional ρ , cumple la condición de tomar valores exclusivamente en el intervalo $[-1; 1]$. Cuando $r = 1$, todos los elementos muestrales (x_i, y_i) se encuentran sobre una recta con pendiente positiva; si $r = -1$, la recta tiene pendiente negativa¹⁴.

¹³ MARTÍN PLIEGO: capítulo 3.

¹⁴ MARTÍN PLIEGO y RUIZ-MAYA: *Fundamentos...* Capítulo 8.

1.5 Estadísticos ordenados

En ocasiones, de los posibles valores que puede tomar una variable aleatoria interesan algunos concretos: en las temperaturas, los valores mínimo y máximo diarios, en el análisis de fiabilidad, la vida máxima de una pieza, las dosis mínimas o máximas de un medicamento, cotizaciones extremas de Bolsa, etc. En todos estos ejemplos, se pretende llevar a cabo inferencias sobre aspectos muy particulares del campo de variación. Los estadísticos ordenados, definidos a continuación, resuelven estos problemas y otros análogos.

Consideremos una variable aleatoria continua ξ con función de densidad $f(x)$ y definida en el intervalo $-\infty < x < \infty$. De esta población se toman muestras aleatorias simples de tamaño n . Se ordenan los elementos muestrales concretos (x_1, \dots, x_n) de menor a mayor denominando u_1 al primero, u_2 al segundo, y u_n al último. El resultado de la ordenación

$$u_1 \leq u_2 \leq \dots \leq u_n$$

recibe el nombre de **muestra ordenada**. Cuando dos elementos de la muestra tienen el mismo valor el orden de su colocación en la muestra ordenada es indiferente.

Evidentemente, el primer elemento de la muestra ordenada, u_1 , es igual al menor valor de la muestra, mín (x_1, \dots, x_n) , y el último, u_n , al mayor, máx (x_1, \dots, x_n) . Las nuevas variables u_i reciben el nombre de **estadísticos ordenados de orden i** en muestras aleatorias simples de tamaño n .

Antes de seguir aclaramos las ideas anteriores mediante un ejemplo.

EJEMPLO 5

El resultado de una muestra aleatoria simple de tamaño cinco ha sido, según el orden en que los elementos fueron obtenidos,

$$x_1 = 0,52; \quad x_2 = 0,34; \quad x_3 = 0,96; \quad x_4 = 0,31; \quad x_5 = 0,71$$

ordenamos los cinco valores de menor a mayor

$$0,31; \quad 0,34; \quad 0,52; \quad 0,71; \quad 0,96$$

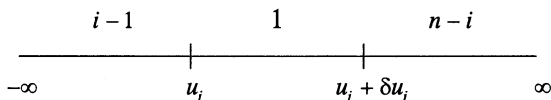
que en la muestra ordenada corresponden a

$$u_1 = 0,31; \quad u_2 = 0,34; \quad u_3 = 0,52; \quad u_4 = 0,71; \quad u_5 = 0,96.$$

1.5.1. DISTRIBUCIÓN DEL ESTADÍSTICO ORDENADO u_i

Si la población es continua, sabemos que $P(\xi = u_i)$ es igual a cero¹⁵, por lo que consideramos que en el intervalo $(u_i; u_i + \delta u_i]$ aparece solamente un valor, debiendo estar simultáneamente los $n - 1$ restantes fuera de él, de tal forma que $i - 1$ valores tienen que ser menores que u_i y los $n - i$ restantes mayores que $u_i + \delta u_i$, verificándose, por tanto, que $(i - 1) + (1) + (n - i) = n$.

Esquemáticamente los n elementos se reparten como sigue



y la probabilidad elemental del suceso $\{u_i < \xi \leq u_i + \delta u_i\}$ es:

$$k [P(-\infty < \xi \leq u_i)]^{i-1} \cdot P(u_i < \xi \leq u_i + \delta u_i) \cdot [P(u_i + \delta u_i < \xi < \infty)]^{n-i} \quad [1]$$

la constante k es el número de ordenaciones que pueden formarse con n elementos de los cuales $i - 1$ son menores que u_i y $n - i$ mayores que $u_i + \delta u_i$, es decir,

$$k = \frac{n!}{(i-1)! 1! (n-i)!} = \frac{n!}{(i-1)! (n-i)!}.$$

Calculemos las tres probabilidades anteriores:

■ $P(1 \text{ valor menor que } u_i) = P(-\infty < \xi \leq u_i) = \int_{-\infty}^{u_i} f(x) dx = F(u_i)$

$P(i - 1 \text{ valores menores que } u_i) = [P(-\infty < \xi \leq u_i)]^{i-1} = [F(u_i)]^{i-1}$

■ $P(1 \text{ valor en } (u_i; u_i + \delta u_i]) = P(u_i < \xi \leq u_i + \delta u_i) =$
 $= \int_{u_i}^{u_i + \delta u_i} f(x) dx = F(u_i + \delta u_i) - F(u_i).$

■ $P(1 \text{ valor mayor que } u_i + \delta u_i) = P(u_i + \delta u_i < \xi < \infty) =$
 $= \int_{u_i + \delta u_i}^{\infty} f(x) dx = 1 - \int_{-\infty}^{u_i + \delta u_i} f(x) dx =$
 $= 1 - F(u_i + \delta u_i)$

$P(n - i \text{ valores mayores que } u_i + \delta u_i) = [P(u_i + \delta u_i < \xi < \infty)]^{n-i} =$
 $= [1 - F(u_i + \delta u_i)]^{n-i}.$

¹⁵ MARTÍN PLIEGO y RUIZ-MAYA: *Fundamentos...* Capítulo 2.

Sustituyendo estas expresiones en [1], llegamos a

$$k [F(u_i)]^{i-1} [F(u_i + \delta u_i) - F(u_i)] [1 - F(u_i + \delta u_i)]^{n-i}.$$

El límite de la probabilidad anterior dividida por δu_i cuando $\delta u_i \rightarrow 0$ es:

$$\begin{aligned} \lim_{\delta u_i \rightarrow 0} \frac{P(u_i < \xi \leq u_i + \delta u_i)}{\delta u_i} &= \\ &= k \lim_{\delta u_i \rightarrow 0} \frac{[F(u_i)]^{i-1} [F(u_i + \delta u_i) - F(u_i)] [1 - F(u_i + \delta u_i)]^{n-i}}{\delta u_i} = \\ &= k [F(u_i)]^{i-1} \lim_{\delta u_i \rightarrow 0} \frac{[F(u_i + \delta u_i) - F(u_i)]}{\delta u_i} [1 - F(u_i)]^{n-i} \end{aligned}$$

donde

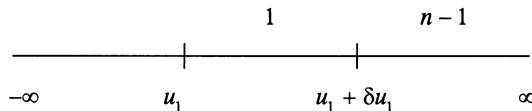
$$\lim_{\delta u_i \rightarrow 0} \frac{[F(u_i + \delta u_i) - F(u_i)]}{\delta u_i} = f(u_i)$$

siendo $f(u_i)$ la función de densidad de la población particularizada para u_i , resultando la densidad en el muestreo de la variable aleatoria u_i

$$g(u_i) = \frac{n!}{(i-1)! (n-i)!} [F(u_i)]^{i-1} f(u_i) [1 - F(u_i)]^{n-i}.$$

1.5.2. DISTRIBUCIÓN DEL MENOR VALOR DE LA MUESTRA, u_1

Hemos denominado al menor valor de la muestra como u_1 , esto es, el mínimo de los n valores muestrales (x_1, \dots, x_n) , $u_1 = \min(x_1, \dots, x_n)$. Esto lleva a que no hay a su izquierda ningún valor menor que u_1 , entre u_1 y $u_1 + \delta u_1$ hay uno y a la derecha de $u_1 + \delta u_1$ hay $n-1$ mayores que u_1 , gráficamente



La función de densidad en el muestreo de esta variable aleatoria, u_1 , la obtenemos particularizando en la distribución del i -ésimo valor para $i-1=0$ y $n-i=n-1$, es decir, haciendo $i=1$, llegando a la expresión

$$g(u_1) = n f(u_1) [1 - F(u_1)]^{n-1}.$$

EJEMPLO 3

Hallamos en la variable aleatoria con función de densidad $f(x) = e^{-x}$, $x \geq 0$, $P(u_1 \geq 0,45)$ en muestras aleatorias simples de tamaño 20.

La función de densidad de u_1 es

$$g(u_1) = n f(u_1) [1 - F(u_1)]^{n-1}$$

siendo

$$\begin{aligned} f(u_1) &= e^{-u_1} & u_1 &\geq 0 \\ F(u_1) &= 1 - e^{-u_1} & u_1 &\geq 0 \end{aligned}$$

luego

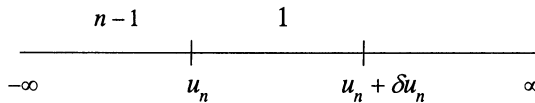
$$\begin{aligned} g(u_1) &= 20 e^{-u_1} [1 - (1 - e^{-u_1})]^{19} = \\ &= 20 e^{-20u_1}, \quad u_1 \geq 0 \end{aligned}$$

Por consiguiente,

$$P(u_1 \geq 0,45) = \int_{0,45}^{\infty} 20 e^{-20u_1} du_1 = 0,0001234.$$

1.5.3. DISTRIBUCIÓN DEL MAYOR VALOR DE LA MUESTRA, u_n

De forma análoga a como se ha obtenido la función de densidad en el muestreo del menor valor de la muestra, hallamos la del mayor valor, $u_n = \max(x_1, \dots, x_n)$. Al ser u_n el mayor valor de la muestra entre él y $u_n + \delta u_n$ hay un solo valor y los $n - 1$ restantes tienen que ser menores que u_n , como podemos apreciar en el gráfico



Particularizando en la función de densidad y de distribución del i -ésimo valor para $i - 1 = n - 1$, es decir, $n = i$ resulta

$$g(u_n) = n f(u_n) [F(u_n)]^{n-1}.$$

EJEMPLO 4

De una población con función de densidad $f(x) = \frac{3x^2}{\theta^3}$, para $0 \leq x \leq \theta$ se toman muestras aleatorias simples de tamaño n y se calcula $P(u_n \leq 0,01)$.

La función de distribución de esta variable aleatoria es $F(x) = \frac{x^3}{\theta^3}$ para $0 \leq x \leq \theta$.

El estadístico ordenado u_n es el mayor valor de la muestra y su función de densidad en el muestreo igual a

$$g(u_n) = \frac{n!}{(n-1)!} f(u_n) [F(u_n)]^{n-1} = n \cdot 3 \frac{u_n^2}{\theta^3} \left[\frac{u_n^3}{\theta^3} \right]^{n-1} = 3n \frac{u_n^{3n-1}}{\theta^{3n}} \quad \text{para } 0 \leq u_n \leq \theta$$

$$P(u_n \leq 0,01) = \int_0^{0,01} 3n \frac{u_n^{3n-1}}{\theta^{3n}} du_n = \frac{1}{\theta^{3n}} (0,01)^{3n}.$$

1.6 Muestreo en poblaciones normales

En páginas anteriores hemos obtenido ciertos momentos muestrales sin indicar nada respecto a su distribución concreta en el muestreo, pues se ignoraba cuál era la de la población de la que procedía la muestra. En las páginas que siguen estudiaremos este extremo para el caso particular de que la población sea normal.

De una variable aleatoria ξ con distribución $N(\mu; \sigma)$ se toman muestras aleatorias simples de tamaño n (x_1, \dots, x_n). Se consideran los estadísticos media y varianza, obteniéndose sus distribuciones en el muestreo. Asimismo, en el caso de una distribución normal bidimensional, se estudiarán los estadísticos muestrales, coeficiente de correlación y de regresión.

1.7 Varianza poblacional conocida

1.7.1. DISTRIBUCIÓN DE LA MEDIA MUESTRAL

La media muestral, como sabemos, tiene por expresión

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{1}{n} x_1 + \dots + \frac{1}{n} x_n.$$

La suma del tercer miembro de la igualdad indica que la media muestral es combinación lineal de variables aleatorias normales e independientes (x_1, \dots, x_n) , por ser la muestra aleatoria simple, implicando que la combinación lineal sigue una distribución normal¹⁶. Los dos parámetros de la distribución muestral han sido obtenidos páginas atrás: la esperanza matemática es la poblacional y la varianza la poblacional dividida por el tamaño muestral, por lo cual

$$\bar{x} \rightarrow N\left(\mu; \frac{\sigma}{\sqrt{n}}\right).$$

1.7.2. DISTRIBUCIÓN DE LA DIFERENCIA DE MEDIAS MUESTRALES

Si en vez de una población se consideran dos, $N(\mu_1; \sigma_1)$ y $N(\mu_2; \sigma_2)$, y de cada una de ellas se extrae una muestra aleatoria simple la primera de tamaño n (x_1, \dots, x_n) y la segunda de tamaño m (y_1, \dots, y_m) independiente de la primera, las medias muestrales son

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}; \quad \bar{y} = \frac{\sum_{i=1}^m y_i}{m}$$

y sus distribuciones muestrales

$$\bar{x} \rightarrow N\left(\mu_1; \frac{\sigma_1}{\sqrt{n}}\right); \quad \bar{y} \rightarrow N\left(\mu_2; \frac{\sigma_2}{\sqrt{m}}\right).$$

Se considera como estadístico de interés la diferencia de las medias muestrales $\bar{x} - \bar{y}$, combinación lineal de $m + n$ variables aleatorias normales e independientes, por lo cual su distribución será normal, con media

$$E(\bar{x} - \bar{y}) = E(\bar{x}) - E(\bar{y}) = \mu_1 - \mu_2$$

y varianza

$$V(\bar{x} - \bar{y}) = V(\bar{x}) + V(\bar{y}) = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}$$

es decir,

$$\bar{x} - \bar{y} \rightarrow N\left[\mu_1 - \mu_2; \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}\right].$$

¹⁶ MARTÍN PLIEGO y RUIZ-MAYA: *Fundamentos...* Capítulo 7.

1.8 Distribución conjunta de la media y la varianza muestrales o Lema de Fisher-Cochran

Presentamos, a continuación sin demostración¹⁷, dos importantes propiedades de la distribución conjunta en el muestreo de la media y varianzas muestrales.

- Las variables aleatorias media y varianza muestrales (\bar{x} y s^2) son estadísticamente independientes.
- El estadístico $\frac{ns^2}{\sigma^2}$ se distribuye como una χ^2 con $(n-1)$ grados de libertad.

Podemos interpretar este resultado de la siguiente manera: la variable $\frac{ns^2}{\sigma^2}$ es igual a la suma de los cuadrados de $n-1$ variables aleatorias $N(0;1)$ e independientes.

1.9 Varianza poblacional desconocida

1.9.1. DISTRIBUCIÓN DE LA MEDIA MUESTRAL

Cuando es conocida la varianza poblacional, la distribución de la media muestral es $N\left(\mu; \frac{\sigma}{\sqrt{n}}\right)$, sin embargo, cuando la varianza poblacional es desconocida no sirve de nada práctico tal conocimiento, puesto que no se pueden calcular probabilidades de sucesos correspondientes a la media muestral. Para obtener la distribución muestral desconociendo σ^2 , es preciso recurrir a una distribución no dependiente de la varianza σ^2 , y esta distribución es la t de Student.

¹⁷ RUIZ-MAYA y MARTÍN PLIEGO: *Estadística II: Inferencia*. Capítulo 2.

Partiendo de la distribución de la media muestral y siendo la variable aleatoria ξ $N(0; 1)$

$$\bar{x} \rightarrow N\left(\mu; \frac{\sigma}{\sqrt{n}}\right); \quad \bar{x} = \mu + \frac{\sigma}{\sqrt{n}} \xi; \quad \sigma\xi = \sqrt{n} (\bar{x} - \mu) \rightarrow N(0; \sigma).$$

La variable aleatoria ns^2 es igual a la suma de cuadrados de $(n - 1)$ variables $N(0; \sigma)$ e independientes, y_i , y puede ser formulada una nueva variable de la siguiente manera

$$\sqrt{\frac{ns^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}.$$

Esta expresión es independiente de la media muestral, \bar{x} , como se expuso en el apartado 1.8 y resulta ser el denominador de la distribución t con $(n - 1)$ grados de libertad¹⁸ y junto con que la variable $\sigma\xi$ se distribuye $N(0; \sigma)$, conduce a una distribución t de Student con $(n - 1)$ grados de libertad

$$\frac{\sqrt{n} (\bar{x} - \mu)}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}} \rightarrow t(n-1).$$

1.9.2. DISTRIBUCIÓN DE LA DIFERENCIA DE MEDIAS MUESTRALES

Obtenemos la distribución en el muestreo de la diferencia de medias muestrales provenientes de la misma distribución $N(\mu; \sigma)$. Se toman dos muestras aleatorias simples, independientes entre sí, la primera (x_1, \dots, x_n) de tamaño n y la segunda (y_1, \dots, y_m) de tamaño m .

Las respectivas medias y varianzas muestrales son

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}; \quad \bar{y} = \frac{\sum_{i=1}^m y_i}{m}$$

$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}; \quad s_y^2 = \frac{\sum_{i=1}^m (y_i - \bar{y})^2}{m}.$$

¹⁸ MARTÍN PLIEGO y RUIZ-MAYA: *Fundamentos...* Capítulo 7.

En primer lugar, se establece por la generalización del lema de Fisher-Cochran que la variable aleatoria $ns_x^2 + ms_y^2$ es igual a la suma de cuadrados de $(n + m - 2)$ variables aleatorias v_i , $N(0; \sigma)$ e independientes

$$ns_x^2 + ms_y^2 = \sum_{i=1}^{n+m-2} v_i^2$$

con lo cual

$$\frac{ns_x^2 + ms_y^2}{\sigma^2}$$

sigue una distribución χ^2 con $n + m - 2$ grados de libertad.

Asimismo, la variable $ns_x^2 + ms_y^2$ es independiente de \bar{x} y de \bar{y} y, por consiguiente, de su diferencia $\bar{x} - \bar{y}$.

Para justificar la distribución en el muestreo de la diferencia de las dos medias, $\bar{x} - \bar{y}$, se procede como en el caso de una sola media: llegar a una distribución que no dependa de la varianza poblacional σ^2 .

La variable $\bar{x} - \bar{y}$ se distribuye

$$\bar{x} - \bar{y} \rightarrow N\left(0; \sigma \sqrt{\frac{n+m}{nm}}\right),$$

y siendo la variable ξ $N(0; 1)$ se obtiene

$$\bar{x} - \bar{y} = \sigma \sqrt{\frac{n+m}{nm}} \xi$$

con lo que

$$\sqrt{\frac{nm}{n+m}} (\bar{x} - \bar{y}) = \sigma \xi$$

se distribuye $N(0; \sigma)$.

Por otra parte, el estadístico

$$\sqrt{\frac{ns_x^2 + ms_y^2}{n+m-2}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^m (y_i - \bar{y})^2}{n+m-2}}$$

independiente de la variable $\bar{x} - \bar{y}$, es el denominador de una distribución t de Student con $(n + m - 2)$ grados de libertad, por lo cual, la variable

$$\frac{\sqrt{\frac{nm}{n+m}} (\bar{x} - \bar{y})}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^m (y_i - \bar{y})^2}{n+m-2}}}$$

se distribuye como $t(n + m - 2)$.

Si en vez de una única población, de la cual proceden las dos muestras, o de dos poblaciones con la misma distribución, partimos de dos poblaciones con medias distintas: $N(\mu_1; \sigma)$ y $N(\mu_2; \sigma)$, la distribución en el muestreo de la diferencia de medias muestrales es

$$\frac{\sqrt{\frac{nm}{n+m}} [(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)]}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^m (y_i - \bar{y})^2}{n+m-2}}}$$

1.10 Distribución en el muestreo del coeficiente de correlación lineal

- Si ρ es el coeficiente de correlación lineal poblacional en una distribución normal bidimensional siendo $\rho \neq 0$ se tiene que, cuando $n \rightarrow \infty$, la variable aleatoria¹⁹

$$Z = \frac{1}{2} \ln \frac{1+r}{1-r}$$

se distribuye asintóticamente

$$N\left(\frac{1}{2} \ln \frac{1+\rho}{1-\rho}; \sqrt{\frac{1}{n-3}}\right)$$

¹⁹ Esta variable aleatoria Z recibe el nombre de **transformación de Fisher**.

- Si se supone que en la población $\rho = 0$, el estadístico coeficiente de correlación lineal muestral r tiene un comportamiento probabilístico que puede derivarse de la distribución

$$t(n-2) = \frac{r}{\sqrt{1-r^2}} \sqrt{n-2},$$

es decir, la anterior transformación de r sigue una ley t de Student con $n-2$ grados de libertad.

1.11 Distribución de los coeficientes de regresión lineal muestrales

El coeficiente de regresión muestral de ξ sobre η ²⁰ es igual a

$$b = r \frac{s_y}{s_x}.$$

La esperanza del estadístico b es

$$E(b) = \beta,$$

siendo β el coeficiente de regresión de ξ sobre η en la población y la variable

$$t = \frac{s_x \sqrt{n-2}}{s_y \sqrt{1-r^2}} (b - \beta)$$

se distribuye como una t de Student con $n-2$ grados de libertad.

La distribución en el muestreo del coeficiente de regresión muestral b' , de ξ sobre η , se obtiene de manera análoga, llegándose a que la variable

$$t = \frac{s_y \sqrt{n-2}}{s_x \sqrt{1-r^2}} (b' - \beta')$$

se distribuye como $t(n-2)$.

²⁰ MARTÍN PLIEGO y RUIZ-MAYA: *Fundamentos...* Capítulo 8.

Ejercicios resueltos

EJERCICIO 1.1 Una variable aleatoria ξ se distribuye uniformemente en el intervalo $0 \leq x \leq 1$. La función de densidad es igual a $f(x) = 1$. Se toman muestras aleatorias simples de tamaño 2. Hállese la función de distribución en el muestreo de la media muestral

$$\bar{x} = \frac{x_1 + x_2}{2}.$$

SOLUCIÓN. Tenemos que

$$G(\bar{x}) = P\left(\frac{x_1 + x_2}{2} \leq \bar{x}\right) = P(x_2 \leq 2\bar{x} - x_1).$$

Para hallar esta probabilidad hemos de tener en cuenta el campo de variación de \bar{x} y cómo se desarrolla en él la recta $x_2 = 2\bar{x} - x_1$, haciéndolo de dos formas distintas, según que $2\bar{x}$ sea menor o mayor que la unidad (o lo que es igual, $\bar{x} \leq \frac{1}{2}$ o $\bar{x} > \frac{1}{2}$).

- Si $\bar{x} \leq \frac{1}{2}$

$$G(\bar{x}) = P(x_2 \leq 2\bar{x} - x_1) = \int_0^{2\bar{x}} \int_0^{2\bar{x}-x_1} dx_1 dx_2 = 2\bar{x}^2.$$

- Si $\bar{x} > \frac{1}{2}$

$$\begin{aligned} G(\bar{x}) &= P(x_2 \leq 2\bar{x} - x_1) = \int_0^{2\bar{x}-1} \int_0^1 dx_1 dx_2 + \int_{2\bar{x}-1}^1 \int_0^{2\bar{x}-x_1} dx_1 dx_2 = \\ &= -2\bar{x}^2 + 4\bar{x} - 1. \end{aligned}$$

La función de densidad en el muestreo de la media muestral, $g(\bar{x}) = G'(\bar{x})$, es igual a

$$g(\bar{x}) = \begin{cases} 4\bar{x} & \text{para } 0 \leq \bar{x} \leq \frac{1}{2} \\ 4(1 - \bar{x}) & \text{para } \frac{1}{2} < \bar{x} \leq 1 \end{cases}$$

En el gráfico siguiente, exponemos la función de densidad de la población, $U(0; 1)$, y la de la media muestral

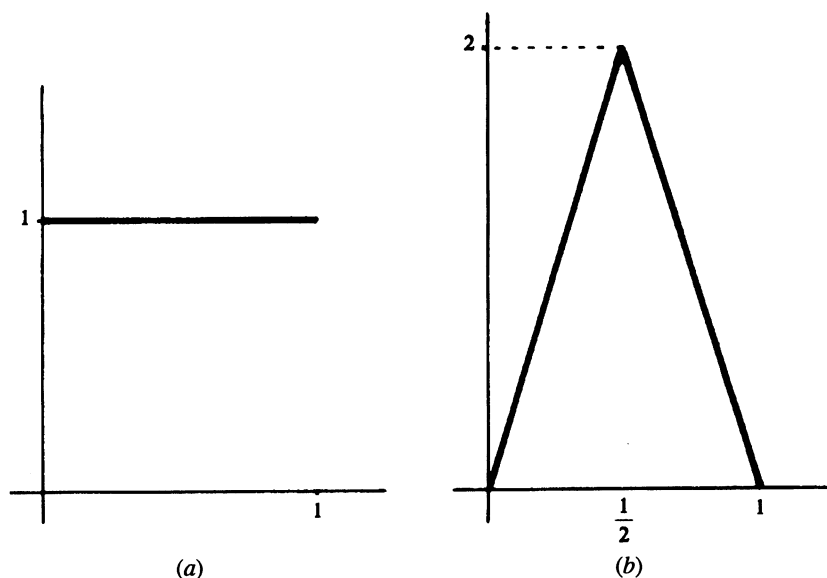


FIGURA 1.3

(a) Función de densidad poblacional.

(b) Función de densidad en el muestreo de las medias de muestras de tamaño 2.

EJERCICIO 1.2 En una bolsa hay 100 bolas, de igual tamaño, marcadas con uno de tres números distintos: cero, uno y dos, con las siguientes frecuencias: 30 ceros, 50 unos y 20 doses. Se toman muestras aleatorias de tamaño dos, primero con reemplazamiento y después sin él.

Determinense las esperanzas y varianzas de la media muestral en los dos casos.

SOLUCIÓN. La media de la población es

$$\mu = E(\xi) = 0 \cdot 0,3 + 1 \cdot 0,5 + 2 \cdot 0,2 = 0,9;$$

y el momento de orden dos

$$E(\xi^2) = 0^2 \cdot 0,3 + 1^2 \cdot 0,5 + 2^2 \cdot 0,2 = 1,3$$

y la varianza poblacional

$$V(\xi) = E(\xi^2) - [E(\xi)]^2 = 1,3 - 0,9^2 = 0,49.$$

■ *Muestreo aleatorio con reemplazamiento* (muestra aleatoria simple)

\bar{x}_i	$P(\bar{x} = \bar{x}_i)$
0,0	0,09
0,5	0,30
1,0	0,37
1,5	0,20
2,0	0,04

$$E(\bar{x}) = 0 \cdot 0,09 + 0,5 \cdot 0,30 + 1 \cdot 0,37 + 1,5 \cdot 0,20 + 2 \cdot 0,04 = 0,9$$

$$E(\bar{x}^2) = 0^2 \cdot 0,09 + 0,5^2 \cdot 0,30 + 1^2 \cdot 0,37 + 1,5^2 \cdot 0,20 + 2^2 \cdot 0,04 = 1,055$$

$$V(\bar{x}) = 1,055 - 0,9^2 = 0,245$$

Vimos que $V(\bar{x}) = \frac{\sigma^2}{n}$. En el ejemplo, el tamaño de la muestra es dos, por lo cual,

$$V(\bar{x}) = \frac{\sigma^2}{2} = \frac{0,49}{2} = 0,245,$$

cantidad que coincide con la calculada.

■ *Muestreo aleatorio sin reemplazamiento* (muestra aleatoria)

\bar{x}_i	$P(\bar{x} = \bar{x}_i)$
0,0	0,088
0,5	0,303
1,0	0,369
1,5	0,202
2,0	0,038

$$E(\bar{x}) = 0 \cdot 0,088 + 0,5 \cdot 0,303 + 1 \cdot 0,369 + 1,5 \cdot 0,202 + 2 \cdot 0,038 = 0,9.$$

La esperanza matemática de la media muestral coincide con la poblacional, cosa que no sucede con la varianza de la media muestral, como apreciamos a continuación

$$E(\bar{x}^2) = 0^2 \cdot 0,088 + 0,5^2 \cdot 0,303 + 1^2 \cdot 0,369 + 1,5^2 \cdot 0,202 + 2^2 \cdot 0,038 = 1,0513$$

$$V(\bar{x}) = 1,0513 - 0,9^2 = 0,2413$$

pues los elementos muestrales no son independientes.

EJERCICIO 1.3 Como continuación del ejercicio 1.1 comprobamos que $E(\bar{x}) = \mu$. (La esperanza matemática de la distribución $U(0; 1)$ ¹ es igual a 0,5).

SOLUCIÓN. La esperanza matemática de la media muestral es

$$E(\bar{x}) = \int_0^{\frac{1}{2}} \bar{x} 4\bar{x} d\bar{x} + \int_{\frac{1}{2}}^1 \bar{x} 4(1-\bar{x}) d\bar{x} = 0,5$$

valor que coincide con el de la esperanza matemática poblacional.

EJERCICIO 1.4 Tenemos una variable aleatoria que toma los valores 1, 2 y 3 con probabilidades 0,1, 0,2 y 0,7, respectivamente.

Calcúlese la esperanza de la varianza muestral.

SOLUCIÓN. En muestras aleatorias simples de tamaño tres, comprobamos que

$$E(s^2) = \frac{(n-1)\sigma^2}{n},$$

en este caso,

$$E(s^2) = \frac{2\sigma^2}{3}.$$

En la tabla siguiente hemos calculado, para cada clase de muestra, su varianza, figurando el número de muestras y las probabilidades de cada clase de muestra.

Muestras tipo	Número de muestras de cada tipo	Varianzas muestrales	Probabilidades de obtención de las muestras tipo
(1, 1, 1)	1	0	$0,1^3 = 0,001$
(1, 1, 2)	3	$2/9$	$0,1^2 \cdot 0,2 = 0,002$
(1, 1, 3)	3	$8/9$	$0,1^2 \cdot 0,7 = 0,007$
(1, 2, 3)	6	$2/3$	$0,1 \cdot 0,2 \cdot 0,7 = 0,014$
(2, 2, 1)	3	$2/9$	$0,1 \cdot 0,2^2 = 0,004$
(2, 2, 2)	1	0	$0,2^3 = 0,008$
(2, 2, 3)	3	$2/9$	$0,2^2 \cdot 0,7 = 0,028$
(3, 3, 1)	3	$8/9$	$0,1 \cdot 0,7^2 = 0,049$
(3, 3, 2)	3	$2/9$	$0,2 \cdot 0,7^2 = 0,098$
(3, 3, 3)	1	0	$0,7^3 = 0,343$

¹ MARTÍN PLIEGO y RUIZ-MAYA: *Fundamentos...* Capítulo 9.

Agrupando las varianzas que presenta el mismo valor, tenemos la tabla final siguiente:

s_i^2	$P(s^2 = s_i^2)$
0	0,352
2/9	0,396
2/3	0,084
8/9	0,168

La esperanza matemática de la varianza muestral es

$$E(s^2) = 0 \cdot 0,352 + \frac{2}{9} \cdot 0,396 + \frac{2}{3} \cdot 0,084 + \frac{8}{9} \cdot 0,168 = 0,2933.$$

Por otra parte, la varianza poblacional es igual a 0,44 y

$$E(s^2) = \frac{2 \cdot 0,44}{3} = 0,2933$$

que coincide con el resultado anterior.

EJERCICIO 1.5 En el ejercicio 1.1 se ha calculado la función de densidad en el muestreo de la media de muestras aleatorias simples de tamaño 2. Verifíquese que, si la varianza poblacional es $\frac{1}{12}$, la de la media muestral es $\frac{1}{24}$.

SOLUCIÓN. La función de densidad en el muestreo de la media es

$$g(\bar{x}) = \begin{cases} 4\bar{x} & \text{para } 0 \leq \bar{x} \leq \frac{1}{2} \\ 4(1 - \bar{x}) & \text{para } \frac{1}{2} < \bar{x} \leq 1 \end{cases}$$

Sabemos que

$$V(\bar{x}) = E[\bar{x} - E(\bar{x})]^2 = E(\bar{x}^2) - E(\bar{x})^2$$

y como

$$E(\bar{x}^2) = \int_0^{\frac{1}{2}} \bar{x}^2 4\bar{x} d\bar{x} + \int_{\frac{1}{2}}^1 \bar{x}^2 4(1 - \bar{x}) d\bar{x} = \frac{1}{16} + \frac{1}{6} + \frac{1}{16} = \frac{7}{24}$$

entonces

$$V(\bar{x}) = E(\bar{x}^2) - E(\bar{x})^2 = \frac{7}{24} - \frac{1}{4} = \frac{1}{24}.$$

Se ha demostrado que

$$V(\bar{x}) = \frac{\sigma^2}{n} = \frac{\sigma^2}{2} = \frac{\frac{1}{12}}{2} = \frac{1}{24}.$$

EJERCICIO 1.6 Si una variable aleatoria se distribuye $U(0; 1)$ su función de densidad es igual a $f(x) = 1$, cuyo campo de variación es $0 \leq x \leq 1$, y la función de distribución $F(x) = x$. Calcúlese $P(u_2 \leq 0,1)$ si el tamaño muestral es cinco.

SOLUCIÓN. Si el tamaño muestral es cinco, la función de densidad del segundo estadístico ordenado, u_2 , es para $i - 1 = 1$ y $n - i = 5 - 2 = 3$

$$g(u_2) = \frac{5!}{1! 3!} F(u_2) f(u_2) [1 - F(u_2)]^3$$

$$F(u_2) = u_2; \quad f(u_2) = 1$$

resultando

$$g(u_2) = 20u_2 [1 - u_2]^3$$

$$P(u_2 \leq 0,1) = \int_0^{0,1} 20u_2(1 - u_2)^3 du_2 = 0,0815$$

EJERCICIO 1.7 Dada una distribución $N(\mu; \sigma)$ y en muestras aleatorias simples de tamaño n , hállese la distribución en el muestreo de la media muestral mediante la función característica.

SOLUCIÓN. La función característica de la media muestral es

$$\begin{aligned} \varphi_{\bar{x}}(t) &= E(e^{it\bar{x}}) = E \left[e^{i t \frac{\sum_{j=1}^n x_j}{n}} \right] = \\ &= E \left[\prod_{j=1}^n e^{i \frac{t}{n} x_j} \right] = \prod_{j=1}^n \varphi_j \left(\frac{t}{n} \right) = \left[\varphi \left(\frac{t}{n} \right) \right]^n. \end{aligned}$$

donde $\varphi \left(\frac{t}{n} \right)$ es la función característica de la distribución $N(\mu; \sigma)$ particularizada para $\frac{t}{n}$, por lo cual,

$$\varphi_{\bar{x}}(t) = \left[e^{i \frac{t}{n} \mu - \frac{1}{2} \frac{t^2}{n^2} \sigma^2} \right]^n = e^{it\mu - \frac{1}{2} t^2 \frac{\sigma^2}{n}}$$

función característica que corresponde a la distribución normal de media μ y varianza $\frac{\sigma^2}{n}$,

$$N\left(\mu; \frac{\sigma}{\sqrt{n}}\right).$$

EJERCICIO 1.8 De una población $N(10; 4)$ se toman muestras aleatorias simples de tamaño 100, calcúlese la probabilidad que la media muestral sea mayor que 10,1.

SOLUCIÓN. La media muestral se distribuye

$$N\left(10; \frac{4}{\sqrt{100}}\right) = N(10; 0,4)$$

$$P(\bar{x} \geq 10,1) = P(10 + 0,4\xi \geq 10,1) = P(\xi \geq 0,25) = 0,4013$$

siendo $\xi \sim N(0; 1)$.

EJERCICIO 1.9 Se parte de dos poblaciones $N(\mu_1; \sigma_1)$ y $N(\mu_2; \sigma_2)$, y se extrae de cada una de ellas una muestra aleatoria simple, la primera de tamaño n (x_1, \dots, x_n) y la segunda de tamaño m (y_1, \dots, y_m), ambas independientes entre sí. Obténgase la distribución en el muestreo de la diferencia de medias mediante la función característica.

SOLUCIÓN. La función característica de la variable $\bar{x} - \bar{y}$ es

$$\begin{aligned} \varphi_{\bar{x}-\bar{y}}(t) &= E[e^{it(\bar{x}-\bar{y})}] = E[e^{it\bar{x}} e^{i(-t)\bar{y}}] = E[e^{it\bar{x}}] E[e^{i(-t)\bar{y}}] = \varphi_{\bar{x}}(t) \varphi_{\bar{y}}(-t) = \\ &= e^{\mu_1 it - \frac{1}{2} \frac{t^2}{n} \sigma_1^2} e^{\mu_2 i(-t) - \frac{1}{2} \frac{t^2}{m} \sigma_2^2} = e^{(\mu_1 - \mu_2) it - \frac{t^2}{2} \left(\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m} \right)} \end{aligned}$$

la última expresión es, precisamente, la función característica de una distribución normal de media $\mu_1 - \mu_2$ y varianza $\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}$.

EJERCICIO 1.10 De la distribución $N(5; 2)$ se toma una muestra aleatoria simple de tamaño 20, con media \bar{x} y de la distribución $N(-4; 10)$ otra muestra

aleatoria simple de tamaño 25 e independiente de la primera, con media \bar{y} . Calcúlese $P(\bar{x} - \bar{y} \leq 15)$.

SOLUCIÓN. La diferencia de las medias se distribuye

$$N\left(9; \sqrt{\frac{4}{20} + \frac{100}{25}}\right) = N(9; 2,0494)$$

y expresada la diferencia de las medias en función de la variable ξ , $N(0; 1)$, se tiene que $\bar{x} - \bar{y} = 9 + 2,0494\xi$, y la probabilidad pedida resulta

$$P(\bar{x} - \bar{y} \leq 15) = P(9 + 2,0494\xi \leq 15) = P(\xi \leq 2,9277) = 0,99827.$$

EJERCICIO 1.11 Como aplicación de la independencia de la media y la varianza muestrales calculamos la probabilidad del suceso conjunto $\{\bar{x} \geq 0,9; s^2 \geq 0,6\}$ donde \bar{x} y s^2 proceden de muestras aleatorias simples de tamaño cinco de una distribución $N(1; 0,5)$.

SOLUCIÓN. Por la propiedad de independencia (lema de Fisher-Cochran) tenemos

$$P(\bar{x} \geq 0,9; s^2 \geq 0,6) = P(\bar{x} \geq 0,9) P(s^2 \geq 0,6)$$

\bar{x} se distribuye $N(1; 0,2236)$ y la variable $\frac{5s^2}{0,5^2}$ como $\chi^2(4)$.

$$P(\bar{x} \geq 0,9) = P(1 + 0,2236\xi \geq 0,9) = P(\xi \geq -0,4472) = 0,6726$$

$$P(s^2 \geq 0,6) = P\left(\frac{5s^2}{0,25} \geq \frac{5 \cdot 0,6}{0,25}\right) = P(\chi^2(4) \geq 12) = 0,019.$$

La probabilidad pedida es igual a

$$P(\bar{x} \geq 0,9; s^2 \geq 0,6) = 0,6726 \cdot 0,019 = 0,0128.$$

EJERCICIO 1.12 De una variable aleatoria $N(-1; \sigma)$ se extrae una muestra aleatoria simple de tamaño 10, cuyo resultado es

1,03	-1,79	1,45	-2,54	0,37
-0,60	0,53	0,28	-2,21	-2,66

Calcúlese $P(\bar{x} \geq -1,2)$.

SOLUCIÓN. Para obtener la probabilidad anterior se construye, paso a paso, la necesaria distribución t . De la muestra se obtiene la media y la suma de cuadrados de las desviaciones con respecto a la media

$$\bar{x} = -0,614; \quad \sum_{i=1}^{10} (x_i - \bar{x})^2 = 21,865$$

$$\begin{aligned} P(\bar{x} \geq -1,2) &= P(\bar{x} - (-1) \geq -1,2 + 1) = P(\bar{x} + 1 \geq -0,2) = \\ &= P\left[\sqrt{10} (\bar{x} + 1) \geq -0,2\sqrt{10}\right] = P\left[\frac{\sqrt{10} (\bar{x} + 1)}{\sqrt{\frac{21,865}{9}}} \geq \frac{-0,2\sqrt{10}}{\sqrt{\frac{21,865}{9}}}\right] = \\ &= P(t(9) \geq -0,4058) = 0,6492. \end{aligned}$$

EJERCICIO 1.13 De una distribución $N(100; 2)$ se toman dos muestras aleatorias simples, independientes entre sí, de tamaños 4 y 5

Muestra 1	Muestra 2
98,0	97,8
103,4	101,3
100,5	97,9
99,7	100,7
	100,3

Calcúlese la probabilidad del suceso $\{\bar{x} - \bar{y} \geq 2\}$.

SOLUCIÓN.

$$\text{En la muestra 1 : } \bar{x} = 100,4; \quad \sum_{i=1}^4 (x_i - \bar{x})^2 = 15,26$$

$$\text{En la muestra 2 : } \bar{y} = 99,6 \quad \sum_{i=1}^5 (y_i - \bar{y})^2 = 10,72$$

Siguiendo de manera análoga los pasos del ejemplo anterior

$$\begin{aligned} P(\bar{x} - \bar{y} \geq 2) &= P\left[\frac{\sqrt{\frac{4 \cdot 5}{4 + 5}} (\bar{x} - \bar{y})}{\sqrt{\frac{15,26 + 10,72}{4 + 5 - 2}}} \geq \frac{\sqrt{\frac{4 \cdot 5}{4 + 5}} \cdot 2}{\sqrt{\frac{15,26 + 10,72}{4 + 5 - 2}}}\right] = \\ &= P[t(7) \geq 1,5476] = 0,0862. \end{aligned}$$

EJERCICIO 1.14 En una población normal bidimensional, el coeficiente de correlación es $\rho = 0,35$. Se toma una muestra aleatoria simple de tamaño 100. Calcúlese $P(r \geq 0,75)$.

SOLUCIÓN. La variable aleatoria transformación de Fisher

$$Z = \frac{1}{2} \ln \frac{1+r}{1-r}$$

se distribuye, en este caso concreto, aproximadamente

$$Z \xrightarrow{d} N\left(\frac{1}{2} \ln \frac{1+0,35}{1-0,35}; \sqrt{\frac{1}{97}}\right) = N(0,3654; 0,1015)$$

$$P(Z \geq a) = P\left(\frac{1}{2} \ln \frac{1+r}{1-r} \geq a\right) = P\left(r \geq \frac{e^{2a} - 1}{e^{2a} + 1}\right) = P(r \geq 0,75)$$

$$\frac{e^{2a} - 1}{e^{2a} + 1} = 0,75; \quad a = 0,973$$

$$P(r \geq 0,75) = P(Z \geq 0,973) = P(0,3654 + 0,1015\xi \geq 0,973) = P(\xi \geq 5,9862) \approx 0.$$

EJERCICIO 1.15 De una distribución normal bidimensional se toma una muestra aleatoria simple de tamaño 6, cuyos valores son

x_i	y_i
1,5	-3,08
-1,1	-5,03
0,3	-3,99
-0,5	-4,56
-0,2	-4,34
0,9	-3,51

Calcúlese la probabilidad del suceso $\{|b - \beta| \leq 0,05\}$

SOLUCIÓN. Para obtener la probabilidad debemos transformar el suceso $\{|b - \beta| \leq 0,05\}$ de manera que lleguemos a la distribución $t(4)$ indicada en el apartado 1.9.4.

Para ello necesitamos hallar las varianzas y coeficiente de correlación muestrales

$$s_x^2 = \frac{\sum_{i=1}^6 x_i^2}{6} - \left(\frac{\sum_{i=1}^6 x_i}{6} \right)^2 = \frac{4,65}{6} - \left(\frac{0,90}{6} \right)^2 = 0,7525; \quad s_x = 0,8675$$

$$s_y^2 = \frac{\sum_{i=1}^6 y_i^2}{6} - \left(\frac{\sum_{i=1}^6 y_i}{6} \right)^2 = \frac{102,6567}{6} - \left(\frac{-24,51}{6} \right)^2 = 0,4222; \quad s_y = 0,6498$$

$$s_{xy} = m_{11} = \frac{\sum_{i=1}^6 x_i y_i}{6} - \bar{x}\bar{y} = \frac{-0,295}{6} - \frac{0,90}{6} \frac{-24,51}{6} = 0,5636; \quad r = 0,9998$$

$$P(|b - \beta| \leq 0,05) = P(-0,05 \leq b - \beta \leq 0,05) =$$

$$= P \left[-\frac{0,8675 \sqrt{4}}{0,6498 \sqrt{1 - 0,9996}} \leq \frac{0,8675 \sqrt{4}}{0,6498 \sqrt{1 - 0,9996}} (b - \beta) \leq \right.$$

$$\left. \leq \frac{0,8675 \sqrt{4}}{0,6498 \sqrt{1 - 0,9996}} \right] =$$

$$= P(-133,5 \leq t(4) \leq 133,5) = 1.$$

CAPÍTULO 2

Suficiencia e información

2.1 Parámetros desconocidos

El objeto de nuestro interés (un fenómeno de la naturaleza, la conducta colectiva de una sociedad, el comportamiento de un individuo, el desarrollo de un juego de azar) es lo que modelizamos a través de una variable aleatoria y denominamos, de manera genérica, **población**.

Para establecer sus normas de funcionamiento y predecir su comportamiento, suponemos que el fenómeno aleatorio poblacional se puede describir a través de un determinado **modelo probabilístico** preestablecido. En esta fase primaria, toda la información que tengamos es fundamental: los preconocimientos que se tengan sobre esta población, nuestras propias reflexiones y creencias razonables serán determinantes para iniciar el proceso inferencial estadístico con ciertas esperanzas de éxito, pues todo el trabajo preparatorio se concretará en la asignación de un modelo probabilístico materializado en una cierta distribución de probabilidad. Esta asignación es conocida como **distribución subyacente de la población**.

Si efectuamos un repaso a los modelos teóricos de probabilidad, encontramos una característica común: su completa especificación depende siempre de la asigna-

ción de valores concretos a los parámetros que los identifican unívocamente. En efecto, en el modelo binomial se debe asignar un valor al parámetro p , la distribución de Poisson estará perfectamente especificada si se conoce el parámetro λ , en el modelo normal no se pueden determinar probabilidades sin el establecimiento de valores concretos para los parámetros μ y σ , etc. Y precisamente el hecho que motiva la necesidad de los métodos de la inferencia estadística es *el desconocimiento del verdadero valor de estos parámetros*.

En general, el modelo probabilístico con el que se pretende representar el comportamiento de la población puede expresarse como

$$f(x; \theta) \quad \text{para} \quad -\infty < x < \infty$$

de manera que $\theta \in \Theta$, siendo Θ el campo de variación del parámetro θ , denominado **espacio paramétrico**, pudiendo ser unidimensional o k -dimensional, según que el modelo probabilístico contenga uno o k parámetros. Así, por ejemplo, en la distribución de Poisson aparece un solo parámetro (λ) y en la normal dos (μ y σ).

En principio, el proceso inferencial estadístico se encamina a la evaluación de parámetros desconocidos, que puede lograrse mediante los planteamientos alternativos de la estimación y contrastación estadística de hipótesis.

En cualquier caso, tanto en el procedimiento de estimación de parámetros como en el de contrastación de hipótesis, el proceso inferencial siempre se basará en la información que nos pueda suministrar una muestra aleatoria simple X extraída de la población.

2.1 Función de verosimilitud

Para construir procedimientos inferenciales acerca del parámetro desconocido θ , es básico el disponer de mecanismos que permitan relacionar la información muestral X con dicho parámetro θ . La **función de verosimilitud** es la vía más utilizada para alcanzar este propósito.

Como se ha estudiado en el capítulo anterior, si la muestra obtenida es aleatoria simple entonces todos y cada uno de los elementos muestrales x_i son variables aleatorias estadísticamente independientes con la misma distribución de probabilidad que la población. Ahora bien, como en ésta se desconoce el verdadero valor del parámetro θ que la concreta, cada uno de los elementos de la muestra x_i pre-

sentará una distribución de probabilidad que también dependerá de este parámetro desconocido θ , es decir¹,

$$x_i \rightarrow iid f(x; \theta)$$

con lo que la distribución de probabilidad conjunta de la muestra es

$$f(x_1, \dots, x_n; \theta) = f(x_1; \theta) \cdots f(x_n; \theta) = L(\mathbf{X}; \theta).$$

Las funciones $f(x_i; \theta)$ representan las funciones de probabilidad del elemento i -ésimo de la muestra, siendo funciones de cuantía en el caso que la población sea discreta y de densidad si es continua.

Esta función conjunta, que de manera compacta se representa por $L(\mathbf{X}; \theta)$, recibe la denominación de *función de verosimilitud*, y tiene una interpretación ambivalente importante.

En efecto, podemos plantear los dos situaciones siguientes:

- Suponer que de las dos componentes de esta función, \mathbf{X} y θ , \mathbf{X} es variable y θ toma un valor, aunque desconocido, concreto y fijo del espacio paramétrico Θ , con lo cual $L(\mathbf{X}; \theta)$ es realmente la función de probabilidad conjunta de la muestra, que permitirá calcular las probabilidades de extracción de las distintas muestras de la población.
- Admitir que el parámetro θ puede tomar cualquier valor de los que comprende su espacio paramétrico y establecer que la muestra \mathbf{X} es fija y determinada, con lo que la función $L(\mathbf{X}; \theta)$ es únicamente función del parámetro θ . Esta es la interpretación que conduce al concepto de *función de verosimilitud*, que proporciona una medida racional de la posibilidad con la que el suceso (muestra \mathbf{X}) ha podido ocurrir según que el verdadero valor del parámetro θ sea uno u otro.

Para entender mejor la diferencia entre las dos situaciones anteriores consideremos el siguiente ejemplo: Supongamos que en una moneda trucada la probabilidad de cara, parámetro desconocido θ , puede tomar solamente los dos valores:

$$\theta_1 = \frac{1}{2}; \quad \theta_2 = \frac{2}{3}.$$

Las posibles muestras de tamaño dos serán: (cara; cara), (cara; cruz), (cruz; cara) y (cruz; cruz).

¹ Las iniciales *iid* resumen la propiedad que cada una de las variables aleatorias x_i son *independientes e idénticamente distribuidas*.

Las probabilidades de obtención de estas muestras para cada uno de los dos valores admitidos de θ son:

<i>Probabilidades</i>		
<i>Muestras</i>	$\theta_1 = 1/2$	$\theta_2 = 2/3$
X_1 (cara; cara)	$P(X_1/\theta_1) = 1/4$	$P(X_1/\theta_2) = 4/9$
X_2 (cara; cruz)	$P(X_2/\theta_1) = 1/4$	$P(X_2/\theta_2) = 2/9$
X_3 (cruz; cara)	$P(X_3/\theta_1) = 1/4$	$P(X_3/\theta_2) = 2/9$
X_4 (cruz; cruz)	$P(X_4/\theta_1) = 1/4$	$P(X_4/\theta_2) = 1/9$
Total	1	1

donde para cada valor fijo de θ se comprueba que

$$\sum_{i=1}^4 P(X_i/\theta_1) = \sum_{i=1}^4 P(X_i/\theta_2) = 1$$

no teniendo ningún sentido la suma de verosimilitudes

$$P(X_1/\theta_1) + P(X_1/\theta_2)$$

puesto que son probabilidades de distintas poblaciones (o, si se quiere, de distintas distribuciones de probabilidad).

Para el caso que la población tuviera una distribución de probabilidad continua, se tendría

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} L(\mathbf{X}; \theta) dx_1 \dots dx_n = 1.$$

La función de verosimilitud permite establecer un orden de preferencia respecto a los diferentes valores del parámetro desconocido θ , tomando como base la información contenida en la muestra obtenida \mathbf{X} , de manera que una determinada muestra informa mejor sobre un cierto valor θ_1 del parámetro que sobre otro valor θ_2 , si la verosimilitud en el primer caso es mayor que en el segundo, es decir, si $L(\mathbf{X}; \theta_1) > L(\mathbf{X}; \theta_2)$.

En este sentido, Fisher considera que el concepto de probabilidad para realizar inferencias partiendo de la muestra obtenida no es lo más idóneo que deba usarse puesto que «la cantidad matemática apropiada para medir nuestro orden de preferencia entre diferentes posibles poblaciones no obedece a las leyes de la probabilidad. Para distinguirla de la probabilidad utilizo el término **verosimilitud**».

2.2 Suficiencia

Para efectuar inferencias sobre el vector paramétrico θ , siempre se parte de la información que pueda suministrar una muestra aleatoria \mathbf{X} , como anteriormente se ha señalado, resumiéndose dicha información muestral en un estadístico $T(\mathbf{X})$. Por ejemplo, puede sintetizarse la información de la muestra $\mathbf{X} = (x_1, \dots, x_n)$ en un solo valor tal como la media muestral

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}.$$

Al actuar así aparecen dos cuestiones de interés:

- ¿El resumen que realiza $T(\mathbf{X})$ sobre la muestra \mathbf{X} , es tal que no se pierde ninguna información que pudiera contener \mathbf{X} acerca de los parámetros desconocidos que componen el vector paramétrico θ ?

Este punto es el contenido del concepto de **suficiencia**.

- ¿Se puede establecer alguna medida de la **información** contenida en la muestra sobre el conjunto de parámetros que aparecen en θ ?

Si la respuesta es afirmativa, ¿el resumen que supone el estadístico $T(\mathbf{X})$ recoge toda esa información? A esta segunda cuestión dedicaremos la parte última de este capítulo.

En cuanto al primero de los temas planteados, según Fisher², un estadístico es suficiente para efectuar inferencias sobre un parámetro θ cuando «*resume el conjunto de información relevante suministrada por la muestra*», y ningún otro estadístico constituido con la misma muestra puede proporcionar información adicional acerca del parámetro desconocido θ .

Los análisis estadísticos se realizan mediante una reducción de la información contenida en la muestra, es decir, sustituyendo la información primaria que proviene de cada elemento muestral por el resumen que supone el estadístico utilizado, síntesis y, a la vez, función de cada uno de esos elementos muestrales. Este resumen o reducción que lleva a cabo el estadístico $T(\mathbf{X})$ debe producirse sin pérdida de información y, en este sentido, la suficiencia reclama la propiedad que el estadístico debe tener en relación a la no pérdida de información.

² Véase FISHER (1922).

Para comprender mejor el alcance del concepto de suficiencia y sus consecuencias, comenzamos estudiando de qué manera cualquier estadístico genera una determinada partición del espacio muestral \mathfrak{X} .

Una función $T(\mathbf{X})$, estadístico, define una partición del espacio muestral si lo divide en una colección de k sucesos disjuntos \mathfrak{X}_i tales que

$$\bigcup_{i=1}^k \mathfrak{X}_i = \mathfrak{X}$$

$$\mathfrak{X}_i \cap \mathfrak{X}_j = \emptyset$$

siendo cada una de las muestras $\mathbf{X} \in \mathfrak{X}$.

Consideremos una población $B(1, \theta)$ ³ de la que se extrae una muestra aleatoria simple de tamaño 3, $\mathbf{X} = \{x_1, x_2, x_3\}$, y los siguientes estadísticos

$$T_1(\mathbf{X}) = \sum_{i=1}^3 x_i$$

$$T_2(\mathbf{X}) = \text{máx}(x_1, x_2, x_3).$$

En la tabla siguiente se relacionan las posibles muestras que pueden obtenerse (el espacio muestral \mathfrak{X}), así como los valores que alcanza cada uno de los dos estadísticos propuestos:

Muestra	Espacio muestral	$T_1(\mathbf{X})$	$T_2(\mathbf{X})$
1. ^a	(0, 0, 0)	0	0
2. ^a	(1, 0, 0)	1	1
3. ^a	(0, 1, 0)	1	1
4. ^a	(0, 0, 1)	1	1
5. ^a	(1, 1, 0)	2	1
6. ^a	(1, 0, 1)	2	1
7. ^a	(0, 1, 1)	2	1
8. ^a	(1, 1, 1)	3	1

En los esquemas siguientes se comprueba cómo cada estadístico produce una partición distinta del espacio muestral \mathfrak{X} .

³ El parámetro p de la distribución binomial lo denominamos θ , como genéricamente se representa cualquier parámetro poblacional.

1. ^a	2. ^a	3. ^a	4. ^a
5. ^a	6. ^a	7. ^a	8. ^a

Partición generada por $T_1(\mathbf{X})$

1. ^a	2. ^a	3. ^a	4. ^a
5. ^a	6. ^a	7. ^a	8. ^a

Partición generada por $T_2(\mathbf{X})$

Un estadístico es suficiente, respecto al parámetro θ , si basta con conocer a qué conjunto de la partición generada por él conduce la muestra obtenida, no añadiendo más información el saber cuál es el punto muestral (o muestra en concreto) que corresponda a esa partición.

Por ejemplo, si $T_1(\mathbf{X})$ fuera suficiente, el saber que procede de la quinta muestra no suministra más información acerca del parámetro θ que saber que $T_1(\mathbf{X})$ es igual a 2.

Cualquier estadístico $T(\mathbf{X})$ supone una reducción de la información que suministra uno a uno cada elemento de la muestra. La propiedad de suficiencia plantea el problema de encontrar una determinada reducción, es decir, un $T(\mathbf{X})$, donde no se pierda información.

Formalmente, podemos caracterizar la suficiencia de un estadístico mediante el siguiente planteamiento:

1. En una población definida por $f(x, \theta)$ existe un parámetro θ desconocido.
2. Para poder asignar un valor a dicho parámetro θ ⁴ partimos de la información que acerca del parámetro suministra una muestra $\mathbf{X} = \{x_1, \dots, x_n\}$.
3. En este proceso se construyen estadísticos $T(\mathbf{X})$ que suponen reducciones de la muestra y, también, particiones del espacio muestral.
4. Un estadístico $T(\mathbf{X})$ será suficiente si el conocimiento pormenorizado de los elementos muestrales que integran \mathbf{X} no añade nada sobre θ que no proporcione la síntesis $T(\mathbf{X})$.
5. Por consiguiente, la distribución de \mathbf{X} (conocimiento pormenorizado de la muestra) condicionada a que el estadístico $T(\mathbf{X})$ tome el valor concreto t , $T(\mathbf{X}) = t$, no debe depender de θ porque, en caso contrario, el conocimiento detallado de \mathbf{X} sí tendría que ver con el parámetro θ , y el estadístico $T(\mathbf{X})$ no habría captado por completo el mensaje que sobre θ pudiese estar contenido en la muestra \mathbf{X} .

El razonamiento anterior permite la siguiente definición formal de un estadístico suficiente:

«Un estadístico es suficiente respecto al parámetro θ si la distribución de probabilidad de la muestra \mathbf{X} condicionada al estadístico no implica el parámetro θ , es decir, si $F(\mathbf{X}/T(\mathbf{X}) = t)$ no depende de θ ».

⁴ Objetivo básico de las técnicas de estimación que se estudian en los capítulos siguientes.

Obtenemos estas distribuciones condicionadas para cada uno de los estadísticos $T_1(\mathbf{X})$ y $T_2(\mathbf{X})$ del ejemplo que introducimos en el concepto de partición del espacio muestral, donde se disponía de una muestra aleatoria simple de tamaño $n = 3$ extraída de una población $B(1, \theta)$.

Calculamos las probabilidades de las correspondientes distribuciones condicionadas a cada uno de los valores que pueda tomar el estadístico $T(\mathbf{X}) = t$, teniendo en cuenta que

$$P[(x_1, x_2, x_3)/T(\mathbf{X}) = t] = \frac{P[(x_1, x_2, x_3) \cap (T(\mathbf{X}) = t)]}{P(T(\mathbf{X}) = t)}$$

Para $T_1(\mathbf{X}) = \sum_{i=1}^3 x_i$ tendremos

$$\blacksquare P[(0, 0, 0)/T_1(\mathbf{X}) = 0] = \frac{P[(0, 0, 0) \cap (T_1(\mathbf{X}) = 0)]}{P(T_1(\mathbf{X}) = 0)} = \frac{(1 - \theta)^3}{(1 - \theta)^3} = 1$$

y para cualquier $(x_1, x_2, x_3) \neq (0, 0, 0)$

$$P[(x_1, x_2, x_3)/T_1(\mathbf{X}) = 0] = \frac{P[(x_1, x_2, x_3) \cap (T_1(\mathbf{X}) = 0)]}{P(T_1(\mathbf{X}) = 0)} = \frac{0}{(1 - \theta)^3} = 0$$

$$\blacksquare P[(1, 0, 0)/T_1(\mathbf{X}) = 1] = \frac{P[(1, 0, 0) \cap (T_1(\mathbf{X}) = 1)]}{P(T_1(\mathbf{X}) = 1)} = \frac{\theta(1 - \theta)^2}{3\theta(1 - \theta)^2} = \frac{1}{3}$$

$$P[(0, 1, 0)/T_1(\mathbf{X}) = 1] = \frac{1}{3}$$

$$P[(0, 0, 1)/T_1(\mathbf{X}) = 1] = \frac{1}{3}$$

siendo nulas el resto de probabilidades condicionadas a $T_1(\mathbf{X}) = 1$.

$$\blacksquare P[(1, 1, 0)/T_1(\mathbf{X}) = 2] = \frac{P[(1, 1, 0) \cap (T_1(\mathbf{X}) = 2)]}{P(T_1(\mathbf{X}) = 2)} = \frac{\theta(1 - \theta)^2}{3\theta(1 - \theta)^2} = \frac{1}{3}$$

$$P[(0, 1, 1)/T_1(\mathbf{X}) = 2] = \frac{1}{3}$$

$$P[(1, 0, 1)/T_1(\mathbf{X}) = 2] = \frac{1}{3}$$

siendo nulas el resto de probabilidades condicionadas a $T_1(\mathbf{X}) = 2$.

$$\blacksquare P[(1, 1, 1)/T_1(\mathbf{X}) = 3] = 1$$

siendo nula la probabilidad de cualquier otra muestra $(x_1, x_2, x_3) \neq (1, 1, 1)$ cuando $T_1(\mathbf{X}) = 3$.

Muestras	$P[\mathbf{X}/T_1(\mathbf{X}) = 0]$	$P[\mathbf{X}/T_1(\mathbf{X}) = 1]$	$P[\mathbf{X}/T_1(\mathbf{X}) = 2]$	$P[\mathbf{X}/T_1(\mathbf{X}) = 3]$
(0, 0, 0)	1	0	0	0
(1, 0, 0)	0	1/3	0	0
(0, 1, 0)	0	1/3	0	0
(0, 0, 1)	0	1/3	0	0
(1, 1, 0)	0	0	1/3	0
(1, 0, 1)	0	0	1/3	0
(0, 1, 1)	0	0	1/3	0
(1, 1, 1)	0	0	0	1

En este ejemplo, las distribuciones condicionadas son las que aparecen en el cuadro anterior que, como puede comprobarse, no dependen del parámetro θ , lo que lleva a que el estadístico $T_1(\mathbf{X})$ es suficiente para realizar inferencias acerca del parámetro θ de una población $B(1, \theta)$.

Estudiemos, a continuación, el otro estadístico $T_2(\mathbf{X}) = \max(x_1, x_2, x_3)$, teniendo en cuenta que

$$P[T_2(\mathbf{X}) = 0] = (1 - \theta)^3$$

$$P[T_2(\mathbf{X}) = 1] = 1 - P[T_2(\mathbf{X}) = 0] = 1 - (1 - \theta)^3$$

con lo que las probabilidades condicionadas son:

Muestras	$P[(x_1, x_2, x_3)/T_2(\mathbf{X}) = 0]$	$P[(x_1, x_2, x_3)/T_2(\mathbf{X}) = 1]$
(0, 0, 0)	1	0
(1, 0, 0)	0	$\frac{\theta(1 - \theta)^2}{1 - (1 - \theta)^3}$
(0, 1, 0)	0	$\frac{\theta(1 - \theta)^2}{1 - (1 - \theta)^3}$
(0, 0, 1)	0	$\frac{\theta(1 - \theta)^2}{1 - (1 - \theta)^3}$
(1, 1, 0)	0	$\frac{\theta^2(1 - \theta)}{1 - (1 - \theta)^3}$
(1, 0, 1)	0	$\frac{\theta^2(1 - \theta)}{1 - (1 - \theta)^3}$
(0, 1, 1)	0	$\frac{\theta^2(1 - \theta)}{1 - (1 - \theta)^3}$
(1, 1, 1)	0	$\frac{\theta^3}{1 - (1 - \theta)^3}$

que, al depender de θ , indican que el estadístico $T_2(\mathbf{X})$ no incorpora toda la información existente en la muestra \mathbf{X} sobre el parámetro θ , lo que haría necesario completar la reducción proporcionada por $T_2(\mathbf{X})$ con la información concreta de la muestra que produce cada valor concreto de $T_2(\mathbf{X})$. Por tanto, $T_2(\mathbf{X})$ no es suficiente respecto al parámetro θ .

Analizado el concepto de suficiencia de un estadístico y dentro de la vasta problemática de la suficiencia, puede plantearse la siguiente pregunta: ¿Es siempre necesario desarrollar las correspondientes distribuciones condicionadas de \mathbf{X} al estadístico $T(\mathbf{X})$ para determinar la suficiencia o no de un estadístico? O, por el contrario, ¿existe algún criterio simplificador que permita, más eficazmente, el análisis de la suficiencia?

A esta cuestión se dedica el epígrafe siguiente.

2.4 Criterio de factorización de Fisher-Neyman

¿Cómo puede determinarse de forma operativa si un estadístico es suficiente? La solución a esta pregunta viene dada por el criterio de factorización de Fisher-Neyman que dice: *la condición necesaria y suficiente para que el estadístico $T(\mathbf{X})$ sea suficiente es que la función de verosimilitud de la muestra $L(\mathbf{X}; \theta)$ se pueda descomponer en el producto de dos funciones, una $g(T(\mathbf{X}); \theta)$ dependiente del parámetro y de la muestra, a través del estadístico $T(\mathbf{X})$, y otra $H(\mathbf{X})$ independiente del parámetro θ , es decir,*

$$L(\mathbf{X}; \theta) = g(T(\mathbf{X}); \theta) \cdot H(\mathbf{X}).$$

Una demostración intuitiva de este teorema⁵ es la siguiente:

Representamos la función de probabilidad de la muestra \mathbf{X} condicionada al estadístico $T(\mathbf{X})$ como

$$f(\mathbf{X}/T(\mathbf{X}); \theta)$$

sabiendo que, si el estadístico $T(\mathbf{X})$ es suficiente, no debe depender del parámetro θ , es decir,

$$f(\mathbf{X}/T(\mathbf{X}); \theta) = H(\mathbf{X}).$$

⁵ Para una demostración rigurosa, véase, entre otros, ARNÁIZ y WILKS.

Por otra parte, toda función de probabilidad condicional se puede expresar como

$$f(\mathbf{X}/T(\mathbf{X}); \theta) = \frac{f(\mathbf{X} \cap T(\mathbf{X}); \theta)}{g(T(\mathbf{X}); \theta)} = \frac{f(\mathbf{X}; \theta)}{g(T(\mathbf{X}); \theta)} = \frac{L(\mathbf{X}; \theta)}{g(T(\mathbf{X}); \theta)}$$

pues la función de probabilidad de la muestra $f(\mathbf{X}; \theta)$ para cada valor de θ coincide con su función de verosimilitud $L(\mathbf{X}; \theta)$. Por tanto,

$$L(\mathbf{X}; \theta) = g(T(\mathbf{X}); \theta) \cdot f(\mathbf{X}/T(\mathbf{X}); \theta)$$

y si el estadístico $T(\mathbf{X})$ es suficiente, resulta

$$L(\mathbf{X}; \theta) = g(T(\mathbf{X}); \theta) \cdot H(\mathbf{X}).$$

Veamos a continuación algunas aplicaciones de este criterio para la determinación de estadísticos suficientes.

EJEMPLO 1

De una población con distribución de Poisson con parámetro λ , se extrae una muestra aleatoria simple de tamaño n . Compruébese si el estadístico $T(\mathbf{X}) = \sum_{i=1}^n x_i$ es suficiente.

La función de verosimilitud es

$$\begin{aligned} L(\mathbf{X}; \lambda) &= f(x_1; \lambda) \cdots f(x_n; \lambda) = \\ &= \frac{e^{-\lambda} \cdot \lambda^{x_1}}{x_1!} \cdots \frac{e^{-\lambda} \cdot \lambda^{x_n}}{x_n!} = \\ &= \frac{e^{-n\lambda} \cdot \lambda^{\sum_{i=1}^n x_i}}{x_1! \cdots x_n!} = \frac{e^{-n\lambda} \cdot \lambda^{T(\mathbf{X})}}{x_1! \cdots x_n!} \end{aligned}$$

al ser la muestra aleatoria simple, cada x_i se distribuye igual que la población, es decir, $x_i \rightarrow iid$ Poisson (λ).

Dado que

$$L(\mathbf{X}; \lambda) = e^{-n\lambda} \lambda^{T(\mathbf{X})} \frac{1}{x_1! \cdots x_n!}$$

entonces

$$e^{-n\lambda} \lambda^{T(\mathbf{X})} = g(T(\mathbf{X}); \lambda)$$

depende del parámetro λ y de la muestra \mathbf{X} a través del estadístico, y la función

$$H(\mathbf{X}) = \frac{1}{x_1! \cdots x_n!}$$

no depende del parámetro λ , llegándose a que el estadístico $T(\mathbf{X}) = \sum_{i=1}^n x_i$ es suficiente.

EJEMPLO 2

En una población continua con distribución exponencial definida a través de la función de densidad

$$f(x; \theta) = \theta e^{-\theta x}, \quad x \geq 0, \quad \theta > 0$$

se considera el estadístico $T(\mathbf{X}) = \sum_{i=1}^n x_i$; estúdiense si es suficiente.

La función de verosimilitud es

$$\begin{aligned} L(\mathbf{X}; \theta) &= f(x_1; \theta) \cdots f(x_n; \theta) = \\ &= \theta e^{-\theta x_1} \cdots \theta e^{-\theta x_n} = \\ &= \theta^n e^{-\theta \sum_{i=1}^n x_i} = \theta^n e^{-\theta T(\mathbf{X})}. \end{aligned}$$

Definiendo la función $H(\mathbf{X}) = 1$, y considerando que

$$g(T(\mathbf{X}); \theta) = \theta^n e^{-\theta T(\mathbf{X})},$$

se puede establecer la factorización

$$L(\mathbf{X}; \theta) = g(T(\mathbf{X}); \theta) \cdot H(\mathbf{X})$$

que conduce a que $T(\mathbf{X}) = \sum_{i=1}^n x_i$ es suficiente.

2.5 Información

Un estadístico es suficiente si aprovecha toda la **información** que respecto al parámetro θ le suministra la muestra; en este epígrafe se propone medir esa información y se analiza hasta qué punto un estadístico suficiente la incorpora en sí.

Antes de efectuar este análisis, parece conveniente tener en cuenta las siguientes consideraciones previas:

- ¿Cómo puede definirse la cantidad de información contenida en una muestra? El término información se considera estrechamente vinculado a la idea de disparidad. Cuanto mayor sea la variabilidad, las discrepancias en la población, mayor información debe contener la muestra aleatoria \mathbf{X} . En este principio básico se fundamentan las medidas que se analizarán.
- Si se desea conocer todo lo referente al parámetro θ que contenga la muestra aleatoria \mathbf{X} , la vía idónea para hacerlo es a través de la función de verosimilitud $L(\mathbf{X}; \theta)$, por ser una relación funcional que liga la muestra con el parámetro.

En esta verosimilitud se basan las medidas que se proponen, pero considerando su transformación logarítmica, $\ln L(\mathbf{X}; \theta)$,⁶ por las siguientes razones:

- La transformación logarítmica es monótona y neutral en los procesos de optimización: el máximo (o mínimo) de una función, si existe, se obtiene en el mismo punto que el máximo (o mínimo) de su transformada logarítmica.
- Dado que, generalmente, se trabaja con muestras aleatorias simples y que la verosimilitud es el producto de las funciones de probabilidad para cada uno de los elementos muestrales, es decir,

$$L(\mathbf{X}; \theta) = f(x_1; \theta) f(x_2; \theta) \cdots f(x_n; \theta)$$

la transformación logarítmica del producto incorpora la propiedad de aditividad, puesto que,

$$\ln L(\mathbf{X}; \theta) = \sum_{i=1}^n \ln f(x_i; \theta)$$

característica que, según veremos a continuación, es de importancia para cualquier medida de la información.

⁶ $\ln L(\mathbf{X}; \theta)$ se conoce como **función de soporte**.

2.5.1. INFORMACIÓN DE FISHER

Para Fisher *una buena medida de la información que una muestra \mathbf{X} contiene respecto a un parámetro θ , viene dada por la sensibilidad (capacidad de reacción o variación) que dicha muestra aleatoria \mathbf{X} muestre ante variaciones del parámetro θ . Es decir, el tamaño de la variación de $L(\mathbf{X}; \theta)$ ante variaciones del parámetro θ indicaría cuánta información sobre dicho parámetro contiene la muestra, recogiendo las discrepancias (información) que se producen respecto a los diferentes valores de dicho parámetro.*

Partiendo de esta idea, Fisher define la variable que denomina **score** como

$$SC(\theta) = \frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta}$$

que incorpora la transformación logarítmica, de acuerdo con lo anteriormente señalado, y recoge las variaciones instantáneas respecto al parámetro θ , a través de la derivada de $\ln L(\mathbf{X}; \theta)$ respecto a ese parámetro θ .⁷

De esta manera, la variable aleatoria **score** sería como un marcador que refleja los movimientos de la verosimilitud, a través de la función $\ln L(\mathbf{X}; \theta)$, ante cambios infinitesimales del parámetro θ .

Elegido el elemento que suministra la información requerida, Fisher adoptó el criterio de medir dicha información mediante la variabilidad⁸ que presentaba la variable aleatoria **score**, a través de la varianza de esta variable, a la que llamó **cantidad de información** y que representaremos por $I(\theta)$.

A fin de obtener la varianza es preciso tener en cuenta que deben cumplirse las siguientes condiciones, denominadas **condiciones de regularidad de Fisher-Wolfowitz**:

- I. La población de donde procede la muestra presenta un campo de variación que no depende del parámetro θ y, por tanto, la muestra \mathbf{X} tampoco.
- II. La función $\ln L(\mathbf{X}; \theta)$ admite, por lo menos, dos primeras derivadas respecto al parámetro θ .
- III. Las operaciones de derivación e integración (o suma en el caso de variable aleatoria discreta) son intercambiables.

⁷ Recuérdese el concepto de derivada.

⁸ Insistimos en la idea de información como variabilidad que antes se apuntó.

Para calcular $I(\theta)$ determinaremos inicialmente la esperanza matemática de la variable aleatoria **score** ⁹

$$\begin{aligned} E[SC(\theta)] &= E\left[\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta}\right] = \\ &= \int_{\mathbf{X}} \frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta} L(\mathbf{X}; \theta) d\mathbf{X} = \\ &= \int_{\mathbf{X}} \frac{1}{L(\mathbf{X}; \theta)} \frac{\partial L(\mathbf{X}; \theta)}{\partial \theta} L(\mathbf{X}; \theta) d\mathbf{X} = \\ &= \int_{\mathbf{X}} \frac{\partial L(\mathbf{X}; \theta)}{\partial \theta} d\mathbf{X} = \\ &= \frac{\partial}{\partial \theta} \int_{\mathbf{X}} L(\mathbf{X}; \theta) d\mathbf{X} = 0 \end{aligned}$$

puesto que

$$\int_{\mathbf{X}} L(\mathbf{X}; \theta) d\mathbf{X} = 1$$

por tanto, la variable **score** es de media nula, y

$$\begin{aligned} I(\theta) &= V[SC(\theta)] = V\left[\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta}\right] = \\ &= E\left[\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta}\right]^2 - \left\{E\left[\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta}\right]\right\}^2 = \\ &= E\left[\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta}\right]^2 \end{aligned}$$

que, dado su carácter de varianza, verificará

$$I(\theta) \geq 0.$$

⁹ Utilizamos notación compacta para resumir las integrales múltiples, es decir,

$$\int_{x_1} \dots \int_{x_n} \frac{\partial \ln L(x_1, \dots, x_n; \theta)}{\partial \theta} L(x_1, \dots, x_n; \theta) dx_1 \dots dx_n = \int_{\mathbf{X}} \frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta} L(\mathbf{X}; \theta) d\mathbf{X}.$$

Para calcular $I(\theta)$ determinaremos inicialmente la esperanza matemática de la variable aleatoria **score** ⁹

$$\begin{aligned} E[SC(\theta)] &= E\left[\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta}\right] = \\ &= \int_{\mathbf{X}} \frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta} L(\mathbf{X}; \theta) d\mathbf{X} = \\ &= \int_{\mathbf{X}} \frac{1}{L(\mathbf{X}; \theta)} \frac{\partial L(\mathbf{X}; \theta)}{\partial \theta} L(\mathbf{X}; \theta) d\mathbf{X} = \\ &= \int_{\mathbf{X}} \frac{\partial L(\mathbf{X}; \theta)}{\partial \theta} d\mathbf{X} = \\ &= \frac{\partial}{\partial \theta} \int_{\mathbf{X}} L(\mathbf{X}; \theta) d\mathbf{X} = 0 \end{aligned}$$

puesto que

$$\int_{\mathbf{X}} L(\mathbf{X}; \theta) d\mathbf{X} = 1$$

por tanto, la variable **score** es de media nula, y

$$\begin{aligned} I(\theta) &= V[SC(\theta)] = V\left[\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta}\right] = \\ &= E\left[\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta}\right]^2 - \left\{E\left[\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta}\right]\right\}^2 = \\ &= E\left[\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta}\right]^2 \end{aligned}$$

que, dado su carácter de varianza, verificará

$$I(\theta) \geq 0.$$

⁹ Utilizamos notación compacta para resumir las integrales múltiples, es decir,

$$\int_{x_1} \dots \int_{x_n} \frac{\partial \ln L(x_1, \dots, x_n; \theta)}{\partial \theta} L(x_1, \dots, x_n; \theta) dx_1 \dots dx_n = \int_{\mathbf{X}} \frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta} L(\mathbf{X}; \theta) d\mathbf{X}.$$

Esta **cantidad de información de Fisher** presenta las siguientes propiedades:

1. $I(\theta) = 0$ si y sólo si la verosimilitud no depende del parámetro θ y, por tanto, no contiene información ninguna respecto a dicho parámetro, es decir, si

$$\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta} = 0, \quad \text{para todo } \theta \in \Theta.$$

2. Es una medida de información aditiva. Esta propiedad puede contemplarse bajo los siguientes prismas:

- Si \mathbf{X}_1 y \mathbf{X}_2 son dos muestras independientes aleatorias simples, procedentes de la misma población

$$I_{\mathbf{X}_1, \mathbf{X}_2}(\theta) = I_{\mathbf{X}_1}(\theta) + I_{\mathbf{X}_2}(\theta)$$

En efecto,

$$\begin{aligned} I_{\mathbf{X}_1, \mathbf{X}_2}(\theta) &= E \left[\frac{\partial \ln L(\mathbf{X}_1, \mathbf{X}_2; \theta)}{\partial \theta} \right]^2 = \\ &= E \left[\frac{\partial \ln L(\mathbf{X}_1; \theta)}{\partial \theta} + \frac{\partial \ln L(\mathbf{X}_2; \theta)}{\partial \theta} \right]^2 = \\ &= E \left[\frac{\partial \ln L(\mathbf{X}_1; \theta)}{\partial \theta} \right]^2 + E \left[\frac{\partial \ln L(\mathbf{X}_2; \theta)}{\partial \theta} \right]^2 + \\ &\quad + 2E \left[\frac{\partial \ln L(\mathbf{X}_1; \theta)}{\partial \theta} \right] E \left[\frac{\partial \ln L(\mathbf{X}_2; \theta)}{\partial \theta} \right] = \\ &= E \left[\frac{\partial \ln L(\mathbf{X}_1; \theta)}{\partial \theta} \right]^2 + E \left[\frac{\partial \ln L(\mathbf{X}_2; \theta)}{\partial \theta} \right]^2 = \\ &= I_{\mathbf{X}_1}(\theta) + I_{\mathbf{X}_2}(\theta) \end{aligned}$$

ya que las muestras son independientes y además

$$E \left[\frac{\partial \ln L(\mathbf{X}_1; \theta)}{\partial \theta} \right] = E \left[\frac{\partial \ln L(\mathbf{X}_2; \theta)}{\partial \theta} \right] = 0$$

según se expuso anteriormente.

- Si consideramos cada uno de los elementos x_i de la muestra como una muestra elemental de tamaño uno, por la aditividad de información de Fisher

$$I(\theta) = \sum_{i=1}^n I_i(\theta)$$

donde, al ser la muestra aleatoria simple

$$I_i(\theta) = E \left[\frac{\partial \ln f(x_i; \theta)}{\partial \theta} \right]^2 = E \left[\frac{\partial \ln f(x; \theta)}{\partial \theta} \right]^2 = i(\theta)$$

siendo $f(x; \theta)$ la función de probabilidad de la población e $i(\theta)$ la cantidad de información sobre θ existente en dicha población, con lo que

$$I(\theta) = \sum_{i=1}^n I_i(\theta) = n i(\theta),$$

es decir, la cantidad de información sobre θ que contiene la muestra es igual a n veces la información de cada una de las variables aleatorias muestrales x_i que, a su vez, coincide con la de la variable aleatoria poblacional, verificándose, por tanto,

$$E \left[\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta} \right]^2 = n E \left[\frac{\partial \ln f(x; \theta)}{\partial \theta} \right]^2.$$

3. La cantidad de información de Fisher puede expresarse también como¹⁰

$$I(\theta) = E \left[\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta} \right]^2 = -E \left[\frac{\partial^2 \ln L(\mathbf{X}; \theta)}{\partial \theta^2} \right],$$

expresión que puede ser útil en algunos casos para facilitar el cálculo de $I(\theta)$.

Cuando la población es multiparamétrica, dada por $f(\mathbf{X}; \theta)$, se obtiene la **matriz de información de Fisher**

$$I(\theta) = -E \left[\frac{\partial^2 \ln L(\mathbf{X}; \theta)}{\partial \theta_i \partial \theta_j} \right],$$

siendo θ un vector de parámetros.

¹⁰ RUIZ-MAYA y MARTÍN PLIEGO: *Estadística II: Inferencia*. Capítulo 4.

2.5.2. ESTADÍSTICO SUFICIENTE E INFORMACIÓN DE FISHER

Si un estadístico $T(\mathbf{X})$ es suficiente se verificará el criterio de factorización de Fisher-Neyman, es decir, la verosimilitud se expresa como

$$L(\mathbf{X}; \theta) = g(T(\mathbf{X}); \theta) \cdot H(\mathbf{X})$$

y, por consiguiente, la variable aleatoria **score** será función del estadístico suficiente, pues

$$\ln L(\mathbf{X}; \theta) = \ln g(T(\mathbf{X}); \theta) + \ln H(\mathbf{X})$$

y

$$\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta} = \frac{\partial \ln g(T(\mathbf{X}); \theta)}{\partial \theta}.$$

Por otra parte, la información suministrada por un estadístico $T(\mathbf{X})$ es siempre igual o inferior a la suministrada por la muestra. Sólo si el estadístico es suficiente contendrá la misma información que la muestra.

En efecto, supongamos que $T(\mathbf{X})$ no es suficiente, por lo que de acuerdo con el criterio de factorización de Fisher-Neyman

$$L(\mathbf{X}; \theta) = g(T(\mathbf{X}); \theta) \cdot H(\mathbf{X}; \theta)$$

donde $H(\mathbf{X}; \theta)$ depende del parámetro al no ser suficiente. Tomando logaritmos y considerando las dos primeras derivadas

$$\frac{\partial^2 \ln L(\mathbf{X}; \theta)}{\partial \theta^2} = \frac{\partial^2 \ln g(T(\mathbf{X}); \theta)}{\partial \theta^2} + \frac{\partial^2 \ln H(\mathbf{X}; \theta)}{\partial \theta^2}$$

tomando esperanzas y multiplicando por -1

$$-E \left[\frac{\partial^2 \ln L(\mathbf{X}; \theta)}{\partial \theta^2} \right] = -E \left[\frac{\partial^2 \ln g(T(\mathbf{X}); \theta)}{\partial \theta^2} \right] - E \left[\frac{\partial^2 \ln H(\mathbf{X}; \theta)}{\partial \theta^2} \right],$$

es decir,

$$I(\theta) = I_T(\theta) - E \left[\frac{\partial^2 \ln H(\mathbf{X}; \theta)}{\partial \theta^2} \right]$$

donde $I_T(\theta)$ representa la cantidad de información que suministra la muestra a través del estadístico $T(\mathbf{X})$. Por otra parte, como

$$-E \left[\frac{\partial^2 \ln H(\mathbf{X}; \theta)}{\partial \theta^2} \right] = E \left[\frac{\partial \ln H(\mathbf{X}; \theta)}{\partial \theta} \right]^2 \geq 0$$

se llega a que

$$I(\theta) \geq I_T(\theta)$$

alcanzándose la igualdad solamente si

$$E \left[\frac{\partial \ln H(\mathbf{X}; \theta)}{\partial \theta} \right]^2 = 0$$

es decir, si

$$\frac{\partial \ln H(\mathbf{X}; \theta)}{\partial \theta} = 0$$

expresión que conduce a que la función $H(\mathbf{X}; \theta)$ no puede depender de θ , esto es, será de la forma $H(\mathbf{X})$ y el estadístico $T(\mathbf{X})$ resulta, por consiguiente, suficiente.

Ejercicios resueltos

EJERCICIO 2.1 Consideremos una variable aleatoria que puede tomar cuatro valores: 1, 2, 3 y 4, y las posibles muestras aleatorias simples de tamaño 2.

Estúdiese la partición generada por el estadístico $T(X) = \sum_{i=1}^2 x_i$ en el espacio muestra.

SOLUCIÓN. El espacio muestral \mathfrak{X} lo representamos de la siguiente forma:

(1, 4)	(2, 4)	(3, 4)	(4, 4)
(1, 3)	(2, 3)	(3, 3)	(4, 3)
(1, 2)	(2, 2)	(3, 2)	(4, 2)
(1, 1)	(2, 1)	(3, 1)	(4, 1)

Consideremos el estadístico $T_1 = x_1 + x_2$. Los valores que puede tomar T_1 y los puntos del espacio muestral que los originan son:

$T_1 = x_1 + x_2$	Muestras que lo originan
2	(1, 1)
3	(1, 2); (2, 1)
4	(1, 3); (2, 2); (3, 1)
5	(1, 4); (2, 3); (3, 2); (4, 1)
6	(2, 4); (3, 3); (4, 2)
7	(3, 4); (4, 3)
8	(4, 4)

y el espacio muestral \mathfrak{X} queda ahora dividido, en sucesos disjuntos, en función de los valores que puede tomar el estadístico, formando la partición generada siete clases de subconjuntos:

(1, 4) $T = 5$	(2, 4) $T = 6$	(3, 4) $T = 7$	(4, 4) $T = 8$
(1, 3) $T = 4$	(2, 3) $T = 5$	(3, 3) $T = 6$	(4, 3) $T = 7$
(1, 2) $T = 3$	(2, 2) $T = 4$	(3, 2) $T = 5$	(4, 2) $T = 6$
(1, 1) $T = 2$	(2, 1) $T = 3$	(3, 1) $T = 4$	(4, 1) $T = 5$

EJERCICIO 2.2 En una población con función de densidad

$$f(\mathbf{X}; \theta) = e^{-(x-\theta)}, \quad x \geq \theta > 0$$

estúdiense si el estadístico $T(\mathbf{X}) = \sum_{i=1}^n x_i$ es suficiente, siendo (x_1, \dots, x_n) una muestra aleatoria simple.

SOLUCIÓN. Formamos la función de verosimilitud

$$\begin{aligned} L(\mathbf{X}; \theta) &= f(x_1; \theta) \cdots f(x_n; \theta) = \\ &= e^{-(x_1-\theta)} \cdots e^{-(x_n-\theta)} = \\ &= e^{-\sum_{i=1}^n x_i + n\theta} = e^{-T(\mathbf{X}) + n\theta} \end{aligned}$$

donde

$$e^{-T(\mathbf{X}) + n\theta} = g(T(\mathbf{X}); \theta).$$

Ahora bien, si hacemos $H(\mathbf{X}) = 1$ no logramos una función que no dependa del parámetro θ pues todos los x_i de la muestra se distribuyen igual que la población y, por consiguiente, $x_i \geq \theta$; es decir, el campo de variación de cada x_i depende del parámetro θ y la función $H(\mathbf{X}) = 1$ no es indiferente ante variaciones de θ .

Como no puede lograrse la factorización requerida el estadístico $T(\mathbf{X}) = \sum_{i=1}^n x_i$ no es suficiente.

EJERCICIO 2.3 Determinése la cantidad de información de Fisher contenida en una muestra aleatoria simple \mathbf{X} , extraída de una población $N(\mu; \sigma)$ con σ conocida.

SOLUCIÓN. La función de verosimilitud es

$$L(\mathbf{X}; \mu) = \sigma^{-n} (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2}$$

tomando logaritmos

$$\ln L(\mathbf{X}; \mu) = -n \ln \sigma - \frac{n}{2} \ln 2\pi - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

y

$$\frac{\partial \ln L(\mathbf{X}; \mu)}{\partial \mu} = -\frac{1}{2\sigma^2} 2 \sum_{i=1}^n (x_i - \mu)(-1) = \frac{\sum_{i=1}^n (x_i - \mu)}{\sigma^2}$$

por lo que

$$\begin{aligned} I(\mu) &= E \left[\frac{\partial \ln L(\mathbf{X}; \mu)}{\partial \mu} \right]^2 = E \left[\frac{\sum_{i=1}^n (x_i - \mu)}{\sigma^2} \right]^2 = \\ &= \frac{1}{\sigma^4} \left[\sum_{i=1}^n E(x_i - \mu)^2 + \sum_{i \neq j} E(x_i - \mu) E(x_j - \mu) \right] = \\ &= \frac{1}{\sigma^4} \sum_{i=1}^n \sigma^2 = \frac{n\sigma^2}{\sigma^4} = \frac{n}{\sigma^2} \end{aligned}$$

que también podría haberse obtenido teniendo en cuenta la aditividad de esta medida,

$$\begin{aligned} f(x; \mu) &= \sigma^{-1} (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \\ \ln f(x; \mu) &= -\ln \sigma - \frac{1}{2} \ln 2\pi - \frac{1}{2\sigma^2} (x - \mu)^2 \\ \frac{\partial \ln f(x; \mu)}{\partial \mu} &= -\frac{1}{2\sigma^2} 2(x - \mu)(-1) = \frac{(x - \mu)}{\sigma^2} \\ i(\mu) &= E \left[\frac{\partial \ln f(x; \mu)}{\partial \mu} \right]^2 = E \left[\frac{x - \mu}{\sigma^2} \right]^2 = \frac{1}{\sigma^4} E(x - \mu)^2 = \frac{1}{\sigma^4} \sigma^2 = \frac{1}{\sigma^2} \end{aligned}$$

y, por tanto,

$$I(\mu) = n i(\mu) = \frac{n}{\sigma^2}.$$

Se puede comprobar cómo la cantidad de información sobre el parámetro μ es tanto mayor cuanto mayor sea el tamaño de la muestra y menor la varianza de la población.

EJERCICIO 2.4 En una población con distribución $N(\mu; \sigma)$, determínese la matriz de información de Fisher respecto a los dos parámetros μ y σ^2 , a partir de una muestra aleatoria simple \mathbf{X} de tamaño n .

SOLUCIÓN. Dado que

$$I(\theta) = -E \left[\frac{\partial^2 \ln L(\mathbf{X}; \theta)}{\partial \theta \partial \theta'} \right] = -E \begin{bmatrix} \frac{\partial^2 \ln L(\mathbf{X}; \theta)}{\partial \mu^2} & \frac{\partial^2 \ln L(\mathbf{X}; \theta)}{\partial \mu \partial \sigma^2} \\ \frac{\partial^2 \ln L(\mathbf{X}; \theta)}{\partial \sigma^2 \partial \mu} & \frac{\partial^2 \ln L(\mathbf{X}; \theta)}{\partial (\sigma^2)^2} \end{bmatrix} =$$

$$= \begin{bmatrix} -E \left[\frac{\partial^2 \ln L(\mathbf{X}; \theta)}{\partial \mu^2} \right] & -E \left[\frac{\partial^2 \ln L(\mathbf{X}; \theta)}{\partial \mu \partial \sigma^2} \right] \\ -E \left[\frac{\partial^2 \ln L(\mathbf{X}; \theta)}{\partial \sigma^2 \partial \mu} \right] & -E \left[\frac{\partial^2 \ln L(\mathbf{X}; \theta)}{\partial (\sigma^2)^2} \right] \end{bmatrix}$$

para determinar esta matriz habrá que calcular previamente las segundas derivadas de la función de soporte $\ln L(\mathbf{X}; \theta)$

$$f(\mathbf{X}; \theta) = f(\mathbf{X}; \mu, \sigma^2) = (\sigma^2)^{-\frac{1}{2}} (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2} (x-\mu)^2}$$

$$L(\mathbf{X}; \theta) = L(\mathbf{X}; \mu, \sigma^2) = (\sigma^2)^{-\frac{n}{2}} (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2}$$

$$\ln L(\mathbf{X}; \theta) = \ln L(\mathbf{X}; \mu, \sigma^2) = -\frac{n}{2} \ln(\sigma^2) - \frac{n}{2} \ln(2\pi) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

siendo

$$\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu)$$

$$\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2$$

y calculando las segundas derivadas

$$\frac{\partial^2 \ln L(\mathbf{X}; \theta)}{\partial \mu^2} = -\frac{n}{\sigma^2}$$

$$\frac{\partial^2 \ln L(\mathbf{X}; \theta)}{\partial \mu \partial \sigma^2} = -\frac{1}{\sigma^4} \sum_{i=1}^n (x_i - \mu)$$

$$\frac{\partial^2 \ln L(\mathbf{X}; \theta)}{\partial (\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (x_i - \mu)^2$$

la matriz de información de Fisher resulta

$$\begin{aligned}
 I(\theta) &= -E \left[\begin{array}{cc} -\frac{n}{\sigma^2} & -\frac{1}{\sigma^4} \sum_{i=1}^n (x_i - \mu) \\ -\frac{1}{\sigma^4} \sum_{i=1}^n (x_i - \mu) & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n E(x_i - \mu)^2 \end{array} \right] = \\
 &= \left[\begin{array}{cc} \frac{n}{\sigma^2} & \frac{1}{\sigma^4} \sum_{i=1}^n E(x_i - \mu) \\ \frac{1}{\sigma^4} \sum_{i=1}^n E(x_i - \mu) & -\frac{n}{2\sigma^4} + \frac{1}{\sigma^6} \sum_{i=1}^n E(x_i - \mu)^2 \end{array} \right] = \\
 &= \left[\begin{array}{cc} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{array} \right]
 \end{aligned}$$

pues $E(x_i - \mu) = 0$ y $E(x_i - \mu)^2 = \sigma^2$ por ser x_i un elemento de una muestra aleatoria simple procedente de una población $N(\mu; \sigma)$.

CAPÍTULO 3

Estimación puntual: propiedades de los estimadores

3.1 Introducción

La definición de la distribución de probabilidad de una variable aleatoria implica, en general, la existencia y conocimiento de parámetros sin cuya concreción no es posible utilizar la distribución. Si, por ejemplo, en una variable aleatoria $B(1; p)$ desconocemos el valor de p , o en una $N(\mu; \sigma)$ se ignora el valor de uno o de los dos parámetros, no podremos calcular probabilidades de sucesos en los que estén implicadas las correspondientes variables.

Las situaciones planteadas se pueden expresar formalmente de la siguiente manera: sea una variable aleatoria ξ cuya función de cuantía o densidad $f(x; \theta)$, incluye un parámetro¹ θ , definido en el **espacio paramétrico**, Θ , ($\theta \in \Theta$). Para cada valor del parámetro θ tendremos una distribución de probabilidad distinta,

¹ O un conjunto de k parámetros, en cuyo caso el espacio paramétrico Θ será un recinto k -dimensional.

pertenciente a una misma familia de distribuciones. Por ejemplo, si la densidad de la variable ξ es $f(x; \theta) = \theta e^{-\theta x}$, $x \geq 0$ y $\theta > 0$, para cada valor de θ hay una función de densidad exponencial.

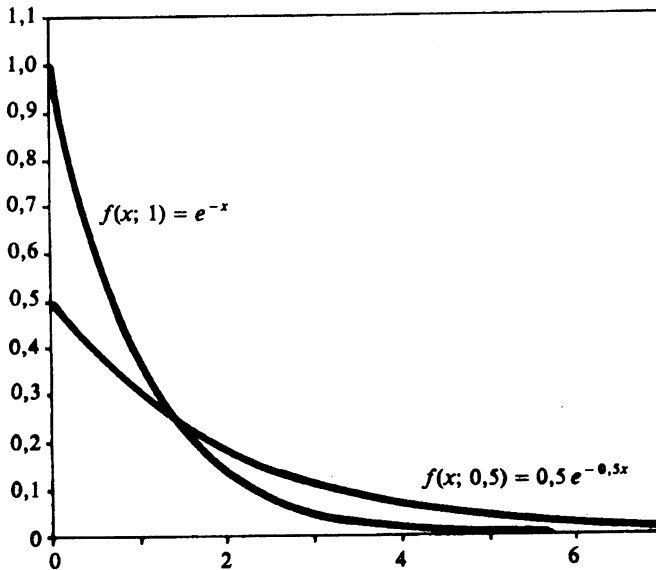


FIGURA 3.1

En la figura 3.1 aparecen las funciones de densidad para dos valores de θ : $\theta = 0,5$ y $\theta = 1$.

Lo expuesto pone de manifiesto la necesidad de asignar valores a los parámetros desconocidos contenidos en una distribución de probabilidad. A este problema está dedicada una parte de la Inferencia Estadística: la **teoría de la estimación**.

3.2 Concepto de estimador, estimación y error de estimación

Sea $f(x; \theta)$ una función de densidad o de cuantía conteniendo un parámetro θ desconocido. El primer objetivo de la Inferencia Estadística es la obtención de un

valor que pueda asignarse a ese parámetro desconocido. Para ello se obtiene de la población la información precisa mediante una muestra aleatoria simple, se establece una función de los valores muestrales², **estimador**, y se asigna al parámetro el valor que tome esta función en la muestra concreta, valor denominado **estimación puntual**.

Por tanto, para efectuar la estimación del parámetro θ , se considera una muestra aleatoria simple $\mathbf{X}(x_1, \dots, x_n)$, de tamaño n , siendo el estimador una función de las variables muestrales $\theta^* = \theta^*(x_1, \dots, x_n)$, es decir, una función que recoge y resume la información que sobre la población le suministra la muestra y que utiliza dicha información para la obtención de valores aproximados del parámetro desconocido θ , aproximaciones que se denominan **estimaciones**.

El estimador θ^* será una variable aleatoria función de las variables aleatorias muestrales (x_1, \dots, x_n) , y se transformará en una estimación del parámetro θ , un valor numérico concreto, cuando las variables muestrales (x_1, \dots, x_n) se conviertan en datos observados al obtenerse una muestra determinada. Como se ha descrito, un estimador siempre es un estadístico, tal como se definió en los capítulos anteriores.

Se ha definido un estimador como una función de las variables muestrales, valores experimentales, por lo cual diversas funciones de éstas pueden considerarse, en principio, como estimadores del parámetro. Lo que lleva a disponer de tantas estimaciones puntuales, en cada muestra, como estimadores hayamos podido construir, por lo que será necesario alguna norma o conjunto de criterios que permitan elegir entre estos posibles estimadores cual es el «mejor» en cada caso.

3.2.1. ERROR CUADRÁTICO MEDIO DEL ESTIMADOR

Al ser desconocido el parámetro nunca sabremos exactamente hasta qué punto cada estimación (valor aproximado) se encuentra lejos o cerca del valor del parámetro, es decir, su utilización conduce a la posibilidad de cometer un error más o menos elevado al trabajar con la estimación puntual como si fuera el valor verdadero.

Para establecer la bondad de un estimador, partimos del hecho de ser deseable conocer si la estimación se encuentra lejos o cerca del valor verdadero, siempre desconocido. En la práctica, tal pretensión no es más que un deseo irrealizable, aunque conveniente, para el planteamiento teórico del problema. El error que podemos cometer, al tomar como valor del parámetro θ el proporcionado por el

² Sin que contenga ningún parámetro de la distribución.

estimador, es la diferencia $\theta^* - \theta$ o, para eliminar el signo de las diferencias, su cuadrado $(\theta^* - \theta)^2$. Si pudiéramos obtener todas las posibles muestras y para cada una la correspondiente estimación, una media global de los errores sería la desviación cuadrática media de todos ellos o, más concretamente, su esperanza matemática, denominada **error cuadrático medio**, **ECM**, **desviación cuadrática media**, o **acuracidad**³ del estimador,

$$ECM(\theta^*) = E(\theta^* - \theta)^2.$$

Un valor pequeño del error cuadrático medio indicará que, en media, el estimador no se encuentra lejos del parámetro θ ; inversamente, cuanto mayor sea más lejos estará, también en media.

En el segundo miembro de $ECM(\theta^*) = E(\theta^* - \theta)^2$ sumamos y restamos la esperanza matemática del estimador, $E(\theta^*)$,⁴ y, efectuando operaciones,

$$\begin{aligned} E(\theta^* - \theta)^2 &= E[\theta^* - \theta + E(\theta^*) - E(\theta^*)]^2 = \\ &= E\{[\theta^* - E(\theta^*)] - [\theta - E(\theta^*)]\}^2 = \\ &= E[\theta^* - E(\theta^*)]^2 + E[\theta - E(\theta^*)]^2 - 2E\{[\theta^* - E(\theta^*)][\theta - E(\theta^*)]\} = \\ &= V(\theta^*) + [\theta - E(\theta^*)]^2 - [\theta - E(\theta^*)] E[\theta^* - E(\theta^*)] = \\ &= V(\theta^*) + [\theta - E(\theta^*)]^2 \end{aligned} \quad [1]$$

la esperanza del tercer término de la penúltima expresión es igual a cero, pues

$$E[\theta^* - E(\theta^*)] = E(\theta^*) - E(\theta^*) = 0.$$

La conclusión a la que llegamos según la relación [1] es la siguiente: el error cuadrático medio, $E(\theta^* - \theta)^2$, resulta igual a la varianza del estimador⁵, $V(\theta^*)$, más el cuadrado de la diferencia entre el valor del parámetro θ y la esperanza matemática del estimador, $[\theta - E(\theta^*)]^2$. De esta descomposición se deduce que un alto valor del error cuadrático medio se puede deber a un alto valor de su varianza, a un alto valor del segundo componente⁶, o a ambas cosas a la vez.

³ Conviene distinguir entre **acuracidad** y **precisión**, términos no equivalentes. La acuracidad hace referencia a la concentración de las estimaciones respecto al verdadero valor del parámetro, la precisión a la concentración de las estimaciones respecto al valor medio del estimador.

⁴ Recordamos que el estimador es variable aleatoria.

⁵ La varianza del estimador mide su **precisión**.

⁶ Posteriormente se verá que la expresión $\theta - E(\theta^*)$ se denomina **sesgo**.

A partir del error cuadrático medio construiremos una buena parte de las propiedades que es razonable exigir a un estimador para ser considerado como «bueno».

3.2.2. PROPIEDADES DE LOS ESTIMADORES

Cuando se llevan a cabo estimaciones se pretende que el resultado se halle lo más cerca posible del valor desconocido del parámetro θ , y una medida de esta pretensión es la minimización del error cometido, expresado en términos del error cuadrático medio. Para que la desviación cuadrática media sea mínima es necesario que los dos sumandos en que se ha descompuesto sean mínimos, dado que la suma de dos números no negativos será mínima cuando lo sean los dos sumandos, es decir que $V(\theta^*)$ (varianza del estimador) sea mínima y $[\theta - E(\theta^*)]^2$ también. Sobre estos dos ejes se construyen las tres primeras propiedades que debe cumplir un estimador para poder considerarlo como «bueno».

La expresión $[\theta - E(\theta^*)]^2$ será mínima cuando valga cero, es decir, cuando la esperanza matemática del estimador sea igual al parámetro θ , es decir, $E(\theta^*) = \theta$. Esto conduce a la propiedad de **insesgadez** coincidiendo, en este caso, el error cuadrático medio con la varianza del estimador.

La condición de mínimo que corresponde a la varianza se logra, para un tamaño de muestra fijo, eligiendo entre todos los posibles estimadores el que presente menor varianza. A esta propiedad se la denomina **eficiencia** del estimador.

Para que las inferencias muestrales resulten correctas, el procedimiento de extracción de la muestra debe conducir a su máxima representatividad y esto se alcanza, también, incrementando el tamaño muestral hasta conseguir que sea «infinito», situación en la que la muestra es la misma población y, por consiguiente, el error cuadrático medio será nulo, por no existir error. De aquí se deduce la propiedad de **consistencia**, que establece el comportamiento probabilístico de los estimadores cuando el tamaño de la muestra es infinito, propiedad límite, y que puede contemplarse como la conveniencia de que la estimación esté próxima al valor desconocido de θ con una probabilidad alta.

Junto a estas tres propiedades, deducidas de la minimización del error cuadrático medio, tenemos otras tres: **suficiencia**, **invarianza** y **robustez**, no inmediatas pero también importantes.

En la generalidad de las situaciones la información que poseemos sobre una población es la proporcionada por una muestra extraída de ella. El estimador que utilizamos para obtener una aproximación del parámetro desconocido conduce a un valor concreto, estimación puntual, valor único que sustituye al conjunto de los valores muestrales individuales. Por otra parte, el tamaño de una muestra suele ser elevado lo que lleva a que sea difícil, costoso, su manejo, por ello es razonable

sustituir los elementos de la muestra por un único valor (estadístico), más manejable que los valores muestrales individuales, deseando que éste valor conserve toda la información contenida en la muestra sobre el parámetro desconocido de la población objeto de estudio. Es evidente que tal deseo es un desideratum, pues es razonable pensar que no siempre tendrá lugar tal situación de salvaguarda de toda la información muestral, sin embargo, cuando un estadístico contiene toda la información proporcionada por la muestra sobre la población resulta más deseable que otro que no disponga de tal propiedad; si ese estadístico existe, recibe el nombre de **suficiente** y la propiedad de **suficiencia**, siendo ésta la cuarta propiedad que se exige a un estimador para poder considerarlo como «bueno». El interés de los estimadores suficientes radica más en sus propiedades óptimas que en la mera reducción de los datos, por muy importante que ésta pueda ser⁷.

La propiedad de **invarianza** se basa en la conveniencia de que obtenido un estimador, θ^* , de un parámetro, θ , el estimador de una función del mismo $g(\theta)$, sea la función del estimador, $g(\theta^*)$, conservándose en esta relación el tipo de transformación representado por la función g .

Cuando se obtienen estimadores es frecuente establecer hipótesis sobre la población bajo estudio, hipótesis que no siempre se cumplen. La propiedad de **robustez** de un estimador se refiere a que las desviaciones de las hipótesis iniciales no afectan a la bondad del estimador, o lo hacen débilmente.

3.3 Estimador insesgado

La función de densidad⁸ en el muestreo de un estimador θ^* , $g(\theta^*; \theta)$, también dependerá del parámetro o parámetros poblacionales, por lo que la esperanza matemática del estimador es igual a

$$E(\theta^*) = \int_{-\infty}^{\infty} \theta^* g(\theta^*; \theta) d\theta^*$$

si su campo de variación es $-\infty < \theta^* < \infty$.

Si tenemos en cuenta que el estimador es función de los elementos muestrales, $\theta^* = \theta^*(x_1, \dots, x_n)$, y que la función de densidad conjunta de la muestra, supuesta

⁷ A esta propiedad se ha dedicado parte del capítulo anterior.

⁸ El planteamiento es similar en el caso discreto.

aleatoria simple de tamaño n , es $f(x_1; \theta) \cdots f(x_n; \theta)$, podemos calcular, también la esperanza del estimador⁹ como

$$\begin{aligned} E(\theta^*) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \theta^*(x_1, \dots, x_n) f(x_1; \theta) \cdots f(x_n; \theta) dx_1 \cdots dx_n = \\ &= \int_{\mathbf{X}} \theta^*(\mathbf{X}) L(\mathbf{X}, \theta) d\mathbf{X}. \end{aligned}$$

En general, la esperanza matemática del estimador de un parámetro dependerá de ese mismo parámetro pudiéndose expresar, en general, como

$$E(\theta^*) = \theta + b(\theta).$$

La cantidad $b(\theta)$ recibe el nombre de **sesgo del estimador**, y es equivalente a un error sistemático, no aleatorio, positivo o negativo. El signo del sesgo de un estimador tiene una interpretación importante. Cuando es positivo indica que si se utiliza tal estimador se tenderá, en media, a sobrestimar el valor del parámetro desconocido, mientras que si es negativo se tenderá a infravalorarlo.

Cuando el sesgo es nulo el estimador se denomina **insesgado**¹⁰ verificándose, por consiguiente, que

$$E(\theta^*) = \theta.$$

En otras palabras, un estimador es insesgado cuando la media de su distribución en el muestreo coincide con el parámetro. La ausencia de sesgo supone la no existencia del error sistemático.

Un estimador es **asintóticamente insesgado** si el sesgo $b(\theta) \rightarrow 0$ cuando $n \rightarrow \infty$.

La insesgades es una propiedad de la variable aleatoria estimador, no de una estimación concreta, y ha de entenderse como propiedad «media», referida al conjunto de todos los valores que pueda presentar el estimador, en el sentido que si tomamos todas las posibles muestras aleatorias simples de un tamaño concreto, se calcula en cada una de ellas el valor del estimador (estimación puntual) y se halla la media de todas las estimaciones, el resultado es igual al valor del parámetro si el estimador es insesgado y distinto de él si no lo es.

La definición de insesgades indica la manera de verificar si un estimador es insesgado, pues no hay más que calcular su esperanza matemática.

⁹ Donde $L(\mathbf{X}, \theta) = f(x_1, \theta) \cdots f(x_n, \theta)$ y $d\mathbf{X} = dx_1 \cdots dx_n$.

¹⁰ También se conoce como **centrado**.

EJEMPLO 1

En la distribución de una variable aleatoria $\xi \sim B(m; p)$, en muestras aleatorias simples de tamaño n , se estima el parámetro p mediante dos estimadores

$$p_1^* = \frac{\bar{x}}{m} \quad \text{y} \quad p_2^* = \frac{\bar{x}}{m+1}$$

Calculamos la esperanza matemática de cada estimador. En el primero

$$E(p_1^*) = E\left(\frac{\bar{x}}{m}\right) = \frac{1}{m} E(\bar{x}) = \frac{1}{m} E(\xi) = \frac{1}{m} mp = p$$

y el estimador es insesgado.

En el segundo

$$E(p_2^*) = E\left(\frac{\bar{x}}{m+1}\right) = \frac{1}{m+1} E(\xi) = \frac{1}{m+1} mp = p - \frac{p}{m+1}$$

siendo sesgado. El signo del sesgo es negativo, implicando que su utilización conducirá en media a infravalorar el valor del parámetro desconocido p , lo que no impide que en algún caso particular, obtuviéramos una estimación puntual que fuera mayor que el valor del parámetro.

Si junto al segundo estimador hubiéramos planteado

$$p_3^* = \frac{\bar{x}}{m-1}$$

sería sesgado con sesgo positivo, $\frac{p}{m-1}$, lo que llevaría a sobrevalorar el parámetro p .

3.3.1. PROPIEDADES DE LOS ESTIMADORES INSESGADOS

1. Si dos estimadores, θ_1^* y θ_2^* , de un parámetro son insesgados, tomamos un número c en el intervalo $(0;1)$ y definimos un nuevo estimador θ^* combinación lineal de los otros dos, el estimador resultante es también insesgado

$$E(\theta_1^*) = E(\theta_2^*) = \theta$$

$$\theta^* = c\theta_1^* + (1-c)\theta_2^*$$

$$E(\theta^*) = E[c\theta_1^* + (1-c)\theta_2^*] = cE(\theta_1^*) + (1-c)E(\theta_2^*) = c\theta + (1-c)\theta = \theta$$

esto quiere decir que a partir de dos estimadores insesgados podemos obtener mediante combinaciones lineales convexas, infinitos estimadores insesgados.

2. El momento muestral de orden r respecto al origen es un estimador insesgado de su correspondiente momento poblacional. En efecto, como según vimos en el epígrafe 1.4

$$E(a_r) = \alpha_r,$$

entonces cuando $\alpha_r^* = a_r$ se verifica que

$$E(\alpha_r^*) = E(a_r) = \alpha_r,$$

es decir, a_r es un estimador insesgado de α_r .

Como caso particular, dado que la media poblacional μ es igual a α_1 , se tendrá que

$$\mu^* = \alpha_1^* = a_1 = \bar{x}$$

obteniéndose que

$$E(\mu^*) = E(\bar{x}) = \alpha_1 = \mu$$

por lo cual, sea cual sea la población, *un estimador insesgado de la media poblacional es la media muestral \bar{x} .*

3. Si se estima la media de una distribución, μ , en muestras aleatorias simples de tamaño n (x_1, \dots, x_n), mediante la combinación lineal

$$\mu^* = \sum_{i=1}^n c_i x_i$$

siendo (c_1, \dots, c_n) números reales, la función muestral es un estimador insesgado de la media poblacional si $\sum_{i=1}^n c_i = 1$, es decir, si dicha combinación lineal es convexa

$$E(\mu^*) = E\left[\sum_{i=1}^n c_i x_i\right] = \sum_{i=1}^n c_i E(x_i) = \sum_{i=1}^n c_i \mu = \mu \sum_{i=1}^n c_i = \mu.$$

Por consiguiente, como caso particular

$$\bar{x} = \sum_{i=1}^n \frac{x_i}{n}, \quad \text{donde} \quad c_i = \frac{1}{n}$$

es un estimador insesgado del parámetro μ , como antes vimos.

4. La varianza de una muestra aleatoria simple es un estimador sesgado de la varianza poblacional, siendo su esperanza¹⁰ igual a

$$E(s^2) = \frac{n-1}{n} \sigma^2 = \sigma^2 - \frac{\sigma^2}{n},$$

donde el sesgo es $-\frac{\sigma^2}{n}$.

Si en vez de tomar s^2 como estimador de σ^2 tomamos la **cuasivarianza muestral**

$$s_1^2 = \frac{n}{n-1} s^2$$

el nuevo estimador es insesgado,

$$E(s_1^2) = E\left[\frac{n}{n-1} s^2\right] = \frac{n}{n-1} \frac{n-1}{n} \sigma^2 = \sigma^2.$$

Tenemos, así, un procedimiento para obtener estimadores insesgados a partir de estimadores sesgados, por simple transformación lineal del estimador sesgado.

5. En los momentos muestrales con respecto a la media, m_r , se obtenía que

$$E(m_r) = \mu_r + O\left(\frac{1}{n}\right)$$

implicando que los momentos muestrales respecto a la media son estimadores sesgados de sus correspondientes momentos poblacionales, siendo el orden del sesgo $O\left(\frac{1}{n}\right)$.

Este sesgo tiende a cero cuando el tamaño de la muestra n se hace suficientemente grande, pues aunque $\mu_r^* = m_r$ no es un estimador insesgado de μ_r , es al menos asintóticamente insesgado.

Como caso particular, recuérdese la varianza muestral analizada en el punto anterior.

6. Entre los momentos centrales en las distribuciones bidimensionales el más importante es la covarianza muestral, m_{11} , cuya esperanza conduce a un estimador sesgado

¹⁰ Véase el capítulo 1.

$$E(m_{11}) = \frac{n-1}{n} \mu_{11} = \mu_{11} - \frac{\mu_{11}}{n}.$$

Para obtener un estimador insesgado, \tilde{m}_{11} , de no hay más que multiplicar m_{11} por la inversa de $\frac{n-1}{n}$

$$E(\tilde{m}_{11}) = E\left(\frac{n}{n-1} m_{11}\right) = \mu_{11}.$$

EJEMPLO 2

En una distribución $N(\mu; \sigma)$ se construye como estimador de μ la función muestral

$$\mu^* = \frac{\sum_{i=1}^n ix_i}{\sum_{i=1}^n i}.$$

Su esperanza matemática es

$$\begin{aligned} E(\mu^*) &= E\left[\frac{\sum_{i=1}^n ix_i}{\sum_{i=1}^n i}\right] = \frac{1}{\sum_{i=1}^n i} \sum_{i=1}^n E(ix_i) = \\ &= \frac{1}{\sum_{i=1}^n i} \sum_{i=1}^n i E(x_i) = \frac{1}{\sum_{i=1}^n i} \mu \sum_{i=1}^n i = \mu \end{aligned}$$

por lo cual el estimador es insesgado.

3.4 Estimador eficiente

Un estimador es eficiente cuando tiene **varianza mínima**. Esta propiedad exige que el estimador que se utilice genere estimaciones parecidas para las diferentes muestras que puedan obtenerse de la población. Al ser pequeña la variabilidad de las estimaciones entre sí, el estimador disfrutará de mayor confianza en su uso que

otro que proporcione estimaciones excesivamente dispares para las distintas muestras posibles.

En principio, para seleccionar el estimador eficiente entre el conjunto de estimadores posibles de un parámetro desconocido θ hay que calcular sus varianzas y comprobar cuál es la menor. Este procedimiento puede ser impracticable si no se es capaz de definir el conjunto de todos los estimadores posibles, por lo que normalmente se investiga la eficiencia de un estimador comparando su varianza con una cantidad, denominada cota de Cramér-Rao, que a continuación va a obtenerse.

3.4.1. COTA DE CRAMÉR-RAO

Sea una población definida por la función de probabilidad $f(x; \theta)$ que contiene el parámetro desconocido θ , estimado mediante θ^* . El estimador recoge la información suministrada por una muestra aleatoria \mathbf{X} de tamaño n . La función de verosimilitud es¹¹

$$L(\mathbf{X}; \theta) = f(x_1, \dots, x_n; \theta)$$

verificándose que

$$\int_{\mathbf{X}} L(\mathbf{X}; \theta) d\mathbf{X} = 1.$$

A partir de la función de verosimilitud se define la variable aleatoria **score**¹²

$$SC(\theta) = \frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta},$$

que presenta las siguientes características¹³

$$E[SC(\theta)] = E\left[\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta}\right] = 0$$

$$V[SC(\theta)] = V\left[\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta}\right] = E\left[\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta}\right]^2 = I(\theta)$$

¹¹ Se utiliza la notación compacta para aligerar la exposición. Recuérdese que,

$$\int_{\mathbf{X}} L(\mathbf{X}; \theta) d\mathbf{X} = \int_{x_1} \dots \int_{x_n} f(x_1, \dots, x_n; \theta) dx_1 \dots dx_n.$$

¹² Véase el capítulo 2.

¹³ Ibídem.

representando $I(\theta)$ la **cantidad de información de Fisher**¹⁴ que la muestra contiene respecto al parámetro θ , siempre y cuando se cumplan las siguientes **condiciones de regularidad de Fisher-Wolfowitz**:

- I. El campo de variación de la población de donde procede la muestra no depende del parámetro θ y, por consiguiente, la muestra \mathbf{X} tampoco¹⁵.
- II. La función $\ln L(\mathbf{X}; \theta)$ admite, por lo menos, las dos primeras derivadas respecto al parámetro θ .
- III. Las operaciones de derivación e integración (o suma en el caso de variable aleatoria discreta) son intercambiables.

Sea un estimador de θ cuya esperanza matemática es

$$E(\theta^*) = \int_{\mathbf{X}} \theta^* L(\mathbf{X}; \theta) d\mathbf{X} = \theta + b(\theta)$$

expresión, que bajo las condiciones de regularidad anteriores, se convierte en

$$\int_{\mathbf{X}} \theta^* \frac{\partial L(\mathbf{X}; \theta)}{\partial \theta} d\mathbf{X} = 1 + b'(\theta). \quad [2]$$

Si ahora se tiene en cuenta la distribución conjunta del estimador θ^* y la variable aleatoria **score**, $SC(\theta)$, por la desigualdad de Schwartz¹⁶ tenemos que

$$\text{cov}^2 \left[\theta^*; \frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta} \right] \leq V(\theta^*) V \left[\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta} \right] \quad [3]$$

y como

$$\begin{aligned} \text{cov} \left[\theta^*; \frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta} \right] &= E \left[\theta^* \frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta} \right] - E(\theta^*) E \left[\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta} \right] = \\ &= E \left[\theta^* \frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta} \right] \end{aligned}$$

pues

$$E \left[\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta} \right] = 0$$

¹⁴ *Ibidem*.

¹⁵ Por ejemplo, una población uniforme con función de densidad $f(x; \theta) = \frac{1}{\theta}$, $0 \leq x \leq \theta$. no cumple esta condición.

¹⁶ MARTÍN PLIEGO y RUIZ-MAYA: *Fundamentos...* Capítulo 4.

llegamos a que la covarianza es igual a $1 + b'(\theta)$, pues

$$\begin{aligned} \text{cov}\left[\theta^*; \frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta}\right] &= E\left[\theta^* \frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta}\right] = \\ &= \int_{\mathbf{X}} \theta^* \frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta} L(\mathbf{X}; \theta) d\mathbf{X} = \\ &= \int_{\mathbf{X}} \theta^* \frac{1}{L(\mathbf{X}; \theta)} \frac{\partial L(\mathbf{X}; \theta)}{\partial \theta} L(\mathbf{X}; \theta) d\mathbf{X} = \\ &= \int_{\mathbf{X}} \theta^* \frac{\partial L(\mathbf{X}; \theta)}{\partial \theta} d\mathbf{X} = \\ &= 1 + b'(\theta) \end{aligned}$$

de acuerdo con [2].

Recordando la definición de cantidad de información de Fisher

$$V\left[\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta}\right] = E\left[\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta}\right]^2 = I(\theta),$$

la desigualdad [3] resulta

$$[1 + b'(\theta)]^2 \leq V(\theta^*) I(\theta)$$

con lo que

$$V(\theta^*) \geq \frac{[1 + b'(\theta)]^2}{I(\theta)} = \frac{[1 + b'(\theta)]^2}{E\left[\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta}\right]^2}.$$

expresión conocida como **cota de Cramér-Rao**, que indica que la varianza de un estimador, para un tamaño de muestra dado, no puede ser menor que esta cota.

En el planteamiento de la cota de Cramér-Rao no se ha especificado nada sobre el tipo de muestreo utilizado, simple o no. Sin embargo, al ser el muestreo que utilizamos el aleatorio simple, estudiamos qué implicaciones tiene en la cota.

Si la muestra es aleatoria simple se verifica la igualdad¹⁷

$$E \left[\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta} \right]^2 = n E \left[\frac{\partial \ln f(\mathbf{X}; \theta)}{\partial \theta} \right]^2$$

y la cota resulta

$$V(\theta^*) \geq \frac{[1 + b'(\theta)]^2}{n E \left[\frac{\partial \ln f(\mathbf{X}; \theta)}{\partial \theta} \right]^2}.$$

Si el estimador es insesgado, $b(\theta) = 0$, la cota se simplifica

$$V(\theta^*) \geq \frac{1}{E \left[\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta} \right]^2}$$

y en muestras aleatorias simples

$$V(\theta^*) \geq \frac{1}{n E \left[\frac{\partial \ln f(\mathbf{X}; \theta)}{\partial \theta} \right]^2}.$$

El denominador de la cota de Cramér-Rao es la cantidad de información de Fisher.

Esta cota depende, en general, del estimador concreto analizado, a través de su sesgo y de la distribución estudiada. Una consecuencia importante en estimadores insesgados, es que la cota únicamente depende de la función de densidad poblacional y del tamaño muestral.

La cota de Cramér-Rao permite determinar si un estimador es eficiente o no. Si su varianza coincide con la cota sabemos que de la clase de estimadores con ese sesgo (o insesgados) ninguno tendrá una varianza menor. De estas palabras no debe deducirse que necesariamente siempre exista un estimador que alcance la cota.

Es preciso destacar que la cota de Cramér-Rao, *CRR*, no tiene por qué tomar siempre un valor muy pequeño (cerca de cero), pues su calidad de mínimo (cota inferior) lo es con respecto a las varianzas de los estimadores, y no con respecto al límite inferior de la varianza, cero.

Un estimador es **asintóticamente eficiente** si

$$\lim_{n \rightarrow \infty} V(\theta^*) = \text{Cota de Cramér-Rao.}$$

¹⁷ Véase el capítulo 4

En la obtención de la cota de Cramér-Rao se ha expuesto, explícitamente, la condición que el campo de variación de la variable aleatoria poblacional no dependiera del parámetro a estimar. La inobservancia de este requisito puede dar lugar a resultados donde las varianzas de los estimadores son menores que las cotas *CCR* indebidamente calculadas.

EJEMPLO 3

En una distribución $N(\mu; \sigma)$ se estima la media poblacional μ mediante la media de una muestra aleatoria simple de tamaño n (x_1, \dots, x_n). El estimador es insesgado y su varianza

$$V(\bar{x}) = \frac{\sigma^2}{n}.$$

Para obtener la cota de Cramér-Rao partimos de la función de densidad poblacional, calculando su logaritmo

$$f(x; \mu) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\ln f(x; \mu) = -\ln \sigma - \frac{1}{2} \ln 2\pi - \frac{(x-\mu)^2}{2\sigma^2}.$$

Derivando respecto al parámetro μ y elevando al cuadrado

$$\frac{\partial \ln f(x; \mu)}{\partial \mu} = \frac{x-\mu}{\sigma^2}$$

$$\left[\frac{\partial \ln f(x; \mu)}{\partial \mu} \right]^2 = \frac{(x-\mu)^2}{\sigma^4}$$

la esperanza matemática es

$$E \left[\frac{\partial \ln f(x; \mu)}{\partial \mu} \right]^2 = E \left[\frac{(x-\mu)^2}{\sigma^4} \right] = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}.$$

Sustituyendo el valor de la esperanza matemática en la expresión de la cota para estimadores insesgados

$$CCR(\mu^*) = \frac{1}{n E \left[\frac{\partial \ln f(x; \mu)}{\partial \mu} \right]^2} = \frac{1}{n \frac{1}{\sigma^2}} = \frac{\sigma^2}{n}.$$

La varianza del estimador coincide con la cota, por lo que concluimos que la media muestral es un estimador eficiente de la media poblacional en la distribución normal.

3.4.2. UNICIDAD DEL ESTIMADOR EFICIENTE

Si existen dos estimadores eficientes con el mismo sesgo son iguales, pues el estimador eficiente es único.

Sean θ_1^* y θ_2^* tales que

$$E(\theta_1^*) = E(\theta_2^*) = \theta + b_0(\theta).$$

Consideramos el estimador de idéntico sesgo

$$\theta^* = \frac{\theta_1^* + \theta_2^*}{2}$$

ya que

$$\begin{aligned} E(\theta^*) &= \frac{1}{2} [E(\theta_1^*) + E(\theta_2^*)] = \\ &= \frac{1}{2} \{[\theta + b_0(\theta)] + [\theta + b_0(\theta)]\} = \\ &= \theta + b_0(\theta) \end{aligned}$$

y definimos las variables

$$d_1 = \theta_1^* - \theta$$

$$d_2 = \theta_2^* - \theta$$

tales que

$$E(d_1)^2 = E(\theta_1^* - \theta)^2 = ECM(\theta_1^*) = V(\theta_1^*) + b_0^2(\theta)$$

$$E(d_2)^2 = E(\theta_2^* - \theta)^2 = ECM(\theta_2^*) = V(\theta_2^*) + b_0^2(\theta)$$

como θ_1^* y θ_2^* son eficientes y tienen el mismo sesgo

$$V(\theta_1^*) = V(\theta_2^*) = V.$$

Consideremos la identidad

$$\left[\frac{d_1 + d_2}{2} \right]^2 + \left[\frac{d_1 - d_2}{2} \right]^2 = \frac{d_1^2 + d_2^2}{2} \quad [4]$$

donde

$$\frac{d_1 + d_2}{2} = \frac{\theta_1^* - \theta + \theta_2^* - \theta}{2} = \frac{\theta_1^* + \theta_2^*}{2} - \theta = \theta^* - \theta$$

$$d_1 - d_2 = (\theta_1^* - \theta) - (\theta_2^* - \theta) = \theta_1^* - \theta_2^*$$

Tomando esperanzas en [4]

$$\begin{aligned} E(\theta^* - \theta)^2 + \frac{1}{4} E(\theta_1^* - \theta_2^*)^2 &= \frac{1}{2} [E(d_1^2) + E(d_2^2)] = \\ &= \frac{1}{2} [V + b_0^2(\theta) + V + b_0^2(\theta)] = \\ &= V + b_0^2(\theta). \end{aligned} \quad [5]$$

Como θ^* en principio no tiene por qué ser eficiente

$$E(\theta^* - \theta)^2 \geq V + b_0^2(\theta),$$

con lo que se tendrá que cumplir en [5]

$$E(\theta_1^* - \theta_2^*)^2 \leq 0$$

cuya única solución es

$$E(\theta_1^* - \theta_2^*)^2 = 0$$

lo que implica la igualdad de los estimadores

$$\theta_1^* = \theta_2^*$$

concluyendo que el estimador eficiente de clase $b_0(\theta)$ es único.

Si $b_0(\theta) = 0$, se obtiene como corolario que si existe un estimador insesgado eficiente, éste es único, y la varianza será mínima cuando sea igual a la cota, es decir, cuando se verifique

$$V(\theta^*) = \frac{1}{I(\theta)}.$$

3.4.3. EFICIENCIA RELATIVA

En ocasiones puede interesar comparar sólo dos estimadores entre sí en cuanto a su dispersión o grado de eficiencia. En este caso no se pretende localizar el mejor de todos los estimadores posibles en relación a esta propiedad sino, simplemente, determinar el mejor entre dos. Para ello se establece la noción de **eficiencia relativa** como cociente de sus correspondientes errores cuadráticos medios.

Si tenemos dos estimadores de un mismo parámetro θ , θ_1^* y θ_2^* , y sus errores cuadráticos medios son

$$ECM(\theta_1^*) = E(\theta_1^* - \theta)^2 = V(\theta_1^*) + b_1^2(\theta)$$

$$ECM(\theta_2^*) = E(\theta_2^* - \theta)^2 = V(\theta_2^*) + b_2^2(\theta)$$

el índice de eficiencia relativa viene dado por

$$e(\theta_1^*; \theta_2^*) = \frac{ECM(\theta_1^*)}{ECM(\theta_2^*)}.$$

Si los estimadores son insesgados la eficiencia relativa es igual al cociente de sus varianzas

$$e(\theta_1^*; \theta_2^*) = \frac{V(\theta_1^*)}{V(\theta_2^*)}.$$

La eficiencia relativa es siempre mayor que cero, pues lo son los errores cuadráticos medios. Es menor que uno si $ECM(\theta_1^*) < ECM(\theta_2^*)$, implicando que deberemos elegir el estimador θ_1^* frente a θ_2^* , y recíprocamente.

3.4.4. ESTIMADORES UMVUE

En el estudio de la propiedad de eficiencia de un estimador aparece como pieza fundamental la cota de Cramér-Rao.

Las posibles situaciones que, respecto a dicha cota, pueden originarse se resumen en el diagrama siguiente:

$$\text{Cota de Cramér-Rao} \begin{cases} \text{Calculable} & \begin{cases} \text{Accesible por algún estimador} \\ \text{No accesible} \end{cases} \\ \text{No calculable} \end{cases}$$

Solamente en un caso, cuando se cumplen las condiciones de regularidad de Fisher-Wolfowitz que permiten calcular dicha cota y cuando exista algún estimador cuya varianza sea igual a la cota (es decir, cuando sea accesible la cota de Cramér-Rao por algún estimador), se puede concluir el análisis de la eficiencia de los estimadores de un parámetro θ . Si la cota no es calculable, o cuando siéndolo no es posible encontrar un estimador con varianza igual a esta cota (no es accesible), es necesario ampliar los métodos de búsqueda del estimador de varianza mínima.¹⁸

¹⁸ Cuya existencia no está asegurada

De todo esto se deduce que el recurso a la cota de Cramér-Rao para el estudio de la eficiencia es una medida parcial para casos concretos, quedando otras situaciones no resueltas.

Una vía de solución general es la construcción de **estimadores insesgados uniformemente de varianza mínima (UMVUE)**¹⁹ definidos de la siguiente manera: consideremos la clase U de estimadores insesgados, θ^* , de un parámetro θ , si un estimador θ_0^* de esta clase verifica, para todo $\theta \in \Theta$, la desigualdad

$$E(\theta_0^* - \theta)^2 \leq E(\theta^* - \theta)^2,$$

diremos que θ_0^* es el estimador insesgado uniformemente de varianza mínima.

La búsqueda de estimadores UMVUE queda fuera del nivel fijado para este manual²⁰; no obstante, se ha querido dejar constancia de su existencia por las dos consideraciones siguientes:

1. La cota de Cramér-Rao puede no ser una solución definitiva en el análisis de la eficiencia de un estimador, bien porque esta cota no pueda calcularse en determinadas poblaciones, bien porque pudiéndose obtener no se encuentre un estimador que alcance dicha cota y con ello se le atribuya el calificativo de estimador eficiente.
2. Existen otras vías de determinación de estimadores eficientes, los UMVUE que si bien limitan la búsqueda del estimador de menor varianza a la clase de estimadores insesgados, implican un mayor nivel de sofisticación. Evidentemente, si un estimador insesgado es eficiente en el sentido Cramér-Rao también puede calificarse de UMVUE, aunque este enfoque alternativo se basa en la pretensión de plantear la determinación de este tipo de estimadores fuera del contexto de esta cota.

3.5 Estimador consistente

A las propiedades de insesgades y eficiencia de los estimadores llegamos minimizando la diferencia entre el estimador y el parámetro, $\theta^* - \theta$, expresada a través del error cuadrático medio, $ECM(\theta^*)$. Estas dos propiedades no están ligadas a un

¹⁹ Uniform Minimum Variante Unbiased Estimator.

²⁰ Véase RUIZ-MAYA y MARTÍN PLIEGO: *Estadística II: Inferencia*. Capítulo 5.

tamaño concreto de la muestra sino a la media de las posibles estimaciones puntuales. La ventaja de que el tamaño muestral aumente radica en que la información proporcionada sobre la población es cada vez mayor llegando, si es infinito, a coincidir la muestra con la población y la estimación puntual con el parámetro.²¹ Resulta evidente la conveniencia de estudiar el comportamiento de los estimadores en función del tamaño muestral.

Expresado formalmente, sea $\theta^* = \theta^*(x_1, \dots, x_n)$ un estimador del parámetro θ de una población con función de densidad o cuantía $f(x; \theta)$, en muestras aleatorias simples de tamaño n . Si damos valores a $n(1, 2, \dots, n, \dots)$ tenemos una sucesión de estimadores $\{\theta_n^*\}$

$$\theta_1^*(x_1); \theta_2^*(x_1, x_2); \dots; \theta_n^*(x_1, \dots, x_n); \dots$$

Dado que un estimador es una variable aleatoria podemos estudiar la sucesión $\{\theta_n^*\}$ bajo los aspectos contemplados en la convergencia de variables aleatorias, es decir, el comportamiento asintótico de la sucesión²².

Un estimador es consistente si converge en probabilidad al parámetro que se quiere estimar, es decir

$$\lim_{n \rightarrow \infty} P(|\theta_n^* - \theta| \geq \varepsilon) = 0$$

o también, si consideramos el suceso complementario del anterior, $|\theta_n^* - \theta| < \varepsilon$,

$$\lim_{n \rightarrow \infty} P(|\theta_n^* - \theta| < \varepsilon) = 1.$$

Cualquiera de estas dos relaciones no es otra cosa que la definición de convergencia en probabilidad de la sucesión $\{\theta_n^*\}$ a la constante θ . El límite en probabilidad del estimador θ_n^* al parámetro θ se designa, también, como

$$\theta_n^* \xrightarrow{p} \theta \quad \text{o} \quad \text{plím } \theta_n^* = \theta$$

Analizamos las consecuencias que tiene la definición de consistencia de un estimador. Aplicando el Teorema de Chebichev a la variable aleatoria estimador θ_n^* se tiene

$$P(|\theta_n^* - \theta| \geq \varepsilon) \leq \frac{E(\theta_n^* - \theta)^2}{\varepsilon^2}.$$

²¹ Teorema de Glivenko-Cantelli. Vid. MARTIN PLIEGO y RUIZ-MAYA: *Fundamentos...* Capítulo 9.

²² *Ibidem.*

Al estudiar el error cuadrático medio llegábamos a su descomposición en dos sumandos no negativos

$$E(\theta_n^* - \theta)^2 = V(\theta_n^*) + b^2(\theta_n)$$

y tomando límites en la desigualdad de Chebichev

$$\lim_{n \rightarrow \infty} P(|\theta_n^* - \theta| \geq \varepsilon) \leq \lim_{n \rightarrow \infty} \frac{E(\theta_n^* - \theta)^2}{\varepsilon^2}$$

$$\lim_{n \rightarrow \infty} E(\theta_n^* - \theta)^2 = \lim_{n \rightarrow \infty} V(\theta_n^*) + \lim_{n \rightarrow \infty} b^2(\theta_n)$$

El resultado anterior dice que para que se cumpla

$$\lim_{n \rightarrow \infty} P(|\theta_n^* - \theta| \geq \varepsilon) = 0$$

es suficiente con que

$$\lim_{n \rightarrow \infty} V(\theta_n^*) = 0 \quad \text{y} \quad \lim_{n \rightarrow \infty} b(\theta_n) = 0$$

equivaliendo la segunda condición a

$$\lim_{n \rightarrow \infty} E(\theta_n^*) = \theta$$

y la primera a

$$\lim_{n \rightarrow \infty} V(\theta_n^*) = 0$$

indicando que en el límite el campo de variación del estimador es un único punto,²² coincidente con el valor desconocido del parámetro θ .

Todo lo anterior supone que si el sesgo y varianza de un estimador tienden a cero cuando el tamaño muestral es infinito el estimador es consistente²³.

3.5.1. PROPIEDADES DE LOS ESTIMADORES CONSISTENTES

Se expone una serie de propiedades de los estimadores consistentes, derivadas de las propiedades de la convergencia en probabilidad:

²² Distribución degenerada o causal: MARTÍN PLIEGO y RUIZ-MAYA: *Fundamentos...* Capítulo 6.

²³ Estrictamente hablando, la consistencia es una propiedad límite de una secuencia de estimadores aunque, para abreviar, se utilice el término estimador consistente.

1. Si $\text{plím } \theta_n^* = \theta$ y g es una función continua se verifica

$$\text{plím } g(\theta_n^*) = g(\theta).$$

es decir, si θ_n^* es un estimador consistente de θ entonces $g(\theta_n^*)$ también lo será respecto a $g(\theta)$.

2. Los momentos muestrales con respecto al origen son estimadores consistentes de los correspondientes poblacionales²⁴, es decir,

$$\lim_{n \rightarrow \infty} P(|a_r - \alpha_r| \geq \epsilon) = 0.$$

En efecto, aplicando el teorema de Chebichev

$$P(|a_r - \alpha_r| \geq \epsilon) \leq \frac{E(a_r - \alpha_r)^2}{\epsilon^2} = \frac{V(a_r)}{\epsilon^2} = \frac{\alpha_{2r} - \alpha_r^2}{n \epsilon^2}$$

y supuesta la existencia de α_{2r} , si $n \rightarrow \infty$, el último término tiende a cero, con lo que está demostrada la consistencia de a_r .

3. Los momentos muestrales centrales, m_r , son estimadores consistentes de los correspondientes poblacionales, μ_r . Según Chebichev

$$P(|m_r - \mu_r| \geq \epsilon) \leq \frac{E(m_r - \mu_r)^2}{\epsilon^2}$$

se demuestra que

$$E(m_r - \mu_r)^{2k} = O\left(\frac{1}{n^k}\right),$$

y si $k = 1$ la esperanza es $O\left(\frac{1}{n}\right)$, por lo tanto, el segundo término de la desigualdad de Chebichev es igual a $\epsilon^{-2}O\left(\frac{1}{n}\right)$, siendo su límite cero cuando $n \rightarrow \infty$, resultando

$$\lim_{n \rightarrow \infty} P(|m_r - \mu_r| \geq \epsilon) = 0.$$

²⁴ Teorema de Kintchine. Vid. MARTÍN PLIEGO y RUIZ-MAYA: *Fundamentos...* Capítulo 9.

EJEMPLO 4

La media muestral es un estimador consistente de la media poblacional, supuesto $\sigma^2 < \infty$,

$$\lim_{n \rightarrow \infty} P(|\bar{x} - \mu| \geq \varepsilon) \leq \lim_{n \rightarrow \infty} \frac{E(\bar{x} - \mu)^2}{\varepsilon^2} = \lim_{n \rightarrow \infty} \frac{V(\bar{x})}{\varepsilon^2} = \lim_{n \rightarrow \infty} \frac{\sigma^2}{n\varepsilon^2} = 0.$$

Podemos seguir otro procedimiento para demostrar la consistencia de la media muestral

$$\begin{aligned} \lim_{n \rightarrow \infty} P(|\bar{x} - \mu| \geq \varepsilon) &= 1 - \lim_{n \rightarrow \infty} P(\mu - \varepsilon \leq \bar{x} \leq \mu + \varepsilon) = \\ &= 1 - \lim_{n \rightarrow \infty} P\left(\mu - \varepsilon \leq \mu + \frac{\sigma\xi}{\sqrt{n}} \leq \mu + \varepsilon\right) = \\ &= 1 - \lim_{n \rightarrow \infty} P\left(-\frac{\varepsilon\sqrt{n}}{\sigma} \leq \xi \leq \frac{\varepsilon\sqrt{n}}{\sigma}\right) = \\ &= 1 - P(-\infty < \xi < \infty) = 1 - 1 = 0. \end{aligned}$$

3.6 Suficiencia

En el capítulo 2 se ha estudiado la necesidad de tomar en consideración el hecho de la posible pérdida de información contenida en una muestra cuando a fin de hacerla operativa, se sustituyen los valores individuales por un estadístico. Se asigna a un estadístico la calificación de suficiente cuando resume el conjunto de información relevante contenida en la muestra, y ningún otro estadístico puede proporcionar información adicional acerca del parámetro desconocido de la población. Al ser el estimador función de los valores muestrales y, por consiguiente, un estadístico, es preciso extender a los estimadores la propiedad de suficiencia.

Si un estimador es suficiente, la distribución de la muestra \mathbf{X} dado el estimador θ^* es independiente del parámetro θ , es decir,

$$f(\mathbf{X}/\theta^*; \theta) = H(\mathbf{X}).$$

EJEMPLO 5

En una distribución de Poisson y en muestras aleatorias simples de tamaño 2 se utiliza como estimador del parámetro λ la función muestral $T = 2x_1 + 3x_2$. Se extrae una muestra ($x_1 = 3, x_2 = 0$) y queremos verificar si el estimador es suficiente.

El estimador T toma en la muestra el valor 6, por tanto, necesitamos comprobar si la probabilidad del suceso $\{\xi_1 = 3, \xi_2 = 0/T = 6\}$ es independiente del parámetro λ

$$\begin{aligned}
 P\{\xi_1 = 3, \xi_2 = 0/T = 6\} &= \frac{P\{\xi_1 = 3, \xi_2 = 0; T = 6\}}{P(T = 6)} = \frac{P(\xi_1 = 3, \xi_2 = 0)}{P(T = 6)} \\
 &= \frac{P(\xi_1 = 3, \xi_2 = 0)}{P[(\xi_1 = 3, \xi_2 = 0) \cup (\xi_1 = 0, \xi_2 = 2)]} = \\
 &= \frac{P(\xi_1 = 3) P(\xi_2 = 0)}{P(\xi_1 = 3) P(\xi_2 = 0) + P(\xi_1 = 0) P(\xi_2 = 2)} = \\
 &= \frac{\frac{e^{-\lambda}\lambda^3}{3!} \frac{e^{-\lambda}\lambda^0}{0!}}{\frac{e^{-\lambda}\lambda^3}{3!} \frac{e^{-\lambda}\lambda^0}{0!} + \frac{e^{-\lambda}\lambda^0}{0!} \frac{e^{-\lambda}\lambda^2}{2!}} = \frac{\lambda}{\lambda + 3}.
 \end{aligned}$$

la probabilidad no es independiente de λ por lo que T no es suficiente como estimador del parámetro.

3.6.1. CRITERIO DE FACTORIZACIÓN DE FISHER-NEYMAN

La aplicación de la definición de suficiencia para determinar si un estadístico es suficiente es poco operativa, por lo cual se utiliza como procedimiento el criterio de factorización de Fisher-Neyman,²⁵ de más sencilla aplicación, estableciendo que para que un estadístico (estimador) sea suficiente la función de verosimilitud de la muestra, $L(\mathbf{X}; \theta)$, debe descomponerse en el producto de dos funciones, una de ellas, $g(\theta^*, \theta)$, dependiente del parámetro θ y de la muestra a través del estimador θ^* , y la otra, $H(\mathbf{X})$, no dependiente del parámetro θ

$$L(\mathbf{X}; \theta) = g(\theta^*; \theta) \cdot H(\mathbf{X}).$$

El criterio de factorización de Fisher-Neyman puede contemplarse desde un punto de vista operativo en la búsqueda de estimadores suficientes. Tal y como se ha planteado lo utilizamos para determinar si es suficiente un estimador conocido, ahora bien, si la función de densidad conjunta de la muestra puede descomponerse de la manera indicada por el criterio de Fisher-Neyman, el estadístico θ^* que aparece en la función $g(\theta^*, \theta)$ es un estimador suficiente del parámetro θ .

²⁵ Véase el capítulo 2.

EJEMPLO 6

En la distribución $N(\mu; \sigma)$, siendo conocida σ , la media muestral es un estimador suficiente de la media poblacional, en muestras aleatorias simples de tamaño n .

La función de densidad conjunta de la muestra es

$$f(x_1, \dots, x_n; \mu) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x_1 - \mu)^2}{2\sigma^2}} \dots \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x_n - \mu)^2}{2\sigma^2}} =$$

$$= [\sigma \sqrt{2\pi}]^{-n} e^{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}}$$

desarrollamos el exponente

$$\sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n x_i^2 + n \mu^2 - 2n \mu \bar{x} = n (\mu^2 - 2\bar{x}\mu) + \sum_{i=1}^n x_i^2$$

definimos las funciones

$$g(\bar{x}; \mu) = e^{-\frac{n(\mu^2 - 2\bar{x}\mu)}{2\sigma^2}}$$

y

$$H(x_1, \dots, x_n) = [\sigma \sqrt{2\pi}]^{-n} e^{-\frac{\sum_{i=1}^n x_i^2}{2\sigma^2}}$$

llegando a la descomposición de la función de densidad conjunta de la muestra en la forma

$$L(\mathbf{X}; \mu) = f(x_1, \dots, x_n; \mu) = g(\bar{x}; \mu) H(x_1, \dots, x_n)$$

por lo que la media muestral es un estimador suficiente de la media poblacional.

3.7 Estimador invariante

Si el estimador de θ es θ^* sería deseable que, por ejemplo, los estimadores de $\theta + k$, $k\theta$ y θ^2 fueran, respectivamente, $\theta^* + k$, $k\theta^*$ y $(\theta^*)^2$. Es decir, que el estimador de la función de un parámetro, $g(\theta)$, sea igual a la función del estimador del parámetro, $[g(\theta)]^* = g(\theta^*)$. Cuando esto sucede, diremos que el **estimador es invariante**.²⁶

²⁶ En el capítulo siguiente se verá que el método de estimación de la máxima verosimilitud es invariante.

Por ejemplo, si para estimar la varianza poblacional σ^2 se utiliza como estimador la varianza muestral, s^2 , entonces, para estimar la desviación típica poblacional, σ , el método utilizado, si es invariante, debería proporcionar como resultado que $\sigma^* = s$, desviación típica muestral.

En realidad los estimadores en sí no pueden calificarse de invariantes, sino más bien debe ser el método de estimación que se utilice el que ha de ser invariante ante cualquier transformación.

La invarianza y la insesgaredad de estimadores de parámetros ligados funcionalmente por algún tipo de transformación no puede lograrse simultáneamente. Por ejemplo, si el método para estimar θ y θ^2 es invariante y, además, los estimadores de cada parámetro son insesgados, tendríamos que $E(\theta^*) = \theta$ y $E[(\theta^*)^2] = \theta^2$, sin embargo, $V(\theta^*) = E[(\theta^*)^2] - [E(\theta^*)]^2 = \theta^2 - \theta^2 = 0$, lo que obliga a que el estimador θ^* tenga una distribución degenerada.

3.8 Estimador robusto

Un estimador es robusto cuando vulneraciones de los supuestos de partida en que se base el proceso de estimación no alteran de manera significativa los resultados que proporciona. Los aspectos más usuales que atañen a la robustez se refieren al alejamiento de la función de distribución atribuida a una variable aleatoria respecto de la verdadera, desconocida.

Supongamos que se desea estimar el parámetro θ de una población que creemos presenta una distribución de probabilidad $F_1(x; \theta)$. Se extrae una muestra aleatoria simple de tamaño n y se construye un estimador θ^* cuya distribución de probabilidad en el muestreo, $G_1(\theta^*; \theta)$, deriva de la poblacional $F_1(x; \theta)$.

Si la distribución de la población no es $F_1(x; \theta)$ sino $F_2(x; \theta)$, la distribución del estimador ya no sería $G_1(\theta^*; \theta)$ sino $G_2(\theta^*; \theta)$. Entonces si $G_1(\theta^*; \theta)$ no se aleja demasiado de $G_2(\theta^*; \theta)$, es de esperar que este cambio en las condiciones de partida no altere sustancialmente el proceso de estimación, por lo que las consecuencias no serían excesivamente desfavorables y el estimador obtenido inicialmente sería robusto. En caso contrario, si $G_1(\theta^*; \theta)$ es bastante dispar con respecto a $G_2(\theta^*; \theta)$ al llevar a cabo inferencias con θ^* se cometerían errores de consideración.

Citamos dos ejemplos. Supongamos que el estadístico $T(\mathbf{X})$ sigue la distribución $N(1; 1)$, aunque realmente la distribución correcta es $\chi^2(1)$. La probabilidad del suceso $\{T(\mathbf{X}) \geq 1\}$ en el primer caso $P(N(1; 1) \geq 1)$ es 0,5 y en el segundo $P(\chi^2(1) \geq 1) = 0,345$. La diferencia de las probabilidades indica, con claridad, lo que supone el tomar como distribución poblacional una u otra distribución.

Si no se conoce la varianza en una población $N(\mu; \sigma)$ se recurre a la distribución t de Student

$$t(n-1) = \frac{\bar{x} - \mu}{\sqrt{s^2}} \sqrt{n-1}$$

alteraciones moderadas en la normalidad respecto a $N(\mu; \sigma)$ no afectan seriamente a las inferencias basadas en $t(n-1)$, sobre todo para valores altos de n , indicando que los procedimientos sustentados en el estadístico t son robustos. Sin embargo, la no normalidad afecta seriamente las inferencias que se realicen sobre σ^2 basadas en $\chi^2(n-1)$.

Ejercicios resueltos

EJERCICIO 3.1 Consideremos una variable aleatoria $N(\mu; \sigma)$ y dos estimadores de la media poblacional μ , en muestras aleatorias simples de tamaño n ,

$$\mu_1^* = \frac{x_1 + \dots + x_n}{n-1} \quad \text{y} \quad \mu_2^* = \frac{x_1 + \dots + x_n}{n+1}.$$

Selecciónese el mejor de ellos teniendo en cuenta su error cuadrático medio.

SOLUCION. Hemos visto que

$$ECM(\mu^*) = E(\mu^* - \mu)^2 = V(\mu^*) + [\mu - E(\mu^*)]^2$$

y, para cada estimador, las expresiones de $E(\mu^*)$, $V(\mu^*)$ son

$$E(\mu_1^*) = \frac{n}{n-1}\mu; \quad V(\mu_1^*) = \frac{n}{(n-1)^2}\sigma^2$$

$$E(\mu_2^*) = \frac{n}{n+1}\mu; \quad V(\mu_2^*) = \frac{n}{(n+1)^2}\sigma^2$$

$$\begin{aligned} ECM(\mu_1^*) &= E(\mu_1^* - \mu)^2 = V(\mu_1^*) + [\mu - E(\mu_1^*)]^2 = \\ &= \frac{n\sigma^2}{(n-1)^2} + \left[\mu - \frac{n}{n-1}\mu \right]^2 = \frac{n\sigma^2 + \mu^2}{(n-1)^2}. \end{aligned}$$

$$\begin{aligned} ECM(\mu_2^*) &= E(\mu_2^* - \mu)^2 = V(\mu_2^*) + [\mu - E(\mu_2^*)]^2 = \\ &= \frac{n\sigma^2}{(n+1)^2} + \left[\mu - \frac{n}{n+1}\mu \right]^2 = \frac{n\sigma^2 + \mu^2}{(n+1)^2}. \end{aligned}$$

Para decidir cuál de los dos estimadores es preferible suponemos que

$$ECM(\mu_1^*) > ECM(\mu_2^*),$$

y, si es así, se ha de verificar que

$$\frac{n\sigma^2 + \mu^2}{(n-1)^2} > \frac{n\sigma^2 + \mu^2}{(n+1)^2}.$$

La desigualdad siempre tiene lugar pues, operando, llegamos a que

$$(n + 1)^2 > (n - 1)^2$$

por lo que se deberá elegir, en este caso, el segundo estimador como «mejor» de los dos.

EJERCICIO 3.2 Una variable aleatoria ξ tiene la función de densidad

$$f(x; \theta) = \theta e^{-\theta x}, \quad \text{para } x \geq 0 \quad \text{y} \quad \theta > 0.$$

Se estima el parámetro θ mediante dos estimadores diferentes en muestras aleatorias simples de tamaño dos

$$\theta_1^* = \frac{2x_1 + 4x_2}{3}; \quad \theta_2^* = \frac{4x_1 + 5x_2}{3}.$$

Compárense a través de sus errores cuadráticos medios.

SOLUCIÓN. Las esperanzas y varianzas de los estimadores son

$$E(\theta_1^*) = E\left(\frac{2x_1 + 4x_2}{3}\right) = 2E(\xi)$$

$$V(\theta_1^*) = V\left(\frac{2x_1 + 4x_2}{3}\right) = \frac{1}{9}[4V(\xi) + 16V(\xi)] = \frac{20}{9}V(\xi)$$

$$E(\theta_2^*) = E\left(\frac{4x_1 + 5x_2}{3}\right) = 3E(\xi)$$

$$V(\theta_2^*) = V\left(\frac{4x_1 + 5x_2}{3}\right) = \frac{41}{9}V(\xi)$$

necesitando obtener la esperanza y varianza de la variable aleatoria poblacional, que como puede comprobarse, son

$$E(\xi) = \frac{1}{\theta};$$

$$V(\xi) = \frac{1}{\theta^2}$$

resultando las correspondientes de los estimadores iguales a

$$E(\theta_1^*) = \frac{2}{\theta}; \quad E(\theta_2^*) = \frac{3}{\theta}$$

$$V(\theta_1^*) = \frac{20}{9\theta^2}; \quad V(\theta_2^*) = \frac{41}{9\theta^2}$$

los errores cuadráticos medios respectivos son

$$\begin{aligned} ECM(\theta_1^*) &= E(\theta_1^* - \theta)^2 = V(\theta_1^*) + [\theta - E(\theta_1^*)]^2 = \\ &= \frac{20}{9\theta^2} + \left[\theta - \frac{2}{\theta}\right]^2 = \frac{9\theta^4 - 36\theta^2 + 56}{9\theta^2} \end{aligned}$$

$$\begin{aligned} ECM(\theta_2^*) &= E(\theta_2^* - \theta)^2 = V(\theta_2^*) + [\theta - E(\theta_2^*)]^2 = \\ &= \frac{41}{9\theta^2} + \left[\theta - \frac{3}{\theta}\right]^2 = \frac{9\theta^4 - 54\theta^2 + 122}{9\theta^2} \end{aligned}$$

Para verificar la influencia del parámetro sobre los errores cuadráticos medios bastará comprobar si, por ejemplo, se cumple la siguiente desigualdad para todo valor de θ

$$\begin{aligned} E(\theta_1^* - \theta)^2 &< E(\theta_2^* - \theta)^2 \\ \frac{9\theta^4 - 36\theta^2 + 56}{9\theta^2} &< \frac{9\theta^4 - 54\theta^2 + 122}{9\theta^2} \\ \theta^2 &< \frac{11}{3} \end{aligned}$$

por lo que la relación entre los dos errores cuadráticos medios se cumple sólo para los valores de θ^2 inferiores a $\frac{11}{3}$, es decir, el valor de θ , desconocido, influye en la elección del estimador, como se aprecia en la figura 3.2.

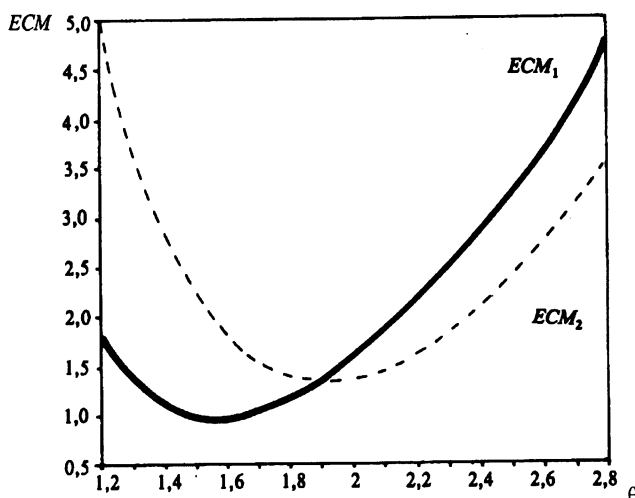


FIGURA 3.2

Se puede comprobar que

$$\text{Para } \theta^2 < \frac{11}{3} \quad \text{es} \quad ECM(\theta_1^*) < ECM(\theta_2^*)$$

$$\text{Para } \theta^2 > \frac{11}{3} \quad \text{es} \quad ECM(\theta_1^*) > ECM(\theta_2^*)$$

EJERCICIO 3.3 Para estimar el parámetro p de una variable aleatoria $B(1; p)$, en muestras aleatorias simples de tamaño 3, se establece el estimador

$$p^* = \frac{2x_1 + x_2 + 4x_3}{10}.$$

Estúdiese si dicho estimador es insesgado.

SOLUCIÓN La esperanza matemática de p^* se puede obtener de dos formas: bien mediante la distribución de probabilidad de la variable aleatoria poblacional junto con las propiedades de la esperanza matemática, bien calculando la función de cuantía en el muestreo de p^* y hallando, a continuación, su esperanza matemática.

Primer procedimiento

$$\begin{aligned} E(p^*) &= E\left[\frac{2x_1 + x_2 + 4x_3}{10}\right] = \\ &= \frac{1}{10}[2E(x_1) + E(x_2) + 4E(x_3)] = \\ &= \frac{1}{10}(2p + p + 4p) = \frac{7p}{10} = \\ &= p - \frac{3}{10}p \neq p. \end{aligned}$$

Al no coincidir $E(p^*)$ con el parámetro p , el estimador planteado es sesgado y el sesgo igual a $-3p$ indicando que, en media, el estimador subestima el valor de p .

Segundo procedimiento

Calculamos la distribución de probabilidad en el muestreo de p^* . Hay ocho muestras aleatorias simples distintas, hallamos para cada una de ellas la correspondiente estimación y su probabilidad. Todos los valores aparecen en el cuadro siguiente

Clases de muestras	p^*	Probabilidades de obtención de cada clase de muestras
(0;0;0)	0	$(1-p)^3$
(1;0;0)	2/10	$p(1-p)^2$
(0;1;0)	1/10	$p(1-p)^2$
(0;0;1)	4/10	$p(1-p)^2$
(1;1;0)	3/10	$p^2(1-p)$
(1;0;1)	6/10	$p^2(1-p)$
(0;1;1)	5/10	$p^2(1-p)$
(1;1;1)	7/10	p^3

La esperanza matemática del estimador p^* es igual a

$$\begin{aligned}
 E(p^*) &= \sum_{i=1}^8 p_i^* P(p^* = p_i^*) = \\
 &= 0(1-p)^3 + \frac{1}{10} p(1-p)^2 + \frac{2}{10} p(1-p)^2 + \dots + \frac{7}{10} p^3 = \\
 &= \frac{7p}{10} \neq p
 \end{aligned}$$

y el estimador es sesgado.

EJERCICIO 3.4 La variable aleatoria ξ que tiene por función de densidad

$$f(x) = \theta e^{-\theta x}, \quad x \geq 0, \quad \theta > 0$$

se considera como estimador, en muestras aleatorias simples de tamaño 2, la función muestral

$$\theta^* = \frac{x_1 + 3x_2}{4}.$$

Compruébese si el estimador es insesgado.

SOLUCIÓN. La esperanza matemática del estimador es

$$E(\theta^*) = E\left[\frac{x_1 + 3x_2}{4}\right] = \frac{1}{4}[E(x_1) + 3E(x_2)] = E(\xi)$$

y la de la variable aleatoria ξ

$$E(\xi) = \int_0^{\infty} \theta x e^{-\theta x} dx = \frac{1}{\theta} = \theta + \frac{1 - \theta^2}{\theta},$$

es decir,

$$E(\theta^*) = \frac{1}{\theta} = \theta + \frac{1 - \theta^2}{\theta}.$$

El estimador es sesgado por no coincidir su esperanza matemática con el parámetro θ , y el sesgo será positivo o negativo dependiendo del valor de θ , positivo cuando $\theta < 1$ y negativo siempre que $\theta > 1$.

También podemos calcular la esperanza matemática del estimador a partir de la distribución en el muestreo de θ^* , cuyo campo de variación es $\theta^* \geq 0$,²⁷

$$\begin{aligned} G(\theta^*; \theta) &= P(\eta \leq \theta^*) = P\left[\frac{\xi_1 + 3\xi_2}{4} \leq \theta^*\right] = P(\xi_1 + 3\xi_2 \leq 4\theta^*) = \\ &= \int_0^{4\theta^*} \int_0^{\frac{4\theta^* - x_1}{3}} \theta e^{-\theta x_1} \theta e^{-\theta x_2} dx_1 dx_2 = \\ &= 1 - \frac{3}{2} e^{-\frac{4}{3}\theta\theta^*} + \frac{1}{2} e^{-4\theta\theta^*}. \end{aligned}$$

La función de densidad del estimador, $g(\theta^*; \theta) = G'(\theta^*; \theta)$, es

$$g(\theta^*; \theta) = 2\theta \left[e^{-\frac{4}{3}\theta\theta^*} - e^{-4\theta\theta^*} \right], \quad \theta^* \geq 0,$$

por tanto

$$E(\theta^*) = \int_0^{\infty} \theta^* g(\theta^*; \theta) d\theta^* = \int_0^{\infty} \theta^* 2\theta \left[e^{-\frac{4}{3}\theta\theta^*} - e^{-4\theta\theta^*} \right] d\theta^* = \frac{1}{\theta}.$$

EJERCICIO 3.5 En la distribución $N(\mu; \sigma)$ y en muestras aleatorias simples de tamaño n , se consideran como estimadores de los parámetros μ y σ^2 la media y la cuasivarianza muestrales, respectivamente. Estúdiese su eficiencia.

SOLUCIÓN. Sabemos²⁸ que

$$V(\bar{x}) = \frac{\sigma^2}{n} \quad \text{y} \quad V(s_1^2) = \frac{2\sigma^4}{n-1}.$$

²⁷ El estimador θ^* puede ser cero, ya que la variable también puede tomar este valor, sin embargo, debemos rechazar que θ^* sea igual a cero pues en el enunciado se establece que $\theta > 0$.

²⁸ Véase el capítulo 1.

La matriz de información correspondiente al parámetro μ y σ^2 en la distribución $N(\mu; \sigma)$ es²⁸

$$I(\theta) = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix}$$

y su inversa

$$[I(\theta)]^{-1} = \begin{bmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{bmatrix}$$

por consiguiente,

$$V(\bar{x}) = CCR(\mu^*) = \frac{\sigma^2}{n}$$

y la media muestral es un estimador eficiente de μ , mientras que la cuasivarianza muestral no lo es de σ^2 , pues

$$V(s_1^2) = \frac{2\sigma^4}{n-1} > CCR[(\sigma^2)^*] = \frac{2\sigma^4}{n}.$$

EJERCICIO 3.6 En la distribución $B(m; p)$ se considera como estimador del parámetro p el estadístico $p^* = \frac{\bar{x}}{m}$, siendo \bar{x} la media muestral en muestras aleatorias simples de tamaño n . Hállese la cota de Cramér-Rao y estúdiense la eficiencia del estimador.

SOLUCIÓN. En primer lugar, determinamos si el estimador es insesgado

$$E(p^*) = E\left[\frac{\bar{x}}{m}\right] = \frac{mp}{m} = p.$$

Al coincidir la esperanza del estimador con el parámetro, el estimador es insesgado, y el numerador de la cota de Cramér-Rao es la unidad.

²⁸ Véase el capítulo 2, ejercicio 2.4.

Calculamos, a continuación, el denominador de la cota. La función de cuantía de la distribución binomial es

$$f(x; p) = \binom{m}{x} p^x (1-p)^{m-x}$$

su logaritmo neperiano y derivada respecto a p

$$\ln f(x; p) = \ln \binom{m}{x} + x \ln p + (m-x) \ln (1-p)$$

$$\frac{\partial \ln f(x; p)}{\partial p} = \frac{x}{p} - \frac{m-x}{1-p} = \frac{x-mp}{p(1-p)}$$

donde

$$E \left[\frac{\partial \ln f(x; p)}{\partial p} \right]^2 = \frac{E(x-mp)^2}{p^2(1-p)^2}$$

y teniendo en cuenta que $E(\xi) = mp$, $E(\xi - mp)^2 = V(\xi) = mp(1-p)$

$$E \left[\frac{\partial \ln f(x; p)}{\partial p} \right]^2 = \frac{mp(1-p)}{p^2(1-p)^2} = \frac{m}{p(1-p)}$$

y la cota de Cramér-Rao

$$CCR = \frac{1}{nE \left[\frac{\partial \ln f(x; p)}{\partial p} \right]^2} = \frac{1}{n \frac{m}{p(1-p)}} = \frac{p(1-p)}{nm}$$

La varianza del estimador es

$$V(p^*) = V \left[\frac{\bar{x}}{m} \right] = \frac{1}{m^2} V(\bar{x}) = \frac{1}{m^2} \frac{mp(1-p)}{n} = \frac{p(1-p)}{nm}$$

La varianza del estimador coincide con la cota de Cramér-Rao por lo cual el estimador estudiado es eficiente.

EJERCICIO 3.7 En la distribución de Poisson estimamos el parámetro λ mediante la media de una muestra aleatoria simple de tamaño n . Estúdiese su eficiencia.

SOLUCIÓN. La media muestral es un estimador insesgado de la media poblacional,²⁹ $E(\bar{x}) = \lambda$, y su varianza la de la población dividida por el tamaño de la muestra,

n , $V(\bar{x}) = \frac{\lambda}{n}$. Indicamos el cálculo de la cota de Cramér-Rao obteniendo previamente

el logaritmo de la función de cuantía de Poisson

$$\ln P(\xi = x; \lambda) = -\lambda + x \ln \lambda + \ln x!$$

y su derivada respecto a λ

$$\frac{\partial \ln P(\xi = x; \lambda)}{\partial \lambda} = -1 + \frac{x}{\lambda} = \frac{x - \lambda}{\lambda}$$

$$E \left[\frac{\partial \ln P(\xi = x; \lambda)}{\partial \lambda} \right]^2 = \frac{E(x - \lambda)^2}{\lambda^2} = \frac{V(\xi)}{\lambda^2} = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$

$$CCR = \frac{1}{n E \left[\frac{\partial \ln P(\xi = x; \lambda)}{\partial \lambda} \right]^2} = \frac{\lambda}{n}$$

La varianza del estimador es igual a $\frac{\lambda}{n}$, por lo cual, es eficiente.

EJERCICIO 3.8 En la función de densidad gamma $G(p, q)$ estímesse el parámetro q mediante la media de una muestra aleatoria simple de tamaño n y estúdiense su eficiencia.

SOLUCIÓN. La función de densidad de la variable aleatoria es³⁰

$$f(x; p, q) = \frac{q^p}{\Gamma(p)} x^{p-1} e^{-qx},$$

su esperanza es igual a $E(\xi) = \frac{p}{q}$ y la varianza $V(\xi) = \frac{p}{q^2}$.

Determinamos si el estimador es insesgado

$$E(q^*) = E(\bar{x}) = E(\xi) = \frac{p}{q} = q + \frac{p - q^2}{q}$$

²⁹ Véase el capítulo 1.

³⁰ MARTÍN PLIEGO y RUIZ-MAYA: *Fundamentos...* Capítulo 7.

el sesgo es

$$\frac{p - q^2}{q}$$

y el numerador de la cota igual a

$$[1 + b'(q)]^2 = \frac{p^2}{q^4}.$$

El logaritmo de la función de densidad

$$\ln f(x; p, q) = p \ln q - \ln \Gamma(p) + (p - 1) \ln x - qx$$

y la derivada respecto al parámetro q

$$\frac{\partial \ln f(x; p, q)}{\partial q} = \frac{p}{q} - x$$

$$E \left[\frac{\partial \ln f(x; p, q)}{\partial q} \right]^2 = E \left[\frac{p}{q} - x \right]^2 = E \left[x - \frac{p}{q} \right]^2 = V(\xi) = \frac{p}{q^2}$$

resultando la cota

$$CCR = \frac{[1 + b'(q)]^2}{n E \left[\frac{\partial \ln f(x; p, q)}{\partial q} \right]^2} = \frac{\frac{p^2}{q^4}}{n \frac{p}{q^2}} = \frac{p}{nq^2}.$$

Como la varianza del estimador, \bar{x} , es

$$V(\bar{x}) = \frac{V(\xi)}{n} = \frac{p}{nq^2}$$

el estimador es eficiente dentro de la clase de estimadores con sesgo

$$\frac{p}{q} - q.$$

EJERCICIO 3.9 En una distribución $N(\mu; \sigma)$ se estima la media poblacional μ en una muestra aleatoria simple de tamaño n (x_1, \dots, x_n) por medio de la función muestral

$$\mu^* = \frac{\sum_{i=1}^n ix_i}{n}.$$

Estúdiese la eficiencia de este estimador.

SOLUCIÓN. Para hallar la cota de Cramér-Rao necesitamos saber, en primer lugar, si el estimador es insesgado o no, calculando para ello su esperanza matemática

$$\begin{aligned} E(\mu^*) &= E\left[\frac{\sum_{i=1}^n ix_i}{n}\right] = \frac{1}{n} \sum_{i=1}^n E(ix_i) = \frac{1}{n} \sum_{i=1}^n i E(x_i) = \\ &= \frac{\mu}{n} \sum_{i=1}^n i = \frac{n+1}{2} \mu = \mu + \frac{n-1}{2} \mu. \end{aligned}$$

El estimador es sesgado y sobrestima, en media, el parámetro. La varianza del estimador es igual a

$$\begin{aligned} V(\mu^*) &= V\left[\frac{\sum_{i=1}^n ix_i}{n}\right] = \frac{1}{n^2} \sum_{i=1}^n i^2 V(x_i) = \\ &= \frac{\sigma^2}{n^2} \sum_{i=1}^n i^2 = \frac{\sigma^2}{n^2} \frac{n(n+1)(2n+1)}{6} = \\ &= \frac{\sigma^2(n+1)(2n+1)}{6n}, \end{aligned}$$

siendo

$$CCR(\mu^*) = \frac{\sigma^2}{n} = \frac{(n+1)^2}{4} \frac{\sigma^2}{n}.$$

puesto que dado que el sesgo $b(\mu) = \frac{n-1}{2} \mu$

$$CCR = \frac{[1 + b'(\mu)]^2}{I(\mu)} = \frac{\left[1 + \frac{n-1}{2}\right]^2}{\frac{n}{\sigma^2}} = \frac{(n+1)^2}{4} \frac{\sigma^2}{n}$$

al ser

$$I(\mu) = \frac{n}{\sigma^2} \quad (\text{ejercicio 3.5})$$

Comparando las dos expresiones, llegamos a la conclusión que al ser la varianza del estimador mayor que la cota de Cramér-Rao el estimador no es eficiente, siempre que $n > 1$.

EJERCICIO 3.10 Una variable aleatoria está definida en el intervalo $0 \leq x \leq \theta$ y tiene la función de densidad

$$f(x; \theta) = \frac{3}{\theta^3} x^2.$$

Consideremos el estimador $\theta^* = \frac{4\bar{x}}{3}$ en muestras aleatorias simples de tamaño n . Estúdiense si es insesgado y eficiente.

SOLUCIÓN. Calculamos la media y la varianza de la población

$$E(\xi) = \int_0^\theta \frac{3}{\theta^3} x^3 dx = \frac{3}{4} \theta$$

$$V(\xi) = \int_0^\theta \left(x - \frac{3}{4} \theta\right)^2 \frac{3}{\theta^3} x^2 dx = \frac{3}{80} \theta^2$$

y las del estimador

$$E(\theta^*) = E\left[\frac{4}{3} \bar{x}\right] = \frac{4}{3} E(\xi) = \frac{4}{3} \frac{3}{4} \theta = \theta$$

$$V(\theta^*) = V\left[\frac{4}{3} \bar{x}\right] = \frac{16}{9} \frac{V(\xi)}{n} = \frac{16}{9} \frac{3}{80} \frac{\theta^2}{n} = \frac{\theta^2}{15n}$$

El estimador es insesgado y la falsa cota de Cramér-Rao, si se calculara ignorando las condiciones de regularidad, es igual a $\frac{\theta^2}{9n}$ que resulta mayor que la varianza del estimador. La explicación de este hecho, radica en que la CCR se ha hallado indebidamente, pues el campo de variación de la variable aleatoria poblacional no es independiente del parámetro desconocido θ .

EJERCICIO 3.11 Compruébese que la varianza muestral, s^2 , es un estimador consistente de la varianza poblacional, σ^2 .

SOLUCIÓN. Como

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2 = a_2 - a_1^2$$

al ser los momentos muestrales estimadores consistentes de los poblacionales

$$\text{plim } a_1 = \alpha_1 = \mu \quad \text{y} \quad \text{plim } a_2 = \alpha_2$$

por las propiedades expuestas y ser la varianza muestral función de los momentos,

$$s^2 = a_2 - a_1^2$$

resulta que

$$\text{plim } s^2 = \text{plim } a_2 - \text{plim } a_1^2 = \alpha_2 - \alpha_1^2 = \sigma^2.$$

EJERCICIO 3.12 La variable aleatoria ξ tiene por función de densidad

$$f(x; \theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}; \quad x \geq 0, \theta > 0.$$

Estudiese la consistencia del estimador de θ , $\theta^* = \bar{x}$.

SOLUCIÓN. Como primer paso del estudio de la consistencia del estimador verificamos su insesgader

$$E(\theta^*) = E(\bar{x}) = E(\xi) = \theta$$

pues

$$E(\xi) = \int_0^{\infty} \frac{x}{\theta} e^{-\frac{x}{\theta}} dx = \theta$$

siendo, por tanto, el estimador insesgado.

La varianza de la variable poblacional es

$$E(\xi^2) = \int_0^{\infty} \frac{x^2}{\theta} e^{-\frac{x}{\theta}} dx = 2\theta^2$$

$$V(\xi) = E(\xi^2) - [E(\xi)]^2 = 2\theta^2 - \theta^2 = \theta^2$$

y la del estimador

$$V(\theta^*) = V(\bar{x}) = \frac{V(\xi)}{n} = \frac{\theta^2}{n}.$$

Pasando a la condición de consistencia

$$\lim_{n \rightarrow \infty} P(|\theta^* - \theta| \geq \varepsilon) = \lim_{n \rightarrow \infty} \frac{E(\theta^* - \theta)^2}{\varepsilon^2} = \lim_{n \rightarrow \infty} \frac{V(\theta^*)}{\varepsilon^2} = \lim_{n \rightarrow \infty} \frac{\theta^2}{n \varepsilon^2} = 0$$

condición que verifica la consistencia del estimador considerado.

EJERCICIO 3.13 Para estimar el parámetro p de una variable aleatoria ξ distribuida $B(1, p)$, en muestras aleatorias simples de tamaño 3, se establece el estadístico $T = x_1 + x_2 + x_3$, sabiendo que en una muestra concreta su valor ha sido 2. Demuéstrese que el estadístico es suficiente.

SOLUCIÓN. Para ello, verificaremos que $P(\xi_1 = x_1, \xi_2 = x_2, \xi_3 = x_3 / T = 2)$ es independiente del parámetro p

$$P = P(\xi_1 = x_1, \xi_2 = x_2, \xi_3 = x_3 / T = 2) = \frac{P(\xi_1 = x_1, \xi_2 = x_2, \xi_3 = x_3; T = 2)}{P(T = 2)}$$

Si $T \neq 2$ la probabilidad es igual a 0, y si $T = 2$

$$P = \frac{P(\xi_1 = x_1, \xi_2 = x_2, \xi_3 = x_3)}{\binom{3}{2} p^2 q} = \frac{p^{\sum_{i=1}^3 x_i} q^{n - \sum_{i=1}^3 x_i}}{\binom{3}{2} p^2 q} = \frac{p^T q^{n-T}}{\binom{3}{2} p^2 q} = \frac{p^2 q}{\binom{3}{2} p^2 q} = \frac{1}{\binom{3}{2}} = \frac{1}{3}.$$

La probabilidad no depende del parámetro p , por lo cual, el estadístico T es suficiente.

EJERCICIO 3.14 En una distribución $B(1, p)$, en muestras aleatorias simples de tamaño n , se utiliza como estimador la suma de los valores muestrales $T = x_1 + \dots + x_n$, demuéstrese que es suficiente.

SOLUCIÓN. La función de cuantía conjunta de la muestra es

$$\begin{aligned} L(\mathbf{X}; p) &= P(\xi_1 = x_1, \dots, \xi_n = x_n) \\ &= p^{x_1} (1-p)^{1-x_1} \dots p^{x_n} (1-p)^{1-x_n} = \\ &= p^{\sum_{i=1}^n x_i} (1-p)^{n - \sum_{i=1}^n x_i} = \\ &= p^T (1-p)^{n-T} = (1-p)^n \left[\frac{p}{1-p} \right]^T = g(T; p) \end{aligned}$$

Definimos una función $H(x_1, \dots, x_n) = 1$, y establecemos la factorización de la función de cuantía conjunta de la muestra

$$L(\mathbf{X}; p) = g(T; \theta) \cdot 1 = g(T; \theta) H(x_1, \dots, x_n)$$

con lo que queda demostrada la suficiencia del estimador suma de los valores muestrales.

CAPÍTULO 4

Métodos de estimación

De la misma manera que planteamos el problema de las condiciones que debe verificar un estimador para ser aceptado como «bueno» (capítulo 3) necesitamos disponer de criterios «objetivos» que permitan obtener, de entre los infinitos estimadores de un parámetro, el más «razonable», como paso previo a la constatación del cumplimiento de las deseables propiedades que debe cumplir un «buen» estimador.

A este fin se establecen distintos métodos de estimación, tratando aquí tres de ellos¹: máxima verosimilitud, de los momentos, mínimos cuadrados.

Cada uno de estos procedimientos tiene base de partida distinta: el de máxima verosimilitud y de los momentos la utilización del conocimiento de la distribución de probabilidad poblacional, directa en el primer caso y a través de sus momentos en el segundo; en los mínimos cuadrados la minimización de la distancia euclídea entre el valor conocido de una observación y su estimación puntual.

4.1 Método de estimación de máxima verosimilitud

El método de estimación máximo-verosímil se fundamenta en el supuesto intuitivo siguiente: de varios sucesos que pueden tener lugar, admitimos que aparecerá el

¹ El método de estimador minimax y el de Bayes se estudian en el capítulo 10. No trataremos la estimación de parámetros mediante estadísticos ordenados, remitiendo para ello a BALASKRISHNAN-COHEN.

más probable², o si ha aparecido uno concreto será razonable suponer que, entre todos los posibles, era el más probable. Es decir, si en una bolsa tenemos más bolas blancas que negras y extraemos una al azar, pensamos que lo más probable es que sea blanca, y si conocido el resultado debemos decidir si hay más bolas blancas que negras, y si la bola extraída ha sido blanca, diremos que el número de bolas blancas es superior al de negras. Este es el planteamiento de base en la estimación de parámetros por máxima verosimilitud y que introducimos con un ejemplo más preciso.

EJEMPLO 1

La probabilidad de obtener cara con una moneda puede tomar dos valores: 0,7 o 0,4. Para decidir entre ambos procedemos de la siguiente manera. La variable aleatoria en cuestión sigue una distribución $B(1; p)$, siendo p la probabilidad desconocida de cara: 0,7 o 0,4. Parece razonable que la elección del valor de la probabilidad en la población (0,7 o 0,4) dependa del número de caras que aparezcan en una muestra. Tomamos, para ello, una muestra aleatoria simple de tamaño 3, siendo el resultado dos caras y una cruz. Como los elementos muestrales son independientes, la probabilidad de aparición del suceso {dos caras y una cruz} es

$$P(\text{dos caras y una cruz}) = P(c \cap c \cap +) = P(c) P(c) P(+) = p p q = p^2 q.$$

Para el cálculo de la probabilidad hemos de tener en cuenta todas las posibles combinaciones, en este caso, tres, (cc+; c+c; +cc).³ Por consiguiente, la probabilidad del suceso es

$$P = 3p^2q.$$

El paso siguiente es la evaluación de esta probabilidad para cada uno de los dos valores que puede tomar el parámetro p , según provenga la muestra de una distribución $B(1; 0,7)$ o $B(1; 0,4)$

$$\text{Población con } p = 0,7: \quad P = 3 \cdot 0,7^2 \cdot 0,3 = 0,441$$

$$\text{Población con } p = 0,4: \quad P = 3 \cdot 0,4^2 \cdot 0,6 = 0,288$$

El resultado dice que es más probable la aparición de la muestra constituida por dos caras y una cruz si en la población el parámetro p toma el valor 0,7 que si es 0,4, por lo cual atribuimos al parámetro p el primer valor, y admitimos que la distribución es $B(1; 0,7)$.

El planteamiento operativo del método de estimación máximo-verosímil es el siguiente: Tenemos una variable aleatoria ξ con función de densidad, $f(x; \theta)$ o de cuantía, $P(\xi = x; \theta)$, siendo θ un parámetro desconocido, cuyo campo de variación es Θ . Tomamos una muestra aleatoria simple \mathbf{X} de tamaño n , (x_1, \dots, x_n) .

² Esto es una formulación de la Ley del Azar. MARTÍN PLIEGO y RUIZ-MAYA: *Fundamentos...* Capítulo 1.

³ Para llegar al resultado no es necesario tomar en consideración el número de formas en que pueden presentarse 2 caras y 1 cruz, aunque didácticamente sea conveniente.

Denominamos a la función conjunta de cuantía o densidad de la muestra **función de verosimilitud**

$$L(\mathbf{X}; \theta) = f(x_1; \theta) \cdots f(x_n; \theta),$$

donde $f(x_i; \theta)$ representa la función de cuantía del i -ésimo elemento muestral si la población es discreta o la función de densidad de este elemento muestral cuando la población sea continua.

■ Para la elección del estimador $\hat{\theta}$ del parámetro θ , de entre los posibles valores que puede tomar éste, se sigue el criterio de elegir $\hat{\theta}$ tal que

$$L(\mathbf{X}; \hat{\theta}) = \max_{\theta \in \Theta} L(\mathbf{X}; \theta).$$

La función de verosimilitud es multiplicativa y complicada para calcular su máximo respecto al parámetro θ por lo que, para mayor sencillez, se halla el valor de $\hat{\theta}$ en su logaritmo (transformación monótona creciente), siendo el estimador máximo-verosímil $\hat{\theta}$ el que verifica la relación

$$\ln L(\mathbf{X}; \hat{\theta}) = \max_{\theta \in \Theta} \ln L(\mathbf{X}; \theta).$$

Hallando el logaritmo de la función, derivamos respecto a θ , la igualamos a cero y llegamos a una ecuación con una incógnita

$$L(\mathbf{X}; \theta) = f(x_1; \theta) \cdots f(x_n; \theta)$$

$$\ln L(\mathbf{X}; \theta) = \ln f(x_1; \theta) \cdots \ln f(x_n; \theta) = \sum_{i=1}^n \ln f(x_i; \theta)$$

$$\frac{\ln L(\mathbf{X}; \theta)}{\partial \theta} = \sum_{i=1}^n \frac{\ln f(x_i; \theta)}{\partial \theta} = 0$$

la solución, $\hat{\theta} = \hat{\theta}(\mathbf{X})$, únicamente función de los elementos muestrales⁴, será el estimador máximo-verosímil del parámetro⁵, siempre que se verifique la condición de máximo

$$\left[\frac{\partial^2 \ln L(\mathbf{X}; \theta)}{\partial \theta^2} \right]_{\theta = \hat{\theta}} < 0.$$

⁴ Se prescinde de las soluciones que conduzcan a que el estimador sea igual a una constante.

⁵ En general, la solución de la ecuación no ofrece grandes dificultades, de no ser así habrá que recurrir a métodos iterativos de cálculo. Remitimos, entre otros, a SILVEY para la exposición de tal forma de actuar.

El número de raíces de la ecuación no tiene por qué ser uno. Si la raíz es única el estimador se denomina **estimador máximo verosímil en sentido estricto**, y corresponde al máximo absoluto de la función de verosimilitud. Cuando hay más de una raíz los estimadores reciben el nombre de **estimadores máximo verosímiles en sentido amplio**.

El campo de variación del estimador máximo-verosímil, $\hat{\theta}(x_1, \dots, x_n)$, es el mismo que el del parámetro θ .

EJEMPLO 2

Calcúlese el estimador máximo verosímil del parámetro p de la distribución $B(1; p)$, en muestras aleatorias simples de tamaño n .

La función de verosimilitud es

$$L(\mathbf{X}; p) = p^{\sum_{i=1}^n x_i} (1 - p)^{n - \sum_{i=1}^n x_i}$$

Tomando logaritmos, derivando respecto a p e igualando la derivada a cero

$$\frac{\partial \ln L(\mathbf{X}; p)}{\partial p} = \frac{\sum_{i=1}^n x_i}{p} - \frac{n - \sum_{i=1}^n x_i}{1 - p} = 0$$

la solución de la ecuación es

$$\hat{p} = \frac{\sum_{i=1}^n x_i}{n} = \bar{x}.$$

Comprobamos que es un máximo

$$\frac{\partial^2 \ln L(\mathbf{X}; p)}{\partial p^2} = \frac{-\sum_{i=1}^n x_i}{p^2} = \frac{n - \sum_{i=1}^n x_i}{(1 - p)^2}$$

expresión que particularizada para $\hat{p} = \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ conduce a

$$\left[\frac{\partial^2 \ln L(\mathbf{X}; p)}{\partial p^2} \right]_{p = \hat{p}} = -\frac{n}{\bar{x}} - \frac{n}{1 - \bar{x}}$$

menor que cero, condición de máximo.

EJEMPLO 3

En la distribución $N(\mu; \sigma)$, con varianza conocida, el parámetro μ desconocido se estima mediante el método de la máxima verosimilitud, en muestras aleatorias simples de tamaño n . La

función de verosimilitud es

$$\begin{aligned} L(\mathbf{X}; \mu) &= f(x_1; \mu, \sigma) \cdots f(x_n; \mu, \sigma) = \\ &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_1 - \mu)^2}{2\sigma^2}} \cdots \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_n - \mu)^2}{2\sigma^2}} = \\ &= \frac{1}{(\sigma^2)^{\frac{n}{2}} (2\pi)^{\frac{n}{2}}} e^{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}} \end{aligned}$$

su logaritmo es igual a

$$\ln L(\mathbf{X}; \mu) = -\frac{n}{2} \ln \sigma^2 - \frac{n}{2} \ln 2\pi - \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}$$

y su derivada respecto al parámetro μ

$$\frac{\partial \ln L(\mathbf{X}; \mu)}{\partial \mu} = \frac{\sum_{i=1}^n (x_i - \mu)}{\sigma^2} = 0.$$

Resuelta la ecuación llegamos a que el estimador máximo verosímil de μ es la media de la muestra

$$\hat{\mu} = \frac{\sum_{i=1}^n x_i}{n} = \bar{x}.$$

La condición de máximo se verifica, pues

$$\left[\frac{\partial^2 \ln L(\mathbf{X}; \mu)}{\partial \mu^2} \right]_{\mu = \hat{\mu}} = \frac{-n}{\sigma^2} < 0.$$

■ No siempre se puede obtener el estimador máximo verosímil por el método analítico expuesto, máximo de una función, en cuyo caso se procede partiendo del fundamento del método: lo sucedido es lo más verosímil que puede suceder.

EJEMPLO 4

Calculamos el estimador máximo verosímil del límite superior del campo de variación de la distribución $U(0; \theta)$, $0 \leq x \leq \theta$, con función de densidad igual a $1/\theta$, en muestras aleatorias simples de tamaño n .

La función de verosimilitud es $L(\mathbf{X}; \theta) = \frac{1}{\theta^n}$, siendo $\ln L(\mathbf{X}; \theta) = -n \ln \theta$. La derivada de esta función respecto a θ es $\frac{-n}{\theta}$ que se anula para $\theta \rightarrow \infty$, solución no admisible por conducir a una función de densidad igual a cero.

Hemos llegado, siguiendo los pasos establecidos, a que el método analítico de estimación máximo-verosímil no conduce a una solución aceptable, sin embargo, la obtendremos recurriendo al fundamento del método: de varios sucesos que pueden tener lugar se da el más verosímil. El parámetro θ es el límite superior del campo de variación de la distribución, lo que implica que la variable aleatoria no puede tomar ningún valor por encima de él, por consiguiente, el mayor valor de la muestra es la estimación máximo-verosímil del parámetro, es decir, $\hat{\theta} = \max(x_1, \dots, x_n)$.

■ Se ha comentado que puede no ser posible obtener el máximo de la función de verosimilitud de forma exacta siendo necesario, por consiguiente, recurrir a procedimientos aproximados. En el ejemplo siguiente presentamos un caso de esta situación llegando, únicamente, al planteamiento de las ecuaciones.

EJEMPLO 5

En la distribución gamma $G(p; q)$ calculamos los estimadores máximo verosímiles de p y q en muestras aleatorias simples de tamaño n . La expresión de la función de densidad es

$$f(x; p, q) = \frac{q^p}{\Gamma(p)} x^{p-1} e^{-qx}$$

La función de verosimilitud resulta

$$L(\mathbf{X}; p, q) = \frac{q^{np}}{[\Gamma(p)]^n} (x_1 \cdots x_n)^{p-1} e^{-q \sum_{i=1}^n x_i}$$

Calculamos el logaritmo de esta función

$$\ln L(\mathbf{X}; p, q) = np \ln q - n \ln \Gamma(p) + (p-1) \sum_{i=1}^n \ln x_i - q \sum_{i=1}^n x_i$$

y las derivadas respecto a p y q

$$\frac{\partial \ln L(\mathbf{X}; p, q)}{\partial p} = n \ln q - \frac{n}{\Gamma(p)} \frac{\partial \Gamma(p)}{\partial p} + \sum_{i=1}^n \ln x_i = 0$$

$$\frac{\partial \ln L(\mathbf{X}; p, q)}{\partial q} = \frac{np}{q} - \sum_{i=1}^n x_i = 0$$

de la segunda ecuación obtenemos

$$\hat{q} = \frac{\hat{p}}{\bar{x}}$$

4.2 Propiedades de los estimadores máximo verosímiles

El método de estimación máximo verosímil consiste en elegir, de entre los posibles valores que puede tomar el parámetro θ , el que hace máxima la función de verosimilitud $L(\mathbf{X}; \theta)$. Tal y como se ha definido el estimador máximo verosímil coincide con la moda de la función de verosimilitud. Por tanto, el método de la máxima verosimilitud pretende hallar la moda de $L(\mathbf{X}; \theta)$. El hecho que la moda sea generalmente un estimador más pobre que la media o la mediana explica por qué las propiedades del método en las muestras de tamaño pequeño son, en general, pobres. Para grandes muestras, sin embargo, la moda tiende a aproximarse a la media (si existe) y a la mediana, y el método tiene propiedades en muestras grandes,⁶ como se verá en las propiedades asintóticas.

Las propiedades de los estimadores obtenidos por el método de máxima verosimilitud son:

1. **Insesgadez.** Los estimadores máximo verosímiles no son, en general, insesgados. Sin embargo, si no son insesgados lo son asintóticamente.
2. **Consistencia.** Bajo condiciones generales, los estimadores máximo-verosímiles son consistentes. Si el estimador máximo-verosímil no es insesgado, al ser consistente será asintóticamente insesgado, puesto que $\hat{\theta}$ converge al parámetro θ que, también es, en el límite, su valor medio.
3. **Eficiencia.** Si existe un estimador cuya varianza es igual a la cota de Cramér-Rao es el obtenido por máxima verosimilitud.

Todo estimador máximo verosímil no tiene porqué ser eficiente, sin embargo, si existe un estimador eficiente es el máximo-verosímil.

4. **Normalidad y eficiencias asintóticas.** Los estimadores máximo-verosímiles son asintóticamente *normales* con esperanza θ y asintóticamente eficientes, es decir,

$$\hat{\theta} \xrightarrow{d} N\left(\theta, \frac{1}{\sqrt{I(\theta)}}\right).$$

⁶ Véase ROHATGI (1976).

- 5. Suficiencia.** Si $T = T(\mathbf{X})$ es un estadístico suficiente del parámetro θ , el estimador máximo-verosímil, $\hat{\theta}$, es función del estadístico suficiente $T(\mathbf{X})$.
- 6. Invarianza.** Si $\hat{\theta}$ es el estimador máximo-verosímil de θ , $g(\hat{\theta})$ es el estimador máximo-verosímil de $g(\theta)$. La estimación máximo-verosímil es invariante bajo una transformación del parámetro, o de los parámetros en su caso.

4.3 Estimación por el método de los momentos

Supóngase que la función de densidad, o de cuantía, $f(x; \theta_1, \dots, \theta_k)$ de una población contiene k parámetros desconocidos. El método de estimación basado en los momentos consiste en establecer la igualdad de los momentos poblacionales a los muestrales. La justificación teórica de este procedimiento de estimación se basa en que $E(a_r) = \alpha_r$, es decir, que a_r es un estimador insesgado de α_r y en el **teorema de Khintchine**⁷ que implica la consistencia de a_r como estimador del momento poblacional α_r .

El momento poblacional de orden r con respecto al origen, α_r , será función, en general, de los k parámetros

- En el caso discreto

$$\alpha_r(\theta_1, \dots, \theta_k) = \sum_{i=1}^{\infty} x_i^r P(\xi = x_i).$$

- En el caso continuo

$$\alpha_r(\theta_1, \dots, \theta_k) = \int_{-\infty}^{\infty} x^r f(x; \theta_1, \dots, \theta_k) dx.$$

De la población tomamos una muestra aleatoria simple de tamaño n (x_1, \dots, x_n), calculando los k primeros momentos muestrales de orden r con respecto al origen

$$a_r = \sum_{i=1}^n \frac{x_i^r}{n}.$$

⁷ Véase MARTÍN PLIEGO y RUIZ-MAYA: *Fundamentos...* Capítulo 9.

EJEMPLO 8

En la distribución $N(\mu, \sigma)$ estimamos la media y la varianza por el método de los momentos. Tenemos que estimar dos parámetros, por lo cual necesitamos dos ecuaciones, resultado de igualar los momentos poblacionales de primer y segundo orden respecto al origen a los muestrales

$$\alpha_1(\mu, \sigma^2) = a_1$$

$$\alpha_2(\mu, \sigma^2) = a_2$$

El primer momento poblacional en la distribución normal es μ , y el segundo α_2 que, expresado en función de la varianza queda $\alpha_2 = \sigma^2 + \mu^2$. El sistema de ecuaciones es

$$\mu = \frac{\sum_{i=1}^n x_i}{n}$$

$$\sigma^2 + \mu^2 = \frac{\sum_{i=1}^n x_i^2}{n}$$

y la solución

$$\mu^* = \frac{\sum_{i=1}^n x_i}{n} = \bar{x}$$

$$\sigma^{2*} = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = s^2$$

4.4 Estimación por el método de los mínimos cuadrados⁸

En muchas situaciones, el comportamiento probabilístico de una variable aleatoria η se puede expresar mediante una función desconocida $g(\Xi; \theta_1, \dots, \theta_k)$, integrada por variables aleatorias y no aleatorias, Ξ , y un conjunto de parámetros, $\theta_1, \dots, \theta_k$, necesitando estimar estos parámetros a fin de disponer operativamente de la función g .

⁸ El contenido de este epígrafe no es una exposición exhaustiva del método de los mínimos cuadrados, sino un simple panorama. Para un estudio pormenorizado tenemos, entre otros, a GRAYBILL y SEARLE.

Los métodos de construcción de estimadores de máxima verosimilitud y de los momentos requieren el conocimiento de la función de probabilidad de la población, el de los mínimos cuadrados recurre a la minimización de distancias, prescindiendo de la función de probabilidad.

En el método de los mínimos cuadrados se parte de una muestra de valores de $(\eta; \Xi)$, observaciones⁹, y se pretende que la función ajustada *pase lo más cerca posible* de todos ellos. *Lo más cerca posible* se mide mediante la distancia global entre los valores observados y los teóricos proporcionados por la función ajustada, calculándose los parámetros desconocidos de modo que esa distancia global sea mínima, teniendo como consecuencia, de ser así, que al utilizar la función ajustada para predecir valores de la variable η los resultados se hallen lo más cerca posible de los teóricos desconocidos.

Si la función a ajustar es $y = g(\mathbf{X}; \theta_1, \dots, \theta_k)$, siendo $\theta_1, \dots, \theta_k$ parámetros desconocidos¹⁰, para cada valor de la variable \mathbf{X} , x_i , tendremos un valor teórico de la variable y , Y_i , proporcionado por la función una vez ajustada; el **error**¹¹ cometido al sustituir cada valor observado, y_i , por el teórico, Y_i , es $e_i = y_i - Y_i$. La evaluación del error global se efectúa mediante la suma de cuadrados de los errores¹²

$$\Phi = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n [y_i - g(x_i; \theta)]^2.$$

Los estimadores de los parámetros $\theta_1, \dots, \theta_k$ son los valores que hacen mínimo Φ , es decir, la solución del sistema cuya ecuación general es

$$\frac{\partial \Phi}{\partial \theta_i} = -2 \sum_{j=1}^n [y_j - g(x_j; \theta)] \frac{\partial g(x_j; \theta)}{\partial \theta_i} = 0$$

o, despejando

$$\sum_{j=1}^n g(x_j; \theta) \frac{\partial g(x_j; \theta)}{\partial \theta_i} = \sum_{j=1}^n y_j \frac{\partial g(x_j; \theta)}{\partial \theta_i}.$$

La distancia, elemento básico de este método de estimación, puede ser la longitud entre el valor observado y el teórico según la paralela al eje de ordena-

⁹ Se le suele dar el nombre de **nube de puntos**.

¹⁰ Representaremos abreviadamente la función $g(\mathbf{X}; \theta_1, \dots, \theta_k)$ por $g(\mathbf{X}; \theta)$.

¹¹ También llamado **residuo**.

¹² En MARTÍN PLIEGO puede seguirse la mecánica de este método: capítulo 8.

das (distancia vertical) o la distancia entre el punto observado y la función a ajustar (distancia ortogonal¹³) Los resultados de uno y otro procedimiento son diferentes.

El valor de

$$\Phi = \sum_{i=1}^n e_i^2$$

es menor si se utiliza la distancia ortogonal frente a la vertical. Sin embargo, en la práctica, es preferible esta última por la simplicidad de las expresiones y sus propiedades.¹⁴

¹³ Distancia perpendicular a la tangente a la curva.

¹⁴ En CASELLA y CRAMÉR pueden estudiarse los mínimos cuadrados ortogonales.

Ejercicios resueltos

EJERCICIO 4.1 La variable aleatoria continua ξ tiene como función de densidad

$$f(x; \theta) = \theta e^{-\theta x}; \quad x \geq 0 \quad \text{y} \quad \theta > 0.$$

El parámetro θ puede adoptar tres valores: 1, 2 o 3. Se toma una muestra aleatoria simple de tamaño 5 con el siguiente resultado: tres valores comprendidos entre 0,003 y 0,06 y dos mayores que 0,5. Con esta información decídase cuál de los tres valores del parámetro θ debe ser elegido bajo el criterio de la máxima verosimilitud.

SOLUCIÓN. La probabilidad de una secuencia cualquiera de los cinco valores es

$$P_1 = P(0,003 < \xi \leq 0,06; 0,003 < \xi \leq 0,06; 0,003 < \xi \leq 0,06; 0,5 \leq \xi; 0,5 \leq \xi)$$

y la probabilidad buscada tomando en consideración todas las posibles ordenaciones de las secuencias bajo las condiciones anteriores es igual a

$$P = \frac{5!}{3! 2!} P_1 = 10 P_1.$$

Como la muestra es aleatoria simple, los elementos muestrales son independientes entre sí, por lo cual, la probabilidad anterior es igual a

$$\begin{aligned} P &= 10 P(0,003 < \xi \leq 0,06)^3 P(0,5 \leq \xi)^2 = \\ &= 10 \left[\int_{0,003}^{0,06} \theta e^{-\theta x} dx \right]^3 \left[\int_{0,5}^{\infty} \theta e^{-\theta x} dx \right]^2. \end{aligned}$$

Evaluamos la probabilidad para cada uno de los tres valores que puede tomar el parámetro θ

■ $\theta = 1$

$$P = 10 \left[\int_{0,003}^{0,06} e^{-x} dx \right]^3 \left[\int_{0,5}^{\infty} e^{-x} dx \right]^2 = 0,00052.$$

■ $\theta = 2$

$$P = 10 \left[\int_{0,003}^{0,06} 2e^{-2x} dx \right]^3 \left[\int_{0,5}^{\infty} 2e^{-2x} dx \right]^2 = 0,00166.$$

■ $\theta = 3$

$$P = 10 \left[\int_{0,003}^{0,06} 3e^{-3x} dx \right]^3 \left[\int_{0,5}^{\infty} 3e^{-3x} dx \right]^2 = 0,00188.$$

De las tres probabilidades calculadas, la tercera es la mayor, por lo cual decidimos estimar el valor del parámetro como $\hat{\theta} = 3$.

EJERCICIO 4.2 Calcúlese el estimador máximo-verosímil del parámetro p de la distribución $B(m; p)$, en muestras aleatorias simples de tamaño n .

SOLUCIÓN. La función de verosimilitud es

$$\begin{aligned} L(\mathbf{X}; p) &= \binom{m}{x_1} p^{x_1} (1-p)^{m-x_1} \cdots \binom{m}{x_n} p^{x_n} (1-p)^{m-x_n} = \\ &= \prod_{i=1}^n \binom{m}{x_i} p^{\sum_{i=1}^n x_i} (1-p)^{nm - \sum_{i=1}^n x_i} \end{aligned}$$

$$\ln L(\mathbf{X}; p) = \sum_{i=1}^n \ln \binom{m}{x_i} + \sum_{i=1}^n x_i \ln p + \left(nm - \sum_{i=1}^n x_i \right) \ln (1-p)$$

$$\frac{\partial \ln L(\mathbf{X}; p)}{\partial p} = \frac{\sum_{i=1}^n x_i}{p} - \frac{nm - \sum_{i=1}^n x_i}{1-p} = 0$$

y el estimador máximo-verosímil

$$\hat{p} = \frac{\sum_{i=1}^n x_i}{nm} = \frac{\bar{x}}{m}.$$

Calculando la segunda derivada respecto a p y sustituyendo en ella p por el estimador se llega a la relación

$$-\frac{nm^2}{\bar{x}} - \frac{nm^2}{m - \bar{x}} < 0$$

mediante la que se comprueba que la solución de la ecuación es un máximo.

EJERCICIO 4.3 Calcúlese el estimador máximo-verosímil del parámetro λ de la distribución de Poisson en muestras aleatorias simples de tamaño n .

SOLUCIÓN. La función de verosimilitud es

$$L(\mathbf{X}; \lambda) = e^{-n\lambda} \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!}$$

$$\ln L(\mathbf{X}; \lambda) = -n\lambda + \ln \lambda \sum_{i=1}^n x_i - \sum_{i=1}^n \ln x_i!$$

$$\frac{\partial \ln L(\mathbf{X}; \lambda)}{\partial \lambda} = -n + \frac{\sum_{i=1}^n x_i}{\lambda} = 0;$$

$$\hat{\lambda} = \frac{\sum_{i=1}^n x_i}{n} = \bar{x}$$

verificándose la condición de máximo

$$\left[\frac{\partial^2 \ln L(\mathbf{X}; \lambda)}{\partial \lambda^2} \right]_{\hat{\lambda}=\bar{x}} = -\frac{n}{\bar{x}} < 0.$$

EJERCICIO 4.4 Si en la distribución $N(\mu; \sigma)$ la varianza es desconocida y la media μ conocida, calcúlese la estimación máximo-verosímil de la varianza, en muestras aleatorias simples de tamaño n .

SOLUCIÓN. El logaritmo de la función de verosimilitud es

$$\ln L(\mathbf{X}; \sigma^2) = -\frac{n}{2} \ln(\sigma^2) - \frac{n}{2} \ln(2\pi) - \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}$$

y la derivada respecto a σ^2 e igualada a cero

$$\frac{\ln L(\mathbf{X}; \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^4} = 0$$

como $\sigma^2 > 0$ el estimador máximo verosímil de σ^2 es

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}$$

debiéndose observar que el estimador no es la varianza muestral, pues las desviaciones de los valores muestrales lo son con respecto a la media poblacional, μ , y no respecto a la media muestral, \bar{x} .

EJERCICIO 4.5 Si en la distribución normal desconocemos los dos parámetros (μ y σ^2), calcúlese los estimadores máximo verosímiles de μ y σ^2 .

SOLUCIÓN. La función de verosimilitud es

$$\begin{aligned} L(\mathbf{X}; \mu, \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_1 - \mu)^2}{2\sigma^2}} \dots \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_n - \mu)^2}{2\sigma^2}} = \\ &= \frac{1}{(\sigma^2)^{\frac{n}{2}} (2\pi)^{\frac{n}{2}}} e^{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}} \end{aligned}$$

su logaritmo

$$\ln L(\mathbf{X}; \mu, \sigma^2) = -\frac{n}{2} \ln(\sigma^2) - \frac{n}{2} \ln(2\pi) - \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}$$

y las derivadas respecto a los parámetros μ y σ^2 , igualadas a cero son

$$\left. \begin{aligned} \frac{\partial \ln L(\mathbf{X}; \mu, \sigma^2)}{\partial \mu} &= \frac{\sum_{i=1}^n (x_i - \mu)}{\sigma^2} = 0 \\ \frac{\partial \ln L(\mathbf{X}; \mu, \sigma^2)}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^4} = 0 \end{aligned} \right\}$$

obteniéndose como solución del sistema

$$\hat{\mu} = \bar{x}; \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = s^2.$$

Las estimaciones máximo-verosímiles de μ y σ^2 son la media y la varianza muestrales.

Se comprueba la condición de máximo calculando el Hessiano

$$H = \begin{vmatrix} \frac{\partial^2 \ln L(\mathbf{X}; \mu, \sigma^2)}{\partial \mu^2} & \frac{\partial^2 \ln L(\mathbf{X}; \mu, \sigma^2)}{\partial \mu \partial \sigma^2} \\ \frac{\partial^2 \ln L(\mathbf{X}; \mu, \sigma^2)}{\partial \mu \partial \sigma^2} & \frac{\partial^2 \ln L(\mathbf{X}; \mu, \sigma^2)}{\partial (\sigma^2)^2} \end{vmatrix}$$

y sustituyendo en él los estimadores obtenidos, se llega a

$$H = \begin{vmatrix} -\frac{n}{s^2} & 0 \\ 0 & -\frac{n}{2s^4} \end{vmatrix}$$

cuyo resultado es $|H| = \frac{n^2}{2s^6}$ cantidad mayor que cero lo que indica un extremo, y al ser $-\frac{n}{s^2}$ menor que cero la solución del sistema de ecuaciones proporciona un máximo.

EJERCICIO 4.6 Calcúlense los estimadores máximo-verosímiles de los k parámetros p_i de una distribución multinomial, habiéndose extraído la muestra aleatoria simple de tamaño n con las frecuencias absolutas siguientes: n_1 valores de x_1 , n_2 iguales a x_2 , ..., y n_k de x_k .

SOLUCIÓN. La función de verosimilitud es

$$L(n_1, \dots, n_k; p_1, \dots, p_k) = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k}$$

y su logaritmo

$$\ln L(n_1, \dots, n_k; p_1, \dots, p_k) = \ln n! - \sum_{i=1}^k \ln n_i! + \sum_{i=1}^k n_i \ln p_i.$$

Si hacemos

$$p_k = 1 - \sum_{i=1}^{k-1} p_i$$

el logaritmo de la función de verosimilitud queda

$$\ln L(n_1, \dots, n_k; p_1, \dots, p_{k-1}) = \ln n! - \sum_{i=1}^k \ln n_i! + \sum_{i=1}^{k-1} n_i \ln p_i + n_k \ln \left(1 - \sum_{i=1}^{k-1} p_i \right)$$

y derivando respecto a los $k - 1$ parámetros e igualando a cero

$$\frac{\partial \ln L(n_1, \dots, n_k; p_1, \dots, p_{k-1})}{\partial p_i} = \frac{n_i}{p_i} - \frac{n_k}{1 - \sum_{i=1}^{k-1} p_i} = 0$$

$$\frac{n_i}{p_i} = \frac{n_k}{1 - \sum_{i=1}^{k-1} p_i} = \frac{n_k}{p_k}$$

y de aquí se deduce, fácilmente, que

$$\frac{n_1}{p_1} = \dots = \frac{n_k}{p_k} = \frac{n_1 + \dots + n_k}{p_1 + \dots + p_k} = n$$

por lo cual, los estimadores máximo-verosímiles tienen como expresión general

$$\hat{p}_i = \frac{n_i}{n}.$$

EJERCICIO 4.7 Sea la variable aleatoria con función de densidad

$$f(x; \theta, \lambda) = \frac{1}{\lambda} e^{-\frac{x-\theta}{\lambda}}, \quad \lambda > 0, \quad -\infty < \theta < \infty, \quad x \geq \theta.$$

Estímense los parámetros θ y λ en muestras aleatorias simples de tamaño n , por el método de máxima verosimilitud.

SOLUCIÓN. La función de verosimilitud es

$$L(\mathbf{X}; \theta, \lambda) = \frac{1}{\lambda^n} e^{-\frac{1}{\lambda} \sum_{i=1}^n (x_i - \theta)}$$

su logaritmo y derivadas respecto a θ y λ , igualadas a cero son

$$\ln L(\mathbf{X}; \theta, \lambda) = -n \ln \lambda - \frac{1}{\lambda} \sum_{i=1}^n (x_i - \theta)$$

$$\left. \begin{aligned} \frac{\partial \ln L(\mathbf{X}; \theta, \lambda)}{\partial \lambda} &= -\frac{n}{\lambda} + \frac{1}{\lambda^2} \sum_{i=1}^n (x_i - \theta) = 0 \\ \frac{\partial \ln L(\mathbf{X}; \theta, \lambda)}{\partial \theta} &= \frac{n}{\lambda} = 0 \end{aligned} \right\}$$

La primera ecuación del sistema conduce a

$$\hat{\lambda} = \frac{\sum_{i=1}^n (x_i - \hat{\theta})}{n}$$

estimador de λ en función del estimador de θ . La segunda ecuación indica que no es posible, siguiendo este procedimiento, hallar $\hat{\theta}$, por lo cual habrá que seguir un camino similar al del ejemplo 4. El valor del parámetro θ que hace máxima la función de verosimilitud

$$\max_{\theta \in \Theta} L(\mathbf{X}; \theta, \lambda) = \max_{\theta \in \Theta} \frac{1}{\lambda^n} e^{-\frac{1}{\lambda} \sum_{i=1}^n (x_i - \theta)}$$

es el mismo para el que se verifican las igualdades

$$\max_{\theta \in \Theta} e^{-\frac{1}{\lambda} \sum_{i=1}^n (x_i - \theta)} = \min_{\theta \in \Theta} \sum_{i=1}^n (x_i - \theta) = \max_{\theta \in \Theta} \theta = \min(x_1, \dots, x_n),$$

es decir, $\hat{\theta}$ es el menor valor de la muestra (x_1, \dots, x_n) , y el estimador máximo-verosímil de λ resulta

$$\hat{\lambda} = \frac{\sum_{i=1}^n (x_i - \hat{\theta})}{n} = \frac{\sum_{i=1}^n [x_i - \min(x_1, \dots, x_n)]}{n} = \bar{x} - \min(x_1, \dots, x_n).$$

EJERCICIO 4.8 La variable aleatoria ξ tiene por función de densidad la función gamma

$$f(x; p, q) = \frac{q^p}{\Gamma(p)} x^{p-1} e^{-qx}; \quad x \geq 0, \quad p > 0, \quad q > 0.$$

Obténganse los estimadores de p y q por el método de los momentos en muestras aleatorias simples de tamaño n .

SOLUCIÓN. Al tener que estimar dos parámetros debemos calcular los dos primeros momentos de la distribución gamma con respecto al origen

$$\alpha_1 = E(\xi) = \frac{p}{q}; \quad \alpha_2 = E(\xi^2) = \frac{(p+1)p}{q^2}.$$

y los correspondientes muestrales a_1 y a_2 .

El sistema de ecuaciones cuya solución son los estimadores es

$$\left. \begin{aligned} \alpha_1(p, q) &= a_1 \\ \alpha_2(p, q) &= a_2 \end{aligned} \right\}$$

y sustituyendo los momentos poblacionales por sus expresiones queda

$$\left. \begin{aligned} \frac{p}{q} &= \frac{\sum_{i=1}^n x_i}{n} \\ \frac{(p+1)p}{q^2} &= \frac{\sum_{i=1}^n x_i^2}{n} \end{aligned} \right\}$$

sistema cuya solución es

$$p^* = \frac{\bar{x}^2}{s^2}$$

$$q^* = \frac{\bar{x}}{\frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2} = \frac{\bar{x}}{s^2}$$

EJERCICIO 4.9 Calcúlese los estimadores por los momentos de los parámetros p y q en la función de densidad beta

$$f(x; p, q) = \frac{1}{B(p, q)} x^{p-1}(1-x)^{q-1}; \quad 0 \leq x \leq 1, \quad p > 0, \quad q > 0.$$

SOLUCIÓN. Los dos primeros momentos poblacionales con respecto al origen son

$$\alpha_1 = E(\xi) = \frac{p}{p+q};$$

$$\alpha_2 = E(\xi^2) = \frac{p(p+1)}{(p+q+1)(p+q)}$$

y el sistema de ecuaciones correspondientes

$$\left. \begin{aligned} \frac{p}{p+q} &= \bar{x} \\ \frac{p(p+1)}{(p+q+1)(p+q)} &= \frac{\sum_{i=1}^n x_i^2}{n} \end{aligned} \right\}$$

con la solución

$$p^* = \frac{\bar{x}(\bar{x} - a_2)}{s^2}$$

$$q^* = \frac{(1 - \bar{x})(\bar{x} - a_2)}{s^2}$$

CAPÍTULO 5

Estimación por intervalos

5.1 Introducción

Mediante los procedimientos estudiados en el capítulo 4 es posible construir un «buen» estimador de un parámetro θ que verifique, incluso, todas las propiedades exigidas al respecto.

Cuando se particulariza el estimador θ^* para una muestra \mathbf{X} concreta, es decir, cuando se obtiene la **estimación puntual** para una muestra determinada, no se sabe si la estimación obtenida está o no próxima al verdadero valor del parámetro θ , desconocimiento debido a la aleatoriedad del muestreo. Por ello, resulta extremadamente útil arbitrar alguna medida que nos aproxime al posible error de esa estimación determinada.

A este respecto, la decisión que se adopta es la de acompañar toda estimación puntual θ^* de un intervalo, que denominamos $[\underline{\theta}(\mathbf{X}); \bar{\theta}(\mathbf{X})]$, donde *confiamos* que se encuentre incluido el verdadero valor del parámetro θ .

Este intervalo, denominado **intervalo de confianza** por las razones que luego se expondrán, presenta dos límites, uno inferior, designado por $\underline{\theta}(\mathbf{X})$, y otro superior $\bar{\theta}(\mathbf{X})$, ambas funciones de la información muestral, y tales que

$$P[\underline{\theta}(\mathbf{X}) \leq \theta \leq \bar{\theta}(\mathbf{X})] = 1 - \alpha \quad [1]$$

donde $1 - \alpha$ recibe el nombre de **nivel de confianza**.

El intervalo $[\underline{\theta}(\mathbf{X}); \bar{\theta}(\mathbf{X})]$, antes de particularizar su valor para una muestra concreta \mathbf{X} , es un **intervalo aleatorio**, dependiente del vector muestral \mathbf{X} ,¹ fijado una muestra concreta el nivel de confianza $1 - \alpha$ y, por tanto, la expresión [1] no debe interpretarse como «la probabilidad de que el parámetro θ tome algún valor entre $\underline{\theta}(\mathbf{X})$ y $\bar{\theta}(\mathbf{X})$ es igual a $1 - \alpha$ », por las dos razones siguientes:

- El parámetro θ siempre será desconocido, lo que impide verificar la afirmación anterior.
- En $P[\underline{\theta}(\mathbf{X}) \leq \theta \leq \bar{\theta}(\mathbf{X})]$ las variables aleatorias son $\underline{\theta}(\mathbf{X})$ y $\bar{\theta}(\mathbf{X})$ y no el parámetro θ .

La interpretación correcta de la expresión [1] es que $1 - \alpha$ es la probabilidad que el intervalo aleatorio $[\underline{\theta}(\mathbf{X}); \bar{\theta}(\mathbf{X})]$ incluya el verdadero valor del parámetro θ antes de extraer la muestra, es decir,

$$P \{ \theta \in [\underline{\theta}(\mathbf{X}); \bar{\theta}(\mathbf{X})] \} = 1 - \alpha .$$

Seleccionada una muestra \mathbf{X} , la probabilidad de que el parámetro θ esté incluido en el intervalo $[\underline{\theta}(\mathbf{X}); \bar{\theta}(\mathbf{X})]$ es 1 o 0, dependiendo de que el parámetro θ esté o no esté entre los dos números en que se convierten $\underline{\theta}(\mathbf{X})$ y $\bar{\theta}(\mathbf{X})$, al particularizarlos para la muestra \mathbf{X} concreta.

En esta situación no se puede hablar de probabilidad del intervalo al nivel $1 - \alpha$ sino de **confianza** puesto que, una vez extraída la muestra \mathbf{X} , la probabilidad será 1 o 0, y no la inicial $1 - \alpha$ que se transforma en confianza.

Un ejemplo que aclara la distinción entre probabilidad inicial y confianza es el siguiente²: *Un estudiante se examina de un programa de 100 lecciones, de esas lecciones hay sólo una que no sabe, no recuerda exactamente el número de esa lección, tendrá que mirarlo en el programa. Se examina sacando una bola y desarrollando la lección que le ha correspondido, si la sabe aprueba y si no le suspenden. Este estudiante antes de sacar la bola tiene una probabilidad del 0,99 de aprobar. Supongamos que comparece ante el tribunal y saca una bola, la probabilidad que tienen ya en ese momento de aprobar no es de 0,99, es de uno si la lección salida es de las que se sabe y de cero si es la que no sabe, ahora bien, mientras consulta el programa él tiene una confianza del 0,99 de que aprobará, ya que antes de sacar la bola tenía una probabilidad de 0,99 de obtener una bola cuya lección correspondiente se sabía.*

¹ Los extremos de cada intervalo variarán para cada muestra.

² Véase ARNAIZ. No conocemos, en otro texto, un ejemplo tan esclarecedor como éste, por lo que lo reproducimos literalmente.

Para otros autores, el concepto de confianza puede interpretarse en el siguiente sentido: si se repitiera el experimento muestral \mathbf{X} un número suficiente de veces, en $100(1 - \alpha)$ de cada 100 casos se confiaría en que el parámetro θ pertenecerá al intervalo $[\underline{\theta}(\mathbf{X}); \bar{\theta}(\mathbf{X})]$.

En algunos casos específicos puede interesar contruir intervalos de confianza **unilaterales**. Para ello se hace $\underline{\theta}(\mathbf{X}) = 0$, con lo que el intervalo de confianza adopta la forma $[0; \bar{\theta}(\mathbf{X})]$, o bien $\bar{\theta}(\mathbf{X}) = 0$ y el intervalo es $[\underline{\theta}(\mathbf{X}) = 0]$.

5.2 Métodos de construcción de intervalos de confianza

Exponemos a continuación los dos métodos más usuales para la construcción de intervalos de confianza, es decir, para la determinación de los límites $\underline{\theta}(\mathbf{X})$ y $\bar{\theta}(\mathbf{X})$ de dichos intervalos.

5.2.1. MÉTODO DE LA CANTIDAD PIVOTAL

Sea \mathbf{X} una muestra aleatoria simple extraída de una población con distribución $f(x; \theta)$ donde $\theta \in \Theta$, siendo Θ cualquier intervalo de la recta real.

Una función $T(\mathbf{X}; \theta)$, que dependa del parámetro θ pero que su distribución de probabilidad no dependa de dicho parámetro θ recibe el nombre de **cantidad pivotal** o, simplemente **pivote**.

Este pivote $T(\mathbf{X}; \theta)$ para cada valor fijo de θ es un estadístico.

Por ejemplo, hemos visto en el capítulo 1 que, en una muestra aleatoria simple extraída de una población $N(\mu; \sigma)$ con σ conocida, la media muestral \bar{x} sigue la siguiente distribución

$$\bar{x} \rightarrow N\left(\mu; \frac{\sigma}{\sqrt{n}}\right)$$

entonces, la función

$$T(\mathbf{X}; \mu) = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \rightarrow N(0; 1)$$

es una cantidad pivotal para el parámetro μ dado que depende de ese parámetro, pero su distribución de probabilidad, $N(0; 1)$, no depende de él.

TEOREMA

Si la cantidad pivotal $T(\mathbf{X}; \theta)$ es función monótona de θ , es posible determinar un intervalo de confianza para el parámetro θ .

En efecto, para un nivel de confianza $1 - \alpha$ se pueden elegir un par de valores $K_1(\alpha)$ y $K_2(\alpha)$, pertenecientes al campo de variación de $T(\mathbf{X}; \theta)$, tales que

$$P[K_1(\alpha) \leq T(\mathbf{X}; \theta) \leq K_2(\alpha)] = 1 - \alpha, \quad \text{para todo } \theta \in \Theta.$$

Si $T(\mathbf{X}; \theta)$ es monótona en θ , se pueden resolver las ecuaciones

$$T(\mathbf{X}; \theta) = K_1(\alpha)$$

$$T(\mathbf{X}; \theta) = K_2(\alpha)$$

obteniéndose los límites $\underline{\theta}(\mathbf{X})$ y $\bar{\theta}(\mathbf{X})$ al despejar θ , con lo que el intervalo buscado es

$$P[\underline{\theta}(\mathbf{X}) \leq \theta \leq \bar{\theta}(\mathbf{X})] = 1 - \alpha.$$

EJEMPLO 1

En una población con función de densidad

$$f(x; \theta) = \frac{3x^2}{\theta^3}; \quad 0 \leq x \leq \theta$$

el estimador máximo-verosímil del parámetro θ es el mayor valor de la muestra, u_n , siendo su función de distribución en el muestreo³

$$F(u_n) = \left[\frac{u_n}{\theta} \right]^{3n}, \quad 0 \leq u_n \leq \theta$$

y, por tanto, la de densidad

$$f(u_n) = \frac{3n}{\theta^{3n}} u_n^{3n-1}; \quad 0 \leq u_n \leq \theta.$$

Determinése el intervalo de confianza del 80% para el parámetro θ .

³ Véase el capítulo 3.

Podemos considerar como cantidad pivotal

$$T(\mathbf{X}; \theta) = \left[\frac{u_n}{\theta} \right]^{3n}$$

que al ser una **transformación integral**⁴ de u_n tendrá una distribución uniforme en el intervalo $[0; 1]$, por lo que sus funciones de distribución y densidad son , respectivamente,

$$F_T(t) = t \quad 0 \leq t \leq 1$$

$$f_T(t) = 1 \quad 0 \leq t \leq 1$$

que no dependen del parámetro θ .

Por tanto,

$$P[K_1(\alpha) \leq T(\mathbf{X}; \theta) \leq K_2(\alpha)] = 1 - \alpha = 0,80$$

si se distribuye el conjunto de la probabilidad de las colas por igual entre las mismas⁵ tendremos que

$$\int_{K_2(\alpha)}^1 f_T(t) dt = \int_{K_2(\alpha)}^1 dt = 0,10$$

$$\int_0^{K_1(\alpha)} f_T(t) dt = \int_0^{K_1(\alpha)} dt = 0,10$$

y

$$\left. \begin{aligned} 1 - K_2(\alpha) &= 0,10 \\ K_1(\alpha) &= 0,10 \end{aligned} \right\}$$

es decir

$$\left. \begin{aligned} K_2(\alpha) &= 0,90 \\ K_1(\alpha) &= 0,10 \end{aligned} \right\}$$

Formando el sistema de ecuaciones

$$\left. \begin{aligned} T(\mathbf{X}; \theta) &= \left[\frac{u_n}{\theta} \right]^{3n} = K_1(\alpha) = 0,10 \\ T(\mathbf{X}; \theta) &= \left[\frac{u_n}{\theta} \right]^{3n} = K_2(\alpha) = 0,90 \end{aligned} \right\}$$

⁴ Véase MARTÍN PLIEGO y RUIZ-MÁYA: *Fundamentos...* Capítulo 7.

⁵ A esta cuestión, longitud de los intervalos de confianza, dedicaremos el epígrafe 5.2.3.

se obtiene como solución

$$\underline{\theta}(\mathbf{X}) = \frac{u_n}{\sqrt[3n]{0,90}}$$

$$\bar{\theta}(\mathbf{X}) = \frac{u_n}{\sqrt[3n]{0,10}}$$

siendo, por consiguiente, el intervalo de confianza del 80% para el parámetro θ

$$\left[\frac{u_n}{\sqrt[3n]{0,90}} ; \frac{u_n}{\sqrt[3n]{0,10}} \right].$$

5.2.2. MÉTODO GENERAL

La aplicación del método anterior implica la búsqueda y localización de una cantidad pivotal $T(\mathbf{X}; \theta)$, cuya distribución en el muestreo sea independiente del parámetro, θ . Este procedimiento no siempre conduce a una sencilla determinación de los extremos del intervalo de confianza por lo cual, en ocasiones, interesa recurrir al método alternativo expuesto a continuación, si bien ambos procedimientos proporcionan los mismos resultados cuando se cumplen las condiciones exigidas a cada uno de ellos para su aplicación.

En el método general de construcción de intervalos de confianza no se exige la propiedad del pivote en que se basa el método anterior, pudiendo tener la función muestral que se utilice una distribución de probabilidad que dependa del parámetro θ .

Sea una población con función de probabilidad $f(x; \theta)$ en la que se quiere construir un intervalo de confianza para el parámetro θ . Se toma un estimador de dicho parámetro, estimador que representamos por θ^* , cuya función de probabilidad en el muestreo es $g(\theta^*; \theta)$.

Dado el nivel de confianza $1 - \alpha$, se determinan los extremos de un intervalo $K_1(\alpha; \theta)$ y $K_2(\alpha; \theta)$ tales que

$$P[K_1(\alpha; \theta) \leq \theta^* \leq K_2(\alpha; \theta)] = 1 - \alpha.$$

En una población continua tenemos

$$\int_{K_1(\alpha; \theta)}^{K_2(\alpha; \theta)} g(\theta^*; \theta) d\theta^* = 1 - \alpha$$

o bien

$$\int_{-\infty}^{K_1(\alpha; \theta)} g(\theta^*; \theta) d\theta^* = \alpha_1$$

$$\int_{K_2(\alpha; \theta)}^{\infty} g(\theta^*; \theta) d\theta^* = \alpha_2$$

donde $\alpha_1 + \alpha_2 = \alpha$, como indica la figura siguiente

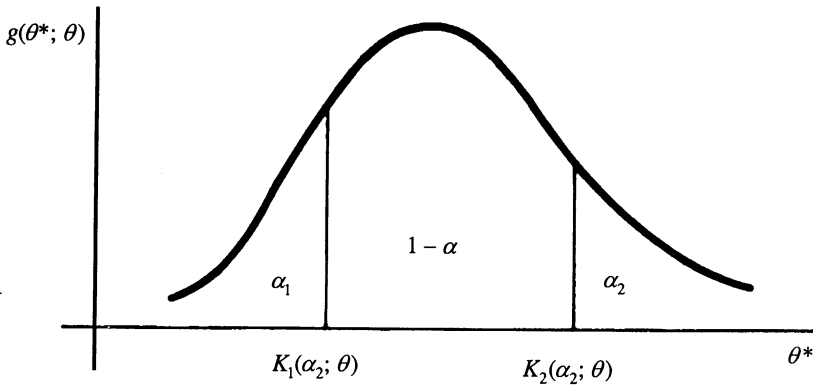


FIGURA 5.1

La asignación de valores a α_1 y α_2 , es decir, el reparto de la probabilidad α entre las dos colas de la distribución, incide en la longitud del intervalo obtenido. Este tema se analizará en el epígrafe siguiente.

Al resolver las integrales anteriores se obtienen los límites $K_1(\alpha_1; \theta)$ y $K_2(\alpha_2; \theta)$ tal que

$$P[K_1(\alpha; \theta) \leq \theta^* \leq K_2(\alpha; \theta)] = 1 - \alpha$$

de donde se deducen las ecuaciones, denominadas **ecuaciones integrales**

$$\left. \begin{aligned} K_1(\alpha_1; \theta) &= \theta^* \\ K_2(\alpha_2; \theta) &= \theta^* \end{aligned} \right\}$$

de manera que, si $K_1(\alpha_1; \theta)$ y $K_2(\alpha_2; \theta)$ son funciones monótonas en θ , permiten despejar el parámetro θ y hallar los límites $\underline{\theta}(X)$ y $\bar{\theta}(X)$ del intervalo buscado.

Por otra parte, conviene recordar que si la población tiene una distribución de probabilidad discreta, no siempre será factible lograr un intervalo de un nivel de confianza exactamente igual a $1 - \alpha$ sino que, a veces, habrá que aceptar un intervalo tal que

$$P[\underline{\theta}(\mathbf{X}) \leq \theta \leq \bar{\theta}(\mathbf{X})] \geq 1 - \alpha.$$

EJEMPLO 2

Consideremos la misma población del ejemplo anterior

$$f(x; \theta) = \frac{3x^2}{\theta^3}; \quad 0 \leq x \leq \theta.$$

El estimador máximo-verosímil es el mayor valor de la muestra, $\theta^* = u_n$, con función de densidad en el muestreo

$$f(u_n) = \frac{3n}{\theta^{3n}} u_n^{3n-1}; \quad 0 \leq u_n \leq \theta.$$

Obtégase un intervalo de confianza del 80% para el parámetro θ por el método general.

Para determinar las ecuaciones integrales adoptamos el criterio

$$\alpha_1 = \alpha_2 = \frac{\alpha}{2} = \frac{1 - 0,80}{2} = 0,10$$

por lo que

$$P[K_1(\alpha; \theta) \leq u_n \leq K_2(\alpha; \theta)] = 0,80$$

conduce a

$$\int_0^{K_1(\alpha; \theta)} \frac{3n}{\theta^{3n}} u_n^{3n-1} du_n = 0,10$$

$$\int_{K_2(\alpha; \theta)}^{\theta} \frac{3n}{\theta^{3n}} u_n^{3n-1} du_n = 0,10$$

es decir

$$\left[\frac{K_1(\alpha; \theta)}{\theta} \right]^{3n} = 0,10$$

$$1 - \left[\frac{K_2(\alpha; \theta)}{\theta} \right]^{3n} = 0,10$$

las ecuaciones integrales se obtienen finalmente como

$$K_1(\alpha; \theta) = \sqrt[3n]{0,10} \theta = \theta^* = u_n$$

$$K_2(\alpha; \theta) = \sqrt[3n]{0,90} \theta = \theta^* = u_n$$

cuya representación gráfica es

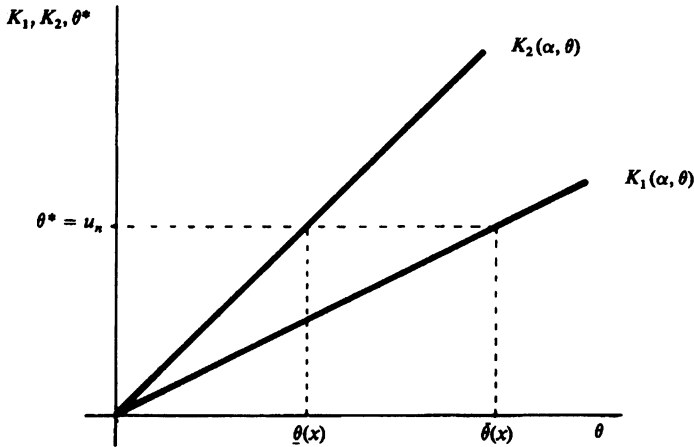


FIGURA 5.2

y los extremos del intervalo de confianza son

$$\underline{\theta}(X) = \frac{u_n}{\sqrt[3n]{0,90}}$$

$$\bar{\theta}(X) = \frac{u_n}{\sqrt[3n]{0,10}}$$

intervalo que coincide con el obtenido por el método del pivote para este parámetro θ de la población considerada, donde u_n representa el mayor valor de una muestra aleatoria simple de tamaño n .

5.2.3. INTERVALOS DE CONFIANZA DE LONGITUD MÍNIMA

Tal como se ha planteado la construcción de intervalos de confianza para un parámetro θ se pueden obtener, para cada muestra \mathbf{X} , un número infinito de intervalos pues los valores de α_1 y α_2 que cumplen la condición $\alpha_1 + \alpha_2 = \alpha$ son infinitos. Esta situación conduce a que con los mismos valores muestrales se obtengan diferentes intervalos (extremos distintos), aunque todos ellos con el mismo nivel de confianza $1 - \alpha$.

De entre todos estos intervalos es deseable elegir uno sólo: el que presente menor longitud, es decir, el intervalo cuya amplitud $D = \bar{\theta}(\mathbf{X}) - \underline{\theta}(\mathbf{X})$ sea la menor posible, siempre que cumpla la condición que

$$p \quad P[\underline{\theta}(\mathbf{X}) \leq \theta \leq \bar{\theta}(\mathbf{X})] = 1 - \alpha.$$

Para determinar dicho intervalo, se utiliza el método de optimización condicionada de los multiplicadores de Lagrange.⁶ Es preciso poner de manifiesto que no siempre es factible determinar un intervalo de nivel de confianza $1 - \alpha$ de longitud mínima, en cuyo caso se recurre al *convenio* de hacer $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$, con lo que el intervalo de confianza es único aunque no sea de longitud mínima.

EJEMPLO 3

Siguiendo con la población

$$f(x; \theta) = \frac{3x^2}{\theta^3}; \quad 0 \leq x \leq \theta,$$

donde la distribución de estimador máximo verosímil $\theta^* = u_n$ viene dada por

$$f(u_n) = \frac{3n}{\theta^{3n}} u_n^{3n-1}; \quad 0 \leq u_n \leq \theta$$

cuya representación gráfica es

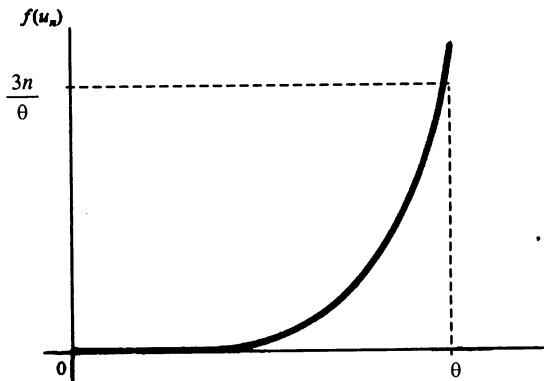


FIGURA 5.3

⁶ Véase el epígrafe 5.3.1., para el caso de intervalo de la media poblacional en una distribución normal.

el intervalo de longitud mínima estaría limitado superiormente por el límite de la variable

$$P[K_1(\alpha; \theta) \leq u_n \leq \theta] = 1 - \alpha$$

que difiere del utilizado en los dos ejemplos anteriores que, para el nivel de confianza $1 - \alpha = 0,80$, se hacía $\alpha_1 = \alpha_2 = 0,10$, y no $\alpha_2 = 0$ y $\alpha_1 = 0,20$ que proporciona el intervalo de longitud mínima.

Una solución alternativa a la minimización de la longitud del intervalo aleatorio es la de minimizar la longitud esperada de esos intervalos, es decir, hacer lo menor posible

$$E[\bar{\theta}(\mathbf{X}) - \underline{\theta}(\mathbf{X})]$$

esperanza sujeta a la misma restricción anterior, nivel de confianza $1 - \alpha$.

Desafortunadamente, este procedimiento tampoco resuelve todos los casos que se presentan, pues no puede asegurarse que exista siempre un intervalo de longitud esperada mínima para todo valor del parámetro θ .

Para solucionar el problema, el criterio más habitualmente empleado es repartir el complementario del nivel de confianza $1 - \alpha$, entre las dos colas por igual, es decir,

$$\alpha_1 = \alpha_2 = \frac{\alpha}{2}$$

situación que, por otra parte, coincide en muchas ocasiones con el intervalo de longitud mínima, como luego se comprobará en las aplicaciones concretas. Una ventaja de esta solución es que conduce a un intervalo único.

5.3 Intervalos de confianza en poblaciones normales

En este epígrafe la información muestral siempre procederá de poblaciones con distribución normal, donde es relativamente sencillo encontrar cantidades pivotaes que faciliten la determinación de cada uno de los intervalos de confianza que se exponen.

Se analizan los casos en que se obtengan una o dos muestras aleatorias simples y se estudian los posibles intervalos que sobre cada uno de los parámetros de una $N(\mu; \sigma)$ puede considerarse.

5.3.1. INTERVALO DE CONFIANZA PARA LA MEDIA μ DE UNA DISTRIBUCIÓN NORMAL SIENDO σ CONOCIDA

Sea una población con distribución normal $N(\mu; \sigma)$ de donde se extrae una muestra aleatoria simple de tamaño n .

Se puede formar el pivote

$$T(\mathbf{X}; \mu) = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}},$$

cuya distribución es $N(0; 1)$ puesto que⁶

$$\bar{x} \rightarrow N\left(\mu; \frac{\sigma}{\sqrt{n}}\right).$$

El intervalo de confianza para el parámetro μ del nivel $1 - \alpha$ vendrá dado por la expresión

$$P[K_1 \leq N(0; 1) \leq K_2] = 1 - \alpha,$$

siendo K_1 y K_2 los valores tabulares de una $N(0; 1)$ correspondientes al nivel de confianza $1 - \alpha$, sobre los que más adelante se estudiará cómo deben ser para que el intervalo obtenido tenga longitud mínima.

Siguiendo con el razonamiento general, determinados K_1 y K_2 se procede de la forma siguiente

$$P\left[K_1 \leq \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq K_2\right] = 1 - \alpha$$

$$P\left[K_1 \frac{\sigma}{\sqrt{n}} \leq \bar{x} - \mu \leq K_2 \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

$$P\left[-\bar{x} + K_1 \frac{\sigma}{\sqrt{n}} \leq -\mu \leq -\bar{x} + K_2 \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

$$P\left[\bar{x} - K_1 \frac{\sigma}{\sqrt{n}} \geq \mu \geq \bar{x} - K_2 \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

⁶ Véase el capítulo 1.

es decir,

$$P\left[\bar{x} - K_2 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} - K_1 \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

y, por tanto, el intervalo obtenido, en principio, será

$$\left[\bar{x} - K_2 \frac{\sigma}{\sqrt{n}}; \bar{x} - K_1 \frac{\sigma}{\sqrt{n}}\right]$$

cuya longitud es

$$D = \left[\bar{x} - K_1 \frac{\sigma}{\sqrt{n}}\right] - \left[\bar{x} - K_2 \frac{\sigma}{\sqrt{n}}\right] = \frac{\sigma}{\sqrt{n}} (K_2 - K_1)$$

distancia que deberá ser mínima pero sujeta a la condición

$$P[K_1 \leq N(0; 1) \leq K_2] = 1 - \alpha.$$

Empleando el método de los multiplicadores de Lagrange se deberá buscar el mínimo de

$$\Phi = \frac{\sigma}{\sqrt{n}} (K_2 - K_1) + \gamma \left[\int_{K_1}^{K_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du - (1 - \alpha) \right],$$

cuyas primeras derivadas parciales son

$$\left. \begin{aligned} \frac{\partial \Phi}{\partial K_2} &= \frac{\sigma}{\sqrt{n}} + \gamma \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}K_2^2} = 0 \\ \frac{\partial \Phi}{\partial K_1} &= -\frac{\sigma}{\sqrt{n}} - \gamma \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}K_1^2} = 0 \\ \frac{\partial \Phi}{\partial \gamma} &= \int_{K_1}^{K_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du - (1 - \alpha) = 0 \end{aligned} \right\}$$

De las dos primeras ecuaciones se obtiene

$$e^{-\frac{1}{2}K_2^2} = e^{-\frac{1}{2}K_1^2},$$

es decir,

$$K_1^2 = K_2^2$$

cuyas posibles soluciones son

- $K_1 = K_2$, incompatible con la existencia del intervalo puesto que hace su longitud nula, $D = 0$.
- $K_1 = -K_2$, conduce a que el intervalo de longitud mínima⁶ sea el simétrico en una distribución $N(0; 1)$, es decir, el formado por los opuestos $(-K; K)$.

Por consiguiente, el intervalo de confianza de longitud mínima para la media μ de una población $N(\mu; \sigma)$ con σ conocida es

$$\left[\bar{x} - K \frac{\sigma}{\sqrt{n}}; \bar{x} + K \frac{\sigma}{\sqrt{n}} \right].$$

EJEMPLO 4

De una población $N(\mu; 3)$ se extrae una muestra aleatoria simple de tamaño 25 siendo

$$\sum_{i=1}^{25} x_i = 60.$$

Determinése el intervalo de confianza del 95% para la media de la población.

Tenemos que

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{60}{25} = 2,4$$

y, por otra parte, en una distribución $N(0; 1)$ si

$$P[-K \leq N(0; 1) \leq K] = 0,95$$

entonces $K = 1,96$.

El intervalo de confianza del 95% para la media μ resulta

$$\left[2,4 - 1,96 \frac{3}{\sqrt{25}}; 2,4 + 1,96 \frac{3}{\sqrt{25}} \right],$$

es decir, la media μ se encontrará incluida previsiblemente en el intervalo

$$[1,224; 3,576]$$

con una confianza del 95%.

⁶ La prueba de la condición suficiente se deja al lector.

5.3.2. INTERVALO DE CONFIANZA PARA LA MEDIA μ DE UNA DISTRIBUCIÓN NORMAL SIENDO σ DESCONOCIDA

Dado el desconocimiento de σ en la distribución poblacional $N(\mu; \sigma)$ la cantidad pivotal a la que se recurre es⁷

$$T(\mathbf{X}; \mu) = \frac{\sqrt{n} (\bar{x} - \mu)}{\sqrt{\frac{1}{n-1} ns^2}} = \frac{\bar{x} - \mu}{\sqrt{\frac{s^2}{n-1}}},$$

con distribución de una variable aleatoria t de Student con $n-1$ grados de libertad, siendo el estadístico s^2 la varianza muestral.

Por la simetría de esta distribución respecto al origen de coordenadas, se llega, al igual que en el caso anterior de la $N(0; 1)$, a que el intervalo de longitud mínima es el formado por los valores opuestos $[-t; t]$. Por tanto, siguiendo el método del pivote, para un intervalo del nivel de confianza $1 - \alpha$ se determinará, en primer lugar, el valor t tal que

$$P[-t \leq t(n-1) \leq t] = 1 - \alpha$$

y, posteriormente,

$$P\left[-t \leq \frac{\bar{x} - \mu}{\sqrt{\frac{s^2}{n-1}}} \leq t\right] \leq 1 - \alpha$$

donde, después de trasponer términos, queda

$$P\left[\bar{x} - t \sqrt{\frac{s^2}{n-1}} \leq \mu \leq \bar{x} + t \sqrt{\frac{s^2}{n-1}}\right] = 1 - \alpha$$

siendo, por consiguiente, el intervalo de confianza de longitud mínima para la media μ el definido por los límites

$$\left[\bar{x} - t \sqrt{\frac{s^2}{n-1}}; \bar{x} + t \sqrt{\frac{s^2}{n-1}}\right].$$

⁷ Véase el capítulo 1.

EJEMPLO 5

En una población $N(\mu; \sigma)$, donde se desconoce el valor del parámetro σ , se obtiene una muestra aleatoria simple de tamaño 10, en la que

$$\sum_{i=1}^{10} x_i = 41 \quad \text{y} \quad \sum_{i=1}^{10} x_i^2 = 229.$$

Hállese el intervalo de confianza del 90% para la media de la población.

Como

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{41}{10} = 4,1$$

$$s^2 = a_2 - \bar{x}^2 = \frac{\sum_{i=1}^n x_i^2}{n} - (4,1)^2 = \frac{229}{10} - (4,1)^2 = 6,09$$

y dado que en una distribución $t(n-1) = t(9)$ el intervalo del 90% se obtiene para un valor t tal que

$$P[-t \leq t(9) \leq t] = 0,90$$

$$t = 1,8331.$$

El intervalo de confianza de longitud mínima para μ en este caso resulta

$$\left[4,1 - 1,8331 \sqrt{\frac{6,09}{9}}; 4,1 + 1,8331 \sqrt{\frac{6,09}{9}} \right] = [2,592; 5,608].$$

5.3.3. INTERVALO DE CONFIANZA PARA LA VARIANZA σ^2 DE UNA POBLACIÓN NORMAL

El pivote lo proporciona el lema de Fisher-Cochran⁸ pues el estadístico

$$T(\mathbf{X}; \sigma^2) = \frac{ns^2}{\sigma^2}$$

sabemos sigue una distribución χ^2 con $n-1$ grados de libertad.

⁸ Véase el capítulo 1.

La distribución χ^2 de Pearson no es simétrica, por lo cual el intervalo de longitud mínima se obtiene para unos límites extremos de cálculo complejo. En este caso, para llegar a un intervalo único, se opta por el criterio de repartir por el igual la probabilidad complementaria a $1 - \alpha$, es decir, se hace

$$\alpha_1 = \alpha_2 = \frac{\alpha}{2}.$$

aunque no sepamos, en muchas ocasiones, si el intervalo es de longitud mínima.

Para calcular los valores tabulares K_1 y K_2 que verifican la expresión

$$P[K_1 \leq \chi^2(n-1) \leq K_2] = 1 - \alpha$$

se tiene en cuenta que

$$P[\chi^2(n-1) \geq K_1] = 1 - \alpha + \frac{\alpha}{2} = 1 - \frac{\alpha}{2}$$

$$P[\chi^2(n-1) \geq K_2] = \frac{\alpha}{2}$$

Determinados K_1 y K_2 , tendremos que

$$P\left[K_1 \leq \frac{ns^2}{\sigma^2} \leq K_2\right] = 1 - \alpha$$

donde, transponiendo términos, se llega a

$$P\left[\frac{ns^2}{K_2} \leq \sigma^2 \leq \frac{ns^2}{K_1}\right] = 1 - \alpha,$$

siendo el intervalo de confianza para la varianza σ^2 de una población normal

$$\left[\frac{ns^2}{K_2}; \frac{ns^2}{K_1}\right].$$

EJEMPLO 6

Con la misma información muestral del ejemplo anterior, determínese el intervalo de confianza del 80% para la varianza σ^2 de una población normal.

Los valores K_1 y K_2 , extremos del intervalo correspondiente a una distribución $\chi^2(n-1)$, $\chi^2(9)$, se obtienen sabiendo que

$$P[\chi^2(9) \geq K_1] = 0,90$$

$$P[\chi^2(9) \geq K_2] = 0,10$$

resultando, pues

$$K_1 = 4,168$$

$$K_2 = 14,684$$

El intervalo deseado es

$$\left[\frac{10 \cdot 6,09}{14,684}; \frac{10 \cdot 6,09}{4,168} \right] = [4,147; 14,611].$$

5.4 Intervalos de confianza en poblaciones no normales

Cuando la población de donde se extrae la información muestral no tiene un comportamiento probabilístico normal, para la construcción de intervalos de confianza de sus respectivos parámetros se seguirán los criterios que a continuación se exponen.

5.4.1. CASO GENERAL

Si la población tiene una distribución $f(x; \theta)$ no normal para la obtención del intervalo de confianza del parámetro θ se puede utilizar cualquiera de los dos métodos expuestos en el epígrafe 5.2 eligiendo aquel del que, en cada caso concreto, resulte un desarrollo más sencillo.

Pero si no se conoce cuál es la distribución de la población es posible determinar un intervalo de confianza acotado inferiormente, mediante la desigualdad de Chebichev, para el parámetro media μ de la población siempre que se conozca el valor de σ .

En efecto, la desigualdad de Chebichev aplicada a *cualquier* variable aleatoria ξ establece que

$$P\left[|\xi - \mu| \leq K\sqrt{V(\xi)}\right] = 1 - \frac{1}{K^2}.$$

En capítulos anteriores se ha expuesto que un «buen» estimador de la media poblacional μ es la correspondiente media muestral \bar{x}

$$\mu^* = \bar{x}$$

siendo

$$E(\bar{x}) = \mu$$

$$V(\bar{x}) = \frac{\sigma^2}{n}$$

entonces, a través de la desigualdad de Chebichev, puede obtenerse la expresión

$$P\left[|\bar{x} - \mu| \leq K \frac{\sigma}{\sqrt{n}}\right] \geq 1 - \alpha$$

donde la cota inferior del nivel de confianza $1 - \alpha$ es tal que $1 - \alpha = 1 - \frac{1}{K^2}$, derivándose como intervalo de confianza para el parámetro μ

$$\left[\bar{x} - K \frac{\sigma}{\sqrt{n}}; \bar{x} + K \frac{\sigma}{\sqrt{n}}\right].$$

EJEMPLO 7

De una población con $\sigma^2 = 11$, se extrae una muestra aleatoria simple de tamaño $n = 36$ siendo la suma de sus elementos

$$\sum_{i=1}^{36} x_i = 180.$$

Determinese el intervalo de confianza para la media poblacional μ de, por lo menos, el 90%.

Como

$$1 - \alpha = 0,90 = 1 - \frac{1}{K^2},$$

tendremos que $K = 3,162$, y como

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{180}{36} = 5$$

el intervalo de confianza para μ de, por lo menos, el 90% es

$$\left[5 - 3,162 \frac{\sqrt{11}}{\sqrt{36}}; 5 + 3,162 \frac{\sqrt{11}}{\sqrt{36}}\right] = [3,255; 6,745].$$

5.4.2. INTERVALOS DE CONFIANZA BASADOS EN GRANDES MUESTRAS

Si el tamaño muestral n es suficientemente grande, se puede hacer uso de las propiedades asintóticas de los estimadores estudiados en capítulos anteriores, permitiendo establecer la siguiente casuística:

A. Intervalos de confianza para un parámetro θ

Para determinar el intervalo de confianza de un parámetro θ , siempre que la muestra sea de gran tamaño, se puede partir del estimador máximo verosímil $\hat{\theta}$ ⁹ pues, como se sabe, tiene el comportamiento asintótico siguiente

$$\hat{\theta} \xrightarrow{d} N\left(\theta; \sqrt{V(\hat{\theta})}\right),$$

es decir, $\hat{\theta}$ es asintóticamente normal, asintóticamente insesgado (si no es insesgado para cualquier tamaño muestral n) y asintóticamente eficiente (si no es eficiente para cualquier n).

Su varianza asintótica coincide con el inverso de la información de Fisher,

$$V(\hat{\theta}) = [I(\theta)]^{-1} = \frac{1}{n E \left[\frac{\partial \ln f(x; \theta)}{\partial \theta} \right]^2}.$$

Estos resultados permiten definir la cantidad pivotal

$$T(\mathbf{X}; \theta) = \frac{\hat{\theta} - \theta}{\sqrt{V(\hat{\theta})}} \xrightarrow{d} N(0; 1)$$

a partir de la cual se genera el intervalo de confianza para el parámetro θ siguiente

$$\left[\hat{\theta} - K \sqrt{V(\hat{\theta})}; \hat{\theta} + K \sqrt{V(\hat{\theta})} \right].$$

Si no se conoce $V(\hat{\theta})$ puede sustituirse por su estimación $\hat{V}(\hat{\theta})$, sin que ello afecte apreciablemente la bondad de la aproximación.

⁹ Véase: propiedades de los estimadores máximo verosímiles en el capítulo 4.

EJEMPLO 8

En una población con distribución exponencial dada por

$$f(x; \theta) = \frac{1}{\theta} e^{-\frac{1}{\theta}x}; \quad x \geq 0, \quad \theta > 0.$$

Se extrae una muestra de tamaño $n = 100$ en la que

$$\sum_{i=1}^{100} x_i = 214 \quad \text{y} \quad \sum_{i=1}^{100} x_i^2 = 1.016.$$

Determinese el intervalo de confianza del 90% para el parámetro θ .

Puede comprobarse, fácilmente, que el estimador máximo verosímil del parámetro θ en esta población es

$$\hat{\theta} = \bar{x}$$

y su varianza

$$V(\hat{\theta}) = V(\bar{x}) = \frac{\sigma^2}{n} = \frac{\theta^2}{n}$$

siendo la varianza poblacional desconocida, puesto que $\sigma^2 = \theta^2$.¹⁰

Por la propiedad de invarianza de los estimadores máximo verosímiles, tendremos que

$$V(\hat{\theta}) = \frac{(\hat{\theta}^2)}{n} = \frac{\bar{x}^2}{n}.$$

Para la muestra obtenida

$$\hat{\theta} = \bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{214}{100} = 2,14$$

$$V(\hat{\theta}) = \frac{\bar{x}^2}{n} = \frac{2,14^2}{100} = 0,045796.$$

Para un nivel de confianza del 90%, el valor tabular K se determina sabiendo que

$$P[-K \leq N(0; 1) \leq K] = 0,90,$$

donde $K = 1,64$, y el intervalo de confianza para θ resulta

$$P[2,14 - 1,64 \sqrt{0,045796}; 2,14 + 1,64 \sqrt{0,045796}] = [2,105; 2,491].$$

¹⁰ La demostración se deja como ejercicio al lector.

B. Intervalo de confianza para la proporción p de una población dicotómica

En una población binomial $B(1; p)$ se toma una muestra aleatoria simple de tamaño n .

El estimador máximo verosímil del parámetro p es

$$\hat{p} = \bar{x} = \frac{\sum_{i=1}^n x_i}{n},$$

donde cada variable x_i toma el valor 1 si presenta la característica estudiada y 0 en caso contrario.

Sabemos que

$$E(\hat{p}) = E(\bar{x}) = \mu = p$$

$$V(\hat{p}) = V(\bar{x}) = \frac{\sigma^2}{n} = \frac{pq}{n} = \frac{p(1-p)}{n}.$$

Por otra parte, por el teorema de Moivre-Laplace, el estadístico \hat{p} tiene una distribución asintóticamente normal dada por

$$\hat{p} \xrightarrow{d} N\left(p; \sqrt{\frac{p(1-p)}{n}}\right)$$

lo que facilita la construcción del pivote

$$T(\mathbf{X}; p) = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \rightarrow N(0; 1),$$

expresión en la que se ha sustituido

$$V(\hat{p}) = \frac{p(1-p)}{n}$$

por su estimación.

Si se fija el nivel de confianza en $1 - \alpha$, se obtiene la constante tabular K como se ha hecho habitualmente, tomando

$$P[-K \leq N(0; 1) \leq K] = 1 - \alpha,$$

siendo el intervalo de confianza para el parámetro p resultante

$$\left[\hat{p} - K \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}; \hat{p} + K \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right].$$

EJEMPLO 9

De una urna con bolas blancas y negras se obtiene una muestra aleatoria simple de tamaño $n = 140$, observándose 80 bolas blancas.

Hállese el intervalo de confianza de la proporción p de bolas blancas de la urna al nivel del 99,7%.

Al repetir el experimento $B(1; p)$, extracción de bolas con reposición de la urna, un número de 140 veces con un resultado de 80 bolas blancas, el estimador máximo verosímil de la proporción p es

$$\hat{p} = \frac{\sum_{i=1}^n x_i}{n} = \frac{80}{140} = 0,57.$$

El valor tabular K se determina teniendo en cuenta que

$$P[-K \leq N(0; 1) \leq K] = 0,997,$$

donde $K \approx 3$.

Por tanto, el intervalo de confianza para p es

$$\left[0,57 - 3 \sqrt{\frac{0,57(1-0,57)}{140}}; 0,57 + 3 \sqrt{\frac{0,57(1-0,57)}{140}} \right] = [0,445; 0,696].$$

5.5 Determinación del tamaño de la muestra

En este epígrafe consideraremos cómo se puede fijar el tamaño de la muestra en los casos de estimación de una media poblacional μ o de una proporción p , parámetros más frecuentemente utilizados en una investigación.

A. Determinación del tamaño de la muestra para estimar la media μ de una población con σ conocida

Para una población normal $N(\mu; \sigma)$ se ha visto anteriormente que el intervalo de confianza del parámetro μ al nivel $1 - \alpha$ viene dado por la expresión

$$P\left[\bar{x} - K \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + K \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

que puede definirse equivalentemente por

$$P\left[|\bar{x} - \mu| \leq K \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

donde al ser $\mu^* = \bar{x}$, la expresión $|\bar{x} - \mu| = |\mu^* - \mu|$ representa el error «e» que se comete en el proceso de estimación.

Si situamos dicho error «e» en el límite aceptado, es decir, si hacemos

$$e = |\bar{x} - \mu| = K \frac{\sigma}{\sqrt{n}},$$

es posible determinar el tamaño de la muestra n a través de la igualdad

$$n = \frac{K^2 \sigma^2}{e^2}$$

que permite calcular, fijado un nivel de confianza $1 - \alpha$ del que depende el valor tabular K y el error máximo «e» que se está dispuesto a aceptar, el tamaño de la muestra n que habría que tomar, bajo el supuesto de que σ^2 es conocida.

EJEMPLO 10

Determinese el tamaño de la muestra necesario para estimar la media μ de una población con $\sigma = 4,2$, si se quiere tener una confianza del 95% de que el error de estimación se sitúe, como mucho, entre $\pm 0,05$.

Como $1 - \alpha = 0,95$, en las tablas de la $N(0; 1)$ se obtiene $K = 1,96$, por tanto,

$$n = \frac{(1,96)^2 \cdot (4,2)^2}{(0,05)^2} = 27.106 \text{ elementos muestrales}$$

B. Determinación del tamaño de la muestra para estimar la media μ de una población con σ desconocida

Partimos del intervalo de confianza de la media μ al nivel $1 - \alpha$

$$P \left[\bar{x} - t \sqrt{\frac{s^2}{n-1}} \leq \mu \leq \bar{x} + t \sqrt{\frac{s^2}{n-1}} \right] = 1 - \alpha .$$

puediendo expresarse este intervalo, también, como

$$P \left[|\bar{x} - \mu| \leq t \sqrt{\frac{s^2}{n-1}} \right] = 1 - \alpha$$

y, por el mismo razonamiento anterior, el tamaño muestral n se determina a partir de la expresión

$$n = \frac{t^2 s^2}{e^2} + 1 .$$

C. Determinación del tamaño de la muestra para estimar la proporción p de una población

Análogamente a lo contemplado hasta ahora, se considera el correspondiente intervalo de confianza para el parámetro p al nivel $1 - \alpha$

$$P \left[\hat{p} - K \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + K \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right] = 1 - \alpha ,$$

donde la probabilidad de cometer un error $e = |\hat{p} - p|$ viene dada por

$$P \left[|\hat{p} - p| \leq K \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right] = 1 - \alpha$$

concluyéndose que

$$n = \frac{K^2 \hat{p}(1-\hat{p})}{e^2} .$$

Si no se dispone de una estimación previa de p , por ejemplo, a través de una muestra piloto de tamaño reducido, se puede utilizar la cota superior

$$n = \frac{K^2}{4e^2}$$

pues

$$\text{máx } \hat{p}(1 - \hat{p}) = \frac{1}{4}.$$

EJEMPLO 11

Determinése el tamaño muestral necesario para estimar la proporción p de una población, con una probabilidad del 99,7% de que el error cometido en la estimación sea como mucho del $\pm 2\%$.

Al nivel $1 - \alpha = 0,997$, le corresponde un valor tabular, en la $N(0; 1)$, $K \approx 3$, por tanto,

$$n = \frac{3^2}{4 \cdot (0,02)^2} = 5.625.$$

CAPÍTULO 6

Contrastación de hipótesis

6.1 Conceptos generales

La función de probabilidad de una variable aleatoria ξ , $f(x; \theta)$, discreta o continua, depende de un parámetro θ que toma valores en el espacio paramétrico Θ , $\theta \in \Theta$, de tal forma que para cada valor de θ en Θ , la función $f(x; \theta)$ es distinta¹. Una **hipótesis estadística** sobre el parámetro es una conjetura sobre los valores concretos que pueda tomar.

El establecimiento de una hipótesis sobre θ supone dividir su espacio paramétrico en dos partes: una, que denominamos Θ_0 , integrada por el conjunto de valores de θ que cumplen la hipótesis otra, Θ_1 , por el conjunto de valores que no la cumplen, es decir, el conjunto complementario del primero. Los dos conjuntos Θ_0 y Θ_1 son disjuntos por definición, $\Theta_0 \cup \Theta_1 = \Theta$, siendo tal división una partición del espacio paramétrico Θ .

¹ θ puede ser un vector integrado por k parámetros, y el espacio paramétrico Θ será k -dimensional.

A la hipótesis que se desea contrastar la denominamos **hipótesis nula** H_0 [$\theta \in \Theta_0$] y, a la otra, **hipótesis alternativa** H_1 [$\theta \in \Theta_1$]. La asignación del término *nula*² o *alternativa* a una u otra hipótesis es arbitraria si bien, tradicionalmente, se denomina nula la hipótesis que implica el valor existente del parámetro, o la que suponemos más estable, siendo precisa una elevada evidencia para rechazarla.

EJEMPLO 1

En la distribución $B(1; p)$ el campo de variación del parámetro p es el intervalo $(0; 1)$. Una hipótesis nula puede ser la pertenencia de p al intervalo $\Theta_0 = (0; 0,3]$ y la alternativa a $\Theta_1 = (0,3; 1)$, es decir, $H_0 = \{0 < p \leq 0,3\}$, $H_1 = \{0,3 < p < 1\}$.

Si los subconjuntos Θ_0 y Θ_1 se componen de un solo elemento (θ_0 y θ_1 , respectivamente) las hipótesis correspondientes se denominan **simples** y, en caso contrario, **compuestas**. En la hipótesis simple, la distribución de probabilidad queda perfectamente determinada (y es única), cosa que no sucede en las compuestas, donde coexiste un cierto número de ellas, número que puede ser infinito.

El problema que se plantea en un contraste de hipótesis es la determinación de cuál de las dos hipótesis puede ser aceptada. A continuación se expone la metodología para dar solución a este problema.

6.2 Contraste de hipótesis

Un contraste o test de hipótesis es una regla de decisión mediante la cual optamos por una u otra hipótesis, a la luz de la información proporcionada por una muestra extraída de la población objeto de estudio.

El procedimiento para llevar a cabo un contraste es el siguiente: se procede a una partición del espacio muestral \mathfrak{X} en dos subconjuntos *disjuntos*, C y C^* , de tal forma que si el punto muestra X pertenece a uno de ellos, por ejemplo a C , llama-

² El término *hipótesis nula* fue introducido por Fisher en su exposición sobre el caso de la dama y el té con leche para representar la hipótesis defendida por este investigador: la nula posibilidad de que la dama pudiera distinguir el orden en que se vertieron el té y la leche.

do **región crítica**, se rechaza la hipótesis nula y si, por el contrario, pertenece al subconjunto complementario C^* , **región de aceptación**, se acepta la hipótesis nula. El rechazo de la hipótesis nula equivale a la aceptación de la alternativa, y viceversa. Debiendo entender la aceptación o rechazo de una hipótesis en el sentido que la muestra ha proporcionado evidencia suficiente, pero no absoluta, para que sea razonable la aceptación o el rechazo de la hipótesis.

EJEMPLO 2

El peso de un producto oscila entre 1 y 4 kg y puede distribuirse con media 2 kg o 3 kg. Se toma una muestra aleatoria de tamaño uno, si el peso es mayor que 2,6 kg se rechaza la hipótesis que la media sea igual a 2 kg y se acepta, por consiguiente, que sea igual a 3 kg.

El espacio muestral \mathfrak{X} es el intervalo [1; 4], la región crítica el subintervalo $C = [2,6; 4]$ y la región de aceptación el subintervalo $C^* = [1; 2,6)$, de tal forma que

$$\mathfrak{X} = C^* \cup C = [1; 2,6) \cup [2,6; 4] = [1; 4].$$

La solución dada al problema de la contrastación de las dos hipótesis implica la posibilidad de acertar o fracasar en la elección al no saber con certeza cuál es la verdadera. La situación queda reflejada en el cuadro siguiente:

		<i>Decisión</i>	
		<i>Aceptar H_0</i>	<i>Rechazar H_0</i>
<i>Hipótesis cierta</i>	H_0	Correcta	Errónea
	H_1	Errónea	Correcta

que expresado de otra manera dice que

- Si la hipótesis nula es cierta y se acepta *la decisión es correcta*.
- Si la hipótesis nula es cierta y se rechaza *la decisión es errónea*, y a este error se le denomina **error de Tipo I**, o de primera especie.
- Si la hipótesis alternativa es cierta y se acepta *la decisión es correcta*.
- Si la hipótesis alternativa es cierta y se rechaza *la decisión es errónea*, error denominado **error de Tipo II** o de segunda especie.

Las situaciones de error, como las de acierto, son desconocidas e incontrolables de manera cierta, sin embargo, procuraremos establecer controles sobre ellas mediante el conocimiento de las probabilidades de cometer los mencionados errores. Para ello distinguiremos el caso de hipótesis simple del relativo a hipótesis compuestas.

6.2.1. HIPÓTESIS SIMPLES

La probabilidad de cometer el error de Tipo I (*rechazar la hipótesis nula siendo cierta*) se llama **nivel de significación** del contraste, o **tamaño de la región crítica** o del contraste, y se designa por α .

La probabilidad de cometer el error de Tipo II no tiene nombre particular y se representa por β , utilizándose preferentemente su complementario a la unidad, probabilidad llamada **potencia del contraste** que es la *probabilidad de rechazar la hipótesis nula siendo falsa*³.

El cálculo de las probabilidades anteriores se realiza de la siguiente forma. El suceso que el punto muestra \mathbf{X} pertenezca a la región crítica, $\mathbf{X} \in C$, implica rechazar la hipótesis nula y si ésta es cierta designaremos el suceso por $\{\mathbf{X} \in C/H_0\}$, y si la alternativa es cierta por $\{\mathbf{X} \in C/H_1\}$.

Según esto, el nivel de significación del contraste es

$$\alpha = P(\text{error de Tipo I}) = P(\text{rechazar } H_0 \text{ siendo cierta}) = P(\mathbf{X} \in C/H_0)$$

y

$$\beta = P(\text{error de Tipo II}) = P(\text{aceptar } H_0 \text{ siendo falsa}) = P(\mathbf{X} \in C^*/H_1).$$

En la práctica, en vez de utilizar $P(\text{error de Tipo II})$, β , se recurre, como dijimos, a su complementario, la potencia del contraste

$$\begin{aligned} 1 - \beta &= P(\text{rechazar } H_0 \text{ siendo falsa}) = 1 - P(\text{aceptar } H_1 \text{ siendo falsa}) = \\ &= 1 - P(\mathbf{X} \in C^*/H_1) = P(\mathbf{X} \in C/H_1). \end{aligned}$$

En el ejemplo siguiente vemos que es posible conocer estas probabilidades.

EJEMPLO 3

En una población $N(\mu; 2)$ tenemos la hipótesis nula $H_0 [\mu = 1]$ y la alternativa $H_1 [\mu = 4]$. Se toma una muestra aleatoria de tamaño uno y se considera como región crítica el intervalo $[2; \infty)$, es decir, si el valor muestral es igual o superior a 2 se rechaza H_0 , en caso contrario se acepta.

La probabilidad del error de Tipo I, nivel de significación, es la probabilidad de que el valor muestral pertenezca a la región crítica, $[2; \infty)$, cuando es cierta la hipótesis nula,

³ Este suceso es equivalente al de aceptar la hipótesis alternativa siendo cierta.

$H_0 [\mu = 1]$. En estas condiciones no tenemos más que hallar en una distribución $N(1; 2)$ la probabilidad del suceso $\{x_1 \geq 2\}$

$$\begin{aligned}\alpha &= P(\text{error de Tipo I}) = P[x_1 \geq 2/N(1; 2)] = \\ &= P(1 + 2\xi \geq 2) = P[N(0; 1) \geq 0,5] = 0,3085\end{aligned}$$

con lo cual comprobamos que, efectivamente, aunque no sepamos si la elección ha sido acertada o no, disponemos de un criterio razonable de información. Información completada con el cálculo de la probabilidad del error de Tipo II, probabilidad de aceptar la hipótesis nula siendo falsa, es decir, si se acepta H_0 será porque el valor muestral no pertenece a la región crítica y sí al intervalo complementario $(-\infty; 2)$, siendo cierta la hipótesis alternativa, $N(4; 2)$,

$$\begin{aligned}\beta &= P(\text{error de Tipo II}) = P[x_1 < 2/N(4; 2)] = \\ &= P(4 + 2\xi < 2) = P[N(0; 1) < -1] = 0,1587.\end{aligned}$$

Dado el desconocimiento que el experimentador tiene sobre qué hipótesis es la correcta no sabrá en cuál de las cuatro situaciones descritas se encuentra, dos correctas y dos incorrectas. Ante la posibilidad de cometer un error el experimentador pretende protegerse⁴ haciendo que la probabilidad de cometerlo sea mínima, siendo la situación ideal fijar el nivel de significación lo menor posible por ser la probabilidad de cometer un error⁵ y, *simultáneamente* hacer la potencia lo mayor posible por ser la probabilidad de un acierto.

Este planteamiento no puede llevarse a cabo *simultáneamente* pues *el nivel de significación y la potencia del contraste no son independientes*, es decir, *no se pueden fijar arbitrariamente por separado* (véase el ejemplo 4).

No obstante, es preciso establecer alguna clase de control sobre las probabilidades de los dos errores a fin de minimizar alguna de ellas. La elección entre el mínimo nivel de significación o la máxima potencia depende del coste que suponga cometer el error de Tipo I o el error de Tipo II. *Convencionalmente*, se fija el nivel de significación en valores reducidos, por ejemplo, 0,10, 0,05, 0,01, etc.

Tal y como se ha definido el procedimiento para contrastar hipótesis, una partición del espacio muestral \mathfrak{X} en dos subconjuntos disjuntos, el número de contrastes para distinguir entre dos hipótesis concretas es elevado⁶, tantos como particiones distintas se puedan hacer en el espacio muestral \mathfrak{X} , por lo cual parece razonable plantearse la posibilidad de que unas particiones conduzcan a mejores resultados que otras, en el sentido que la conclusión sea más o menos segura.

⁴ Esta precaución no es otra cosa que un intento de controlar la aparición de errores.

⁵ De hecho estamos planteando la existencia de un suceso raro y, como tal, poco probable (Ley del azar).

⁶ Incluso infinito.

EJEMPLO 4

Consideremos la variable aleatoria con función de densidad $\theta e^{-\theta x}$, para $\theta > 0$ y $x \geq 0$. Se contrasta la hipótesis nula $H_0 [\theta = 1]$ frente a la alternativa $H_1 [\theta = 5]$, mediante una muestra aleatoria de tamaño uno, siendo la región crítica el intervalo $\{0 \leq x_1 \leq K\}$, lo que implica que la de aceptación es $\{x_1 > K\}$ y

$$\{0 \leq x_1 \leq K\} \cup \{x_1 > K\} = \{x_1 \geq 0\}.$$

Si el nivel de significación es α , el valor de K , constante que delimita la región crítica⁷, se calcula del siguiente modo

$$\begin{aligned} \alpha &= P(\text{rechazar } H_0 \text{ siendo cierta}) = P(x_1 \leq K/H_0 [\theta = 1]) = \\ &= \int_0^K e^{-x_1} dx_1 = 1 - e^{-K} = \alpha \end{aligned}$$

y despejando de la ecuación $1 - e^{-K} = \alpha$ la constante K , tenemos $K = -\ln(1 - \alpha)$. La solución indica que la región crítica depende, como sabíamos, del nivel de significación. Hacemos, a continuación, la potencia del contraste

$$\begin{aligned} \text{Potencia} &= 1 - \beta = P(X \in C/H_1 [\theta = 5]) = P(x_1 \leq -\ln(1 - \alpha)/H_1) = \\ &= \int_0^{-\ln(1-\alpha)} 5e^{-5x_1} dx_1 = 1 - e^{5 \ln(1-\alpha)} = 1 - (1 - \alpha)^5 = 1 - \beta, \end{aligned}$$

es decir, $\beta = (1 - \alpha)^5$. Del resultado se deduce, claramente, que la relación entre la potencia del contraste y el nivel de significación es directa (Fig. 6.1); cuando disminuye α , situación deseable, disminuye también $1 - \beta$, situación no deseable.

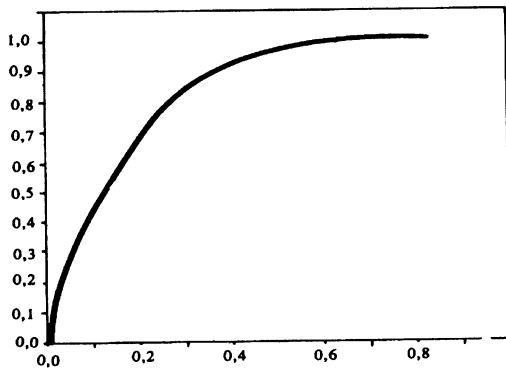


FIGURA 6.1

Relación entre el nivel de significación y la potencia del contraste

⁷ Partición del espacio muestral en dos intervalos.

Si para un mismo nivel de significación⁸ un contraste presenta mayor potencia que la de cualquier otro, diremos que el primero es la **mejor región crítica**, siendo preciso destacar que no siempre existe este tipo de contraste.

Si en un experimento estuviera involucrado un número de contrastes reducido, todos del mismo tamaño α , la obtención del más potente no ofrece especial dificultad. La situación cambia completamente cuando precisamos el contraste más potente de entre todos los posibles contrastes del mismo tamaño. La solución a este problema, cuando existe el contraste uniformemente más potente, la proporciona el **lema de Neyman-Pearson**.

6.2.2. LEMA DE NEYMAN-PEARSON

Sobre el parámetro θ de una función de cuantía o densidad $f(x; \theta)$ se establecen dos hipótesis simples: nula $H_0 [\theta = \theta_0]$ y alternativa $H_1 [\theta = \theta_1]$. Se toma una muestra aleatoria simple de tamaño n cuya función de verosimilitud es $L(\mathbf{X}; \theta)$ y se particulariza para cada una de las hipótesis, $L(\mathbf{X}; \theta_0)$ en la nula y $L(\mathbf{X}; \theta_1)$ en la alternativa.

Se divide el espacio muestral \mathfrak{X} en dos subconjuntos disjuntos C y C^* , $\mathfrak{X} = C \cup C^*$, siendo C la región crítica y C^* la región de aceptación.

Si cuando la muestra \mathbf{X} pertenece a C ($\mathbf{X} \in C$) se verifica

$$\frac{L(\mathbf{X}; \theta_0)}{L(\mathbf{X}; \theta_1)} \leq K$$

mientras que si $\mathbf{X} \in C^*$

$$\frac{L(\mathbf{X}; \theta_0)}{L(\mathbf{X}; \theta_1)} > K,$$

el contraste que se obtiene es óptimo, el que proporciona la mejor región crítica.

En estas condiciones, si el contraste se realiza con un nivel de significación α , la región crítica C tiene mayor o igual potencia que la de cualquier otra región de ese mismo tamaño⁹.

⁸ Al fijar el nivel de significación se eliminan todos los contrastes cuyo tamaño sea distinto.

⁹ Esto no supone necesariamente que sea la única región con potencia mayor o igual que el resto sino que es *una* de las regiones que tiene potencia mayor o igual, de tal manera que si hay varias regiones con la misma potencia, es indiferente utilizar una u otra para el contraste. Tampoco implica que la potencia sea elevada sino que es la mayor de todas ellas.

Consideramos otra región crítica A , del mismo tamaño α , no siendo C y A disjuntas, $C \cap A \neq \emptyset$, y establecemos las dos igualdades siguientes

$$C = \{C \cap A\} \cup \{C \cap A^*\}$$

$$A = \{C \cap A\} \cup \{C^* \cap A\}$$

como se indica en la figura 6.2

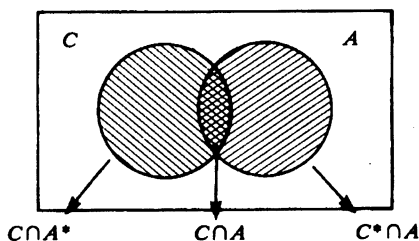


FIGURA 6.2

Teniendo en cuenta las particiones efectuadas en cada región, la probabilidad de que un punto muestra \mathbf{X} pertenezca a la región crítica C es

$$P(\mathbf{X} \in C) = P(\mathbf{X} \in \{C \cap A\}) + P(\mathbf{X} \in \{C \cap A^*\})$$

y a la región crítica A

$$P(\mathbf{X} \in A) = P(\mathbf{X} \in \{C \cap A\}) + P(\mathbf{X} \in \{C^* \cap A\})$$

restándolas

$$P(\mathbf{X} \in C) - P(\mathbf{X} \in A) = P(\mathbf{X} \in \{C \cap A^*\}) - P(\mathbf{X} \in \{C^* \cap A\}). \quad [1]$$

Bajo la hipótesis nula $P(\mathbf{X} \in C/H_0)$ y $P(\mathbf{X} \in A/H_0)$ son los niveles de significación del contraste en cada región crítica pero al ser, por hipótesis, las dos regiones críticas del mismo tamaño α se verifica que

$$P(\mathbf{X} \in C/H_0) - P(\mathbf{X} \in A/H_0) = 0$$

y de [1] tenemos

$$P(\mathbf{X} \in \{C \cap A^*\}/H_0) - P(\mathbf{X} \in \{C^* \cap A\}/H_0) = 0$$

que, en el caso continuo¹⁰, estas probabilidades se pueden expresar como

$$\int_{C \cap A^*} L(\mathbf{X}; \theta_0) d\mathbf{X} - \int_{C^* \cap A} L(\mathbf{X}; \theta_0) d\mathbf{X} = 0. \quad [2]$$

Si la hipótesis alternativa es cierta, las probabilidades $P(\mathbf{X} \in C/H_1)$ y $P(\mathbf{X} \in A/H_1)$ son las potencias de cada una de las regiones críticas, debiéndose demostrar que la potencia de C es, al menos, igual que la de A , es decir, que en las condiciones del lema se verifica

$$P(\mathbf{X} \in C/H_1) \geq P(\mathbf{X} \in A/H_1).$$

La diferencia de probabilidades bajo la hipótesis alternativa según [1] es

$$\begin{aligned} P(\mathbf{X} \in C/H_1) - P(\mathbf{X} \in A/H_1) &= P(\mathbf{X} \in \{C \cap A^*\}/H_1) - P(\mathbf{X} \in \{C^* \cap A\}/H_1) = \\ &= \int_{C \cap A^*} L(\mathbf{X}; \theta_1) d\mathbf{X} - \int_{C^* \cap A} L(\mathbf{X}; \theta_1) d\mathbf{X}. \end{aligned} \quad [3]$$

Según las condiciones iniciales del lema, si el punto muestra pertenece a la región crítica C y, en particular, a $C \cap A^*$, tiene lugar que

$$L(\mathbf{X}; \theta_0) \leq K L(\mathbf{X}; \theta_1)$$

y fuera de ella, en C^* y, en particular, si pertenece a $C^* \cap A$, se verifica que

$$L(\mathbf{X}; \theta_0) > K L(\mathbf{X}; \theta_1).$$

Integrando en estas dos desigualdades

$$\int_{C \cap A^*} L(\mathbf{X}; \theta_0) d\mathbf{X} \leq K \int_{C \cap A^*} L(\mathbf{X}; \theta_1) d\mathbf{X} \quad [4]$$

$$\int_{C \cap A^*} L(\mathbf{X}; \theta_0) d\mathbf{X} > K \int_{C \cap A^*} L(\mathbf{X}; \theta_1) d\mathbf{X}. \quad [5]$$

Multiplicando en [3] por la constante K , sustituyendo las integrales por las cotas obtenidas en [4] y [5], respectivamente,

$$\begin{aligned} K [P(\mathbf{X} \in C/H_1) - P(\mathbf{X} \in A/H_1)] &= \\ &= K \int_{C \cap A^*} L(\mathbf{X}; \theta_1) d\mathbf{X} - K \int_{C^* \cap A} L(\mathbf{X}; \theta_1) d\mathbf{X} \geq \\ &\geq \int_{C \cap A^*} L(\mathbf{X}; \theta_0) d\mathbf{X} - \int_{C^* \cap A} L(\mathbf{X}; \theta_0) d\mathbf{X}. \end{aligned}$$

¹⁰ La demostración para una variable aleatoria discreta es similar.

En la expresión [2] vimos que las dos últimas integrales son iguales, por lo que,

$$K [P(\mathbf{X} \in C/H_1) - P(\mathbf{X} \in A/H_1)] \geq 0$$

eliminando la constante K ¹¹

$$P(\mathbf{X} \in C/H_1) - P(\mathbf{X} \in A/H_1) \geq 0$$

o, equivalentemente,

$$P(\mathbf{X} \in C/H_1) \geq P(\mathbf{X} \in A/H_1)$$

y el lema queda demostrado.

EJEMPLO 5

Sea la función de densidad

$$f(x; \theta) = \theta e^{-\theta x}, \quad \text{para } x \geq 0, \theta > 0.$$

Determinése la mejor región crítica en muestras aleatorias de tamaño n , para contrastar la hipótesis nula $H_0 [\theta = 1]$ frente a la alternativa $H_1 [\theta = 2]$.

La función de verosimilitud es

$$L(\mathbf{X}; \theta) = \theta e^{-\theta x_1} \dots \theta e^{-\theta x_n} = \theta^n e^{-\theta \sum_{i=1}^n x_i}$$

y el cociente de las funciones de verosimilitud bajo cada una de las hipótesis

$$\frac{L(\mathbf{X}; \theta_0)}{L(\mathbf{X}; \theta_1)} = \frac{\theta_0^n e^{-\theta_0 \sum_{i=1}^n x_i}}{\theta_1^n e^{-\theta_1 \sum_{i=1}^n x_i}} = \left(\frac{\theta_0}{\theta_1}\right)^n e^{-(\theta_0 - \theta_1) \sum_{i=1}^n x_i} \leq K_1$$

tomando logaritmos

$$\sum_{i=1}^n x_i \geq K = \frac{n \ln \frac{\theta_1}{\theta_0} - \ln K_1}{\theta_0 - \theta_1}$$

y la forma de la mejor región crítica resulta

$$\sum_{i=1}^n x_i \geq K.$$

¹¹ Que, por ser cota de cociente de funciones de verosimilitud, no es negativa.

Conocida la distribución en el muestreo del estadístico $\sum_{i=1}^n x_i$, bajo H_0 y H_1 , podemos determinar K y, por consiguiente, la mejor región crítica.

Por la definición de nivel de significación

$$\begin{aligned}\alpha &= P(\text{rechazar la } H_0 \text{ cuando es cierta}) = \\ &= P\left[\sum_{i=1}^n x_i \geq K/H_0\right]\end{aligned}$$

necesitando hallar la distribución en el muestreo de $\sum_{i=1}^n x_i$. El estadístico $\sum_{i=1}^n x_i$, suma de n variables aleatorias independientes con la misma distribución, $\theta e^{-\theta x}$, es una variable gamma $G(n; \theta)$ distribuida¹² como

$$f(x; \theta) = \frac{\theta^n}{\Gamma(n)} x^{n-1} e^{-\theta x}.$$

Si disponemos del nivel de significación tenemos que

$$\begin{aligned}\alpha &= P\left[\sum_{i=1}^n x_i \geq K/H_0\right] = P[G(n; \theta) \geq K/H_0] = \\ &= \int_K^{\infty} \frac{\theta^n}{\Gamma(n)} x^{n-1} e^{-\theta_0 x} dx\end{aligned}$$

y de esta ecuación despejamos K .

Como la muestra es de tamaño uno y el nivel de significación igual a 0,05

$$\begin{aligned}0,05 &= \int_K^{\infty} \theta_0 e^{-\theta_0 x} dx = e^{-\theta_0 K}; \\ K &= -\frac{\ln 0,05}{\theta_0}\end{aligned}$$

y siendo $\theta_0 = 1$

$$K = -\ln 0,05 = 2,996.$$

¹² Véase MARTÍN PLIEGO y RUIZ-MAYA: *Fundamentos ...* Capítulo 7.

El lema de Neyman-Pearson no solamente dice que cuando la muestra pertenece a la región crítica, la región del espacio muestra \mathfrak{X} donde se verifica

$$\frac{L(\mathbf{X}; \theta_0)}{L(\mathbf{X}; \theta_1)} \leq K,$$

sea la mejor región crítica entre todas las del mismo tamaño, sino que proporciona un camino para determinar cuál es, y lo hace mediante un estadístico, pudiéndose prescindir de la necesidad de comprobar la pertenencia del punto muestra a la región crítica C , tal como se indica en la cadena de sucesos siguiente

$$\{\mathbf{X} \in C\} \Rightarrow \left\{ \frac{L(\mathbf{X}; \theta_0)}{L(\mathbf{X}; \theta_1)} \leq K \right\} \Rightarrow \{T(\mathbf{X}; \theta_0, \theta_1) \leq K_1\},^{12}$$

por lo que la pertenencia de la muestra a la región C equivale a que el valor del estadístico $T(\mathbf{X}; \theta_0, \theta_1)$ sea menor que una constante K_1 , que habrá que determinar en cada caso.

Como los sucesos anteriores son equivalentes, sus probabilidades son iguales

$$P(\mathbf{X} \in C) = P \left[\frac{L(\mathbf{X}; \theta_0)}{L(\mathbf{X}; \theta_1)} \leq K \right] = P[T(\mathbf{X}; \theta_0, \theta_1) \leq K_1].$$

Si conocemos la distribución en el muestreo del estadístico $T(\mathbf{X}; \theta_0, \theta_1)$, podemos determinar el valor de K_1 , bien al haber establecido el nivel de significación α o la potencia del contraste $1 - \beta$, obteniendo la mejor región crítica

$$\alpha = P(\mathbf{X} \in C/H_0) = P \left[\frac{L(\mathbf{X}; \theta_0)}{L(\mathbf{X}; \theta_1)} \leq K / H_0 \right] = P[T(\mathbf{X}; \theta_0, \theta_1) \leq K_1/H_0]$$

$$1 - \beta = P(\mathbf{X} \in C/H_1) = P \left[\frac{L(\mathbf{X}; \theta_0)}{L(\mathbf{X}; \theta_1)} \leq K / H_1 \right] = P[T(\mathbf{X}; \theta_0, \theta_1) \leq K_1/H_1].$$

En la mayoría de las ocasiones, la determinación de la distribución en el muestreo de $T(\mathbf{X}; \theta_0, \theta_1)$ puede ser complicada, por lo cual, el camino es el de la simplificación hasta llegar a un estadístico $T(\mathbf{X}; \theta_0, \theta_1)$ con distribución en el muestreo conocida o fácil de obtener. Actuar de esta manera implica que sólo nece-

¹² El estadístico $T(\mathbf{X}; \theta_0, \theta_1)$ es el resultado de simplificar el cociente $\frac{L(\mathbf{X}; \theta_0)}{L(\mathbf{X}; \theta_1)}$, pudiendo cambiar el sentido de la desigualdad inicial.

sitamos efectuar el cálculo de la constante K en el último eslabón de la cadena. Bien entendido que, sea cual sea el punto donde nos detengamos, el resultado del contraste será siempre el mismo, pues los sucesos que componen la cadena son equivalentes.

Mediante el ejemplo siguiente aclaramos lo que supone la cadena de sucesos equivalentes o de probabilidades iguales.

EJEMPLO 5

Sea la variable aleatoria ξ con función de densidad

$$f(x; \theta) = (\theta + 1) x^\theta, \quad \text{para } 0 \leq x \leq 1 \text{ y } \theta > 0.$$

Hallaremos la mejor región crítica para el contraste de la hipótesis nula $H_0 [\theta = 1]$ frente a la alternativa $H_1 [\theta = 2]$, tomando una muestra aleatoria de tamaño uno, y suponiendo que el cociente de funciones de verosimilitud sea menor que 0,8

$$\frac{L(x_1; \theta_0)}{L(x_1; \theta_1)} \leq 0,8.$$

En términos del lema de Neyman-Pearson, esta relación dice que si la muestra pertenece a una región crítica donde se verifica esta desigualdad esa región crítica es la mejor, en el sentido que su potencia será la mayor, de entre todas las del mismo tamaño,

$$\frac{L(x_1; \theta_0)}{L(x_1; \theta_1)} = \frac{2x_1}{3x_1^2} = \frac{2}{3x_1} \leq 0,8.$$

Si nos detenemos aquí la mejor región crítica es de la forma $C \Rightarrow \left\{ \frac{2}{3x_1} \leq 0,8 \right\}$, y si seguimos simplificando, llegamos a

$$\left\{ \frac{2}{3x_1} \leq 0,8 \right\} \Rightarrow \left\{ \frac{1}{x_1} \leq 1,2 \right\} \Rightarrow \left\{ \frac{1}{1,2} \leq x_1 \right\}$$

siendo, entonces, la forma de la mejor región crítica $C \Rightarrow \left\{ \frac{1}{1,2} \leq x_1 \right\}$.

Comprobaremos que el nivel de significación calculado en cada una de ellas es el mismo, es decir,

$$P \left[\frac{2}{3x_1} \leq 0,8 / H_0 \right] = P \left[\frac{1}{1,2} \leq x_1 / H_0 \right].$$

Nivel de significación en la región crítica $\left\{ \frac{2}{3x_1} \leq 0,8 \right\}$

$$\begin{aligned} \alpha &= P(\text{rechazar la } H_0 \text{ siendo cierta}) = \\ &= P\left[\frac{L(x_1; \theta_0)}{L(x_1; \theta_1)} \leq 0,8 / H_0 \right] = \\ &= P\left[\frac{2}{3x_1} \leq 0,8 / H_0 \right]. \end{aligned}$$

Como nos hemos detenido en la región crítica $\left\{ \frac{2}{3x_1} \leq 0,8 \right\}$ necesitamos hallar su distribución en el muestreo. Para ello consideramos la variable aleatoria $\eta = \frac{2}{3\xi}$ y calculamos su función de distribución

$$\begin{aligned} G(y; \theta) &= P(\eta \leq y) = P\left[\frac{2}{3\xi} \leq y \right] = P\left[\frac{2}{3y} \leq \xi \right] = \\ &= 1 - P\left[\xi < \frac{2}{3y} \right] = 1 - F\left[\frac{2}{3y}; \theta \right] \end{aligned}$$

y la función de densidad es

$$\begin{aligned} g(y; \theta) &= G'(y; \theta) = -f\left(\frac{2}{3y}; \theta \right) \left(-\frac{2}{3y^2} \right) = \\ &= (\theta + 1) \left(\frac{2}{3} \right)^{\theta+1} y^{-(\theta+2)}. \end{aligned}$$

El campo de variación de η es el intervalo $\left[\frac{2}{3}, \infty \right)$.

Bajo la hipótesis nula, la función $g(y; \theta/H_0)$ es igual a $\frac{8}{9} y^{-3}$.

Volviendo al cálculo del nivel de significación

$$\begin{aligned} \alpha &= P\left[\frac{2}{3x_1} \leq 0,8 / H_0 \right] = P(\eta \leq 0,8 / H_0) = \\ &= \int_{\frac{2}{3}}^{0,8} g(y; \theta/H_0) dy = \int_{\frac{2}{3}}^{0,8} \frac{8}{9} y^{-3} dy = \\ &= 0,3056. \end{aligned}$$

Nivel de significación en la región crítica $\left\{ \frac{1}{1,2} \leq x_1 \right\}$

$$\begin{aligned} \alpha &= P \left[\frac{L(x_1; \theta_0)}{L(x_1; \theta_1)} \leq 0,8 / H_0 \right] = P \left[\frac{1}{1,2} \leq x_1 / H_0 \right] = \\ &= \int_{\frac{1}{1,2}}^1 f(x_1; \theta / H_0) dx = \int_{\frac{1}{1,2}}^1 2x_1 dx = \\ &= 0,3056. \end{aligned}$$

Como vemos se llega, necesariamente, al mismo nivel de significación en ambas regiones críticas.

6.3 Hipótesis compuestas

Una hipótesis es compuesta, como hemos indicado, cuando el subconjunto del espacio paramétrico Θ definido por ella contiene más de un elemento pudiendo ser compuesta la hipótesis nula, la alternativa o ambas. De las distintas situaciones que pueden darse, estudiaremos el caso en el que la hipótesis nula es simple y la alternativa compuesta¹³.

De las posibles hipótesis alternativas compuestas, consideraremos dos tipos

$$(\theta > \theta_0) \text{ ó } (\theta < \theta_0) \text{ y } \theta \neq \theta_0.$$

Esta diferencia de hipótesis compuestas conduce a **contrastos o tests unilaterales** en el primer caso y **bilaterales** en el segundo¹⁴.

En los contrastes unilaterales la región crítica se compone de un solo intervalo, siendo del tipo $T(\mathbf{X}) \geq K$ o $T(\mathbf{X}) \leq K$, pues la hipótesis alternativa se verifica para valores del parámetro mayores o menores que θ_0 , según sea el caso, sin que tenga que haber necesaria correspondencia entre el sentido de la desigualdad inicial y el final de la región crítica.

¹³ El caso en el que la hipótesis nula es compuesta no lo tendremos en cuenta en este manual.

¹⁴ Se conocen, también, como contrastes de **una cola** y de **dos colas**.

Por el contrario, en los tests bilaterales la hipótesis alternativa se verifica para cualquier valor del parámetro distinto de θ_0 , mayores o menores que θ_0 , y la región crítica adopta la forma $|T(\mathbf{X})| \geq K$, es decir,

$$\{[T(\mathbf{X}) \leq -K] \cup [T(\mathbf{X}) \geq K]\}.$$

6.3.1. FUNCIÓN DE POTENCIA

Sea una hipótesis alternativa compuesta $H_1[\theta \in \Theta_1]$ que se opone a la nula $H_0[\theta = \theta_0]$. La función de potencia se define como¹⁵

$$P(\theta) = P(\text{rechazar la hipótesis nula } H_0) = P(\mathbf{X} \in C),$$

resultando que $P(\theta)$ es función del parámetro θ y obteniéndose para cada valor de θ compatible con la hipótesis alternativa, $\theta \in \Theta_1$, la potencia que tendría la alternativa simple correspondiente a ese valor de θ .

Cuando la hipótesis nula es simple, $\theta = \theta_0$, la función de potencia particularizada para ella es igual al nivel de significación, $P(\theta_0) = \alpha$, puesto que

$$P(\theta_0) = P(\mathbf{X} \in C / H_0) = P(\mathbf{X} \in C / \theta = \theta_0) = \alpha.$$

EJEMPLO 6

En la función de densidad

$$f(x; \theta) = \theta e^{-\theta x}, \quad \text{para } x \geq 0, \theta > 0,$$

se plantean las hipótesis $H_0[\theta = 2]$ y $H_1[\theta > 2]$, tomando para el contraste una muestra aleatoria de tamaño uno.

La mejor región crítica según el lema de Neyman-Pearson es $x_1 \leq K$ y si, por ejemplo, el nivel de significación es 0,07

$$\alpha = P(x_1 \leq K / H_0) = \int_0^K 2e^{-2x_1} dx_1 = 1 - e^{-2K} = 0,07$$

y el valor de K es igual a 0,036.

¹⁵ Este concepto sustituye al de potencia del contraste que se utiliza en los contrastes de hipótesis simples. Al ser la hipótesis alternativa H_1 compuesta no se puede asignar un único valor numérico a la probabilidad de rechazar la hipótesis nula cuando es falsa (aceptar H_1 cuando es cierta), por lo que la probabilidad de rechazar la hipótesis nula quedará como función del parámetro θ .

Según esto, la potencia del contraste es

$$\begin{aligned}
 1 - \beta &= P(\text{aceptar } H_1 \text{ siendo cierta}) = \\
 &= P(X \in C/H_1) = P(x_1 \leq 0,036/H_1) = \\
 &= \int_0^{0,036} \theta e^{-\theta x_1} dx_1 = 1 - e^{-0,036\theta}.
 \end{aligned}$$

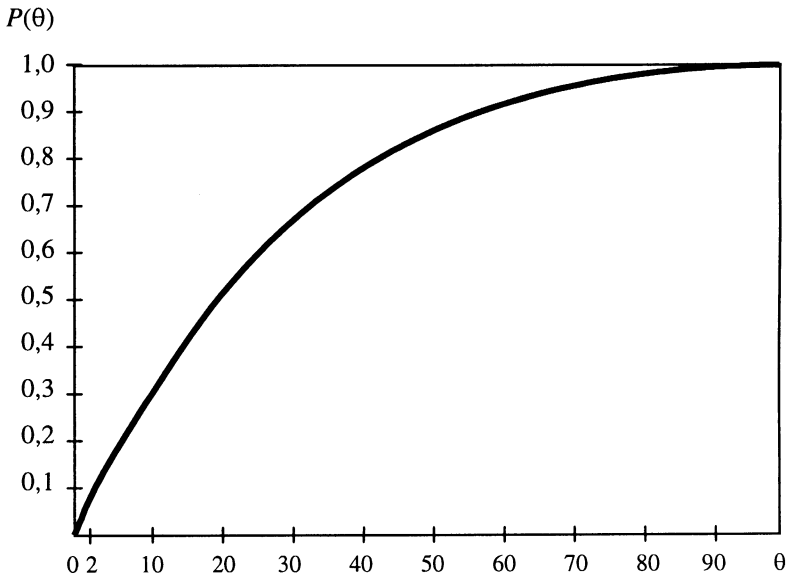


FIGURA 6.3

El resultado indica que no es posible obtener un valor concreto para la potencia del contraste, sino que se llega a una función del parámetro θ , es decir, para cada valor de θ compatible con la hipótesis alternativa (en este caso, $\theta > 2$) se obtiene un valor que será el correspondiente a esa hipótesis simple. En la figura 6.3 apreciamos cómo a medida que el parámetro θ toma valores más alejados de 2, hipótesis nula, la potencia de los posibles contrastes simples tiende a la unidad.

Si $\theta = 10$, la potencia resulta 0,302, valor que corresponde a la potencia del contraste en el que la hipótesis alternativa fuera simple, $H_1 [\theta = 10]$.

■ Contraste inesgado¹⁶

Un contraste es inesgado cuando

$$\max_{\theta \in \Theta_0} P(\text{rechazar } H_0) \leq \min_{\theta \in \Theta_1} P(\text{rechazar } H_0).$$

En el caso de hipótesis simple, el mínimo de la función de potencia coincide con el nivel de significación; es decir, *el contraste será inesgado cuando la probabilidad de rechazar la hipótesis nula siendo cierta siempre es menor que la de rechazarla siendo falsa.*

En el caso de hipótesis nula simple, se tendrá que el nivel de significación (probabilidad de rechazar la hipótesis nula siendo cierta) será siempre menor que la de rechazar H_0 en cualquier otro caso, es decir,

$$\alpha \leq P(\theta) \quad \text{para} \quad \theta \in \Theta_1.$$

El interés de los contrastes inesgados no reside tanto en el hecho, con ser importante, de que el nivel de significación no sea superior a la potencia del contraste, como en que constituyen una clase de contrastes que permiten llegar a la obtención de regiones críticas óptimas.

■ Contraste consistente

Un contraste se denomina consistente si su función de potencia es tal que si $\theta \in \Theta_1$

$$\lim_{n \rightarrow \infty} P(\theta) = 1.$$

6.3.2. CONTRASTES UNIFORMEMENTE MÁS POTENTES (CUMP)

Si existe un contraste óptimo, denominado **contraste uniformemente más potente**, viene determinado por el lema de Neyman-Pearson que establece que la región crítica óptima C debe cumplir las siguientes condiciones¹⁷:

Cuando la muestra \mathbf{X} pertenece a la región crítica C se verifica

$$\frac{L(\mathbf{X}; \theta_0)}{L(\mathbf{X}; \theta)} \leq K$$

¹⁶ Se le denomina, también, **centrado**.

¹⁷ La demostración del lema de Neyman-Pearson en el caso de hipótesis compuestas es análoga a la que se ofrece para hipótesis simples.

mientras que si $\mathbf{X} \in C^*$

$$\frac{L(\mathbf{X}; \theta_0)}{L(\mathbf{X}; \theta)} > K$$

donde $\theta_0 \in \Theta_0$ y $\theta \in \Theta_1$.

El que una hipótesis sea compuesta tiene como consecuencia que al poder tomar el parámetro más de un valor, la distribución de probabilidad de la población queda indeterminada, no siendo posible calcular probabilidades de ciertos sucesos. Una de las consecuencias de plantear hipótesis alternativas compuestas podemos apreciarla en los ejemplos siguientes.

EJEMPLO 7

En la función de densidad

$$f(x; \theta) = \theta e^{-\theta x}, \quad x \geq 0 \quad \text{y} \quad \theta > 0,$$

y para contrastar la hipótesis nula $H_0 [\theta = \theta_0]$ frente a las alternativas $[\theta > \theta_0]$, $[\theta < \theta_0]$ o $[\theta \neq \theta_0]$, hallamos la mejor región crítica mediante el lema de Neyman-Pearson.

El cociente de funciones de verosimilitud es

$$\begin{aligned} \frac{L(\mathbf{X}; \theta_0)}{L(\mathbf{X}; \theta)} &= \left(\frac{\theta_0}{\theta}\right)^n e^{-(\theta_0 - \theta) \sum_{i=1}^n x_i} \leq K_2 \\ e^{-(\theta_0 - \theta) \sum_{i=1}^n x_i} &\leq \left(\frac{\theta}{\theta_0}\right)^n K_2 \\ (\theta_0 - \theta) \sum_{i=1}^n x_i &\geq K_1 \end{aligned}$$

siendo

$$K_1 = -\ln \left(\frac{\theta}{\theta_0}\right)^n K_2$$

Hipótesis alternativa $[\theta < \theta_0]$

Como $\theta_0 > \theta$, $\theta_0 - \theta > 0$ y la mejor región crítica es $\sum_{i=1}^n x_i \geq K$.

Hipótesis alternativa $[\theta > \theta_0]$

Como $\theta_0 < \theta$, $\theta_0 - \theta < 0$, la mejor región crítica es $\sum_{i=1}^n x_i \leq K$.

Hipótesis alternativa [$\theta \neq \theta_0$]

Al no saber si $\theta_0 > \theta$ o $\theta_0 < \theta$ se desconoce el signo de $\theta_0 - \theta$ y no se puede obtener la mejor región crítica mediante el lema de Neyman-Pearson, debiendo recurrir a otros procedimientos, como se expondrá en el capítulo siguiente.

EJEMPLO 8

En una distribución $N(\mu; \sigma)$ con varianza conocida, hallamos la mejor región crítica para el contraste de la hipótesis nula $H_0 [\mu = \mu_0]$ y de la alternativa $H_1 [\mu \neq \mu_0]$. Para ello aplicamos el lema de Neyman-Pearson.

El cociente de funciones de verosimilitud particularizadas para las dos hipótesis es

$$\begin{aligned} \frac{L(\mathbf{X}; \mu_0)}{L(\mathbf{X}; \mu)} &= e^{-\frac{1}{2\sigma^2} \left[\sum_{i=1}^n (x_i - \mu_0)^2 - \sum_{i=1}^n (x_i - \mu)^2 \right]} = \\ &= e^{-\frac{n(\mu_0^2 - \mu^2)}{2\sigma^2} + \frac{n(\mu_0 - \mu)}{\sigma^2} \bar{x}} \leq K_1 \end{aligned}$$

tomando logaritmos

$$\begin{aligned} -\frac{n(\mu_0^2 - \mu^2)}{2\sigma^2} + \frac{n(\mu_0 - \mu)}{\sigma^2} \bar{x} &\leq \ln K_1 \\ (\mu_0 - \mu)\bar{x} \leq K &= \frac{\sigma^2}{n} \left[\ln K_1 + \frac{n(\mu_0^2 - \mu^2)}{2\sigma^2} \right]. \end{aligned}$$

La región crítica adopta la forma $(\mu_0 - \mu)\bar{x} \leq K$.

Para despejar la media muestral necesitamos conocer el signo de la diferencia $\mu_0 - \mu$, pues de él depende el sentido de la desigualdad: si $\mu_0 - \mu > 0$ la región crítica sería $\bar{x} \leq K$ y si $\mu_0 - \mu < 0$ se llegaría a $\bar{x} \geq K$, y dado el desconocimiento del signo no podemos obtener la mejor región crítica aplicando el lema de Neyman-Pearson.

Ejercicios resueltos

EJERCICIO 6.1 Tenemos una población $N(\mu; 1)$. Sobre el parámetro μ se establecen dos hipótesis: hipótesis nula que $\mu = 1$, hipótesis alternativa $\mu = 2$. La región crítica es el intervalo $[2,282; \infty)$. El contraste se efectúa mediante una muestra aleatoria de tamaño uno.

Determinése el nivel de significación y la potencia del contraste.

SOLUCIÓN.

Nivel de significación

Según la definición de nivel de significación es igual a

$$\begin{aligned}\alpha &= P(\text{rechazar la hipótesis nula siendo cierta}) = P(\mathbf{X} \in C/H_0) = \\ &= P[x_1 \geq 2,282/N(1; 1)] = P(\xi + 1 \geq 2,282) = P(\xi \geq 1,282) = 0,10\end{aligned}$$

la variable aleatoria ξ es $N(0; 1)$

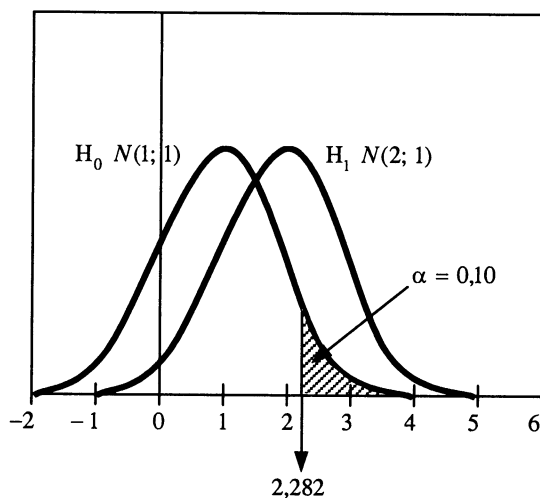


FIGURA 6.4

Potencia del contraste

La potencia del contraste, $1 - \beta$, es igual a

$$\begin{aligned} P(\text{rechazar la hipótesis nula siendo falsa}) &= P(\text{aceptar } H_1 \text{ siendo cierta}) = \\ &= P(\mathbf{X} \in C/H_1) = P(x_1 \geq 2,282/N(2; 1)) = \\ &= P(\xi + 2 \geq 2,282) = P(\xi \geq 0,282) = 0,3889. \end{aligned}$$

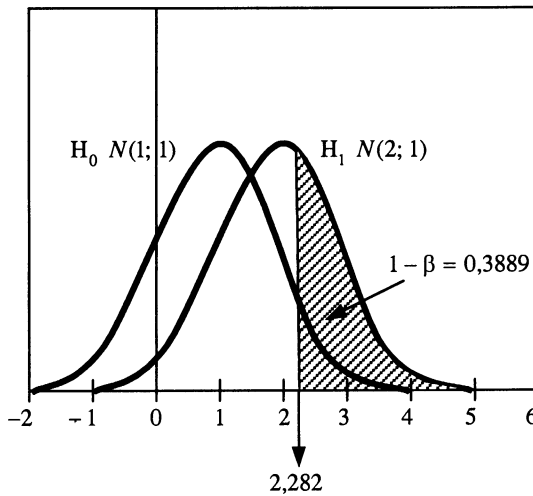


FIGURA 6.5

EJERCICIO 6.2 En el ejercicio 6.1 tenemos una población $N(\mu; 1)$, sobre cuya media μ se establece la hipótesis nula que $\mu = 1$, y la alternativa $\mu = 2$, contrastándolas mediante la región crítica $[2,282; \infty)$, que denominaremos C_1 . Consideramos otra región crítica $C_2 = [0,5; 1,296) \cup [2,96; \infty)$.

Con una muestra aleatoria de tamaño uno, calcúlese en esta última región crítica su nivel de significación y potencia, comparando los resultados con los obtenidos en el ejercicio anterior.

SOLUCIÓN. Los resultados de la región crítica $C_1 = [2,282; \infty)$ son: nivel de significación $\alpha = 0,10$, potencia del contraste $1 - \beta = 0,3889$.

Para la región crítica C_2 $[0,5; 1,296) \cup [2,96; \infty)$

Nivel de significación

$$\begin{aligned} \alpha &= P(\text{rechazar la hipótesis nula siendo cierta}) = P(\mathbf{X} \in C_2/H_0) = \\ &= P(x_1 \in \{[0,5; 1,296) \cup [2,96; \infty)\}/H_0) = \\ &= P(x_1 \in [0,5; 1,296)/H_0) + P(x_1 \in [2,96; \infty)/H_0) = \\ &= P[0,5 < x_1 \leq 1,296/N(1; 1)] + P[x_1 \geq 2,96/N(1; 1)] = \\ &= 0,3079 + 0,025 = 0,3329. \end{aligned}$$

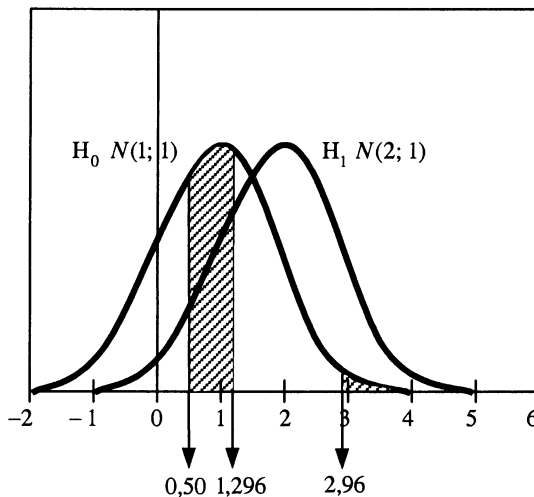


FIGURA 6.6

Potencia del contraste

$$\begin{aligned} 1 - \beta &= P(\text{aceptar } H_1 \text{ siendo cierta}) = \\ &= P(\mathbf{X} \in C_2/H_1) = \\ &= P[0,5 \leq x_1 < 1,296/N(2; 1)] + P[x_1 \geq 2,96/N(2; 1)] = \\ &= 0,1740 + 0,1685 = 0,3425. \end{aligned}$$

Desde el punto de vista del nivel de significación es preferible el primer contraste, pues $\alpha_1 = 0,10 < \alpha_2 = 0,3329$; la potencia del primer contraste es ligeramente mayor que la del segundo $1 - \beta_1 = 0,3889 > 1 - \beta_2 = 0,3425$. Por todo ello, elegiremos el primer contraste al segundo.

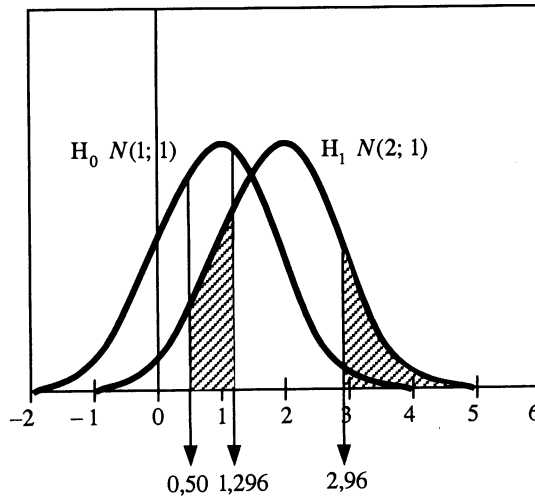


FIGURA 6.7

EJERCICIO 6.3 Hállese la mejor región crítica para contrastar en una población con distribución $B(1; p)$ las hipótesis nula $H_0 [p = p_0]$ y la alternativa $H_1 [p = p_1]$, en muestras aleatorias simples de tamaño n .

SOLUCIÓN. La función de verosimilitud es

$$L(\mathbf{X}; p) = p^{x_1} (1-p)^{1-x_1} \dots p^{x_n} (1-p)^{1-x_n} = p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i}$$

El cociente de las funciones de verosimilitud particularizadas para p_0 y p_1 es igual a

$$\begin{aligned} \frac{L(\mathbf{X}; p_0)}{L(\mathbf{X}; p_1)} &= \frac{p_0^{\sum_{i=1}^n x_i} (1-p_0)^{n-\sum_{i=1}^n x_i}}{p_1^{\sum_{i=1}^n x_i} (1-p_1)^{n-\sum_{i=1}^n x_i}} = \left(\frac{p_0}{p_1}\right)^{\sum_{i=1}^n x_i} \left(\frac{1-p_0}{1-p_1}\right)^{n-\sum_{i=1}^n x_i} = \\ &= \left(\frac{p_0(1-p_1)}{p_1(1-p_0)}\right)^{\sum_{i=1}^n x_i} \left(\frac{1-p_0}{1-p_1}\right)^n \leq K_2 \\ &\left(\frac{p_0(1-p_1)}{p_1(1-p_0)}\right)^{\sum_{i=1}^n x_i} \leq K_1 = K_2 \left(\frac{1-p_1}{1-p_0}\right)^n \end{aligned}$$

y la mejor región crítica resulta ser

$$\sum_{i=1}^n x_i \leq K \quad \text{si} \quad p_0 > p_1$$

y

$$\sum_{i=1}^n x_i \geq K \quad \text{si} \quad p_0 < p_1.$$

EJERCICIO 6.4 En la distribución de Poisson contrástese $H_0 [\lambda = \lambda_0]$ frente a $H_1 [\lambda = \lambda_1]$, $\lambda_0 < \lambda_1$, en muestras aleatorias simples de tamaño n .

SOLUCIÓN. La función de verosimilitud resulta

$$L(\mathbf{X}; \lambda) = \frac{e^{-\lambda} \lambda^{x_1}}{x_1!} \cdots \frac{e^{-\lambda} \lambda^{x_n}}{x_n!} = \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i}}{x_1! \cdots x_n!}.$$

El cociente de funciones de verosimilitud para cada una de las hipótesis es

$$\frac{L(\mathbf{X}; \lambda_0)}{L(\mathbf{X}; \lambda_1)} = \frac{e^{-n\lambda_0} (\lambda_0)^{\sum_{i=1}^n x_i}}{e^{-n\lambda_1} (\lambda_1)^{\sum_{i=1}^n x_i}} = e^{-n(\lambda_0 - \lambda_1)} \left(\frac{\lambda_0}{\lambda_1} \right)^{\sum_{i=1}^n x_i} \leq K_1$$

despejando la suma de los valores muestrales, queda

$$\sum_{i=1}^n x_i \geq K = \frac{\ln K_1 + n(\lambda_0 - \lambda_1)}{\ln \lambda_0 - \ln \lambda_1}.$$

El sentido de la desigualdad cambia porque, al ser $\lambda_0 < \lambda_1$, la diferencia $\ln \lambda_0 - \ln \lambda_1$ es negativa.

EJERCICIO 6.5 La función de densidad de una variable aleatoria es

$$f(x; \theta) = (1 + \theta)x^\theta, \quad \text{para} \quad 0 \leq x \leq 1, \quad \theta > 0.$$

Sobre el parámetro θ se hacen las hipótesis $H_0 [\theta = \theta_0]$ y $H_1 [\theta = \theta_1]$, $\theta_0 > \theta_1$.

Hállese la mejor región crítica en muestras aleatorias simples de tamaño n . Si se conoce el nivel de significación y el tamaño muestral es igual a 1, hállese la mejor región crítica y la potencia del contraste.

SOLUCIÓN. La función de verosimilitud resulta

$$L(\mathbf{X}; \theta) = (1 + \theta)^n (x_1 \cdots x_n)^\theta,$$

y el cociente de funciones de verosimilitud para una y otra hipótesis

$$\frac{L(\mathbf{X}; \theta_0)}{L(\mathbf{X}; \theta_1)} = \left(\frac{1 + \theta_0}{1 + \theta_1} \right)^n (x_1 \cdots x_n)^{\theta_0 - \theta_1} \leq K_1.$$

Tomando logaritmos llegamos a la mejor región crítica

$$\sum_{i=1}^n \ln x_i \leq K.$$

Si la muestra es de tamaño uno, la mejor región crítica es $\ln x_1 \leq K$. Hallamos la distribución en el muestreo de la variable $\eta = \ln \xi$, calculando su función de distribución

$$G(y) = P(\eta \leq y) = P(\ln \xi \leq y) = P(\xi \leq e^y) = F(e^y)$$

$$g(y) = e^y f(e^y) = e^y (1 + \theta) e^{\theta y} = (1 + \theta) e^{(1+\theta)y}.$$

El campo de variación de la variable ξ es el intervalo $[0; 1]$ y el de $\eta = \ln \xi$ el intervalo $(-\infty; 0]$.

Si conocemos el nivel de significación

$$\begin{aligned} \alpha &= P(\ln \xi \leq K/H_0) = P(\eta \leq K/H_0) = \\ &= \int_{-\infty}^K (1 + \theta_0) e^{(1+\theta_0)y} dx = e^{(1+\theta_0)K} \end{aligned}$$

$$K = \frac{\ln \alpha}{1 + \theta_0}$$

y la potencia

$$\begin{aligned} 1 - \beta &= P \left[\ln \xi \leq \frac{\ln \alpha}{1 + \theta_0} / H_1 \right] = \\ &= \int_{-\infty}^{\frac{\ln \alpha}{1 + \theta_0}} (1 + \theta_1) e^{(1+\theta_1)y} dx = \\ &= e^{\frac{1 + \theta_1}{1 + \theta_0} \ln \alpha}. \end{aligned}$$

CAPÍTULO 7

Contrastes de significación

7.1 Principios básicos

En el capítulo anterior se ha estudiado la teoría de Neyman-Pearson relativa a la construcción de **contrastes de hipótesis óptimos**, donde la región crítica utilizada proporcionaba la mayor probabilidad de rechazar la hipótesis nula H_0 cuando ésta era falsa, para un nivel de significación fijado α .

Esta teoría para la determinación de regiones críticas óptimas no siempre conduce a un resultado operativo. Por ejemplo, en una población normal $N(\mu; \sigma)$ con σ conocida si se desea verificar la hipótesis nula $H_0 [\mu = \mu_0]$ frente a la hipótesis alternativa $H_1 [\mu \neq \mu_0]$ no puede obtenerse una región crítica uniformemente más potente que permita realizar dicho contraste mediante los criterios de optimalidad de Neyman-Pearson.¹

Para dar respuesta a situaciones donde no se consigue la formación de una región crítica óptima, puede recurrirse al criterio propuesto por Fisher que conduce a los llamados **contrastes de significación**.

El esquema básico de este tipo de contrastes de significación es el siguiente:

¹ Véase el Capítulo 6.

■ Se plantea por el investigador la hipótesis que considera plausible sobre el valor de un parámetro θ , que determina la distribución de probabilidad de la población $f(x; \theta)$.

Esta hipótesis, llamada hipótesis nula H_0 , se expresa por $H_0[\theta = \theta_0]$.

Como consecuencia de la hipótesis nula aparece la hipótesis alternativa que en lo que estudiaremos en este capítulo será de tipo compuesto, incluso no especificada, presentada por $H_1[\theta \neq \theta_0]$.

■ Mediante una muestra aleatoria simple de tamaño n , $\mathbf{X} (x_1, \dots, x_n)$, se recoge información de la población.

■ Se construye una **medida de discrepancia o estadístico del contraste** que mide la discrepancia entre el valor del parámetro establecido en la hipótesis nula, θ_0 , y el que le podríamos asignar fundamentándonos en la evidencia muestral \mathbf{X} , valor de la estimación del parámetro θ^* . Si denotamos esta medida de discrepancia por D , entonces

$$D = D(\theta_0, \theta^*).$$

La especificación de la forma funcional de D como señala Fisher es, en parte, producto de la imaginación del investigador estadístico que, razonablemente, actuará en el estudio que realiza de manera acorde al interés de lograr una decisión adecuada, lo menos errónea posible. Junto a esto, es necesario que la medida D tenga una distribución de probabilidad determinable bajo el supuesto que H_0 sea cierta.

Como ejemplos de la forma funcional que podría tener esta medida de discrepancia reseñamos los siguientes:

- Desviaciones entre el valor asignado a θ , en función de la evidencia muestral \mathbf{X} , θ^* , y el valor que recoge la hipótesis nula, θ_0 ,

$$D = \theta^* - \theta_0.$$

- Desviaciones absolutas

$$D = |\theta^* - \theta_0|.$$

- Desviaciones cuadráticas

$$D = (\theta^* - \theta_0)^2.$$

- Valor tipificado de la discrepancia como medida adimensional

$$D = \frac{\theta^* - \theta_0}{\sqrt{V(\theta^*)}}.$$

- En el supuesto que la estimación que se utilice se base en un estimador máximo-verosímil, si el tamaño de la muestra $n \rightarrow \infty$ sabemos que

$$\hat{\theta} \xrightarrow{d} N\left(\theta; \sqrt{V(\hat{\theta})}\right)$$

por lo que puede utilizarse como medida de discrepancia la expresión

$$D = \frac{(\hat{\theta} - \theta)^2}{V(\hat{\theta})}$$

que tiene una distribución asintótica χ^2 de Pearson con 1 grado de libertad.

- También pueden recurrirse a otras medidas como las que estudiaremos en los contrastes razón de verosimilitud en los epígrafes siguientes.

■ **Región crítica.** Si se fija un nivel de significación α , la región crítica que se utiliza será $D \geq d_\alpha$, siendo d_α el valor crítico del **estadístico de contraste** D bajo el supuesto de que la hipótesis nula es cierta. Este valor crítico se determina de manera que

$$P[D \geq d_\alpha / H_0] = \alpha.$$

■ La justificación intuitiva de la manera de proceder en los contrastes de significación es la siguiente.

Dado que el nivel de significación α , probabilidad del error de Tipo I, se elige cercano a 0, se rechaza la hipótesis nula H_0 cuando la probabilidad de que se presente una discrepancia igual o mayor que D sea menor que el nivel de significación fijado α , dando a aquélla la denominación de **p-valor**.

Este razonamiento se representa esquemáticamente en la figura 7.1 siguiente.

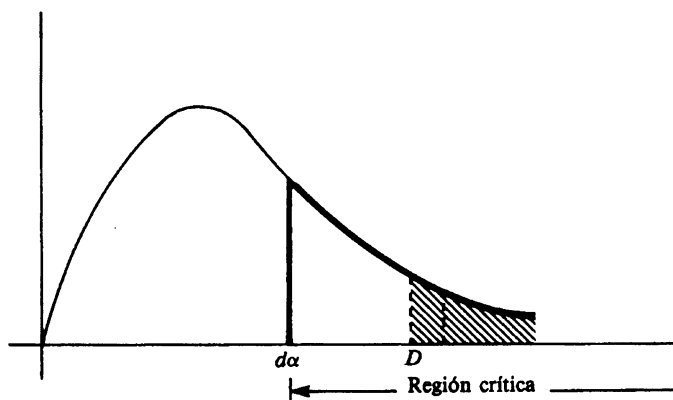


FIGURA 7.1

Si la probabilidad de obtener una discrepancia igual o mayor que D es muy pequeña, D sería un suceso raro, y dado que en la inferencia estadística no se admite, como norma general, el que un suceso raro pueda presentarse² se tiene que concluir que la hipótesis considerada como cierta, H_0 no debe serlo y, por consiguiente, ha de rechazarse. No parece que, en este caso, la evidencia empírica suministrada por la muestra \mathbf{X} sea compatible con la hipótesis nula H_0 y, por ello, no se puede seguir manteniendo esta hipótesis, pues, una probabilidad tan pequeña impide aceptar que H_0 sea cierta. En este caso se dice que la discrepancia es **estadísticamente significativa**.

Si no se procediera de tal manera, se estaría admitiendo la posibilidad de incoherencias entre la hipótesis y los datos observados que reflejarían tales «sucesos raros».

Por el contrario, si $D < d_\alpha$, la discrepancia medida por D se dice que **no es estadísticamente significativa**, entendiéndose esta expresión en el sentido no de la importancia numérica de la misma discrepancia sino en el de la plausibilidad de la hipótesis nula H_0 , como veremos en los casos particulares que se desarrollan a continuación.

Los contrastes de significación no son, generalmente, uniformemente más potentes aunque, en algunos casos, puedan coincidir con los tests óptimos que determina el teorema de Neyman-Pearson.

A título de ejemplo, se consideran dos casos donde puede obtenerse un adecuado contraste de significación.

7.1.1. CONTRASTE DE LA MEDIA DE UNA POBLACIÓN NORMAL CON σ CONOCIDA

En una población normal $N(\mu; \sigma)$ con desviación típica conocida σ se contrasta la hipótesis nula $H_0 [\mu = \mu_0]$ frente a la alternativa $H_1 [\mu > \mu_0]$.

El diseño de un contraste de significación partiendo de una muestra aleatoria simple \mathbf{X} de tamaño n es el siguiente.

Se adopta como medida de discrepancia el estadístico

$$D = \mu^* - \mu_0 = \bar{x} - \mu_0$$

pues un buen estimador de la media poblacional es la media muestral \bar{x} .

² Ley del azar. MARTÍN PLIEGO y RUIZ-MAYA: *Fundamentos...* Capítulo 1.

Si la probabilidad de obtener una discrepancia igual o mayor que D es muy pequeña, D sería un suceso raro, y dado que en la inferencia estadística no se admite, como norma general, el que un suceso raro pueda presentarse² se tiene que concluir que la hipótesis considerada como cierta, H_0 no debe serlo y, por consiguiente, ha de rechazarse. No parece que, en este caso, la evidencia empírica suministrada por la muestra X sea compatible con la hipótesis nula H_0 y, por ello, no se puede seguir manteniendo esta hipótesis, pues, una probabilidad tan pequeña impide aceptar que H_0 sea cierta. En este caso se dice que la discrepancia es **estadísticamente significativa**.

Si no se procediera de tal manera, se estaría admitiendo la posibilidad de incoherencias entre la hipótesis y los datos observados que reflejarían tales «sucesos raros».

Por el contrario, si $D < d_\alpha$, la discrepancia medida por D se dice que **no es estadísticamente significativa**, entendiéndose esta expresión en el sentido no de la importancia numérica de la misma discrepancia sino en el de la plausibilidad de la hipótesis nula H_0 , como veremos en los casos particulares que se desarrollan a continuación.

Los contrastes de significación no son, generalmente, uniformemente más potentes aunque, en algunos casos, puedan coincidir con los tests óptimos que determina el teorema de Neyman-Pearson.

A título de ejemplo, se consideran dos casos donde puede obtenerse un adecuado contraste de significación.

7.1.1. CONTRASTE DE LA MEDIA DE UNA POBLACIÓN NORMAL CON σ CONOCIDA

En una población normal $N(\mu; \sigma)$ con desviación típica conocida σ se contrasta la hipótesis nula $H_0 [\mu = \mu_0]$ frente a la alternativa $H_1 [\mu > \mu_0]$.

El diseño de un contraste de significación partiendo de una muestra aleatoria simple X de tamaño n es el siguiente.

Se adopta como medida de discrepancia el estadístico

$$D = \mu^* - \mu_0 = \bar{x} - \mu_0$$

pues un buen estimador de la media poblacional es la media muestral \bar{x} .

² Ley del azar. MARTÍN PLIEGO y RUIZ-MAYA: *Fundamentos...* Capítulo 1.

Para una muestra procedente de una población normal, la distribución de la media muestral \bar{x} es

$$\bar{x} \rightarrow N\left(\mu; \frac{\sigma}{\sqrt{n}}\right).$$

Por consiguiente, suponiendo que hipótesis nula sea cierta

$$\bar{x} - \mu_0/H_0 \rightarrow N\left(0; \frac{\sigma}{\sqrt{n}}\right)$$

determinándose el valor crítico d_α para el contraste teniendo en cuenta que

$$P(D \geq d_\alpha/H_0) = P(\bar{x} - \mu_0 \geq d_\alpha/H_0) = \alpha.$$

EJEMPLO 1

En una población normal $N(\mu; 2)$ se quiere contrastar $H_0 [\mu = 1,8]$ frente a la alternativa $H_1 [\mu > 1,8]$, a partir de una muestra aleatoria simple de tamaño $n = 30$, donde $\sum_{i=1}^{30} x_i = 56$, con un nivel de significación $\alpha = 0,05$.

En este caso,

$$\bar{x} - \mu_0/H_0 \rightarrow N\left(0; \frac{2}{\sqrt{30}}\right) = N(0; 0,36).$$

El valor crítico d_α es tal que

$$\begin{aligned} P(D \geq d_{0,05}/H_0) &= P(\bar{x} - \mu_0 \geq d_{0,05}/H_0) = \\ &= P(0,36\xi \geq d_{0,05}) = \\ &= P\left[\xi \geq \frac{d_{0,05}}{0,36}\right] = 0,05 \end{aligned}$$

siendo $\xi \rightarrow N(0; 1)$, donde

$$\frac{d_{0,05}}{0,36} = 1,64$$

es decir,

$$d_{0,05} = 0,5904.$$

Como

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{56}{30} = 1,87,$$

entonces

$$D = \bar{x} - \mu_0 = 1,87 - 1,8 = 0,07$$

y al ser $D = 0,07 < d_{0,05} = 0,5904$ no se rechaza la hipótesis nula H_0 , pues la discrepancia registrada para la muestra que se obtuvo no es estadísticamente significativa para suponer que $H_0 [\mu = 1,8]$ no sea cierta al nivel de significación $\alpha = 0,05$.

7.1.2. CONTRASTE DE LA MEDIA DE UNA POBLACIÓN NORMAL CON σ DESCONOCIDA

En esta población, el estadístico media muestral tiene como distribución

$$\bar{x} \rightarrow N\left(\mu; \frac{\sigma}{\sqrt{n}}\right)$$

y

$$\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \rightarrow N(0; 1)$$

pero al desconocerse σ , ha de considerarse una estimación de dicho parámetro.

Si la estimación se basa en la estimación insesgada de la varianza poblacional

$$(\sigma^2)^* = s_1^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{ns^2}{n-1},$$

donde s_1^2 y s^2 son, respectivamente, la cuasivarianza y varianza muestrales, entonces

$$\frac{\bar{x} - \mu}{\frac{s_1}{\sqrt{n}}} = \frac{\sqrt{n}(\bar{x} - \mu)}{\sqrt{\frac{ns^2}{n-1}}}$$

es un estadístico que, como estudiamos en el capítulo 1, tiene una distribución t de Student con $n - 1$ grados de libertad.

Por tanto, en este caso, para contrastar, por ejemplo, $H_0 [\mu = \mu_0]$ frente a la alternativa $H_1 [\mu \neq \mu_0]$, se puede tomar como medida de la discrepancia la desviación tipificada

$$D = \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sqrt{\frac{ns^2}{n-1}}}$$

pues presenta una distribución en el muestreo conocida, $t(n - 1)$, independiente del parámetro σ que se desconoce.

El contraste se realiza siguiendo la pauta general expuesta, tomándose $|D|$, pues tanto las discrepancias negativas como las positivas, si son significativamente grandes, llevan a rechazar H_0

$$P(|D| \geq d_{\alpha}/H_0) = P(|t(n - 1)| \geq d_{\alpha}/H_0) = \alpha.$$

EJEMPLO 2

En una población normal $N(\mu; \sigma)$ se quiere verificar la hipótesis nula $H_0 [\mu = 3,5]$ frente a la alternativa $H_1 [\mu \neq 3,5]$. A tal fin se toma una muestra aleatoria simple de tamaño $n = 25$, donde

$$\sum_{i=1}^{25} x_i = 100 \quad \text{y} \quad \sum_{i=1}^{25} x_i^2 = 600.$$

Constrúyase el contraste de significación con $\alpha = 5\%$.

Como se desconoce σ se utiliza como medida de discrepancia o estadístico del contraste

$$D = \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sqrt{\frac{ns^2}{n-1}}}$$

que, en este caso, sigue una distribución t de Student con 24 grados de libertad.

El valor crítico se determina teniendo en cuenta que

$$P(|D| \geq d_{0,05}/H_0) = P(|t(24)| \geq d_{0,05}/H_0) = 0,05$$

donde, en las tablas correspondientes, se obtiene

$$d_{0,05} = 2,0639.$$

En este caso, la región crítica se toma $|D| \geq d_{0,05}$ pues discrepancias cuantitativamente grandes, tanto positivas como negativas, inducen a pensar que la hipótesis nula sea falsa. Por consiguiente, la región crítica es bilateral, o de dos colas, representándose, gráficamente, como

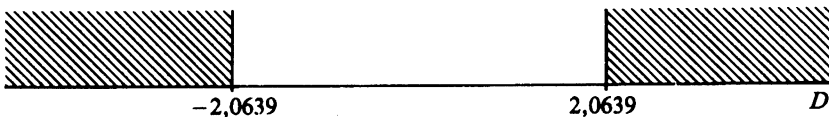


FIGURA 7.2

Como

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{100}{25} = 4$$

$$s^2 = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2 = \frac{600}{25} - 4^2 = 8$$

tenemos que

$$D = \frac{\sqrt{n} (\bar{x} - \mu_0)}{\sqrt{\frac{ns^2}{n-1}}} = \frac{\sqrt{25} (4 - 3,5)}{\sqrt{\frac{25 \cdot 8}{24}}} = 0,86$$

y dado que $|D| = 0,86 < d_{0,05} = 2,0639$, no se rechaza la hipótesis nula H_0 , pues la evidencia muestral no parece contradecir que la media de la población sea $\mu = 3,5$, al nivel de significación del 5%.

7.2 Contraste razón de verosimilitud

Una clase de contrastes de significación de especial interés son los llamados contrastes de la razón de verosimilitud, donde la medida de la discrepancia no está definida en términos de la diferencia entre el valor asignado al parámetro en la hipótesis nula H_0 y el que se obtiene a partir de un estimador del mismo, sino que tal discrepancia se basa en la comparación por cociente³ de la función de verosimilitud evaluada en la hipótesis nula $H_0 [\theta = \theta_0]$ y la función de verosimilitud en un punto máximo.

El estadístico del contraste, llamado **razón de verosimilitud**, es

$$\lambda(\mathbf{X}) = \frac{L(\mathbf{X}; \theta_0)}{\max_{\theta} L(\mathbf{X}; \theta)}$$

³ Una transformación logarítmica de tal cociente lo convierte en una medida de discrepancia por diferencia.

donde $\max_{\Theta} L(\mathbf{X}; \theta)$ representa el valor máximo que puede alcanzar $L(\mathbf{X}; \theta)$ dentro del campo de variación del parámetro θ . La hipótesis nula es simple $H_0[\theta = \theta_0]$, siendo la hipótesis alternativa compuesta $H_1[\theta \in \Theta_1]$ tal que $\Theta = \{\theta_0 \cup \Theta_1\}$.

Si la hipótesis nula también es compuesta, $H_0[\theta \in \Theta_0]$, la razón de verosimilitud debe definirse como⁵

$$\lambda(\mathbf{X}) = \frac{\max_{\Theta_0} L(\mathbf{X}; \theta)}{\max_{\Theta} L(\mathbf{X}; \theta)}$$

Ciñéndonos al caso más sencillo, $H_0[\theta = \theta_0]$, es fácil comprobar que este estadístico cumple las siguientes propiedades:

- $\lambda(\mathbf{X}) \geq 0$, dado que $L(\mathbf{X}; \theta_0) \geq 0$ y $\max_{\Theta} L(\mathbf{X}; \theta) > 0$.
- $\lambda(\mathbf{X}) \leq 1$, pues $L(\mathbf{X}; \theta_0) \leq \max_{\Theta} L(\mathbf{X}; \theta)$.

Por tanto, la razón de verosimilitud está acotada

$$0 \leq \lambda(\mathbf{X}) \leq 1.$$

- Si existe estimador máximo-verosímil del parámetro θ , entonces

$$\max_{\Theta} L(\mathbf{X}; \theta) = L(\mathbf{X}; \hat{\theta})$$

siendo $\hat{\theta}$ dicho estimador y

$$\lambda(\mathbf{X}) = \frac{L(\mathbf{X}; \theta_0)}{L(\mathbf{X}; \hat{\theta})}$$

Puesto que $\lambda(\mathbf{X})$ mide la discrepancia relativa entre $L(\mathbf{X}; \theta_0)$ y $L(\mathbf{X}; \hat{\theta})$, siendo α el nivel de significación, se toma como región crítica $\lambda(\mathbf{X}) \leq K$ tal que

$$P(\lambda(\mathbf{X}) \leq K | H_0) = \alpha.$$

⁵ Existe expresiones alternativas a $\lambda(\mathbf{X})$, entre ellas,

$$\lambda^*(\mathbf{X}) = \frac{\max_{\Theta_0} L(\mathbf{X}; \theta)}{\max_{\Theta_1} L(\mathbf{X}; \theta)}$$

La justificación intuitiva de la elección de esta región crítica es la siguiente:

Como $0 \leq \lambda(\mathbf{X}) \leq 1$, la región crítica es el intervalo $0 \leq \lambda(\mathbf{X}) \leq K$, que supone valores del estadístico del contraste $\lambda(\mathbf{X})$ pequeños (o próximos a cero) puesto que, si la hipótesis nula es falsa, $L(\mathbf{X}; \theta_0)$ no será muy parecido al valor más verosímil de $L(\mathbf{X}; \hat{\theta})$, es decir, mostrará una discrepancia importante respecto a $L(\mathbf{X}; \hat{\theta})$, con lo que el cociente $\lambda(\mathbf{X})$ tomará un valor pequeño. Por tanto, valores pequeños de $\lambda(\mathbf{X})$ indicarán que la hipótesis nula H_0 no es cierta.

Por el contrario, si H_0 es cierta entonces $L(\mathbf{X}; \theta_0)$ estará bastante cercano al valor más verosímil $L(\mathbf{X}; \hat{\theta})$ y el cociente $\lambda(\mathbf{X})$ se situará próximo a la unidad, con lo que valores de $\lambda(\mathbf{X})$ altos no mostrarán una clara evidencia para rechazar la hipótesis nula H_0 .

No obstante, la mayor o menor discrepancia entre $L(\mathbf{X}; \theta_0)$ y $L(\mathbf{X}; \hat{\theta})$ también dependerá de la curvatura que muestre la función de verosimilitud. En la figura siguiente, se refleja este hecho para dos funciones de verosimilitud con distinta curvatura

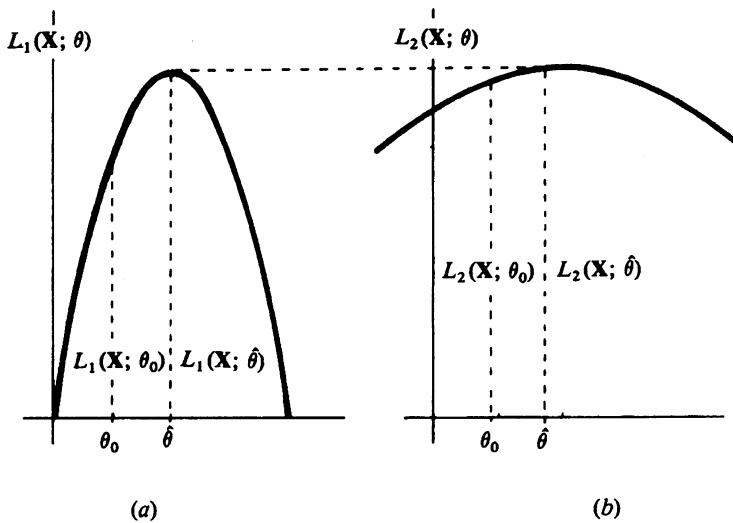


FIGURA 7.3

donde la capacidad de discriminar entre θ_0 y $\hat{\theta}$ a través de sus verosimilitudes (a pesar que la distancia entre esos valores sea la misma) depende de la curvatura

de $L(\mathbf{X}; \theta)$, pues se comprueba, en el caso recogido en el gráfico, que

$$\frac{L_1(\mathbf{X}; \theta_0)}{L_1(\mathbf{X}; \hat{\theta})} \leq \frac{L_2(\mathbf{X}; \theta_0)}{L_2(\mathbf{X}; \hat{\theta})},$$

por lo que la capacidad de discriminación de $\lambda(\mathbf{X})$ es menor cuanto menor sea la curvatura de $L(\mathbf{X}; \theta)$, siendo más improbable que el test razón de verosimilitud rechace la hipótesis nula en el caso (b) que en el (a).

A pesar de esta limitación, el test razón de verosimilitud es preferible a otros contrastes de significación, tales como los vistos en el epígrafe anterior, pues la medida de discrepancia en que se basa, $\lambda(\mathbf{X})$, es menos arbitraria que aquellas al fundamentarse en la información proporcionada por la función de verosimilitud de la muestra.

A continuación presentamos algunos teoremas referentes al contraste razón de verosimilitud.

TEOREMA. Si para un contraste de hipótesis simple existe un contraste razón de verosimilitud, es equivalente al contraste óptimo de Neyman-Pearson.

En efecto, si ambas hipótesis son simples, $H_0[\theta = \theta_0]$ frente a $H_1[\theta = \theta_1]$, el espacio paramétrico es $\Theta = \{\theta_0, \theta_1\}$, pudiendo ocurrir que

■ $\max_{\Theta} L(\mathbf{X}; \theta) = L(\mathbf{X}; \theta_0)$ y

$$\lambda(\mathbf{X}) = \frac{L(\mathbf{X}; \theta_0)}{\max_{\Theta} L(\mathbf{X}; \theta)} = \frac{L(\mathbf{X}; \theta_0)}{L(\mathbf{X}; \theta_1)} = 1$$

y no hay contraste razón de verosimilitud.

■ $\max_{\Theta} L(\mathbf{X}; \theta) = L(\mathbf{X}; \theta_1)$ y

$$\lambda(\mathbf{X}) = \frac{L(\mathbf{X}; \theta_0)}{\max_{\Theta} L(\mathbf{X}; \theta)} = \frac{L(\mathbf{X}; \theta_0)}{L(\mathbf{X}; \theta_1)},$$

siendo, en este caso, la región crítica

$$\lambda(\mathbf{X}) = \frac{L(\mathbf{X}; \theta_0)}{L(\mathbf{X}; \theta_1)} \leq K$$

que coincide con la mejor región crítica que proporciona Neyman-Pearson.

TEOREMA. Si existe un estadístico suficiente para el parámetro θ , el contraste razón de verosimilitud es función del estadístico suficiente.

Por el criterio de factorización de Fisher-Neyman para estadísticos suficientes, si $T = T(\mathbf{X})$ lo es

$$L(\mathbf{X}; \theta) = g(T; \theta) \cdot H(\mathbf{X})$$

por tanto,

$$\lambda(\mathbf{X}) = \frac{L(\mathbf{X}; \theta_0)}{L(\mathbf{X}; \hat{\theta})} = \frac{g(T; \theta_0) \cdot H(\mathbf{X})}{g(T; \hat{\theta}) \cdot H(\mathbf{X})} = \frac{g(T; \theta_0)}{g(T; \hat{\theta})} = \lambda(T),$$

con lo que $\lambda(\mathbf{X})$ es función del estadístico suficiente $T = T(\mathbf{X})$ y $\lambda(\mathbf{X}) = \lambda(T)$ pues θ_0 y $\hat{\theta}$ no son variables.

TEOREMA. Si el contraste de hipótesis se refiere a un solo parámetro θ , se verifica, bajo condiciones de regularidad, la propiedad asintótica

$$-2 \ln \lambda(\mathbf{X}) \xrightarrow{d} \chi^2(1),$$

es decir, la transformación $-2 \ln \lambda(\mathbf{X})$ converge en distribución a una χ^2 de Pearson con 1 grado de libertad.

Si el contraste de hipótesis se construye sobre el vector paramétrico θ donde existen r parámetros independientes incluidos en H_0 , entonces la convergencia se obtiene hacia una $\chi^2(r)$.

EJEMPLO 3

En una población $N(\mu; \sigma)$ con σ conocida se desea contrastar la hipótesis $H_0 [\mu = \mu_0]$ frente a la alternativa $H_1 [\mu \neq \mu_0]$, a partir de la evidencia empírica que proporciona una muestra aleatoria simple de tamaño n . Constrúyase el test razón de verosimilitud.

El estimador de máxima verosimilitud del parámetro μ es \bar{x} , $\hat{\mu} = \bar{x}$, y la razón de verosimilitud

$$\begin{aligned} \lambda &= \frac{L(\mathbf{X}; \mu_0)}{\underset{\mu}{\text{máx}} L(\mathbf{X}; \mu)} = \frac{L(\mathbf{X}; \mu_0)}{L(\mathbf{X}; \hat{\mu})} = \frac{L(\mathbf{X}; \mu_0)}{L(\mathbf{X}; \bar{x})} = \\ &= \frac{(\sigma^2)^{-\frac{n}{2}} (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_0)^2}}{(\sigma^2)^{-\frac{n}{2}} (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{x})^2}} = e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n [\mu_0^2 - \bar{x}^2 - 2(\mu_0 - \bar{x}) x_i]} = \\ &= e^{-\frac{1}{2\sigma^2} \left[n\mu_0^2 - n\bar{x}^2 - 2(\mu_0 - \bar{x}) \sum_{i=1}^n x_i \right]} = e^{-\frac{1}{2\sigma^2} [n\mu_0^2 - n\bar{x}^2 - 2(\mu_0 - \bar{x}) n\bar{x}]} = \\ &= e^{-\frac{1}{2\sigma^2} (n\mu_0^2 - n\bar{x}^2 - 2n\mu_0\bar{x} + 2n\bar{x}^2)} = e^{-\frac{n}{2\sigma^2} (\mu_0^2 - 2\mu_0\bar{x} + \bar{x}^2)} = \\ &= e^{-\frac{n}{2\sigma^2} (\bar{x} - \mu_0)^2} \end{aligned}$$

Como la región crítica viene dada por $\lambda(\mathbf{X}) \leq K$, entonces,

$$\begin{aligned}
 e^{-\frac{n}{2\sigma^2}(\bar{x} - \mu_0)^2} &\leq K_3 \\
 -\frac{n}{2\sigma^2}(\bar{x} - \mu_0)^2 &\leq K_2 \\
 (\bar{x} - \mu_0)^2 &\geq K_1 \\
 |\bar{x} - \mu_0| &\geq K
 \end{aligned}$$

y la región crítica, bilateral (de dos colas), es

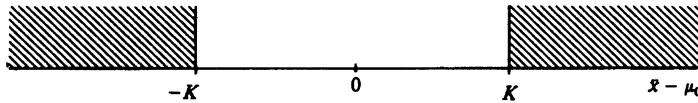


FIGURA 7.3

siendo el estadístico del contraste $\bar{x} - \mu_0$ cuya distribución en el muestreo es

$$\bar{x} - \mu_0 / H_0 \rightarrow N\left(0; \frac{\sigma}{\sqrt{n}}\right)$$

pues

$$\bar{x} / H_0 \rightarrow N\left(\mu_0; \frac{\sigma}{\sqrt{n}}\right).$$

EJEMPLO 4

En una población con distribución de probabilidad exponencial

$$f(x; \theta) = \frac{1}{\theta} e^{-\frac{1}{\theta}x}; \quad x \geq 0 \quad \text{y} \quad \theta > 0$$

se desea contrastar la hipótesis nula $H_0[\theta = \theta_0]$ frente a la alternativa $H_1[\theta > \theta_0]$, considerando una muestra aleatoria simple de tamaño n .

En esta distribución, el parámetro θ es el valor medio o esperanza y su estimación máximo-verosímil

$$\hat{\theta} = \bar{x}.$$

Como la función de verosimilitud es

$$L(\mathbf{X}; \theta) = \theta^{-n} e^{-\frac{1}{\theta} \sum_{i=1}^n x_i} = \theta^{-n} e^{-\frac{n\bar{x}}{\theta}},$$

el estadístico del contraste razón de verosimilitud tiene por expresión

$$\lambda(\mathbf{X}) = \frac{L(\mathbf{X}; \theta_0)}{L(\mathbf{X}; \hat{\theta})} = \frac{\theta_0^{-n} e^{-\frac{n\bar{x}}{\theta_0}}}{\bar{x}^{-n} e^{-\frac{n\bar{x}}{\bar{x}}}} = \left(\frac{\bar{x}}{\theta_0}\right)^n e^{n\left(1 - \frac{\bar{x}}{\theta_0}\right)}$$

Como $\lambda(\mathbf{X})$ es función decreciente respecto a \bar{x} , cuando $\bar{x} > \theta_0$, la región crítica del contraste

$$\lambda(\mathbf{X}) \leq K_3$$

será equivalente a

$$\bar{x} \geq K_2$$

o bien

$$n\bar{x} \geq K_1$$

El estadístico

$$n\bar{x} = \sum_{i=1}^n x_i = x_1 + x_2 + \cdots + x_n,$$

al ser la muestra aleatoria simple, donde los elementos x_i siguen una distribución exponencial y son independientes, tiene⁶ una distribución gamma $G\left(n; \frac{1}{\theta}\right)$.

También se demuestra que

$$\frac{2n\bar{x}}{\theta} \rightarrow \chi^2(2n).$$

Por tanto, la región crítica puede tomarse, alternativamente, como

$$\frac{2n\bar{x}}{\theta_0} \geq K$$

determinando los valores críticos K_1 o K , según sea el estadístico del contraste que se utilice, haciendo

$$P(n\bar{x} \geq K_1 / H_0) = \alpha$$

⁶ Véase MARTÍN PLIEGO y RUIZ-MAYA, *Fundamentos...* Capítulo 7.

o bien

$$P\left[\frac{2n\bar{x}}{\theta_0} \geq K/H_0\right] = P[\chi^2(2n) \geq K/H_0] = \alpha,$$

donde α es el nivel de significación.

Las constantes K y K_1 serán distintas aunque la decisión sobre la hipótesis a aceptar será la misma.

7.3 Contrastes asintóticos basados en la función de verosimilitud

En este epígrafe se estudian dos contrastes alternativos al de razón de verosimilitud que, en ocasiones, son de más cómoda instrumentación y con el que guardan ciertas analogías.

7.3.1. TEST DE WALD

Se desea efectuar un contraste de hipótesis compuesta acerca de un parámetro θ , para el que existe un estimador de máxima verosimilitud $\hat{\theta}$.

Por las propiedades asintóticas de los estimadores máximo verosímiles puede establecerse que

$$\hat{\theta} \xrightarrow{d} N\left(\theta; \sqrt{V(\hat{\theta})}\right)$$

siendo la varianza asintótica de este estimador $\hat{\theta}$ el inverso de la cantidad de información de Fisher $I(\theta)$

$$V(\hat{\theta}) = [I(\theta)]^{-1}$$

por otra parte, dado que $\hat{\theta}$ como estimador máximo-verosímil es consistente, $I(\hat{\theta})$ función de dicho estimador también será un estimador consistente de $I(\theta)$.

Pudiéndose aceptar, por otra parte, que

$$\frac{\hat{\theta} - \theta}{\sqrt{[I(\hat{\theta})]^{-1}}} \xrightarrow{d} N(0; 1).$$

El estadístico del contraste del test de Wald se define como

$$W = (\hat{\theta} - \theta_0)^2 I(\hat{\theta})$$

que sigue una distribución asintótica χ^2 de Pearson con 1 grado de libertad, al ser el cuadrado de la variable normal anterior, en este caso, bajo el supuesto que la hipótesis nula sea cierta y, por tanto, $\theta = \theta_0$.

De manera intuitiva, el test de Wald se basa en la discrepancia entre la hipótesis nula θ_0 y el valor más verosímil que se le puede asignar a dicho parámetro dada la información suministrada por una muestra \mathbf{X} ; este último valor no es otro que el que se determina a través del estimador $\hat{\theta}$. La distancia, tomada en términos cuadráticos para obviar su signo, está ponderada por la cantidad $I(\hat{\theta})$, estimación consistente, de la cantidad de información de Fisher $I(\theta)$.

Hemos señalado en el epígrafe anterior que el contraste razón de verosimilitud venía afectado tanto por la distancia entre θ_0 y $\hat{\theta}$ como por la curvatura que tuviera la función de verosimilitud. En este sentido, recordemos que la cantidad de información de Fisher⁷ $I(\theta)$ es una medida de la curvatura de la función $\ln L(\mathbf{X}; \theta)$ en cada punto, en términos de la varianza de estos valores, pues⁸

$$I(\theta) = V[SC(\theta)] = V\left[\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta}\right]$$

por tanto, si la función $\ln L(\mathbf{X}; \theta)$ presenta poca curvatura, sus diferentes derivadas en cada uno de sus puntos no serán muy dispares y su varianza será reducida, tomando un valor relativamente pequeño $I(\theta)$. En el caso de que esta función tuviera una mayor curvatura, llegaríamos a que $I(\theta)$ tomaría un valor mayor, existiendo puntos de $\ln L(\mathbf{X}; \theta)$ con derivadas de valores altos.

El test de Wald recoge la discrepancia entre θ_0 y $\hat{\theta}$ corrigiéndola del efecto de curvatura de la función de verosimilitud $L(\mathbf{X}; \theta)$ mediante el coeficiente corrector $I(\hat{\theta})$.

Para la especificación completa del test de Wald señalemos que su región crítica es del tipo

$$W \geq K$$

⁷ Véase el capítulo 2.

⁸ *Ibidem*.

pues valores pequeños de W , próximos a cero, evidenciarán que tanto la distancia entre θ_0 y $\hat{\theta}$ como la curvatura de $L(\mathbf{X}; \theta)$, medida a través de la derivada de $\ln L(\mathbf{X}; \theta)$, son cercanas a cero, y puede mantenerse θ_0 como cierto.

Fijado el nivel de significación α , el valor crítico K se determina teniendo en cuenta que

$$P[\chi^2(1) \geq K/H_0] = \alpha .$$

Si se efectúan hipótesis sobre un vector paramétrico θ , el estadístico del contraste del test de Wald viene dado por la expresión

$$W = [\hat{\theta} - \theta_0]' [I(\hat{\theta})] (\hat{\theta} - \theta_0)$$

que sigue una distribución asintótica χ^2 de Pearson con r grados de libertad, siendo r el número de parámetros independientes que figuren en el vector paramétrico θ (total de parámetros menos restricciones).

EJEMPLO 5

En una población $N(\mu; \sigma)$ con σ conocida, se desea contrastar $H_0 [\mu = \mu_0]$ frente a la hipótesis alternativa $H_1 [\mu \neq \mu_0]$. Determinése el correspondiente test de Wald.

Sabemos que la cantidad de información de Fisher sobre el parámetro μ , correspondiente a la muestra \mathbf{X} , es tal que

$$I(\theta) = n i(\theta) ,$$

donde $i(\theta)$ es la cantidad de información de Fisher de la variable aleatoria poblacional, que se determina como

$$\begin{aligned} f(x; \mu, \sigma^2) &= (\sigma^2)^{-\frac{1}{2}} (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}} \\ \ln f(x; \mu, \sigma^2) &= -\frac{1}{2} \ln \sigma^2 - \frac{1}{2} \ln(2\pi) - \frac{1}{2} \frac{(x-\mu)^2}{\sigma^2} \\ \frac{\partial \ln f(x; \mu, \sigma^2)}{\partial \mu} &= \frac{x-\mu}{\sigma^2} \\ i(\mu) &= E \left[\frac{\partial \ln f(x; \mu, \sigma^2)}{\partial \mu} \right]^2 = E \left[\frac{x-\mu}{\sigma^2} \right]^2 = \frac{1}{\sigma^4} E[x-\mu]^2 = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2} \end{aligned}$$

por tanto,

$$I(\mu) = n \frac{1}{\sigma^2} .$$

El estimador máximo-verosímil de la media μ en una población $N(\mu; \sigma)$ es igual a

$$\hat{\mu} = \bar{x},$$

con lo que el estadístico del contraste del test de Wald resulta

$$W = (\hat{\mu} - \mu_0)^2 I(\mu) = (\bar{x} - \mu_0)^2 \frac{n}{\sigma^2},$$

siendo la región crítica

$$(\bar{x} - \mu_0)^2 \frac{n}{\sigma^2} \geq K_2,$$

es decir,

$$\begin{aligned} (\bar{x} - \mu_0)^2 &\geq K_1 \\ |\bar{x} - \mu_0| &\geq K \end{aligned}$$

región crítica coincidente con la determinada a través de un contraste razón de verosimilitud, según vimos en el epígrafe anterior, siendo la distribución asintótica, tanto del test razón de verosimilitud como del test de Wald, una χ^2 de Pearson con 1 grado de libertad.

7.3.2. TEST DE LOS MULTIPLICADORES DE LAGRANGE

El test de los multiplicadores de Lagrange parte del multiplicador que se obtiene en el proceso de optimización de verosimilitud restringida⁹ al supuesto que la hipótesis nula es cierta.

Se pretende, pues, hallar el valor del parámetro θ que haga máxima la función $\ln L(\mathbf{X}; \theta)$ sujeto a la condición que la hipótesis nula $H_0[\theta = \theta_0]$ es cierta. Para ello se forma el lagrangiano

$$M = \ln L(\mathbf{X}; \theta) - \gamma(\theta - \theta_0)$$

donde el vector gradiente, igualado a cero, viene dado por

$$\left. \begin{aligned} \frac{\partial M}{\partial \theta} &= \frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta} - \gamma = 0 \\ \frac{\partial M}{\partial \gamma} &= \theta - \theta_0 = 0 \end{aligned} \right\}$$

⁹ Condicionada.

y el multiplicador γ que satisface ambas ecuaciones es

$$\left[\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta} \right]_{\theta=\theta_0} = SC(\theta_0)$$

es decir, la variable *score* particularizada en la hipótesis nula.

Si la muestra \mathbf{X} es aleatoria simple, el *score* es

$$SC(\theta) = \frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta} = \sum_{i=1}^n \frac{\partial \ln f(x_i; \theta)}{\partial \theta}$$

siendo x_i cada uno de los elementos muestrales.

Sabemos que ¹⁰

$$E[SC(\theta)] = 0$$

$$V[SC(\theta)] = I(\theta)$$

y por el Teorema Central del Límite el *score*, suma de variables aleatorias independientes igualmente distribuidas, tiene la distribución límite:

$$SC(\theta) \xrightarrow{d} N(0; \sqrt{I(\theta)})$$

o, equivalentemente

$$\frac{SC(\theta)}{\sqrt{I(\theta)}} \xrightarrow{d} N(0; 1).$$

Tomando esta distribución como referencia, el estadístico del contraste del test de los multiplicadores de Lagrange (M.L.) es

$$ML = [SC(\theta_0)]^2 [I(\theta_0)]^{-1}$$

siguiendo una distribución asintótica χ^2 de Pearson con un grado de libertad.

Obsérvese que el *score* es la derivada de la función $\ln L(\mathbf{X}; \theta)$ que, geométricamente, puede medirse a través de la pendiente de la tangente a dicha función en

¹⁰ Véase el Capítulo 2

el punto que corresponda, es decir,

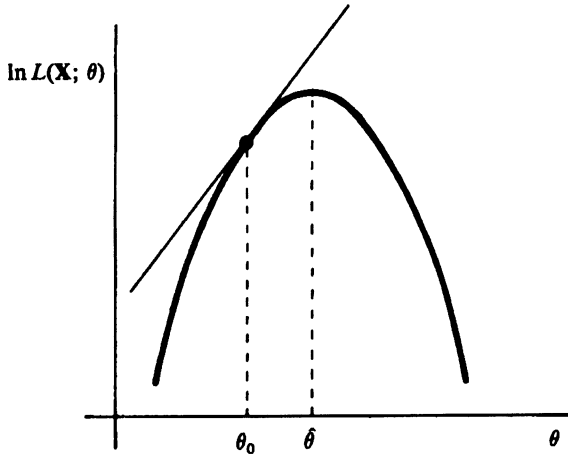


FIGURA 7.5

dicha pendiente para $\hat{\theta}$ será cero puesto que al obtenerse en ese punto el máximo no restringido de la función, la correspondiente tangente será paralela al eje de abscisas y su pendiente, por tanto, es nula.

Por todo ello, θ_0 será cierto en la medida que se acerque lo más posible al estimador de máxima verosimilitud $\hat{\theta}$, valor más verosímil que puede determinarse a partir de la evidencia empírica suministrada por la muestra. Si esto es así, entonces $SC(\theta_0)$, para que no sea rechazada la H_0 , deberá encontrarse también próximo a cero. Esta consideración lleva a que la región crítica del test de los multiplicadores de Lagrange será

$$ML \geq K$$

pues valores próximos a cero de ML invitarán a no rechazar la hipótesis nula por lo antes dicho.

Fijado el nivel de significación α , la constante K quedará determinada como solución de

$$P[\chi^2(1) \geq K/H_0] = \alpha.$$

El test de los multiplicadores de Lagrange también recoge el efecto curvatura de la función $\ln L(\mathbf{X}; \theta)$ en θ_0 al incluir el término cantidad de información de Fisher $I(\theta)$ particularizado en dicho valor θ_0 .

Si se efectúa el contraste de hipótesis sobre el vector paramétrico θ , la expresión del estadístico del contraste del test de los multiplicadores de Lagrange resulta

$$ML = [SC(\theta_0)]' [I(\theta_0)]^{-1} [SC(\theta_0)]$$

que sigue, asintóticamente, una distribución χ^2 de Pearson con r grados de libertad, siendo r el número de parámetros independientes, al igual que en los tests razón de verosimilitud y de Wald.

EJEMPLO 6

Contrúyase el estadístico del contraste del test de los multiplicadores de Lagrange para el mismo caso que el propuesto en el test de Wald.

Como en una población con distribución $N(\mu; \sigma)$

$$\ln L(\mathbf{X}; \mu, \sigma^2) = -\frac{n}{2} \ln \sigma^2 - \frac{n}{2} \ln(2\pi) - \frac{1}{2} \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma^2}$$

entonces

$$SC(\mu) = \frac{\partial \ln L(\mathbf{X}; \mu, \sigma^2)}{\partial \mu} = \frac{\sum_{i=1}^n (x_i - \mu)}{\sigma^2} = \frac{n\bar{x} - n\mu}{\sigma^2} = n \frac{(\bar{x} - \mu)}{\sigma^2},$$

tomando $I(\mu)$ el valor, según se calculó en el ejemplo anterior,

$$I(\mu) = \frac{n}{\sigma^2}$$

y

$$\begin{aligned} ML &= [SC(\mu_0)]^2 [I(\mu_0)]^{-1} = \\ &= \frac{n^2 (\bar{x} - \mu_0)^2}{\sigma^4} \frac{\sigma^2}{n} = \\ &= \frac{n(\bar{x} - \mu_0)^2}{\sigma^2} \end{aligned}$$

y la región crítica del test de los multiplicadores de Lagrange viene determinada a partir de

$$\begin{aligned} \frac{n(\bar{x} - \mu_0)^2}{\sigma^2} &\geq K_2 \\ (\bar{x} - \mu_0)^2 &\geq K_1 \\ |\bar{x} - \mu_0| &\geq K \end{aligned}$$

región crítica que coincide con las correspondientes a este contraste que proporcionan los tests razón de verosimilitud y de Wald.

7.3.3. COMPARACIÓN DE LOS TESTS RAZÓN DE VEROSIMILITUD, WALD Y MULTIPLICADORES DE LAGRANGE

- El test asintótico de razón de verosimilitud (*RV*) definido por

$$-2 \ln \lambda(\mathbf{X}) = -2[\ln L(\mathbf{X}; \theta_0) - \ln L(\mathbf{X}; \hat{\theta})]$$

se basa en la distancia entre verosimilitudes, necesitando la determinación previa del estimador de máxima verosimilitud $\hat{\theta}$.

- El test de Wald presenta una medida de discrepancia definida en términos de la diferencia $\hat{\theta} - \theta_0$, siendo necesario, para obtener la cantidad de información de Fisher $I(\hat{\theta})$, determinar también el estimador $\hat{\theta}$, además de las dos primeras derivadas de la función $\ln L(\mathbf{X}; \theta)$ ¹¹.
- El test de los multiplicadores de Lagrange (*ML*) no necesita la obtención del estimador $\hat{\theta}$, aunque sí se deben calcular las dos primeras derivadas de la función $\ln L(\mathbf{X}; \theta)$ para determinar $I(\theta)$.
- Por otra parte, si el test razón de verosimilitud se define bajo su forma más general, en el caso de una hipótesis nula compuesta

$$RV = \lambda(\mathbf{X}) = \frac{\max_{\Theta_0} L(\mathbf{X}; \theta)}{\max_{\Theta} L(\mathbf{X}; \theta)}$$

observamos que necesita tanto la estimación de θ restringida al subespacio paramétrico Θ_0 , que define la hipótesis nula, como la estimación no restringida respecto a Θ . El test de Wald únicamente hace uso de la estimación sin restricciones $\hat{\theta}$ que determina el procedimiento de máxima verosimilitud. Y para el test de los multiplicadores de Lagrange, por el contrario, hay que obtener sólo la estimación restringida en Θ_0 , condición que se impone en el proceso de optimización.

La utilización de uno u otro test depende de la mayor o menor dificultad en la determinación del estimador máximo-verosímil y/o en las derivadas necesarias de la función $\ln L(\mathbf{X}; \theta)$.

¹¹ Recuérdese que:

$$I(\theta) = E \left[\frac{\partial \ln L(\mathbf{X}; \theta)}{\partial \theta} \right]^2 = -E \left[\frac{\partial^2 \ln L(\mathbf{X}; \theta)}{\partial \theta^2} \right].$$

■ Puede demostrarse que, asintóticamente, bajo condiciones generales, los tres tipos de contrastes son equivalentes, *aunque su utilización en aproximaciones para tamaños muestrales no muy elevados puede llevar a diferentes decisiones en el contraste.*

■ Los tres tests son consistentes, pues su función de potencia verifica

$$P(\theta) \xrightarrow{n \rightarrow \infty} 1, \quad \text{para } \theta \in \Theta_1.$$

7.4 Relación entre los contrastes de hipótesis y los intervalos de confianza

Siendo C y C^* las regiones crítica y de aceptación de un contraste de hipótesis con nivel de significación α , se verifica que

$$\begin{aligned}\alpha &= P(\mathbf{X} \in C/H_0) \\ 1 - \alpha &= P(\mathbf{X} \in C^*/H_0)\end{aligned}$$

Cuando existe una región de aceptación C^* tal que

$$P(\mathbf{X} \in C^*/\theta) = 1 - \alpha$$

siendo $Q(\mathbf{X})$ un intervalo se puede establecer una correspondencia biunívoca entre los sucesos

$$(\mathbf{X} \in C^*/\theta) \Leftrightarrow [Q(\mathbf{X}) \supset \theta]$$

es decir, que el intervalo $Q(\mathbf{X})$ incluya el parámetro θ , estableciéndose que

$$P(\mathbf{X} \in C^*/\theta) = P[Q(\mathbf{X}) \supset \theta] = 1 - \alpha.$$

Si para cada muestra \mathbf{X} el conjunto $Q(\mathbf{X})$ es cerrado y no vacío, $Q(\mathbf{X})$ puede definirse por el intervalo $[\underline{\theta}(\mathbf{X}); \bar{\theta}(\mathbf{X})]$, de manera que para un contraste de hipótesis bilateral

$$P(\mathbf{X} \in C^*/\theta) = P[\underline{\theta}(\mathbf{X}) \leq \theta \leq \bar{\theta}(\mathbf{X})] = 1 - \alpha$$

siendo $[\underline{\theta}(\mathbf{X}); \bar{\theta}(\mathbf{X})]$ un intervalo para θ al nivel de confianza $1 - \alpha$.

Esta relación permite efectuar un contraste al nivel de significación α sin más que comprobar que el valor θ_0 establecido en la hipótesis nula H_0 pertenece al intervalo de confianza $1 - \alpha$

EJEMPLO 7

De una población $N(\mu; 2)$ se extrae una muestra aleatoria simple de tamaño $n = 40$ sien-

$$\text{do } \sum_{i=1}^{40} x_i = 84.$$

- Determinése el intervalo de confianza del 95% para el parámetro μ .
 - Contrástese la hipótesis $H_0 [\mu = 3]$ frente $H_1 [\mu \neq 3]$ al nivel de significación $\alpha = 5\%$.
 - Compárese los resultados anteriores.
- (a) Según vimos en el capítulo 5, el intervalo de confianza para el parámetro μ de una población $N(\mu; \sigma)$ con σ conocida es

$$\left[\bar{x} - K \frac{\sigma}{\sqrt{n}}; \bar{x} + K \frac{\sigma}{\sqrt{n}} \right]$$

donde

$$P(-K \leq N(0; 1) \leq K) = 1 - \alpha = 0,95$$

siendo el valor tabular $K = 1,96$, con lo que el intervalo de confianza para μ es

$$\left[2,1 - 1,96 \frac{2}{\sqrt{40}}; 2,1 + 1,96 \frac{2}{\sqrt{40}} \right] = [1,48; 2,72]$$

pues

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{84}{40} = 2,1.$$

- (b) Para verificar la hipótesis nula $H_0 [\mu = 3]$ frente a $H_1 [\mu \neq 3]$ se puede usar un contraste razón de verosimilitud, tal como el que se obtuvo en el epígrafe 7.2, cuya región crítica venía dada por

$$|\bar{x} - \mu_0| \geq K$$

siendo el nivel de significación $\alpha = 0,05$, y se determina K haciendo

$$\begin{aligned} \alpha &= P(|\bar{x} - \mu_0| \geq K/H_0) = \\ &= 1 - P(|\bar{x} - \mu_0| < K/H_0) = \\ &= 1 - P\left(-K \leq \frac{\sigma}{\sqrt{n}}\xi \leq K\right) = \\ &= 1 - P\left(-\frac{K\sqrt{n}}{\sigma} \leq \xi \leq \frac{K\sqrt{n}}{\sigma}\right) = 0,05 \end{aligned}$$

dado que

$$\bar{x} - \mu_0/H_0 \rightarrow N\left(0; \frac{\sigma}{\sqrt{n}}\right)$$

donde $\xi \rightarrow N(0; 1)$.

En las tablas de la distribución $N(0; 1)$ se obtiene

$$\frac{K\sqrt{n}}{\sigma} = 1,96,$$

por tanto,

$$K = 1,96 \frac{\sigma}{\sqrt{n}} = 1,96 \frac{2}{\sqrt{40}} = 0,62.$$

La región crítica para el contraste resulta

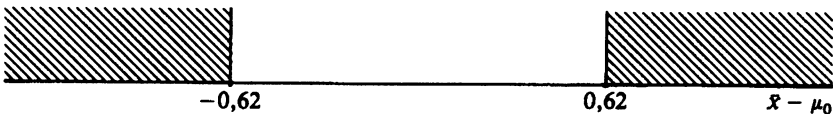


FIGURA 7.6

y como $\bar{x} - \mu_0 = 2,1 - 3 = -0,9$ es un valor de la región crítica se rechaza H_0 , lo que es equivalente a verificar si μ_0 pertenece al intervalo

$$[\bar{x} - 0,62; \bar{x} + 0,62] = [1,48; 2,72].$$

En este caso $\mu_0 = 3$ y se rechaza $H_0[\mu = 3]$ al nivel de significación $\alpha = 0,05$.

- (c) Como ha podido comprobarse, existe una relación biunívoca entre que el parámetro μ pertenezca al intervalo de confianza $[1,48; 2,72]$ y que el estadístico del contraste $\bar{x} - \mu_0$ pertenezca a la región de aceptación $[-0,62; 0,62]$.

CAPÍTULO 8

Contrastes paramétricos

8.1 Contrastación de hipótesis sobre la media de una población $N(\mu; \sigma)$

Para efectuar la contrastación de la hipótesis sobre la media poblacional de una distribución normal distinguimos, en primer lugar, dos casos: población con varianza conocida y con varianza desconocida. La hipótesis nula será simple, $H_0[\mu = \mu_0]$, mientras que la alternativa será simple $H_1[\mu = \mu_1]$ o compuesta, $H_1[\mu > \mu_0]$, $H_1[\mu < \mu_0]$ y $H_1[\mu \neq \mu_0]$. Los contrastes se efectúan tomando muestras aleatorias simples $\mathbf{X}(x_1, \dots, x_n)$ de tamaño n , de una población $N(\mu; \sigma)$.

8.1.1. VARIANZA POBLACIONAL CONOCIDA

8.1.1.1. Hipótesis simples: $H_0[\mu = \mu_0]$, $H_1[\mu = \mu_1]$

Al ser las dos hipótesis simples hallaremos la mejor región crítica mediante el lema de Neyman-Pearson¹. Partimos de una población con distribución de probabilidad

¹ Véase el capítulo 6.

$N(\mu; \sigma)$, con varianza σ^2 conocida. La función de verosimilitud es

$$L(x_1, \dots, x_n; \mu) = L(\mathbf{X}; \mu) = \frac{1}{\sqrt{\sigma^2 2\pi}} e^{-\frac{(x_1 - \mu)^2}{2\sigma^2}} \dots \frac{1}{\sqrt{\sigma^2 2\pi}} e^{-\frac{(x_n - \mu)^2}{2\sigma^2}} =$$

$$= \frac{1}{(\sigma^2)^{\frac{n}{2}} (2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2} \quad [1]$$

El cociente de funciones de verosimilitud cuando sustituimos el parámetro μ por sus valores establecidos en las hipótesis resulta

$$\frac{L(\mathbf{X}; \mu_0)}{L(\mathbf{X}; \mu_1)} = \frac{\frac{1}{(\sigma^2)^{\frac{n}{2}} (2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_0)^2}}{\frac{1}{(\sigma^2)^{\frac{n}{2}} (2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_1)^2}} = e^{-\frac{1}{2\sigma^2} \left[\sum_{i=1}^n (x_i - \mu_0)^2 - \sum_{i=1}^n (x_i - \mu_1)^2 \right]} =$$

$$= e^{-\frac{1}{2\sigma^2} \left[n(\mu_0^2 - \mu_1^2) + 2(\mu_1 - \mu_0) \sum_{i=1}^n x_i \right]} = e^{-\frac{n}{2\sigma^2} [(\mu_0^2 - \mu_1^2) + 2(\mu_1 - \mu_0) \bar{x}]} \leq K_3$$

Tomando logaritmos,

$$\ln \frac{L(\mathbf{X}; \mu_0)}{L(\mathbf{X}; \mu_1)} = -\frac{n}{2\sigma^2} [(\mu_0^2 - \mu_1^2) + 2(\mu_1 - \mu_0) \bar{x}] \leq \ln K_3$$

$$(\mu_0^2 - \mu_1^2) + 2(\mu_1 - \mu_0) \bar{x} \geq \frac{-2\sigma^2 \ln K_3}{n} = K_2 \quad [2]$$

$$(\mu_1 - \mu_0) \bar{x} \geq \frac{K_2 - (\mu_0^2 - \mu_1^2)}{2} = K_1$$

Para despejar \bar{x} en [2] y obtener la región crítica es necesario conocer el signo de la diferencia $\mu_1 - \mu_0$, pues según sea positivo ($\mu_0 < \mu_1$) o negativo ($\mu_0 > \mu_1$) cambia el sentido de la desigualdad y, por consiguiente, la forma del contraste. Estudiaremos en detalle el caso de $\mu_0 < \mu_1$.

A. $\mu_0 < \mu_1$

Si el valor de la media establecido en la hipótesis nula es menor que el de la alternativa, $\mu_0 < \mu_1$, la diferencia $\mu_1 - \mu_0$ es positiva y la mejor región crítica expresada de acuerdo con [2] queda de la forma

$$\bar{x} \geq \frac{K_1}{\mu_1 - \mu_0} = K$$

es decir,

$$\bar{x} \geq K .$$

Ante este resultado la forma de actuar es la siguiente: obtenida la muestra, se calcula su media, \bar{x} , y si su valor es mayor que el de la constante K se rechaza la hipótesis nula, por no proporcionar la muestra suficiente evidencia para sostener la hipótesis nula aceptándose, por consiguiente, la alternativa.

En la figura 8.1 representamos las dos hipótesis sobre la media poblacional, cuando $\mu_0 < \mu_1$, destacando la región crítica. A fin de evitar equívocos recalamos que, si bien en la contrastación de las hipótesis están involucradas distribuciones $N(\mu; \sigma)$, en las regiones críticas aparecen distribuciones $N\left(\mu; \frac{\sigma}{\sqrt{n}}\right)$ correspondientes a la media muestral, estadístico definitorio de la región crítica.

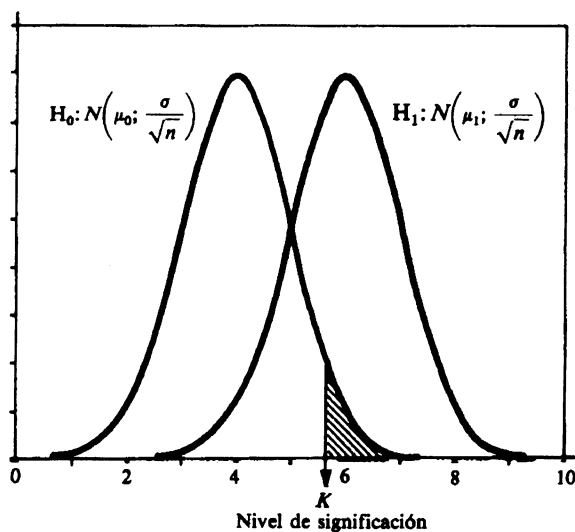


FIGURA 8.1

Funciones de densidad de la media muestral bajo cada una de las dos hipótesis:

$$H_0 [\mu = \mu_0], \quad H_1 [\mu = \mu_1], \quad \mu_0 < \mu_1 .$$

Establecida la forma del contraste, y conocida la distribución en el muestreo del estadístico necesaria para la determinación del valor de la constante K , en su cálculo pueden intervenir dos de los tres elementos siguientes: nivel de significación α , potencia $(1 - \beta)$ o tamaño muestral (n) .

En general, en un contraste de hipótesis se fija el nivel de significación, α , y el tamaño n de la muestra, viniendo impuesta la potencia, $1 - \beta : (\alpha; n) \Rightarrow 1 - \beta$.

Sin embargo, se puede actuar de otras dos maneras: fijando la potencia y el tamaño muestral y determinando el nivel de significación: $(1 - \beta; n) \Rightarrow \alpha$, o fijando el nivel de significación y la potencia hallando el tamaño necesario de la muestra: $(\alpha; 1 - \beta) \Rightarrow n$. En todas las situaciones se calcula el valor de la constante, K , que determina la región crítica. A continuación estudiamos en detalle el primer caso.

■ Nivel de significación, α , y tamaño muestral, n , conocidos

Cálculo de la constante K

Partimos de la definición de nivel de significación

$$\alpha = P(\text{rechazar } H_0 \text{ siendo cierta}) = P(\mathbf{X} \in C/H_0) = P(\bar{x} \geq K/H_0).$$

Si la hipótesis nula es cierta, la población es $N(\mu_0; \sigma)$, la media muestral \bar{x} se distribuye $N\left(\mu_0; \frac{\sigma}{\sqrt{n}}\right)$ pudiéndose expresar \bar{x} como

$$\bar{x} = \frac{\sigma}{\sqrt{n}} \xi + \mu_0$$

siendo $\xi \sim N(0; 1)$, y

$$P(\bar{x} \geq K/H_0) = P\left(\frac{\sigma}{\sqrt{n}} \xi + \mu_0 \geq K\right) = P\left[\xi \geq \frac{\sqrt{n}}{\sigma}(K - \mu_0)\right] = P(\xi \geq z_\alpha) = \alpha$$

Hallando el valor tabular z_α de la distribución $N(0; 1)$ que deja a su derecha un área igual a α , el cálculo de la constante crítica K es inmediato

$$z_\alpha = \frac{\sqrt{n}(K - \mu_0)}{\sigma}; \quad K = \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha$$

y la región crítica resulta

$$\bar{x} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha$$

Potencia del contraste

$$\begin{aligned} 1 - \beta &= P(\text{rechazar } H_1 \text{ siendo cierta}) = P(\mathbf{X} \in C/H_1) = \\ &= P(\bar{x} \geq K/H_1) = P\left(\bar{x} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha/H_1\right). \end{aligned}$$

Si la hipótesis alternativa es cierta la media muestral \bar{x} se distribuye $N\left(\mu_1; \frac{\sigma}{\sqrt{n}}\right)$ conduciendo, de manera análoga a la anterior, a

$$P\left(\bar{x} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha / H_1\right) = P\left(\frac{\sigma}{\sqrt{n}} \xi + \mu_1 \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha\right) =$$

$$= P\left[\xi \geq z_\alpha + \frac{\sqrt{n}}{\sigma} (\mu_0 - \mu_1)\right] = 1 - \beta$$

En la figura 8.2 representamos la potencia del contraste

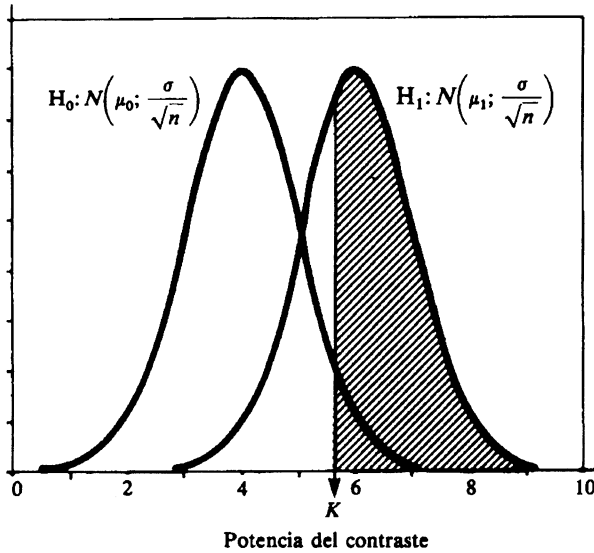


FIGURA 8.2

Funciones de densidad de la media muestral bajo cada una de las dos hipótesis:
 $H_0 [\mu = \mu_0], H_1 [\mu = \mu_1], \mu_0 < \mu_1.$

EJEMPLO 1

En una población $N(\mu; 2)$ se desconoce el valor de la media estableciéndose dos hipótesis sobre él: hipótesis nula $[\mu_0 = 4]$ e hipótesis alternativa $[\mu_1 = 6]$. Calcularemos la mejor región crítica con un nivel de significación $\alpha = 0,05$, mediante una muestra aleatoria simple de tamaño 4, cuyos elementos son

1,8	6,6	2,4	3,5
-----	-----	-----	-----

determinándose si se acepta o rechaza la hipótesis nula.

Como $\mu_0 < \mu_1$, la mejor región crítica es de la forma $\bar{x} \geq K$ y para hallar el valor de la constante K debemos conocer la distribución en el muestreo del estadístico \bar{x} que, por ser la media de una muestra aleatoria simple de tamaño 4 procedente de una población $N(\mu; 2)$, es

$$N\left(\mu; \frac{2}{\sqrt{4}}\right) = N(\mu; 1).$$

El cálculo de K se lleva a cabo a partir de la definición del nivel de significación

$$\alpha = P(\text{rechazar } H_0 \text{ siendo cierta}) = P(\bar{x} \geq K/H_0 [\mu = 4])$$

y si la hipótesis nula es cierta \bar{x} se distribuye en el muestreo $N(4; 1)$ por lo cual,

$$0,05 = P[\bar{x} \geq K/H_0 [\mu = 4]] = P[N(4; 1) \geq K] = P(\xi + 4 \geq K) = P(\xi \geq K - 4).$$

En la distribución $N(0; 1)$ el valor tabular $z_{0,05}$ es 1,645 y el de la constante K , como $z_{0,05} = K - 4 = 1,645$, es $K = 5,645$ y la mejor región crítica es de la forma $\bar{x} \geq 5,645$.

El lema de Neyman-Pearson indica que la mejor región crítica está determinada por el conjunto de valores muestrales que verifican $\bar{x} \geq 5,645$. En la muestra obtenida la media es igual a 3,58 valor que no cumple la condición anterior ($3,58 < 5,645$), es decir, la media no pertenece a la región crítica por lo cual aceptamos la hipótesis nula, $\mu = 4$, y rechazamos que la media poblacional sea igual a 6.

Hallamos la potencia del contraste, $1 - \beta$,

$$\begin{aligned} 1 - \beta &= P(\text{rechazar } H_0 \text{ siendo falsa}) = \\ &= P(\mathbf{X} \in C/H_1 [\mu = 6]) = P(\bar{x} \geq 5,645/H_1 [\mu = 6]) = \\ &= P(N(0; 1) \geq -0,355) = 0,6387. \end{aligned}$$

B. $\mu_0 > \mu_1$

Cuando μ_0 es mayor que μ_1 , la diferencia $\mu_1 - \mu_0$ es negativa y la mejor región crítica es de la forma $\bar{x} \leq K$.

La obtención del valor de la constante K y el de la potencia del contraste se efectúa de manera similar a lo expuesto en el caso anterior, $\mu_0 < \mu_1$, sin más que tener en cuenta que ahora $\mu_0 > \mu_1$.

La mejor región crítica es

$$\bar{x} \leq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha}$$

y la potencia del contraste

$$1 - \beta = P \left[N(0; 1) \leq z_{1-\alpha} + \frac{\sqrt{n}}{\sigma} (\mu_0 - \mu_1) \right].$$

8.1.1.2. Hipótesis compuestas: $H_0 [\mu = \mu_0]$, $H_1 [\mu > \mu_0]$ o $H_1 [\mu < \mu_0]$

En este tipo de contrastación contemplamos hipótesis alternativas compuestas de la forma $\mu > \mu_0$ o $\mu < \mu_0$. Para hallar la mejor región crítica recurrimos al lema de Neyman-Pearson. En la expresión [1], del cociente de verosimilitudes, hacemos $\mu_1 = \mu$, llegándose a [2] como

$$(\mu - \mu_0)\bar{x} \geq K_1$$

y aunque no conozcamos en la hipótesis alternativa el valor concreto de μ sí sabemos que $\mu > \mu_0$, lo que supone que la diferencia $\mu - \mu_0$ es mayor que cero, por lo cual al dividir por $\mu - \mu_0$ no cambia el sentido de la desigualdad, y la mejor región crítica es $\bar{x} \geq K$.

De manera análoga, cuando la hipótesis alternativa es $H_1 [\mu < \mu_0]$, en el primer miembro de la expresión [2] como $\mu < \mu_0$ la diferencia $\mu - \mu_0$ es negativa cambiando el sentido de la desigualdad al despejar \bar{x} , y la mejor región crítica resulta $\bar{x} \leq K$.

EJEMPLO 2

En la distribución $N(\mu; 10)$ hallamos la mejor región crítica para la hipótesis $H_0 [\mu = 2]$ y $H_1 [\mu > 2]$, con un nivel de significación igual a 0,10, en muestras aleatorias simples de tamaño 16. Si los valores muestrales son

-15,21	0,89	-3,25	-5,05
-1,70	-4,21	-14,61	7,85
-12,34	20,10	-2,95	18,52
4,52	-7,07	10,47	-5,61

determinar la hipótesis a elegir.

La mejor región crítica es de la forma $\bar{x} \geq K$. Para hallar el valor de la constante K partimos del nivel de significación. Bajo la hipótesis nula la media muestral se distribuye $N(2; 2,5)$

$$\alpha = 0,10 = P(\bar{x} \geq K/H_0) = P[\bar{x} \geq K/N(2; 2,5)] = 5,205 .$$

Según la muestra extraída la media es igual a $-0,603$ y por ser menor que el valor crítico $5,205$ conduce a no rechazar la hipótesis nula, es decir, a aceptar que la media poblacional es igual a 2 .

Al ser la hipótesis alternativa compuesta, $\mu > 2$, no podemos calcular la potencia del contraste, sino su función de potencia, como vemos a continuación,

$$P(\mu) = P(\text{rechazar } H_0) = P[\bar{x} \geq 5,205/N(\mu; 2,5)]$$

tomando la función de potencia, $P(\mu)$, un valor distinto para cada uno de los posibles valores de la hipótesis alternativa, $\mu > 2$.

8.1.1.3. Hipótesis compuesta: $H_0 [\mu = \mu_0]$, $H_1 [\mu \neq \mu_0]$

Cuando la hipótesis alternativa es $\mu < \mu_0$ o $\mu > \mu_0$ el conocimiento del signo de la diferencia $\mu - \mu_0$ permite en [2] obtener la mejor región crítica mediante la aplicación del lema de Neyman-Pearson, sin embargo, cuando la hipótesis alternativa es de la forma $\mu \neq \mu_0$ desconocemos el signo de $\mu - \mu_0$ y, por tanto, el sentido de la desigualdad final cuando pretendemos dividir por la diferencia, no siendo posible determinar la mejor región crítica a través del lema de Neyman-Pearson. Para obtener un contraste tenemos que recurrir a otro procedimiento, el contraste razón de verosimilitud².

En la distribución normal, la función de verosimilitud de una muestra aleatoria simple de tamaño n es

$$L(\mathbf{X}; \mu) = \frac{1}{(\sigma^2)^{\frac{n}{2}} (2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2}.$$

El máximo de la función $L(\mathbf{X}; \mu)$ bajo la hipótesis nula, $H_0 [\mu = \mu_0]$, se alcanza en el punto μ_0 , puesto que el espacio paramétrico de la media μ se reduce al punto μ_0

$$L(\mathbf{X}; \mu_0) = \frac{1}{(\sigma^2)^{\frac{n}{2}} (2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_0)^2}.$$

Para el caso de la hipótesis alternativa $H_1 [\mu \neq \mu_0]$, calculamos el logaritmo de la función de verosimilitud

$$\ln L(\mathbf{X}; \mu) = -\frac{n}{2} \ln(\sigma^2) - \frac{n}{2} \ln(2\pi) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \quad [3]$$

² Véase el Capítulo 7.

hallamos la derivada de $\ln L(\mathbf{X}; \mu)$ respecto a μ para obtener el valor del parámetro que hace máxima la función $L(\mathbf{X}; \mu)$ en el espacio paramétrico ($-\infty < \mu < \infty$)

$$\frac{\partial \ln L(\mathbf{X}; \mu)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0$$

La solución de esta ecuación es el estimador máximo-verosímil de μ , la media muestral \bar{x} , y particularizada para ésta la función de verosimilitud es igual a

$$L(\mathbf{X}; \bar{x}) = \frac{1}{(\sigma^2)^{\frac{n}{2}} (2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{x})^2}$$

La razón de estas funciones de verosimilitud conduce al estadístico $\lambda(\mathbf{X})$

$$\lambda(\mathbf{X}) = \frac{L(\mathbf{X}; \mu_0)}{L(\mathbf{X}; \bar{x})} = \frac{e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_0)^2}}{e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{x})^2}}$$

y haciendo operaciones tenemos

$$\lambda(\mathbf{X}) = e^{-\frac{n}{2\sigma^2} (\bar{x} - \mu_0)^2}$$

La región crítica es $\lambda(\mathbf{X}) \leq K_2$, y para conocer el valor de la constante recurrimos a la definición del nivel de significación,

$$\alpha = P(0 \leq \lambda(\mathbf{X}) \leq K_2 / H_0) = \int_0^{K_2} g(\mathbf{X}/H_0) d\mathbf{X}$$

siendo $g(\mathbf{X}/H_0)$ la función de densidad en el muestreo de $\lambda(\mathbf{X})$ bajo la hipótesis nula. Sustituyendo $\lambda(\mathbf{X})$ por la expresión anterior tenemos

$$\begin{aligned} \alpha &= P(0 \leq \lambda(\mathbf{X}) \leq K_2) = P\left[0 \leq e^{-\frac{n}{2\sigma^2} (\bar{x} - \mu_0)^2} \leq K_2\right] = \\ &= P\left[-\infty < -\frac{n}{2\sigma^2} (\bar{x} - \mu_0)^2 \leq \ln K_2\right] = \\ &= P\left[\left(\frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right)^2 \geq -2 \ln K_2\right] \end{aligned}$$

ahora bien, la variable aleatoria \bar{x} , media muestral, bajo la hipótesis nula se distribuye $N\left(\mu_0; \frac{\sigma}{\sqrt{n}}\right)$, por lo cual la variable $\frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ es $N(0; 1)$.

Como la hipótesis alternativa es $\mu \neq \mu_0$ en la decisión que hayamos de tomar deberán ser válidos valores de μ tanto mayores que μ_0 como menores, por lo cual el contraste debe ser bilateral o de dos colas.

Operativamente se rechaza H_0 si

$$K_1 \leq \left(\frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right)^2$$

y este suceso tiene lugar cuando se da la unión de los sucesos

$$\frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \leq -K; \quad K \leq \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

donde $K = \sqrt{K_1}$ o, equivalentemente,

$$\left| \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right| \geq K.$$

Para hallar la constante K procedemos de la siguiente manera

$$P \left[\left| \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right| \geq K/H_0 \right] = P[|N(0; 1)| \geq K] = \alpha; \quad P[N(0; 1) \geq z_{\alpha/2}] = \frac{\alpha}{2}$$

y la región crítica es la unión de los intervalos

$$\bar{x} \leq \mu_0 - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}; \quad \bar{x} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}.$$

EJEMPLO 6

En una población $N(\mu; 7)$ se contrasta la hipótesis nula $H_0 [\mu = 20]$ frente a la alternativa $H_1 [\mu \neq 20]$, con un nivel de significación igual a 0,01 y mediante la muestra de tamaño 10

14	17	15	16	19	27	16	12	17	24
----	----	----	----	----	----	----	----	----	----

La región crítica es

$$\left| \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right| \geq K .$$

La media muestral es igual a 17,7, el valor tabular de la distribución $N(0; 1)$ tal que $P[N(0; 1) \geq z_{0,005}]$ es 2,575.

La región crítica resulta

$$\bar{x} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{\alpha/2} = 20 + \frac{7}{\sqrt{10}} 2,575 = 25,700$$

$$\bar{x} \leq \mu_0 - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} = 20 - \frac{7}{\sqrt{10}} 2,575 = 14,300$$

y la de aceptación

$$14,300 < \bar{x} < 25,700$$

comprobamos que la media de la muestra obtenida (17,7) está contenida en la región de aceptación por lo cual no se rechaza la hipótesis nula, $\mu = 20$, con un nivel de significación igual a 0,01.

8.1.2. VARIANZA POBLACIONAL DESCONOCIDA

8.1.2.1. Hipótesis simples: $H_0 [\mu = \mu_0], H_1 [\mu = \mu_1]$

Para resolver este problema recurrimos al contraste razón de verosimilitud. Los valores de μ que hacen máxima la función de verosimilitud en cada hipótesis son μ_0 y μ_1 . Al ser desconocida la varianza poblacional, σ^2 , la estimamos bajo cada una de las hipótesis.

Derivando el logaritmo de la función de verosimilitud [3] respecto a la varianza σ^2 , e igualando la derivada a cero tenemos la ecuación

$$\frac{\partial \ln L(\mathbf{X}; \mu)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 = 0$$

que conduce a

$$s^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}$$

y particularizando μ para μ_0 y μ_1 llegamos a los dos estimadores de la varianza σ^2 , bajo una y otra hipótesis

$$s_0^2 = \frac{\sum_{i=1}^n (x_i - \mu_0)^2}{n}; \quad s_1^2 = \frac{\sum_{i=1}^n (x_i - \mu_1)^2}{n} \quad [4]$$

con lo que el estadístico razón de verosimilitud $\lambda(\mathbf{X})$ es igual a

$$\lambda(\mathbf{X}) = \frac{L(\mathbf{X}; \mu_0, s_0^2)}{L(\mathbf{X}; \mu_1, s_1^2)} = \frac{\frac{1}{(s_0^2)^{\frac{n}{2}} (2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2s_0^2} \sum_{i=1}^n (x_i - \mu_0)^2}}{\frac{1}{(s_1^2)^{\frac{n}{2}} (2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2s_1^2} \sum_{i=1}^n (x_i - \mu_1)^2}}$$

Sustituyendo s_0^2 y s_1^2 por sus expresiones dadas en [4]

$$\lambda(\mathbf{X}) = \left[\frac{\sum_{i=1}^n (x_i - \mu_1)^2}{\sum_{i=1}^n (x_i - \mu_0)^2} \right]^{\frac{n}{2}} e^{-\frac{1}{2} \left[\frac{\sum_{i=1}^n (x_i - \mu_0)^2}{\sum_{i=1}^n (x_i - \mu_0)^2} - \frac{\sum_{i=1}^n (x_i - \mu_1)^2}{\sum_{i=1}^n (x_i - \mu_1)^2} \right]} = \left[\frac{\sum_{i=1}^n (x_i - \mu_1)^2}{\sum_{i=1}^n (x_i - \mu_0)^2} \right]^{\frac{n}{2}} \quad [5]$$

Distinguimos las dos situaciones que pueden darse $\mu_0 < \mu_1$ y $\mu_0 > \mu_1$.

A. $\mu_0 < \mu_1$

Haciendo operaciones³ en [5] llegamos a la variable aleatoria

$$t = \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}}$$

distribuida según $t(n-1)$.

³ RUIZ-MAYA y MARTÍN PLIEGO. *Estadística II: Inferencia*. Capítulo 10, especialmente por la referencia al teorema de Basu.

Cuando $t \geq t(n-1; \alpha)$ se rechaza la hipótesis nula al nivel de significación α .

La potencia del contraste es

$$1 - \beta = P(\mathbf{X} \in C/H_1) = P \left[\frac{\sqrt{n}(\bar{x} - \mu_0)}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}} \geq t(n-1; \alpha) / H_1 \right]$$

bajo la hipótesis nula, la variable

$$t = \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}}$$

se distribuye como $t(n-1)$. Pero si es cierta la hipótesis alternativa, $\mu = \mu_1$, en el numerador de la variable la esperanza de la diferencia $\bar{x} - \mu_0$ no se anula para μ_1 , y el cociente sigue una distribución t no central⁴.

B. $\mu_0 > \mu_1$

Siguiendo el mismo procedimiento que para la hipótesis alternativa cuando $\mu_0 < \mu_1$, la región crítica si $\mu_0 > \mu_1$ es

$$\frac{\sqrt{n}(\bar{x} - \mu_0)}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}} \leq t(n-1; 1 - \alpha).$$

8.1.2.2. Hipótesis compuestas: $H_0 [\mu = \mu_0]$, $H_1 [\mu > \mu_0]$ o $H_1 [\mu < \mu_0]$

Para deducir la región crítica precisa para contrastar estas hipótesis el método es similar al anterior, $H_1 (\mu = \mu_1)$, y la región crítica cuando $\mu > \mu_0$ es

$$\frac{\sqrt{n}(\bar{x} - \mu_0)}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}} \geq t(n-1; \alpha)$$

⁴ MARTÍN PLIEGO y RUIZ-MAYA: *Fundamentos...* Capítulo 7.

y si, por el contrario, $\mu < \mu_0$, el signo de la desigualdad cambia

$$\frac{\sqrt{n}(\bar{x} - \mu_0)}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}} \leq t(n-1; 1-\alpha).$$

La hipótesis alternativas son compuestas por lo que no tenemos potencias concretas sino funciones de potencia, dependiendo de la distribución t no central.

8.1.2.3. Hipótesis compuesta: $H_0 [\mu = \mu_0], H_1 [\mu \neq \mu_0]$

La razón de verosimilitud, en este caso, es el cociente de la función de verosimilitud particularizada para los valores de μ y σ^2 que la hacen máxima en el espacio paramétrico de la hipótesis nula $\{\mu_0; \sigma^2 > 0\}$ y en el espacio paramétrico total $\{-\infty < \mu < \infty; \sigma^2 > 0\}$.

Hallamos las expresiones de los estimadores de μ y σ^2 bajo cada hipótesis. En el espacio paramétrico de la hipótesis nula $\{\mu = \mu_0; \sigma^2 > 0\}$ la media toma un solo valor, μ_0 , por lo que será el que hace máxima $L(\mathbf{X}; \mu, \sigma^2)$. El valor correspondiente a σ^2 es su estimador máximo-verosímil

$$s_0^2 = \frac{\sum_{i=1}^n (x_i - \mu_0)^2}{n}.$$

En el espacio paramétrico total, los estimadores de μ y σ^2 son

$$\hat{\mu} = \frac{\sum_{i=1}^n x_i}{n} = \bar{x}; \quad \hat{\sigma}^2 = s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}.$$

Llevando estos valores a la razón de verosimilitudes, el estadístico $\lambda(\mathbf{X})$ es

$$\lambda(\mathbf{X}) = \frac{L(\mathbf{X}; \mu_0, s_0^2)}{L(\mathbf{X}; \hat{\mu}, s^2)} = \frac{\frac{1}{(s_0^2)^{\frac{n}{2}} (2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2s_0^2} \sum_{i=1}^n (x_i - \mu_0)^2}}{\frac{1}{(s^2)^{\frac{n}{2}} (2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2s^2} \sum_{i=1}^n (x_i - \hat{\mu})^2}}$$

sustituyendo s_0^2 , s^2 y $\hat{\mu}$ por sus expresiones en función de los valores muestrales y haciendo operaciones tenemos que

$$\lambda(\mathbf{X}) = \left[\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \mu_0)^2} \right]^{\frac{n}{2}}.$$

Teniendo como resultado final que la expresión del estadístico utilizado en el contraste es igual a

$$t = \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}}. \tag{6}$$

y la región crítica $|t| \geq t(n-1; \alpha/2)$.

Operativamente el proceso es el siguiente: fijado el nivel de significación, α , y tomada una muestra aleatoria simple de tamaño n , se calcula el estadístico t y se halla el valor tabular de la distribución $t\left(n-1; \frac{\alpha}{2}\right) = t_{\alpha/2}$. Si $|t| \geq t_{\alpha/2}$, o lo equivalente $t \leq -t_{\alpha/2}$, $t \geq t_{\alpha/2}$, se rechaza la hipótesis nula, $\mu = \mu_0$, y se acepta la alternativa, $\mu \neq \mu_0$.

EJEMPLO 4

Para contrastar la hipótesis nula $H_0 [\mu = 3]$ frente a la alternativa $H_1 [\mu \neq 3]$ en la distribución $N(\mu; \sigma)$, con un nivel de significación igual a 0,013, siendo la varianza desconocida, se toma una muestra aleatoria simple de tamaño 18, cuyo resultado es

4,19	3,98	3,65	3,25	3,86	4,03	3,93	3,83	3,98
4,08	3,93	3,75	3,93	3,98	3,84	4,01	4,16	4,01

Hallamos el valor muestral del estadístico [6]

$$t = \frac{\sqrt{18}(\bar{x} - 3)}{\sqrt{\frac{\sum_{i=1}^{18} (x_i - \bar{x})^2}{17}}},$$

siendo $\bar{x} = 3,911$ y $\sum_{i=1}^{18} (x_i - 3,911)^2 = 0,762698$, por lo que

$$t = \frac{\sqrt{18} (3,911 - 3)}{\sqrt{\frac{0,762698}{17}}} = 18,247 .$$

El valor tabular de la distribución t de Student es $t(17; 0,0065) = 2,7988$ y como el valor muestral (18,247) resulta mayor que el tabular (2,7988) la muestra pertenece a la región crítica por lo que se rechaza la hipótesis nula y se acepta que la media de la distribución poblacional es distinta de 3.

8.1.4 CUADRO RESUMEN DE LOS CONTRASTES DE MEDIAS EN POBLACIONES NORMALES

Resumimos en la página siguiente los contrastes estudiados para la media de una distribución normal.

Contrastación de hipótesis sobre la media de una población $N(\mu; \sigma)$

Hipótesis		Región crítica		
Nula	Alternativa			
<i>Varianza poblacional conocida</i>				
$\mu = \mu_0$	$\mu = \mu_1$	$\mu_0 < \mu_1$	$\bar{x} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha$	(*)
		$\mu_0 > \mu_1$	$\bar{x} \leq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha}$	(*)
	$\mu > \mu_0$		$\bar{x} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha$	(*)
	$\mu < \mu_0$		$\bar{x} \leq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha}$	(*)
	$\mu \neq \mu_0$		$ \bar{x} - \mu_0 \geq \frac{\sigma}{\sqrt{n}} z_{\alpha/2}$	(**)
$\mu \leq \mu_0$	$\mu > \mu_0$		$\bar{x} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha$	(*)
$\mu \geq \mu_0$	$\mu < \mu_0$		$\bar{x} \leq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha}$	(*)
<i>Varianza poblacional desconocida</i>				
$\mu = \mu_0$	$\mu = \mu_1$	$\mu_0 < \mu_1$	$\bar{x} \geq \mu_0 + \frac{s}{\sqrt{n-1}} t_\alpha$	(*)
		$\mu_0 > \mu_1$	$\bar{x} \leq \mu_0 + \frac{s}{\sqrt{n-1}} t_{1-\alpha}$	(*)
	$\mu > \mu_0$		$\bar{x} \geq \mu_0 + \frac{s}{\sqrt{n-1}} t_\alpha$	(*)
	$\mu < \mu_0$		$\bar{x} \leq \mu_0 + \frac{s}{\sqrt{n-1}} t_{1-\alpha}$	(*)
	$\mu \neq \mu_0$		$ \bar{x} - \mu_0 \geq \frac{s}{\sqrt{n-1}} t_{\alpha/2}$	(***)
$\mu \leq \mu_0$	$\mu > \mu_0$		$\bar{x} \geq \mu_0 + \frac{s}{\sqrt{n-1}} t_\alpha$	(***)
$\mu \geq \mu_0$	$\mu < \mu_0$		$\bar{x} \leq \mu_0 + \frac{s}{\sqrt{n-1}} t_{1-\alpha}$	(***)
$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$				
$z_\alpha: P[N(0; 1) \geq z_\alpha] = \alpha; \quad z_{1-\alpha}: P[N(0; 1) \geq z_{1-\alpha}] = 1 - \alpha; \quad z_{\alpha/2}: P[N(0; 1) \geq z_{\alpha/2}] = \alpha/2$				
$t_\alpha: P[t(n-1) \geq t_\alpha] = \alpha; \quad t_{1-\alpha}: P[t(n-1) \geq t_{1-\alpha}] = 1 - \alpha; \quad t_{\alpha/2}: P[t(n-1) \geq t_{\alpha/2}] = \alpha/2$				
(*) Uniformemente más potente. (**) Uniformemente más potente insesgado. (***) Uniformemente más potente, insesgado e invariante.				

8.2 Contratación de hipótesis sobre la varianza de una población $N(\mu; \sigma)$

8.2.1. MEDIA POBLACIONAL CONOCIDA

8.2.1.1. Hipótesis simples: $H_0 [\sigma^2 = \sigma_0^2]$, $H_1 [\sigma^2 = \sigma_1^2]$

Para la determinación de la mejor región crítica al ser las dos hipótesis simples, aplicamos el lema de Neyman-Pearson, obteniendo previamente la función de verosimilitud que, para muestras aleatorias simples de tamaño n de la población $N(\mu; \sigma)$, es

$$L(\mathbf{X}; \sigma^2) = \frac{1}{(\sigma^2)^{\frac{n}{2}} (2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2}.$$

El cociente de funciones de verosimilitud particularizadas para la hipótesis nula $[\sigma^2 = \sigma_0^2]$ y alternativa $[\sigma^2 = \sigma_1^2]$ es igual a

$$\frac{L(\mathbf{X}; \sigma_0^2)}{L(\mathbf{X}; \sigma_1^2)} = \left(\frac{\sigma_1^2}{\sigma_0^2} \right)^{\frac{n}{2}} e^{-\frac{1}{2} \left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2} \right) \sum_{i=1}^n (x_i - \mu)^2} \leq K_2$$

tomando logaritmos

$$\frac{n}{2} \ln \frac{\sigma_1^2}{\sigma_0^2} - \frac{1}{2} \left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2} \right) \sum_{i=1}^n (x_i - \mu)^2 \leq \ln K_2$$

y haciendo operaciones

$$\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2} \right) \sum_{i=1}^n (x_i - \mu)^2 \geq K_1 \quad [7]$$

Como en el caso del contraste de medias, debemos tomar en consideración el signo de la diferencia $\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}$, es decir, distinguir el que σ_0^2 sea menor que σ_1^2 o mayor.

A. $\sigma_0^2 < \sigma_1^2$

Si $\sigma_0^2 < \sigma_1^2$, sus valores inversos verifican $\frac{1}{\sigma_0^2} > \frac{1}{\sigma_1^2}$ y la diferencia $\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}$ es mayor que cero, por lo cual al despejar en [7] $\sum_{i=1}^n (x_i - \mu)^2$ no cambia el sentido de la desigualdad, quedando como mejor región crítica

$$\sum_{i=1}^n (x_i - \mu)^2 \geq K.$$

Para calcular el valor de la constante K , y que la región crítica quede perfectamente determinada, partimos del nivel de significación, α ,

$$\alpha = P(\mathbf{X} \in C/H_0) = P\left(\sum_{i=1}^n (x_i - \mu)^2 \geq K / H_0\right) \quad [8]$$

requiriendo conocer la distribución en el muestreo del estadístico $\sum_{i=1}^n (x_i - \mu)^2$ bajo la hipótesis nula, donde las variables aleatorias independientes x_i se distribuyen $N(\mu; \sigma_0)$. Para ello dividimos la variable $x_i - \mu$ por la desviación típica σ_0 , $\frac{x_i - \mu}{\sigma_0}$, siendo la distribución de la nueva variable $N(0; 1)$.

Por todo esto, si dividimos por σ_0^2 la desigualdad $\sum_{i=1}^n (x_i - \mu)^2 \geq K$, que define la región crítica,

$$\sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma_0}\right)^2 \geq \frac{K}{\sigma_0^2}$$

tenemos en el primer miembro la suma de cuadrados de n variables aleatorias $N(0; 1)$ e independiente, distribuyéndose la suma según una variable $\chi^2(n)$, por lo cual estamos en condiciones de hallar el valor de la contante K que delimita la mejor región crítica, una vez fijado el nivel de significación.

Volviendo a [8]

$$\alpha = P\left[\sum_{i=1}^n (x_i - \mu)^2 \geq K / H_0\right] = P\left[\sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma_0}\right)^2 \geq \frac{K}{\sigma_0^2}\right] = P\left[\chi^2(n) \geq \frac{K}{\sigma_0^2}\right]$$

la constante $\frac{K}{\sigma_0^2}$ es el valor tabular, c_α , que verifica la ecuación

$$P[\chi^2(n) \geq c_\alpha] = \alpha$$

y de aquí deducimos (Fig 8.3)

$$K = c_\alpha \sigma_0^2.$$

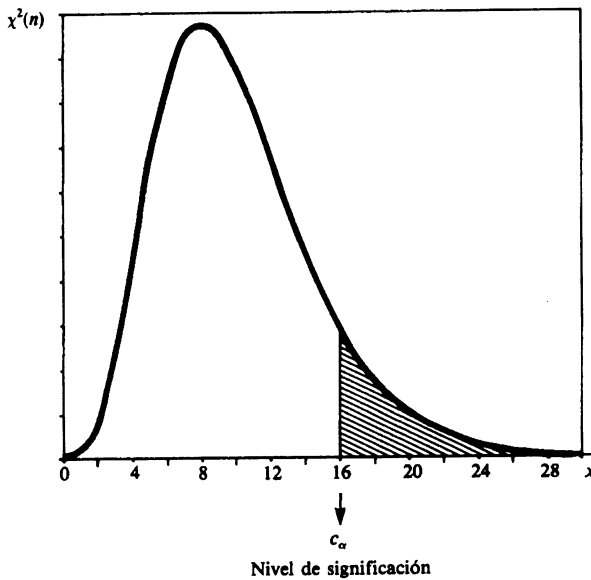


FIGURA 8.3

Función de densidad de la variable $\chi^2(n)$ para contrastar $H_0[\sigma^2 = \sigma_0^2]$ y $H_1[\sigma^2 = \sigma_1^2]$, $\sigma_0^2 < \sigma_1^2$.

La potencia del contraste (Fig. 8.4) se calcula de la siguiente manera

$$\begin{aligned} 1 - \beta &= P\left[\sum_{i=1}^n (x_i - \mu)^2 \geq K / H_1\right] = P\left[\sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma_1}\right)^2 \geq \frac{K}{\sigma_1^2}\right] = \\ &= P\left[\chi^2(n) \geq \frac{K}{\sigma_1^2}\right] = P\left[\chi^2(n) \geq c_\alpha \frac{\sigma_0^2}{\sigma_1^2}\right] \end{aligned}$$

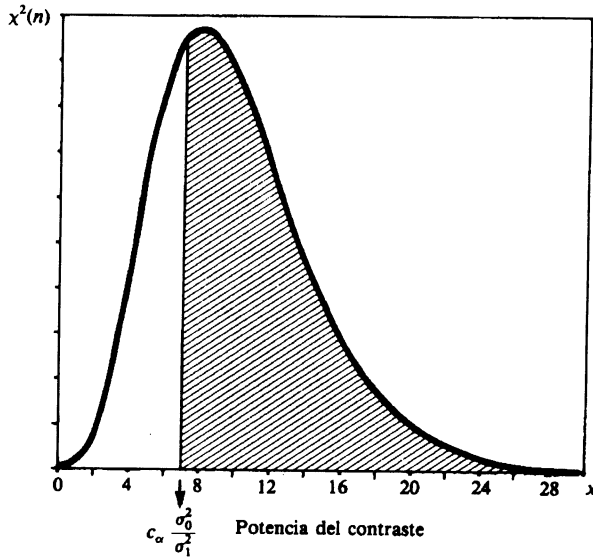


FIGURA 8.4
 Función de densidad de la variable $\chi^2(n)$ para contrastar $H_0[\sigma^2 = \sigma_0^2]$ y $H_1[\sigma^2 = \sigma_1^2]$, $\sigma_0^2 < \sigma_1^2$.

EJEMPLO 4

En una población $N(5; \sigma)$ la hipótesis nula es $\sigma^2 = \sigma_0^2 = 2$ y la alternativa $\sigma^2 = \sigma_1^2 = 4$, $\sigma_0^2 < \sigma_1^2$. El valor del nivel de significación es 0,025. La contrastación se efectúa por medio de la siguiente muestra aleatoria simple de tamaño 10

5,1	6,2	4,0	2,5	3,7
5,5	3,8	3,5	5,7	2,7

La mejor región crítica es de la forma

$$\sum_{i=1}^{10} (x_i - 5)^2 \geq K.$$

Hallamos el valor de la constante K

$$\alpha = 0,025 = P\left[\chi^2(10) \geq \frac{K}{\sigma_0^2}\right] = P\left[\chi^2(10) \geq \frac{K}{2}\right]$$

$\frac{K}{2} = 20,483$ y $K = 40,966$, siendo la mejor región crítica

$$\sum_{i=1}^{10} (x_i - 5)^2 \geq 40,966 .$$

El valor muestral del estadístico $\sum_{i=1}^{10} (x_i - 5)^2$ es igual a 20,11 que, por ser menor que 40,966, conduce a no rechazar la hipótesis nula, $N(5; \sqrt{2})$, pues se halla en la región de aceptación, $\sum_{i=1}^{10} (x_i - 5)^2 < 40,966$,

Potencia del contraste

$$1 - \beta = P \left[\chi^2(10) \geq \frac{K}{\sigma_1^2} \right] = P \left[\chi^2(10) \leq \frac{40,966}{4} \right] = P[\chi^2(10) \geq 10,2415] = 0,43 .$$

B. $\sigma_0^2 > \sigma_1^2$

La diferencia $\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}$ es negativa y la mejor región crítica pasa a ser

$$\sum_{i=1}^n (x_i - \mu)^2 \leq K$$

hallándose la constante K como solución de la ecuación

$$\alpha = P \left[\chi^2(n) \leq \frac{K}{\sigma_0^2} \right]$$

o también

$$P[\chi^2(n) \geq c_{1-\alpha}] = 1 - \alpha$$

siendo $c_{1-\alpha}$ igual a $\frac{K}{\sigma_0^2}$, o $K = c_{1-\alpha} \sigma_0^2$.

La potencia del contraste es

$$1 - \beta = 1 - P \left[\chi^2(n) \geq c_{1-\alpha} \frac{\sigma_0^2}{\sigma_1^2} \right] .$$

8.2.1.2. Hipótesis compuestas: $H_0[\sigma^2 = \sigma_0^2]$, $H_1[\sigma^2 > \sigma_0^2]$ o $H_1[\sigma^2 < \sigma_0^2]$

Tanto un contraste como otro se llevan a cabo recurriendo al tema de Neyman-Pearson y llegando a dos regiones críticas según que σ_0^2 sea mayor o menor que σ_1^2 .

A. $H_1[\sigma^2 > \sigma_0^2]$

Región crítica

$$\sum_{i=1}^n (x_i - \mu)^2 \geq K$$

el valor de K se obtiene como solución de la ecuación

$$\alpha = P \left[\sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma_0} \right)^2 \geq \frac{K}{\sigma_0^2} \right] = P \left[\chi^2(n) \geq \frac{K}{\sigma_0^2} \right] = P[\chi^2(n) \geq c_\alpha]$$

siendo $K = c_\alpha \sigma_0^2$.

B. $H_1[\sigma^2 < \sigma_0^2]$

Análogamente al caso anterior la región crítica es

$$\sum_{i=1}^n (x_i - \mu)^2 \leq K$$

$$\alpha = P \left[\sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma_0} \right)^2 \leq \frac{K}{\sigma_0^2} \right] = P \left[\chi^2(n) \leq \frac{K}{\sigma_0^2} \right] = P[\chi^2(n) \leq c_{1-\alpha}]$$

y $K = c_{1-\alpha} \sigma_0^2$.

8.2.1.3. Hipótesis compuesta: $H_0[\sigma^2 = \sigma_0^2]$, $H_1[\sigma^2 \neq \sigma_0^2]$

Siguiendo los pasos indicados por el lema de Neyman-Pearson llegamos a la desigualdad

$$\left[\frac{1}{\sigma_0^2} - \frac{1}{\sigma^2} \right] \sum_{i=1}^n (x_i - \mu)^2 \geq K_1$$

y al desconocer el signo de $\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}$ no podemos obtener la mejor región crítica mediante este procedimiento, siendo preciso recurrir a un test razón de verosimilitud.

Como desconocemos σ^2 en la hipótesis alternativa lo estimamos mediante el estimador máximo-verosímil, en muestra aleatorias simples de tamaño n ,

$$\hat{\sigma}^2 = s^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n} .$$

El cociente de funciones de verosimilitud es

$$\lambda(\mathbf{X}) = \frac{L(\mathbf{X}; \sigma_0^2)}{L(\mathbf{X}; s^2)} = \frac{\frac{1}{(\sigma_0^2)^2 (2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma_0^2} \sum_{i=1}^n (x_i - \mu)^2}}{\frac{1}{(s^2)^2 (2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2s^2} \sum_{i=1}^n (x_i - \mu)^2}} = \left(\frac{s^2}{\sigma_0^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2 \left(\frac{1}{\sigma_0^2} - \frac{1}{s^2}\right)}$$

sustituyendo s^2 por su expresión

$$\begin{aligned} \lambda(\mathbf{X}) &= \left[\frac{1}{n} \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_0^2} \right]^{\frac{n}{2}} e^{-\frac{1}{2} \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_0^2} + \frac{n}{2}} = \\ &= \left[\frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma_0} \right)^2 \right]^{\frac{n}{2}} e^{-\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma_0} \right)^2 + \frac{n}{2}} \end{aligned}$$

y como bajo la hipótesis nula, $N(\mu; \sigma_0)$,

$$\sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma_0} \right)^2$$

es una variable χ^2 con n grados de libertad resulta

$$\lambda(\mathbf{X}) = \left[\frac{1}{n} \chi^2(n) \right]^{\frac{n}{2}} e^{-\frac{1}{2} \chi^2(n) + \frac{n}{2}} .$$

La razón de verosimilitud, $\lambda(\mathbf{X})$, y, por consiguiente, la región crítica $\lambda(\mathbf{X}) \leq K$, es función de $\chi^2(n)$.

En la figura 8.5 vemos que $\lambda(\mathbf{X})$ será menor que K cuando $\chi^2(n) \leq L_1$ o cuando $\chi^2(n) \geq L_2$, es decir,

$$\{\lambda(\mathbf{X}) \leq K\} \Rightarrow \{\chi^2(n) \leq L_1\} \cup \{\chi^2(n) \geq L_2\}$$

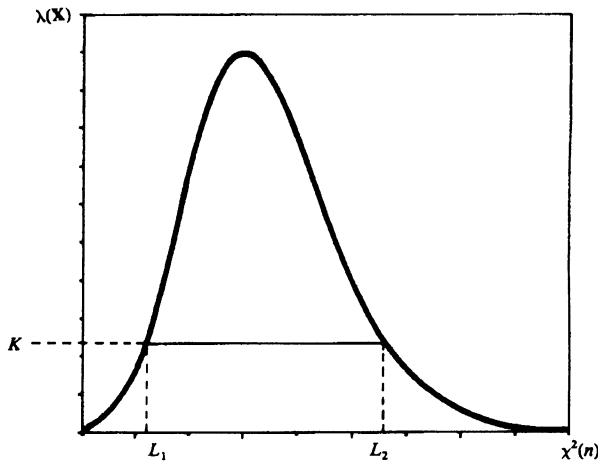


FIGURA 8.5

y como los sucesos $\{\chi^2(n) \leq L_1\}$ y $\{\chi^2(n) \geq L_2\}$ son disjuntos el nivel de significación será

$$\alpha = P[\lambda(\mathbf{X}) \leq K/H_0] = P[\chi^2(n) \leq L_1] + P[\chi^2(n) \geq L_2] = \alpha_1 + \alpha_2 .$$

Desde el punto de vista operativo, y para obtener un intervalo único de tamaño $1 - \alpha$, tomamos en la distribución $\chi^2(n)$ $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$, siendo L_1 la solución de $P[\chi^2(n) \leq L_1] = \frac{\alpha}{2}$ y L_2 la de $P[\chi^2(n) \geq L_2] = \frac{\alpha}{2}$.

Rechazaremos la hipótesis nula, $H_0[\sigma^2 = \sigma_0^2]$, cuando se verifique

$$\sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma_0} \right)^2 \leq \chi^2 \left(n; 1 - \frac{\alpha}{2} \right)$$

o bien

$$\sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma_0} \right)^2 \geq \chi^2 \left(n; \frac{\alpha}{2} \right)$$

y se aceptará cuando

$$\chi^2 \left(n; 1 - \frac{\alpha}{2} \right) \sigma_0^2 < \sum_{i=1}^n (x_i - \mu)^2 < \chi^2 \left(n; \frac{\alpha}{2} \right) \sigma_0^2.$$

La hipótesis alternativa de este contraste, $H_1 [\sigma^2 \neq \sigma_0^2]$, es compuesta por lo cual no podemos hablar de potencia sino de función de potencia⁵

$$P(\sigma^2) = P \left[\chi^2(n; \delta) \leq \chi^2 \left(n; 1 - \frac{\alpha}{2} \right) \right] + P \left[\chi^2(n; \delta) \geq \chi^2 \left(n; \frac{\alpha}{2} \right) \right]$$

la variable $\chi^2(n; \delta)$ es la distribución χ^2 no central⁶ con parámetro δ de no centralidad.

EJEMPLO 5

En una población $N(15; \sigma)$ se contrasta la hipótesis nula $H_0 [\sigma^2 = 4]$ frente a la alternativa $H_1 [\sigma^2 \neq 4]$ con un nivel de significación igual al 10%. La muestra aleatoria simple obtenida es

12	16	9	17
----	----	---	----

El valor del estadístico muestral $\sum_{i=1}^4 (x_i - 15)^2$ es igual a 50, y dividiéndolo por la varianza establecida en la hipótesis nula, $\sigma^2 = 4$, resulta 12,5. Los valores tabulares son

$$\chi^2 \left(n; \frac{\alpha}{2} \right) = \chi^2(4; 0,05) = 9,488$$

$$\chi^2 \left(n; 1 - \frac{\alpha}{2} \right) = \chi^2(4; 0,95) = 0,717$$

La región de aceptación es el intervalo $[0,717; 9,488]$, y al no pertenecer el valor muestral del estadístico (12,5) a la región de aceptación rechazamos la hipótesis nula, por lo que aceptamos que la varianza de la distribución es distinta de 4, $\sigma^2 \neq 4$.

⁵ RUIZ-MAYA y MARTÍN PLIEGO: *Estadística II: Inferencia*. Capítulo 10.

⁶ MARTÍN PLIEGO y RUIZ-MAYA: *Fundamentos...* Capítulo 7.



8.2.2. MEDIA POBLACIONAL DESCONOCIDA

8.2.2.1. Hipótesis simple: $H_0 [\sigma^2 = \sigma_0^2], H_1 [\sigma^2 = \sigma_1^2]$

Para efectuar este contraste no podemos aplicar el lema de Neyman-Pearson pues las hipótesis son compuestas al desconocerse la media poblacional: el espacio paramétrico en la hipótesis nula es $\{-\infty < \mu < \infty; \sigma^2 = \sigma_0^2\}$, y en la alternativa $\{-\infty < \mu < \infty; \sigma^2 = \sigma_1^2\}$, teniendo que recurrir al contraste razón de verosimilitud, cuya región crítica no es uniformemente más potente.

Al desconocer la media poblacional, μ , calculamos su estimador máximo-verosímil, la media muestral \bar{x} . Los valores que hacen máxima la función de verosimilitud son en la hipótesis nula $(\bar{x}; \sigma_0^2)$ y en la alternativa $(\bar{x}; \sigma_1^2)$. El cociente de los valores máximos de la función es

$$\lambda(\mathbf{X}) = \frac{L(\mathbf{X}; \bar{x}, \sigma_0^2)}{L(\mathbf{X}; \bar{x}, \sigma_1^2)} = \left(\frac{\sigma_1^2}{\sigma_0^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2}\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right) \sum_{i=1}^n (x_i - \bar{x})^2} \leq K_2$$

$$e^{-\frac{1}{2}\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right) \sum_{i=1}^n (x_i - \bar{x})^2} \leq \left(\frac{\sigma_0^2}{\sigma_1^2}\right)^{\frac{n}{2}} K_2$$

y tomando logaritmos

$$\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right) \sum_{i=1}^n (x_i - \bar{x})^2 \geq K_1$$

Para determinar la región crítica distinguimos entre $\sigma_0^2 < \sigma_1^2$ y $\sigma_0^2 > \sigma_1^2$.

A. $\sigma_0^2 < \sigma_1^2$

Si $\sigma_0^2 < \sigma_1^2$ la diferencia $\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}$ es mayor que cero y la región crítica toma la forma

$$\sum_{i=1}^n (x_i - \bar{x})^2 \geq K$$

deduciéndose el valor de K de la ecuación

$$\alpha = P(\mathbf{X} \in C/H_0) = P\left[\sum_{i=1}^n (x_i - \bar{x})^2 \geq K / H_0\right].$$

Si la hipótesis nula es cierta, $\sigma^2 = \sigma_0^2$, la variable $\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2}$ se distribuye $\chi^2(n-1)$, por lo cual

$$\alpha = P\left[\sum_{i=1}^n (x_i - \bar{x})^2 \geq K / H_0\right] = P\left[\chi^2(n-1) \geq \frac{K}{\sigma_0^2}\right]$$

$\frac{K}{\sigma_0^2}$ es el valor tabular de $\chi^2(n-1; \alpha)$, d_α , y $K = d_\alpha \sigma_0^2$, resultando la región crítica

$$\sum_{i=1}^n (x_i - \bar{x})^2 \geq d_\alpha \sigma_0^2.$$

La potencia del contraste es

$$1 - \beta = P(\mathbf{X} \in C / H_1) = P\left[\sum_{i=1}^n (x_i - \bar{x})^2 \geq K / H_1\right] = P\left[\chi^2(n-1) \geq \frac{d_\alpha \sigma_0^2}{\sigma_1^2}\right].$$

EJEMPLO 6

Para contrastar, con un nivel de significación $\alpha = 0,01$, en una distribución $N(\mu; \sigma)$ la hipótesis nula $H_0[\sigma_0^2 = 25]$ y la alternativa $H_1[\sigma_1^2 = 36]$, se toma una muestra aleatoria simple de tamaño 18, cuya varianza es igual a 30. Hállese la potencia del contraste.

La región crítica es

$$\sum_{i=1}^{18} (x_i - \bar{x})^2 \geq d_\alpha \sigma_0^2 = d_{0,01} 25$$

La varianza muestral es igual a

$$s^2 = \frac{\sum_{i=1}^{18} (x_i - \bar{x})^2}{18} = 30$$

y de aquí

$$\sum_{i=1}^{18} (x_i - \bar{x})^2 = 540.$$

Para hallar $d_{0,01}$ sabemos que $P[\chi^2(17) \geq d_{0,01}] = 0,01$, y el valor tabular $d_{0,01} = 33,409$.

Como $\sum_{i=1}^{18} (x_i - \bar{x})^2 = 540$ es menor que $d_{0,01} \sigma_0^2 = 33,409 \cdot 25 = 835,225$ no se rechaza la hipótesis nula.

La potencia del contraste es igual a

$$P\left[\chi^2(17) \geq \frac{d_{0,01} \sigma_0^2}{\sigma_1^2}\right] = P\left[\chi^2(17) \geq \frac{835,225}{36}\right] = P[\chi^2(17) \geq 23,2] = 0,195.$$

B. $\sigma_0^2 > \sigma_1^2$

Al ser mayor la varianza bajo la hipótesis nula, σ_0^2 , que bajo la alternativa, σ_1^2 , la diferencia de los inversos de las varianzas es negativa y la región crítica igual a

$$\sum_{i=1}^n (x_i - \bar{x})^2 \leq K$$

Siguiendo un procedimiento análogo al del caso anterior, $\sigma_0^2 < \sigma_1^2$, llegamos a que la constante K se calcula mediante la ecuación

$$P\left[\chi^2(n-1) \leq \frac{K}{\sigma_0^2}\right] = \alpha$$

o también

$$P[\chi^2(n-1) \geq d_{1-\alpha}] = 1 - \alpha$$

y $K = d_{1-\alpha} \sigma_0^2$

La potencia del contraste resulta

$$1 - \beta = P\left[\chi^2(n-1) \leq d_{1-\alpha} \frac{\sigma_0^2}{\sigma_1^2}\right] = 1 - P\left[\chi^2(n-1) \geq d_{1-\alpha} \frac{\sigma_0^2}{\sigma_1^2}\right].$$

8.2.2.2. Hipótesis compuesta:⁷ $H_0[\sigma^2 = \sigma_0^2]$, $H_1[\sigma^2 > \sigma_0^2]$ o $H_1[\sigma^2 < \sigma_0^2]$

- Si la hipótesis alternativa es $H_1[\sigma^2 < \sigma_0^2]$

La región crítica resulta

$$\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2} \leq \chi^2(n-1; 1-\alpha)$$

- Si la hipótesis alternativa es $H_1[\sigma^2 > \sigma_0^2]$ se llega a la región crítica

$$\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2} \geq \chi^2(n-1; \alpha).$$

8.2.2.3. Hipótesis compuesta: $H_0[\sigma^2 = \sigma_0^2]$, $H_1[\sigma^2 \neq \sigma_0^2]$

Como en casos anteriores tenemos que recurrir al contraste razón de verosimilitud.

El espacio paramétrico en cada una de las dos hipótesis es: en la nula $\{-\infty < \mu < \infty; \sigma^2 = \sigma_0^2\}$, y en la alternativa $\{-\infty < \mu < \infty; \sigma^2 \neq \sigma_0^2\}$.

Rechazaremos la hipótesis nula, $H_0(\sigma^2 = \sigma_0^2)$, cuando se verifique

$$\sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma_0} \right)^2 \leq \chi^2 \left(n-1; 1 - \frac{\alpha}{2} \right),$$

o bien

$$\sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma_0} \right)^2 \geq \chi^2 \left(n-1; \frac{\alpha}{2} \right).$$

8.2.3 CUADRO RESUMEN DE LOS CONTRASTES DE VARIANZA EN POBLACIONES NORMALES

Los contrastes anteriores esquemáticamente se pueden resumir en el siguiente cuadro:

⁷ RUIZ-MAYA y MARTÍN PLIEGO. *Estadística II: Inferencia*. Capítulo 10.

Contrastación de hipótesis sobre la varianza de una población $N(\mu; \sigma)$.

Hipótesis		Región crítica		
Nula	Alternativa			
<i>Media poblacional conocida</i>				
$\sigma^2 = \sigma_0^2$	$\sigma^2 = \sigma_1^2$	$\sigma_0^2 < \sigma_1^2$	$\sum_{i=1}^n (x_i - \mu)^2 \geq c_{\alpha} \sigma_0^2$	(*)
		$\sigma_0^2 > \sigma_1^2$	$\sum_{i=1}^n (x_i - \mu)^2 \leq c_{1-\alpha} \sigma_0^2$	(*)
	$\sigma^2 > \sigma_0^2$	$\sum_{i=1}^n (x_i - \mu)^2 \geq c_{\alpha} \sigma_0^2$	(*)	
	$\sigma^2 < \sigma_0^2$	$\sum_{i=1}^n (x_i - \mu)^2 \leq c_{1-\alpha} \sigma_0^2$	(*)	
	$\sigma^2 \neq \sigma_0^2$	$\sum_{i=1}^n (x_i - \mu)^2 \leq c_{1-\alpha/2} \sigma_0^2$ $\sum_{i=1}^n (x_i - \mu)^2 \geq c_{\alpha/2} \sigma_0^2$		
$\sigma^2 \leq \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\sum_{i=1}^n (x_i - \mu)^2 \geq c_{\alpha} \sigma_0^2$	(*)	
$\sigma^2 \geq \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\sum_{i=1}^n (x_i - \mu)^2 \leq c_{1-\alpha} \sigma_0^2$	(*)	
<i>Media poblacional desconocida</i>				
$\sigma^2 = \sigma_0^2$	$\sigma^2 = \sigma_1^2$	$\sigma_0^2 < \sigma_1^2$	$\sum_{i=1}^n (x_i - \bar{x})^2 \geq d_{1-\alpha} \sigma_0^2$	(*)
		$\sigma_0^2 > \sigma_1^2$	$\sum_{i=1}^n (x_i - \bar{x})^2 \leq d_{\alpha} \sigma_0^2$	(*)
	$\sigma^2 > \sigma_0^2$	$\sum_{i=1}^n (x_i - \bar{x})^2 \geq d_{\alpha} \sigma_0^2$	(*)	
	$\sigma^2 < \sigma_0^2$	$\sum_{i=1}^n (x_i - \bar{x})^2 \leq d_{1-\alpha} \sigma_0^2$	(*)	
	$\sigma^2 \neq \sigma_0^2$	$\sum_{i=1}^n (x_i - \bar{x})^2 \leq d_{1-\alpha/2} \sigma_0^2$ $\sum_{i=1}^n (x_i - \bar{x})^2 \geq d_{\alpha/2} \sigma_0^2$		
$\sigma^2 \leq \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\sum_{i=1}^n (x_i - \bar{x})^2 \geq d_{\alpha} \sigma_0^2$	(**)	
$\sigma^2 \geq \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\sum_{i=1}^n (x_i - \bar{x})^2 \leq d_{1-\alpha} \sigma_0^2$	(**)	
$c_{\alpha} : P[\chi^2(n) \geq c_{\alpha}] = \alpha; \quad c_{1-\alpha} : P[\chi^2(n) \geq c_{1-\alpha}] = 1 - \alpha; \quad c_{\alpha/2} : P[\chi^2(n) \geq c_{\alpha/2}] = \alpha/2$ $d_{\alpha} : P[\chi^2(n-1) \geq d_{\alpha}] = \alpha; \quad d_{1-\alpha} : P[\chi^2(n-1) \geq d_{1-\alpha}] = 1 - \alpha; \quad d_{\alpha/2} : P[\chi^2(n-1) \geq d_{\alpha/2}] = \alpha/2$				
(*) Uniformemente más potente. (**) Uniformemente más potente, insesgado.				

8.3 Contratación de hipótesis sobre el coeficiente de correlación lineal de una población normal bidimensional

De la variable aleatoria bidimensional $(\xi; \eta)$ con función de densidad conjunta normal⁸ y coeficiente de correlación poblacional ρ , se toman muestras aleatorias simples de tamaño n , $(x_1; y_1), \dots, (x_n; y_n)$. La expresión del coeficiente de correlación muestral, r , es

$$r = \frac{m_{11}}{s_x s_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

8.3.1. CONTRASTE DE INCORRELACIÓN LINEAL:

$$H_0 [\rho = 0], \quad H_1 [\rho \neq 0]$$

El coeficiente de correlación, poblacional o muestral, toma valores en el intervalo $[-1; 1]$ y de ellos se considera en particular el cero, correlación nula. Este valor implica que las variables aleatorias ξ y η son independientes por ser normales. Bajo esta hipótesis, $\rho = 0$, según lo estudiado en el capítulo 1 la variable

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

sigue en el muestreo una distribución t con $(n - 2)$ grados de libertad.

Para contrastar la hipótesis establecida sobre el coeficiente de correlación poblacional recurrimos a un contraste de significación⁹. El estadístico t sigue en el

⁸ MARTÍN PLIEGO y RUIZ-MAYA: *Fundamentos...* Capítulo 8.

⁹ Véase el Capítulo 7.

muestreo una distribución $t(n - 2)$ por lo cual, fijado el nivel de significación α , hallamos el valor tabular $t_{\alpha/2}$ que verifica

$$P(|t(n - 2)| \geq t_{\alpha/2}) = \alpha$$

calculamos el valor muestral del estadístico t y si, en estas condiciones, se verifica que $|t| \geq t_{\alpha/2}$ rechazamos la hipótesis nula, $\rho = 0$, pues ha tenido lugar un suceso raro¹⁰. En caso contrario, $|t| \leq t_{\alpha/2}$, aceptaríamos esta hipótesis.

8.3.2. CONTRASTE DE ASOCIACIÓN LINEAL: $H_0 [\rho = \rho_0]$, $H_1 [\rho \neq \rho_0]$

Para valores de n moderados la variable aleatoria¹¹

$$Z = \frac{1}{2} \ln \frac{1+r}{1-r}$$

se distribuye aproximadamente

$$N\left(\frac{1}{2} \ln \frac{1+\rho}{1-\rho}; \sqrt{\frac{1}{n-3}}\right).$$

Formamos la nueva variable diferencia

$$\frac{1}{2} \ln \frac{1+r}{1-r} - \frac{1}{2} \ln \frac{1+\rho_0}{1-\rho_0}$$

distribuyéndose

$$N\left(0; \sqrt{\frac{1}{n-3}}\right)$$

por lo cual, la variable

$$v = \frac{\frac{1}{2} \ln \frac{1+r}{1-r} - \frac{1}{2} \ln \frac{1+\rho_0}{1-\rho_0}}{\frac{1}{\sqrt{n-3}}}$$

¹⁰ Ley del azar.

¹¹ Transformación de Fisher, véase capítulo 1.

es aproximadamente $N(0; 1)$. Partiendo de aquí, y con un nivel de significación α , determinamos el valor tabular $z_{\alpha/2}$ tal que

$$P[|N(0; 1)| \geq z_{\alpha/2}] = \alpha$$

si el valor absoluto muestral del estadístico v es mayor que $z_{\alpha/2}$ rechazamos la hipótesis nula, $\rho = \rho_0$, y aceptamos que el coeficiente de correlación poblacional ρ es distinto de ρ_0 .

8.4 Contrastación de hipótesis sobre los coeficientes de regresión de una población normal bidimensional

■ $H_0 [\beta_1 = \beta_1^0], \quad H_1 [\beta_1 \neq \beta_1^0]$.

En el capítulo 1 se ha expuesto que si β_1 es el coeficiente de regresión de (η/ξ) , la variable

$$t = \frac{s_x \sqrt{n-2}}{s_y \sqrt{1-r^2}} (b_1 - \beta_1)$$

se distribuye como una $t(n-2)$, siendo

$$s_x^2 = \frac{\sum_{i=1}^n x_i^2 - n\bar{x}^2}{n}; \quad s_y^2 = \frac{\sum_{i=1}^n y_i^2 - n\bar{y}^2}{n}; \quad s_{xy} = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{n}$$

$$b_1 = \frac{s_{xy}}{s_x^2} = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}$$

Por consiguiente, si comparamos el valor muestral del estadístico t con el tabular, fijado el nivel de significación α , obtenido de

$$P[|t(n-2)|; t_{\alpha/2}] = \alpha$$

tenemos el contraste de significación preciso para decidir por cual de las dos hipótesis optamos.

■ El contraste del coeficiente de regresión de (ξ/η) sigue los mismos pasos sin más que considerar el estadístico

$$t' = \frac{s_y \sqrt{n-2}}{s_x \sqrt{1-r^2}} (b'_1 - \beta'_1).$$

EJEMPLO 7

De una población normal bidimensional $(\xi; \eta)$ se ha tomado una muestra aleatoria simple de tamaño 5, cuyos valores aparecen en las dos primeras columnas del cuadro siguiente. Se desea contrastar, con un nivel de significación del 5%, la hipótesis nula que el coeficiente de regresión de (η/ξ) , β_1 , es igual a -1 , frente a la alternativa que es distinto de este valor.

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
3,2	13	10,24	169	41,6
2,6	12	6,76	144	31,2
5,2	8	27,04	64	41,6
1,5	13	2,25	169	19,5
3,9	11	15,21	121	42,9
16,4	57	61,50	667	176,8

El coeficiente de regresión muestral resulta

$$b_1 = \frac{\sum_{i=1}^5 x_i y_i - 5\bar{x}\bar{y}}{\sum_{i=1}^5 x_i^2 - 5\bar{x}^2} = \frac{176,8 - \frac{16,4 \cdot 57}{5}}{61,5 - \frac{16,4^2}{5}} = -1,318.$$

El estadístico necesario para el contraste es

$$t = \frac{s_x \sqrt{n-2}}{s_y \sqrt{1-r^2}} (b_1 - \beta_1)$$

$$s_x = \sqrt{\frac{\sum_{i=1}^5 x_i^2 - 5\bar{x}^2}{5}} = 1,2416; \quad s_y = \sqrt{\frac{\sum_{i=1}^5 y_i^2 - 5\bar{y}^2}{5}} = 1,8547; \quad r = -0,882.$$

El valor del estadístico

$$t = \frac{1,2416 \sqrt{3}}{1,8547 \sqrt{1 - (-0,882)^2}} [-1,318 - (-1)] = -0,782$$

Se acepta la hipótesis nula, $\beta_1 = -1$, frente a la alternativa, $\beta_1 \neq -1$, pues $|-0,782|$ es menor que $t(3; 0,025) = 3,1824$.

8.5 Contrastes entre dos poblaciones

En dos distribuciones normales, con medias μ_1 y μ_2 desconocidas, contrastamos la hipótesis nula que las medias son iguales, $\mu_1 = \mu_2$, frente a la alternativa que son distintas, $\mu_1 \neq \mu_2$. La contrastación se efectúa bajo dos supuestos: varianzas poblacionales conocidas o desconocidas. Pasamos después a la comparación de las dos varianzas, distinguiendo si las medias poblacionales son conocidas o no.

8.5.1. CONTRASTE DE IGUALDAD DE MEDIAS

8.5.1.1. Varianzas poblacionales conocidas

Sobre las medias de dos poblaciones $N(\mu_1; \sigma_1)$ y $N(\mu_2; \sigma_2)$, con varianzas conocidas y distintas, se establece la hipótesis nula de la igualdad de medias, $H_0[\mu_1 = \mu_2]$, y la alternativa de su diferencia, $H_1[\mu_1 \neq \mu_2]$. Para llevar a cabo el contraste se toman dos muestras aleatorias simples, independientes entre sí, de tamaños n y m ($n + m = N$), $\mathbf{X}(x_1, \dots, x_n)$ e $\mathbf{Y}(y_1, \dots, y_m)$.

Para contrastar las hipótesis recurrimos al test razón de verosimilitud, necesitando para ello obtener los valores de los parámetros μ_1 y μ_2 que hacen máxima la función de verosimilitud en cada espacio paramétrico. El espacio paramétrico total es $\{-\infty < \mu_1 < \infty; -\infty < \mu_2 < \infty\}$, y el definido por la hipótesis nula $\{-\infty < \mu < \infty\}$.

En primer lugar, establecemos la función de verosimilitud conjunta de las dos muestras. La función de verosimilitud de una muestra aleatoria simple es igual al producto de la función de densidad poblacional particularizada para cada elemento muestral. Al ser las dos muestras independientes la función de verosimilitud conjunta es igual al producto de las funciones de verosimilitud de cada muestra.

$$\begin{aligned} L(\mathbf{X}, \mathbf{Y}; \mu_1, \mu_2) &= L(\mathbf{X}; \mu_1) L(\mathbf{Y}; \mu_2) = \\ &= f(x_1; \mu_1) \cdots f(x_n; \mu_1) f(y_1; \mu_2) \cdots f(y_m; \mu_2) = \\ &= \frac{1}{(\sigma_1^2)^{\frac{n}{2}} (\sigma_2^2)^{\frac{m}{2}} (2\pi)^{\frac{N}{2}}} e^{-\frac{1}{2} \left[\frac{\sum_{i=1}^n (x_i - \mu_1)^2}{\sigma_1^2} + \frac{\sum_{i=1}^m (y_i - \mu_2)^2}{\sigma_2^2} \right]} \end{aligned}$$

Calculamos su logaritmo neperiano

$$\begin{aligned} \ln L(\mathbf{X}, \mathbf{Y}; \mu_1, \mu_2) &= -\frac{n}{2} \ln \sigma_1^2 - \frac{m}{2} \ln \sigma_2^2 - \frac{N}{2} \ln (2\pi) - \\ &\quad - \frac{1}{2} \left[\frac{\sum_{i=1}^n (x_i - \mu_1)^2}{\sigma_1^2} + \frac{\sum_{i=1}^m (y_i - \mu_2)^2}{\sigma_2^2} \right]. \end{aligned} \quad [11]$$

Bajo la hipótesis nula, $\mu_1 = \mu_2 = \mu$, el logaritmo de la función de verosimilitud conjunta es

$$\begin{aligned} \ln L(\mathbf{X}, \mathbf{Y}; \mu) &= -\frac{n}{2} \ln \sigma_1^2 - \frac{m}{2} \ln \sigma_2^2 - \frac{N}{2} \ln (2\pi) - \\ &\quad - \frac{1}{2} \left[\frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_1^2} + \frac{\sum_{i=1}^m (y_i - \mu)^2}{\sigma_2^2} \right]. \end{aligned} \quad [12]$$

A continuación hallamos, bajo las condiciones establecidas para cada hipótesis, los valores de los parámetros que hacen máximas las funciones de verosimilitud. Deri-

vamos [11] respecto a μ_1 y μ_2 , obteniendo como solución de las ecuaciones, en el espacio paramétrico total, los estimadores máximo verosímiles de μ_1 y μ_2

$$\hat{\mu}_1 = \bar{x} \quad \text{y} \quad \hat{\mu}_2 = \bar{y}$$

En el caso de la hipótesis nula, $\mu_1 = \mu_2 = \mu$, derivamos [12] respecto a μ e igualamos a cero, llegando a

$$\hat{\mu} = \frac{n\sigma_2^2 \bar{x} + m\sigma_1^2 \bar{y}}{m\sigma_1^2 + n\sigma_2^2}.$$

La media obtenida es la media ponderada de las medias muestrales utilizando como ponderaciones los tamaños de las muestras y las varianzas poblacionales.

En el caso particular de poblaciones con varianzas iguales, $\sigma_1 = \sigma_2 = \sigma^2$, la estimación de la media es

$$\hat{\mu} = \frac{n\bar{x} + m\bar{y}}{m + n}$$

media ponderada de las medias muestrales utilizando como ponderaciones únicamente los tamaños muestrales.

Hallamos los valores que hacen máxima la función de verosimilitud en cada espacio paramétrico, formamos su cociente $\lambda(\mathbf{X}, \mathbf{Y})$, la razón de verosimilitud,

$$\lambda(\mathbf{X}, \mathbf{Y}) = e^{-\frac{1}{2} \left[\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_1^2} + \frac{\sum_{i=1}^m (y_i - \bar{y})^2}{\sigma_2^2} - \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_1^2} - \frac{\sum_{i=1}^m (y_i - \bar{y})^2}{\sigma_2^2} \right]}. \quad [13]$$

Simplificamos¹² el exponente de [13] obteniendo como expresión de la razón de verosimilitud conjunta

$$\lambda(\mathbf{X}, \mathbf{Y}) = e^{-\frac{1}{2} \frac{nm}{m\sigma_1^2 + n\sigma_2^2} (\bar{x} - \bar{y})^2} \leq K_2.$$

Para determinar la región crítica necesitamos conocer la distribución en el muestreo de $\lambda(\mathbf{X}, \mathbf{Y})$ o la de un estadístico derivado de $\lambda(\mathbf{X}, \mathbf{Y})$. Para ello, tomamos logaritmos en la expresión anterior

$$-\frac{1}{2} \frac{nm}{m\sigma_1^2 + n\sigma_2^2} (\bar{x} - \bar{y})^2 \leq \ln K_2$$

¹² Véase RUIZ-MAYA y MARTÍN PLIEGO: *Estadística II: Inferencia*. Capítulo 11.

y operando se llega a que la región crítica es igual a

$$(\bar{x} - \bar{y})^2 \geq K_1$$

o, lo que es equivalente, a

$$|\bar{x} - \bar{y}| \geq K.$$

Conocida la distribución en el muestreo del estadístico $\bar{x} - \bar{y}$ la forma de proceder es la siguiente: si $\bar{x} - \bar{y} \leq -K$ o $\bar{x} - \bar{y} \geq K$ se rechaza la hipótesis nula, $\mu_1 = \mu_2$.

Para hallar K , y que la región crítica quede totalmente determinada, fijamos el nivel de significación α y partiendo de su definición calculamos la constante K que delimita la región crítica

$$\begin{aligned} \alpha &= P(\text{rechazar } H_0 \text{ siendo cierta}) = P(0 \leq \lambda(\mathbf{X}, \mathbf{Y}) \leq K_2/H_0) = P[(\bar{x} - \bar{y})^2 \geq K_1/H_0] = \\ &= P(|\bar{x} - \bar{y}| \geq K/H_0) = P(\bar{x} - \bar{y} \leq -K/H_0) + P(\bar{x} - \bar{y} \geq K/H_0). \end{aligned}$$

La región crítica de dos colas, $|\bar{x} - \bar{y}| \geq K$, es función de la diferencia de las medias muestrales, necesitando obtener su distribución en el muestreo. Si es cierta la hipótesis nula, $\mu_1 = \mu_2 = \mu$, las distribuciones de las medias muestrales son

$$\bar{x} \Rightarrow N\left(\mu; \frac{\sigma_1}{\sqrt{n}}\right); \quad \bar{y} \Rightarrow N\left(\mu; \frac{\sigma_2}{\sqrt{m}}\right)$$

y la diferencia de las dos medias, $\bar{x} - \bar{y}$, se distribuye

$$N\left(0; \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}\right)$$

Definiendo una variable aleatoria ξ , $N(0; 1)$,

$$\begin{aligned} P(\bar{x} - \bar{y} \leq -K/H_0) &= P\left[\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} \xi \leq -K\right] = \\ &= P\left[\xi \leq -\frac{K}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}\right] = \frac{\alpha}{2} \end{aligned}$$

y, análogamente,

$$P(\bar{x} - \bar{y} \geq K/H_0) = P\left[\xi \geq \frac{K}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}\right] = \frac{\alpha}{2}.$$

Si llamamos $z_{\alpha/2}$ al valor tabular de la distribución $N(0; 1)$ que deja a su derecha una probabilidad igual a $\alpha/2$, el valor de la constante K delimitador de la región crítica es

$$-z_{\alpha/2} = -\frac{K}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}; \quad K = z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$$

resultando

$$|\bar{x} - \bar{y}| \geq z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$$

Cuando las dos poblaciones tienen varianzas iguales, $\sigma_1^2 = \sigma_2^2 = \sigma^2$, la región crítica se reduce a

$$|\bar{x} - \bar{y}| \geq z_{\alpha/2} \sigma \sqrt{\frac{n+m}{nm}}$$

EJEMPLO 9

Para contrastar la hipótesis nula de igualdad de medias de dos poblaciones $N(\mu_1; 5)$ y $N(\mu_2; 5,5)$ se han tomado dos muestras aleatorias simples e independientes de tamaños 6 y 5, respectivamente,

Muestra 1: x_i	Muestra 2: y_i
21,3	21,0
11,8	10,4
9,5	18,8
9,8	26,5
6,9	10,1
13,1	

Determinése qué decisión se toma si el nivel de significación se fija en $\alpha = 0,0472$.

La región crítica es de la forma $|\bar{x} - \bar{y}| \geq K$, distribuyéndose la variable $\bar{x} - \bar{y}$

$$N\left(0; \sqrt{\frac{25}{6} + \frac{30,25}{5}}\right) = N(0; 3,196).$$

El valor de la constante K es

$$K = z_{\frac{0,0472}{2}} \sqrt{\frac{25}{6} + \frac{30,25}{5}} = 3,196 z_{0,0236}$$

y hallamos el valor de $z_{0,0236}$ que verifica $P(N(0; 1) \geq z_{0,0236}) = 0,0236$, siendo $z_{0,0236} = 1,985$, con lo que la constante K es igual a $3,196 \cdot 1,985 = 6,344$.

La región crítica resulta $|\bar{x} - \bar{y}| \geq 6,344$. Los valores de las medias muestrales son $\bar{x} = 12,07$ e $\bar{y} = 17,36$ y el valor absoluto de la diferencia $|\bar{x} - \bar{y}| = 5,29$ que, por ser menor que $6,344$, indica su no pertenencia a la región crítica, por lo cual no rechazamos la hipótesis que las dos poblaciones tengan el mismo valor medio, $\mu_1 \neq \mu_2$.

La hipótesis nula propuesta, $\mu_1 = \mu_2$, se puede expresar como $\mu_1 - \mu_2 = 0$, dando pie a una generalización: la diferencia de medias igual a una cantidad conocida, $\mu_1 - \mu_2 = \tau$, teniendo como hipótesis alternativa que $\mu_1 - \mu_2 \neq \tau$. Relacionamos los contrastes fundamentales que en este caso pueden hacerse

Hipótesis nula	Hipótesis alternativa	Región crítica
$\mu_1 - \mu_2 \leq \tau$	$\mu_1 - \mu_2 > \tau$	$\bar{x} - \bar{y} \geq \tau + z_\alpha \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$
$\mu_1 - \mu_2 \geq \tau$	$\mu_1 - \mu_2 < \tau$	$\bar{x} - \bar{y} \leq \tau - z_\alpha \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$
$\mu_1 - \mu_2 = \tau$	$\mu_1 - \mu_2 \neq \tau$	$ \bar{x} - \bar{y} - \tau \geq z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$

8.5.1.2. Varianza poblacional común desconocida

Sean dos poblaciones $N(\mu_1; \sigma)$ y $N(\mu_2; \sigma)$ cuya varianza común es desconocida. Las hipótesis a contrastar son $H_0 [\mu_1 = \mu_2]$ y $H_1 [\mu_1 \neq \mu_2]$, y para ello se parte de dos muestras aleatorias simples e independientes de tamaños n y m , $n + m = N$, $\mathbf{X} (x_1, \dots, x_n)$ y $\mathbf{Y} (y_1, \dots, y_m)$.

La función de verosimilitud conjunta es igual a

$$L(\mathbf{X}, \mathbf{Y}; \mu_1, \mu_2, \sigma^2) = \frac{1}{(\sigma^2)^{\frac{N}{2}} (2\pi)^{\frac{N}{2}}} e^{-\frac{1}{2\sigma^2} \left[\sum_{i=1}^n (x_i - \mu_1)^2 + \sum_{i=1}^m (y_i - \mu_2)^2 \right]}$$

y su logaritmo

$$\ln L(\mathbf{X}, \mathbf{Y}; \mu_1, \mu_2, \sigma^2) = -\frac{N}{2} \ln(\sigma^2) - \frac{N}{2} \ln(2\pi) - \frac{1}{2\sigma^2} \left[\sum_{i=1}^n (x_i - \mu_1)^2 + \sum_{i=1}^m (y_i - \mu_2)^2 \right].$$

Calculamos los estimadores máximo-verosímiles de μ_1 , μ_2 y σ^2 en el espacio paramétrico total $\{ -\infty < \mu_1 < \infty; -\infty < \mu_2 < \infty; \sigma^2 > 0 \}$

$$\frac{\partial \ln L(\mathbf{X}, \mathbf{Y}; \mu_1, \mu_2, \sigma^2)}{\partial \mu_1} = \frac{\sum_{i=1}^n (x_i - \mu_1)}{\sigma^2} = 0 \quad [14]$$

$$\frac{\partial \ln L(\mathbf{X}, \mathbf{Y}; \mu_1, \mu_2, \sigma^2)}{\partial \mu_2} = \frac{\sum_{i=1}^m (y_i - \mu_2)}{\sigma^2} = 0 \quad [15]$$

$$\frac{\partial \ln L(\mathbf{X}, \mathbf{Y}; \mu_1, \mu_2, \sigma^2)}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} \left[\sum_{i=1}^n (x_i - \mu_1)^2 + \sum_{i=1}^m (y_i - \mu_2)^2 \right] = 0 \quad [16]$$

De las ecuaciones [14] y [15] obtenemos los estimadores máximo-verosímiles de μ_1 y μ_2

$$\hat{\mu}_1 = \bar{x}; \quad \hat{\mu}_2 = \bar{y}$$

y de la ecuación [16] el estimador, s^2 , de la varianza común desconocida σ^2 ,

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^m (y_i - \bar{y})^2}{N}.$$

Pasando al espacio paramétrico definido por la hipótesis nula, $\mu_1 = \mu_2 = \mu$, $\{ -\infty < \mu < \infty; \sigma^2 > 0 \}$, la función de verosimilitud conjunta es

$$L(\mathbf{X}, \mathbf{Y}; \mu, \sigma^2) = \frac{1}{(\sigma^2)^{\frac{N}{2}} (2\pi)^{\frac{N}{2}}} e^{-\frac{1}{2\sigma^2} \left[\sum_{i=1}^n (x_i - \mu)^2 + \sum_{i=1}^m (y_i - \mu)^2 \right]}$$

y su logaritmo

$$\ln L(\mathbf{X}, \mathbf{Y}; \mu, \sigma^2) = -\frac{N}{2} \ln(\sigma^2) - \frac{N}{2} \ln(2\pi) - \frac{1}{2\sigma^2} \left[\sum_{i=1}^n (x_i - \mu)^2 + \sum_{i=1}^m (y_i - \mu)^2 \right].$$

Derivando respecto a los dos parámetros μ y σ^2 e igualando a cero las derivadas, obtenemos el estimador de la media común μ

$$\hat{\mu} = \frac{\sum_{i=1}^n x_i + \sum_{i=1}^m y_i}{N} = \frac{n\bar{x} + m\bar{y}}{N}$$

y el de la varianza σ^2 , que denominamos s_0^2 ,

$$s_0^2 = \frac{\sum_{i=1}^n (x_i - \hat{\mu})^2 + \sum_{i=1}^m (y_i - \hat{\mu})^2}{N}.$$

El cociente de la función de verosimilitud conjunta, particularizada para los valores que la hacen máxima en cada espacio paramétrico, es

$$\lambda(\mathbf{X}, \mathbf{Y}) = \frac{\frac{1}{(s_0^2)^{\frac{N}{2}} (2\pi)^{\frac{N}{2}}} e^{-\frac{1}{2s_0^2} \left[\sum_{i=1}^n (x_i - \hat{\mu})^2 + \sum_{i=1}^m (y_i - \hat{\mu})^2 \right]}}{\frac{1}{(s^2)^{\frac{N}{2}} (2\pi)^{\frac{N}{2}}} e^{-\frac{1}{2s^2} \left[\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^m (y_i - \bar{y})^2 \right]}} \leq K_1$$

y haciendo operaciones¹¹ llegamos a la variable aleatoria

$$t = \frac{\frac{n m}{n + m} (\bar{x} - \bar{y})^2}{\frac{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^m (y_i - \bar{y})^2}{n + m - 2}}$$

¹¹ Véase RUIZ-MAYA y MARTÍN PLIEGO: *Estadística II: Inferencia*. Capítulo 11.

variable aleatoria distribuida en el muestreo como una $t(n + m - 2)$, dando lugar a la región crítica

$$\left| \frac{\sqrt{\frac{nm}{n+m}}(\bar{x} - \bar{y})}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^m (y_i - \bar{y})^2}{n+m-2}}} \right| \geq K$$

El valor de K verifica $t(n + m - 2; \alpha/2) = K$, siendo α el nivel de significación.

EJEMPLO 10

Se dispone de dos muestras aleatorias simples e independientes de tamaños 9 y 7, procedentes de dos poblaciones $N(\mu_1; \sigma)$ y $N(\mu_2; \sigma)$

Muestra 1 : x_i	181	145	150	123	156	160	115	150	134
Muestra 2 : y_i	121	144	123	116	156	142	101		

Se contrasta, con un nivel de significación del 10%, la hipótesis de igualdad de medias, $\mu_1 = \mu_2$.

El estadístico que permite el contraste es

$$t = \left| \frac{\sqrt{\frac{9 \cdot 7}{9+7}}(146 - 129)}{\sqrt{\frac{3188 + 2176}{14}}} \right| = 1,7234$$

el valor tabular $t_{0,05}$ tal que $P(t(14) \geq t_{0,05}) = 0,05$ es 1,7613 que al ser mayor que el muestral lleva a no rechazar la hipótesis nula, es decir, a la vista de la evidencia proporcionada por las muestras éstas pueden proceder de poblaciones con igual media.

De la misma forma que en el caso de las varianzas poblacionales conocidas, los contrastes generalizados son

Hipótesis nula	Hipótesis alternativa	Región crítica
$\mu_1 - \mu_2 \leq \tau$	$\mu_1 - \mu_2 > \tau$	$\bar{x} - \bar{y} \geq \tau + st(n + m - 2; \alpha)$
$\mu_1 - \mu_2 \geq \tau$	$\mu_1 - \mu_2 < \tau$	$\bar{x} - \bar{y} \leq \tau - st(n + m - 2; \alpha)$
$\mu_1 - \mu_2 = \tau$	$\mu_1 - \mu_2 \neq \tau$	$ \bar{x} - \bar{y} - \tau \geq st(n + m - 2; \alpha/2)$

donde

$$s^2 = \frac{\frac{n+m}{nm} \left[\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^m (y_i - \bar{y})^2 \right]}{n+m-2}$$

8.5.2. CONTRASTE DE IGUALDAD DE VARIANZAS

8.5.2.1. Medias poblacionales conocidas

En dos poblaciones $N(\mu_1; \sigma_1)$ y $N(\mu_2; \sigma_2)$, con medias conocidas, las hipótesis a contrastar son $H_0 [\sigma_1^2 = \sigma_2^2]$ y $H_1 [\sigma_1^2 \neq \sigma_2^2]$. Se toman dos muestras aleatorias simples e independientes entre sí de tamaños n y m , ($n + m = N$) $\mathbf{X} (x_1, \dots, x_n)$ y $\mathbf{Y} (y_1, \dots, y_m)$.

La función de verosimilitud conjunta es igual a

$$L(\mathbf{X}, \mathbf{Y}; \sigma_1^2, \sigma_2^2) = \frac{1}{(\sigma_1^2)^{\frac{n}{2}} (\sigma_2^2)^{\frac{m}{2}} (2\pi)^{\frac{N}{2}}} e^{-\frac{1}{2} \left[\frac{\sum_{i=1}^n (x_i - \mu_1)^2}{\sigma_1^2} + \frac{\sum_{i=1}^m (y_i - \mu_2)^2}{\sigma_2^2} \right]}$$

y su logaritmo

$$\ln L(\mathbf{X}, \mathbf{Y}; \sigma_1^2, \sigma_2^2) = -\frac{n}{2} \ln \sigma_1^2 - \frac{m}{2} \ln \sigma_2^2 - \frac{N}{2} \ln (2\pi) - \frac{1}{2} \left[\frac{\sum_{i=1}^n (x_i - \mu_1)^2}{\sigma_1^2} + \frac{\sum_{i=1}^m (y_i - \mu_2)^2}{\sigma_2^2} \right]$$

Derivamos la función $\ln L(\mathbf{X}, \mathbf{Y}; \sigma_1^2, \sigma_2^2)$ respecto a σ_1^2 y σ_2^2 para obtener el máximo en el espacio paramétrico total $\{\sigma^2 > 0\}$ e igualamos a cero

$$\left. \begin{aligned} \frac{\partial \ln L(\mathbf{X}, \mathbf{Y}; \sigma_1^2, \sigma_2^2)}{\partial \sigma_1^2} &= -\frac{n}{2\sigma_1^2} + \frac{1}{2\sigma_1^4} \sum_{i=1}^n (x_i - \mu_1)^2 = 0 \\ \frac{\partial \ln L(\mathbf{X}, \mathbf{Y}; \sigma_1^2, \sigma_2^2)}{\partial \sigma_2^2} &= -\frac{m}{2\sigma_2^2} + \frac{1}{2\sigma_2^4} \sum_{i=1}^m (y_i - \mu_2)^2 = 0 \end{aligned} \right\}$$

y los estimadores máximo-verosímiles de las varianzas σ_1^2 y σ_2^2 , solución del sistema precedente, son

$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \mu_1)^2}{n} \quad \text{y} \quad s_y^2 = \frac{\sum_{i=1}^m (y_i - \mu_2)^2}{m}$$

Si la hipótesis nula es cierta, $\sigma_1^2 = \sigma_2^2 = \sigma^2$, la función de verosimilitud conjunta queda

$$L(\mathbf{X}, \mathbf{Y}; \sigma^2) = \frac{1}{(\sigma^2)^{\frac{N}{2}} (2\pi)^{\frac{N}{2}}} e^{-\frac{1}{2\sigma^2} \left[\sum_{i=1}^n (x_i - \mu_1)^2 + \sum_{i=1}^m (y_i - \mu_2)^2 \right]}$$

tomando logaritmos, derivando respecto a σ^2 e igualando a cero la derivada, llegamos al estimador máximo-verosímil, s^2 , de la varianza común

$$s^2 = \frac{\sum_{i=1}^n (x_i - \mu_1)^2 + \sum_{i=1}^m (y_i - \mu_2)^2}{n + m} = \frac{ns_x^2 + ms_y^2}{N}$$

El cociente de la función de verosimilitud conjunta, $\lambda(\mathbf{X}, \mathbf{Y})$, particularizada para valores de σ_1^2, σ_2^2 y σ^2 que la hacen máxima es

$$\lambda(\mathbf{X}, \mathbf{Y}) = \frac{\frac{1}{(\sigma^2)^{\frac{N}{2}} (2\pi)^{\frac{N}{2}}} e^{-\frac{1}{2s^2} \left[\sum_{i=1}^n (x_i - \mu_1)^2 + \sum_{i=1}^m (y_i - \mu_2)^2 \right]}}{\frac{1}{(s_x^2)^{\frac{n}{2}} (s_y^2)^{\frac{m}{2}} (2\pi)^{\frac{N}{2}}} e^{-\frac{1}{2} \left[\frac{\sum_{i=1}^n (x_i - \mu_1)^2}{s_x^2} + \frac{\sum_{i=1}^m (y_i - \mu_2)^2}{s_y^2} \right]}}$$

sustituyendo en la expresión los sumatorios por sus expresiones en función de s_x^2, s_y^2 y s^2 tenemos

$$\lambda(\mathbf{X}, \mathbf{Y}) = N^{\frac{N}{2}} \frac{(s_x^2)^{\frac{n}{2}} (s_y^2)^{\frac{m}{2}}}{(ns_x^2 + ms_y^2)^{\frac{N}{2}}} e^{-\frac{1}{2} \frac{[ns_x^2 + ms_y^2]}{N}} e^{-\frac{1}{2} \left[\frac{ns_x^2}{s_x^2} + \frac{ms_y^2}{s_y^2} \right]} =$$

y operando¹²

$$= \frac{N^{\frac{N}{2}}}{n^{\frac{n}{2}} m^{\frac{m}{2}}} \left[\frac{\frac{\sum_{i=1}^n (x_i - \mu_1)^2}{\sum_{i=1}^m (y_i - \mu_2)^2}}{\frac{\sum_{i=1}^n (x_i - \mu_1)^2}{\sum_{i=1}^m (y_i - \mu_2)^2} + 1} \right]^{\frac{n}{2}} \left[\frac{1}{\frac{\sum_{i=1}^n (x_i - \mu_1)^2}{\sum_{i=1}^m (y_i - \mu_2)^2} + 1} \right]^{\frac{m}{2}} \quad [17]$$

Consideremos las variables

$$\frac{\sum_{i=1}^n (x_i - \mu_1)^2}{\sigma_1^2} \quad \text{y} \quad \frac{\sum_{i=1}^m (y_i - \mu_2)^2}{\sigma_2^2}$$

distribuidas, respectivamente, como $\chi^2(n)$ y $\chi^2(m)$ e independientes por serlo las dos muestras. La variable

$$\frac{\frac{\chi^2(n)}{n}}{\frac{\chi^2(m)}{m}}$$

se distribuye como una F de Fisher-Snedecor con n y m grados de libertad, $F(n; m)$, variable que podemos expresar como

$$\frac{\frac{\sum_{i=1}^n (x_i - \mu_1)^2}{n\sigma^2}}{\frac{\sum_{i=1}^m (y_i - \mu_2)^2}{m\sigma^2}} = \frac{m}{n} \frac{\sum_{i=1}^n (x_i - \mu_1)^2}{\sum_{i=1}^m (y_i - \mu_2)^2} \Rightarrow F(n; m)$$

si la hipótesis es nula es cierta, es decir si

$$\sigma_1^2 = \sigma_2^2 = \sigma^2$$

¹² Véase RUIZ-MAYA y MARTÍN PLIEGO: *Estadística II: Inferencia*. Capítulo 11.

por lo que

$$\frac{\sum_{i=1}^n (x_i - \mu_1)^2}{\sum_{i=1}^m (y_i - \mu_2)^2} = \frac{n}{m} F(n; m)$$

y sustituyendo este cociente de sumatorios en la expresión [17], llegamos a la razón de verosimilitudes

$$\lambda(\mathbf{X}, \mathbf{Y}) = \frac{N^{\frac{N}{2}}}{n^{\frac{n}{2}} m^{\frac{m}{2}}} \frac{\left[\frac{n}{m} F(n; m) \right]^{\frac{n}{2}}}{\left[\frac{n}{m} F(n; m) + 1 \right]^{\frac{N}{2}}}$$

La variable $\lambda(\mathbf{X}, \mathbf{Y})$ tiene una distribución en el muestreo complicada por lo que la obtención de la región crítica de manera directa lo es también. Sin embargo, el que la razón de verosimilitud sea función de la variable $F(n; m)$ permite obtener la región crítica de forma sencilla.

Según el contraste razón de verosimilitud la región crítica es

$$0 \leq \lambda(\mathbf{X}, \mathbf{Y}) \leq K$$

y como $\lambda(\mathbf{X}, \mathbf{Y})$ resulta función de la distribución $F(n; m)$ se verifica, como indica la figura 8.6, la equivalencia de los sucesos

$$\{0 \leq \lambda(\mathbf{X}, \mathbf{Y}) \leq K\} \Rightarrow \{0 \leq F(n; m) \leq L_1\} \cup \{L_2 \leq F(n; m) < \infty\}$$

siendo disjuntos los intervalos del segundo miembro. Tomando probabilidades, como $\{0 \leq \lambda(\mathbf{X}, \mathbf{Y}) \leq K\}$ es la región crítica,

$$\begin{aligned} P(0 \leq \lambda(\mathbf{X}, \mathbf{Y}) \leq K) &= \alpha = P(0 \leq F(n; m) \leq L_1) + P(L_2 \leq F(n; m) < \infty) = \\ &= P[F(n; m) \leq L_1] + P[L_2 \leq F(n; m)] \end{aligned}$$

y, por convenio, se consideran como valores de L_1 y L_2 los que verifican

$$\begin{aligned} P[F(n; m) \leq L_1] &= \frac{\alpha}{2} \\ P[L_2 \leq F(n; m)] &= \frac{\alpha}{2} \end{aligned}$$

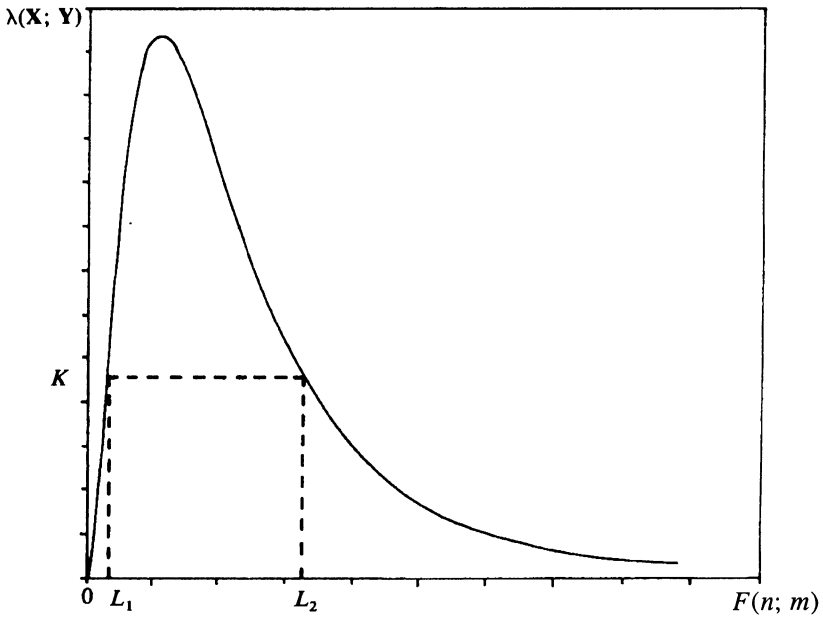


FIGURA 8.6

es decir,

$$L_1 = F\left(n; m; 1 - \frac{\alpha}{2}\right) = \frac{1}{F\left(m; n; \frac{\alpha}{2}\right)}$$

$$L_2 = F\left(n; m; \frac{\alpha}{2}\right)$$

y el valor muestral del estadístico que permite la contrastación es

$$F = \frac{m \sum_{i=1}^n (x_i - \mu_1)^2}{n \sum_{i=1}^m (y_i - \mu_2)^2}$$

EJEMPLO 11

De una población $N(-2; \sigma_1)$ se ha tomado una muestra de tamaño 3: $(-6; -1; -3)$ y de otra, $N(4; \sigma_2)$ una de tamaño 4: $(1; 7; 4; 5)$, ambas independientes. Con un nivel de significación del 10% contrástese la hipótesis de igualdad de varianzas.

Para conocer el valor muestral del estadístico F necesitamos los de los sumatorios $\sum_{i=1}^3 (x_i - \mu_1)^2$ y $\sum_{i=1}^4 (y_i - \mu_2)^2$, cuyos valores son, respectivamente, 18 y 19, y el estadístico F resulta igual a

$$F = \frac{4 \cdot 18}{3 \cdot 19} = 1,263$$

La región crítica es

$$\left\{ 0 \leq F \leq \frac{1}{F(3, 4; 0,05)} \right\} \cup \{F(3, 4; 0,05) < F\} = \{0 < F < 0,152\} \cup \{6,59 < F\}$$

como $F = 1,263$ no está comprendido en ninguno de los dos intervalos, pertenece a la región de aceptación y no rechazamos la hipótesis de igualdad de varianzas, al nivel de significación fijado.

8.5.2.2. Medias poblacionales desconocidas

Sean dos poblaciones $N(\mu_1; \sigma_1)$ y $N(\mu_2; \sigma_2)$, con medias desconocidas. La hipótesis nula establecida sobre las varianzas es $H_0 [\sigma_1^2 = \sigma_2^2]$ y la alternativa $H_1 [\sigma_1^2 \neq \sigma_2^2]$. Los espacios paramétricos son: el total $\{-\infty < \mu_i < \infty; \sigma_i^2 > 0\}$ y el definido por la hipótesis nula $\{-\infty < \mu_i < \infty; \sigma^2 > 0\}$. Para efectuar el contraste, se toman dos muestras aleatorias simples e independientes entre sí de tamaños n y m , ($n + m = N$) $\mathbf{X}(x_1, \dots, x_n)$ e $\mathbf{Y}(y_1, \dots, y_m)$. Siguiendo un procedimiento análogo al anterior llegamos¹⁵ a que el estadístico

$$F = \frac{m-1}{n-1} \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^m (y_i - \bar{y})^2}$$

¹⁵ Véase RUIZ-MAYA y MARTÍN PLIEGO: *Estadística II: Inferencia*. Capítulo 11.

se distribuye como una $F(n-1; m-1)$. De la misma forma que cuando conocemos las medias poblacionales, la región crítica está formada (Fig. 8.6) por la unión de los intervalos disjuntos

$$\{0 \leq F(n-1; m-1) \leq L_1\}, \quad \{L_2 \leq F(n-1; m-1) < \infty\}$$

obteniéndose los valores de L_1 y L_2 como solución de las ecuaciones

$$P[F(n-1; m-1) \leq L_1] = \frac{\alpha}{2}$$

$$P[L_2 \leq F(n-1; m-1)] = \frac{\alpha}{2}$$

es decir,

$$L_1 = F\left(n-1; m-1; 1 - \frac{\alpha}{2}\right) = \frac{1}{F\left(m-1; n-1; \frac{\alpha}{2}\right)}$$

$$L_2 = F\left(n-1; m-1; \frac{\alpha}{2}\right)$$

y como valor muestral del estadístico que permite la contrastación

$$F = \frac{m-1}{n-1} \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^m (y_i - \bar{y})^2}.$$

8.6 Contrastación de hipótesis en muestras grandes

8.6.1. PLANTEAMIENTO GENERAL

Todos los casos estudiados en este capítulo se refieren a hipótesis sobre parámetros de distribuciones normales: media, varianza, coeficientes de correlación y regresión. El conocimiento del tipo de distribución posibilita en todos ellos la obtención de la región crítica. En lo que sigue tomamos en consideración hipótesis sobre parámetros de poblaciones no normales cuyos estimadores se construyen con

muestras de tamaño grande que permiten llegar en sus distribuciones en el muestreo a la situación asintótica normal.

Sea un parámetro θ de una población no normal, cuyo estimador máximo-verosímil $\hat{\theta}$ se basa en una muestra aleatoria simple de tamaño n y $V(\hat{\theta})$ su varianza. Se puede demostrar que la variable

$$\frac{\hat{\theta} - \theta}{\sqrt{V(\hat{\theta})}}$$

converge en distribución a la $N(0; 1)$ por las propiedades asintóticas de estos estimadores.

Tenemos casos particulares que conducen directamente a la posición asintótica normal: cuando en el cociente de verosimilitudes se llega a la media muestral que, por el Teorema Central del Límite, también es asintóticamente normal.

8.6.2. CONTRASTE DE PROPORCIONES

Como caso particular de población no normal estudiaremos, a continuación, la contrastación de algunas hipótesis que pueden hacerse sobre el parámetro de la distribución $B(1; p)$.

■ $H_0 [p = p_0], \quad H_1 [p = p_1]$

Para llevar a cabo este contraste aplicamos el lema de Neyman-Pearson, pues las dos hipótesis son simples, en muestras aleatorias simples de tamaño n , cuyos valores x_i son 0 o 1.

Formamos el cociente de verosimilitudes

$$\begin{aligned} \frac{L(\mathbf{X}; p_0)}{L(\mathbf{X}; p_1)} &= \frac{p_0^{x_1} (1 - p_0)^{1-x_1} \cdots p_0^{x_n} (1 - p_0)^{1-x_n}}{p_1^{x_1} (1 - p_1)^{1-x_1} \cdots p_1^{x_n} (1 - p_1)^{1-x_n}} \\ &= \left(\frac{p_0}{p_1} \right)^{\sum_{i=1}^n x_i} \left(\frac{1 - p_0}{1 - p_1} \right)^{n - \sum_{i=1}^n x_i} = \\ &= \left(\frac{1 - p_0}{1 - p_1} \right)^n \left(\frac{p_0 (1 - p_1)}{p_1 (1 - p_0)} \right)^{\sum_{i=1}^n x_i} \leq K_3 \end{aligned}$$

$$\left[\frac{p_0(1-p_1)}{p_1(1-p_0)} \right]^{\sum_{i=1}^n x_i} \leq K_2. \tag{18}$$

Si la expresión

$$\frac{p_0(1-p_1)}{p_1(1-p_0)} \tag{19}$$

es mayor que 1, equivalente a que $p_0(1-p_1) > p_1(1-p_0)$ y de aquí se deduce que $p_0 > p_1$, el logaritmo es positivo y al despejar $\sum_{i=1}^n x_i$ no cambia el signo de la desigualdad [18], siendo la región crítica

$$\sum_{i=1}^n x_i \leq K_1$$

o al dividir por el tamaño de la muestra, n , $\bar{x} \leq K$.

La determinación de la contante K se efectúa por el procedimiento habitual, fijado el nivel de significación α ,

$$\alpha = P(\bar{x} \leq K/H_0)$$

necesitando, en este punto, conocer la distribución en el muestreo de \bar{x} . Si el tamaño muestral es elevado, por el teorema de Moivre-Laplace, la media muestral se distribuye en la hipótesis nula aproximadamente,

$$N\left(p_0; \sqrt{\frac{p_0(1-p_0)}{n}}\right)$$

por lo cual podemos pasar de $P[\bar{x} \leq K/H_0]$ a

$$P\left[\bar{x} \leq K / N\left(p_0; \sqrt{\frac{p_0(1-p_0)}{n}}\right)\right]$$

y tenemos el problema resuelto, región crítica definida, así como la potencia del contraste.

Si, por el contrario, la expresión [19] es menor que la unidad (p_0 será menor que p_1) su logaritmo negativo y la región crítica pasa a ser $\bar{x} \geq K$, pudiéndose aplicar todo lo expuesto referente al cálculo de la constante K y de la potencia del contraste.

EJEMPLO 8

En un proceso de fabricación se detecta que el 17% de las piezas son defectuosas. Mediante una puesta a punto se espera bajar esta cifra al 10%. Contrastar estas hipótesis con un nivel de significación del 1%, en una muestra aleatoria simple de tamaño 100, en la que se han obtenido seis piezas defectuosas.

La variable aleatoria estado de la pieza es binominal (pieza defectuosa, 1, pieza en buen estado, 0) con probabilidad bajo la hipótesis nula $p_0 = 0,17$ y en la alternativa $p_1 = 0,10$.

En estas condiciones y como $p_0 > p_1$, ($0,17 > 0,10$), la región crítica es $\bar{x} \leq K$. Para hallar K sabemos que

$$\alpha = 0,001 = P[\bar{x} \leq K/B(1; 0,17)] \approx P[\bar{x} \leq K/N(0,17; 0,034563)]$$

pues

$$\sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0,17 \cdot 0,83}{100}} = 0,037563$$

y si la variable ξ es $N(0; 1)$

$$\begin{aligned} 0,001 &\approx P(0,17 + 0,037563 \xi \leq K) = \\ &= P\left(\xi \leq \frac{K - 0,17}{0,037563}\right) = \\ &= P(\xi \leq z_{0,001}) \end{aligned}$$

el valor tabular correspondiente a la probabilidad 0,001 es $z_{0,001} = -3,08$, por lo que $(K - 0,17)/0,037563 = -3,08$, y de aquí obtenemos $K = 0,0543$, y la región crítica

$$\bar{x} \leq 0,0543.$$

En la muestra han aparecido seis piezas defectuosas, por lo que $\bar{x} = \frac{6}{100} = 0,06$ cantidad mayor que 0,0543, es decir, pertenece a la región de aceptación ($\bar{x} \geq 0,0543$) y se acepta que el porcentaje de piezas defectuosas sigue siendo igual a 0,17: la puesta a punto del proceso de fabricación no ha dado lugar a un descenso significativo del promedio de piezas defectuosas.

La potencia del contraste es igual a

$$\begin{aligned} 1 - \beta &= P(\bar{x} \leq 0,0543/H_1) = \\ &= P(\bar{x} \leq 0,0543/B(1; 0,10)) \approx \\ &\approx P(\bar{x} \leq 0,0543/N(0,10; 0,03)) \end{aligned}$$

la desviación típica de la distribución normal bajo la hipótesis alternativa es

$$\sqrt{\frac{p_1(1-p_1)}{n}} = \sqrt{\frac{0,1 \cdot 0,9}{100}} = 0,03$$

y la potencia

$$P(0,1 + 0,03\xi \leq 0,0543) = 0,0638.$$

Ejercicios resueltos

EJERCICIO 8.1 En una población $N(\mu; 2)$ se establecen dos hipótesis: $H_0 [\mu_0 = 0]$ y $H_1 [\mu_1 = 1]$; el tamaño de la muestra aleatoria simple es $n = 100$ y la potencia del contraste $1 - \beta = 0,90$. Calcúlese la constante de la mejor región crítica, $\bar{x} \geq K$, y el nivel de significación, α .

SOLUCIÓN. Tenemos que para este contraste la región crítica es $\bar{x} \leq K$, luego

$$1 - \beta = P(\xi \geq z_{0,90}) = 0,90; \quad z_{0,90} = -1,282$$

$$K = \mu_1 + \frac{\sigma}{\sqrt{n}} z_{1-\beta} = 1 + \frac{2}{10} (-1,282) = 0,7436$$

la mejor región crítica es $\bar{x} \geq 0,7436$, y el nivel de significación

$$\alpha = P(\bar{x} \geq K/H_0) = P[N(0; 2) \geq 0,7436] = 0,3557.$$

EJERCICIO 8.2 Si en la población $N(\mu; 2)$ las hipótesis son $H_0 [\mu = 6]$ y $H_1 [\mu = 4]$, hállese la mejor región crítica y la potencia del contraste si el nivel de significación es igual a 0,05 y la muestra aleatoria simple es de tamaño 4.

SOLUCIÓN. Tenemos que

$$\alpha = P[\bar{x} \leq K/H_0 (\mu = 6)] = P[N(6; 1) \leq K] = 0,05$$

y $K = 4,355$, siendo la mejor región crítica el intervalo $\{\bar{x} \leq 4,355\}$.

La potencia del contraste

$$\begin{aligned} 1 - \beta &= P[\bar{x} \leq 4,355/H_1 (\mu = 4)] = P[N(4; 1) \leq 4,355] = \\ &= P[N(0; 1) \leq 0,355] = 0,6387. \end{aligned}$$

EJERCICIO 8.3 En la distribución $N(\mu; 12)$, contrástese la hipótesis nula $\mu = -5$ frente a la alternativa $\mu < -5$, en muestras aleatorias simples de tamaño 9 y con un nivel de significación $\alpha = 0,15$, siendo la muestra extraída

-20,06	6,05	0,63
4,56	3,17	-15,26
-17,20	-0,28	-3,16

SOLUCIÓN. La mejor región crítica es de la forma $\bar{x} \leq K$. Bajo la hipótesis nula, la media muestral se distribuye $N(-5; 4)$, y como $\alpha = 0,15$

$$0,15 = P[\bar{x} \leq K/N(-5; 4)] = P[N(-5; 4) \leq K]$$

y el valor de K que verifica la ecuación es $K = -9,148$.

La media de la muestra obtenida es igual a $-4,617$, valor no incluido en la región crítica, $-4,617 > -9,148$, aceptándose por consiguiente la hipótesis nula, es decir, se admite que la verdadera media poblacional es $\mu = -5$.

La función de potencia es igual a

$$P(\mu) = P[\bar{x} \leq -9,148/N(\mu; 4)] = P[N(\mu; 4) \leq -9,148].$$

EJERCICIO 8.4 En la distribución $N(-3; \sigma)$ contrástese la hipótesis $H_0[\sigma^2 = 10]$ frente a la alternativa $H_1[\sigma^2 = 5]$, con un nivel de significación del 7,5% y en una muestra aleatoria simple de tamaño 6 en la que el valor muestral de $\sum_{i=1}^6 (x_i + 3)^2$ es 7,42.

SOLUCIÓN. La mejor región crítica es $\sum_{i=1}^6 (x_i + 3)^2 \leq K$, y obtenemos el valor de K de la ecuación

$$\alpha = 0,075 = P\left[\chi^2(6) = \frac{K}{10}\right]$$

el valor tabular de $\frac{K}{10}$ es 1,9195 por lo cual, $K = 19,195$ y la mejor región crítica queda

$$\sum_{i=1}^6 (x_i + 3)^2 \leq 19,195$$

y se rechaza la hipótesis nula, $N(-3; \sqrt{10})$, aceptándose la alternativa, $N(3; \sqrt{5})$, pues el valor muestral (7,42) es menor que el crítico (19,195).

EJERCICIO 8.5 En la distribución $N(200; \sigma)$ se contrasta la hipótesis nula $H_0[\sigma^2 = 25]$ frente a la alternativa $H_1[\sigma^2 > 25]$ con un nivel de significación $\alpha = 0,017$, en muestras aleatorias simples de tamaño 15. Determínese cuál de las dos hipótesis se acepta si extraída una muestra resultó en ella

$$\sum_{i=1}^{15} (x_i - 20)^2 = 874,2.$$

SOLUCIÓN. Para que la región crítica quede determinada calculamos el valor de la constante K

$$\alpha = 0,017 = P\left[\chi^2(15) \geq \frac{K}{\sigma_0^2}\right] = P\left[\chi^2(15) \geq \frac{K}{25}\right]$$

$\frac{K}{25} = 27,876$ y $K = 696,9$. El valor muestral del estadístico (874,2) está incluido en la región crítica por lo que se rechaza la hipótesis nula, $\sigma = 25$, y se acepta que la varianza de la distribución pueda tomar cualquier valor superior a 25.

La función de potencia es igual a

$$1 - \beta = P\left[\chi^2(15) \geq \frac{K}{\sigma^2}\right] = P\left[\chi^2(15) \geq \frac{696,9}{\sigma^2}\right].$$

EJERCICIO 8.6 Contrástese con un nivel de significación del 20%, las hipótesis nula $H_0[\sigma^2 = 4]$ y alternativa $H_1[\sigma^2 \neq 4]$. Tomemos una muestra aleatoria simple de tamaño 7, cuyo resultado es

7,1	5,3	4,7	8,0	9,9	3,4	3,6
-----	-----	-----	-----	-----	-----	-----

SOLUCIÓN. La región de aceptación es

$$\chi^2\left(6; 1 - \frac{0,20}{2}\right) 4 < \sum_{i=1}^7 (x_i - \bar{x})^2 < \chi^2\left(6; \frac{0,20}{2}\right) 4$$

$$\chi^2(6; 0,90) = 2,204; \quad \chi^2(6; 0,10) = 10,645$$

y resulta

$$[8,816; 42,580]$$

El valor del estadístico $\sum_{i=1}^7 (x_i - \bar{x})^2$ es 35,12 comprendido en la región de aceptación por lo que no se rechaza la hipótesis nula que la varianza sea igual a 4.

EJERCICIO 8.7 De una población normal bidimensional $(\xi; \eta)$ se ha tomado una muestra aleatoria simple de tamaño 6

x_i	8	11	7	10	12	9
y_j	108	82	63	87	91	88

y se desea contrastar la hipótesis de incorrelación poblacional, $\rho = 0$, entre las dos variables, con un nivel de significación $\alpha = 0,20$.

SOLUCIÓN. El valor del coeficiente de correlación muestral es

$$\begin{aligned}
 r &= \frac{\sum_{i=1}^6 x_i y_i - 6\bar{x}\bar{y}}{\sqrt{\left(\sum_{i=1}^6 x_i^2 - 6\bar{x}^2\right)\left(\sum_{i=1}^6 y_i^2 - 6\bar{y}^2\right)}} = \\
 &= \frac{4961 - \frac{57 \cdot 519}{6}}{\sqrt{\left(559 - \frac{57^2}{6}\right)\left(45951 - \frac{519^2}{6}\right)}} = 0,224
 \end{aligned}$$

El valor del estadístico utilizado en el contraste

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{2 \cdot 0,224}{\sqrt{1-0,224^2}} = 0,460.$$

El valor tabular $t_{0,10}$ con el que se ha de comparar el valor de t se obtiene de

$$P(|t(4; 0,10)| \geq t_{0,10}) = 0,20$$

pues el contraste es bilateral, resultando $t_{0,10} = 1,5332$.

Como el valor de t obtenido en la muestra (0,460) es menor que el crítico (1,5332) no pertenece a la región crítica, está en la de aceptación, y no se rechaza la hipótesis de incorrelación entre las dos variables.

EJERCICIO 8.8 Se ha tomado una muestra aleatoria simple de tamaño 100 de una población normal bidimensional, siendo el coeficiente de correlación muestral $r = 0,58$. Contrástese la hipótesis nula que el coeficiente de correlación poblacional es $\rho = 0,75$, frente a la alternativa que es distinto de esta cifra. El contraste se realiza con un nivel de significación $\alpha = 0,06$.

SOLUCIÓN. El valor del estadístico que permitirá efectuar el contraste es

$$v = \frac{\frac{1}{2} \ln \frac{1 + 0,58}{1 - 0,58} - \frac{1}{2} \ln \frac{1 + 0,75}{1 - 0,75}}{\sqrt{100 - 3}} = -0,0315.$$

El tabular correspondiente, $z_{0,06/2}$, en la distribución $N(0; 1)$ se obtiene de

$$P(|N(0; 1)| \geq z_{0,03}) = 0,06$$

y es $z_{0,03} = 1,891$.

Al ser $|-0,035| = 0,035$ menor que 1,891 se acepta la hipótesis nula: el coeficiente de correlación puede ser igual a 0,75.

EJERCICIO 8.9 Para contratar la hipótesis de igualdad de varianzas de las distribuciones $N(\mu_1; \sigma_1)$ y $N(\mu_2; \sigma_2)$, con un nivel de significación igual al 10% se toman dos muestra aleatorias simples independientes de tamaños 5 y 10

Muestras 1: x_i	Muestras 2: y_i	
25,9	16,7	13,5
22,3	13,6	18,6
26,4	22,8	18,9
24,4	17,2	15,4
27,8	8,9	10,8

SOLUCIÓN. Los valores tabulares que delimitan la región crítica son

$$L_1 = F(4; 9; 1 - 0,05) = \frac{1}{F(9; 4; 0,05)} = \frac{1}{3,63} = 0,275$$

$$L_2 = F(4; 9; 0,05) = 6,00$$

siendo la región crítica la unión de los intervalos $\{F \leq 0,275\}$ y $\{6,00 \leq F\}$.

El estadístico cuyo valor muestral necesitamos es

$$F = \frac{10 - 1}{4 - 1} \frac{\sum_{i=1}^5 (x_i - \bar{x})^2}{\sum_{i=1}^{10} (y_i - \bar{y})^2}$$

$$\sum_{i=1}^5 (x_i - \bar{x})^2 = 3233,26 - 5 \cdot 25,36^2 = 17,612$$

$$\sum_{i=1}^{10} (y_i - \bar{y})^2 = 2597,96 - 10 \cdot 15,64^2 = 151,864$$

y el valor de F es igual a 0,348, no incluido en la región crítica ($0,275 < 0,348 < 6,00$) por lo que no rechazamos la hipótesis nula de igualdad de varianzas, con un nivel de significación del 10%.

CAPÍTULO 9

Contrastes no paramétricos

9.1 Introducción

En los capítulos anteriores se han estudiado los contrastes más usuales donde las hipótesis venían referidas a valores que se asignaban a un parámetro (o conjunto de parámetros) definitorio de una distribución específica de probabilidad $f(x; \theta)$. Todo el proceso, pues, derivaba de la hipótesis de partida de que la muestra obtenida procedía de una población cuya distribución de probabilidad se suponía conocida y donde los procedimientos estadísticos inferenciales se centraban en los parámetros desconocidos de la supuesta distribución poblacional.

La asignación previa de un determinado tipo de distribución de probabilidad a la población (denominada distribución subyacente) gravitaba sobre el proceso inferencial, lo que hacía necesario exigir la propiedad de robustez a las técnicas inferenciales a fin de que vulneraciones de esta hipótesis de partida no afectaran significativamente el resultado de una estimación o de un test de hipótesis.

Para obviar este problema se han desarrollado las técnicas no paramétricas, donde el conjunto de hipótesis de partida se reducen o, incluso, desaparecen, con lo cual disminuye el riesgo de contaminación del proceso inferencial por una errónea especificación del cuadro de hipótesis iniciales.

Dentro de las técnicas no paramétricas se incluyen dos tipos de situaciones, no mutuamente excluyentes:

- Técnicas no paramétricas, en sentido estricto, donde no aparece ningún tipo de hipótesis acerca de un determinado parámetro θ de una población, sino que el proceso se basa en un estadístico sin referencia a ningún parámetro poblacional.
- Métodos de distribución libre, en los que el estadístico utilizado presenta una distribución de probabilidad que no depende de la distribución de probabilidad de la población de la que se ha extraído la muestra que suministra información al estadístico.

En este tipo de técnicas no suele utilizarse directamente la información muestral sobre los valores de la variable objeto de estudio, sino más bien la frecuencia con que aparecen dichos valores en la muestra, la posición de los mismos en una muestra ordenada, el orden o rango que ocupa un valor en la muestra ordenada.

Por otra parte, las técnicas no paramétricas son las únicas que pueden utilizarse si las observaciones se recogen en una escala ordinal o nominal.

Es común, en gran parte de los tests que se desarrollan a continuación, el efectuar el supuesto elemental de la continuidad de población, supuesto que, teóricamente, impide la existencia de valores repetidos¹ en la población investigada. Sin embargo, la intrínseca «discreticidad» de la muestra, dados el tamaño muestral finito con que se trabaja y la imperfección en la medición de los valores de la variable observada, hace posible encontrar con elementos repetidos en la muestra. Esta situación es tratada en los tests que se ven afectados por repeticiones.

Por último, reseñamos que los contrastes no paramétricos son algo menos eficientes² que sus correspondientes paramétricos, cuando la población tiene una distribución normal, y más eficientes cuando la distribución de la población no es normal.

Antes de iniciar esta breve introducción a la contrastación no paramétrica se exponen una serie de conceptos necesarios.

■ **Frecuencia absoluta:** Número de veces que aparece un determinado valor de una variable (modalidad o categoría de un atributo). Se representa por n_i para la i -ésima observación, siendo tal que

$$\sum_{i=1}^r n_i = n .$$

¹ Ties en inglés.

² En el sentido de presentar una menor potencia del contraste para un mismo nivel de significación α .

■ **Frecuencia absoluta acumulada:** Número de observaciones que hay iguales o inferiores a un cierto valor de la variable considerada; se representa por N_i , siendo:

$$\begin{aligned} N_1 &= n_1 \\ N_2 &= n_1 + n_2 \\ N_3 &= n_1 + n_2 + n_3 \\ &\dots\dots\dots \\ N_r &= n_1 + n_2 + \dots + n_r = n \end{aligned}$$

si existen r clases excluyentes de observaciones.

■ **Rango:** Número de orden que le corresponde a cada observación si el conjunto de observaciones se ordenan de menor a mayor, empezando en el 1.

■ **Función indicador:** Función que asigna un código, arbitrario, a una expresión si en ella aparece una observación determinada x_i de la muestra. Por ejemplo

$$\psi_i = \begin{cases} 1 & \text{si } x_i \in A \\ 0 & \text{si } x_i \notin A \end{cases}$$

donde A representa una partición del espacio muestral o la presencia de una determinada característica.

En este capítulo se desarrollan contrastes basados en una muestra, siendo los objetivos básicos verificar si dicha muestra procede de una determinada población (bondad del ajuste), si la muestra es aleatoria (tests de aleatoriedad), que pertenezca a una población con una medida de localización especificada (contrastos de localización), existencia de asociación entre las variables o caracteres que componen una muestra bidimensional (tests de asociación), existencia de independencia entre los distintos niveles de dos factores o variables si la muestra se puede disponer en forma de tabla de doble entrada (Tablas de Contingencia).³

9.2 Contrastes de bondad del ajuste

El objetivo de estos contrastes de bondad del ajuste es verificar si una muestra procede de una población con una determinada distribución de probabilidad.

³ En el epígrafe correspondiente a Tablas de Contingencia, también se incluye el contraste de homogeneidad, aunque su diseño parte del conocimiento de dos o más muestras.

A la información de la muestra $\mathbf{X} (x_1, \dots, x_n)$ se ajusta una cierta distribución de probabilidad, siendo la hipótesis nula que dicho ajuste es «bueno», en el sentido que la información muestral no evidencie que debe rechazarse esta hipótesis nula. La hipótesis alternativa recoge el enunciado opuesto al de la hipótesis nula, es decir, que dicho ajuste no es correcto.

Veremos a continuación los tests más usados para este fin.

9.2.1. TEST χ^2

Como se ha señalado, la hipótesis nula consiste en que las observaciones constituyen una muestra aleatoria simple extraída de una población con una distribución de probabilidad, P .

Se distinguen dos casos: el primero supone que P está completamente especificada, de modo que no contiene parámetros desconocidos y, por tanto, puede calcularse la probabilidad de cualquier suceso en P . El segundo, contempla lo contrario, cuando es necesaria la estimación previa de algunos parámetros desconocidos a fin de poder determinar probabilidades de subconjuntos del campo de variación de P .

9.2.1.1. Test χ^2 cuando no se estiman parámetros

Se dispone de una muestra cuyas observaciones pueden clasificarse en r clases o categorías mutuamente excluyentes: categorías que pueden venir dadas en intervalos de tipo cuantitativo o simplemente en una escala ordinal o nominal. Si representamos las categorías por A_1, A_2, \dots, A_r , debe cumplirse que su unión configure el campo de variación de la variable poblacional.

En la muestra existen n_1 elementos que pertenecen a la clase A_1 , n_2 a la A_2 , y así, sucesivamente, n_r a la última clase A_r , tal que $n_1 + n_2 + \dots + n_r = n$, siendo n el tamaño de la muestra.

Bajo el supuesto que la hipótesis nula es cierta, de acuerdo con la distribución de probabilidad P , puede establecerse que

$$P(A_i) = p_i$$

donde

$$p_1 + p_2 + \dots + p_r = 1.$$

Al ser independientes las observaciones, por proceder de una muestra aleatoria simple, la probabilidad de la aparición de n_i elementos de la clase A_i , n_2 ele-

mentos de la clase A_2 , etc., viene dada por una distribución multinomial⁴ a través de la expresión

$$P(n_1, \dots, n_r) = \frac{n!}{n_1! \dots n_r!} p_1^{n_1} \dots p_r^{n_r}$$

donde cada n_i sigue una distribución marginal de tipo binomial $B(n; p_i)$ cuyo valor esperado, que denominamos E_i , es

$$E(n_i) = np_i = E_i$$

este valor esperado representa el número de observaciones pertenecientes a la clase A_i que cabe esperar se obtengan en la muestra, si la distribución de probabilidad de la población es la que se incluye en la hipótesis nula.

Debido a las fluctuaciones aleatorias en el proceso del muestreo es razonable que las **frecuencias observadas** n_i y las **frecuencias esperadas** E_i no coincidan, aunque para valores grandes del tamaño muestral n es previsible, si la hipótesis nula es cierta, que las discrepancias entre dichas frecuencias no sean grandes.

Como medida de las diferencias, o discrepancias entre las frecuencias observadas y esperadas, Pearson propuso el estadístico

$$X^2 = \sum_{i=1}^r \frac{(n_i - E_i)^2}{E_i} = \sum_{i=1}^r \frac{(n_i - np_i)^2}{np_i}$$

llamado estadístico de la bondad del ajuste.

En esta expresión cada diferencia $n_i - E_i$ aparece elevada al cuadrado, para evitar que signos contrarios compensen la medida global, pues el interés se centra en la cuantía de la desviación y no en su dirección o signo. La diferencia se pondera por el inverso de la frecuencia esperada E_i , puesto que una discrepancia grande podría llevar a rechazar el modelo de probabilidad recogido en la hipótesis nula, aunque la categoría A_i sea de probabilidad no muy grande. No es lo mismo encontrar una diferencia $(n_i - E_i)^2 = 25$ en una categoría donde $E_i = 500$ que en otra cuyo valor esperado sea $E_i = 5$.

⁴ Véase MARTÍN PLIEGO y RUIZ-MAYA : *Fundamentos...* Capítulo 7.

Distribución del estadístico X^2

La distribución conjunta multinomial de las frecuencias muestrales n_1, \dots, n_r puede obtenerse como la distribución condicional de r variables independientes n_1, \dots, n_r con distribución de Poisson de parámetro np_i (donde $p_1 + \dots + p_r = 1$), dada la condición $n_1 + \dots + n_r = n$.

Por tanto, las variables tipificadas

$$\xi_i = \frac{n_i - np_i}{\sqrt{np_i}}$$

son asintóticamente normales $N(0; 1)$ e independientes.⁵

Por otra parte, la condición $\sum_{i=1}^r n_i = n$ obliga a que

$$\sum_{i=1}^r \sqrt{p_i} \xi_i = 0 \quad [1]$$

pues, como

$$\sqrt{np_i} \xi_i = \sqrt{n} \sqrt{p_i} \xi_i = n_i - np_i$$

al sumar esta expresión para todo valor de i se obtiene

$$\sqrt{n} \sum_{i=1}^r \sqrt{p_i} \xi_i = \sum_{i=1}^r n_i - n \sum_{i=1}^r p_i = \sum_{i=1}^r n_i - n = 0.$$

Estas consideraciones conducen a que el estadístico

$$\begin{aligned} X^2 &= \sum_{i=1}^r \frac{(n_i - np_i)^2}{np_i} = \\ &= \sum_{i=1}^r \xi_i^2 \end{aligned}$$

siga una distribución χ^2 con $r - 1$ grados de libertad, reduciéndose los grados de libertad en uno por la restricción lineal dada por [1] que, de manera implícita, contiene el estadístico.⁶

⁵ Véase ROHATGI.

⁶ Para una demostración formal puede consultarse KENDALL-STUART o CRAMÉR, que ofrecen dos exposiciones diferentes.

Región crítica

Valores elevados del estadístico X^2 evidencian discrepancias relevantes entre las frecuencias observadas n_i y las esperadas E_i , por lo que deberá rechazarse la hipótesis nula de que dicha muestra procede de una población con distribución de probabilidad P , siendo la región crítica del test de la forma $X^2 \geq K$.

El valor de K se determina de manera que si α es el nivel de significación fijado se verifique

$$P(X^2 \geq K/H_0) = \alpha$$

obteniéndose K mediante las tablas de la distribución χ^2 .

Observaciones

■ Dado que la distribución del estadístico X^2 es asintótica se utiliza, comúnmente, como regla de aproximación aceptable, a condición que todas las frecuencias esperadas sean mayores o iguales a cinco ($E_i \geq 5$). En el caso que no se dé esta condición hay que proceder a un reagrupamiento de clases o categorías hasta lograr el cumplimiento de la regla. Este reagrupamiento produce, a su vez, una reducción de los grados de libertad de la distribución del estadístico X^2 , al disminuir el número r de clases o categorías.

■ El test X^2 puede aplicarse a cualquier distribución poblacional, continua o discreta. Asimismo, las categorías o clases pueden estar definidas en términos de modalidades de un atributo, por lo que su uso no excluye situaciones donde las observaciones sean ordinales o nominales.

■ Es aconsejable, para una mejor calidad de la aproximación a través de la distribución asintótica, que las clases o categorías se construyan de tal manera que sus probabilidades sean aproximadamente iguales.

EJEMPLO 1

Con un dado de póker se han realizado 600 tiradas con el siguiente resultado:

<i>Caras del dado</i>	n_i
As	70
K	115
Q	122
J	98
Rojo	85
Negro	110

Verifíquese al 5% de nivel de significación la hipótesis que el dado está bien construido.

La hipótesis que el dado está bien construido equivale a que la muestra de 600 tiradas procede de una población uniforme discreta con probabilidad igual a $1/6$ para cada cara del dado.

Es decir, se puede postular que

$$H_0 \left[\text{es "bueno" el ajuste de una distribución de probabilidad con } p_i = \frac{1}{6} \right].$$

En primer lugar, para la realización del contraste, se determinan las frecuencias esperadas E_i que, en este ejemplo donde existe equiprobabilidad para las seis alternativas

$$p_i = \frac{1}{6}, \text{ son}$$

$$E_i = n p_i = 600 \frac{1}{6} = 100$$

Como todas las $E_i \geq 5$ se llega a que el estadístico del contraste sigue aproximadamente una distribución χ^2 con $r - 1 = 6 - 1 = 5$ grados de libertad, cuya región crítica del 5% es tal que

$$P(\chi^2(5) \geq K/H_0) = 0,05$$

obteniéndose en la correspondiente tabla de la distribución χ^2 que el valor crítico K es igual a 11,070.

El valor muestral del estadístico resulta

$$\begin{aligned} X^2 &= \frac{(70 - 100)^2}{100} + \frac{(115 - 100)^2}{100} + \frac{(122 - 100)^2}{100} + \\ &+ \frac{(98 - 100)^2}{100} + \frac{(85 - 100)^2}{100} + \frac{(110 - 100)^2}{100} = \\ &= 19,38 \end{aligned}$$

Como $X^2 = 19,38 > 11,070$ se puede rechazar la hipótesis nula, pues la muestra parece evidenciar una contradicción entre dicha hipótesis y las observaciones muestrales; las frecuencias observadas y las esperadas son significativamente distintas con $\alpha = 5\%$.

9.2.1.2. Text χ^2 cuando se estiman parámetros

La hipótesis nula especifica que la muestra procede de una población cuya distribución de probabilidad contiene parámetros desconocidos, donde en cada categoría A_i

$$P(A_i; \theta_1, \dots, \theta_k) = p_i(\theta_1, \dots, \theta_k)$$

probabilidades que no se podrán determinar si previamente no se estiman estos parámetros desconocidos.

Fisher demostró que si se estiman los parámetros θ_j con la misma información muestral con que se realiza el contraste, el estadístico del test

$$X^2 = \sum_{i=1}^r \frac{[n_i - n p_i(\theta_1, \dots, \theta_k)]^2}{n p_i(\theta_1, \dots, \theta_k)} = \sum_{i=1}^r \frac{(n_i - \hat{E}_i)^2}{\hat{E}_i}$$

tiene una distribución asintótica que sigue una ley χ^2 con $r - k - 1$ grados de libertad, siendo k el número de parámetros que es necesario estimar y $\hat{E}_i = n \hat{p}_i(\theta_1, \dots, \theta_k)$ las frecuencias estimadas de cada categoría.

La forma de la región crítica es análoga a la del caso anterior.

EJEMPLO 2

Dada la siguiente información muestral de tamaño 1.000:

<i>Intervalos</i>	n_i
De 0 hasta 20	110
Más de 20 hasta 40	200
Más de 40 hasta 60	300
Más de 60 hasta 80	250
Más de 80 hasta 100	140

ajústese una distribución normal y verifíquese si este ajuste es «bueno», al nivel de significación del 5%.

Una distribución normal viene especificada en función de dos parámetros: la media poblacional μ y su desviación típica σ .

Como no son conocidos los parámetros deben estimarse. Si recurrimos a los estimadores máximo-verosímiles⁵

$$\hat{\mu} = \bar{x}$$

$$\hat{\sigma}^2 = s^2$$

⁵ Véase el Capítulo 4.

para la muestra facilitada se obtiene:

<i>Intervalos</i>	n_i	<i>Marcas de clase x_i</i>	$x_i n_i$	$x_i^2 n_i$
Desde 0 hasta 20	110	10	1.100	11.000
Más de 20 hasta 40	200	30	6.000	180.000
Más de 40 hasta 60	300	50	15.000	750.000
Más de 60 hasta 80	250	70	17.500	1.225.000
Más de 80 hasta 100	140	90	12.600	1.134.000
	$n=1.000$		52.200	3.300.000

donde

$$\hat{\mu} = \bar{x} = \frac{\sum_{i=1}^r x_i n_i}{n} = \frac{52.200}{1.000} = 52,2$$

$$\hat{\sigma}^2 = s^2 = a_2 - \bar{x}^2 = \frac{\sum_{i=1}^r x_i^2 n_i}{n} - \bar{x}^2 = \frac{3.300.000}{1.000} - 52,2^2 = 575,16$$

$$\hat{\sigma} = s = \sqrt{575,16} = 23,98$$

Por tanto, se procede al ajuste de una distribución normal $N(52,2; 23,98)$, siendo la hipótesis nula que la muestra procede de la población ahora especificada.

Para estimar las frecuencias esperadas E_i es necesario calcular las probabilidades de las distintas clases o intervalos siendo preciso, pues, determinarlas mediante las tablas de la distribución normal

$$p_1 = P(0 < \eta \leq 20) = P\left[\frac{-52,2}{23,98} < \xi \leq \frac{20 - 52,2}{23,98}\right] =$$

$$= P(-2,18 < \xi \leq -1,34) = 0,0901 - 0,0146 = 0,0755$$

$$p_2 = P(20 < \eta \leq 40) = P(-1,34 < \xi \leq -0,51) = 0,3050 - 0,0901 = 0,2149$$

$$p_3 = P(40 < \eta \leq 60) = P(-0,51 < \xi \leq 0,33) = 1 - (0,3707 + 0,3050) = 0,3243$$

$$p_4 = P(60 < \eta \leq 80) = P(0,33 < \xi \leq 1,16) = 0,3707 - 0,1230 = 0,2477$$

$$p_5 = P(80 < \eta \leq 100) = P(1,16 < \eta \leq 1,99) = 0,1230 - 0,0233 = 0,0997$$

Podemos formar ahora el cuadro con la información requerida para el contraste:

Intervalos	n_i	\hat{p}_i	$\hat{E}_i = n\hat{p}_i$	$n_i - \hat{E}_i$	$(n_i - \hat{E}_i)^2$	$\frac{(n_i - \hat{E}_i)^2}{\hat{E}_i}$
Desde 0 hasta 20	110	0,0755	75,5	34,5	1.190,25	15,76
Más de 20 hasta 40	200	0,2149	214,9	-14,9	222,01	1,03
Más de 40 hasta 60	300	0,3243	324,3	-24,3	590,49	1,82
Más de 60 hasta 80	250	0,2477	247,7	2,3	5,29	0,02
Más de 80 hasta 100	140	0,0997	99,7	40,3	1.624,09	16,29
	$n = 1.000$					34,92

Como todas las estimaciones \hat{E}_i son mayores que 5, el estadístico del contraste sigue aproximadamente una distribución χ^2 con $r - k - 1 = 5 - 2 - 1 = 2$ grados de libertad, pues $k = 2$ dado que han sido dos los parámetros estimados.

El valor de K que define la región crítica satisface la ecuación

$$P(\chi^2(2) \geq K/H_0) = \alpha = 0,05$$

y es $K = 5,991$.

Al ser el valor muestral de estadístico del contraste

$$X^2 = \sum_{i=1}^r \frac{(n_i - \hat{E}_i)^2}{\hat{E}_i} = 34,92$$

mayor que el crítico $K = 5,991$ se rechaza la hipótesis nula, pues la muestra presenta evidencia clara de no proceder de una población $N(52,2; 23,98)$.

9.2.2 TEST G^2 DE LA RAZÓN DE VEROSIMILITUD

Con la hipótesis nula de que una muestra procede de una población con una distribución de probabilidad P , el diseño de este contraste parte del test razón de verosimilitud, que para la distribución multinomial, que se asigna a la muestra agrupada en clases o categorías excluyentes, es

$$\lambda(\mathbf{X}) = \frac{L(p_1^0, p_2^0, \dots, p_r^0)}{L(\hat{p}_1, \hat{p}_2, \dots, \hat{p}_r)}$$

donde p_i^0 representa la probabilidad de la clase i -ésima de acuerdo con la hipótesis nula H_0 y \hat{p}_i su correspondiente estimación máximo-verosímil.

Obtengamos, en primer lugar, estas estimaciones. La función de verosimilitud de la muestra es

$$\begin{aligned} L(p_1, p_2, \dots, p_r) &= \frac{n!}{n_1! n_2! \dots n_r!} p_1^{n_1} \cdot p_2^{n_2} \dots p_r^{n_r} = \\ &= Q \cdot p_1^{n_1} \cdot p_2^{n_2} \dots p_{r-1}^{n_{r-1}} (1 - p_1 - \dots - p_{r-1})^{n_r} \end{aligned}$$

siendo

$$Q = \frac{n!}{n_1! n_2! \dots n_r!}$$

una expresión que no depende de los parámetros p_i , y teniendo en cuenta la restricción entre los parámetros que conduce a que $p_r = 1 - p_1 - \dots - p_{r-1}$.

El logaritmo de la función de verosimilitud es

$$\begin{aligned} \ln L(p_1, p_2, \dots, p_r) &= \ln Q + n_1 \ln p_1 + n_2 \ln p_2 + \dots + \\ &\quad + n_r \ln (1 - p_1 - \dots - p_{r-1}) \end{aligned}$$

donde la condición de existencia de extremo genera el sistema de ecuaciones

$$\frac{\partial \ln L(p_1, p_2, \dots, p_r)}{\partial p_i} = \frac{n_i}{p_i} - \frac{n_r}{1 - p_1 - \dots - p_{r-1}} = 0$$

que implica

$$\frac{n_r}{1 - p_1 - \dots - p_{r-1}} = \frac{n_1}{p_1} = \frac{n_2}{p_2} = \dots = \frac{n_{r-1}}{p_{r-1}} = n$$

siendo los estimadores máximo-verosímiles

$$\hat{p}_i = \frac{n_i}{n}.$$

El estadístico razón de verosimilitud es

$$\begin{aligned} \lambda(\mathbf{X}) &= \frac{Q \cdot (p_1^0)^{n_1} (p_2^0)^{n_2} \dots (p_r^0)^{n_r}}{Q \cdot (\hat{p}_1)^{n_1} (\hat{p}_2)^{n_2} \dots (\hat{p}_r)^{n_r}} = \\ &= \frac{\prod_{i=1}^r (p_i^0)^{n_i}}{\prod_{i=1}^r \binom{n_i}{n}} = n^{\sum_{i=1}^r n_i} \prod_{i=1}^r \left(\frac{p_i^0}{n_i} \right)^{n_i} = \\ &= n^n \prod_{i=1}^r \left(\frac{p_i^0}{n_i} \right)^{n_i}. \end{aligned}$$

Por otra parte, $-2 \ln \lambda(\mathbf{X})$ tiene una distribución asintótica que sigue una ley χ^2 cuyos grados de libertad son el número de parámetros menos el número de restricciones que presenten estos parámetros, es decir, el número de grados de libertad será $r - 1$.⁶

El test basado en lo anterior se conoce como test G^2 de la razón de verosimilitud para la bondad de ajuste.

Por tanto,

$$G^2 = -2 \ln \lambda(\mathbf{X}) = -2 \left[n \ln n + \sum_{i=1}^r n_i \ln \left(\frac{p_i^0}{n_i} \right) \right].$$

La región crítica tiene la forma $G^2 \geq K$ y el valor de K viene determinado por

$$P(G^2 \geq K/H_0) = P(\chi^2(r-1) \geq K/H_0) = \alpha$$

siendo α el nivel de significación.

EJEMPLO 3

Para la misma muestra de 600 tiradas del dado del ejemplo 1 (pág 277), siendo el nivel de significación $\alpha = 0,05$, se tendría el valor crítico $K = 11,070$, donde el valor muestral del estadístico es

$$\begin{aligned} G^2 &= -2 \left[600 \ln 600 + 70 \ln \left(\frac{1/6}{70} \right) + 115 \ln \left(\frac{1/6}{115} \right) + 122 \ln \left(\frac{1/6}{122} \right) \right. \\ &\quad \left. + 98 \ln \left(\frac{1/6}{98} \right) + 85 \ln \left(\frac{1/6}{85} \right) + 110 \ln \left(\frac{1/6}{110} \right) \right] = \\ &= 20,1107 \end{aligned}$$

lo que conduce a rechazar la hipótesis nula H_0 que el dado está bien construido, no evidenciándose que todas las caras tengan la misma probabilidad de aparición.

⁶ Véase el Capítulo 7.

9.2.3. TEST DE KOLMOGOROV-SMIRNOV

La función de distribución empírica de la muestra ⁷ es

$$F_n(x_i) = \frac{N_i}{n},$$

donde N_i es la frecuencia absoluta acumulada o número de observaciones menores o iguales a x_i , $N_i = n_1 + \dots + n_i$.

Si la hipótesis nula es que la muestra procede de una población con función de distribución $F(x)$ de tipo continuo, las diferencias entre $F_n(x)$ y $F(x)$, para una muestra de tamaño suficientemente grande, es de esperar que no sean significativas.

La mayor discrepancia vertical entre $F_n(x)$ y $F(x)$, es decir, el estadístico

$$D_n = \sup_{-\infty < x < \infty} |F_n(x) - F(x)|$$

recibe el nombre de **estadístico de Kolmogorov-Smirnov**, y tiene la siguientes propiedades:

- Según el teorema de Glivenko-Cantelli⁸ la función de distribución empírica converge casi seguro a la función de distribución $F(x)$, por lo que,

$$P\left[\lim_{n \rightarrow \infty} D_n = 0\right] = 1.$$

- El estadístico D_n presenta una distribución libre y, por consiguiente, la distribución de probabilidad de D_n es la misma cualquiera que sea $F(x)$. Téngase en cuenta que la variable aleatoria $y = F(x)$, por la propiedad de la transformación integral⁹, tiene siempre una distribución uniforme continua en el intervalo $[0;1]$, con lo que

$$D_n = \sup_{-\infty < x < \infty} |F_n(x) - F(x)| = \sup_{-\infty < x < \infty} |F_n(x) - y|.$$

La distribución asintótica del estadístico D_n también es una distribución libre de $F(x)$.

⁷ Véase el Capítulo 1.

⁸ Véase MARTÍN PLIEGO y RUIZ-MAYA: *Fundamentos...* Capítulo 9.

⁹ Véase MARTÍN PLIEGO y RUIZ-MAYA: *Fundamentos...* Capítulo 7.

Para verificar la hipótesis nula H_0 [es «bueno» el ajuste de la distribución continua $F(x)$] se utiliza la región crítica dada por

$$P(D_n \geq K/H_0) = \alpha,$$

siendo α el nivel de significación.

Los valores críticos K para diferentes niveles de significación y distintos tamaños muestrales se recogen en la tabla 5 del Apéndice de Tablas Estadísticas, titulada Test de Kolmogorov-Smirnov.

EJEMPLO 4

Verifíquese al 2% de nivel de significación si la muestra de tamaño 50

Intervalos	n_i
0,00-0,20	1
0,20-0,40	7
0,40-0,60	10
0,60-0,80	18
0,80-1,00	14
	$n = 50$

procede de una población continua con función de distribución

$$F(x) = x^2, \quad \text{para } 0 \leq x \leq 1$$

Dispongamos los cálculos necesarios para determinar el valor muestral del estadístico con D_n de Kolmogorov-Smirnov

Intervalos	n_i	p_i	N_i	$F_n(x) = \frac{N_i}{n}$	$F(x)$	$D = F_n(x) - F(x) $
0,00-0,20	1	$0,2^2 = 0,04$	1	0,02	0,04	0,02
0,20-0,40	7	$0,4^2 - 0,2^2 = 0,12$	8	0,16	0,16	0,00
0,40-0,60	10	$0,6^2 - 0,4^2 = 0,20$	18	0,36	0,36	0,00
0,60-0,80	18	$0,8^2 - 0,6^2 = 0,28$	36	0,72	0,64	0,08
0,80-1,00	14	$1 - 0,8^2 = 0,36$	50	1,00	1,00	0,00
	$n = 50$					

siendo

$$D_n = \sup_{-\infty < x < \infty} |F_n(x) - F(x)| = 0,08.$$

Como en la Tabla 5 del Apéndice de Tablas Estadísticas para $n = 50$ y $\alpha = 0,02$ se obtiene un valor crítico aproximado $K = 0,214$, al ser $D_n = 0,08 < 0,214$ no se rechaza la hipótesis nula que la muestra analizada pueda proceder de una población con función de distribución

$$F(x) = x^2 \quad \text{para} \quad 0 \leq x \leq 1.$$

9.2.4. COMPARACIÓN DE LOS TESTS χ^2 , G^2 Y DE KOLMOGOROV-SMIRNOV

A modo de resumen, se pueden establecer las siguientes conclusiones:

- El test de Kolmogorov-Smirnov es de más sencilla aplicación que los tests G^2 y χ^2 de bondad del ajuste.

- El test de Kolmogorov-Smirnov no se ve afectado por reagrupaciones en las categorías en que se clasifican las observaciones, mientras que en los tests G^2 y χ^2 se pierde información con tales reagrupamientos, disminuyendo los grados de libertad de la distribución de los estadísticos utilizados.

- El contraste de Kolmogorov-Smirnov es aplicable a muestras pequeñas pues utiliza la distribución exacta del estadístico D_n mientras que los tests G^2 y χ^2 están diseñados para muestras de gran tamaño al basarse en las distribuciones asintóticas de sus correspondientes estadísticos.

- La potencia del contraste de Kolmogorov-Smirnov es mayor que las de los tests G^2 y χ^2 , si bien tienden a similares potencias cuando n se hace suficientemente grande.

- Los tests G^2 y χ^2 pueden adaptarse fácilmente cuando hay que estimar parámetros de la población, mientras que el de Kolmogorov-Smirnov se hace excesivamente conservador en esta situación.

- Los contrastes G^2 y χ^2 son aplicables a poblaciones con distribuciones discretas o continuas y cuando la información muestral es cualitativa. Sin embargo, el test Kolmogorov-Smirnov requiere la condición que la función de distribución $F(x)$ de la población donde se extrae la muestra sea continua.

9.3 Test de rachas para el contraste de aleatoriedad

El objetivo de estos contrastes es verificar que las observaciones x_1, \dots, x_n constituyen una muestra aleatoria procedente de una población continua.

Una **racha**¹⁰ es una sucesión de símbolos del mismo tipo limitada por símbolo de tipo distinto. Consideraremos el caso más elemental que sólo existan dos tipos de símbolos; en este caso, las observaciones se pueden designar como de tipo A o B. Las rachas extremas no están precedidas o seguidas de ningún símbolo distinto.

Como ejemplo, consideremos la secuencia

AAA BBB AA BBB AAAA B

que consta de seis rachas.

Si las dos clases de observaciones, A y B, proceden aleatoriamente de una misma población, los símbolos A y B en la secuencia aparecerán mezclados y, por consiguiente, el número de rachas será elevado. Mientras que, por el contrario, si las observaciones no han aparecido aleatoriamente, el número de rachas tenderá a ser reducido, en el límite solamente dos, es decir,

AA ... A BB ... B .

Si en la muestra de tamaño n existen n_1 observaciones del tipo A y n_2 del B, siendo $n_1 + n_2 = n$, las distribuciones condicionadas de probabilidad de que se presenten R rachas son ¹¹

■ Para R par:

$$P(R/n_1, n_2) = \frac{2 \binom{n_1 - 1}{\frac{R}{2} - 1} \binom{n_2 - 1}{\frac{R}{2} - 1}}{\binom{n}{n_1}}$$

¹⁰ En inglés, *run*.

¹¹ La laboriosa obtención de la distribución de probabilidad de las rachas puede verse, entre otros, en FISZ y GIBBONS-CHAKRABORTI.

en este caso, el número de rachas de cada tipo de símbolo A y B son las mismas. Si denotamos por R_1 y R_2 el número de estas rachas se tiene que

$$R_1 = R_2 = \frac{R}{2}.$$

■ Para R impar:

$$P(R/n_1, n_2) = \frac{\binom{n_1 - 1}{\frac{R - 1}{2}} \binom{n_2 - 1}{\frac{R - 3}{2}} + \binom{n_1 - 1}{\frac{R - 3}{2}} \binom{n_2 - 1}{\frac{R - 1}{2}}}{\binom{n}{n_1}},$$

verificándose que $R_1 = R_2 \pm 1$.

Si representamos por $\gamma = \frac{n_1}{n}$, la proporción de elementos de tipo A en la muestra, con lo que $\frac{n_2}{n}$ es igual a $1 - \gamma$, Wald y Wolfowitz demostraron que la variable aleatoria R , número de rachas, sigue una distribución asintótica normal dada por

$$R \xrightarrow{d} N[2\gamma(1 - \gamma)n; 2\gamma(1 - \gamma)\sqrt{n}].$$

siendo la aproximación aceptable si $n_1 > 10$ y $n_2 > 10$.

La región crítica que se utiliza en estos tests suele ser, en general, de tipo unilateral pues pocas rachas indican cierta tendencia a que las dos clases de elementos A y B no procedan aleatoriamente de una misma población continua. Por tanto, la región crítica vendrá definida a través de

$$P(R \leq K/H_0) = \alpha,$$

siendo α el nivel de significación.

Los valores críticos de K para este test se presentan en la Tabla 6(a) del Apéndice de Tablas Estadísticas para un nivel de significación $\alpha = 0,05$ y, valores de n_1 y n_2 menores o iguales a 20.

También puede considerarse una región crítica de tipo bilateral, pues una secuencia como

ABABAB ... ABABABAB ...

parece indicar cierta periodicidad sistemática que no concuerda con la hipótesis de aleatoriedad de la muestra.

En este caso la región crítica se obtiene a través de las expresiones

$$P(R \leq K_1/H_0) = \frac{\alpha}{2}$$

$$P(R \geq K_2/H_0) = \frac{\alpha}{2}$$

donde $K_1 < R < K_2$ es la región de aceptación para la hipótesis nula. El valor crítico K_1 se obtendrá de la Tabla 6(a) del Apéndice y el correspondiente a K_2 de la Tabla 6(b) para $\alpha = 0,05$.

EJEMPLO 4

En un colectivo de 15 niños, de ambos sexos, se obtienen las medidas de sus estaturas con el resultado siguiente

Niñas	1,43	1,54	1,40	1,55	1,50	1,60	1,41	1,47	1,51
Niños	1,62	1,45	1,58	1,61	1,48	1,63			

Verifíquese al 5% de nivel de significación que estas estaturas constituyen una muestra aleatoria.

Ordenamos los valores observados en la muestra de la variable estatura afectando a cada valor de un símbolo indicativo del sexo del niño (A:niña, O: niño) cuya estatura se ha medido

1,40 1,41 1,43 1,45 1,47 1,48 1,50 1,51 1,54 1,55 1,58 1,60 1,61 1,62 1,63
A A A O A O A A A A O A O O O

presentando $R = 8$ rachas, con $n_1 = 9$ y $n_2 = 6$.

Parece indicado en este ejemplo el uso de un contraste unilateral donde la presencia de pocas rachas mostrarían que las poblaciones de estaturas de niñas y niños serían diferentes.

Por tanto, el valor crítico que cumple

$$P(R \leq K/H_0) = \alpha = 0,05$$

lo encontramos en la Tabla 6(a) donde para $n_1 = 9$ y $n_2 = 6$ se obtiene $K = 4$.

Como en la muestra $R = 8 > 4$, no se rechaza la hipótesis nula que las observaciones muestrales proceden aleatoriamente de una misma población continua.

9.4 Test de Wilcoxon: Contraste de localización

Mediante el test se contrasta la hipótesis nula que la muestra procede de una población continua con una determinada medida de posición o localización. Si la hipótesis nula es que el percentil¹² p de la población es θ , se tendrán que localizar $p\%$ observaciones muestrales inferiores a θ y $(100 - p)\%$ superiores a esta medida de localización θ .¹³

Sea X una muestra aleatoria simple de tamaño n procedente de una población continua simétrica respecto a la mediana Me . Este enunciado constituye la hipótesis nula.

Se determinan las diferencias $z_i = x_i - Me$, y se asignan los rangos R_i a la secuencia ordenada de las diferencias en valor absoluto, $|z_i|$. Este contraste no solamente tiene en cuenta el signo de las diferencias de la variable auxiliar z_i , sino también la magnitud de tales diferencias.

Siendo la función indicador

$$\psi_i = \begin{cases} 1 & \text{si } z_i > 0 \\ 0 & \text{si } z_i < 0 \end{cases}$$

el estadístico T^+ de Wilcoxon se define como

$$T^+ = \sum_{i=1}^n R_i \psi_i$$

que proporciona la suma de los rangos de las diferencias con signo positivo, pues el producto $R_i \psi_i$ es igual a cero si z_i es negativo, e igual a R_i si la diferencia z_i es positiva.

Un valor elevado de T^+ tiene lugar cuando muchas de las mayores desviaciones absolutas respecto a Me son positivas y, por tanto, la población no parece que pueda ser simétrica respecto a ese valor de la mediana. De la misma manera, un

¹² Véase MARTÍN PLIEGO.

¹³ Estos métodos son de distribución libre.

valor reducido de T^+ indica que la mayor parte de las menores desviaciones absolutas respecto a Me son negativas, llegándose a la misma conclusión que antes.

Si todas las diferencias tuvieran signo positivo

$$T^+ = \sum_{i=1}^n R_i \psi_i = \sum_{i=1}^n R_i = 1 + 2 + \dots + n = \frac{n(n+1)}{2}$$

mientras que si todas fueran negativas

$$T^+ = \sum_{i=1}^n R_i \psi_i = \sum_{i=1}^n R_i \cdot 0 = 0$$

Por todo lo anterior, la región crítica es

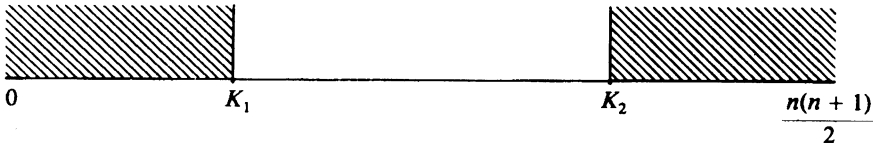


FIGURA 9.1

cumpliéndose

$$P(T^+ \leq K_1 / H_0) = \frac{\alpha}{2}$$

$$P(T^+ \geq K_2 / H_0) = \frac{\alpha}{2}$$

Un valor muestral T^+ tal que $K_1 < T^+ < K_2$ permite no rechazar la hipótesis nula H_0 . Para $n \leq 30$, existe una tabla con probabilidades exactas.¹⁴ Para $n \geq 15$, es posible recurrir a la distribución asintótica

$$T^+ \xrightarrow{d} N\left(\frac{n(n+1)}{4}; \sqrt{\frac{n(n+1)(2n+1)}{24}}\right).$$

¹⁴ Véase DANIEL.

EJEMPLO 5

De un colectivo de profesionales se toma una muestra de la cuota íntegra satisfecha a la Hacienda por el impuesto personal.

Compruébese que la variable que define los pagos a la Hacienda, por este impuesto directo, tiene una distribución simétrica con mediana $Me = 2,4$ millones de pesetas. Nivel de significación $\alpha = 3\%$.

La información muestral se recoge en la tabla siguiente:

Importe Cuota íntegra x_i	$z_i = x_i - 2,4$	ψ_i	$ z_i $	R_i	$R_i \psi_i$
3,8	1,4	1	1,4	11	11
4,5	2,1	1	2,1	14	14
1,3	-1,1	0	1,1	10	0
4,3	1,9	1	1,9	13	13
2,8	0,4	1	0,4	4	4
5,1	2,7	1	2,7	16	16
0,8	-1,6	0	1,6	12	0
2,3	-0,1	0	0,1	1	0
5,3	2,9	1	2,9	17	17
5,4	3,0	1	3,0	18	18
2,7	0,3	1	0,3	3	3
1,8	-0,6	0	0,6	6	0
3,3	0,9	1	0,9	9	9
3,2	0,8	1	0,8	8	8
1,7	-0,7	0	0,7	7	0
1,9	-0,5	0	0,5	5	0
2,6	0,2	1	0,2	2	2
4,8	2,4	1	2,4	15	15

Según la información muestral, el valor del estadístico T^+ es

$$T^+ = \sum_{i=1}^n R_i \psi_i = 130.$$

Para determinar la región crítica tenemos en cuenta la distribución asintótica del estadístico T^+

$$T^+ \xrightarrow{d} N\left(\frac{n(n+1)}{4}; \sqrt{\frac{n(n+1)(2n+1)}{24}}\right) = N(85,5; 22,96)$$

para $n = 18$.

Con las tablas de la distribución $N(0; 1)$ se llega a que la región crítica del 3% de nivel de significación es

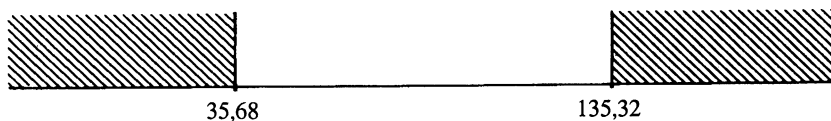


FIGURA 9.2

Como $T^+ = 130$ no se encuentra incluido en la región crítica, no rechazamos la hipótesis nula que la distribución de los pagos por este tipo de impuesto sea de carácter simétrico con mediana igual a 2,4 millones de pesetas, en el colectivo estudiado.

9.5 Test de correlación por rangos de Spearman para asociación

En este epígrafe se expone un procedimiento de distribución libre que parte de observaciones muestrales obtenidas de una población bidimensional, a fin de verificar la existencia o no de asociación entre las dos variables que componen la población.

Sólo cuando la población sea normal bidimensional los contrastes de asociación pueden interpretarse como contrastes de independencia.

A partir de una muestra aleatoria bidimensional $(x_i; y_i)$ de tamaño n , extraída de una población continua, se desea verificar la hipótesis nula H_0 de no existencia de correlación entre las variables ξ y η que definen la población, frente a la hipótesis alternativa de que exista algún grado de asociación.

Para ello, se ordenan por separado en cada observación bidimensional los elementos x_i e y_i de menor a mayor y se asignan los rangos $R_i(x)$ y $R_i(y)$ a cada uno.

Se define el coeficiente de correlación por rangos de Spearman¹⁵ como el coeficiente de correlación lineal entre los rangos $R_i(x)$ y $R_i(y)$, que denotamos por r_s .

¹⁵ Véase MARTÍN PLIEGO.

Calculamos, previamente, el coeficiente dado por

$$r_s = \frac{\sum_{i=1}^n [R_i(x) - \bar{R}(x)] [R_i(y) - \bar{R}(y)]}{\sqrt{\sum_{i=1}^n [R_i(x) - \bar{R}(x)]^2 \sum_{i=1}^n [R_i(y) - \bar{R}(y)]^2}}$$

donde

$$\begin{aligned} \bar{R}(x) = \bar{R}(y) &= \frac{\sum_{i=1}^n R_i(x)}{n} = \frac{\sum_{i=1}^n R_i(y)}{n} = \\ &= \frac{1 + 2 + \dots + n}{n} = \frac{(1+n)n}{2n} = \frac{n+1}{2} \end{aligned}$$

por otra parte¹⁶

$$\sum_{i=1}^n R_i^2(x) = \sum_{i=1}^n R_i^2(y) = 1^2 + 2^2 + \dots + n^2 = \frac{n(n+1)(2n+1)}{6}$$

con lo que

$$\begin{aligned} \sum_{i=1}^n [R_i(x) - \bar{R}(x)]^2 &= \sum_{i=1}^n [R_i(y) - \bar{R}(y)]^2 = \\ &= \sum_{i=1}^n R_i^2(x) - n\bar{R}^2(x) = \\ &= \frac{n(n+1)(2n+1)}{6} - n\left(\frac{n+1}{2}\right)^2 = \\ &= \frac{n^3 - n}{12}. \end{aligned}$$

Si consideramos las diferencias

$$d_i = R_i(x) - R_i(y),$$

al ser

$$\bar{R}(x) = \bar{R}(y)$$

¹⁶ Véase MARTÍN PLIEGO y RUIZ-MAYA: *Fundamentos...* Apéndice Matemático.

entonces

$$\begin{aligned}
 \sum_{i=1}^n d_i^2 &= \sum_{i=1}^n [R_i(x) - R_i(y)]^2 = \\
 &= \sum_{i=1}^n \{[R_i(x) - \bar{R}(x)] - [R_i(y) - \bar{R}(y)]\}^2 = \\
 &= \sum_{i=1}^n [R_i(x) - \bar{R}(x)]^2 + \sum_{i=1}^n [R_i(y) - \bar{R}(y)]^2 - \\
 &\quad - 2 \sum_{i=1}^n [R_i(x) - \bar{R}(x)][R_i(y) - \bar{R}(y)]
 \end{aligned}$$

luego

$$\begin{aligned}
 \sum_{i=1}^n [R_i(x) - \bar{R}(x)][R_i(y) - \bar{R}(y)] &= \\
 &= \frac{\sum_{i=1}^n [R_i(x) - \bar{R}(x)]^2 + \sum_{i=1}^n [R_i(y) - \bar{R}(y)]^2 - \sum_{i=1}^n d_i^2}{2} = \\
 &= \frac{\frac{n^3 - n}{12} + \frac{n^3 - n}{12} - \sum_{i=1}^n d_i^2}{2} = \frac{n^3 - n}{12} - \frac{\sum_{i=1}^n d_i^2}{2}.
 \end{aligned}$$

y el coeficiente de correlación por rangos de Spearman resulta igual a

$$r_s = \frac{\frac{n^3 - n}{12} - \frac{\sum_{i=1}^n d_i^2}{2}}{\sqrt{\frac{n^3 - n}{12} \cdot \frac{n^3 - n}{12}}} = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n}.$$

Se demuestra que la esperanza del estadístico r_s bajo la hipótesis nula de inco-relación, es cero

$$E(r_s / H_0) = 0.$$

Cuando existe un grado perfecto de asociación entre los rangos se cumple que, para todo i ,

$$R_i(x) = R_i(y)$$

lo que implica que $d_i = 0$ para cada par de observaciones y , por tanto, $r_s = 1$.

Cuando los rangos son equidistantes, dándose la igualdad $R_i(x) = n + 1 - R_i(y)$, por ejemplo

$R_i(x)$	4	2	3	5	1
$R_i(y)$	2	4	3	1	5

se obtiene que $r_s = -1$.

El estadístico r_s sigue una distribución simétrica respecto al valor cero, siendo su campo de variación el intervalo $[-1; 1]$

$$-1 \leq r_s \leq 1$$

Dado que la hipótesis nula se refiere a la ausencia de asociación, un valor muestral de r_s que se aparte demasiado de cero esperamos que contradiga la hipótesis nula y, por tanto, la evidencia empírica conduce a rechazar esta hipótesis H_0 .

La región crítica del contraste la formarán valores de r_s próximos a 1 o a -1, con lo que esta región crítica será bilateral, determinándose a partir de la relación

$$P(|r_s| \geq K/H_0) = \alpha,$$

siendo α el nivel de significación.

Cuando $n \geq 10$ puede utilizarse la distribución asintótica propuesta por Kendall, dada a través de la variable aleatoria

$$t(n-2) = \frac{r_s \sqrt{n-2}}{\sqrt{1-r_s^2}}$$

que sigue una ley t de Student con $n-2$ grados de libertad.

En caso de existir valores de x_i y/o y_i repetidos y, por consiguiente, con el mismo rango, el estadístico r_s debe ser corregido, aunque si las repeticiones no son abundantes, su efecto no resulta significativo.

EJEMPLO 6

Se somete a 15 estudiante a un test de conocimientos sobre matemáticas y música, obteniendo las siguiente puntuaciones

<i>Estudiante</i>	<i>Puntuación en matemáticas</i>	<i>Puntuación en música</i>
1	4,2	F
2	5,0	K
3	8,3	D
4	2,5	G
5	7,4	E
6	6,1	H
7	7,5	C
8	3,4	L
9	1,5	Ñ
10	6,5	J
11	8,8	A
12	3,3	M
13	6,3	I
14	9,0	B
15	5,8	N

La escala de puntuaciones en música es ordinal y sigue el orden inverso del abecedario.

Verifíquese la hipótesis nula de ausencia de asociación entre las puntuaciones de ambas materias. Nivel de significación: $\alpha = 0,05$.

Para determinar el valor muestral del estadístico en este ejemplo, formamos la tabla de la página siguiente, siendo

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n} = 1 - \frac{6 \cdot 124}{15^3 - 15} = 0,779.$$

Como $n = 15 > 10$ utilizamos la distribución asintótica dada por

$$t(n - 2) = \frac{r_s \sqrt{n - 2}}{\sqrt{1 - r_s^2}} = \frac{0,779 \sqrt{15 - 2}}{\sqrt{1 - 0,779^2}} = 4,479.$$

La región crítica es

$$P(|t(n - 2)| \geq K/H_0) = 0,05$$

Estudiante	Puntuación en matemáticas (x_i)	Puntuación en música (y_i)	$R_i(x)$	$R_i(y)$	d_i	d_i^2
1	4,2	F	11	6	5	25
2	5,0	K	10	11	-1	1
3	8,3	D	3	4	-1	1
4	2,5	G	14	7	7	49
5	7,4	E	5	5	0	0
6	6,1	H	8	8	0	0
7	7,5	C	4	3	1	1
8	3,4	L	12	12	0	0
9	1,5	Ñ	15	15	0	0
10	6,5	J	6	10	-4	16
11	8,8	A	2	1	1	1
12	3,3	M	13	13	0	0
13	6,3	I	7	9	-2	4
14	9,0	B	1	2	-1	1
15	5,8	N	9	14	-5	25

donde para $n = 15$, el valor tabular de una t de Student con 13 grados de libertad es $K = 2,160$, siendo la región de rechazo

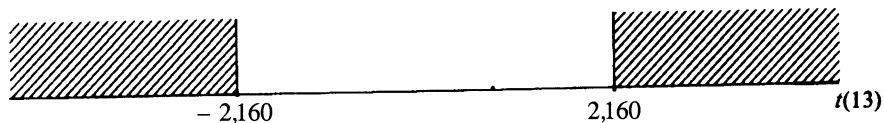


FIGURA 9.3

Como en la muestra $t(13) = 4,58 > 2,160$ se rechaza la hipótesis nula de ausencia de asociación entre las puntuaciones de matemáticas y música. La evidencia muestral parece indicar que, por el contrario, dichas puntuaciones están asociadas en los alumnos.

9.6 Tablas de Contingencia

Una tabla de contingencia $r \times c$ es una tabla de doble entrada donde se recogen la frecuencia conjunta n_{ij} con que aparece el nivel i -ésimo de un factor A y el nivel j -ésimo de un factor B . El número de niveles de A es r , siendo c el de B . La disposición de la tabla es la siguiente

FRECUENCIAS CONJUNTAS OBSERVADAS

		Factor B						Frecuencias marginales
		B ₁	B ₂	...	B _j	...	B _c	
Factor A	A ₁	n ₁₁	n ₁₂	...	n _{1j}	...	n _{1c}	n _{1•}
	A ₂	n ₂₁	n ₂₂	...	n _{2j}	...	n _{2c}	n _{2•}
	⋮	⋮	⋮	...	⋮	...	⋮	⋮
	A _i	n _{i1}	n _{i2}	...	n _{ij}	...	n _{ic}	n _{i•}
	⋮	⋮	⋮	...	⋮	...	⋮	⋮
	A _r	n _{r1}	n _{r2}	...	n _{rj}	...	n _{rc}	n _{r•}
Frecuencias marginales		n _{•1}	n _{•2}	...	n _{•j}	...	n _{•c}	n

donde se cumplen las igualdades

$$\sum_{i=1}^r n_{ij} = n_{•j} \quad \sum_{j=1}^c n_{ij} = n_{i•}$$

$$\sum_{i=1}^r n_{i•} = \sum_{j=1}^c n_{•j} = \sum_{i=1}^r \sum_{j=1}^c n_{ij} = n$$

los niveles de los factores A y B se identifican con modalidades o categorías del atributo de carácter cualitativo o con clases o intervalos de una variable cuantitativa.

Aunque los diseños muestrales que pueden dar origen a una tabla de contingencia $r \times c$ son variados,¹⁷ trataremos sólo los contrastes más típicos de independencia y de homogeneidad.

9.6.1. CONTRASTE DE INDEPENDENCIA

La hipótesis nula que se quiere contrastar es la de independencia entre los niveles del factor A respecto a los niveles del factor B.

¹⁷ Véase RUIZ-MAYA y otros, donde se efectúa un análisis completo de estos diseños en tablas de contingencia $r \times c$, exponiéndose además el tratamiento de dichas tablas a través de modelos logarítmico lineales y de la metodología de los modelos de respuesta discreta.

Si $p_{ij} = P(A_i \cap B_j)$, la hipótesis nula de independencia se especifica¹⁸ como

$$p_{ij} = P(A_i \cap B_j) = P(A_i) \cdot P(B_j) = p_{i\cdot} \cdot p_{\cdot j}$$

para todo $i, j : (i = 1, \dots, r; j = 1, \dots, c)$.

La distribución conjunta de los factores A y B contiene $r + c - 2$ parámetros desconocidos, pues como se tiene que verificar

$$\sum_{i=1}^r p_{i\cdot} = 1 \quad \sum_{j=1}^c p_{\cdot j} = 1$$

uno de los parámetros $p_{i\cdot}$ y otro de los parámetros $p_{\cdot j}$ se obtienen por diferencia.

Estimamos dichos parámetros bajo la condición de independencia impuesta por la hipótesis nula

$$p_{ij} = p_{i\cdot} \cdot p_{\cdot j}, \quad \text{para todo } i, j$$

para lo que formamos la función de verosimilitud

$$L(p_{i\cdot}, p_{\cdot j}) = \prod_{i=1}^r \prod_{j=1}^c p_{ij}^{n_{ij}} = \prod_{i=1}^r \prod_{j=1}^c (p_{i\cdot} \cdot p_{\cdot j})^{n_{ij}}$$

siendo su logaritmo

$$\ln L(p_{i\cdot}, p_{\cdot j}) = \sum_{i=1}^r \sum_{j=1}^c n_{ij} \ln (p_{i\cdot} \cdot p_{\cdot j}).$$

Para obtener los valores de $p_{i\cdot}$ y $p_{\cdot j}$ que hacen máxima esta expresión consideramos las restricciones

$$\sum_{i=1}^r p_{i\cdot} = 1 \quad \sum_{j=1}^c p_{\cdot j} = 1$$

y el lagrangiano

$$M(p_{i\cdot}, p_{\cdot j}) = \sum_{i=1}^r \sum_{j=1}^c (n_{ij} \ln p_{i\cdot} + n_{ij} \ln p_{\cdot j}) - \gamma_1 \left(\sum_{i=1}^r p_{i\cdot} - 1 \right) - \gamma_2 \left(\sum_{j=1}^c p_{\cdot j} - 1 \right)$$

¹⁸ Véase MARTÍN PLIEGO y RUÍZ-MAYA. *Fundamentos...* Capítulo 1.

donde

$$\left. \begin{aligned} \frac{\partial M}{\partial p_{i\cdot}} &= \sum_{j=1}^c n_{ij} \frac{1}{p_{i\cdot}} - \gamma_1 = \frac{n_{i\cdot}}{p_{i\cdot}} - \gamma_1 = 0 \\ \frac{\partial M}{\partial p_{\cdot j}} &= \sum_{i=1}^r n_{ij} \frac{1}{p_{\cdot j}} - \gamma_2 = \frac{n_{\cdot j}}{p_{\cdot j}} - \gamma_2 = 0 \\ \frac{\partial M}{\partial \gamma_1} &= \sum_{i=1}^r p_{i\cdot} - 1 = 0 \\ \frac{\partial M}{\partial \gamma_2} &= \sum_{j=1}^c p_{\cdot j} - 1 = 0 \end{aligned} \right\}$$

con lo que

$$\begin{aligned} \frac{n_{i\cdot}}{p_{i\cdot}} &= \gamma_1; & \frac{n_{i\cdot}}{\gamma_1} &= p_{i\cdot} \\ \frac{n_{\cdot j}}{p_{\cdot j}} &= \gamma_2; & \frac{n_{\cdot j}}{\gamma_2} &= p_{\cdot j} \end{aligned}$$

sumando respecto a i y respecto a j

$$\begin{aligned} \frac{1}{\gamma_1} \sum_{i=1}^r n_{i\cdot} &= \sum_{i=1}^r p_{i\cdot} = 1 \\ \frac{1}{\gamma_2} \sum_{j=1}^c n_{\cdot j} &= \sum_{j=1}^c p_{\cdot j} = 1 \end{aligned}$$

es decir, $\gamma_1 = \gamma_2 = n$.

Obteniéndose los estimadores máximo-verosímiles de $p_{i\cdot}$ y $p_{\cdot j}$

$$\begin{aligned} \hat{p}_{i\cdot} &= \frac{n_{i\cdot}}{n} \\ \hat{p}_{\cdot j} &= \frac{n_{\cdot j}}{n} \end{aligned}$$

Para efectuar el contraste de independencia utilizamos el estadístico X^2 , como medida de las desviaciones entre la muestra y la hipótesis poblacional

$$X^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}}$$

siendo \hat{E}_{ij} las frecuencias estimadas bajo el supuesto que la hipótesis nula es cierta, es decir

$$\hat{E}_{ij} = n\hat{p}_{ij} = n\hat{p}_{i\cdot}\hat{p}_{\cdot j} = n \frac{n_{i\cdot}}{n} \frac{n_{\cdot j}}{n} = \frac{n_{i\cdot}n_{\cdot j}}{n}$$

Este estadístico, como vimos en el epígrafe 9.2.1.2, sigue asintóticamente una distribución χ^2 de Pearson con los siguientes grados de libertad:

$$\begin{aligned} \text{g.l.} &= \text{Total clases de observaciones} - \text{n}^\circ \text{ de parámetros estimados} - 1 = \\ &= (r \cdot c) - (r + c - 2) - 1 = \\ &= rc - r - c + 1 = \\ &= r(c - 1) - (c - 1) = \\ &= (r - 1)(c - 1) \end{aligned}$$

pues el número total de clases o grupos de observaciones es igual al número total de celdas o casillas de la tabla de contingencia $r \cdot c$.

La región crítica para el contraste de independencia se determina a través de la expresión

$$P(\chi^2[(r - 1)(c - 1)] \geq K/H_0) = \alpha,$$

siendo α el nivel de significación.

EJEMPLO 7

Estúdiase si existe asociación entre el nivel educativo de los individuos y la preferencia por un determinado medio de comunicación, a partir de la información muestral:

FRECUENCIAS OBSERVADAS

		Medio de comunicación			Frecuencias marginales
		Prensa	Radio	T. V.	
Nivel educativo	Básico	15	10	25	50
	Medio	40	25	45	110
	Superior	45	30	55	130
Frecuencias marginales		100	65	125	290

Nivel de significación $\alpha = 2,5\%$.

Al tomar como hipótesis nula la independencia entre los diferentes niveles de las variables medio de comunicación y nivel educativo, la presencia de asociación se puede mantener si se rechaza esta hipótesis.

Para hallar el valor muestral del estadístico

$$X^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}}$$

se deben calcular las frecuencias estimadas \hat{E}_{ij}

$$\hat{E}_{11} = \frac{n_{1 \cdot} \cdot n_{\cdot 1}}{n} = \frac{50 \cdot 100}{290} = 17,24$$

$$\hat{E}_{12} = \frac{n_{1 \cdot} \cdot n_{\cdot 2}}{n} = \frac{50 \cdot 65}{290} = 11,21$$

$$\hat{E}_{21} = \frac{n_{2 \cdot} \cdot n_{\cdot 1}}{n} = \frac{110 \cdot 100}{290} = 37,93$$

$$\hat{E}_{22} = \frac{n_{2 \cdot} \cdot n_{\cdot 2}}{n} = \frac{110 \cdot 65}{290} = 24,66$$

$$\hat{E}_{13} = n_{1 \cdot} - \hat{E}_{11} - \hat{E}_{12} = 50 - 17,24 - 11,21 = 21,55$$

$$\hat{E}_{23} = n_{2 \cdot} - \hat{E}_{21} - \hat{E}_{22} = 110 - 37,93 - 24,66 = 47,41$$

$$\hat{E}_{31} = n_{\cdot 1} - \hat{E}_{11} - \hat{E}_{21} = 100 - 17,24 - 37,93 = 44,83$$

$$\hat{E}_{32} = n_{\cdot 2} - \hat{E}_{12} - \hat{E}_{22} = 65 - 11,21 - 24,66 = 29,13$$

$$\hat{E}_{33} = n_{\cdot 3} - \hat{E}_{13} - \hat{E}_{23} = 125 - 21,55 - 47,41 = 56,04$$

la tabla de frecuencias estimadas, de acuerdo con la hipótesis de independencia, es

FRECUENCIAS ESTIMADAS

		Medio de comunicación			Frecuencias marginales
		Prensa	Radio	T. V.	
Nivel educativo	Básico	17,24	11,21	21,55	50,00
	Medio	37,93	24,66	47,41	110,00
	Superior	44,83	29,13	56,04	130,00
Frecuencias marginales		100,00	65,00	125,00	290,00

y el valor muestral del estadístico X^2 se obtiene teniendo en cuenta los resultados siguientes:

n_{ij}	\hat{E}_{ij}	$n_{ij} - \hat{E}_{ij}$	$(n_{ij} - \hat{E}_{ij})^2$	$\frac{(n_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}}$
15	17,24	-2,24	5,0176	0,2910
10	11,21	-1,21	1,4641	0,1306
25	21,55	3,45	11,9025	0,5523
40	37,93	2,07	4,2849	0,1130
25	24,66	0,34	0,1156	0,0047
45	47,41	-2,41	5,8081	0,1225
45	44,83	0,17	0,0289	0,0006
30	29,13	0,87	0,7569	0,0260
55	53,04	-1,04	1,0816	0,0193
290	290,00			1,2600

Estadístico que sigue una distribución asintótica χ^2 de Pearson con $(r - 1)(c - 1) = (3 - 1)(3 - 1) = 4$ grados de libertad, y la región crítica viene dada por

$$P(\chi^2(4) \geq K/H_0) = \alpha = 0,025$$

siendo el valor tabular $K = 11,143$.

Como $X^2 = 1,26 < 11,143$, no puede rechazarse la hipótesis de independencia entre los niveles de estas dos variables y, por consiguiente, la evidencia empírica no marca asociación entre los factores analizados.

9.6.2. CONTRASTE DE HOMOGENEIDAD

Se dispone de c muestras clasificadas según los r niveles de una variable A y se quiere contrastar la hipótesis que las c muestras proceden de la misma población.

La información relativa a las frecuencias conjuntas observadas se dispone de acuerdo con una tabla de doble entrada como la siguiente

		Muestras						Frecuencias marginales
		1	2	...	j	...	c	
Variable A	A_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1c}	$n_{1\bullet}$
	A_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2c}	$n_{2\bullet}$
	\vdots	\vdots	\vdots	...	\vdots	...	\vdots	\vdots
	A_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{ic}	$n_{i\bullet}$
	\vdots	\vdots	\vdots	...	\vdots	...	\vdots	\vdots
	A_r	n_{r1}	n_{r2}	...	n_{rj}	...	n_{rc}	$n_{r\bullet}$
Tamaño de las muestras		$n_{\bullet 1}$	$n_{\bullet 2}$...	$n_{\bullet j}$...	$n_{\bullet c}$	n

que formalmente es una tabla de contingencia, hecha la salvedad del distinto origen de formación.

La frecuencia n_{ij} es el número de elementos de la muestra j que pertenecen al nivel i -ésimo de A ; $n_{\cdot j}$ el número total de observaciones de la muestra j -ésima y $n_{i\cdot}$ el número total de elementos de todas las muestras que pertenecen al nivel i -ésimo de A , siendo $n = \sum_{j=1}^c n_{\cdot j}$ el número total de observaciones del conjunto de las c muestras.

Contrastar la hipótesis nula H_0 que todas las muestras proceden de la misma población equivale a contrastar la hipótesis que, para todas las muestras, se verifica

$$P(A_1) = p_1, \dots, P(A_r) = p_r$$

verificándose que

$$\sum_{i=1}^r p_i = 1.$$

Para realizar el contraste se utiliza el estadístico X^2 como medida de las desviaciones entre la muestra y la población siendo

$$X^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}},$$

donde E_{ij} son las frecuencias estimadas de acuerdo con la hipótesis nula, es decir,

$$\hat{E}_{ij} = n_{\cdot j} \hat{p}_i.$$

Como los parámetros p_i son desconocidos se procede a su estimación máximo-verosímil, \hat{p}_i , siendo como vimos en el epígrafe 9.2.2.,

$$\hat{p}_i = \frac{n_{i\cdot}}{n}$$

con lo que la estimación de las frecuencias esperadas es

$$\hat{E}_{ij} = n_{\cdot j} \hat{p}_i = \frac{n_{\cdot j} n_{i\cdot}}{n},$$

siendo el estadístico del contraste

$$X^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}}$$

que se distribuye asintóticamente según una χ^2 con los siguientes grados de libertad:

$$\begin{aligned} \text{g.l.} &= n^\circ \text{ de muestras [n}^\circ \text{ de clases} - 1] - n^\circ \text{ de parámetros estimados} = \\ &= c(r - 1) - (r - 1) = \\ &= (r - 1)(c - 1) \end{aligned}$$

pues, por la restricción $\sum_{i=1}^r p_i = 1$, sólo son $r - 1$ los parámetros que se estiman.

La región crítica está determinada por la expresión

$$P(\chi^2 [(r - 1)(c - 1)] \geq K/H_0) = \alpha,$$

siendo α el nivel de significación.¹⁹

EJEMPLO 8

Unos naranjos se abonan con dos marcas distintas de fertilizantes F_1 y F_2 : 50 naranjos con el fertilizante F_1 y 60 con F_2 , resultando que unos naranjos aumentaron su producción, otros la disminuyeron y otros no la variaron.

Contraste la hipótesis nula de que ambos tipos de fertilizantes producen los mismos efectos con un nivel de significación $\alpha = 0,10$, teniendo en cuenta que la información muestral es la siguiente

FRECUENCIAS OBSERVADAS

		Muestra con		Frecuencias marginales
		F_1	F_2	
Producción	+	20	35	55
	=	20	15	35
	-	10	10	20
Tamaño de las muestras		50	60	110

La hipótesis nula consiste en que los dos fertilizantes producen efectos homogéneos o, lo que es lo mismo, que ambas muestras proceden de una misma población.

¹⁹ Aunque el diseño del contraste de homogeneidad difiere del contraste de independencia, formalmente, por el estadístico utilizado y región crítica, ambos tests son similares.

Iniciamos, en primer lugar, el cálculo de las frecuencias estimadas \hat{E}_{ij} bajo el supuesto de homogeneidad, es decir, teniendo en cuenta que

$$\hat{E}_{ij} = \frac{n_{i \cdot} \cdot n_{\cdot j}}{n}$$

por lo que

$$\hat{E}_{11} = \frac{n_{1 \cdot} \cdot n_{\cdot 1}}{n} = \frac{55 \cdot 50}{110} = 25$$

$$\hat{E}_{21} = \frac{n_{2 \cdot} \cdot n_{\cdot 1}}{n} = \frac{35 \cdot 50}{110} = 15,9$$

$$\hat{E}_{31} = n_{\cdot 1} - \hat{E}_{11} - \hat{E}_{21} = 50 - 25 - 15,9 = 9,1$$

$$\hat{E}_{12} = n_{1 \cdot} - \hat{E}_{11} = 55 - 25 = 30$$

$$\hat{E}_{22} = n_{2 \cdot} - \hat{E}_{21} = 35 - 15,9 = 19,1$$

$$\hat{E}_{32} = n_{3 \cdot} - \hat{E}_{31} = 20 - 9,1 = 10,9$$

y la tabla de frecuencias estimadas la que mostramos en la página siguiente

FRECUENCIAS ESTIMADAS

		Muestra con		Frecuencias marginales
		F_1	F_2	
Producción	+	25,0	30,0	55
	=	15,9	19,1	35
	-	9,1	10,9	20
Total		50,0	60,0	110

El valor muestral del estadístico

$$X^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}}$$

se obtiene según indica la tabla siguiente:

n_{ij}	\hat{E}_{ij}	$n_{ij} - \hat{E}_{ij}$	$(n_{ij} - \hat{E}_{ij})^2$	$\frac{(n_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}}$
20	25,0	-5,0	25,00	1,00
35	30,0	5,0	25,00	1,20
20	15,9	4,1	16,81	1,06
15	19,1	-4,1	16,81	0,88
10	9,1	0,9	0,81	0,09
10	10,9	-0,9	0,81	0,07
110	110,0			4,30

siguiendo el estadístico del contraste asintóticamente una distribución χ^2 con

$$(r - 1)(c - 1) = (3 - 1)(2 - 1) = 2 \text{ grados de libertad;}$$

por consiguiente, la región crítica es

$$P(\chi^2(2) \geq K/H_0) = \alpha = 0,10$$

y el valor crítico K el correspondiente a las tablas de la χ^2 con 2 grados de libertad y una probabilidad igual a 0,10, $K = 4,605$.

Como en la muestra conjunta $\chi^2 = 4,30 < 4,605$ no se rechaza la hipótesis nula de que ambos fertilizantes producen efectos similares.

CAPÍTULO 10

Introducción a la teoría de la decisión e inferencia bayesiana

10.1 Teoría de la decisión

En gran parte de los fenómenos físicos, biológicos y sociales se desconocen las leyes que regulan su comportamiento por lo que no es arriesgado pensar que estén gobernados por leyes de naturaleza estadística.

Bajo esta premisa, a todo decisor se le plantea el problema de elegir entre las **acciones alternativas** posibles en cada caso, sin tener el conocimiento preciso para optar con certeza por la acción más adecuada, basándose en la mayor o menor especificación que realice sobre las leyes de naturaleza estadística que han de reemplazar a las **leyes deterministas** desconocidas sobre dicho fenómeno.

En general, una **ley estadística** estará representada por la distribución de probabilidad $F(x; \theta)$ del fenómeno analizado donde, para cada posible valor del parámetro θ , se obtiene uno u otro comportamiento concreto del fenómeno analizado. En la terminología de este enfoque, el conjunto de valores que puede tomar el parámetro θ recibe el nombre de **Estados de la Naturaleza**.

Sucintamente, se exponen los componentes fundamentales de un **proceso de decisión**:

- Conjunto de **acciones, alternativas o estrategias** que el decisor puede adoptar. Se representa por $A = \{a_1, \dots, a_h\}$.
- Sin control por parte del decisor, condicionando el éxito o fracaso de la toma de decisiones, se halla el **Estado de la Naturaleza** conjunto de los valores de los parámetros o parámetro (según que la ley estadística dependa de más de un parámetro o de sólo uno) que caracterizan cada situación del fenómeno (en términos inferenciales, cada hipótesis sobre la población), designándose el espacio paramétrico como $\Theta = \{\theta_1, \dots, \theta_k\}$.
- Derivados de la elección de una determinada acción, según cual sea el Estado de la Naturaleza θ , se producen una serie de resultados que se representan por $l(\theta; a)$, donde cada valor será la **pérdida** en que se incurre cuando el decisor toma la acción a y la Naturaleza se encuentra en el estado θ . La expresión $l(\theta; a)$ recibe el nombre de **función de pérdida**. Una pérdida negativa es equivalente a una ganancia.

Suponiendo que las acciones y los Estados de la Naturaleza son finitos, el proceso de decisión se recoge esquemáticamente en el cuadro siguiente:

		<i>Estados de la naturaleza</i>			
		θ_1	θ_2	...	θ_k
<i>Acciones</i>	a_1	$l(\theta_1; a_1)$	$l(\theta_2; a_1)$...	$l(\theta_k; a_1)$
	a_2	$l(\theta_1; a_2)$	$l(\theta_2; a_2)$...	$l(\theta_k; a_2)$
	\vdots	\vdots	\vdots	\ddots	\vdots
	a_h	$l(\theta_1; a_h)$	$l(\theta_2; a_h)$...	$l(\theta_k; a_h)$

EJEMPLO 1

Consideremos el esquema de decisión resumido en el cuadro

		<i>Estados de la naturaleza</i>	
		$\theta_1 =$ llueve	$\theta_2 =$ no llueve
<i>Acciones</i>	No salir de casa (a_1)	3	3
	Salir sin paraguas (a_2)	5	0
	Salir con paraguas (a_3)	1	4

La función de pérdida está explicitada numéricamente; por ejemplo, si se sale sin paraguas y no llueve, la pérdida es $l(\theta_2; a_2) = 0$, y si se sale con paraguas y llueve, entonces, $l(\theta_1; a_3) = 1$.

Las pérdidas pueden medirse en términos de utilidad negativa de cada combinación Acción-Estado de la Naturaleza; cuantificando el coste económico, en caso de ser factible, o asignándole valores, que optimicen alguna medida deseable, según veremos más adelante.

10.1.1. TIPOLOGÍA DE LA DECISIÓN

Los problemas de decisión se pueden clasificar atendiendo a un doble criterio:

A. Por el conocimiento que el decisor tenga acerca de los Estados de la Naturaleza:

- **Decisión con certidumbre.** El decisor dispone de conocimiento exacto sobre cuál es el Estado de la Naturaleza en que tiene que decidir. Este caso no lo analizaremos aquí, pues su resolución se reduce a un problema de optimización matemática.
- **Decisión en ambiente de incertidumbre.** Cada Estado de la Naturaleza θ_i se presenta aleatoriamente, sin que resulte factible determinar la probabilidad con que lo hace. No es posible, por tanto, probabilizar los Estados de la Naturaleza; sólo se conocen las distintas situaciones en que puede situarse la Naturaleza.
- **Decisión con riesgo.** Se establece a priori una ley de probabilidad para los Estados de la Naturaleza incluidos en Θ , de tal manera que

$$p_i = P(\theta = \theta_i)$$

$$\sum_{i=1}^k P(\theta = \theta_i) = 1$$

es decir, se reconoce un comportamiento estocástico sobre Θ .

B. Según se disponga de alguna información sobre el fenómeno bajo estudio:

- **Decisión sin experimentación.** No se dispone de ningún tipo de información sobre el fenómeno (no existe información muestral sobre la población analizada).
- **Decisión con experimentación.** Se dispone de información suministrada por una muestra aleatoria simple de tamaño n , $\mathbf{X} (x_1, \dots, x_n)$.

10.1.2. DECISIÓN SIN EXPERIMENTACIÓN

Sólo se exponen los criterios más relevante en esta situación.

10.1.2.1. Decisión en ambiente de incertidumbre: Criterio minimax

Este criterio está basado en que el decisor presenta una actitud de precaución manifiesta pues selecciona la acción cuya pérdida máxima sea la menor. Es decir, en principio, el decisor se sitúa para cada acción en la peor de las situaciones, optando por la acción que presente la menor pérdida máxima; por tanto, selecciona la acción a_j tal que

$$\min_a \max_{\theta} l(\theta_i; a_j)$$

EJEMPLO 2

En el ejemplo propuesto anterior se tiene que:

$$\max_{\theta} l(\theta_i; a_1) = 3$$

$$\max_{\theta} l(\theta_i; a_2) = 5$$

$$\max_{\theta} l(\theta_i; a_3) = 4$$

luego, como

$$\min_a \max_{\theta} l(\theta_i; a_j) = 3,$$

opta por la acción a_1 (no salir de casa).

10.1.2.2. Decisión con riesgo: Solución de Bayes

Se supone, como antes se señaló, que el espacio paramétrico, conjunto de los Estados de la Naturaleza, es probabilizable y, por tanto, es posible asignar probabilidades a cada valor θ_i que pertenezca a Θ .

Bajo este supuesto, se podrá determinar, para cada acción, la **pérdida esperada** definida por

$$B(a_j) = E[l(\theta_i; a_j)] = \sum_{i=1}^k l(\theta_i; a_j) P(\theta = \theta_i).$$

La **solución o acción de Bayes** es la acción cuya pérdida esperada sea menor.

EJEMPLO 3

En el mismo ejemplo anterior se considera que

$$p_1 = P(\theta = \theta_1) = 0,3 \quad \text{y} \quad p_2 = P(\theta = \theta_2) = 0,7$$

las pérdidas esperadas son

$$B(a_1) = l(\theta_1; a_1) \cdot p_1 + l(\theta_2; a_1) \cdot p_2 = 3 \cdot 0,3 + 3 \cdot 0,7 = 3$$

$$B(a_2) = l(\theta_1; a_2) \cdot p_1 + l(\theta_2; a_2) \cdot p_2 = 5 \cdot 0,3 + 0 \cdot 0,7 = 1,5$$

$$B(a_3) = l(\theta_1; a_3) \cdot p_1 + l(\theta_2; a_3) \cdot p_2 = 1 \cdot 0,3 + 4 \cdot 0,7 = 3,1$$

donde a_2 es la acción que presenta la menor pérdida esperada y, por tanto, la que se adopta como solución de Bayes, de este problema de decisión para la distribución a priori $g(\theta)$ establecida.

10.1.3. DECISIÓN CON EXPERIMENTACIÓN

Para facilitar y mejorar la adopción de decisiones se toma una muestra aleatoria simple de tamaño n , $\mathbf{X} (x_1, \dots, x_n)$, que proporciona información acerca del Estado de la Naturaleza θ , lo que incidirá en la elección de la estrategia más adecuada.

Antes de proceder al análisis de los criterios más usuales es preciso tener en cuenta los siguientes conceptos.

10.1.3.1. Función de decisión

La función de decisión es una función de la información muestral \mathbf{X} , representada por $d(\mathbf{X})$, y que define una aplicación del espacio muestral \mathfrak{X} sobre el espacio de las posibles acciones a adoptar.

Una función de decisión podría ser, fijado un valor constante k para un espacio de acciones $A = \{a_1, a_2\}$, la siguiente

$$d(\mathbf{X}) = \sum_{i=1}^n x_i < k \Rightarrow a_1$$

$$d(\mathbf{X}) = \sum_{i=1}^n x_i \geq k \Rightarrow a_2$$

10.1.3.2. Función de pérdida

Se ha visto que el éxito o fracaso en la toma de una decisión se mide a través de la función de pérdida.

En este caso, con la información muestral disponible, la función de pérdida es

$$l(\theta; a) = l[\theta; d(\mathbf{X})] = l(\theta; d).$$

La función de pérdida a utilizar debe ser tal que aumente cuando el error en que se pueda incurrir al tomar una decisión crezca, y disminuya si lo hace el error. Bajo este sencillo principio se justifica la utilización de las siguientes funciones de pérdida, por otra parte las más habituales:

- Si no se quiere cometer grandes errores

$$l(\theta; d) = c(d - \theta)^2,$$

siendo c una constante positiva.

- Si se pretende que la pérdida sea proporcional al error

$$l(\theta; d) = c |d - \theta|$$

con c positivo.

10.1.3.3. Función de riesgo

La función de decisión y la de pérdida dependen de la muestra \mathbf{X} , por lo que tanto una función como otra son variables aleatorias.

En este sentido, la función de riesgo es la esperanza matemática de la función de pérdida, calculada para la distribución de la muestra \mathbf{X} considerando θ como un valor fijo, es decir,

$$R(\theta; d) = E[l(\theta; d)]$$

donde, si la distribución de la muestra \mathbf{X} es continua, se obtiene

$$R(\theta; d) = E[l(\theta; d)] = \int_{\mathbf{X}} l[\theta; d(\mathbf{X})] f(\mathbf{X}/\theta) d\mathbf{X}.$$

Si existiera una función de riesgo $R(\theta; d_1)$ tal que

$$R(\theta; d_1) \leq R(\theta; d_2) \quad \text{para todo } \theta$$

representando d_2 cualquiera otra función de decisión, d_1 sería la función de decisión óptima para tomar decisiones.

Como habitualmente encontraremos una función de decisión cuyo riesgo siempre sea el menor para todo valor del parámetro θ , será preciso recurrir a criterios de compromiso, como los que a continuación se exponen.

10.1.3.4. Decisión en ambiente de incertidumbre: Criterio minimax

Se elige la relación de decisión donde se obtuvo

$$\min_d \max_{\theta} R(\theta; d),$$

es decir, se adopta la función de decisión cuyo riesgo máximo sea menor.

Gráficamente, si tenemos tres funciones de decisión posibles como

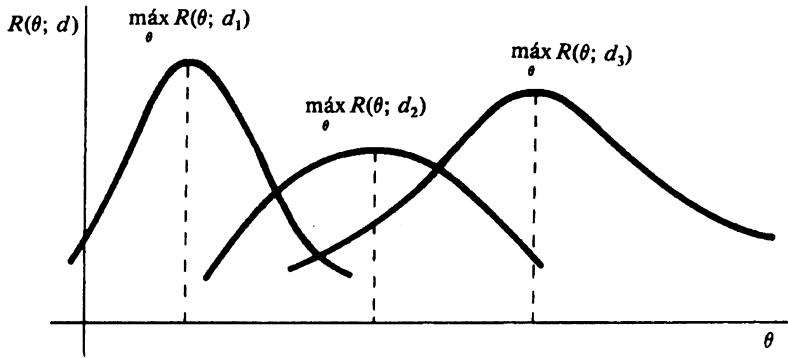


FIGURA 10.1

se utilizará, según el criterio, la función d_2 .

Si la función de decisión se utiliza para asignar algún valor al parámetro θ , bajo la información muestral que proporciona \mathbf{X} , la función de decisión minimax recibe el nombre de **estimador minimax**, confundiendo el espacio de las acciones con el espacio paramétrico, es decir, $A = \Theta$.

EJEMPLO 4

Sea una población con función de densidad

$$f(x/\theta) = \frac{1}{\theta} e^{-\frac{1}{\theta}x}; \quad x \geq 0$$

definiendo el Estado de la Naturaleza (o espacio paramétrico) por $\theta > 0$.

Con referencia a una muestra aleatoria simple de tamaño n se proponen las siguientes funciones de decisión para estimar θ

$$d_1(\mathbf{X}) = \bar{x}$$

$$d_2(\mathbf{X}) = x_n$$

Elíjase la solución minimax si la función de pérdida es cuadrática

$$l(\theta; d) = (d - \theta)^2.$$

Fácilmente puede calcularse que en la población

$$\mu = E(\xi) = \theta$$

$$\sigma^2 = V(\xi) = \theta^2$$

Al ser la muestra aleatoria simple se verifica

$$E[d_1(\mathbf{X})] = E(\bar{x}) = \mu = \theta$$

$$E[d_2(\mathbf{X})] = E(x_n) = \mu = \theta$$

y los riesgos son

$$\begin{aligned} R(\theta; d_1) &= E[l(\theta; d_1)] = E(d_1 - \theta)^2 = E(\bar{x} - \theta)^2 = \\ &= V(\bar{x}) = \frac{\sigma^2}{n} = \frac{\theta^2}{n} \end{aligned}$$

$$\begin{aligned} R(\theta; d_2) &= E[l(\theta; d_2)] = E(d_2 - \theta)^2 = E(x_n - \theta)^2 = \\ &= V(x_n) = V(X) = \theta^2 \end{aligned}$$

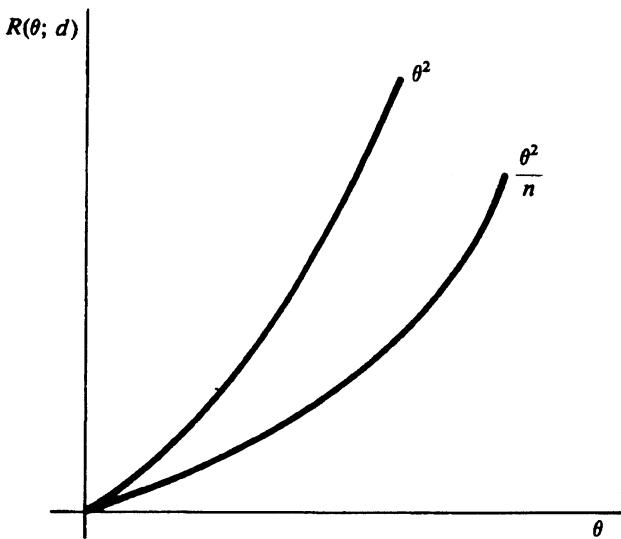


FIGURA 10.2

las funciones de riesgo aparecen en la figura 10.2, obteniéndose que si la función de decisión adoptada es $d_1(\mathbf{X})$

$$\min_d \max_{\theta} R(\theta; d), \quad \text{para todo } \theta$$

o, lo que es lo mismo, el estimador minimax del parámetro θ es $\theta^* = \bar{x}$, para la función de pérdida utilizada.

10.1.3.5. Decisión con riesgo: Decisión de Bayes

El esquema de decisión que se analizará a continuación se basa simultáneamente en:

- La información suministrada por una muestra aleatoria simple de tamaño n , \mathbf{X} .
- La asignación de una distribución de probabilidad a priori sobre el espacio paramétrico, que representamos por $g(\theta)$.

Bajo estas premisas, el riesgo $R(\theta; d)$ es una variable aleatoria que dependerá de la variable aleatoria θ .

Se entiende por **Riesgo de Bayes** el valor esperado de la función de riesgo $R(\theta; d)$ calculado sobre la distribución a priori de θ . Si se denota por $B(d)$ tendremos:

- Para $g(\theta)$ con distribución discreta

$$B(d) = E[R(\theta; d)] = \sum_{\Theta} R(\theta_i; d) P(\theta = \theta_i).$$

- Para $g(\theta)$ con distribución continua

$$B(d) = E[R(\theta; d)] = \int_{\Theta} R(\theta; d) g(\theta) d\theta.$$

La **decisión de Bayes** es la decisión que presente un riesgo de Bayes menor, la que minimice

$$\begin{aligned} B(d) &= E[R(\theta; d)] = \int_{\Theta} R(\theta; d) g(\theta) d\theta = \\ &= \int_{\Theta} \left[\int_{\mathbf{X}} l[\theta; d(\mathbf{X})] f(\mathbf{X}/\theta) d\mathbf{X} \right] g(\theta) d\theta = \\ &= \int_{\Theta} \int_{\mathbf{X}} l[\theta; d(\mathbf{X})] f(\mathbf{X}/\theta) g(\theta) d\theta d\mathbf{X} \end{aligned}$$

Veremos en el epígrafe 10.3 que el estimador de Bayes es precisamente la decisión de Bayes que se acaba de proponer. Estudiaremos, por tanto, el método más operativo para determinar este estimador y sus propiedades, comenzando en el siguiente epígrafe con la determinación las distribuciones de probabilidad asignables al parámetro θ de acuerdo con la concepción Bayesiana.

10.2 Inferencia bayesiana: distribuciones a priori y a posteriori

A lo largo de los capítulos anteriores se han recogido técnicas y procedimientos de la denominada «inferencia estadística clásica» donde, en los casos en que se exponían métodos basados en parámetros poblacionales, se suponía que éstos eran desconocidos pero no aleatorios.

La diferencia fundamental que plantea el «enfoque inferencial bayesiano» radica en que considera el parámetro desconocido θ como variable aleatoria, a la que se asigna, por tanto, una determinada distribución de probabilidad, que recibe el nombre de **distribución a priori** y se representa por $g(\theta)$.

La información suministrada por una muestra aleatoria simple \mathbf{X} , bajo la óptica bayesiana, puede cambiar la idea del comportamiento probabilístico que se tuviera sobre el parámetro θ y, en este sentido, es posible aceptar que exista una **distribución a posteriori** del parámetro θ , representada por $g(\theta/\mathbf{X})$, donde se recoge la modificación de la distribución de probabilidad de dicho parámetro θ cuando se dispone de la información muestral \mathbf{X} .

Si la distribución a priori del parámetro θ se postula de tipo discreto, aplicando el teorema de Bayes pueden determinarse las correspondientes probabilidades a posteriori para cada uno de los valores que tome el parámetro θ . En efecto

$$P(\theta = \theta_i/\mathbf{X}) = \frac{P(\mathbf{X}/\theta = \theta_i) P(\theta = \theta_i)}{\sum_{i=1}^k P(\mathbf{X}/\theta = \theta_i) P(\theta = \theta_i)}.$$

Si la distribución a priori $g(\theta)$ es continua, y la población de donde se extrae la muestra también, puede obtenerse la distribución a posteriori del parámetro θ teniendo en cuenta que:

- $f(\mathbf{X}; \theta)$ representa la función de densidad conjunta del vector aleatorio muestral \mathbf{X} y del parámetro θ .
- $f(\mathbf{X}/\theta)$ es la función de densidad de la muestra \mathbf{X} para un θ dado.
- $f(\mathbf{X})$ es la distribución marginal de la muestra \mathbf{X} .

Como se sabe

$$f(\mathbf{X}/\theta) = \frac{f(\mathbf{X}; \theta)}{g(\theta)}$$

$$g(\theta/\mathbf{X}) = \frac{f(\mathbf{X}; \theta)}{f(\mathbf{X})}$$

donde

$$f(\mathbf{X}; \theta) = f(\mathbf{X}/\theta) g(\theta)$$

$$f(\mathbf{X}; \theta) = g(\theta/\mathbf{X}) f(\mathbf{X})$$

por lo que

$$f(\mathbf{X}/\theta) g(\theta) = g(\theta/\mathbf{X}) f(\mathbf{X})$$

y la distribución a posteriori del parámetro θ resulta

$$g(\theta/\mathbf{X}) = \frac{f(\mathbf{X}/\theta) g(\theta)}{f(\mathbf{X})}$$

obteniéndose la densidad marginal $f(\mathbf{X})$ como

$$f(\mathbf{X}) = \int_{\Theta} f(\mathbf{X}; \theta) d\theta = \int_{\Theta} f(\mathbf{X}/\theta) g(\theta) d\theta$$

llegándose, finalmente, a que

$$g(\theta/\mathbf{X}) = \frac{f(\mathbf{X}/\theta) g(\theta)}{\int_{\Theta} f(\mathbf{X}/\theta) g(\theta) d\theta}$$

versión continua del Teorema de Bayes.

10.3 Estimación de Bayes

El **estimador de Bayes** es la función de decisión, establecida sobre el espacio paramétrico Θ , que produzca el menor riesgo esperado.

El estimador será $\hat{\theta} = d^*(\mathbf{X})$ tal que minimice la expresión

$$B(d) = \int_{\Theta} \int_{\mathbf{X}} l[\theta; d(\mathbf{X})] f(\mathbf{X}/\theta) g(\theta) d\theta d\mathbf{X}.$$

Existe un procedimiento operativo para determinar la decisión de Bayes sin tener que describir todo el espacio muestral (y, por tanto, trabajar en él).

En el epígrafe anterior establecimos la distribución a posteriori del parámetro θ , calculada como

$$g(\theta/\mathbf{X}) = \frac{f(\mathbf{X}/\theta) g(\theta)}{f(\mathbf{X})}$$

por consiguiente, el riesgo de Bayes queda

$$\begin{aligned} B(d) &= \int_{\Theta} \int_{\mathbf{X}} l[\theta; d(\mathbf{X})] f(\mathbf{X}/\theta) g(\theta) d\theta d\mathbf{X} = \\ &= \int_{\Theta} \int_{\mathbf{X}} l[\theta; d(\mathbf{X})] g(\theta/\mathbf{X}) f(\mathbf{X}) d\theta d\mathbf{X} = \\ &= \int_{\mathbf{X}} f(\mathbf{X}) \left[\int_{\Theta} l[\theta; d(\mathbf{X})] g(\theta/\mathbf{X}) d\theta \right] d\mathbf{X} = \\ &= \int_{\mathbf{X}} f(\mathbf{X}) E_{\Theta/\mathbf{X}} [l(\theta; d(\mathbf{X}))] d\mathbf{X} \end{aligned}$$

siendo $E_{\Theta/\mathbf{X}} \{l[\theta; d(\mathbf{X})]\}$, riesgo de Bayes a posteriori, es decir,

$$E_{\Theta/\mathbf{X}} \{l[\theta; d(\mathbf{X})]\} = \int_{\Theta} l[\theta; d(\mathbf{X})] g(\theta/\mathbf{X}) d\theta. \quad [1]$$

Dado que la función de decisión $d(\mathbf{X})$ que minimiza $B(d)$ es la misma que hace mínima la expresión [1], pues $d(\mathbf{X})$ no depende de $f(\mathbf{X})$, puede utilizarse esta última para determinar el estimador de Bayes, de manera que dicho estimador será aquel que verifique

$$\begin{aligned} \frac{\partial E_{\Theta/\mathbf{X}} [l(\theta; d(\mathbf{X}))]}{\partial d(\mathbf{X})} &= \psi[\theta; d(\mathbf{X})] = 0 \Rightarrow \theta^* = d^*(\mathbf{X}) \\ \left[\frac{\partial^2 E_{\Theta/\mathbf{X}} [l(\theta; d(\mathbf{X}))]}{\partial d^2(\mathbf{X})} \right]_{d(\mathbf{X})=d^*(\mathbf{X})} &> 0 \end{aligned}$$

EJEMPLO 5

Determinése el estimador de Bayes para el parámetro θ de una distribución de Poisson, en muestras aleatorias simples de tamaño n , supuesta una distribución a priori de θ dada por

$$g(\theta) = 3e^{-3\theta}, \quad \theta > 0$$

siendo la pérdida cuadrática

$$l[\theta; d(\mathbf{X})] = l(\theta; d) = (d - \theta)^2.$$

Para determinar la distribución a posteriori $g(\theta/\mathbf{X})$ calculamos, previamente,

$$f(\mathbf{X}/\theta) = f(x_1/\theta) \cdots f(x_n/\theta) = \frac{e^{-\theta}\theta^{x_1}}{x_1!} \cdots \frac{e^{-\theta}\theta^{x_n}}{x_n!} = \frac{e^{-n\theta} \theta^{\sum_{i=1}^n x_i}}{x_1! \cdots x_n!}$$

siendo

$$\begin{aligned} f(\mathbf{X}) &= \int_0^\infty f(\mathbf{X}/\theta) g(\theta) d\theta = \int_0^\infty \frac{e^{-n\theta} \theta^{\sum_{i=1}^n x_i}}{x_1! \cdots x_n!} 3e^{-3\theta} d\theta = \\ &= \frac{3}{x_1! \cdots x_n!} \int_0^\infty \theta^{\sum_{i=1}^n x_i} e^{-(n+3)\theta} d\theta = \frac{3}{x_1! \cdots x_n!} \int_0^\infty \theta^s e^{-(n+3)\theta} d\theta \end{aligned}$$

donde

$$s = \sum_{i=1}^n x_i$$

que siempre será un número natural, dados los posibles valores x_i que puede tomar esta población con distribución de Poisson.

Por tanto,

$$f(\mathbf{X}) = \frac{3}{x_1! \cdots x_n!} \cdot \frac{\Gamma(s+1)}{(n+3)^{s+1}}$$

y la distribución a posteriori de θ

$$\begin{aligned} g(\theta/\mathbf{X}) &= \frac{f(\mathbf{X}/\theta) g(\theta)}{f(\mathbf{X})} = \\ &= \frac{\frac{e^{-n\theta}\theta^s}{x_1! \cdots x_n!} 3e^{-3\theta}}{\frac{3}{x_1! \cdots x_n!} \cdot \frac{\Gamma(s+1)}{(n+3)^{s+1}}} = \frac{\theta^s (n+3)^{s+1} e^{-(n+3)\theta}}{\Gamma(s+1)} = \\ &= \frac{(n+3)^{s+1} \theta^s e^{-(n+3)\theta}}{s!}, \quad \theta > 0 \end{aligned}$$

es una distribución gamma

$$G(s+1; n+3).$$

La función que habrá que minimizar es

$$\begin{aligned}
 \Phi &= E_{\theta/X} [l(\theta; d(\mathbf{X}))] = \int_0^\infty l(\theta; d) g(\theta/X) d\theta = \\
 &= \int_0^\infty (d - \theta)^2 \frac{(n+3)^{s+1} \theta^s e^{-(n+3)\theta}}{s!} d\theta = \\
 &= \int_0^\infty (d^2 - 2d\theta + \theta^2) \frac{(n+3)^{s+1} \theta^s e^{-(n+3)\theta}}{s!} d\theta = \\
 &= \frac{(n+3)^{s+1}}{s!} \left[d^2 \frac{\Gamma(s+1)}{(n+3)^{s+1}} - 2d \frac{\Gamma(s+2)}{(n+3)^{s+2}} + \frac{\Gamma(s+3)}{(n+3)^{s+3}} \right] = \\
 &= \frac{1}{s!} \left[d^2 \cdot s! - 2d \frac{(s+1)s!}{n+3} + \frac{(s+2)(s+1)s!}{(n+3)^2} \right] = \\
 &= d^2 - 2d \frac{s+1}{n+3} + \frac{(s+2)(s+1)}{(n+3)^2}
 \end{aligned}$$

y

$$\frac{\partial \Phi}{\partial d} = 2d - 2 \frac{s+1}{n+3} = 0$$

siendo

$$\frac{s+1}{n+3} = \frac{\sum_{i=1}^n x_i + 1}{n+3} = d^*(\mathbf{X})$$

comprobándose que

$$\frac{\partial^2 \Phi}{\partial d^2} = 2 > 0$$

luego

$$\theta^* = d^*(\mathbf{X}) = \frac{\sum_{i=1}^n x_i + 1}{n+3}$$

es el estimador de Bayes que hace mínimo el riesgo esperado.

TEOREMAS

1. Si la función de riesgo $R[\theta; d(\mathbf{X})]$ es constante respecto a θ , el estimador de Bayes coincide con el estimador minimax.

En efecto,

$$B(d) = E_{\Theta} \{R[\theta; d(\mathbf{X})]\} = R[\theta; d(\mathbf{X})]$$

y la función $\theta^* = d^*(\mathbf{X})$ que minimiza $B(d)$ también es la que determina el menor riesgo $R[\theta; d(\mathbf{X})]$.

2. Si la función de pérdida es de tipo cuadrático, el estimador de Bayes es el valor medio de la distribución a posteriori,

$$\theta^* = E_{\Theta/\mathbf{X}} [\theta/\mathbf{X}].$$

En efecto, el estimador de Bayes $\theta^* = d^*(\mathbf{X})$ es el que hace mínima la expresión

$$\begin{aligned} E_{\Theta/\mathbf{X}} \{l[\theta; d(\mathbf{X})]\} &= \int_{\Theta} l[\theta; d(\mathbf{X})] g(\theta/\mathbf{X}) d\theta = \\ &= \int_{\Theta} (d(\mathbf{X}) - \theta)^2 g(\theta/\mathbf{X}) d\theta = \\ &= E_{\Theta/\mathbf{X}} \{[\theta - d(\mathbf{X})]^2/\mathbf{X}\} \end{aligned}$$

y teniendo en cuenta las propiedades de la esperanza matemática¹

$$\begin{aligned} \min_d E_{\Theta/\mathbf{X}} \{l[\theta; d(\mathbf{X})]\} &= \min_d E_{\Theta/\mathbf{X}} \{[\theta - d(\mathbf{X})]^2/\mathbf{X}\} = \\ &= E_{\Theta/\mathbf{X}} \{[\theta - E_{\Theta/\mathbf{X}}(\theta/\mathbf{X})]^2/\mathbf{X}\} \end{aligned}$$

luego

$$d^*(\mathbf{X}) = E_{\Theta/\mathbf{X}}(\theta/\mathbf{X}).$$

3. Análogamente, si la función de pérdida es proporcional

$$l(\theta; d) = |\theta - d(\mathbf{X})|,$$

el estimador de Bayes es la mediana de la distribución a posteriori de θ dado \mathbf{X} , pues el mínimo del valor medio de las desviaciones absolutas respecto al origen de referencia se obtiene cuando se toma la mediana como ese origen de referencia.

¹ Véase MARTÍN PLIEGO y RUIZ-MAYA: *Fundamentos...* Capítulo 4.

10.4 Contrastación bayesiana de hipótesis

Al igual que en la estimación bayesiana, la contrastación bayesiana se desarrolla bajo el supuesto que el experimentador puede asignar una distribución de probabilidad al parámetro θ .

En esta situación, se desea contrastar la hipótesis nula $H_0 (\theta \in \Theta_0)$ frente a la alternativa $H_1 (\theta \in \Theta_1)$, cumpliéndose que

$$\Theta_0 \cup \Theta_1 = \Theta \quad \text{y} \quad \Theta_0 \cap \Theta_1 = \emptyset .$$

Las probabilidades a priori de que sean ciertas cada una de estas hipótesis son

$$P(\theta \in \Theta_0) = g(\theta_0)$$

$$P(\theta \in \Theta_1) = g(\theta_1) = 1 - g(\theta_0)$$

El contraste de hipótesis se puede plantear como un problema típico de decisión, tal como

		<i>Estados de la Naturaleza</i>	
		Θ_0	Θ_1
<i>Decisión</i>	d_1 : Rechazar H_0	$l(\Theta_0; d_1)$	$l(\Theta_1; d_1)$
	d_2 : No rechazar H_0	$l(\Theta_0; d_2)$	$l(\Theta_1; d_2)$

Tomamos como función de pérdida

$$l(\theta; d) = \begin{cases} 0 & \text{si no se incurre en error} \\ 1 & \text{si se comete error} \end{cases}$$

y la de decisión

$$d(\mathbf{X} \in C) = d_1$$

$$d(\mathbf{X} \in C^*) = d_2$$

siendo C la región crítica y C^* la de aceptación.

La función de riesgo resulta

$$\begin{aligned} R(\theta_0; d) &= E[l(\theta_0; d)] = \\ &= 1 \cdot P(\text{Rechazar } H_0/\Theta_0) + 0 \cdot P(\text{No rechazar } H_0/\Theta_0) = \\ &= P(\text{Rechazar } H_0/\Theta_0) = P(\mathbf{X} \in C/\Theta_0) = \alpha \end{aligned}$$

$$\begin{aligned} R(\theta_1; d) &= E[l(\theta_1; d)] = \\ &= 0 \cdot P(\text{Rechazar } H_0/\Theta_1) + 1 \cdot P(\text{No rechazar } H_0/\Theta_1) = \\ &= P(\text{No rechazar } H_0/\Theta_1) = P(\mathbf{X} \in C^*/\Theta_1) = \beta \end{aligned}$$

con lo que el riesgo esperado, o riesgo de Bayes, para la distribución a priori de θ es igual a

$$\begin{aligned} B(d) &= E[R(\theta; d)] = \\ &= g(\theta_0) \alpha + g(\theta_1) \beta = \\ &= g(\theta_0) P(\mathbf{X} \in C/\Theta_0) + g(\theta_1) P(\mathbf{X} \in C^*/\Theta_1) = \\ &= g(\theta_0) \int_C f(\mathbf{X}/\Theta_0) d\mathbf{X} + g(\theta_1) \int_{C^*} f(\mathbf{X}/\Theta_1) d\mathbf{X} = \\ &= g(\theta_0) \int_C f(\mathbf{X}/\Theta_0) d\mathbf{X} + g(\theta_1) \left[1 - \int_C f(\mathbf{X}/\Theta_1) d\mathbf{X} \right] = \\ &= g(\theta_1) + \int_C [g(\theta_0) f(\mathbf{X}/\Theta_0) - g(\theta_1) f(\mathbf{X}/\Theta_1)] d\mathbf{X}, \end{aligned}$$

como $g(\theta_1)$ está fijado a priori y es positivo, el mínimo del riesgo de Bayes se obtendrá para una región crítica C tal que

$$g(\theta_0) f(\mathbf{X}/\Theta_0) - g(\theta_1) f(\mathbf{X}/\Theta_1) < 0$$

es decir, cuando

$$g(\theta_0) f(\mathbf{X}/\Theta_0) < g(\theta_1) f(\mathbf{X}/\Theta_1).$$

Por otra parte, como

$$f(\mathbf{X}/\theta) g(\theta) = g(\theta/\mathbf{X}) f(\mathbf{X})$$

la condición anterior se transforma, en términos de distribución a posteriori, en

$$g(\theta_0/\mathbf{X}) f(\mathbf{X}) < g(\theta_1/\mathbf{X}) f(\mathbf{X})$$

es decir,

$$g(\theta_0/\mathbf{X}) < g(\theta_1/\mathbf{X}).$$

Consecuentemente, se rechazará la hipótesis nula H_0 si la probabilidad de que el parámetro pertenezca a Θ_0 , dada la muestra \mathbf{X} , es menor que la probabilidad de que dicho parámetro pertenezca a Θ_1 dada \mathbf{X} cuando

$$P(\theta \in \Theta_0/\mathbf{X}) < P(\theta \in \Theta_1/\mathbf{X})$$

Puede comprobarse que las reglas equivalentes de decisión en la contrastación bayesiana de hipótesis, bien en función de la distribución a priori del parámetro θ en términos de su distribución posteriori, no necesitan que se fije de antemano un determinado nivel de significación α , pues las probabilidades de que el parámetro se encuentre en la hipótesis nula o en la alternativa pueden ser evaluadas directamente.

EJEMPLO 6

En una población $B(1; \theta)$ se contrasta

$$H_0 \left[\theta \leq \frac{1}{4} \right] \quad \text{frente a} \quad H_1 \left[\theta > \frac{1}{4} \right].$$

Se supone que la distribución a priori del parámetro θ es uniforme en el intervalo $[0; 1]$.

Realícese el contraste bayesiano de hipótesis para una muestra aleatoria de tamaño uno.

Sabemos que

$$f(\mathbf{X}/\theta) = f(x_1/\theta) = \theta^{x_1} (1-\theta)^{1-x_1}, \quad x_1 = \{0; 1\}$$

como

$$g(\theta) = 1, \quad 0 \leq \theta \leq 1$$

entonces

$$f(\mathbf{X}) = f(x_1) = \int_0^1 f(\mathbf{X}/\theta) g(\theta) d\theta = \int_0^1 \theta^{x_1} (1-\theta)^{1-x_1} d\theta$$

donde, al ser $x_1 = \{0; 1\}$,

$$f(x_1 = 0) = \int_0^1 (1-\theta) d\theta = \left[\theta - \frac{\theta^2}{2} \right]_0^1 = \frac{1}{2}$$

$$f(x_1 = 1) = \int_0^1 \theta d\theta = \left[\frac{\theta^2}{2} \right]_0^1 = \frac{1}{2}$$

es decir,

$$f(x_1) = \frac{1}{2} \quad \text{para} \quad x_1 = \{0; 1\}$$

por tanto, como

$$g(\theta/\mathbf{X}) = \frac{f(\mathbf{X}/\theta) g(\theta)}{f(\mathbf{X})}$$

se tiene que

$$g(\theta/x_1) = \frac{\theta^{x_1} (1-\theta)^{1-x_1}}{1/2} = 2\theta^{x_1} (1-\theta)^{1-x_1}, \quad 0 < \theta < 1$$

Como la muestra x_1 puede tomar el valor cero o el uno

$$g(\theta/x_1 = 0) = 2(1-\theta), \quad 0 < \theta < 1$$

$$g(\theta/x_1 = 1) = 2\theta, \quad 0 < \theta < 1$$

siendo

$$P(\theta \leq \theta_0/x_1 = 0) = \int_0^{\theta_0} 2(1-\theta) d\theta = 2\left(\theta_0 - \frac{\theta_0^2}{2}\right)$$

$$P(\theta \leq \theta_0/x_1 = 1) = \int_0^{\theta_0} 2\theta d\theta = \theta_0^2$$

y

$$P(\theta > \theta_0/x_1 = 0) = 1 - 2\left(\theta_0 - \frac{\theta_0^2}{2}\right)$$

$$P(\theta > \theta_0/x_1 = 1) = 1 - \theta_0^2$$

Si la muestra es $x_1 = 0$, como para la hipótesis nula $H_0\left[\theta \leq \frac{1}{4}\right]$

$$P\left(\theta \leq \frac{1}{4}/x_1 = 0\right) = 1 - 2\left(\frac{1}{4} - \frac{(\frac{1}{4})^2}{2}\right) = \frac{7}{16}$$

y para la hipótesis alternativa

$$P\left(\theta > \frac{1}{4}/x_1 = 0\right) = 1 - \frac{7}{16} = \frac{9}{16}$$

y dado que

$$\frac{7}{16} < \frac{9}{16},$$

entonces se rechazará la hipótesis nula $H_0\left[\theta \leq \frac{1}{4}\right]$.

Ejercicios resueltos

EJERCICIO 10.1 Considérese el esquema de decisión cuya matriz de pérdidas es

		Estados de la naturaleza	
		$\theta = \theta_1$	$\theta = \theta_2$
Acciones	a_1	0	5
	a_2	5	0
	a_3	2	6
	a_4	3	4

Si la distribución a priori de θ es

$$P(\theta = \theta_1) = p$$

$$P(\theta = \theta_2) = 1 - p$$

estúdiense la solución de Bayes.

SOLUCIÓN. Las pérdidas esperadas para cada posible acción son

$$B(a_1) = 0p + 5(1 - p) = 5 - 5p$$

$$B(a_2) = 5p + 0(1 - p) = 5p$$

$$B(a_3) = 2p + 6(1 - p) = 6 - 4p$$

$$B(a_4) = 3p + 4(1 - p) = 4 - p$$

Representamos gráficamente en la página siguiente las pérdidas esperadas, $B(a_i)$, en función del parámetro p .

Como p es tal que $0 < p < 1$ entonces, dentro de este campo de variación, las pérdidas esperadas mínimas se obtienen si se adoptan las siguientes opciones

$$0 < p \leq \frac{1}{2} \Rightarrow \min_a B(a) = B(a_2) \Rightarrow \text{se elige } a_2$$

$$\frac{1}{2} \leq p < 1 \Rightarrow \min_a B(a) = B(a_1) \Rightarrow \text{se elige } a_1$$

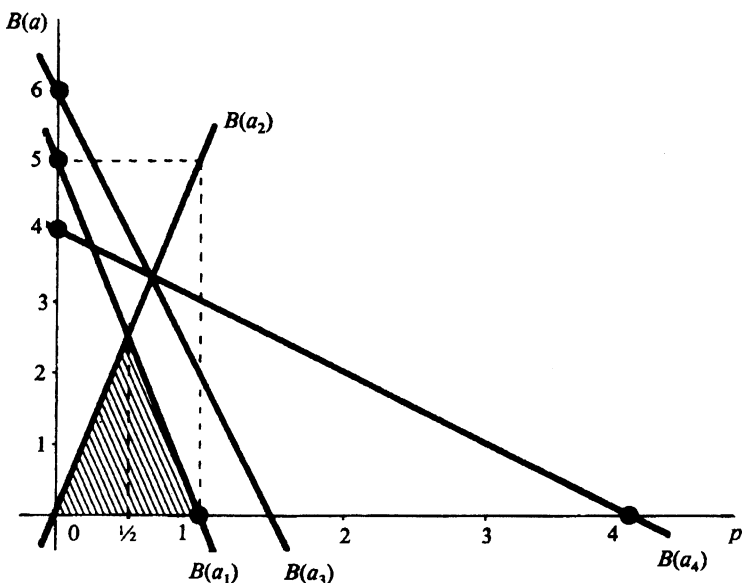


FIGURA 10.3

EJERCICIO 10.2 Determinése, a partir de una muestra aleatoria simple de tamaño n , la estimación bayesiana del parámetro θ de una distribución binomial $B(1; \theta)$, si se asigna una distribución a priori uniforme en el intervalo $[0; 1]$.

La función de pérdida se fija como

$$l(\theta; d) = \frac{(\theta - d)^2}{\theta(1 - \theta)}$$

SOLUCIÓN. Para hallar la distribución a posteriori $g(\theta/\mathbf{X})$ procedemos previamente a calcular

$$\begin{aligned} f(\mathbf{X}/\theta) &= f(x_1/\theta) \cdots f(x_n/\theta) = \\ &= \theta^{x_1}(1 - \theta)^{1-x_1} \cdots \theta^{x_n}(1 - \theta)^{1-x_n} = \\ &= \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i} = \theta^s (1 - \theta)^{n-s} \end{aligned}$$

siendo

$$s = \sum_{i=1}^n x_i$$

Al ser uniforme en $[0; 1]$ la distribución a priori

$$g(\theta) = 1, \quad 0 \leq \theta \leq 1$$

con lo que

$$\begin{aligned} f(\mathbf{X}) &= \int_0^1 f(\mathbf{X}/\theta) g(\theta) d\theta = \\ &= \int_0^1 \theta^s (1 - \theta)^{n-s} \cdot 1 d\theta = \\ &= \frac{\Gamma(s+1) \Gamma(n-s+1)}{\Gamma(n+2)} = \\ &= \frac{s! (n-s)!}{(n+1)!} \end{aligned}$$

y se deduce que la distribución a posteriori es

$$\begin{aligned} g(\theta/\mathbf{X}) &= \frac{f(\mathbf{X}/\theta) g(\theta)}{f(\mathbf{X})} = \\ &= \frac{\theta^s (1 - \theta)^{n-s}}{\frac{s! (n-s)!}{(n+1)!}} = \\ &= \frac{(n+1)!}{s! (n-s)!} \theta^s (1 - \theta)^{n-s}, \quad 0 \leq \theta \leq 1 \end{aligned}$$

que es la distribución beta $B(s+1; n-s+1)$.

Para obtener el estimador de Bayes habrá que hacer mínima la expresión

$$\begin{aligned} \Phi &= E_{\theta/\mathbf{X}} \{[l(\theta; d(\mathbf{X}))]\} = \int_{\Theta} l(\theta; d) g(\theta/\mathbf{X}) d\theta = \\ &= \int_0^1 \frac{(\theta - d)^2}{\theta(1 - \theta)} \cdot \frac{(n+1)!}{s! (n-s)!} \theta^s (1 - \theta)^{n-s} d\theta = \\ &= \frac{(n+1)!}{s! (n-s)!} \int_0^1 (\theta - d)^2 \theta^{s-1} (1 - \theta)^{n-s-1} d\theta = \\ &= \frac{(n+1)!}{s! (n-s)!} \int_0^1 (\theta^2 - 2d\theta + d^2) \theta^{s-1} (1 - \theta)^{n-s-1} d\theta = \\ &= \frac{(n+1)!}{s! (n-s)!} \left[\frac{\Gamma(s+2) \Gamma(n-s)}{\Gamma(n+2)} - 2d \frac{\Gamma(s+1) \Gamma(n-s)}{\Gamma(n+1)} + d^2 \frac{\Gamma(s) \Gamma(n-s)}{\Gamma(n)} \right] = \\ &= \frac{(n+1)!}{s! (n-s)!} \left[\frac{(s+1)! (n-s-1)!}{(n+1)!} - 2d \frac{s! (n-s-1)!}{n!} + d^2 \frac{(s-1)! (n-s-1)!}{(n-1)!} \right] = \\ &= \frac{s+1}{n-s} - 2d \frac{n+1}{n-s} + d^2 \frac{(n+1)n}{s(n-s)} \end{aligned}$$

derivando respecto a $d = d(\mathbf{X})$

$$\frac{\partial \Phi}{\partial d} = -2 \frac{n+1}{n-s} + 2d \frac{(n+1)n}{s(n-s)} = 0$$

y el estimador de Bayes resulta

$$\theta^* = d^*(\mathbf{X}) = \frac{s}{n} = \frac{\sum_{i=1}^n x_i}{n} = \bar{x}.$$

Comprobamos a continuación la condición de mínimo

$$\frac{d^2 \Phi}{\partial d^2} = 2 \frac{(n+1)n}{s(n-s)} > 0$$

pues $s < n$, dado que $s = \sum_{i=1}^n x_i$ y cada x_i es cero o uno en la distribución poblacional $B(1; \theta)$.

Sólo en el caso que $s = n$, $\bar{x} = 1$ y $\theta^* = 1$, la distribución $B(1; \theta)$ se convierte en una distribución degenerada, donde $P(\theta^* = 1) = 1$, y no existe suceso dicotómico alguno.

EJERCICIO 10.3 En una población $N(\theta; 1)$ se desea contrastar la hipótesis nula $H_0 [\theta \leq 0,8]$ frente a la alternativa $H_1 [\theta > 0,8]$. Para ello se toma una muestra aleatoria simple de tamaño $n = 24$ cuya media es 0,95.

Efectúese el contraste bayesiano de estas hipótesis si se supone que la distribución a priori de θ es $N(0; 1)$.

SOLUCIÓN. Podemos comprobar que la distribución a posteriori de θ , dada la muestra \mathbf{X} , tiene por expresión

$$\theta/\mathbf{X} \rightarrow N\left(\frac{n\bar{x}}{n+1}; \frac{1}{\sqrt{n+1}}\right)$$

con lo que, en nuestro caso,

$$N(0,912; 0,25).$$

Rechazaremos la hipótesis nula H_0 cuando

$$P(\theta \leq \theta_0/\mathbf{X}) = \frac{1}{2};$$

y al verificarse que

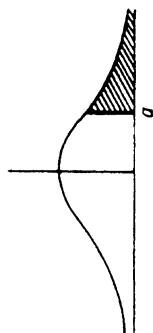
$$P(\theta \leq 0,8/\mathbf{X}) = P(0,25\xi + 0,912 \leq 0,8) = P(\xi \leq -0,45) = 0,3264 < \frac{1}{2}$$

rechazamos H_0 [$\theta \leq 0,8$].

TABLAS ESTADÍSTICAS

TABLA 1

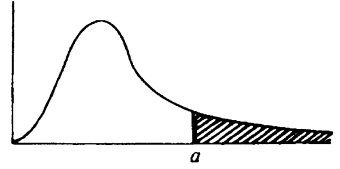
Distribución Normal (0; 1). $P(X \geq a)$



	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,4960	0,4920	0,4880	0,4841	0,4801	0,4761	0,4721	0,4681	0,4641
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
0,2	0,4207	0,4168	0,4129	0,4091	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
0,6	0,2743	0,2709	0,2676	0,2644	0,2611	0,2579	0,2546	0,2514	0,2483	0,2451
0,7	0,2420	0,2389	0,2358	0,2327	0,2297	0,2266	0,2236	0,2207	0,2177	0,2148
0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
1,0	0,1587	0,1563	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
1,2	0,1151	0,1131	0,1112	0,1094	0,1075	0,1057	0,1038	0,1020	0,1003	0,0985
1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
1,8	0,0359	0,0352	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233

TABLA 2

Distribución χ^2 . $P(\chi^2 \geq a)$

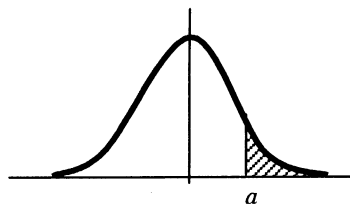


Grados de libertad	Probabilidades			
	0,10	0,05	0,025	0,01
1	2,706	3,841	5,024	6,635
2	4,605	5,991	7,378	9,210
3	6,251	7,815	9,348	11,345
4	7,779	9,488	11,143	13,277
5	9,236	11,070	12,833	15,086
6	10,645	12,592	14,449	16,812
7	12,017	14,067	16,013	18,475
8	13,362	15,507	17,535	20,090
9	14,684	16,919	19,023	21,666
10	15,987	18,307	20,483	23,209
11	17,275	19,675	21,920	24,725
12	18,549	21,026	23,337	26,217
13	19,812	22,362	24,736	27,688
14	21,064	23,685	26,119	29,141
15	22,307	24,996	27,488	30,578
16	23,542	26,296	28,845	32,000
17	24,769	27,587	30,191	33,409
18	25,989	28,869	31,526	34,805
19	27,204	30,143	32,852	36,191
20	28,412	31,410	34,170	37,566
21	29,615	32,670	35,479	38,932
22	30,813	33,924	36,781	40,289
23	32,007	35,172	38,076	41,638
24	33,196	36,415	39,364	42,080
25	34,382	37,652	40,646	44,314
26	35,563	38,885	41,923	45,642
27	36,741	40,113	43,194	46,963
28	37,916	41,337	44,461	48,278
29	39,087	42,557	45,722	49,588
30	40,256	43,773	46,979	50,892

Para valores elevados de n (grados de libertad) mayores que 30 pueden obtenerse las probabilidades a través de la transformación $\sqrt{2\chi^2(n)} - \sqrt{2n-1}$ que es aproximadamente $N(0; 1)$.

TABLA 3

Distribución *t* de Student . $P[t(n) \geq a]$



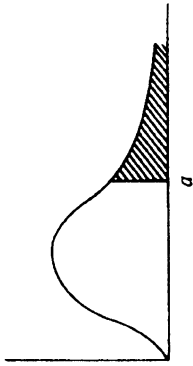
Grados de libertad	Probabilidades			
	0,10	0,05	0,025	0,01
1	3,0777	6,3138	12,7062	31,8205
2	1,8856	2,9200	4,3027	6,9646
3	1,6377	2,3534	3,1824	4,5407
4	1,5332	2,1318	2,7764	3,7469
5	1,4759	2,0150	2,5706	3,3649
6	1,4398	1,9432	2,4469	3,1427
7	1,4149	1,8946	2,3646	2,9980
8	1,3968	1,8595	2,3060	2,8965
9	1,3830	1,8331	2,2622	2,8214
10	1,3722	1,8125	2,2281	2,7638
11	1,3634	1,7959	2,2010	2,7181
12	1,3562	1,7823	2,1788	2,6810
13	1,3502	1,7709	2,1604	2,6503
14	1,3450	1,7613	2,1448	2,6245
15	1,3406	1,7531	2,1314	2,6025
16	1,3368	1,7459	2,1199	2,5835
17	1,3334	1,7396	2,1098	2,5669
18	1,3304	1,7341	2,1009	2,5524
19	1,3277	1,7291	2,0930	2,5395
20	1,3253	1,7247	2,0860	2,5280
21	1,3232	1,7207	2,0796	2,5176
22	1,3212	1,7171	2,0739	2,5083
23	1,3195	1,7139	2,0687	2,4999
24	1,3178	1,7109	2,0639	2,4922
25	1,3163	1,7081	2,0595	2,4851
26	1,3150	1,7056	2,0555	2,4786
27	1,3137	1,7033	2,0518	2,4727
28	1,3125	1,7011	2,0484	2,4671
29	1,3114	1,6991	2,0452	2,4620
30	1,3104	1,6973	2,0423	2,4573

Para valores elevados de *n* (grados de libertad) mayores que 30, la variable aleatoria *t*(*n*) es, aproximada-

mente, $N\left(0; \sqrt{\frac{n}{n-2}}\right)$.

TABLA 4 (Continuación)

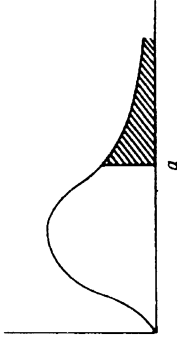
Distribución F . $P[F(m; n) \geq a] = 0,10$.



Grados de libertad del denominador	Grados de libertad del numerador																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	39,86	49,50	53,59	55,83	57,24	58,20	58,91	59,44	59,86	60,19	60,71	61,22	61,74	62,00	62,26	62,53	62,79	63,06	63,33
2	8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38	9,39	9,41	9,42	9,44	9,45	9,46	9,47	9,47	9,48	9,49
3	5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24	5,23	5,22	5,20	5,18	5,17	5,17	5,16	5,15	5,14	5,13
4	4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94	3,92	3,90	3,87	3,84	3,83	3,82	3,80	3,79	3,78	3,76
5	4,06	3,78	3,62	3,52	3,45	3,40	3,37	3,34	3,32	3,30	3,27	3,24	3,21	3,19	3,17	3,16	3,14	3,12	3,10
6	3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96	2,94	2,90	2,87	2,84	2,82	2,80	2,78	2,76	2,74	2,72
7	3,59	3,26	3,07	2,96	2,88	2,83	2,78	2,75	2,72	2,70	2,67	2,63	2,59	2,58	2,56	2,54	2,51	2,49	2,47
8	3,46	3,11	2,92	2,81	2,73	2,67	2,62	2,59	2,56	2,54	2,50	2,46	2,42	2,40	2,38	2,36	2,34	2,32	2,29
9	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44	2,42	2,38	2,34	2,30	2,28	2,25	2,23	2,21	2,18	2,16
10	3,29	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35	2,32	2,28	2,24	2,20	2,18	2,16	2,13	2,11	2,08	2,06
11	3,23	2,86	2,66	2,54	2,45	2,39	2,34	2,30	2,27	2,25	2,21	2,17	2,12	2,10	2,08	2,05	2,03	2,00	1,97
12	3,18	2,81	2,61	2,48	2,39	2,33	2,28	2,24	2,21	2,19	2,15	2,10	2,06	2,04	2,01	1,99	1,96	1,93	1,90
13	3,14	2,76	2,56	2,43	2,35	2,28	2,23	2,20	2,16	2,14	2,10	2,05	2,01	1,98	1,96	1,93	1,90	1,88	1,85
14	3,10	2,73	2,52	2,39	2,31	2,24	2,19	2,15	2,12	2,10	2,05	2,01	1,96	1,94	1,91	1,89	1,86	1,83	1,80

TABLA 4 (Continuación)

Distribución F . $P[F(m; n) \geq a] = 0,10$.



Grados de libertad del denominador	Grados de libertad del numerador																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
15	3,07	2,70	2,49	2,36	2,27	2,21	2,16	2,12	2,09	2,06	2,02	1,97	1,92	1,90	1,87	1,85	1,82	1,79	1,76
16	3,05	2,67	2,46	2,33	2,24	2,18	2,13	2,09	2,06	2,03	1,99	1,94	1,89	1,87	1,84	1,81	1,78	1,75	1,72
17	3,03	2,64	2,44	2,31	2,22	2,15	2,10	2,06	2,03	2,00	1,96	1,91	1,86	1,84	1,81	1,78	1,75	1,72	1,69
18	3,01	2,62	2,42	2,29	2,20	2,13	2,08	2,04	2,00	1,98	1,93	1,89	1,84	1,81	1,78	1,75	1,72	1,69	1,66
19	2,99	2,61	2,40	2,27	2,18	2,11	2,06	2,02	1,98	1,96	1,91	1,86	1,81	1,79	1,76	1,73	1,70	1,67	1,63
20	2,97	2,59	2,38	2,25	2,16	2,09	2,04	2,00	1,96	1,94	1,89	1,84	1,79	1,77	1,74	1,71	1,68	1,64	1,61
21	2,96	2,57	2,36	2,23	2,14	2,08	2,02	1,98	1,95	1,92	1,87	1,83	1,78	1,75	1,72	1,69	1,66	1,62	1,59
22	2,95	2,56	2,35	2,22	2,13	2,06	2,01	1,97	1,93	1,90	1,86	1,81	1,76	1,73	1,70	1,67	1,64	1,60	1,57
23	2,94	2,55	2,34	2,21	2,11	2,05	1,99	1,95	1,92	1,89	1,84	1,80	1,74	1,72	1,69	1,66	1,62	1,59	1,55
24	2,93	2,54	2,33	2,19	2,10	2,04	1,98	1,94	1,91	1,88	1,83	1,78	1,73	1,70	1,67	1,64	1,61	1,57	1,53
25	2,92	2,53	2,32	2,18	2,09	2,02	1,97	1,93	1,89	1,87	1,82	1,77	1,72	1,69	1,66	1,63	1,59	1,56	1,52
26	2,91	2,52	2,31	2,17	2,08	2,01	1,96	1,92	1,88	1,86	1,81	1,76	1,71	1,68	1,65	1,61	1,58	1,54	1,50
27	2,90	2,51	2,30	2,17	2,07	2,00	1,95	1,91	1,87	1,85	1,80	1,75	1,70	1,67	1,64	1,60	1,57	1,53	1,49
28	2,89	2,50	2,29	2,16	2,06	2,00	1,94	1,90	1,87	1,84	1,79	1,74	1,69	1,66	1,63	1,59	1,56	1,52	1,48
29	2,89	2,50	2,28	2,15	2,06	1,99	1,93	1,89	1,86	1,83	1,78	1,73	1,68	1,65	1,62	1,58	1,55	1,51	1,47
30	2,88	2,49	2,28	2,14	2,05	1,98	1,93	1,88	1,85	1,82	1,77	1,72	1,67	1,64	1,61	1,57	1,54	1,50	1,46
40	2,84	2,44	2,23	2,09	2,00	1,93	1,87	1,83	1,79	1,76	1,71	1,66	1,61	1,57	1,54	1,51	1,47	1,42	1,38
60	2,79	2,39	2,18	2,04	1,95	1,87	1,82	1,77	1,74	1,71	1,66	1,60	1,54	1,51	1,48	1,44	1,40	1,35	1,29
120	2,75	2,35	2,13	1,99	1,90	1,82	1,77	1,72	1,68	1,65	1,60	1,55	1,48	1,45	1,41	1,37	1,32	1,26	1,19
∞	2,71	2,30	2,08	1,94	1,85	1,77	1,72	1,67	1,63	1,60	1,55	1,49	1,42	1,38	1,34	1,30	1,24	1,17	1,10

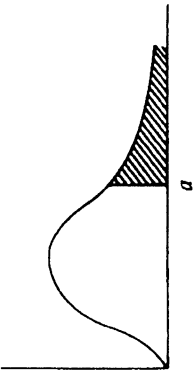
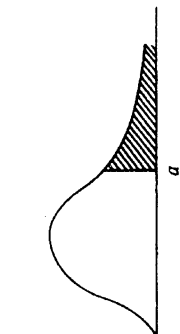


TABLA 4 (Continuación)
 Distribución F . $P[F(m; n) \geq a] = 0,05$.

Grados de libertad del denominador	Grados de libertad del numerador																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161,40	199,50	215,70	224,60	230,20	234,00	236,80	238,90	240,50	241,90	243,90	245,90	248,00	249,10	250,10	251,10	252,20	253,30	254,30
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,43	19,45	19,45	19,46	19,47	19,48	19,49	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,70	8,66	8,64	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	5,91	5,86	5,80	5,77	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	4,68	4,62	4,56	4,53	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	4,00	3,94	3,87	3,84	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	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,12	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	3,07	3,01	2,94	2,90	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	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13

TABLA 4 (Continuación)

Distribución F . $P[F(m; n) \geq a] = 0,05$.



Grados de libertad del denominador	Grados de libertad del numerador																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,13	2,05	2,01	1,96	1,91	1,86	1,81	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79	1,73
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77	1,71
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,07	1,99	1,95	1,90	1,85	1,80	1,75	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,13	2,06	1,97	1,93	1,88	1,84	1,79	1,73	1,67
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,04	1,96	1,91	1,87	1,82	1,77	1,71	1,65
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,10	2,03	1,94	1,90	1,85	1,81	1,75	1,70	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93	1,89	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	2,00	1,92	1,84	1,79	1,74	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	1,92	1,84	1,75	1,70	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	1,83	1,75	1,66	1,61	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	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22	1,00

TABLA 5

Valores críticos del test de Kolmogorov-Smirnov para una muestra

Tamaño muestral	mues- n	Nivel de significación α			
		0,10	0,05	0,02	0,01
1		0,95000	0,97500	0,99000	0,99500
2		0,77639	0,84189	0,90000	0,92929
3		0,63604	0,70760	0,78456	0,82900
4		0,56522	0,62394	0,68887	0,73424
5		0,50945	0,56328	0,62718	0,66853
6		0,46799	0,51926	0,57741	0,61661
7		0,43607	0,48342	0,53844	0,57581
8		0,40962	0,45427	0,50654	0,54179
9		0,38746	0,43001	0,47960	0,51332
10		0,36866	0,40925	0,45662	0,48893
11		0,35242	0,39122	0,43670	0,46770
12		0,33815	0,37543	0,41918	0,44905
13		0,32549	0,36143	0,40362	0,43247
14		0,31417	0,34890	0,38970	0,41762
15		0,30397	0,33760	0,37713	0,40420
16		0,29472	0,32733	0,36571	0,39201
17		0,28627	0,31796	0,35528	0,38086
18		0,27851	0,30936	0,34569	0,37062
19		0,27136	0,30143	0,33685	0,36117
20		0,26473	0,29408	0,32866	0,35241
21		0,25858	0,28724	0,32104	0,34427
22		0,25283	0,28087	0,31394	0,33666
23		0,24746	0,27490	0,30728	0,32954
24		0,24242	0,26931	0,30104	0,32286
25		0,23768	0,26404	0,29516	0,31657
26		0,23320	0,25907	0,28962	0,31064
27		0,22898	0,25438	0,28438	0,30502
28		0,22497	0,24993	0,27942	0,29971
29		0,22117	0,24571	0,27471	0,29466
30		0,21756	0,24170	0,27023	0,28987

Para tamaños muestrales $\hat{n} > 30$, el valor crítico D puede calcularse como $D = \sqrt{-\ln(\alpha/2)/2n}$.

Para $\alpha = 0,05$, $D = 1,358/\sqrt{n}$ y para $\alpha = 0,01$, $D = 1,628/\sqrt{n}$.

Fuente: Adaptada de MILLER, L. H.: «Table of percentage points of Kolmogorov Statistics», JASA; vol. 51. Reproducida con permiso de JASA. Copyright 1956 por American Statistical Association. Todos los derechos reservados.

TABLA 6(a)

Valores críticos de la prueba R de rachas.

n_2	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
n_1																				
2																				
3																				
4																				
5																				
6																				
7																				
8																				
9																				
10																				
11																				
12																				
13																				
14																				
15																				
16																				
17																				
18																				
19																				
20																				

Fuente: SWED, F. S.; EISENHART, C.: «Tables for testing randomness of grouping in a sequence of alternatives». *Ann. Math. Stat.* Vol. 14. Reproducida con permiso del editor. Copyright 1943 por *Institute of Mathematical Statistics*. Todos los derechos reservados.

TABLA 6(b)

Valores críticos de la prueba R de rachas.

n_2	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
n_1																				
2																				
3																				
4				9	9															
5				9	10	10	11													
6				9	10	11	12	12												
7					11	12	13	13	14	14	14	14	15	15						
8					11	12	13	14	14	14	15	16	16	17	17	17	17	17	17	17
9						13	14	14	15	16	16	16	17	17	18	18	18	18	18	18
10						13	14	15	16	16	17	17	18	18	18	19	19	19	20	20
11						13	14	15	16	17	17	18	19	19	20	20	21	21	21	21
12						13	14	16	16	17	18	19	19	20	20	21	21	22	22	22
13						15	16	17	18	19	19	20	20	21	21	22	22	23	23	23
14						15	16	17	18	19	20	20	21	22	22	23	23	23	24	24
15						15	16	18	18	19	20	21	22	22	23	23	24	24	25	25
16							17	18	19	20	21	21	22	23	23	24	25	25	25	25
17							17	18	19	20	21	22	23	23	24	25	25	26	26	26
18							17	18	19	20	21	22	23	24	25	25	26	26	27	27
19							17	18	20	21	22	23	24	25	26	26	27	27	27	27
20							17	18	20	21	22	23	24	25	25	26	27	27	27	28

Fuente: SWED, F. S.; EISENHART, C.: «Tables for testing randomness of grouping in a sequence of alternatives». *Ann. Math. Stat.* Vol. 14. Reproducida con permiso del editor. Copyright 1943 por *Instit. of Mathematical Statistics*. Todos los derechos reservados.