



# Principios de Inferencia Estadística

Ricardo Vélez Ibarrola  
Alfonso Garía Pérez

# *Principios de Inferencia Estadística*

RICARDO VÉLEZ IBARROLA  
ALFONSO GARCÍA PÉREZ

UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA

**PRINCIPIOS DE INFERENCIA ESTADÍSTICA**  
**(6102310GR01A01)**

*Quedan rigurosamente prohibidas, sin la autorización escrita de los titulares del copyright, bajo las sanciones establecidas en las leyes, la reproducción total o parcial de esta obra por cualquier medio o procedimiento, comprendidos la reprografía y el tratamiento informático y la distribución de ejemplares de ella mediante alquiler o préstamos públicos.*

© Universidad Nacional de Educación a Distancia  
Madrid 2012

*Librería UNED: c/ Bravo Murillo, 38 - 28015 Madrid  
Tels.: 91 398 75 60 / 73 73  
e-mail: libreria@adm.uned.es*

© Ricardo Vélez Ibarrola, Alfonso García Pérez

*Todas nuestras publicaciones han sido sometidas a un sistema de evaluación antes de ser editadas.*

*ISBN: 978-84-362-6577-4  
Depósito legal: M-35165-2012*

*Primera edición: noviembre de 2012*

*Impreso en España - Printed in Spain  
Imprime y encuaderna: Grafo, S.A.  
Avda. Cervantes, 51, edif. 21, (Vizcaya)*



# Prólogo

Este libro está concebido como texto para el primer curso de Inferencia Estadística en el Grado de Matemáticas de la U.N.E.D. Pretende exponer los conceptos fundamentales de la Estadística matemática, que han de servir de base a la elaboración posterior de modelos más específicos y que deberían ser tenidos en cuenta en toda aplicación estadística. Presupone que el lector conoce el cálculo de probabilidades al nivel habitual de los primeros cursos del Grado.

Parafraseando la máxima de Tukey<sup>1</sup>: En un problema estadístico,

Es importante entender lo que se *puede hacer* antes de entrar a juzgar lo que se *debe hacer*.

Los sucesivos capítulos del libro tratan de describir las principales líneas de conducta frente a un problema estadístico —en circunstancias generales— y las diferentes metodologías que pueden emplearse. La aplicación real de tales técnicas requiere casi siempre de ciertos prolegómenos (Muestreo y Análisis de datos, principalmente) que no serán tratados aquí, por tener su lugar independiente en el plan de estudios.

El contenido del libro es, por otra parte, muy clásico y requiere, en consecuencia, escasa justificación. Recorre el esquema tradicional: introducir los conceptos de base y, en particular, las *distribuciones en el muestreo* (Cap. 1, 2 y 3), para tratar sucesivamente los *intervalos de confianza* (Cap. 4), la *estimación puntual* (Cap. 5, 6 y 7) y el *contraste de hipótesis* (Cap. 8, 9 y 10).

En cuanto a la forma, hemos procurado huir de formalismos hueros, cuyo pretendido rigor se reduce a emplear el estilo: *definición / lema / teorema / demostración*, sin ninguna aclaración intermedia. En su lugar, hemos pretendido utilizar el lenguaje para explicar “de palabra” —en la medida de lo

---

<sup>1</sup>Tukey, J.W. (1977): *Exploratory data analysis*. Addison-Wesley. pág. v.

posible— el significado y el motivo de cada concepto y resultado. El único inconveniente de tal opción es que incrementa el volumen del texto, aunque, como contrapartida, debería facilitar su lectura. De hecho, la opción parece clara, pues, por muy breve que sea una página de jeroglífico, leerla puede costar años.

Quizás existen desarrollos matemáticos que no han nacido para ser utilizados, sino con el propósito estético —nada desdeñable— de aclarar en profundidad alguna cuestión meramente teórica. Está claro, sin embargo, que no es éste el caso de la Inferencia estadística que goza, además, del sentido que le confiere su utilidad. Así pues, es recomendable que el lector, después de haber adquirido cada idea con claridad, antes que a pulir y perfeccionar sus aspectos teóricos, dedique sus esfuerzos a asegurarse de que sabe usarla. Ello sólo puede lograrse ejercitándose en la resolución de ejemplos y problemas. El libro contiene un centenar de ejemplos y otros tantos ejercicios resueltos que deben servir para tal fin. Ni que decir tiene, que ello también ha contribuido a aumentar considerablemente su volumen, pero la experiencia enseña que es más difícil encontrar material práctico que buenas exposiciones de la teoría.

Todo texto está en deuda con los que lo precedieron —algunos de los cuales se citan en la bibliografía— y, también, con el ambiente en que se ha desarrollado. En este último sentido debemos gratitud a muchos de nuestros maestros y colegas, así como a diversas generaciones de estudiantes de distintas Universidades. Especial mención merece el agradecimiento al Prof. I. Yáñez, con el que hemos mantenido interesantes discusiones sobre muchas cuestiones aquí tratadas; su claridad de ideas sobre el tema, sin escribirlo, es superior a la nuestra, después de escrito.

La composición del texto hubiese sido con seguridad mucho más deficiente sin la ayuda inestimable del *TEX* y del *LATEX*, de manera que también estamos en deuda con sus creadores.

Los autores  
Madrid, Septiembre 1992

# Índice

<b>1. Conceptos básicos</b>	<b>1</b>
1.1. Introducción . . . . .	1
1.2. Concepto de población . . . . .	4
1.3. Concepto de muestra aleatoria . . . . .	7
1.4. Estadísticos y su distribución en el muestreo . . . . .	15
<b>2. Distribución muestral</b>	<b>33</b>
2.1. Introducción . . . . .	33
2.2. Distribución muestral y sus características . . . . .	34
2.3. Distribución en el muestreo y comportamiento asintótico . . . . .	36
2.4. Distribución en el muestreo de los momentos muestrales . . . . .	39
2.5. Comportamiento asintótico de los cuantiles muestrales . . . . .	41
2.6. Comportamiento asintótico de los momentos muestrales . . . . .	45
<b>3. Poblaciones normales</b>	<b>59</b>
3.1. Introducción . . . . .	59
3.2. Distribución de la media y varianza muestrales. Distribución $\chi^2$ de Pearson . . . . .	60
3.3. Estadístico de Student. Distribución $t$ de Student . . . . .	67
3.4. Distribución de la diferencia de medias muestrales . . . . .	71
3.5. Distribución del cociente de cuasivarianzas muestrales. Distribución $F$ de Snedecor . . . . .	75
3.6. El coeficiente de correlación muestral . . . . .	79
<b>4. Intervalos de confianza</b>	<b>91</b>
4.1. Introducción . . . . .	91
4.2. Definición de intervalo de confianza . . . . .	92
4.3. Método de la cantidad pivotal para la construcción de intervalos de confianza . . . . .	96

4.4. Método de Neyman para la construcción de intervalos de confianza . . . . .	101
4.5. Intervalos de confianza para los parámetros de distribuciones normales . . . . .	114
4.6. Intervalos de confianza basados en distribuciones asintóticas .	120
<b>5. Estimación Puntual</b>	<b>139</b>
5.1. Introducción . . . . .	139
5.2. Comparación de estimadores: Riesgo y error cuadrático medio	142
5.3. Propiedades deseables de los estimadores . . . . .	151
5.3.1. Estimadores insesgados . . . . .	151
5.3.2. Estimadores consistentes . . . . .	155
5.3.3. Estimadores invariantes . . . . .	157
5.4. Estadísticos suficientes . . . . .	161
5.4.1. Estadísticos suficientes minimales . . . . .	170
5.5. Criterios de selección de estimadores . . . . .	175
5.5.1. Estimadores minimax . . . . .	176
5.5.2. Estimadores Bayes . . . . .	179
<b>6. Estimadores de mínimo riesgo</b>	<b>209</b>
6.1. Introducción . . . . .	209
6.2. Estimadores centrados de mínima varianza . . . . .	210
6.3. Acotaciones de la varianza de un estimador . . . . .	223
6.4. Estimación de parámetros de posición y escala . . . . .	238
6.4.1. Parámetros de posición . . . . .	238
6.4.2. Parámetros de escala . . . . .	244
6.4.3. Parámetros de posición y escala . . . . .	249
<b>7. Métodos de estimación</b>	<b>279</b>
7.1. Introducción . . . . .	279
7.2. El método de los momentos . . . . .	279
7.3. El método de la máxima verosimilitud . . . . .	281
7.4. Propiedades asintóticas de los estimadores de máxima verosimilitud . . . . .	292
7.5. Estimación Bayesiana . . . . .	303
7.6. Estimación mínimo cuadrática . . . . .	306
7.7. Conclusión . . . . .	308

<b>8. Contraste de hipótesis</b>	<b>327</b>
8.1. Introducción . . . . .	327
8.2. Planteamiento general de los contrastes de hipótesis . . . . .	328
8.3. Contraste de hipótesis simple frente a simple . . . . .	336
8.4. Contraste de hipótesis unilaterales y bilaterales . . . . .	347
<b>9. Métodos de contraste</b>	<b>387</b>
9.1. Introducción . . . . .	387
9.2. Tests de razón de verosimilitudes . . . . .	388
9.3. Relación entre estimación confidencial y contraste de hipótesis	400
<b>10. Contrastes globales</b>	<b>429</b>
10.1. Introducción . . . . .	429
10.2. Contrastos $\chi^2$ de bondad del ajuste . . . . .	431
10.2.1. Hipótesis simple . . . . .	431
10.2.2. Hipótesis compuesta . . . . .	438
10.3. Contraste $\chi^2$ de homogeneidad . . . . .	445
10.4. Contraste $\chi^2$ de independencia . . . . .	450
10.5. Contraste de Kolmogorov-Smirnov de bondad del ajuste . . .	454
10.6. Contraste de homogeneidad de Kolmogorov-Smirnov . . . . .	461
10.7. Contrastos de posición . . . . .	467
10.7.1. El test de los signos . . . . .	468
10.7.2. El test de Wilcoxon de los rangos signados . . . . .	470
10.7.3. El test de la mediana . . . . .	473
10.7.4. El test de Mann-Whitney . . . . .	476
10.8. Contrastos de independencia . . . . .	478
10.8.1. El test $\tau$ de Kendall . . . . .	478
10.8.2. El test del coeficiente de correlación entre rangos de Spearman . . . . .	481
10.8.3. El test de rachas . . . . .	483

# Capítulo 1

## Conceptos básicos de la Estadística Matemática

### 1.1. Introducción

El objeto de este primer capítulo es presentar los elementos protagonistas que intervienen en los problemas estadísticos, a la vez que llevar a cabo una descripción inicial de los mismos, que sirva, entre otras cosas, para recalcar su conexión con el Cálculo de Probabilidades.

Es bien sabido que el Cálculo de Probabilidades proporciona una teoría matemática que permite analizar las propiedades de los fenómenos en los que interviene el azar. A tal fin utiliza como modelo básico, común a cualquier situación aleatoria, el concepto de espacio probabilístico,  $(\Omega, \mathcal{A}, P)$ , junto con la noción de variable aleatoria definida sobre él. A partir de esta base se deducen todas las consecuencias que se utilizan en el análisis de las situaciones aleatorias y también, en otra dirección, se proponen resultados que facilitan la construcción de espacios probabilísticos adaptados a fenómenos aleatorios concretos.

Por ejemplo, una aseveración fundamental, en esta segunda dirección, es que para construir un espacio probabilístico, que sirva de modelo de una determinada variable aleatoria real, basta conocer la función de distribución de dicha variable. En este sentido, el Cálculo de Probabilidades cumple la mitad de su papel al permitir la comprensión de afirmaciones como: “la velocidad del movimiento caótico de las moléculas de un gas sigue una distribución normal de parámetros...”<sup>1</sup> o “la vida de un determinado tipo de

---

<sup>1</sup> A. Vassiliev (1985): *Introduction a la physique statistique*. Ed. MIR. Moscú.

componente eléctrica tiene distribución exponencial de media...”; papel que completa ofreciendo herramientas para la deducción de consecuencias de este tipo de premisas.

Lo que interesa destacar en este momento, acerca de las consideraciones anteriores, es que, para el Cálculo de Probabilidades, el espacio probabilístico es un dato —el dato de partida— y, en los problemas de su estricta competencia, ha de disponerse de información suficiente —expresada de una u otra forma— sobre el fenómeno en estudio, para poder construir el espacio probabilístico adecuado.

Si se desea profundizar en la forma de adaptar un modelo probabilístico, a cada situación aleatoria extraída de la realidad, es perfectamente claro que no cabe más alternativa que llevar a cabo observaciones del fenómeno en cuestión.

Así por ejemplo, no existe mejor procedimiento para tratar de averiguar si un dado está bien equilibrado, que realizar con él un buen número de lanzamientos y “verificar estadísticamente” si los resultados obtenidos confirman o invalidan la corrección del dado. Hasta entonces —o, más exactamente, hasta que no se conozcan las probabilidades,  $p_i$ , de las diversas caras— el Cálculo de Probabilidades sólo podrá actuar en función de los parámetros  $p_i$ , sin alcanzar por tanto soluciones numéricas.

De la misma manera, si se desea estudiar la distribución de la estatura de los individuos de una determinada población, es necesario, sin duda, conocer las tallas de un gran número de ellos. Quizás no sea imprescindible conocerlas todas —puesto que ello sería prácticamente inviable— pero tampoco es posible deducir la distribución exacta mediante consideraciones meramente teóricas.

El disponer de un conjunto de observaciones acerca del fenómeno considerado, en lugar de un espacio probabilístico totalmente especificado, hace abandonar los dominios del Cálculo de Probabilidades, para introducirse en el terreno de la *Estadística Matemática* o *Inferencia Estadística*, cuya finalidad es obtener información acerca de la ley de probabilidad de un fenómeno, a partir de una observación no exhaustiva del mismo —ya que una observación exhaustiva es a menudo imposible, o incluso carece de sentido (como en el caso del dado antes citado).

Frecuentemente se entiende por Estadística el proceso de recolección de datos u observaciones, así como su tratamiento numérico; etapas que com-

portan, en muchas ocasiones, considerables complicaciones técnicas. Sin embargo, el interés de la Inferencia Estadística se centra exclusivamente en el análisis e interpretación de las observaciones, como método para obtener conclusiones sobre la ley de probabilidad del fenómeno en estudio.

Precisamente, un factor clave de diferenciación entre los diversos problemas de Inferencia Estadística que pueden plantearse, es el tipo de conclusiones que quieren establecerse sobre la situación aleatoria en consideración. Pueden distinguirse así:

- Situaciones en las que el objetivo es utilizar la información suministrada por las observaciones para obtener *un pronóstico numérico único* acerca de un determinado parámetro de la distribución (media, varianza, coeficiente de correlación, o cualquier otra cosa). Ello caracteriza los problemas de *Estimación puntual* (Capítulos 5 a 7) en los que se obtienen conclusiones del tipo: “la probabilidad de cara con esta moneda es 0’37” o “la duración media de esta clase de componentes eléctricas es de 137 horas” etc.; lo cual obliga a analizar las cualidades de tales estimaciones o, mejor dicho, de los métodos empleados para obtenerlas.
- En otras ocasiones el objetivo puede ser proporcionar un margen de variación para un determinado parámetro de la distribución desconocida; es decir *precisar un intervalo numérico en el que pueda razonablemente afirmarse que varía el parámetro en cuestión*. Se habla entonces de estimación por *Intervalos de confianza* (Capítulo 4) y conclusiones típicas en este caso serán: “la probabilidad de obtener cara con esta moneda está comprendida entre 0’37 y 0’385” o bien “la desviación típica de las estaturas de la población española tiene un valor entre 3’5 y 4’5 cm.”
- Por el contrario, la finalidad puede ser tratar de corroborar o *invalidar una determinada afirmación acerca de la distribución de probabilidad del fenómeno estudiado*. Se trata en este caso de problemas de *Contraste de hipótesis* (Capítulos 8 a 10), tales como comprobar si un determinado dado está cargado o verificar si cierto fármaco aumenta la supervivencia ante una enfermedad.

La sistematización de estos distintos tipos de problemas y la obtención de procedimientos adecuados para su solución, configura la mayor parte de

la teoría de la Estadística Matemática; si bien, obviamente, serán necesarias algunas precisiones básicas, antes de poder abordar el planteamiento exacto y la resolución de los problemas enumerados.

## 1.2. Concepto de población

Según se acaba de indicar, todo problema de Inferencia Estadística está motivado por un cierto grado de desconocimiento acerca de la ley de probabilidad que rige un determinado fenómeno aleatorio.

El caso más simple es que el interés se centre en una cierta variable aleatoria unidimensional,  $X$ , cuya distribución,  $F$ , es más o menos desconocida. Puede tratarse, por ejemplo, de la puntuación obtenida en el lanzamiento de un dado; de la respuesta, favorable o no, a un cierto tratamiento médico; de la duración de un cierto tipo de componente electrónico; del volumen mensual de ventas de un cierto establecimiento; etc... Las situaciones en que la variable  $X$ , y consecuentemente su distribución, sean multidimensionales son técnicamente más complicadas pero conceptualmente idénticas.

De una manera genérica la distribución desconocida  $F$  de la variable aleatoria involucrada en un problema de Inferencia Estadística, recibe el nombre de *distribución teórica* o *distribución de la población*, siendo esta segunda denominación debida a la gran frecuencia con que la Estadística Matemática se utiliza para la determinación de la distribución de alguna característica de los individuos de una población.

En los modelos de Inferencia Estadística, el mayor o menor grado de desconocimiento acerca de la distribución teórica  $F$  se refleja mediante la familia  $\mathcal{F}$  de distribuciones, candidatas a ser realmente la distribución de la población. Así  $\mathcal{F}$  puede interpretarse como la información previa de que se dispone ante un problema estadístico: si  $\mathcal{F} = \{F_1, F_2\}$  sólo se trata de dilucidar si la distribución teórica es una u otra, mientras que si  $\mathcal{F}$  es la familia de todas las distribuciones unidimensionales, el desconocimiento acerca del fenómeno es total.

La situación a la vez más simple y más útil es aquella en que la familia  $\mathcal{F}$  está compuesta por distribuciones de forma funcional fija y conocida, dependientes de un parámetro  $\theta$ , de una o más dimensiones, que varía dentro de un subconjunto  $\Theta$  de  $\mathbb{R}^k$ , denominado *espacio paramétrico*. Simbólicamente:

$$\mathcal{F} = \{F_\theta \mid \theta \in \Theta \subset \mathbb{R}^k\}.$$

Tales situaciones, que se engloban comúnmente bajo la denominación de *Inferencia estadística paramétrica*, se presentan con gran frecuencia debido a que es posible a menudo conocer la forma funcional de  $F$ , mediante consideraciones teóricas, y concentrar la indeterminación en los valores de determinados parámetros de la distribución. Algunos ejemplos pueden aclarar esta idea:

### Ejemplo 1.1

Supongamos que la variable  $X$  en estudio es la duración de una cierta componente eléctrica que no envejece; es decir que, si la componente sobrevive en el instante  $t$  su estado es el mismo que inicialmente y la distribución del tiempo que falta para su rotura sigue siendo la misma que al principio. Debido a que la distribución exponencial es la única con esta propiedad de “falta de memoria”, la distribución de  $X$  será exponencial:

$$F_\theta(x) = \begin{cases} 1 - e^{-\theta x} & \text{si } x \geq 0, \\ 0 & \text{si } x < 0, \end{cases}$$

o bien, en términos de densidades,

$$f_\theta(x) = \begin{cases} \theta e^{-\theta x} & \text{si } x \geq 0, \\ 0 & \text{si } x < 0. \end{cases}$$

Todo el desconocimiento acerca de la distribución teórica  $F$ , queda entonces centrado en el valor del parámetro  $\theta \in (0, \infty)$ , o, equivalentemente, en el valor de la duración media  $1/\theta$  de la distribución.

### Ejemplo 1.2

Estamos interesados en la variable  $X$ , número de accidentes de un cierto tipo que se producen por unidad de tiempo entre los individuos de un cierto colectivo. Si se sabe que es muy pequeña la probabilidad de que cada uno individualmente sufra dicho accidente, pero que es muy grande el número de candidatos a padecerlo, la conocida aproximación de la distribución binomial hacia la distribución de Poisson, permite concluir que la distribución teórica  $F$  será de la forma

$$F_\theta(x) = \sum_{k=0}^{[x]} \frac{\theta^k e^{-\theta}}{k!} \quad \text{para } x \geq 0,$$

o más claramente, en términos de la función de probabilidad,

$$f_\theta(x) = \frac{\theta^x e^{-\theta}}{x!} \quad \text{para } x = 0, 1, 2, \dots$$

Nuevamente el desconocimiento acerca de la distribución teórica queda entonces reducido al valor del número medio de accidentes por unidad de tiempo,  $\theta$ , que varía en el espacio paramétrico  $\Theta = (0, \infty)$ .

### Ejemplo 1.3

Es un hecho empíricamente probado que la inmensa mayoría de las características morfológicas de los individuos de una población biológica (longitudes, pesos, diámetros, concentraciones, etc.) siguen una distribución normal. La razón profunda de este hecho es, sin duda, el teorema central del límite, debido a que tales características son el resultado de la acumulación de gran número de factores. Por consiguiente, pese a que en cada caso particular la regla requeriría teóricamente una comprobación estadística, en la práctica es muy frecuente dar por hecho que tal tipo de características tienen una distribución teórica con función de densidad

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

siendo  $\theta = (\mu, \sigma)$  el parámetro bidimensional en el que se concentra el desconocimiento de la población. Así, en este caso el espacio paramétrico es pues  $\Theta = \mathbb{R} \times (0, \infty) \subset \mathbb{R}^2$  o algún subconjunto suyo.

Como ocurre en los ejemplos anteriores, las familias de distribuciones teóricas,  $\{F_{\theta} \mid \theta \in \Theta\}$ , que se presentan en los problemas de Inferencia estadística nunca son distribuciones muy raras. Concretamente, casi sin excepción son familias de distribuciones continuas, con función de densidad, o familias de distribuciones discretas, caracterizadas por su función de probabilidad; como corresponde respectivamente a que la característica poblacional que se desea analizar sea continua (como la estatura, duración, etc.) o discreta (número de accidentes, resultado de un dado, etc.). Naturalmente entonces, es cómodo manejar la distribución teórica a través de su función de densidad o de probabilidad, según el caso. La misma notación:  $f_{\theta}(x)$ , se empleará, a efectos teóricos, en las dos situaciones puesto que ambos tratamientos pueden efectuarse conjuntamente.

La Estadística Matemática comprende también situaciones en las cuales, a diferencia de las anteriores, la familia de distribuciones  $\mathcal{F}$  a la que pertenece la distribución teórica es no paramétrica. Los ejemplos más frecuentes relativos a este caso son que  $\mathcal{F}$  sea la familia de todas las funciones de distribución continuas, o la familia de todas las distribuciones absolutamente continuas, etc. Es fácil comprender que tales problemas de Inferencia

estadística no paramétrica comportan mayor número de dificultades y admiten soluciones menos completas que los casos paramétricos.

### 1.3. Concepto de muestra aleatoria

Una vez precisadas las ideas básicas acerca del concepto de población, debemos ocuparnos de las observaciones a partir de las cuales ha de intentarse disminuir el desconocimiento de la distribución teórica  $F$  de la variable aleatoria  $X$  en estudio. En principio tales observaciones pueden obtenerse llevando a cabo repeticiones del experimento aleatorio que da lugar a la variable aleatoria  $X$  y anotando los valores de  $X$  en cada una de ellas. Se obtendrá así un conjunto de valores numéricos  $(x_1, x_2, \dots, x_n)$  que constituyen lo que se denomina una *muestra aleatoria* de  $X$ . El número  $n$  de repeticiones efectuadas, y de observaciones obtenidas, se denomina *tamaño de la muestra*.

La idea que subyace en la elección de las observaciones por procedimientos aleatorios es la siguiente: Obviamente sería deseable disponer de una muestra “representativa” de la población y, si se conociese la distribución de ésta, se podría tratar de seleccionar aquella con este requisito. Sin embargo, puesto que la distribución teórica es desconocida, no es posible siquiera prejuzgar si una determinada muestra es “representativa” de algo que se desconoce. Ante tal círculo vicioso, la única solución es confiar en que el azar se encargará de suministrar una muestra con la deseada representatividad. De hecho, el próximo capítulo establecerá que es muy razonable esperar que así ocurra, en el sentido de que las leyes del azar juegan a favor de conseguirlo. Ello es fácil de comprender en situaciones concretas: Al lanzar 100 veces un dado cuyo equilibrio se pretende analizar, es posible que, aunque el dado estuviese bien equilibrado, se produzca el suceso “poco representativo” de obtener más de 95 seises; lo cual llevaría sin duda a concluir estadísticamente que el dado está cargado. Sin embargo, la probabilidad de tal suceso es muy pequeña (si el dado está realmente bien equilibrado) y no sólo eso, sino que la ley fuerte de los grandes números garantiza que, si el tamaño de la muestra es suficientemente grande, es sumamente probable que se obtenga una muestra representativa, con proporción de cada puntuación muy próxima a la probabilidad real de cada cara.

Aún aceptando, de acuerdo con lo anterior, el principio del muestreo aleatorio, los procedimientos posibles para obtener una muestra aleatoria

pueden ser diversos, dependiendo de las condiciones en que se efectúen las repeticiones del experimento aleatorio. Lo más simple es llevar a cabo repeticiones *independientes* y en tal caso la muestra obtenida se denomina una *muestra aleatoria simple*.

Ello se reduce en el ejemplo del dado a hacer los lanzamientos siempre en las mismas condiciones, sin tener en cuenta los resultados anteriores; y, para obtener una muestra de estaturas de los individuos de una población, a seleccionar en cada ocasión al individuo a tallar con igual probabilidad entre todos sus miembros.

En este segundo caso, la posibilidad de tallar dos veces al mismo individuo no parece razonable y sería más lógico utilizar un *muestreo sin reemplazamiento*, excluyendo de cada elección a los individuos que ya hayan sido seleccionados. Por supuesto, el efecto es que las repeticiones no son independientes unas de otras y la muestra que se obtenga no será aleatoria simple. En la práctica relativa a situaciones de este tipo, se emplea también el *muestreo estratificado*, dividiendo la población por regiones geográficas (o con cualquier otro criterio) y eligiendo una submuestra dentro de cada uno de tales estratos; el *muestreo bietápico*, eligiendo aleatoriamente algunos estratos y, dentro de ellos, algunos individuos; el *muestreo sistemático* que, tras ordenar a los individuos, selecciona a los que ocupan lugares múltiples de un número fijo empezando por uno elegido al azar; etc.

Dado que los principios teóricos de la Estadística Matemática son comunes para cualquiera de ellas, no insistiremos en la descripción de estas técnicas de muestreo utilizadas en la práctica, por no añadir dificultades “de cálculo” a las dificultades conceptuales. Pensemos, pues, que las muestras son siempre aleatorias simples y, aunque tal hipótesis sea a menudo superflua, se mantendrá, casi sin excepción, sobre todo en los ejemplos. De los demás casos, sólo importa, en este momento, dejar constancia de que todo problema de Inferencia estadística, para estar bien formulado, tiene que incluir la definición precisa del *procedimiento de muestreo*, con el que se obtienen las observaciones.

Hay otra consideración importante acerca de las muestras que conviene tener presente desde el principio:

Cuando ya se ha obtenido efectivamente una muestra aleatoria de tamaño  $n$ :  $(x_1, x_2, \dots, x_n)$ , se dispone simplemente con un conjunto de  $n$  valores numéricos con los cuales tratar de cumplir el objetivo de precisar la distribución teórica. Sin embargo, cuando se está planificando la obtención de la muestra aleatoria y diseñando el procedimiento que se empleará pa-

ra intentar conocer la distribución teórica, todavía no se sabe que valores numéricos ( $x_1, x_2, \dots, x_n$ ) resultarán, de manera que deben considerarse como  $n$  variables aleatorias:  $(X_1, X_2, \dots, X_n)$ , una de cuyas realizaciones se va a observar.

Desde este segundo punto de vista, la muestra aleatoria — como conjunto de  $n$  variables aleatorias — tiene una distribución  $n$ -dimensional (supuesto que las observaciones individuales son unidimensionales), que gobierna las probabilidades de aparición de cada muestra efectiva. En el caso fundamental del muestreo aleatorio simple, puesto que se efectúan repeticiones independientes del experimento aleatorio, observando en cada una el valor de la variable  $X$  en estudio, cuya distribución teórica es  $F$ , podemos dar la siguiente definición precisa:

*Una muestra aleatoria simple, de tamaño  $n$ , de una variable aleatoria  $X$  con distribución teórica  $F$ , son  $n$  variables aleatorias  $(X_1, X_2, \dots, X_n)$ , independientes e igualmente distribuidas, con distribución común  $F$ .*

*Consecuentemente la función de distribución conjunta de una muestra aleatoria simple  $(X_1, X_2, \dots, X_n)$ , correspondiente a una distribución de la población  $F$ , es*

$$F(x_1, x_2, \dots, x_n) = F(x_1)F(x_2) \cdots F(x_n).$$

No hay ninguna ambigüedad en representar mediante el mismo símbolo,  $F$ , la función de distribución teórica y la función de distribución de la muestra, puesto que el número de variables distingue claramente una de otra: la segunda siempre tiene  $n$  veces más variables que la primera. En cambio, se hace énfasis así en que la función de distribución de la muestra depende totalmente de la distribución poblacional, e incluso coincide con ella si  $n = 1$ .

Aunque no haya de resultar de ninguna importancia práctica, cabe señalar que cualquier otro procedimiento de muestreo, mediante el cual se vaya a obtener una muestra aleatoria (no simple)  $(X_1, X_2, \dots, X_n)$  estará bien definido si determina la *distribución conjunta de la muestra*, considerada como conjunto de  $n$  variables aleatorias. Mejor que insistir en esa dirección, tratemos de ilustrar las ideas relativas al muestreo aleatorio simple mediante algunos ejemplos:

### Ejemplo 1.4

Supongamos que estamos interesados en conocer la probabilidad,  $\theta$ , de obtener cara con una cierta moneda. Es decir, se trata de estudiar la variable aleatoria

$$X = \begin{cases} 1 & \text{si se obtiene cara} \\ 0 & \text{si se obtiene cruz} \end{cases}$$

cuya distribución, caracterizada por:

$$P\{X = 1\} = \theta, \quad P\{X = 0\} = 1 - \theta,$$

depende del parámetro desconocido  $\theta$ , que varía en el espacio paramétrico  $\Theta = [0, 1]$ . Imaginemos que, por alguna razón, sólo se pueden efectuar 3 lanzamientos y obtener una muestra aleatoria simple  $(X_1, X_2, X_3)$ , de tamaño 3. Tamaño ciertamente escaso a efectos estadísticos, pero que permite prever exhaustivamente todas las muestras posibles, junto con la probabilidad de que se presente cada una de ellas:

$X_1$	$X_2$	$X_3$	Probabilidad
1	1	1	$\theta^3$
1	1	0	$\theta^2(1 - \theta)$
1	0	1	$\theta^2(1 - \theta)$
0	1	1	$\theta^2(1 - \theta)$
1	0	0	$\theta(1 - \theta)^2$
0	1	0	$\theta(1 - \theta)^2$
0	0	1	$\theta(1 - \theta)^2$
0	0	0	$(1 - \theta)^3$

Ello especifica la distribución de la muestra, a través de su función de probabilidad. Es difícil en esta situación, imaginar un procedimiento de muestreo diferente, que no sea aleatorio simple, puesto que la independencia entre los lanzamientos tiene por efecto que la probabilidad de cada muestra sea el producto de las probabilidades de los resultados que la componen.

A menudo, es útil manejar la distribución de la muestra mediante una fórmula explícita —aún cuando ello dé la misma información que la tabla anterior. En este caso concreto, puede lograrse escribiendo la función de probabilidad de la población en la forma

$$f_\theta(x) = \theta^x(1 - \theta)^{1-x} \quad \text{para } x = 0 \text{ ó } 1.$$

De forma que la función de probabilidad de la muestra se expresa

$$\begin{aligned} f_{\theta}(x_1, x_2, x_3) &= \theta^{x_1}(1-\theta)^{1-x_1}\theta^{x_2}(1-\theta)^{1-x_2}\theta^{x_3}(1-\theta)^{1-x_3} \\ &= \theta^{x_1+x_2+x_3}(1-\theta)^{3-x_1-x_2-x_3}. \end{aligned}$$

### Ejemplo 1.5

En la situación del ejemplo 1.2, si se puede aceptar que los números de accidentes que se producen en unidades de tiempo distintas son independientes entre sí, la obtención de una muestra aleatoria simple de tamaño  $n$  consiste sencillamente en observar el número de accidentes que se producen durante  $n$  días. La muestra aleatoria simple  $(X_1, X_2, \dots, X_n)$  puede tomar entonces cada uno de los valores  $(x_1, x_2, \dots, x_n)$ , ( $x_i \in \mathbb{N} \cup \{0\}$ ) con probabilidad:

$$f_{\theta}(x_1, x_2, \dots, x_n) = \frac{\theta^{x_1} e^{-\theta}}{x_1!} \cdot \frac{\theta^{x_2} e^{-\theta}}{x_2!} \cdots \frac{\theta^{x_n} e^{-\theta}}{x_n!} = \frac{\theta^{x_1+x_2+\cdots+x_n} e^{-n\theta}}{x_1! x_2! \cdots x_n!}.$$

Dicha función de probabilidad determina la distribución de la muestra, que depende lógicamente de la distribución de la población, a través del parámetro  $\theta$ .

### Ejemplo 1.6

Supongamos que se desea estudiar la talla y el peso de los individuos de una población. Consideramos entonces la variable aleatoria bidimensional  $(X, Y)$  que asocia tales características a cada uno de los individuos de la población. Su distribución teórica supondremos que es una normal bidimensional, con función de densidad  $f_{\theta}(x, y)$  dada por

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-\frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]}$$

y dependiente del parámetro desconocido :

$$\theta = (\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) \in \Theta = \mathbb{R}^2 \times (0, \infty)^2 \times [-1, 1]$$

que agrupa las medias de ambas características, sus desviaciones típicas y su coeficiente de correlación.

A fin de tratar de obtener información sobre el parámetro  $\theta$ , pensamos en seleccionar  $n$  individuos, al azar e independientemente, para disponer de una muestra aleatoria simple  $((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$  con los datos de talla y peso de cada uno de ellos.

La muestra aleatoria simple  $((X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n))$ , considerada como variable aleatoria, tendrá una distribución  $2n$ -dimensional, caracterizada por la función de densidad  $f_\theta(x_1, y_1, x_2, y_2, \dots, x_n, y_n)$ :

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-\mu_1)(y_1-\mu_2)}{\sigma_1\sigma_2} + \frac{(y_1-\mu_2)^2}{\sigma_2^2}\right]}$$

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_2-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_2-\mu_1)(y_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(y_2-\mu_2)^2}{\sigma_2^2}\right]}$$

$$\dots \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_n-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_n-\mu_1)(y_n-\mu_2)}{\sigma_1\sigma_2} + \frac{(y_n-\mu_2)^2}{\sigma_2^2}\right]}$$

que se expresa

$$\frac{1}{(2\pi\sigma_1\sigma_2\sqrt{1-\rho^2})^n} e^{-\frac{1}{2(1-\rho^2)} \sum_{i=1}^n \left[ \frac{(x_i-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_i-\mu_1)(y_i-\mu_2)}{\sigma_1\sigma_2} + \frac{(y_i-\mu_2)^2}{\sigma_2^2} \right]}$$

y naturalmente depende de los parámetros desconocidos de la distribución de la población.

Como puede verse en los ejemplos anteriores, la *distribución de la muestra* suele manejarse a través de la correspondiente función de probabilidad o de densidad —según que la distribución poblacional sea discreta o continua— que proporciona la probabilidad o densidad de probabilidad con que puede presentarse cada muestra concreta, al llevar a cabo el muestreo. Generalmente, dicha función de probabilidad o de densidad será denotada por  $f_\theta(x_1, x_2, \dots, x_n)$ .

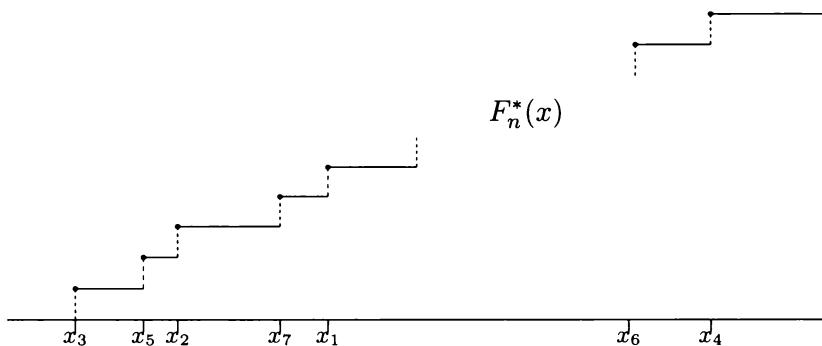
En un sentido totalmente distinto, se puede asociar a cada muestra concreta  $(x_1, x_2, \dots, x_n)$ , una *distribución muestral* con la pretensión de emular a la distribución teórica  $F$ , a partir únicamente de la información contenida en los  $n$  valores numéricos que componen la muestra.

Así, puesto que es  $F(x) = P\{X \leq x\}$ , para cada  $x \in \mathbb{R}$ , la *distribución muestral* asociada a la muestra  $(x_1, x_2, \dots, x_n)$  de  $X$  se define como

$$F_n^*(x) = \frac{\text{número de elementos muestrales} \leq x}{n}$$

para cada  $x \in \mathbb{R}$ ; es decir,  $F_n^*(x)$  es la frecuencia de elementos de la muestra que son menores o iguales que  $x$ .

La representación gráfica de la distribución muestral es de la forma:



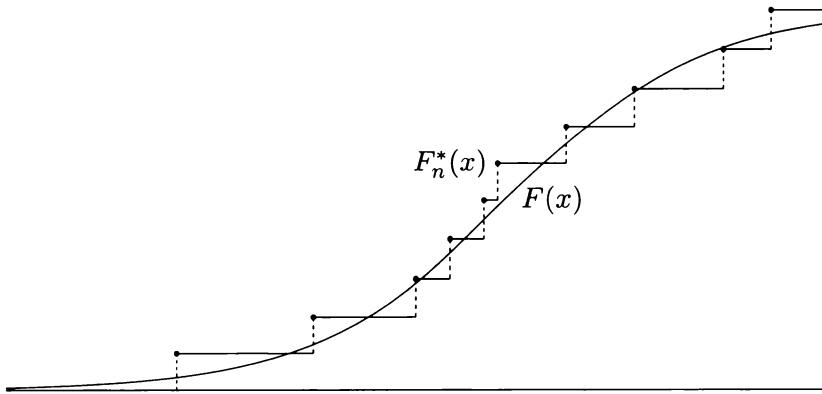
con saltos de magnitud  $1/n$  en los sucesivos valores muestrales ordenados de menor a mayor, supuesto que sean distintos, y saltos múltiples en el caso de que varios valores muestrales coincidiesen. Es decir, que la distribución muestral es siempre una distribución discreta, concentrada en los valores que aparecen en la muestra y cuya función de probabilidad es

$$p_n^*(x) = \frac{j}{n} \quad \text{si } j \text{ valores muestrales coinciden con } x,$$

para cada  $x \in I\!\!R$ .

Como se ve, la *distribución muestral*  $F_n^*(x)$  refleja exclusivamente, de una manera peculiar, los valores numéricos contenidos en la muestra concreta a la que está asociada y no tiene relación directa con la *distribución de la muestra* ni con la *distribución de la población*, en el sentido de que una muestra determinada tendría la misma distribución muestral fuese cual fuese la población de origen.

Pese a ello, es muy razonable esperar que la distribución muestral proporcione una imagen aproximada, más o menos tosca, de la distribución de la población de la que se haya extraído la muestra. Como se indicaba anteriormente, las leyes del azar juegan a nuestro favor, haciendo más verosímil, en caso de que el tamaño de la muestra sea grande, una situación del tipo:



en que la distribución muestral está muy próxima a la distribución de la población, que aquellas en que ocurriese lo contrario.

El próximo capítulo analizará con detenimiento esta cuestión, que sólo obedece al hecho de que cada valor de la muestra se obtiene aleatoriamente de la distribución de la población, y no de otra. De forma que, en el fondo, sí existe una relación, indirecta pero importante, entre la distribución muestral y la distribución poblacional de la que proviene la muestra.

El concepto de distribución muestral se extiende fácilmente al caso de muestras de una población con mayor número de dimensiones. Por ejemplo, si  $((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$  es una muestra aleatoria de una variable bidimensional, la distribución muestral es la distribución sobre el plano  $\mathbb{R}^2$  que tiene como función de probabilidad

$$p_n^*(x, y) = \frac{j}{n} \quad \text{si } j \text{ valores muestrales } (x_i, y_i) \text{ coinciden con } (x, y),$$

para cada  $(x, y) \in \mathbb{R}^2$ ; o en términos de la función de distribución

$$F_n^*(x, y) = \frac{1}{n} (\text{número de valores muestrales } (x_i, y_i) \text{ con } x_i \leq x \text{ e } y_i \leq y)$$

para cada  $(x, y) \in \mathbb{R}^2$ .

Conviene precisar, antes de acabar con las ideas genéricas sobre el muestreo aleatorio, la terminología que se empleará más adelante:

El conjunto de muestras posibles que pueden obtenerse al seleccionar una muestra aleatoria, de un tamaño determinado, de una cierta población, se denomina *espacio muestral* y habitualmente será designado por  $\mathcal{X}$ .

Por ejemplo, si se trata de una muestra aleatoria, de tamaño  $n$ , de una variable aleatoria unidimensional, el espacio muestral será  $\mathbb{R}^n$  o algún subconjunto suyo; en el caso de que la población fuese bidimensional, el espacio muestral sería  $\mathbb{R}^{2n}$  o algún subconjunto suyo; etc. . . Así, en el ejemplo 1.4 el espacio muestral es:

$$\mathcal{X} = \{(1, 1, 1), (1, 1, 0), (1, 0, 1), (0, 1, 1), (1, 0, 0), (0, 1, 0), (0, 0, 1), (0, 0, 0)\}$$

mientras que en el ejemplo 1.5 es  $(\mathbb{N} \cup \{0\})^n \subset \mathbb{R}^n$ ; y en el ejemplo 1.6,  $\mathbb{R}^{2n}$ .

En cualquier caso, se trata siempre de un subconjunto de un espacio euclídeo  $\mathbb{R}^{m \cdot n}$ , de manera que podemos considerar en  $\mathcal{X}$  la  $\sigma$ -álgebra restringida de la  $\sigma$ -álgebra de Borel  $\mathbb{B}^{m \cdot n}$ , que representaremos por  $\mathcal{B}$ . Obviamente, la distribución de la muestra determina una medida de probabilidad,  $\mathcal{P}$ , en el espacio muestral  $(\mathcal{X}, \mathcal{B})$ , que permite expresar las probabilidades de sucesos relativos a la muestra :  $\mathcal{P}\{(X_1, X_2, \dots, X_n) \in B\}$ , para cada  $B \in \mathcal{B}$ . De la misma manera que la medida de probabilidad,  $P$ , asociada a la distribución poblacional, permite expresar las probabilidades de los sucesos relativos a la población:  $P\{X \in B\}$ , para cada conjunto de Borel  $B$  en el espacio euclídeo  $(\mathbb{R}^m)$  en el que toma valores la variable  $X$  que describe la población. En el contexto de los problemas paramétricos, suele ponerse el parámetro como subíndice:  $\mathcal{P}_\theta$ , para indicar que se trata de la distribución de la muestra, asociada a la distribución teórica  $F_\theta$ .

## 1.4. Estadísticos y su distribución en el muestreo

Aunque todavía no se hayan concretado con precisión los problemas que pretende resolver la Inferencia Estadística, sí se puede precisar cuál ha de ser el instrumento empleado para su solución. Si a partir de una muestra aleatoria  $(x_1, x_2, \dots, x_n)$  se desea disminuir el desconocimiento acerca de la distribución de la población, sea lo que sea lo que haya que hacer, habrá de hacerse en función de los resultados observados en la muestra. En otras palabras, cualquiera que sea el método de inferencia que se utilice, tendrá que estar basado en una función  $T(x_1, x_2, \dots, x_n)$  de los valores muestrales. Tal idea se recoge en la siguiente definición precisa:

*Se denomina estadístico a cualquier función  $T$  del espacio muestral  $(\mathcal{X}, \mathcal{B})$  en un espacio euclídeo  $(\mathbb{R}^k, \mathbb{B}^k)$  que sea medible (respecto a las  $\sigma$ -álgebras  $\mathcal{B}$  y  $\mathbb{B}^k$ ). La dimensión  $k$  del espacio euclídeo imagen se denomina dimensión del estadístico.*

Desde el punto de vista práctico la condición de medibilidad impuesta a los estadísticos no tiene ninguna relevancia, puesto que las funciones que se usan habitualmente como tales no ofrecen ninguna duda respecto a su medibilidad. Así, por ejemplo, son de uso frecuente los estadísticos:

- 1)  $T_1(x_1, x_2, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i$ , media aritmética de la muestra o media muestral, que suele designarse por  $\bar{x}$ .
- 2)  $T_2(x_1, x_2, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ , varianza de los valores muestrales, que suele designarse por  $s^2$ .
- 3)  $T_3(x_1, x_2, \dots, x_n) = \min(x_1, x_2, \dots, x_n)$ , el más pequeño de los valores muestrales, que suele designarse por  $x_{(1)}$ .
- 4)  $T_4(x_1, x_2, \dots, x_n) = \max(x_1, x_2, \dots, x_n)$ , el mayor valor muestral, que suele designarse por  $x_{(n)}$ .
- 5)  $T_5(x_1, x_2, \dots, x_n) = \sum_{i=1}^n \log x_i$ .
- 6)  $T_6(x_1, x_2, \dots, x_n) = \begin{cases} 1 & \text{si } x_1 \leq 2, \\ 0 & \text{si } x_1 > 2. \end{cases}$
- 7)  $T_7(x_1, x_2, \dots, x_n) = \frac{1}{n} (\text{número de } x_i \leq 3) = F_n^*(3)$ .

Así como combinaciones multidimensionales de los mismos:

$$(\bar{x}, s^2), \quad (x_{(1)}, x_{(n)}), \quad (\bar{x}, x_{(1)}, x_{(n)}), \quad \text{etc.}$$

En cambio, desde el punto de vista teórico, la condición de medibilidad que deben satisfacer los estadísticos es importante. La razón es que, cuando se considera la muestra  $(X_1, X_2, \dots, X_n)$  como variable aleatoria, con valores en el espacio muestral  $(\mathcal{X}, \mathcal{B})$ , el estadístico  $T(X_1, X_2, \dots, X_n)$  es a su vez una variable aleatoria<sup>2</sup>, puesto que no puede preverse el valor que adoptará después de hacer efectivo el muestreo. En este sentido, debe considerarse su distribución de probabilidad, que indique con qué probabilidad (o densidad de probabilidad) puede tomar cada posible valor. Concretamente:

---

<sup>2</sup>Como función medible de  $(\mathcal{X}, \mathcal{B})$  en  $(\mathbb{R}^k, \mathcal{B}^k)$ .

Se denomina distribución en el muestreo de un estadístico  $T$ , definido en el espacio muestral  $(\mathcal{X}, \mathcal{B})$ , a la distribución de la variable aleatoria  $T(X_1, X_2, \dots, X_n)$ .

Es decir, la distribución en el muestreo del estadístico  $T$  es la medida de probabilidad que induce la distribución de la muestra,  $\mathcal{P}$ , mediante la función  $T : (\mathcal{X}, \mathcal{B}) \rightarrow (\mathbb{R}^k, \mathcal{B}^k)$ .

La noción de distribución en el muestreo asociada a un estadístico es fundamental en Estadística Matemática, ya que es dicha distribución la que permite juzgar las cualidades del estadístico en relación con los objetivos propuestos; comparar las ventajas o inconvenientes de unos estadísticos frente a otros y evaluar probabilísticamente los resultados que proporcione. Es conveniente, por tanto, tratar de aclarar este concepto mediante algunos ejemplos.

### Ejemplo 1.7

En el ejemplo 1.4 hemos considerado una muestra aleatoria simple, de tamaño 3, de lanzamientos de una moneda con probabilidad de cara  $\theta$ , obteniendo todas las muestras posibles y sus probabilidades. Imaginemos que nos interesa ahora el estadístico media muestral:

$$\bar{X} = \frac{X_1 + X_2 + X_3}{3}.$$

De acuerdo con los datos de la tabla del ejemplo 1.4, se tiene

$\bar{X}$	Probabilidad
1	$\theta^3$
2/3	$3\theta^2(1 - \theta)$
1/3	$3\theta(1 - \theta)^2$
0	$(1 - \theta)^3$

como posibles valores del estadístico, junto con sus probabilidades. Ello especifica la distribución en el muestreo del estadístico  $\bar{X}$ , mediante su función de probabilidad. Naturalmente, depende de la distribución teórica (a través del parámetro  $\theta$ ), igual que ocurría con la distribución de la muestra.

### Ejemplo 1.8

A fin de estudiar la duración de un cierto tipo de componente eléctrica, que tiene una distribución teórica exponencial de parámetro  $\theta$  como la des-

crita en el ejemplo 1.1, se obtiene una muestra aleatoria simple de tamaño  $n$ :  $(X_1, X_2, \dots, X_n)$ .

La distribución de la muestra en el espacio  $\mathbb{R}^n$  tendrá entonces función de densidad

$$f_\theta(x_1, x_2, \dots, x_n) = \begin{cases} \theta^n e^{-\theta \sum_{i=1}^n x_i} & \text{si } x_i \geq 0 \text{ para cada } i = 1, 2, \dots, n \\ 0 & \text{si } x_i < 0 \text{ para algún } i = 1, 2, \dots, n. \end{cases}$$

Consideremos el estadístico unidimensional

$$T(X_1, X_2, \dots, X_n) = X_1 + X_2 + \dots + X_n$$

cuya distribución en  $\mathbb{R}$  tendrá densidad

$$h(t) = \begin{cases} \frac{\theta^n t^{n-1} e^{-\theta t}}{(n-1)!} & \text{si } t \geq 0 \\ 0 & \text{si } t < 0, \end{cases}$$

puesto que la suma de  $n$  variables aleatorias independientes con distribución exponencial de parámetro  $\theta$  tiene distribución gamma  $\gamma(n, \theta)$ .

### Ejemplo 1.9

Dada una muestra aleatoria simple  $(X_1, X_2, \dots, X_n)$  de una variable aleatoria con distribución teórica  $F$ , en muchas ocasiones, tiene interés considerar los estadísticos ordenados, obtenidos al colocar los valores de la muestra en orden creciente:

$$X_{(1)} \leq X_{(2)} \leq X_{(3)} \leq \dots \leq X_{(n-1)} \leq X_{(n)}.$$

Así:

$$X_{(1)} = \min\{X_1, X_2, \dots, X_n\} \quad \text{menor valor muestral,}$$

$$X_{(2)} = \min(\{X_1, X_2, \dots, X_n\} - \{X_{(1)}\}) \quad \text{segundo valor muestral,}$$

$$X_{(3)} = \min(\{X_1, X_2, \dots, X_n\} - \{X_{(1)}, X_{(2)}\}) \quad \text{tercer valor muestral,}$$

$$\vdots$$

$$X_{(n)} = \max\{X_1, X_2, \dots, X_n\} \quad \text{mayor valor muestral.}$$

Hay que hacer notar que, como variables aleatorias, los estadísticos ordenados no coinciden con ninguno de los elementos de la muestra; por ejemplo:

$$X_{(1)} = \begin{cases} X_1 & \text{si } X_1 \leq X_2, X_3, \dots, X_n \\ X_2 & \text{si } X_2 \leq X_1, X_3, \dots, X_n \\ \vdots & \\ X_n & \text{si } X_n \leq X_1, X_2, \dots, X_{n-1}. \end{cases}$$

De forma que, aunque  $X_1, X_2, \dots, X_n$  sean variables aleatorias independientes e igualmente distribuidas,  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$  no son ni independientes ni igualmente distribuidas.

Podemos, a modo de ejemplo, expresar la distribución en el muestreo del estadístico bidimensional  $(X_{(1)}, X_{(n)})$ :

$$\begin{aligned} H(s, t) &= \mathcal{P}\{X_{(1)} \leq s, X_{(n)} \leq t\} = \mathcal{P}\{X_{(n)} \leq t\} - \mathcal{P}\{X_{(1)} > s, X_{(n)} \leq t\} = \\ &= \begin{cases} [F(t)]^n - [F(t) - F(s)]^n & \text{si } s \leq t \\ [F(t)]^n & \text{si } t < s, \end{cases} \end{aligned}$$

donde  $\mathcal{P}$  representa la distribución de la muestra y  $F$  la distribución teórica. ( Nótese que  $\{X_{(n)} \leq t\}$  es el suceso de que los  $n$  valores muestrales estén en  $(-\infty, t]$ , mientras que  $\{X_{(1)} > s, X_{(n)} \leq t\}$  es el suceso de que los  $n$  estén en  $(s, t]$ ).

En el caso en que  $F$  tuviese densidad  $f$ ,  $(X_{(1)}, X_{(n)})$  tendría densidad en  $\mathbb{R}^2$ :

$$h(s, t) = \begin{cases} n(n-1)[F(t) - F(s)]^{n-2}f(t)f(s) & \text{si } s \leq t \\ 0 & \text{si } s > t. \end{cases}$$

La distribución en el muestreo de un estadístico  $T$  permite considerar sus momentos:  $E[T]$ ,  $E[T^2]$ ,  $V(T)$ , etc., mediante los cuales se pueden expresar propiedades de los estadísticos que resultan esenciales desde el punto de vista de la Inferencia. Por ejemplo, es fácil establecer que: “la media en el muestreo de la media muestral coincide con la media poblacional”. Simbólicamente,  $E[\bar{X}] = \mu$ . En este momento, lo interesante de este enunciado es que su interpretación exige distinguir con claridad: el estadístico media de la distribución muestral ( $\bar{X}$ ), la distribución en el muestreo de dicho estadístico —o más exactamente, la media de tal distribución ( $E[\bar{X}]$ )— y la media de la distribución de la población ( $\mu$ ).

La distinción también es delicada en el caso del estadístico

$$T_7(x_1, x_2, \dots, x_n) = F_n^*(3)$$

de los listados al principio de esta sección. Si consideramos una muestra concreta fija  $(x_1, x_2, \dots, x_n)$ ,  $F_n^*(x)$  son valores numéricos que describen, al variar  $x$  en  $\mathbb{R}$ , la distribución muestral asociada a dicha muestra. En cambio, fijado  $x \in \mathbb{R}$ , al variar la muestra dentro del espacio muestral,

$F_n^*(x)$  es un estadístico —una variable aleatoria— cuya distribución en el muestreo será considerada más adelante.

Teóricamente la distribución en el muestreo de cualquier estadístico es fácil de obtener, al menos en el caso del muestreo aleatorio simple:

Si  $(X_1, X_2, \dots, X_n)$  es una muestra aleatoria simple de una variable aleatoria con distribución teórica  $F$  y  $T = (T_1, T_2, \dots, T_k)$  es un estadístico  $k$ -dimensional, la función de distribución en el muestreo de  $T$ ,  $H(t_1, t_2, \dots, t_k)$  es, por definición, igual a

$$\mathcal{P}\{T_1(X_1, \dots, X_n) \leq t_1, T_2(X_1, \dots, X_n) \leq t_2, \dots, T_k(X_1, \dots, X_n) \leq t_k\}$$

$$= \int_A F(dx_1)F(dx_2) \cdots F(dx_n),$$

siendo  $A$  el subconjunto del espacio muestral  $\mathcal{X}$  en el que se cumplen las desigualdades:  $T_1(x_1, \dots, x_n) \leq t_1, T_2(x_1, \dots, x_n) \leq t_2, \dots, T_k(x_1, \dots, x_n) \leq t_k$ .

Sin embargo, dicha expresión teórica raramente permite obtener resultados explícitos, utilizables a efectos prácticos; de forma que la obtención explícita de la distribución en el muestreo de un estadístico plantea a menudo serias dificultades. Tendremos que ocuparnos más adelante de esta cuestión, sobre todo en relación con distribuciones teóricas de tipo normal, antes de poder abordar los verdaderos problemas estadísticos.

## Ejercicios

### Ejercicio 1.1

De una población con distribución de Poisson de parámetro  $\lambda$ , se obtiene una muestra aleatoria simple, de tamaño  $n$ . Determinar la distribución en el muestreo de la media muestral.

### Ejercicio 1.2

Determinar la distribución en el muestreo de  $\bar{X}$  y de  $s^2$ , correspondientes a una muestra aleatoria simple, de tamaño  $n$ , de la variable aleatoria

$$X = \begin{cases} 1 & \text{con probabilidad } \theta, \\ 0 & \text{con probabilidad } 1 - \theta. \end{cases}$$

**Ejercicio 1.3**

En un dado de quinielas están marcados los resultados 1, 2 y  $X$ . Para hacer inferencias acerca de las probabilidades  $p_1$ ,  $p_2$  y  $p_X$  de cada resultado, se lanza tres veces el dado. Construir el espacio muestral y la distribución de la muestra (de tamaño 3). Hallar la distribución del estadístico  $(T_1, T_2, T_X)$ , siendo  $T_i$  la frecuencia de resultados iguales a  $i$ . Determinar la distribución de  $T_X$ , su media y su varianza. Calcular la covarianza entre  $T_1$  y  $T_2$ .

**Ejercicio 1.4**

Si  $X_1, \dots, X_n$  es una muestra aleatoria simple de una distribución absolutamente continua, con densidad  $f$ , determinar la distribución en el muestreo de la muestra ordenada:  $(X_{(1)}, \dots, X_{(n)})$ . Aplicar el resultado al caso en el que la población sea: (a)  $N(\mu, \sigma)$ , y (b) Uniforme en  $(0, 1)$ .

**Ejercicio 1.5**

Si  $X_1, \dots, X_n$  es una muestra aleatoria simple de una distribución absolutamente continua, con función de distribución  $F$  y función de densidad  $f$ , determinar la distribución en el muestreo del estadístico de orden  $r$ :  $X_{(r)}$ . Aplicar el resultado al caso en el que  $X$  sea uniforme en  $(0, 1)$ .

**Ejercicio 1.6**

Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de una distribución uniforme en  $(\theta - 1/2, \theta + 1/2)$ . Determinar la distribución en el muestreo del mínimo  $X_{(1)}$  y del máximo  $X_{(n)}$ . Determinar además sus esperanzas y deducir un estadístico, no nulo, cuya esperanza sea cero.

**Ejercicio 1.7**

Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de una distribución con función de densidad

$$f_\theta(x) = \frac{\theta}{(1+x)^{1+\theta}} \quad \text{si } x > 0,$$

siendo  $\theta$  un parámetro mayor que cero. Determinar la distribución en el muestreo del estadístico  $T = \sum_{i=1}^n \log(1 + X_i)$ .

**Ejercicio 1.8**

Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de una variable aleatoria discreta con función de probabilidad

$$f_\theta(x) = \frac{1}{\theta}, \quad \text{si } x = 1, 2, \dots, \theta,$$

siendo  $\theta$  un número natural. Si  $T_1 = 2X_1 - 1$  y  $T = X_{(n)}$ , calcular  $E[T_1|T = t]$ . Deducir que  $E[T_1|T]$  es un estadístico, función de  $T$  y calcular su esperanza.

### Ejercicio 1.9

Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de una variable aleatoria con función de densidad

$$f_\theta(x) = e^{-(x-\theta)} \exp\{-e^{-(x-\theta)}\} \quad \text{para } x \in \mathbb{R},$$

siendo  $\theta$  un número real. Determinar la distribución en el muestreo del estadístico  $T = \sum_{i=1}^n e^{-X_i}$  y calcular  $E[\log T]$  y  $V(\log T)$ .

### Ejercicio 1.10

Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de una distribución de Rayleigh cuya función de densidad es

$$f_\theta(x) = \frac{2}{\theta} x e^{-x^2/\theta} \quad \text{para } x > 0,$$

donde  $\theta$  es un parámetro positivo. Determinar la distribución en el muestreo del estadístico  $T = \sum_{i=1}^n X_i^2$ . Hallar su media y su varianza.

## Solución de los ejercicios

### Ejercicio 1.1

La muestra aleatoria simple,  $X_1, \dots, X_n$ , está compuesta por  $n$  variables aleatorias discretas, independientes y con distribución de Poisson de parámetro  $\lambda$ . Por tanto, el estadístico media muestral

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

será también una variable aleatoria discreta, que puede tomar cada uno de los valores  $r = 0, 1/n, 2/n, \dots, k/n, \dots$ . Su distribución en el muestreo puede caracterizarse, entonces, mediante su función de probabilidad:

$$\mathcal{P}_\lambda\{\bar{X} = r\} = \mathcal{P}_\lambda\left\{\sum_{i=1}^n X_i = n r\right\} = \frac{e^{-n\lambda} (n\lambda)^{nr}}{(nr)!}$$

ya que la variable aleatoria  $\sum_{i=1}^n X_i$  tiene distribución de Poisson de parámetro  $n\lambda$ . De hecho, su función característica

$$\varphi(t) = \prod_{i=1}^n \varphi_{X_i}(t) = \prod_{i=1}^n e^{\lambda(e^{it}-1)} = e^{n\lambda(e^{it}-1)}$$

es la de una distribución de Poisson, de parámetro  $n\lambda$ .

### Ejercicio 1.2

La variable aleatoria  $X$  sigue una distribución discreta, concretamente binomial  $B(1, \theta)$ , por lo que la media muestral  $\bar{X}$  (que en este caso proporciona la frecuencia de unos en la muestra), también será una variable discreta, con función de probabilidad

$$\mathcal{P}_\theta\left\{\bar{X} = \frac{r}{n}\right\} = \mathcal{P}_\theta\left\{\sum_{i=1}^n X_i = r\right\} = \binom{n}{r} \theta^r (1-\theta)^{n-r}$$

para  $r = 0, 1, 2, \dots, n$ ; puesto que la suma,  $S_n$ , de  $n$  variables independientes, binomiales  $B(1, \theta)$  tiene distribución binomial  $B(n, \theta)$ .

La varianza muestral

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

también será una variable aleatoria discreta. Obsérvese que al tomar  $X_i$  solamente los valores 1 y 0, será  $\sum_{i=1}^n X_i^2 = \sum_{i=1}^n X_i$ , de forma que

$$s^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 = \frac{S_n(n - S_n)}{n^2};$$

con lo cual  $s^2$  toma el mismo valor cuando sea  $S_n = r$  que cuando sea  $S_n = n - r$  y, en definitiva, la función de probabilidad de  $s^2$  será

$$\begin{aligned} \mathcal{P}_\theta \left\{ s^2 = \frac{r(n - r)}{n^2} \right\} &= \mathcal{P}_\theta \{S_n = r\} + \mathcal{P}_\theta \{S_n = n - r\} \\ &= \binom{n}{r} [\theta^r (1 - \theta)^{n-r} + \theta^{n-r} (1 - \theta)^r] \end{aligned}$$

para  $r = 0, 1, \dots, n$ . Nótese que, según lo anterior, es  $s^2 = \bar{X}(1 - \bar{X})$ , de manera que, en este caso, existe una dependencia funcional entre  $s^2$  y  $\bar{X}$ .

### Ejercicio 1.3

La distribución teórica queda descrita por la variable

$$Y = \begin{cases} 1 & \text{con probabilidad } p_1, \\ 2 & \text{con probabilidad } p_2, \\ X & \text{con probabilidad } p_X. \end{cases}$$

La lista de las 27 muestras posibles, de tamaño 3, junto con sus probabilidades, aparece en la tabla siguiente:

Muestra	Prob.	Muestra	Prob.	Muestra	Prob.
111	$p_1^3$	211	$p_1^2 p_2$	X11	$p_1^2 p_X$
112	$p_1^2 p_2$	212	$p_1 p_2^2$	X12	$p_1 p_2 p_X$
11X	$p_1^2 p_X$	21X	$p_1 p_2 p_X$	X1X	$p_1 p_X^2$
121	$p_1^2 p_2$	221	$p_1 p_2^2$	X21	$p_1 p_2 p_X$
122	$p_1 p_2^2$	222	$p_2^3$	X22	$p_2^2 p_X$
12X	$p_1 p_2 p_X$	22X	$p_2^2 p_X$	X2X	$p_2 p_X^2$
1X1	$p_1^2 p_X$	2X1	$p_1 p_2 p_X$	XX1	$p_1 p_X^2$
1X2	$p_1 p_2 p_X$	2X2	$p_2^2 p_X$	XX2	$p_2 p_X^2$
1XX	$p_1 p_X^2$	2XX	$p_2 p_X^2$	XXX	$p_X^3$

que representa la función de probabilidad de la muestra.

El estadístico  $(T_1, T_2, T_X)$  puede tomar los valores siguientes, con las probabilidades que se indican:

$T_1$	$T_2$	$T_X$	Prob.	$T_1$	$T_2$	$T_X$	Prob.
1	0	0	$p_1^3$	1/3	1/3	1/3	$6p_1p_2p_X$
2/3	1/3	0	$3p_1^2p_2$	0	1	0	$p_2^3$
2/3	0	1/3	$3p_1^2p_X$	0	2/3	1/3	$3p_2^2p_X$
1/3	2/3	0	$3p_1p_2^2$	0	1/3	2/3	$3p_2p_X^2$
1/3	0	2/3	$3p_1p_X^2$	0	0	1	$p_X^3$

lo cual proporciona la función de probabilidad del estadístico tridimensional.

La marginal de  $T_X$  se expresa

$$T_X = \begin{cases} 0 & \text{con probabilidad } (1 - p_X)^3 \\ 1/3 & \text{con probabilidad } 3p_X(1 - p_X)^2 \\ 2/3 & \text{con probabilidad } 3p_X^2(1 - p_X) \\ 1 & \text{con probabilidad } p_X^3 \end{cases}$$

como no era difícil de prever. Por consiguiente,

$$\mathbb{E}[T_X] = p_X \quad \text{y} \quad V(T_X) = p_X(1 - p_X)/3.$$

Por otra parte,

$$\mathbb{E}[T_1T_2] = \frac{2}{9}3p_1^2p_2 + \frac{2}{9}3p_1p_2^2 + \frac{1}{9}6p_1p_2p_X = \frac{2}{3}p_1p_2.$$

Y, como  $\mathbb{E}[T_1] = p_1$  y  $\mathbb{E}[T_2] = p_2$ , resulta

$$\text{Cov}(T_1, T_2) = -\frac{1}{3}p_1p_2.$$

#### Ejercicio 1.4

La densidad de la muestra aleatoria simple  $(X_1, X_2, \dots, X_n)$  vale

$$f(x_1, x_2, \dots, x_n) = f(x_1)f(x_2) \cdots f(x_n)$$

en cualquier punto  $(x_1, x_2, \dots, x_n)$  de  $\mathbb{R}^n$ .

En cambio, la densidad  $g(y_1, y_2, \dots, y_n)$  de la muestra ordenada, está concentrada en la región  $\Delta$  de  $\mathbb{R}^n$  en la que se cumple:  $y_1 < y_2 < \dots < y_n$ ; de forma que es

$$g(y_1, y_2, \dots, y_n) = 0 \quad \text{si } (y_1, y_2, \dots, y_n) \notin \Delta.$$

Por el contrario, al ordenar la muestra, cada punto  $(y_1, y_2, \dots, y_n)$  de  $\Delta$  acumula la densidad de probabilidad  $f(x_1, x_2, \dots, x_n)$  de cada uno de los puntos  $(x_1, x_2, \dots, x_n)$  obtenidos al permutar de forma arbitraria las coordenadas de  $(y_1, y_2, \dots, y_n)$ . Puesto que hay  $n!$  de tales puntos y, en todos ellos, es  $f(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(y_i)$ , resulta

$$g(y_1, y_2, \dots, y_n) = n! f(y_1)f(y_2) \cdots f(y_n) \quad \text{para } (y_1, y_2, \dots, y_n) \in \Delta.$$

También puede razonarse: Dados  $y_1 < y_2 < \dots < y_n$ , para construir una muestra en la cual  $y_1 < X_{(1)} < y_1 + dy_1$ ,  $y_2 < X_{(2)} < y_2 + dy_2, \dots$ ,  $y_n < X_{(n)} < y_n + dy_n$ , uno de los  $n$  elementos muestrales ha de escogerse en  $(y_1, y_1 + dy_1)$  —lo cual tiene probabilidad  $nf(y_1)dy_1$ —; de los  $n - 1$  restantes ha de escogerse uno en  $(y_2, y_2 + dy_2)$  —lo cual tiene probabilidad  $(n - 1)f(y_2)dy_2$ —; y así sucesivamente. Es decir

$$\begin{aligned} \mathcal{P}\{y_1 < X_{(1)} < y_1 + dy_1, \dots, y_n < X_{(n)} < y_n + dy_n\} = \\ nf(y_1)(n - 1)f(y_2) \cdots f(y_n)dy_1dy_2 \cdots dy_n, \end{aligned}$$

de donde la densidad  $g(y_1, y_2, \dots, y_n)$  resulta

$$\begin{cases} n! f(y_1)f(y_2) \cdots f(y_n) & \text{si } y_1 < y_2 < \dots < y_n, \\ 0 & \text{en caso contrario.} \end{cases}$$

En particular, si  $X$  tiene distribución  $N(\mu, \sigma)$ ,  $g(y_1, y_2, \dots, y_n)$  vale:

$$\begin{cases} \frac{n!}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right\} & \text{si } y_1 < y_2 < \dots < y_n, \\ 0 & \text{en caso contrario.} \end{cases}$$

Y, si  $X$  es uniforme en  $(0, 1)$ , la densidad será

$$g(y_1, y_2, \dots, y_n) = \begin{cases} n! & \text{si } 0 < y_1 < y_2 < \dots < y_n < 1 \\ 0 & \text{en caso contrario.} \end{cases}$$

**Ejercicio 1.5**

La densidad en el muestreo del estadístico  $X_{(r)}$  puede obtenerse como marginal de la densidad del vector  $(X_{(1)}, \dots, X_{(n)})$ , integrando con respecto a las variables  $y_1 < \dots < y_{r-1}$  en  $(-\infty, y_r)$  y con respecto a las variables  $y_{r+1} < \dots < y_n$  en  $(y_r, \infty)$ . Como

$$\int_{-\infty}^{y_r} \cdots \int_{-\infty}^{y_3} \int_{-\infty}^{y_2} f(y_1) f(y_2) \cdots f(y_{r-1}) dy_1 dy_2 \cdots dy_{r-1} = \frac{F(y_r)^{r-1}}{(r-1)!}$$

e

$$\int_{y_r}^{\infty} \cdots \int_{y_{n-2}}^{\infty} \int_{y_{n-1}}^{\infty} f(y_{r+1}) \cdots f(y_{n-1}) f(y_n) dy_n dy_{n-1} \cdots dy_{r+1} = \frac{[1 - F(y_r)]^{n-r}}{(n-r)!}$$

la densidad de  $X_{(r)}$  resulta ser

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

Alternativamente se puede determinar la función de distribución:

$$G(y_r) = \mathcal{P}\{X_{(r)} \leq y_r\} = \mathcal{P}\{\text{Al menos } r \text{ de las } X_i \text{ sean } \leq y_r\} =$$

$$\sum_{j=r}^n \mathcal{P}\{j \text{ de las } X_i \text{ sean } \leq y_r\} = \sum_{j=r}^n \binom{n}{j} [F(y_r)]^j [1 - F(y_r)]^{n-j}$$

puesto que cada  $X_i$  tiene probabilidad  $F(y_r)$  de ser  $\leq y_r$  y probabilidad  $1 - F(y_r)$  de ser  $> y_r$ .

Derivando respecto a  $y_r$  se obtiene

$$\begin{aligned} g(y_r) &= f(y_r) \left[ \sum_{j=r}^n \binom{n}{j} j F(y_r)^{j-1} (1 - F(y_r))^{n-j} \right. \\ &\quad \left. - \sum_{j=r}^{n-1} \binom{n}{j} (n-j) F(y_r)^j (1 - F(y_r))^{n-j-1} \right]; \end{aligned}$$

lo cual, una vez cancelados los sumandos iguales, produce la misma expresión de  $g(y_r)$  obtenida antes.

Más simplemente, si  $h$  es suficientemente pequeño para que se pueda despreciar la probabilidad de que en  $(y_r, y_r + h)$  haya dos o más de las  $X_i$ ,  $\mathcal{P}\{y_r < X_{(r)} < y_r + h\}$  valdrá

$$\frac{n!}{(r-1)!(n-r)!} F(y_r)^{r-1} [F(y_r + h) - F(y_r)] [1 - F(y_r + h)]^{n-r}$$

puesto que tiene que haber  $r - 1$  de las  $X_i$  en  $(-\infty, y_r)$ , una en  $(y_r, y_r + h)$  y las  $n - r$  restantes en  $(y_r, \infty)$ . Por lo tanto, dividiendo por  $h$  y haciendo  $h$  tender a 0, se obtiene nuevamente:

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

Si  $X$  es uniforme en  $(0, 1)$ , sustituyendo en la expresión anterior  $F(y_r) = y_r$ , resulta

$$g(y_r) = \frac{n!}{(r-1)!(n-r)!} y_r^{r-1} (1-y_r)^{n-r} \quad \text{para } y_r \in (0, 1),$$

que es la función de densidad de una beta  $\beta(r, n-r+1)$ .

### Ejercicio 1.6

A partir del ejercicio 1.5 se obtiene directamente que las funciones de densidad del mínimo y del máximo son respectivamente:

$$g(y_1) = n [1 - F(y_1)]^{n-1} f(y_1) = n [1/2 + \theta - y_1]^{n-1}$$

para  $\theta - 1/2 < y_1 < \theta + 1/2$ , y

$$h(y_n) = n [F(y_n)]^{n-1} f(y_n) = n [y_n - \theta + 1/2]^{n-1}$$

para  $\theta - 1/2 < y_n < \theta + 1/2$ .

Sus esperanzas serán

$$\mathbb{E}[X_{(1)}] = \int_{\theta-1/2}^{\theta+1/2} y_1 n \left[ \frac{1}{2} + \theta - y_1 \right]^{n-1} dy_1 = \theta - \frac{1}{2} + \frac{1}{n+1},$$

$$\mathbb{E}[X_{(n)}] = \int_{\theta-1/2}^{\theta+1/2} y_n n \left[ y_n - \theta + \frac{1}{2} \right]^{n-1} dy_n = \theta + \frac{1}{2} - \frac{1}{n+1}.$$

Por último, como es

$$\mathbb{E}[X_{(n)} - X_{(1)}] = \frac{n-1}{n+1},$$

el estadístico buscado puede ser

$$X_{(n)} - X_{(1)} - \frac{n-1}{n+1}.$$

**Ejercicio 1.7**

La variable aleatoria  $Y = \log(1 + X)$  tiene por función de densidad

$$g(y) = f(e^y - 1) \cdot e^y = \frac{\theta}{(1 + e^y - 1)^{1+\theta}} e^y = \theta e^{-\theta y} \quad \text{para } y > 0.$$

Es decir, una distribución exponencial  $\gamma(1, \theta)$ . La función característica de  $T$  será, por tanto,

$$\varphi_T(t) = \prod_{i=1}^n \varphi_{Y_i}(t) = \left(1 - \frac{it}{\theta}\right)^{-n};$$

con lo cual  $T$  tiene distribución gamma  $\gamma(n, \theta)$ .

**Ejercicio 1.8**

Como la función de distribución de las  $X_i$  es

$$F_\theta(x) = \begin{cases} 0 & \text{si } x < 1 \\ j/\theta & \text{si } j \leq x < j + 1, \quad \text{para } j = 1, \dots, \theta - 1 \\ 1 & \text{si } x \geq \theta, \end{cases}$$

la función de distribución del máximo,  $T$ , será

$$[F_\theta(t)]^n = \begin{cases} 0 & \text{si } t < 1 \\ (j/\theta)^n & \text{si } j \leq t < j + 1, \quad \text{para } j = 1, \dots, \theta - 1 \\ 1 & \text{si } t \geq \theta; \end{cases}$$

de forma que, su función de probabilidad resulta

$$\mathcal{P}\{T = t\} = \frac{t^n - (t - 1)^n}{\theta^n} \quad \text{para } t = 1, 2, \dots, \theta.$$

Por otro lado, para  $x_1 = 1, 2, \dots, \theta$ ,  $\mathcal{P}\{T = t | X_1 = x_1\}$  vale

$$\begin{cases} 0 & \text{si } t < x_1 \\ \prod_{i=2}^n P\{X_i \leq t\} = \left(\frac{t}{\theta}\right)^{n-1} & \text{si } t = x_1 \\ \mathcal{P}\{\max_{2 \leq i \leq n} X_i = t\} = \frac{t^{n-1} - (t - 1)^{n-1}}{\theta^{n-1}} & \text{si } t > x_1 \end{cases}$$

luego

$$\begin{aligned}\mathcal{P}\{X_1 = x_1 | T = t\} &= \frac{\mathcal{P}\{X_1 = x_1\} \mathcal{P}\{T = t | X_1 = x_1\}}{\mathcal{P}\{T = t\}} \\ &= \begin{cases} 0 & \text{si } x_1 > t \\ \frac{t^{n-1}}{t^n - (t-1)^n} & \text{si } x_1 = t \\ \frac{t^{n-1} - (t-1)^{n-1}}{t^n - (t-1)^n} & \text{si } x_1 < t. \end{cases}\end{aligned}$$

Entonces

$$\begin{aligned}\mathbb{E}[T_1 | T = t] &= \sum_{x_1=1}^{\theta} (2x_1 - 1) \mathcal{P}\{X_1 = x_1 | T = t\} \\ &= \sum_{x_1=1}^{t-1} (2x_1 - 1) \frac{t^{n-1} - (t-1)^{n-1}}{t^n - (t-1)^n} + (2t - 1) \frac{t^{n-1}}{t^n - (t-1)^n} \\ &= \frac{(2t-1)t^{n-1} + [t^{n-1} - (t-1)^{n-1}](t-1)^2}{t^n - (t-1)^n} \\ &= \frac{t^{n+1} - (t-1)^{n+1}}{t^n - (t-1)^n}.\end{aligned}$$

Y, por consiguiente,

$$\mathbb{E}[T_1 | T] = \frac{T^{n+1} - (T-1)^{n+1}}{T^n - (T-1)^n}$$

que es un estadístico,  $T^*$ , función de la muestra a través de  $T$ .

La esperanza de  $T^* = \mathbb{E}[T_1 | T]$  debe coincidir con la esperanza de  $T_1$ :

$$\mathbb{E}[T_1] = \sum_{j=1}^{\theta} (2j - 1) \frac{1}{\theta} = \theta.$$

Para hacer la comprobación directa, es útil observar que

$$\sum_{t=1}^{\theta} [t^n - (t-1)^n] = \theta^n,$$

puesto que  $[t^n - (t-1)^n]/\theta^n$  es la función de probabilidad de  $T$ . Entonces

$$\begin{aligned} E[T^*] &= \sum_{t=1}^{\theta} \frac{t^{n+1} - (t-1)^{n+1}}{t^n - (t-1)^n} \cdot \frac{t^n - (t-1)^n}{\theta^n} \\ &= \frac{1}{\theta^n} \sum_{t=1}^{\theta} [t^{n+1} - (t-1)^{n+1}] = \frac{1}{\theta^n} \theta^{n+1} = \theta. \end{aligned}$$

### Ejercicio 1.9

La función de densidad de  $Y = e^{-X}$  es

$$g_{\theta}(y) = e^{\theta} e^{-e^{\theta} y} \quad \text{para } y > 0;$$

es decir, la de una exponencial de parámetro  $e^{\theta}$ . Por tanto,  $T = \sum_{i=1}^n e^{-X_i}$  tendrá una distribución gamma  $\gamma(n, e^{\theta})$  de densidad

$$f_{\theta}(t) = \frac{e^{n\theta} t^{n-1} e^{-e^{\theta} t}}{\Gamma(n)} \quad \text{para } t > 0.$$

Utilizando las derivadas de la función  $\Gamma(n) = \int_0^{\infty} x^{n-1} e^{-x} dx$ , se obtiene

$$\begin{aligned} E[\log T] &= \frac{1}{\Gamma(n)} \int_0^{\infty} \log t e^{n\theta} t^{n-1} e^{-e^{\theta} t} dt \\ &= \frac{1}{\Gamma(n)} \int_0^{\infty} (\log u - \theta) u^{n-1} e^{-u} du = \frac{\Gamma'(n)}{\Gamma(n)} - \theta \end{aligned}$$

y

$$\begin{aligned} E[\log^2 T] &= \frac{1}{\Gamma(n)} \int_0^{\infty} \log^2 t e^{n\theta} t^{n-1} e^{-e^{\theta} t} dt \\ &= \frac{1}{\Gamma(n)} \int_0^{\infty} (\log u - \theta)^2 u^{n-1} e^{-u} du = \frac{\Gamma''(n)}{\Gamma(n)} - 2\theta \frac{\Gamma'(n)}{\Gamma(n)} + \theta^2. \end{aligned}$$

Por último, será

$$V(\log T) = E[\log^2 T] - E[\log T]^2 = \frac{\Gamma''(n)}{\Gamma(n)} - \left( \frac{\Gamma'(n)}{\Gamma(n)} \right)^2.$$

### Ejercicio 1.10

La función de densidad de  $Y = X^2$  es

$$g_{\theta}(y) = f_{\theta}(\sqrt{y}) \frac{1}{2\sqrt{y}} = \frac{1}{\theta} e^{-y/\theta} \quad \text{para } y > 0;$$

es decir, una gamma  $\gamma(1, 1/\theta)$ . La de  $T = \sum_{i=1}^n X_i^2$  será  $\gamma(n, 1/\theta)$ .  
Por consiguiente

$$\mathbb{E}[T] = n\theta \quad \text{y} \quad V(T) = n\theta^2.$$

# Capítulo 2

## Propiedades de la distribución muestral

### 2.1. Introducción

En este capítulo se analizan con detenimiento las propiedades de la distribución muestral y de sus características, consideradas como variables aleatorias dependientes de los elementos aleatorios que componen la muestra.

Los resultados en esta dirección, y en particular los relativos al comportamiento de la distribución muestral cuando el tamaño de la muestra crece, proporcionan una justificación muy completa de las afirmaciones imprecisas del capítulo anterior, acerca de por qué puede esperarse obtener información sobre la distribución de la población a partir de la distribución muestral que, una vez realizado el muestreo, sólo depende de los valores numéricos incluidos en la muestra. Sin duda por tal razón, el teorema de Glivenko-Cantelli, que se incluye en la sección 2.3, recibe a menudo la denominación de teorema central de la Estadística.

La imposibilidad, en la mayor parte de los casos, de determinar explícitamente la distribución exacta en el muestreo de las principales características de la distribución muestral (cuantiles y momentos), hace que cobren interés las distribuciones asintóticas de tales características, que se estudian, respectivamente, en las secciones 2.5 y 2.6. Ello permite, al menos, obtener resultados aproximados, basados en dichos estadísticos, cuando el tamaño muestral sea grande, y asegura que las características muestrales tendrán, entonces, muy poca dispersión alrededor de las características poblacionales análogas.

## 2.2. Distribución muestral y sus características

Consideremos una población descrita por una variable aleatoria unidimensional  $X$  con función de distribución teórica  $F$ .

La distribución muestral asociada a una muestra aleatoria  $(x_1, x_2, \dots, x_n)$  ha sido ya introducida, como aproximación de la distribución teórica a partir de las observaciones muestrales, mediante la expresión:

$$F_n^*(x) = \frac{\text{número de observaciones } x_i \text{ en el intervalo } (-\infty, x]}{n} \quad (2.1)$$

para cada  $x \in \mathbb{R}$ .

Equivalentemente, en términos de los valores  $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ , de la muestra ordenada en orden creciente, será:

$$F_n^*(x) = \begin{cases} 0 & \text{si } x < x_{(1)} \\ \vdots & \\ j/n & \text{si } x_{(j)} \leq x < x_{(j+1)} \quad (j = 1, 2, \dots, n) \\ \vdots & \\ 1 & \text{si } x \geq x_{(n)}. \end{cases} \quad (2.2)$$

Por último, también se puede expresar<sup>1</sup>:

$$F_n^*(x) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x]}(x_i) \quad \text{para cada } x \in \mathbb{R}. \quad (2.3)$$

Mientras que (2.1) recalca la semejanza formal entre la distribución muestral y la teórica, y la expresión (2.2) sugiere cómodamente la representación gráfica de la distribución muestral, la fórmula (2.3) tiene la ventaja de expresarla mediante una suma, que se convertirá en suma de variables aleatorias independientes cuando se reemplace cada  $x_i$  por el correspondiente elemento aleatorio  $X_i$  de una muestra aleatoria simple.

En cualquier caso,  $F_n^*$  es una distribución discreta, que asigna probabilidad  $1/n$  a cada una de las observaciones muestrales, de manera que todos

<sup>1</sup>  $I_A(x)$  siempre representará la función indicatriz del conjunto  $A$ , es decir

$$I_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases}$$

De forma algo incorrecta, se escribirá también  $I_{\{x \in A\}}$ .

sus momentos, llamados *momentos muestrales*, existen y valen:

$$a_k = \frac{1}{n} \sum_{i=1}^n x_i^k, \quad b_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k,$$

según que se trate del momento de orden  $k$  respecto al origen, o del momento de orden  $k$  respecto a la media:

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

Por supuesto, entre los momentos tienen particular importancia la *media muestral*,  $\bar{x}$ , y la *varianza muestral*,

$$s^2 = b_2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

ya introducidos en la sección 1.4.

También tiene interés considerar los *cuantiles muestrales*,  $c_p$ , definidos, para cada  $p \in (0, 1)$  como aquellos valores que verifican simultáneamente<sup>2</sup>:

$$F_n^*(c_p) \geq p \quad \text{y} \quad F_n^*(c_p^-) \leq p.$$

Como  $F_n^*$  es una distribución discreta, los cuantiles muestrales se expresan cómodamente en función de los elementos de la muestra ordenada:

▷ Si  $np \notin \mathbb{N}$ , y  $[np]$  es la parte entera de  $np$ , será:

$$c_p = x_{([np]+1)}$$

puesto que

$$F_n^*(x_{([np]+1)}) = \frac{[np] + 1}{n} > p \quad \text{y} \quad F_n^*(x_{([np]+1)}^-) = F_n^*(x_{([np])}) = \frac{[np]}{n} < p.$$

▷ Si  $np \in \mathbb{N}$ ,  $c_p$  puede no quedar únicamente determinado, al poder ser  $F_n^*(x) = p$  para cualquier  $x \in [x_{(np)}, x_{(np+1)})$ . Se deshace entonces la ambigüedad fijando:

$$c_p = \frac{x_{(np)} + x_{(np+1)}}{2}.$$

Entre los cuantiles, destaca la mediana,  $M = c_{1/2}$ , que es simplemente el valor central de la muestra ordenada, en el caso de que  $n$  sea impar, y la semisuma de los dos valores centrales cuando  $n$  sea par. (Véase el ejercicio 2.1.)

---

<sup>2</sup>  $F(x^-) = \lim_{y \uparrow x} F(y).$

## 2.3. Distribución en el muestreo y comportamiento asintótico

Supondremos a lo largo del resto del capítulo que la muestra va a elegirse por muestreo aleatorio simple, de manera que sus elementos deben considerarse como variables aleatorias,  $(X_1, X_2, \dots, X_n)$ , independientes entre sí y con la misma distribución teórica  $F$ . Ello determina la distribución conjunta  $\mathcal{P}$  de  $(X_1, X_2, \dots, X_n)$ .

En tal caso, el valor de la función de distribución muestral, para cada  $x \in \mathbb{R}$  fijo, debe considerarse también como una variable aleatoria, que se puede expresar en la forma:

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

y cuya distribución en el muestreo es sencilla de obtener. De hecho las variables  $I_{(-\infty, x]}(X_i)$  son independientes y con distribución común:

$$\begin{cases} \mathcal{P}\{I_{(-\infty, x]}(X_i) = 1\} = \mathcal{P}\{X_i \leq x\} = F(x), \\ \mathcal{P}\{I_{(-\infty, x]}(X_i) = 0\} = \mathcal{P}\{X_i > x\} = 1 - F(x); \end{cases}$$

luego,  $nF_n^*(x) = \sum_{i=1}^n I_{(-\infty, x]}(X_i)$  tiene distribución binomial  $B(n, F(x))$ . Es decir, para  $k = 0, 1, 2, \dots, n$ ,

$$\mathcal{P}\left\{F_n^*(x) = \frac{k}{n}\right\} = \binom{n}{k} F(x)^k [1 - F(x)]^{n-k}.$$

Por consiguiente

$$\mathbb{E}[F_n^*(x)] = F(x) \quad \text{y} \quad V(F_n^*(x)) = \frac{F(x)(1 - F(x))}{n}.$$

Además, puesto que  $nF_n^*(x)$  es suma de variables aleatorias independientes e igualmente distribuidas, el teorema central del límite permite afirmar que, cuando  $n \rightarrow \infty$ ,

$$\frac{nF_n^*(x) - nF(x)}{\sqrt{nF(x)(1 - F(x))}} = \sqrt{n} \frac{F_n^*(x) - F(x)}{\sqrt{F(x)(1 - F(x))}} \xrightarrow{d} N(0, 1);$$

de manera que, cuando el tamaño muestral es grande, la distribución de  $F_n^*(x)$  es aproximadamente  $N(F(x), \sqrt{F(x)(1 - F(x))/n})$ .

Ello indica claramente que, a medida que  $n$  crece, la distribución en el muestreo de  $F_n^*(x)$  es cada vez más concentrada alrededor del valor  $F(x)$ ; o, dicho de otra manera, con un tamaño de muestra grande es muy probable que la muestra que se obtenga proporcione un valor de  $F_n^*(x)$  muy próximo al valor desconocido de  $F(x)$ . Los ejercicios 2.2 y 2.3 muestran como utilizar estas conclusiones.

Para precisar más, supongamos que el proceso de muestreo se va a prolongar indefinidamente observando sucesivamente los valores de una sucesión  $\{X_i\}_{i=1}^\infty$  de variables aleatorias independientes y con distribución común  $F$ . En tal caso,  $\mathcal{P}$  designará la distribución conjunta de la sucesión muestral  $\{X_i\}_{i=1}^\infty$ . Puesto que  $\{I_{(-\infty,x]}(X_i)\}_{i=1}^\infty$  es una sucesión de variables aleatorias independientes e igualmente distribuidas con media  $F(x)$ , la ley fuerte de los grandes números permite concluir que, cuando  $n \rightarrow \infty$ ,

$$\frac{1}{n} \sum_{i=1}^n I_{(-\infty,x]}(X_i) = F_n^*(x) \longrightarrow F(x) \quad \mathcal{P}\text{-casi seguro.}$$

Es decir que hay probabilidad cero de obtener una sucesión muestral a lo largo de la cual, los sucesivos valores de  $F_n^*(x)$ , ( $n = 1, 2, 3, \dots$ ), no constituyan una sucesión convergente a  $F(x)$ .

Aunque esta consecuencia de la ley fuerte sea muy satisfactoria, por cuanto garantiza que es posible conocer, con la precisión que se desee, el valor de  $F(x)$ , en cada  $x \in \mathbb{R}$  fijo, interesaría poder asegurar que dicha aproximación se produce simultáneamente para todos los valores de  $x$ ; lo cual no es inmediato pues, aunque el suceso relativo a  $x$  se produzca con probabilidad 1, el número de valores de  $x$  no es numerable. En ese sentido, el siguiente resultado prueba que  $F_n^*(x)$  converge a  $F(x)$ , uniformemente en  $x$ , con probabilidad uno.

**Teorema de Glivenko-Cantelli:** Sea  $\{X_i\}_{i=1}^\infty$  una sucesión de variables aleatorias independientes y con distribución común  $F$ . Si  $F_n^*$  es la función de distribución muestral asociada a la muestra aleatoria simple  $(X_1, X_2, \dots, X_n)$  y

$$\Delta_n = \sup_{x \in \mathbb{R}} |F_n^*(x) - F(x)|,$$

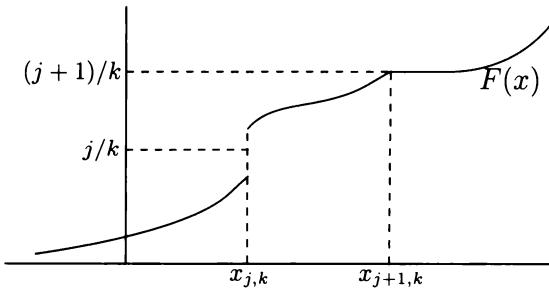
entonces

$$\lim_{n \rightarrow \infty} \Delta_n = 0 \quad \mathcal{P}\text{-casi seguro.}$$

Demostración: Ya se sabe que  $F_n^*(x) \rightarrow F(x)$ , con probabilidad 1, para cada  $x \in \mathbb{R}$  fijo. De la misma manera, considerando las variables aleatorias independientes e idénticamente distribuidas  $I_{(-\infty, x)}(X_i)$  de media  $F(x^-)$ , se obtiene que  $F_n^*(x^-) \rightarrow F(x^-)$ ,  $\mathcal{P}$ -c.s., para cada  $x \in \mathbb{R}$  fijo.

Para cada número natural  $k$ , y  $j = 1, 2, \dots, k$ , consideremos los puntos

$$x_{j,k} = \min \left\{ x \in \mathbb{R} \mid F(x^-) \leq \frac{j}{k} \leq F(x) \right\}$$



y los sucesos del espacio muestral:

$$A_{j,k} = \{F_n^*(x_{j,k}) \rightarrow F(x_{j,k})\},$$

$$B_{j,k} = \{F_n^*(x_{j,k}^-) \rightarrow F(x_{j,k}^-)\},$$

$$D_k = \bigcap_{j=1}^k (A_{j,k} \cap B_{j,k}),$$

$$D = \bigcap_{k=1}^{\infty} D_k.$$

Nótese que  $D_k$  es el suceso definido por la condición de que todas las diferencias  $|F_n^*(x_{j,k}) - F(x_{j,k})|$  y  $\left|F_n^*(x_{j,k}^-) - F(x_{j,k}^-)\right|$ , para un  $k$  fijo, converjan a cero, al tender  $n$  a infinito.  $D$  es el suceso de que esto ocurra simultáneamente para todos los valores de  $k$ .

Según lo anterior,  $\mathcal{P}(A_{j,k}) = \mathcal{P}(B_{j,k}) = 1$  para cualquier  $j$  y cualquier  $k$ , luego  $\mathcal{P}(D_k) = 1$ , para cualquier  $k$  y, también,  $P(D) = 1$ .

Obsérvese ahora que, si  $x \in [x_{j,k}, x_{j+1,k}]$ , es

$$\begin{cases} F(x_{j,k}) \leq F(x) \leq F(x_{j+1,k}) \\ F_n^*(x_{j,k}) \leq F_n^*(x) \leq F_n^*(x_{j+1,k}) \end{cases}$$

y por tanto, como  $F(x_{j+1,k}^-) - F(x_{j,k}) \leq 1/k$ , es

$$\begin{cases} F_n^*(x) - F(x) \leq F_n^*(x_{j+1,k}^-) - F(x_{j,k}) \leq F_n^*(x_{j+1,k}^-) - F(x_{j+1,k}^-) + \frac{1}{k} \\ F_n^*(x) - F(x) \geq F_n^*(x_{j,k}) - F(x_{j+1,k}^-) \geq F_n^*(x_{j,k}) - F(x_{j,k}) - \frac{1}{k}; \end{cases}$$

con lo cual, si  $\delta_n^{(k)}$  es la mayor entre todas las diferencias  $|F_n^*(x_{j,k}) - F(x_{j,k})|$  y  $|F_n^*(x_{j,k}^-) - F(x_{j,k}^-)|$  (para  $n$  y  $k$  fijos), se tiene

$$\begin{cases} F_n^*(x) - F(x) \leq \delta_n^{(k)} + \frac{1}{k}, \\ F_n^*(x) - F(x) \geq -\delta_n^{(k)} - \frac{1}{k}. \end{cases}$$

Por consiguiente, cualquiera que sea  $k \in \mathbb{N}$ , es

$$\Delta_n \leq \delta_n^{(k)} + \frac{1}{k}.$$

Ahora bien, si se verifica el suceso  $D$ , para cualquier  $k \in \mathbb{N}$  y cualquier  $\varepsilon > 0$ , es  $\delta_n^{(k)} < \varepsilon$ , a partir de un  $n$  en adelante. De forma que  $\Delta_n < \varepsilon + 1/k$  a partir de un  $n$  en adelante. Es decir  $\lim_{n \rightarrow \infty} \Delta_n = 0$  siempre que se verifique el suceso  $D$ . Como  $\mathcal{P}(D) = 1$ , resulta que  $\lim_{n \rightarrow \infty} \Delta_n = 0$ ,  $\mathcal{P}$ -c.s.

Imaginando una banda de amplitud  $\varepsilon$ , arbitrariamente estrecha, alrededor de la distribución teórica  $F$ , el resultado anterior garantiza que hay probabilidad uno, de que la distribución muestral  $F_n^*$  llegue a estar contenida dentro de la banda, si se deja crecer suficientemente el tamaño muestral. Aunque la distribución teórica sea desconocida, esto permite conocerla con la precisión que se desee, a costa de tomar muestras suficientemente extensas.

## 2.4. Distribución en el muestreo de los momentos muestrales

Las distribuciones en el muestreo de los momentos muestrales (considerados como variables aleatorias definidas a partir de una muestra aleatoria simple  $(X_1, X_2, \dots, X_n)$ ) no admiten una expresión general, sino que habrían de ser estudiadas en cada caso particular, según la forma funcional de la distribución poblacional. Aún así, la mayor parte de las veces su obtención no

es nada sencilla como podrá apreciarse, en el próximo capítulo, cuando nos ocupemos de esta cuestión para el caso de poblaciones normales.

En cambio los momentos de los momentos muestrales —o, mejor dicho, de sus distribuciones en el muestreo— admiten expresiones generales, mediante cálculos a menudo laboriosos, algunos de los cuales se remitirán a los ejercicios.

Concretamente, para los momentos muestrales respecto al origen,

$$a_k = \frac{1}{n} \sum_{i=1}^n X_i^k$$

se tiene:

1)  $E[a_k] = \frac{1}{n} \sum_{i=1}^n E[X_i^k] = \alpha_k$ , supuesto que existe  $\alpha_k$ , momento de orden  $k$  respecto al origen, de la distribución teórica.

2)  $V(a_k) = \frac{1}{n^2} \sum_{i=1}^n V(X_i^k) = \frac{1}{n} V(X_1^k) = \frac{1}{n} [E[X_1^{2k}] - E[X_1^k]^2] = \frac{\alpha_{2k} - \alpha_k^2}{n}$

habida cuenta de que  $X_1, \dots, X_n$  son variables aleatorias independientes e igualmente distribuidas y supuesto que existe el momento poblacional  $\alpha_{2k}$ .

En particular para la media muestral  $\bar{X} = a_1$ , es

$$E[\bar{X}] = \mu \quad \text{y} \quad V(\bar{X}) = \frac{\alpha_2 - \alpha_1^2}{n} = \frac{\sigma^2}{n},$$

supuesto respectivamente que existan la media poblacional  $\mu = \alpha_1$ , y la varianza poblacional,  $\sigma^2$ . En los ejercicios 2.4 y 2.5 pueden verse las expresiones de los momentos de orden 3 y 4 de  $\bar{X}$ .

En cuanto a los momentos muestrales  $b_k$ , respecto de la media  $\bar{X}$ , tiene interés sobre todo el caso de la varianza muestral

$$s^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2,$$

para la cual

$$E[s^2] = \frac{1}{n} \sum_{i=1}^n E[X_i^2] - E[\bar{X}^2] = \alpha_2 - \frac{\sigma^2}{n} - \mu^2 = \frac{n-1}{n} \sigma^2,$$

dado que  $E[\bar{X}^2] = V(\bar{X}) + E[\bar{X}]^2 = \sigma^2/n + \mu^2$ .

Debido a este resultado es muy frecuente considerar la *cuasivarianza muestral*:

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

que verificará  $E[S^2] = n/(n-1) E[s^2] = \sigma^2$ .

El ejercicio 2.6 contiene el cálculo de la varianza de  $s^2$ , mientras que en los ejercicios 2.7 y 2.8 se calculan la esperanza de  $b_3$  y  $b_4$  respectivamente.

Acerca de la distribución en el muestreo propiamente dicha, el único momento muestral cuya distribución se puede determinar con cierta generalidad es la media muestral,  $\bar{X}$ . De hecho, si  $(X_1, X_2, \dots, X_n)$  es una muestra aleatoria simple de una distribución teórica con función característica  $\varphi$ , la función característica de  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  es

$$\psi(t) = E[e^{it\bar{X}}] = E[e^{it/n \sum_{j=1}^n X_j}] = \left[ \varphi\left(\frac{t}{n}\right) \right]^n,$$

puesto que  $X_1, \dots, X_n$  son variables independientes y con función característica  $\varphi$ .

Así por ejemplo, si la distribución de la población es una distribución gamma  $\gamma(p, a)$ , con función característica  $\varphi(t) = (1 - it/a)^{-p}$ ,  $\bar{X}$  tiene por función característica

$$\psi(t) = \left(1 - \frac{it}{an}\right)^{-pn};$$

es decir que  $\bar{X}$  tiene distribución en el muestreo gamma  $\gamma(pn, an)$ .

## 2.5. Comportamiento asintótico de los cuantiles muestrales

Puesto que a medida que el tamaño muestral crece, la función de distribución muestral se aproxima, con probabilidad 1 y uniformemente, hacia la distribución teórica, parece lógico pensar que ha de suceder lo mismo entre los cuantiles muestrales y los de la distribución teórica. De hecho, ello es así cuando los cuantiles poblacionales son únicos, como muestra el siguiente corolario del teorema de Glivenko-Cantelli:

**Teorema 2.1:** Sea  $\{X_i\}_{i=1}^{\infty}$  una sucesión de variables aleatorias independientes y con distribución común  $F$ , que tenga un único cuantil de orden  $p \in (0, 1)$ :  $x_p$ . Si  $c_p(n)$  es el cuantil muestral de orden  $p$ , asociado a la muestra aleatoria simple  $(X_1, X_2, \dots, X_n)$ , se verifica

$$\lim_{n \rightarrow \infty} c_p(n) = x_p \quad \text{casi seguro}$$

(respecto a la distribución  $\mathcal{P}$  de  $\{X_i\}_{i=1}^{\infty}$ ).

Demostración: Fijado  $k \in \mathbb{N}$ , como el cuantil  $x_p$ , que verifica  $F(x_p^-) \leq p \leq F(x_p)$ , es único, ha de ser  $F(x_p + 1/k) > p$ . Entonces, casi seguro,

$$\lim_{n \rightarrow \infty} F_n^* \left( x_p + \frac{1}{k} \right) > p,$$

de manera que  $F_n^*(x_p + 1/k) > p$  a partir de cierto  $n_0$  en adelante, o lo que es lo mismo,  $c_p(n) \leq x_p + 1/k$ , a partir de cierto  $n_0$  en adelante.

Análogamente, como  $F(x_p - 1/k) < p$ , con probabilidad 1 será,

$$F_n^* \left( x_p - \frac{1}{k} \right) < p$$

o, equivalentemente,  $c_p(n) \geq x_p - 1/k$ , a partir cierto  $n_1$  en adelante.

En definitiva, para cada  $k \in \mathbb{N}$ ,

$$\mathcal{P} \left\{ |c_p(n) - x_p| \leq \frac{1}{k}, \quad \text{a partir de un } n \text{ en adelante} \right\} = 1;$$

de forma que  $\lim_{n \rightarrow \infty} c_p(n) = x_p$   $\mathcal{P}$ -casi seguro.

Al igual que ocurre con los momentos muestrales, la distribución en el muestreo de los cuantiles muestrales no puede determinarse, en general, para una distribución teórica arbitraria. En tal caso, tiene interés conocer, al menos, su distribución asintótica cuando el tamaño muestral tiende a infinito. Ello es posible gracias al siguiente lema, de frecuente utilidad.

**Lema 2.1:** Si  $\{T_n\}_{n=1}^{\infty}$  es una sucesión de variables aleatorias tales que

$$\sqrt{n} (T_n - \theta) \xrightarrow{d} N(0, \sigma)$$

y  $g$  es una función derivable de  $\mathbb{R}$  en  $\mathbb{R}$  tal que  $g'(\theta) \neq 0$ , entonces,

$$\sqrt{n} (g(T_n) - g(\theta)) \xrightarrow{d} N(0, |g'(\theta)| \sigma).$$

Demostración: En virtud del teorema del valor medio,

$$g(T_n) - g(\theta) = (T_n - \theta)(g'(\theta) + \varepsilon_n),$$

siendo  $\varepsilon_n \rightarrow 0$  cuando  $T_n \rightarrow \theta$ . Pero, según la hipótesis,  $T_n \xrightarrow{P} \theta$ , luego  $\varepsilon_n \xrightarrow{P} 0$ . Entonces,

$$\sqrt{n}(g(T_n) - g(\theta)) - \sqrt{n}(T_n - \theta)g'(\theta) = \sqrt{n}(T_n - \theta)\varepsilon_n \xrightarrow{P} 0$$

(por ser  $\sqrt{n}(T_n - \theta)$  convergente en distribución); de ello se deduce que  $\sqrt{n}(g(T_n) - g(\theta))$  y  $\sqrt{n}(T_n - \theta)g'(\theta)$  tienen la misma distribución límite, siendo la de este último término  $N(0, |g'(\theta)|\sigma)$ .

**Teorema 2.2:** Si  $c_p(n)$  es el cuantil muestral de orden  $p$  de una muestra aleatoria simple  $(X_1, X_2, \dots, X_n)$  de una distribución teórica  $F$ , que tiene función de densidad continua  $f$  y cuyo cuantil de orden  $p$ ,  $x_p$ , es único, se verifica que

$$\sqrt{n}(c_p(n) - x_p) \xrightarrow{d} N\left(0, \frac{\sqrt{p(1-p)}}{f(x_p)}\right).$$

Demostración: Supongamos, para evitar complicaciones triviales, que  $np \notin \mathbb{N}$  y sea  $m$  el primer entero mayor que  $np$ ; de manera que  $c_p(n)$  coincide con el  $m$ -ésimo estadístico ordenado  $X_{(m)}$  cuya densidad es (véase ejercicio 1.4):

$$\frac{n!}{(m-1)!(n-m)!} F(x)^{m-1} [1 - F(x)]^{n-m} f(x) \quad \text{para } x \in \mathbb{R}.$$

La variable  $Y_n = F(c_p(n))$  tiene entonces densidad

$$\frac{n!}{(m-1)!(n-m)!} y^{m-1} (1-y)^{n-m} \quad \text{para } y \in (0, 1)$$

y, por consiguiente,  $Z_n = \sqrt{n}(Y_n - p)/\sqrt{p(1-p)}$  tiene densidad

$$\frac{n! p^{m-1/2} (1-p)^{n-m+1/2}}{(m-1)!(n-m)! \sqrt{n}} \left(1 + z \sqrt{\frac{1-p}{np}}\right)^{m-1} \left(1 - z \sqrt{\frac{p}{n(1-p)}}\right)^{n-m}$$

Cuando  $n$  tiende a infinito, usando la aproximación de Stirling, el primer factor es equivalente<sup>3</sup> a

$$\begin{aligned} & \frac{p^{m-1/2}(1-p)^{n-m+1/2}\sqrt{2\pi} n^n e^{-n}}{\sqrt{2\pi(m-1)}(m-1)^{m-1}e^{-m+1}\sqrt{2\pi(n-m)}(n-m)^{n-m}e^{-n+m}} \\ = & \frac{1}{\sqrt{2\pi} e} \left( \frac{np}{m-1} \right)^{m-1/2} \left( \frac{n(1-p)}{n-m} \right)^{n-m+1/2} \\ = & \frac{1}{\sqrt{2\pi} e} \left( 1 + \frac{np - m + 1}{m-1} \right)^{m-1/2} \left( 1 + \frac{n(1-p) - n + m}{n-m} \right)^{n-m+1/2} \end{aligned}$$

que es equivalente a

$$\frac{1}{\sqrt{2\pi} e} e^{np-m+1} e^{n(1-p)-n+m} = \frac{1}{\sqrt{2\pi}},$$

puesto que  $m$  y  $n-m$  tienden a infinito, pero  $np - m + 1$  y  $n(1-p) - n + m$  están comprendidos entre 0 y 1 para cualquier  $n$ .

Por otra parte, teniendo en cuenta que  $m-1 \sim np$  y  $(n-m) \sim n(1-p)$ , será

$$\begin{aligned} & \log \left[ \left( 1 + z \sqrt{\frac{1-p}{np}} \right)^{m-1} \left( 1 - z \sqrt{\frac{p}{n(1-p)}} \right)^{n-m} \right] \\ & \sim (m-1) \left( z \sqrt{\frac{1-p}{np}} - \frac{z^2}{2} \frac{1-p}{np} \right) + (n-m) \left( -z \sqrt{\frac{p}{n(1-p)}} - \frac{z^2}{2} \frac{p}{n(1-p)} \right) \\ & \sim z \sqrt{np(1-p)} - \frac{z^2}{2}(1-p) - z \sqrt{np(1-p)} - \frac{z^2}{2}p = -\frac{z^2}{2}. \end{aligned}$$

Resulta pues  $Z_n \xrightarrow{d} N(0, 1)$ , o bien  $\sqrt{n}(Y_n - p) \xrightarrow{d} N(0, \sqrt{p(1-p)})$ .

Aplicando el lema anterior con  $g = F^{-1}$  se obtiene en definitiva que

$$\sqrt{n}(c_p(n) - x_p) \xrightarrow{d} N \left( 0, \frac{\sqrt{p(1-p)}}{f(x_p)} \right).$$

En particular, para la mediana muestral,  $M$ , el resultado anterior proporciona:

$$\sqrt{n}(M - x_{1/2}) \xrightarrow{d} N \left( 0, \frac{1}{2f(x_{1/2})} \right)$$

---

<sup>3</sup> $a_n$  es equivalente a  $b_n$  si  $a_n/b_n \rightarrow 1$ . Se escribe entonces  $a_n \sim b_n$ .

siendo  $x_{1/2}$  la mediana poblacional. El ejercicio 2.9 muestra como utilizar estas conclusiones.

## 2.6. Comportamiento asintótico de los momentos muestrales

En el caso de los momentos de la distribución muestral, el teorema de Glivenko-Cantelli sugiere, nuevamente, su convergencia casi segura hacia los momentos poblacionales. De hecho, la conclusión se sigue ahora directamente de la ley fuerte de los grandes números.

En efecto, si  $\{X_i\}_{i=1}^{\infty}$  es una sucesión de variables aleatorias independientes, con distribución común  $F$ , y  $a_k(n) = \frac{1}{n} \sum_{i=1}^n X_i^k$ , la ley fuerte de los grandes números permite afirmar que

$$\lim_{n \rightarrow \infty} a_k(n) = \alpha_k \quad \mathcal{P}\text{-casi seguro},$$

supuesto que  $F$  tenga momento finito de orden  $k$ ,  $\alpha_k = E[X_i^k]$  y siendo  $\mathcal{P}$  la distribución conjunta de la sucesión  $\{X_i\}_{i=1}^{\infty}$ .

Es inmediato entonces, que si  $g$  es una función continua de  $\mathbb{R}^k$  en  $\mathbb{R}$  será

$$\lim_{n \rightarrow \infty} g(a_1(n), a_2(n), \dots, a_k(n)) = g(\alpha_1, \alpha_2, \dots, \alpha_k) \quad \mathcal{P}\text{-c.s.}$$

En particular, como los momentos centrales se expresan en función de los momentos respecto al origen mediante funciones polinómicas:

$$b_k = a_k - \binom{k}{1} a_{k-1} a_1 + \binom{k}{2} a_{k-2} a_1^2 + \dots + (-1)^k a_1^k,$$

se concluye que

$$\lim_{n \rightarrow \infty} b_k(n) = \mu_k \quad \mathcal{P}\text{-casi seguro},$$

supuesto que existe el momento central poblacional de orden  $k$ ,  $\mu_k$ .

La distribución asintótica de los momentos muestrales respecto al origen también se obtiene de forma casi inmediata a partir del teorema central del límite e, incluso, la versión multidimensional de este resultado permite obtener la distribución asintótica conjunta de varios de ellos:

**Teorema 2.3:** Dada una muestra aleatoria simple  $(X_1, X_2, \dots, X_n)$  de una población con momentos finitos de orden  $2k$ , se verifica

$$\sqrt{n}(a_k(n) - \alpha_k) \xrightarrow{d} N\left(0, \sqrt{\alpha_{2k} - \alpha_k^2}\right).$$

Además

$$[\sqrt{n}(a_1(n) - \alpha_1), \sqrt{n}(a_2(n) - \alpha_2), \dots, \sqrt{n}(a_k(n) - \alpha_k)] \xrightarrow{d} N_k(0, \Sigma),$$

siendo

$$\Sigma_{jj} = \alpha_{2j} - \alpha_j^2, \quad \Sigma_{jl} = \alpha_{j+l} - \alpha_j \alpha_l$$

los términos de la matriz de covarianzas de la distribución normal  $k$ -dimensional límite.

Por supuesto, el segundo resultado engloba al primero, pero este es sin duda el de mayor relevancia. Sobre todo en el caso particular  $k = 1$ , que garantiza que  $\bar{X}$  es aproximadamente  $N(\mu, \sigma/\sqrt{n})$ , cuando  $n$  es grande, cualquiera que sea  $F$ . (Véase el ejercicio 2.10.)

Demostración: Las variables  $k$ -dimensionales  $Y_i = (X_i, X_i^2, \dots, X_i^k)$  son independientes e idénticamente distribuidas, con

$$E[Y_i] = (\alpha_1, \alpha_2, \dots, \alpha_k) = \alpha$$

y matriz de covarianzas  $\Sigma$  (puesto que  $V(X_i^j) = \alpha_{2j} - \alpha_j^2$  y  $\text{Cov}(X_i^j, X_i^l) = \alpha_{j+l} - \alpha_j \alpha_l$ ). Mediante ellas se expresa

$$[\sqrt{n}(a_1(n) - \alpha_1), \sqrt{n}(a_2(n) - \alpha_2), \dots, \sqrt{n}(a_k(n) - \alpha_k)] = \frac{1}{\sqrt{n}} \left( \sum_{i=1}^n Y_i - n\alpha \right).$$

Y, aplicando el teorema central del límite, se obtiene

$$\frac{1}{\sqrt{n}} \left( \sum_{i=1}^n Y_i - n\alpha \right) \xrightarrow{d} N_k(0, \Sigma).$$

La obtención de la distribución asintótica de los momentos muestrales  $b_k$ , respecto a la media  $\bar{X}$ , requiere de una versión multidimensional del lema de la sección anterior, cuya demostración es similar, utilizando, naturalmente, la versión multidimensional del teorema del valor medio.

**Lema 2.2:** Si  $\{T_n\}_{n=1}^{\infty}$  es una sucesión de variables aleatorias  $k$ -dimensionales, tales que

$$\sqrt{n}(T_n - \theta) \xrightarrow{d} N_k(0, \Sigma)$$

y  $g$  es una función de  $\mathbb{R}^k$  en  $\mathbb{R}$ , con derivadas parciales de primer orden, tales que<sup>4</sup>  $\sigma^2(\theta) = \nabla g(\theta) \Sigma \nabla g(\theta)^t \neq 0$ , entonces

$$\sqrt{n}(g(T_n) - g(\theta)) \xrightarrow{d} N(0, \sigma(\theta)).$$

Para aplicar este resultado a la determinación de la distribución asintótica del momento muestral

$$b_k(n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$$

hay que tomar  $T_n = (a_1(n), a_2(n), \dots, a_k(n))$ ,  $\theta = (\alpha_1, \alpha_2, \dots, \alpha_k)$  y

$$g(x_1, x_2, \dots, x_k) = x_k - \binom{k}{1} x_{k-1} x_1 + \binom{k}{2} x_{k-2} x_1^2 + \dots + (-1)^k x_1^k;$$

de manera que será  $g(T_n) = b_k(n)$  y  $g(\theta) = \mu_k$ .

Para facilitar el cálculo, se puede hacer una translación de manera a conseguir que  $\alpha_1 = 0$  (restando  $\alpha_1$  tanto a la variable poblacional como a las variables de la muestra). En tal caso,  $\nabla g(\theta) = (-k\alpha_{k-1}, 0, 0, \dots, 1)$  y combinando el lema previo con el teorema anterior, resulta

$$\sqrt{n}(b_k(n) - \mu_k) \xrightarrow{d} N(0, \sigma(\theta)),$$

donde

$$\begin{aligned} \sigma(\theta) &= (-k\alpha_{k-1}, 0, 0, \dots, 1) \begin{pmatrix} \alpha_2 & & \alpha_{k+1} \\ \vdots & & \vdots \\ \alpha_{k+1} & & \alpha_{2k} - \alpha_k^2 \end{pmatrix} \begin{pmatrix} -k\alpha_{k-1} \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \\ &= k^2 \alpha_{k-1}^2 \alpha_2 - 2k\alpha_{k-1} \alpha_{k+1} + \alpha_{2k} - \alpha_k^2 \\ &= \alpha_{2k} - \alpha_k^2 - 2k\alpha_{k+1} \alpha_{k-1} + k^2 \alpha_{k-1}^2 \alpha_2. \end{aligned}$$

---

<sup>4</sup>  $\nabla g(\theta)$  es el vector cuyos términos son las derivadas parciales de  $g$ , en el punto  $\theta$ .

Para deshacer el cambio que permitió hacer  $\alpha_1 = 0$ , los momentos poblacionales  $\alpha_j$  deben reemplazarse por los momentos centrales  $\mu_j$ , y la conclusión que se obtiene es:

**Teorema 2.4:** Si  $b_k(n)$  es el momento central de orden  $k$  de una muestra aleatoria simple  $(X_1, X_2, \dots, X_n)$  de una población con momentos finitos de orden  $2k$ , se verifica

$$\sqrt{n}(b_k(n) - \mu_k) \xrightarrow{d} N\left(0, \sqrt{\mu_{2k} - \mu_k^2 - 2k \mu_{k+1} \mu_{k-1} + k^2 \mu_{k-1}^2}\right).$$

Cualitativamente, todos los resultados anteriores aseguran que cualquier momento muestral tiene, cuando  $n$  es grande, una distribución muy concentrada alrededor del correspondiente momento poblacional.

En particular, para la varianza muestral  $s^2$ , como  $\mu_1 = E[X - \mu] = 0$ , se obtiene que

$$\sqrt{n}(s^2 - \sigma^2) \xrightarrow{d} N\left(0, \sqrt{\mu_4 - \sigma^4}\right)$$

siendo  $\sigma^2$  la varianza poblacional.

## Ejercicios

### Ejercicio 2.1

Experimentalmente se miden los tiempos de sedimentación de determinadas partículas coloreadas, flotando en un cierto líquido. Los tiempos observados, ya ordenados, fueron: 0'19, 0'78, 0'96, 1'31, 2'78, 3'16, 4'15, 4'67, 4'85, 6'5, 7'35, 8'01, 8'27, 12'06, 31'75, 32'52, 33'91, 36'71 y 72'89.

(a) Construir la función de distribución muestral.

(b) Hallar los tres primeros momentos muestrales respecto al origen y respecto a la media.

(c) Determinar los cuartiles de la distribución muestral.

### Ejercicio 2.2

Se dispone de una muestra aleatoria simple de tamaño 40 de una distribución exponencial de media 3, ¿qué probabilidad hay de que los valores de la función de distribución muestral y teórica, en  $x = 1$ , difieran en menos de 0'01? Dar una respuesta aproximada del resultado. ¿Cuál debería ser el tamaño muestral para que dicha probabilidad fuese aproximadamente 0'98?.

**Ejercicio 2.3**

Se dispone de una muestra aleatoria simple de tamaño 50 de una distribución de Poisson de media 2. ¿Qué probabilidad hay de que los valores de la función de distribución muestral y teórica, en el punto  $x = 2$ , difieran en menos de 0'03? ¿Qué tamaño muestral hay que tomar para que dicha probabilidad sea aproximadamente 0'99?

**Ejercicio 2.4**

Probar que, si existen los momentos poblacionales de tercer orden, es

$$E[\bar{X}^3] = \frac{\alpha_3 + 3(n-1)\alpha_2\alpha_1 + (n-1)(n-2)\alpha_1^3}{n^2}$$

y

$$E[(\bar{X} - \mu)^3] = \frac{\mu_3}{n^2}.$$

**Ejercicio 2.5**

Probar que, si existen los momentos poblacionales de cuarto orden, es

$$\begin{aligned} E[\bar{X}^4] &= \frac{1}{n^3} \left\{ \alpha_4 + 4(n-1)\alpha_3\alpha_1 + 6(n-1)(n-2)\alpha_2\alpha_1^2 \right. \\ &\quad \left. + 3(n-1)\alpha_2^2 + (n-1)(n-2)(n-3)\alpha_1^4 \right\} \end{aligned}$$

y

$$E[(\bar{X} - \mu)^4] = \frac{\mu_4}{n^3} + 3 \frac{n-1}{n^3} \sigma^4.$$

**Ejercicio 2.6**

Probar que, si existen los momentos poblacionales de cuarto orden, es

$$E[s^4] = \frac{(n^2 - 2n + 1)\mu_4 + (n-1)(n^2 - 2n + 3)\sigma^4}{n^3}$$

y

$$V(s^2) = \frac{(n-1)^2}{n^3} \mu_4 - \frac{n^2 - 4n + 3}{n^3} \sigma^4.$$

**Ejercicio 2.7**

Probar que, si existen los momentos poblacionales de tercer orden, el momento muestral  $b_3 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^3$  verifica

$$E[b_3] = \frac{(n-1)(n-2)}{n^2} \mu_3.$$

**Ejercicio 2.8**

Probar que, si existen los momentos poblacionales de cuarto orden, el momento muestral  $b_4 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^4$  verifica

$$\mathbb{E}[b_4] = \frac{(n-1)(n^2 - 3n + 3)}{n^3} \mu_4 + \frac{3(n-1)(2n-3)}{n^3} \sigma^4.$$

**Ejercicio 2.9**

Dada una muestra aleatoria simple de tamaño 100 de una distribución  $N(\mu, 1'5)$  determinar la probabilidad de que la mediana muestral difiera de  $\mu$  en menos de 0'1.

¿De qué tamaño habría que elegir la muestra para poder afirmar que con probabilidad 0'9 la mediana muestral difiere de  $\mu$  en menos de 0'01?

**Ejercicio 2.10**

Dada una muestra aleatoria simple de tamaño 500 de una distribución  $\gamma(p, \sqrt{p})$  determinar la probabilidad de que la media muestral difiera de la media poblacional en menos de 0'1.

¿De qué tamaño habría que elegir la muestra para poder afirmar que con probabilidad 0'99 la media muestral difiere de la poblacional en menos de 0'01?

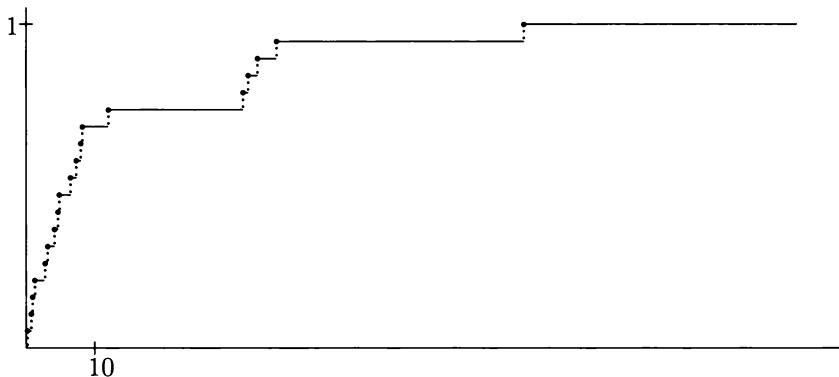
## Solución de los ejercicios

### Ejercicio 2.1

(a) La función de distribución muestral será una función escalonada, con saltos de  $1/19$  en cada valor observado,

$$F_n^*(x) = \begin{cases} 0 & \text{si } x < 0'19 \\ 1/19 & \text{si } 0'19 \leq x < 0'78 \\ \dots & \dots \dots \dots \\ 18/19 & \text{si } 36'71 \leq x < 72'89 \\ 1 & \text{si } x \geq 72'89, \end{cases}$$

siendo su representación gráfica



(b) Los momentos muestrales respecto al origen son:

$$a_1 = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{272'82}{19} = 14'36,$$

$$a_2 = \frac{1}{n} \sum_{i=1}^n x_i^2 = \frac{10.333'896}{19} = 543'89,$$

$$a_3 = \frac{1}{n} \sum_{i=1}^n x_i^3 = \frac{545.971'92}{19} = 28.735'36,$$

y los momentos muestrales respecto a la media, además de  $b_1 = 0$ ,

$$b_2 = s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = a_2 - a_1^2 = 337'71,$$

$$b_3 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3 = a_3 - 3a_2 a_1 + 2a_1^3 = 11.227'369.$$

(c) Por último, respecto a los cuartiles, para el primero de ellos  $np = 19/4 = 4'75$  y  $[np] = 4$ ; de forma que

$$c_{1/4} = x_{([np]+1)} = x_{(5)} = 2'78.$$

Para el segundo cuartil, es decir la mediana,  $np = 19/2 = 9'5$  y, por tanto,  $[np] = 9$ ; luego

$$c_{1/2} = x_{(10)} = 6'5.$$

Por último, para el tercer cuartil,  $p = 3/4$ , es  $[np] = [14'25] = 14$ ; con lo cual

$$c_{3/4} = x_{(15)} = 31'75.$$

### Ejercicio 2.2

Se trata de hallar, con  $n = 40$ ,

$$\mathcal{P}\{|F_n^*(1) - F(1)| < 0'01\}$$

siendo  $F$  la función de distribución de la exponencial de parámetro  $1/3$ .

Por un lado,  $F(1) = 1 - e^{-1/3} = 0'28347$  y, por otra parte,  $Y = nF_n^*(1)$  tiene distribución binomial  $B(40, 0'28347)$ . Luego

$$\mathcal{P}\{|F_n^*(1) - F(1)| < 0'01\} = P\{|Y - 11'3388| < 0'4\} = P\{Y = 11\}$$

$$= \binom{40}{11} 0'28347^{11} 0'71653^{29} = 0'139.$$

Como  $Z = (Y - 11'3388)/2'85$  es aproximadamente  $N(0, 1)$ , se puede obtener una aproximación del resultado anterior en la forma

$$P\{|Y - 11'3388| < 0'4\} \simeq P\{|Z| < \frac{0'4}{2'85}\} = P\{|Z| < 0'14\} = 0'1114.$$

La aproximación no es muy buena, pero mejorará al aumentar el tamaño muestral. Para hallar el valor de  $n$  tal que

$$P\{|F_n^*(1) - F(1)| < 0'01\} \simeq 0'98,$$

se sabe que  $F_n^*(1)$  es aproximadamente  $N(F(1), 0'45/\sqrt{n})$ . Si  $Z$  es  $N(0, 1)$ , es  $P\{|Z| < 2'33\} = 0'98$ . Luego debe ser

$$\sqrt{n} \frac{0'01}{0'45} \simeq 2'33;$$

es decir,  $n \simeq 10.933'5$ . Con muestras de tamaño 10.934, sólo el 2% ellas darán un valor de  $F_n^*(1)$  que se diferencie de  $F(1)$  en más de 0'01.

### Ejercicio 2.3

Se trata de calcular

$$\mathcal{P} \{ |F_n^*(2) - F(2)| < 0'03 \}$$

siendo  $F$  la función de distribución de una distribución de Poisson de parámetro 2 y  $n = 50$ . Por un lado es

$$F(2) = e^{-2} + e^{-2} \frac{2}{1!} + e^{-2} \frac{2^2}{2!} = 5e^{-2} = 0'67667.$$

Por otra parte  $Y = nF_n^*(2)$  tiene distribución binomial  $B(50, 0'67667)$ ; luego

$$\mathcal{P} \{ |F_n^*(2) - F(2)| < 0'03 \} = P \{ |Y - 33'8335| < 1'5 \} = P\{33 \leq Y \leq 35\}$$

$$= \sum_{i=33}^{35} \binom{50}{i} 0'67667^i 0'32333^{50-i} = 0'349175.$$

Como  $Z = (Y - 33'8335)/3'3075$  es aproximadamente  $N(0, 1)$ , una aproximación de esta probabilidad es

$$P\{|Y - 33'8335| < 1'5\} \simeq P\{|Z| < 0'45\} = 0'3472.$$

Si se quiere que  $P\{|F_n^*(2) - F(2)| < 0'03\} \simeq 0'99$ , como  $F_n^*(2)$  es aproximadamente  $N(F(2), 0'4677/\sqrt{n})$  y  $P\{|Z| < 2'58\} = 0'99$ , tendrá que ser

$$\sqrt{n} \frac{0'03}{0'4677} \simeq 2'58$$

de donde  $n \simeq 1617'8$ . Con muestras de tamaño 1618, sólo el 1% darán un valor de  $F_n^*(2)$  que se diferencie de  $F(2)$  en más de 0'03.

### Ejercicio 2.4

$$\left( \sum_{i=1}^n X_i \right)^3 = \sum_{i=1}^n X_i^3 + 3 \sum_{i \neq j=1}^n X_i^2 X_j + \sum_{i \neq j \neq k=1}^n X_i X_j X_k,$$

como las  $X_i$  son independientes e idénticamente distribuidas, tomando esperanzas se tiene

$$\begin{aligned} \mathbb{E}[\bar{X}^3] &= \frac{1}{n^3} \mathbb{E}\left[\left(\sum_{i=1}^n X_i\right)^3\right] = \\ &= \frac{1}{n^3} \left\{ \sum_{i=1}^n \mathbb{E}[X_i^3] + 3 \sum_{i \neq j=1}^n \mathbb{E}[X_i^2]\mathbb{E}[X_j] + \sum_{i \neq j \neq k=1}^n \mathbb{E}[X_i]\mathbb{E}[X_j]\mathbb{E}[X_k] \right\} \\ &= \frac{1}{n^3} \{ n\alpha_3 + 3n(n-1)\alpha_2\alpha_1 + n(n-1)(n-2)\alpha_1^3 \} \\ &= \frac{\alpha_3 + 3(n-1)\alpha_2\alpha_1 + (n-1)(n-2)\alpha_1^3}{n^2}. \end{aligned}$$

Por otro lado, llamando  $Y_i = X_i - \mu$ , es  $\bar{X} - \mu = \bar{Y}$ , luego

$$\mathbb{E}[(\bar{X} - \mu)^3] = \mathbb{E}[\bar{Y}^3]$$

y, aplicando el resultado anterior con  $\alpha_1 = \mathbb{E}[Y_i] = 0$ , resulta

$$\mathbb{E}[(\bar{X} - \mu)^3] = \frac{\mathbb{E}[Y_i^3]}{n^2} = \frac{\mu_3}{n^2}.$$

### Ejercicio 2.5

De manera análoga al ejercicio anterior, es

$$\begin{aligned} \left(\sum_{i=1}^n X_i\right)^4 &= \sum_{i=1}^n X_i^4 + 4 \sum_{i \neq j=1}^n X_i^3 X_j + 3 \sum_{i \neq j=1}^n X_i^2 X_j^2 + 6 \sum_{i \neq j \neq k=1}^n X_i^2 X_j X_k \\ &\quad + \sum_{i \neq j \neq k \neq l=1}^n X_i X_j X_k X_l, \end{aligned}$$

luego

$$\begin{aligned} \mathbb{E}[\bar{X}^4] &= \frac{1}{n^4} \mathbb{E}\left[\left(\sum_{i=1}^n X_i\right)^4\right] = \\ &= \frac{1}{n^3} \{ \alpha_4 + 4(n-1)\alpha_3\alpha_1 + 3(n-1)\alpha_2^2 + 6(n-1)(n-2)\alpha_2\alpha_1^2 \\ &\quad + (n-1)(n-2)(n-3)\alpha_1^4 \}. \end{aligned}$$

Actuando de nuevo como en el ejercicio 2.4,

$$\mathbb{E}[(\bar{X} - \mu)^4] = \mathbb{E}[\bar{Y}^4];$$

de forma que el resultado anterior, con  $\alpha_1 = 0$ , se reduce a

$$\mathbb{E}[(\bar{X} - \mu)^4] = \frac{\mathbb{E}[Y_i^4] + 3(n-1)\mathbb{E}[Y_i^2]^2}{n^3} = \frac{\mu_4 + 3(n-1)\sigma^4}{n^3}.$$

### Ejercicio 2.6

Llamando  $Y_i = X_i - \mu$ , se puede expresar

$$s^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2 = \frac{1}{n} \sum_{i=1}^n Y_i^2 - \bar{Y}^2.$$

Por tanto será

$$\mathbb{E}[s^4] = \mathbb{E} \left[ \frac{1}{n^2} \left( \sum_{i=1}^n Y_i^2 \right)^2 + \bar{Y}^4 - \frac{2}{n} \bar{Y}^2 \sum_{i=1}^n Y_i^2 \right].$$

Pero

$$\mathbb{E} \left[ \left( \sum_{i=1}^n Y_i^2 \right)^2 \right] = \mathbb{E} \left[ \sum_{i=1}^n Y_i^4 + \sum_{i \neq j=1}^n Y_i^2 Y_j^2 \right] = n\mu_4 + n(n-1)\sigma^4.$$

Utilizando el resultado del ejercicio 2.5,

$$\mathbb{E}[\bar{Y}^4] = \frac{\mu_4 + 3(n-1)\sigma^4}{n^3}.$$

Y, por último,

$$\begin{aligned} \mathbb{E} \left[ \bar{Y}^2 \sum_{i=1}^n Y_i^2 \right] &= \frac{1}{n^2} \mathbb{E} \left[ \left( \sum_{i=1}^n Y_i^2 + \sum_{i \neq j=1}^n Y_i Y_j \right) \sum_{i=1}^n Y_i^2 \right] \\ &= \frac{1}{n^2} \mathbb{E} \left[ \sum_{i=1}^n Y_i^4 + \sum_{i \neq j=1}^n Y_i^2 Y_j^2 + 2 \sum_{i \neq j=1}^n Y_i^3 Y_j + \sum_{i \neq j \neq k=1}^n Y_i Y_j Y_k^2 \right] \\ &= \frac{n\mu_4 + n(n-1)\sigma^4}{n^2}. \end{aligned}$$

Con lo cual

$$E[s^4] = \frac{(n-1)^2\mu_4 + (n-1)(n^2 - 2n + 3)\sigma^4}{n^3}.$$

Restando  $E[s^2]^2 = (n-1)^2\sigma^4/n^2$ , se obtiene

$$V(s^2) = \frac{(n-1)^2}{n^3} \mu_4 - \frac{(n-1)(n-3)}{n^3} \sigma^4.$$

### Ejercicio 2.7

Pasando a las variables  $Y_i = X_i - \mu$ , como en los ejercicios anteriores, es

$$\begin{aligned} E[b_3] &= E\left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^3\right] = E\left[\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^3\right] \\ &= E\left[\frac{1}{n} \sum_{i=1}^n (Y_i^3 - 3Y_i^2\bar{Y} + 3Y_i\bar{Y}^2 - \bar{Y}^3)\right] \\ &= E\left[\frac{1}{n} \sum_{i=1}^n Y_i^3 - 3\bar{Y} \frac{1}{n} \sum_{i=1}^n Y_i^2 + 2\bar{Y}^3\right] \\ &= \mu_3 - \frac{3}{n^2} E\left[\sum_{i=1}^n Y_i \sum_{i=1}^n Y_i^2\right] + 2E[\bar{Y}^3] \\ &= \mu_3 - \frac{3}{n^2} n\mu_3 + 2\frac{\mu_3}{n^2} \\ &= \frac{(n-1)(n-2)}{n^2} \mu_3, \end{aligned}$$

utilizando el resultado del ejercicio 2.4 y

$$E\left[\sum_{i=1}^n Y_i \sum_{i=1}^n Y_i^2\right] = E\left[\sum_{i=1}^n Y_i^3 + \sum_{i \neq j=1}^n Y_i Y_j^2\right] = n\mu_3.$$

### Ejercicio 2.8

Nuevamente, en términos de las variables  $Y_i = X_i - \mu$ ,

$$\begin{aligned} E[b_4] &= E\left[\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^4\right] \\ &= E\left[\frac{1}{n} \sum_{i=1}^n (Y_i^4 - 4\bar{Y}Y_i^3 + 6\bar{Y}^2Y_i^2 - 4\bar{Y}^3Y_i + \bar{Y}^4)\right] \end{aligned}$$

$$\begin{aligned}
&= \mu_4 - \frac{4}{n^2} E \left[ \sum_{i=1}^n Y_i \sum_{i=1}^n Y_i^3 \right] + \frac{6}{n^3} E \left[ \left( \sum_{i=1}^n Y_i \right)^2 \sum_{i=1}^n Y_i^2 \right] - 3E[\bar{Y}^4] \\
&= \mu_4 - \frac{4}{n} \mu_4 + \frac{6}{n^2} (\mu_4 + (n-1)\sigma^4) - \frac{3}{n^3} (\mu_4 + 3(n-1)\sigma^4) \\
&= \frac{(n-1)(n^2-3n+3)}{n^3} \mu_4 + \frac{3(n-1)(2n-3)}{n^3} \sigma^4,
\end{aligned}$$

después de sustituir el valor de  $E[\bar{Y}^4]$  por su valor, obtenido en el ejercicio 2.5, y utilizando

$$\begin{aligned}
E \left[ \sum_{i=1}^n Y_i \sum_{i=1}^n Y_i^3 \right] &= E \left[ \sum_{i=1}^n Y_i^4 + \sum_{i \neq j=1}^n Y_i Y_j^3 \right] = n\mu_4, \\
E \left[ \left( \sum_{i=1}^n Y_i \right)^2 \sum_{i=1}^n Y_i^2 \right] &= E \left[ \left( \sum_{i=1}^n Y_i^2 + \sum_{i \neq j=1}^n Y_i Y_j \right) \sum_{i=1}^n Y_i^2 \right] \\
&= E \left[ \sum_{i=1}^n Y_i^4 + \sum_{i \neq j=1}^n Y_i^2 Y_j^2 + 2 \sum_{i \neq j=1}^n Y_i^3 Y_j + \sum_{i \neq j \neq k=1}^n Y_i Y_j Y_k^2 \right] \\
&= n\mu_4 + n(n-1)\sigma^4.
\end{aligned}$$

### Ejercicio 2.9

La muestra es de un tamaño suficientemente grande, para poder utilizar la aproximación normal a la distribución de la mediana muestral. Como en este caso la distribución teórica es normal, y, por tanto, simétrica, será  $x_{1/2} = \mu$  y  $f(x_{1/2}) = f(\mu) = 1/(1'5\sqrt{2\pi})$ .

Por consiguiente, si  $Z$  es  $N(0, 1)$ , será

$$P\{|M - \mu| < 0'1\} \simeq P\{|Z| < 0'532\} = 0'4052.$$

Para conseguir que sea  $P\{|M - \mu| < 0'01\} = 0'9$ , admitiendo que el tamaño muestral va a ser grande, después de tipificar, la aproximación normal transforma esta condición en

$$P\{|Z| < 0'00532\sqrt{n}\} = 0'9.$$

Pero como  $P\{|Z| < 1'645\} = 0'9$ , debe ser

$$0'00532\sqrt{n} = 1'645;$$

es decir,  $n = 95\,611'1$ . Tomando un tamaño muestral de 95 612 unidades se obtendrá, por tanto, la precisión deseada, con probabilidad 0'9.

### Ejercicio 2.10

Como  $X$  es  $\gamma(p, \sqrt{p})$ , la media poblacional es  $\mu = \sqrt{p}$  y la desviación típica poblacional  $\sigma = 1$ . Aunque se sabe que la distribución exacta en el muestreo de  $\bar{X}$  es  $\gamma(500p, 500\sqrt{p})$  (cf. pág. 41), sería precisa así una tabla de la distribución gamma para determinar  $\mathcal{P}\{|\bar{X} - \sqrt{p}| < 0'1\}$ .

Utilizando, en cambio, la aproximación normal de la sección 2.6,  $\bar{X}$  será aproximadamente  $N(\sqrt{p}, 0'04472)$ . Por tanto, si  $Z$  tiene distribución  $N(0, 1)$ ,

$$\mathcal{P}\{|\bar{X} - \sqrt{p}| < 0'1\} \simeq P\{|Z| < 2'236\} = 0'97468.$$

El tamaño muestral necesario para que se cumpla el segundo requisito es, por supuesto, mayor que 500. Según la aproximación normal,  $\bar{X}$  es aproximadamente  $N(\sqrt{p}, 1/\sqrt{n})$  y ha de ser

$$P\{|\bar{X} - \sqrt{p}| < 0'01\} = 0'99;$$

es decir, aproximadamente,

$$P\{|Z| < 0'01\sqrt{n}\} = 0'99.$$

Como  $P\{|Z| < 2'576\} = 0'99$ , tiene que ser  $0'01\sqrt{n} = 2'576$ ; es decir,  $n = 66.357'7$ . Por tanto, tomando un tamaño muestral igual o superior a  $n = 66\,358$ , se obtendrá la precisión requerida en el 99 % de los casos.

## Capítulo 3

# Distribuciones en el muestreo de poblaciones normales

### 3.1. Introducción

La distribución normal juega un papel central en Estadística debido a la gran frecuencia con que las características aleatorias, cuyo estudio se aborda, tienen distribución normal o aproximadamente normal. Incluso el propio nombre indica que la “normalidad” ha sido comprobada para multitud de variables de tipo biológico, agrícola, sociológico, económico, etc..., además de todas aquellas situaciones en las que la variabilidad se produce por efecto de errores o imprecisiones en las medidas de las cantidades en estudio.

Como ya se ha apuntado, las razones por las cuales la Naturaleza parece conocer la distribución normal y amoldarse a ella con frecuencia, tienen sin duda que ver con el teorema central del límite, el cual pone de manifiesto que, siempre que un determinado efecto sea producido por un gran número de causas independientes, cuyas repercusiones individuales son despreciables (y de varianza finita), pero que se acumulan para dar lugar al resultado final, la distribución de éste será aproximadamente normal.

Por supuesto hay que evitar generalizaciones excesivas que lleven a emplear la distribución normal en cualquier circunstancia, sin ninguna comprobación estadística previa. Lo cual a veces se hace por motivos interesados: el caso normal puede ser estudiado con más detalle, admite soluciones analíticas más completas y en definitiva conduce a mejores resultados estadísticos.

Algunos de los motivos iniciales a las que se debe esta situación quedarán patentes en las próximas secciones, cuya finalidad es obtener la distribución en el muestreo de los principales estadísticos de utilidad, correspondientes a

una distribución poblacional de tipo normal.

En una primera aproximación, los problemas estadísticos relativos a una población con distribución teórica  $N(\mu, \sigma)$ , sólo pueden provenir del desconocimiento de  $\mu$ , de  $\sigma$  o de ambos parámetros simultáneamente. Parece natural entonces que en su solución intervengan la media muestral,  $\bar{x}$ , y la varianza muestral  $s^2$  y que el principal interés se centre en conocer la distribución en el muestreo de dichos estadísticos, o de alguna función de ambos. Tal será el objetivo de las secciones 3.2 y 3.3.

La problemática se enriquece por la posibilidad de tratar de efectuar comparaciones entre dos poblaciones independientes con distribuciones  $N(\mu_1, \sigma_1)$  y  $N(\mu_2, \sigma_2)$ . Habrá que atender también a la determinación de la distribución en el muestreo de los estadísticos necesarios para ello; de forma que las secciones 3.4, 3.5 y 3.6 tendrán esta finalidad.

### 3.2. Distribución de la media y varianza muestrales. Distribución $\chi^2$ de Pearson

En vista del objetivo planteado, consideraremos en primer lugar una población descrita por una variable aleatoria unidimensional  $X$ , con distribución teórica  $N(\mu, \sigma)$ , de la cual se obtiene una muestra aleatoria simple de tamaño  $n : (X_1, X_2, \dots, X_n)$ .

La distribución en el muestreo de la media muestral,  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ , se obtiene de forma inmediata: Como combinación lineal de variables aleatorias independientes con distribución normal,  $\bar{X}$  tiene distribución normal; además sus parámetros son

$$\begin{aligned}\mathbb{E}[\bar{X}] &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] = \mathbb{E}[X] = \mu, \\ V(\bar{X}) &= \frac{1}{n^2} \sum_{i=1}^n V(X_i) = V[X] = \frac{\sigma^2}{n}.\end{aligned}$$

Luego, la distribución en el muestreo de  $\bar{X}$  es  $N(\mu, \sigma/\sqrt{n})$ .

La obtención de la distribución en el muestreo de la varianza muestral,  $s^2$ , exige previamente estudiar la distribución de la suma de cuadrados:  $\sum_{i=1}^n X_i^2$ .

Supongamos concretamente que  $X$  es  $N(0, 1)$ , en cuyo caso  $X^2$  tiene por

función de distribución:

$$\begin{aligned} P\{X^2 \leq y\} &= P\{-\sqrt{y} \leq X \leq \sqrt{y}\} = 2 \int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\ &= \int_0^y \frac{1}{\sqrt{2\pi}} e^{-u/2} u^{-1/2} du \quad \text{para } y > 0; \end{aligned}$$

es decir, una distribución gamma  $\gamma(1/2, 1/2)$ , de densidad  $\frac{1}{\sqrt{2\pi}} y^{-1/2} e^{-y/2}$ , para  $y > 0$ .

Por consiguiente, si  $X_1, X_2, \dots, X_n$  son variables aleatorias independientes con distribución  $N(0, 1)$ ,  $\sum_{i=1}^n X_i^2$  tiene distribución gamma  $\gamma(n/2, 1/2)$ .

Aún cuando se trate de un caso particular de una distribución conocida, debido a su utilidad posterior, esta distribución exige un estudio detenido y ha merecido un nombre propio:

*Si  $X_1, X_2, \dots, X_n$  son variables aleatorias independientes y con distribución  $N(0, 1)$ , la distribución de la variable*

$$\sum_{i=1}^n X_i^2$$

*se denomina distribución  $\chi^2$  (ji-cuadrado) de Pearson, con  $n$  grados de libertad. En realidad, coincide con la distribución gamma  $\gamma(n/2, 1/2)$  de función de densidad*

$$\frac{1}{2^{n/2} \Gamma(\frac{n}{2})} y^{\frac{n}{2}-1} e^{-y/2} \quad \text{para } y > 0.$$

Como se ve, dicha distribución depende de un único parámetro,  $n$ , cuya denominación de “grados de libertad” hace referencia al número de sumandos que aportan su variabilidad a la suma. En función de  $n$  se expresan de forma inmediata la media y la varianza de la distribución  $\chi_n^2$ :

$$E\left[\sum_{i=1}^n X_i^2\right] = \sum_{i=1}^n E[X_i^2] = n, \quad V\left(\sum_{i=1}^n X_i^2\right) = \sum_{i=1}^n V(X_i^2) = 2n$$

(puesto que  $E[X_i^4] = 3$ ). Valores que pueden obtenerse alternativamente a partir de los correspondientes momentos de la distribución gamma.

Conviene recordar otra propiedad que la distribución  $\chi^2$  hereda de la distribución gamma: Si  $Y_1$  e  $Y_2$  son variables aleatorias independientes con distribución  $\chi_n^2$  y  $\chi_m^2$  respectivamente, entonces  $Y_1 + Y_2$  tiene distribución  $\chi_{n+m}^2$ . Aunque, en este caso, la comprobación es inmediata a partir de la definición; si

$$Y_1 = \sum_{i=1}^n X_i^2 \quad \text{e} \quad Y_2 = \sum_{i=n+1}^{n+m} X_i^2,$$

siendo todas las  $X_i$  independientes y  $N(0, 1)$ , entonces  $Y_1 + Y_2 = \sum_{i=1}^{n+m} X_i^2$  tiene distribución  $\chi_{n+m}^2$ .

La representación gráfica de la función de densidad de la distribución  $\chi_n^2$  es de la forma:

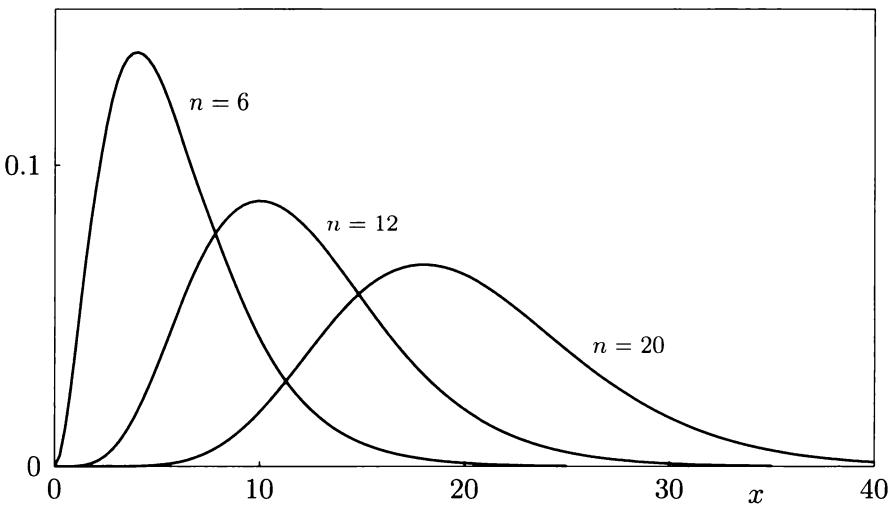


Figura 3.1: Densidades  $\chi^2$

que presenta un máximo en el punto de abscisa  $n - 2$ , supuesto que  $n > 2$ . (Para  $n = 1$  ó  $2$ , la densidad es monótona decreciente).

La función de distribución

$$F(x) = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} \int_0^x y^{\frac{n}{2}-1} e^{-y/2} dy$$

no admite una expresión explícita; de manera que sus valores, obtenidos por integración numérica, están tabulados en la tabla 4.

La disposición de la tabla es peculiar; mientras que el margen izquierdo contiene el parámetro  $n$ , en lugar de situar en el margen superior la variable

independiente  $x$  y, en el cuerpo de la tabla, el valor de  $F(x)$  (o equivalentemente de  $1 - F(x)$ ), el margen superior contiene los valores más útiles de  $1 - F(x)$  y la tabla proporciona la abscisa para la cual se obtiene dicho valor de  $1 - F(x)$ . Así por ejemplo, si  $Y_7, Y_{11}, Y_{23}$  tienen distribuciones  $\chi^2_7, \chi^2_{11}$  y  $\chi^2_{23}$  respectivamente, la tabla da:

$$P\{Y_7 > 2'167\} = 0'95,$$

$$P\{Y_{11} \leq 5'578\} = 1 - 0'9 = 0'1,$$

$$P\{Y_{23} \leq 38'08\} = 1 - 0'025 = 0'975.$$

Tal disposición de la tabla obedece a que su utilización más frecuente suele ser así: dado un valor fijo de  $p \in (0, 1)$ , hallar la abscisa  $x$  tal que

$$\int_x^\infty \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} y^{\frac{n}{2}-1} e^{-y/2} dy = p$$

(véase la figura de la tabla 4).

De todas formas nada impide su utilización al revés; buscar  $x$  en la fila correspondiente a la distribución  $\chi^2$  considerada y leer en el margen superior el valor de  $1 - F(x)$ . Por ejemplo, si  $Y_{13}$  e  $Y_{19}$  tienen distribuciones  $\chi^2_{13}$  y  $\chi^2_{19}$  respectivamente:

$$P\{Y_{13} > 21\} = 0'0767, \quad P\{Y_{19} \leq 25\} = 0'82,$$

mediante una sencilla interpolación lineal (que lógicamente dará peores aproximaciones cuanto más distantes sean los puntos utilizados).

La tabla contiene los datos de la distribución  $\chi^2$  con menos de 31 grados de libertad. ¿Qué sucede si  $n > 30$ ? Simplemente que la distribución  $\chi^2$  se puede aproximar mediante una distribución normal. Por ejemplo, según el teorema central del límite

$$\frac{\sum_{i=1}^n X_i^2 - n}{\sqrt{2n}} \xrightarrow{d} N(0, 1);$$

de forma que, para  $n$  grande,  $\chi_n^2$  se aproxima a una  $N(n, \sqrt{2n})$ .

Sin embargo, existe una aproximación mejor. Si  $Y$  es una variable con distribución  $\chi_n^2$ ,

$$\begin{aligned} P\left\{\sqrt{2Y} - \sqrt{2n-1} \leq x\right\} &= P\left\{Y \leq \frac{x^2}{2} + x\sqrt{2n-1} + n - \frac{1}{2}\right\} \\ &= P\left\{\frac{Y-n}{\sqrt{2n}} \leq \frac{x^2-1}{2\sqrt{2n}} + x\sqrt{1-\frac{1}{2n}}\right\} = F_n(x_n) \end{aligned}$$

donde  $x_n = x\sqrt{1 - 1/(2n)} + (x^2 - 1)/(2\sqrt{2n})$  y  $F_n$  representa la función de distribución de  $(Y - n)/\sqrt{2n}$ .

Puesto que  $x_n \rightarrow x$  y, como acabamos de ver,  $F_n$  converge en distribución hacia la función de distribución continua<sup>1</sup>  $\phi$  de una  $N(0, 1)$ , resulta que, cuando  $n \rightarrow \infty$ ,

$$P\{\sqrt{2Y} - \sqrt{2n-1} \leq x\} \longrightarrow \phi(x);$$

o, dicho de otra manera,

$$\sqrt{2Y} - \sqrt{2n-1} \xrightarrow{d} N(0, 1).$$

Según esto, si  $z_p$  es un valor tal que  $\phi(z_p) = 1 - p$ ,

$$P\left\{Y \leq (z_p + \sqrt{2n-1})^2/2\right\} = P\{\sqrt{2Y} - \sqrt{2n-1} \leq z_p\} \simeq 1 - p.$$

Compárense los valores  $(z_p + \sqrt{2n-1})^2/2$  obtenidos a partir de la tabla de la  $N(0, 1)$  con los valores que proporciona la última fila de la tabla de la distribución  $\chi^2$ , para ver que la aproximación es buena incluso para 30 grados de libertad.

Una última observación sobre la distribución  $\chi^2$  es que, según la ley fuerte de los grandes números, cuando  $n \rightarrow \infty$ ,

$$\frac{Y}{n} \xrightarrow{c.s.} 1.$$

La razón es que  $Y$  tiene la misma distribución que  $\sum_{i=1}^n X_i^2$ , siendo  $X_i^2$  ( $i = 1, \dots, n$ ) independientes e igualmente distribuidas con  $E[X_i^2] = 1$ .

Esta afirmación es comparable con la aproximación normal dada anteriormente, ya que  $(Y - n)/\sqrt{2n} \xrightarrow{d} N(0, 1)$  puede expresarse en la forma

$$\sqrt{\frac{n}{2}} \left( \frac{Y}{n} - 1 \right) \xrightarrow{d} N(0, 1);$$

de forma que, para  $n$  grande,  $Y/n$  es aproximadamente  $N(1, \sqrt{2/n})$ , distribución cada vez más concentrada alrededor de 1 a medida que  $n$  crece.

Para retomar el hilo del problema estadístico que ha conducido a introducir la distribución  $\chi^2$ , se puede enunciar la conclusión obtenida en la forma:

---

<sup>1</sup> Recuérdese que si  $F_n \xrightarrow{d} F$  y  $F$  es continua,  $F_n(x) \rightarrow F(x)$  uniformemente en  $x$ .

Si  $(X_1, X_2, \dots, X_n)$  es una muestra aleatoria simple de una población con distribución teórica  $N(0, \sigma)$ , el momento muestral  $a_2 = \frac{1}{n} \sum_{i=1}^n X_i^2$  verifica:

$$\frac{n a_2}{\sigma^2} \quad \text{tiene distribución } \chi_n^2;$$

puesto que  $na_2/\sigma^2 = \sum_{i=1}^n (X_i/\sigma)^2$  es suma de cuadrados de  $n$  variables independientes y  $N(0, 1)$ .

Sin embargo, como este resultado no se generaliza cómodamente al caso  $\mu \neq 0$ , el interés debe centrarse más bien en la varianza muestral

$$s^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = a_2 - \bar{X}^2,$$

que presenta la ventaja de no depender de  $\mu$  (por ser invariante por translaciones), aunque también la dificultad que supone la sustracción del cuadrado de la variable normal  $\bar{X}$ , no independiente de  $a_2$ . Habrá que examinar entonces el problema de forma más general:

Supongamos que  $X = (X_1, X_2, \dots, X_n)$  es un vector de variables aleatorias independientes y  $N(0, \sigma)$ . Consideremos  $p$  combinaciones lineales<sup>2</sup>

$$Z_i = c_i X^t, \quad i = 1, 2, \dots, p$$

definidas por vectores  $c_i \in \mathbb{R}^n$  tales que  $c_i c_j^t = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$ . Y tratemos de determinar la distribución de la variable aleatoria unidimensional

$$Y = \sum_{i=1}^n X_i^2 - \sum_{i=1}^p Z_i^2 = X \cdot X^t - \sum_{i=1}^p Z_i^2.$$

Para ello, en primer lugar, seleccionemos  $n - p$  vectores,  $c_{p+1}, \dots, c_n$ , que, junto con  $c_1, c_2, \dots, c_p$  constituyan una base ortonormal de  $\mathbb{R}^n$ ; con lo cual la matriz  $n \times n$ :

$$C = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} \quad \text{verificará} \quad CC^t = I.$$

---

<sup>2</sup>  $X^t$  representa  $X$  traspuesteo.

Entonces,  $Z = XC$  será un vector  $(1 \times n)$  de variables aleatorias normales, de media 0 y matriz de covarianzas

$$\mathbb{E}[Z^t Z] = \mathbb{E}[C^t X^t XC] = C^t \mathbb{E}[X^t X] C = C^t \sigma^2 I C = \sigma^2 I.$$

Es decir que  $Z_1, Z_2, \dots, Z_n$  son variables independientes y con distribución  $N(0, \sigma)$ . Ahora bien, como  $X = ZC^t$ , será

$$Y = ZC^t CZ^t - \sum_{i=1}^p Z_i^2 = Z Z^t - \sum_{i=1}^p Z_i^2 = \sum_{i=p+1}^n Z_i^2,$$

de manera que  $Y$  es independiente de  $Z_1, Z_2, \dots, Z_p$  y, además,  $Y/\sigma^2$  tiene distribución  $\chi_{n-p}^2$ .

En particular la situación anterior (con  $p = 1$ ) se puede aplicar a la varianza muestral, ya que

$$ns^2 = \sum_{i=1}^n X_i^2 - (\sqrt{n}\bar{X})^2 \quad \text{siendo } \sqrt{n}\bar{X} = \left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}}\right) X^t.$$

Se obtiene así que  $ns^2/\sigma^2$  es independiente de  $\bar{X}$  y tiene distribución  $\chi_{n-1}^2$ .

Además, para cualquier valor  $\mu \in \mathbb{R}$ , si se considera  $X'_i = X_i + \mu$ , ( $i = 1, \dots, n$ ), será  $\bar{X}' = \bar{X} + \mu$  y  $s'^2 = s^2$ . Luego  $ns'^2/\sigma^2$  es independiente de  $\bar{X}'$  y tiene distribución  $\chi_{n-1}^2$ . Tal es el contenido del siguiente enunciado:

**Teorema de Fisher:** Si  $(X_1, X_2, \dots, X_n)$  es una muestra aleatoria simple de una población  $N(\mu, \sigma)$ , entonces  $s^2$  y  $\bar{X}$  son variables aleatorias independientes y la distribución en el muestreo de  $ns^2/\sigma^2$  es  $\chi_{n-1}^2$  (mientras que  $\bar{X}$  tiene distribución  $N(\mu, \sigma/\sqrt{n})$ ).

Este resultado confirma, en el caso normal, las conclusiones obtenidas en la sección 2.6 para una distribución arbitraria. Para una distribución normal (en la cual  $\mu_4 = 3\sigma^4$ ) los resultados allí obtenidos se expresan:  $\bar{X}$  es asintóticamente  $N(\mu, \sigma/\sqrt{n})$  y  $s^2$  es asintóticamente  $N(\sigma^2, \sqrt{2/n} \sigma^2)$ .

El teorema de Fisher conduce a la misma conclusión si se hace la aproximación de  $\chi_{n-1}^2$  por  $N(n, \sqrt{2n})$ . La diferencia entre ambos casos estriba en que el teorema de Fisher proporciona la distribución exacta para cualquier  $n$  y no sólo la distribución asintótica. Los ejercicios 3.1, 3.2 y 3.3 muestran la utilidad del teorema anterior.

### 3.3. Estadístico de Student. Distribución $t$ de Student

Saber que  $\bar{X}$  tiene distribución en el muestreo  $N(\mu, \sigma/\sqrt{n})$ , o equivalentemente que

$$\sqrt{n} \frac{\bar{X} - \mu}{\sigma} \text{ tiene distribución } N(0, 1)$$

puede resultar de poca utilidad si la varianza poblacional  $\sigma^2$  es un parámetro desconocido de la distribución teórica, ya que no se puede, entonces, usar esta conclusión para hacer previsiones acerca de la diferencia  $\bar{X} - \mu$ . En tal caso, cabe pensar que el resultado no será muy distinto si se sustituye  $\sigma$  por el valor de la desviación típica muestral,  $s$ , puesto que, al menos para muestras grandes,  $\sigma^2$  y  $s^2$  tendrán valores similares. Tal idea llevó a Student a considerar el estadístico:

$$t = \sqrt{n-1} \frac{\bar{X} - \mu}{s}.$$

La sustitución de  $\sqrt{n}$  por  $\sqrt{n-1}$  quedará explicada un poco más adelante, aunque de momento, si se quiere reforzar la analogía, se puede expresar

$$t = \sqrt{n} \frac{\bar{X} - \mu}{S},$$

donde  $S^2$  es la cuasivarianza muestral ( $S^2 = ns^2/(n-1)$ ) que, a efectos de dar una aproximación de  $\sigma^2$ , tiene tantas ventajas, si no más, que  $s^2$ .

En cualquier caso, para poder usar el estadístico  $t$  en sustitución de  $\sqrt{n}(\bar{X} - \mu)/\sigma$ , la cuestión se centra en determinar su distribución en el muestreo (confiando en que el resultado no dependa de  $\sigma^2$ , pues, de lo contrario, no se habría adelantado nada). Para ello, parece natural prescindir de  $\mu$  y adoptar el siguiente planteamiento:

Sean  $X, X_1, X_2, \dots, X_n$ ,  $n+1$  variables aleatorias independientes, con distribución  $N(0, \sigma)$ ; sea  $Y = \sum_{i=1}^n X_i^2$ , y  $U = \sqrt{Y/n}$ . Se pretende hallar la distribución de  $t = X/U$ .

La distribución de  $Y/\sigma^2$  es  $\chi_n^2$ , así que la densidad de  $Y$  será

$$\frac{1}{2^{n/2}\Gamma(\frac{n}{2})\sigma^n} y^{\frac{n}{2}-1} e^{-y/(2\sigma^2)} \quad \text{para } y > 0,$$

con lo cual la densidad de  $U$  es

$$\frac{n^{n/2}}{2^{(n-2)/2}\Gamma(\frac{n}{2})\sigma^n} u^{n-1} e^{-nu^2/(2\sigma^2)} \quad \text{para } u > 0.$$

Como  $X$  y  $U$  son independientes, su densidad conjunta será

$$Ke^{-x^2/(2\sigma^2)}u^{n-1}e^{-nu^2/(2\sigma^2)} = Ku^{n-1}e^{-(x^2+nu^2)/(2\sigma^2)}$$

para  $x \in \mathbb{R}$  y  $u > 0$ , siendo  $K = \frac{1}{\sqrt{2\pi}} \frac{n^{n/2}}{2^{(n-2)/2}\Gamma(\frac{n}{2})\sigma^{n+1}}$ .

El cambio bidimensional  $z = U$ ,  $t = X/U$ , da como densidad de  $(z, t)$

$$Kz^n e^{-z^2(t^2+n)/(2\sigma^2)} \quad \text{para } t \in \mathbb{R} \text{ y } z > 0;$$

con lo cual la densidad marginal de  $t$  resulta ser

$$\begin{aligned} K \int_0^\infty z^n e^{-z^2(t^2+n)/(2\sigma^2)} dz &= \frac{K}{2} \int_0^\infty v^{(n-1)/2} e^{-v(t^2+n)/(2\sigma^2)} dv \\ &= \frac{K}{2} \left( \frac{t^2 + n}{2\sigma^2} \right)^{-(n+1)/2} \Gamma\left(\frac{n+1}{2}\right) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} \end{aligned}$$

para  $t \in \mathbb{R}$  (después de sustituir la constante  $K$  y simplificar).

Nuevamente la situación merece un nombre propio:

*Si  $X, X_1, X_2, \dots, X_n$  son variables aleatorias independientes y con distribución  $N(0, \sigma)$ , la distribución de*

$$\frac{X}{\sqrt{\frac{1}{n} \sum_{i=1}^n X_i^2}},$$

*de densidad*

$$\frac{1}{\sqrt{n\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} \quad \text{para } t \in \mathbb{R},$$

*se denomina distribución  $t$  de Student con  $n$  grados de libertad.*

Suele decirse que la distribución  $t_n$  es la de una  $N(0, 1)$  dividida por la raíz cuadrada de una  $\chi_n^2$  dividida por sus grados de libertad, ambas independientes entre sí. Lo cual refleja la propiedad fundamental de la distribución de Student: su independencia de la varianza  $\sigma^2$  de las variables que intervienen en su definición.

Antes de examinar otras propiedades de la distribución  $t$ , se puede ver que resuelve el problema que ha suscitado su definición.

**Teorema de Student:** Si  $(X_1, X_2, \dots, X_n)$  es una muestra aleatoria simple de una población  $N(\mu, \sigma)$ , el estadístico de Student:

$$\sqrt{n-1} \frac{\bar{X} - \mu}{s} = \sqrt{n} \frac{\bar{X} - \mu}{S}$$

tiene distribución  $t$  de Student con  $n - 1$  grados de libertad.

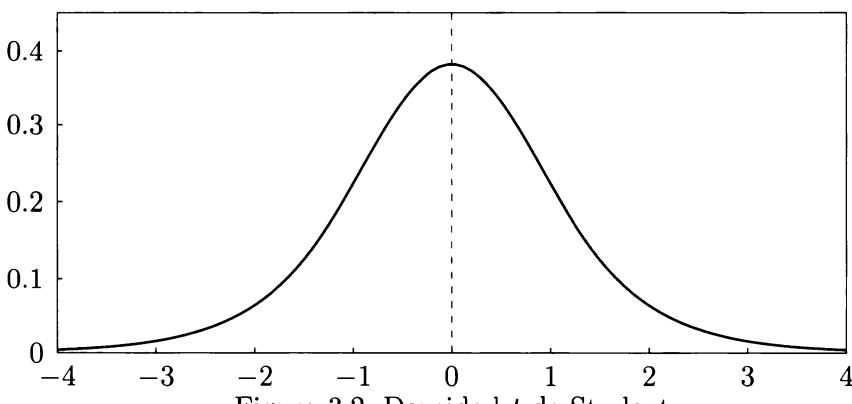
En efecto,

$$\sqrt{n-1} \frac{\bar{X} - \mu}{s} = \frac{\sqrt{n}(\bar{X} - \mu)/\sigma}{\sqrt{\frac{1}{n-1} ns^2/\sigma^2}}$$

y basta observar que  $\sqrt{n}(\bar{X} - \mu)/\sigma$  tiene distribución  $N(0, 1)$ , mientras que, según el teorema de Fisher,  $ns^2/\sigma^2$  tiene la misma distribución que  $\sum_{i=1}^{n-1} X'_i{}^2$ , siendo  $X'_1, \dots, X'_{n-1}$  variables  $N(0, 1)$ , independientes entre sí y de la media muestral  $\bar{X}$ . Por tanto,  $\sqrt{n-1}(\bar{X} - \mu)/s$  tiene distribución  $t_{n-1}$ .

Cuando  $\sigma^2$  sea desconocida, el resultado permite juzgar que valores son previsibles para  $\bar{X} - \mu$ , utilizando  $S$  en lugar de  $\sigma$  y la distribución  $t_{n-1}$  en lugar de la  $N(0, 1)$ . Véase el ejercicio 3.3.

Acerca de la distribución  $t$  de Student, cabe señalar que, al igual que la distribución  $\chi^2$ , depende de un único parámetro  $n$  (el número de sumandos que intervienen en la suma del denominador) que nuevamente se denomina “grados de libertad” de la distribución. En cambio, la distribución de Student tiene como soporte el intervalo  $(-\infty, \infty)$ , siendo su densidad simétrica respecto al origen, de manera que su representación gráfica (para  $n = 10$ ) es de la forma:



de apariencia semejante a la normal.

Para  $n = 1$ , la densidad se reduce a  $\pi^{-1}/(1 + t^2)$ , que corresponde a una distribución de Cauchy sin momentos de primer orden. Si  $n > 1$ , la integral  $\int_0^\infty t(1+t^2/n)^{-(n+1)/2} dt$  es finita, de manera que la media existe y vale cero (por simetría), junto con todos los momentos de orden impar que existan (concretamente los de orden  $\leq n$ ). La varianza existe para  $n > 2$  y (haciendo el cambio  $(1 + t^2/n)^{-1} = u$ ) vale

$$\frac{2\Gamma(\frac{n+1}{2})}{\sqrt{n}\pi\Gamma(\frac{n}{2})} \int_0^\infty t^2 \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} dt = \frac{n\Gamma(\frac{n+1}{2})}{\sqrt{\pi}\Gamma(\frac{n}{2})} \beta\left(\frac{n-2}{2}, \frac{3}{2}\right) = \frac{n}{n-2};$$

de forma que la dispersión decrece rápidamente hacia 1 cuando  $n$  crece.

De hecho, suponiendo, sin pérdida de generalidad, que las variables  $X, X_1, \dots, X_n$  que intervienen en la definición de la distribución  $t$  tienen varianza 1, cuando  $n \rightarrow \infty$ , será

$$\frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{c.s.} 1 \quad \text{y, por tanto,} \quad t_n \xrightarrow{d} N(0, 1).$$

La función de distribución asociada a la densidad de  $t_n$  tampoco puede expresarse explícitamente y sus valores numéricos aparecen en la tabla 5, con una disposición influida, como en el caso de la  $\chi^2$ , por la utilización habitual que se hace de ella. Concretamente, en el cuerpo de la tabla aparece la abscisa  $t_p$  tal que  $1 - F(t_p)$  es igual a la probabilidad  $p$  indicada en el margen superior, según el número de grados de libertad que figuran en el margen izquierdo.

Claro que, por simetría,  $F(-t_p) = 1 - F(t_p) = p$ ; lo cual permite hacer todas las combinaciones precisas. Así por ejemplo, si  $T_5$ ,  $T_{12}$  y  $T_{17}$  tienen distribución  $t$  con 5, 12 y 17 grados de libertad respectivamente:

$$P\{T_{17} > 0'698\} = 0'25,$$

$$P\{T_{17} < -2'11\} = 0'025,$$

$$P\{T_{17} < 0'395\} = 1 - 0'35,$$

$$P\{T_5 > -0'559\} = 1 - 0'3,$$

$$P\{T_{12} < -0'539\} + P\{T_{12} > 0'539\} = 0'6,$$

$$P\{-1'114 < T_{12} < 1'114\} = 1 - 0'3,$$

$$P\{-1'198 < T_5 < 2'571\} = 1 - 0'025 - 0'15,$$

$$P\{0'413 < T_5 < 1'476\} = 1 - 0'1 - 1 + 0'35, \text{ etc...}$$

De acuerdo con la convergencia en distribución señalada, la última fila de la tabla corresponde a la distribución  $N(0, 1)$  y contiene (en forma bastante más imprecisa) la misma información que la tabla de esta distribución.

Cabe observar que las columnas de la tabla decrecen monótonamente hacia el valor correspondiente a la  $N(0, 1)$ ; lo cual tiene una repercusión interesante en el problema estadístico subyacente.

Por una parte, el hecho de que  $\sqrt{n}(\bar{X} - \mu)/S$  tenga distribución  $t_{n-1}$ , permite evaluar probabilísticamente las discrepancias entre  $\bar{X}$  y  $\mu$  mediante afirmaciones del tipo:

$$P\left\{|\bar{X} - \mu| < t_p \frac{S}{\sqrt{n}}\right\} = 1 - p,$$

donde  $p$  se escoge convenientemente pequeño y el valor de  $t_p$  es suministrado por la fila  $n$ -ésima de la tabla.

Por otra parte, en el caso en que la varianza poblacional  $\sigma^2$  fuese conocida (y sólo en ese caso), el mismo tipo de conclusiones puede extraerse del hecho de que  $\sqrt{n}(\bar{X} - \mu)/\sigma$  tenga distribución  $N(0, 1)$ :

$$P\left\{|\bar{X} - \mu| < z_p \frac{\sigma}{\sqrt{n}}\right\} = 1 - p,$$

siendo ahora  $z_p$  el valor en la última fila de la tabla.

Pues bien, a igualdad entre  $S$  y  $\sigma$ , la segunda afirmación es más precisa que la primera, por ser  $z_p < t_p$ . Ello indica que, cuando se conoce  $\sigma^2$ , no es conveniente ignorar tal información y emplear el estadístico de Student, pues se obtendrán, generalmente, peores resultados.

### 3.4. Distribución de la diferencia de medias muestrales

Hay que abordar ahora la situación en la cual las características poblacionales en estudio constituyen una variable aleatoria bidimensional  $(X, Y)$ , cuyas componentes son independientes y tienen distribuciones  $N(\mu_1, \sigma_1)$  y

$N(\mu_2, \sigma_2)$  respectivamente. Una muestra aleatoria simple, de tamaño  $n$ , estará entonces compuesta por  $n$  réplicas independientes  $(X_1, Y_1), \dots, (X_n, Y_n)$  de la variable  $(X, Y)$ .

Sin embargo, debido a la independencia entre las componentes  $X$  e  $Y$ , resulta más clara la interpretación de que se trata de efectuar el estudio comparativo de dos poblaciones (independientes)  $X$  e  $Y$ , con distribuciones normales  $N(\mu_1, \sigma_1)$  y  $N(\mu_2, \sigma_2)$  respectivamente. Incluso ello tiene la ventaja de que no obliga a suponer que los tamaños muestrales en ambas poblaciones sean iguales, sino que se puede contar, como observaciones, con dos muestras aleatorias simples  $(X_1, X_2, \dots, X_n)$  e  $(Y_1, Y_2, \dots, Y_m)$ , una de cada población.

Puesto que se trata de comparar ambas poblaciones y, en primer lugar, sus medias desconocidas,  $\mu_1$  y  $\mu_2$ , parece natural que el análisis esté basado en las dos medias muestrales  $\bar{X}$  e  $\bar{Y}$  y, más concretamente, en su diferencia:  $\bar{X} - \bar{Y}$ . Tiene por tanto interés determinar la distribución en el muestreo del estadístico  $\bar{X} - \bar{Y}$  (definido en el espacio muestral  $\mathcal{X} = \mathbb{R}^{n+m}$ , en el que se considera la distribución de “la” muestra  $(X_1, \dots, X_n, Y_1, \dots, Y_m)$ ).

En principio, la solución no puede ser más simple:  $\bar{X}$  tiene distribución  $N(\mu_1, \sigma_1/\sqrt{n})$ ,  $\bar{Y}$  tiene distribución  $N(\mu_2, \sigma_2/\sqrt{m})$  y ambas son variables aleatorias independientes (puesto que la primera es función de  $(X_1, \dots, X_n)$  y la segunda de  $(Y_1, \dots, Y_m)$ ).

Por tanto  $\bar{X} - \bar{Y}$  tendrá distribución normal de media  $E[\bar{X} - \bar{Y}] = \mu_1 - \mu_2$  y varianza  $V(\bar{X} - \bar{Y}) = V(\bar{X}) + V(\bar{Y}) = \sigma_1^2/n + \sigma_2^2/m$ . Es decir que

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n + \sigma_2^2/m}} \quad \text{tiene distribución } N(0, 1), \quad (3.1)$$

lo cual permitirá, supuesto que  $\sigma_1^2$  y  $\sigma_2^2$  son conocidas, analizar probabilísticamente las diferencias entre  $\bar{X} - \bar{Y}$  y  $\mu_1 - \mu_2$ , con el fin de obtener información sobre ésta a partir de aquella.

Supuesto que las varianzas poblacionales son desconocidas, el resultado anterior, aunque sea cierto, no puede cumplir el mismo cometido y hay que buscar una solución de tipo Student, que reemplace la afirmación anterior por otra que no dependa de las varianzas poblacionales.

Si  $S_1^2$  y  $S_2^2$  son las cuasivarianzas de cada muestra, de acuerdo con el teorema de Fisher,  $(n - 1)S_1^2/\sigma_1^2$  tiene distribución  $\chi_{n-1}^2$  y  $(m - 1)S_2^2/\sigma_2^2$  tiene distribución  $\chi_{m-1}^2$ . Luego, según la propiedad de aditividad de la dis-

tribución  $\chi^2$ ,

$$\frac{(n-1)S_1^2}{\sigma_1^2} + \frac{(m-1)S_2^2}{\sigma_2^2} \quad \text{tiene distribución } \chi_{n+m-2}^2.$$

Puesto que  $\bar{X}$  e  $\bar{Y}$  son independientes de  $S_1^2$  y  $S_2^2$  (en virtud nuevamente del teorema de Fisher), la definición de la distribución  $t$  asegura que

$$\frac{[\bar{X} - \bar{Y} - (\mu_1 - \mu_2)] / \sqrt{\sigma_1^2/n + \sigma_2^2/m}}{\sqrt{\frac{1}{n+m-2} [(n-1)S_1^2/\sigma_1^2 + (m-1)S_2^2/\sigma_2^2]}} \quad \text{tiene distribución } t_{n+m-2}.$$

Sin embargo esta conclusión no es más aprovechable que la anterior, a no ser que se pudiese eliminar de la expresión del estadístico los parámetros desconocidos  $\sigma_1^2$  y  $\sigma_2^2$ . La única alternativa para que ello sea posible, es suponer que, aunque desconocidas, *las varianzas poblacionales son iguales*. En tal caso, la afirmación anterior se reduce a:

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2} \sqrt{\frac{1}{n} + \frac{1}{m}}}} \quad \text{tiene distribución } t_{n+m-2}, \quad (3.2)$$

que ya no involucra ningún parámetro desconocido.

En el caso en que las varianzas poblacionales, además de ser desconocidas, no puedan suponerse iguales, no hay solución exacta al problema de determinar la distribución en el muestreo de  $\bar{X} - \bar{Y}$ , lo cual obliga a adoptar soluciones aproximadas.

*Si los tamaños muestrales no son muy pequeños, ( $n, m \geq 15$ , suele emplearse como regla práctica), cabe argumentar que  $S_1^2$  y  $S_2^2$  constituirán "buenas" aproximaciones de  $\sigma_1^2$  y  $\sigma_2^2$  respectivamente. Con lo cual, sustituyendo en (3.1), las varianzas poblacionales por las cuasivarianzas muestrales, se obtiene que*

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{S_1^2/n + S_2^2/m}} \quad \text{será aproximadamente } N(0, 1). \quad (3.3)$$

Para el caso en que  $n$  ó  $m$  sean pequeños, se han propuesto diferentes aproximaciones, la más popular de las cuales es la debida a Welch, según la cual

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{S_1^2/n + S_2^2/m}} \quad \text{tiene aproximadamente distribución } t_f, \quad (3.4)$$

siendo el número de grados de libertad,  $f$ , el entero más próximo a

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

En resumen, las afirmaciones (3.1)–(3.4) cubren todos los casos posibles (con mayor o menor precisión) y permiten siempre predecir el tamaño de la diferencia entre  $\bar{X} - \bar{Y}$  y  $\mu_1 - \mu_2$ . (Véase, a este propósito, el ejercicio 3.4.)

Es necesario sin embargo hacer una advertencia importante. A menudo se plantea una situación formalmente muy similar a la anterior pero teóricamente distinta.

Si  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$  es una muestra aleatoria simple de una población bidimensional  $(X, Y)$ , cuyas componentes no son independientes, sino que se distribuye con distribución normal de medias  $(\mu_1, \mu_2)$  y matriz de covarianzas

$$\begin{pmatrix} \sigma_1^2 & \sigma_{11} \\ \sigma_{11} & \sigma_2^2 \end{pmatrix}$$

$(\sigma_{11} \neq 0)$ , puede pretenderse comparar las medias  $\mu_1$  y  $\mu_2$ , sobre la base de las observaciones realizadas.

Un ejemplo muy típico de esta situación consiste en comparar las medias de una cierta característica de los individuos de una población, antes ( $X$ ) y después ( $Y$ ) de un cierto tratamiento. Presumiblemente no hay entonces independencia entre cada observación  $X_i$  y la correspondiente  $Y_i$ , lo cual obliga, en primer lugar, a presentar los *datos apareados* (nombre con el que se denomina familiarmente a esta situación).

Ni que decir tiene que, en tal caso, los resultados (3.1)–(3.4) no son aplicables, al no cumplirse las condiciones de independencia en que han sido deducidos. De hecho, la solución a adoptar ahora, consiste en reducir el problema al caso unidimensional analizado en la sección 3.3.

En efecto,  $X - Y$  tendrá distribución  $N(\mu_1 - \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2 - 2\sigma_{11}})$ , y las diferencias  $(X_1 - Y_1, X_2 - Y_2, \dots, X_n - Y_n)$  constituirán una muestra aleatoria simple de la población  $X - Y$ . Por tanto, su media muestral  $\bar{X} - \bar{Y}$  será  $N(\mu_1 - \mu_2, \sqrt{(\sigma_1^2 + \sigma_2^2 - 2\sigma_{11})/n})$  o, dicho de otra manera,

$$\sqrt{n} \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\sigma_{11}}} \quad \text{tiene distribución } N(0, 1).$$

Si alguno de los elementos de la matriz de covarianzas poblacional no fuese conocido, el resultado anterior no es utilizable, pero, según el teorema

de Student,

$$\sqrt{n-1} \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{s^*} \quad \text{tiene distribución } t_{n-1},$$

donde  $s^{*2}$  es ahora la varianza muestral de la muestra  $(X_i - Y_i)$ ; es decir,  $s^{*2} = \frac{1}{n} \sum_{i=1}^n (X_i - Y_i)^2 - (\bar{X} - \bar{Y})^2$ .

### 3.5. Distribución del cociente de cuasivarianzas muestrales. Distribución $F$ de Snedecor

En la misma situación de independencia considerada al principio de la sección anterior y supuesto que las varianzas poblacionales,  $\sigma_1^2$  y  $\sigma_2^2$ , sean desconocidas, el estudio comparativo de las poblaciones  $X$  e  $Y$  exige también la comparación de las varianzas poblacionales. De hecho, en la práctica, convendrá ocuparse en primer lugar de las varianzas, puesto que, como se ha visto, la distribución a utilizar del estadístico  $\bar{X} - \bar{Y}$ , adecuado para analizar la diferencia entre  $\mu_1$  y  $\mu_2$ , depende de que  $\sigma_1^2$  y  $\sigma_2^2$  puedan considerarse iguales o no.

El método para obtener información acerca de la relación entre  $\sigma_1^2$  y  $\sigma_2^2$  estará naturalmente basado en las cuasivarianzas muestrales  $S_1^2$  y  $S_2^2$  (o lo que es equivalente en  $s_1^2$  y  $s_2^2$ ). Pero en vez de considerar  $S_1^2 - S_2^2$ , conviene utilizar el estadístico  $S_1^2/S_2^2$ , cuya distribución en el muestreo puede determinarse explícitamente.

Para ello, el problema a resolver es, en esencia, el siguiente:

Si  $X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_m$  son variables aleatorias independientes con distribución común  $N(0, 1)$ , hay que hallar la distribución de

$$\frac{1/n \sum_{i=1}^n X_i^2}{1/m \sum_{i=1}^m Y_i^2}.$$

Pero  $U = \sum_{i=1}^n X_i^2$  y  $V = \sum_{i=1}^m Y_i^2$  son variables independientes con distribución  $\chi_n^2$  y  $\chi_m^2$  respectivamente; así que la densidad conjunta de  $(U, V)$  será

$$\frac{1}{2^{(n+m)/2} \Gamma(\frac{n}{2}) \Gamma(\frac{m}{2})} u^{(n-2)/2} v^{(m-2)/2} e^{-(u+v)/2} \quad \text{para } u, v \geq 0.$$

Y la cuestión se reduce a hacer el cambio bidimensional

$$\begin{cases} S = V \\ T = \frac{m}{n} \frac{U}{V} \end{cases}$$

que da como densidad conjunta de  $(S, T)$ :

$$\frac{(n/m)^{n/2}}{2^{(n+m)/2}\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} s^{(n+m-2)/2} t^{(n-2)/2} e^{-s(1+nt/m)/2} \quad \text{para } s, t \geq 0.$$

De donde se obtiene como densidad marginal de  $T$

$$\begin{aligned} & \frac{(n/m)^{n/2}}{2^{(n+m)/2}\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} t^{(n-2)/2} \int_0^\infty s^{(n+m-2)/2} e^{-s(1+nt/m)/2} ds \\ &= \frac{(n/m)^{n/2}}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} t^{(n-2)/2} \left(1 + \frac{n}{m}t\right)^{-(n+m)/2} \int_0^\infty v^{(n+m-2)/2} e^{-v} dv \\ &= \frac{\Gamma(\frac{n+m}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} \left(\frac{n}{m}\right)^{n/2} t^{(n-2)/2} \left(1 + \frac{n}{m}t\right)^{-(n+m)/2} \quad \text{para } t \geq 0. \end{aligned}$$

La distribución obtenida depende de dos parámetros  $n$  y  $m$ , que representan respectivamente el número de sumandos del numerador y del denominador, y no depende en cambio de la varianza común de las variables  $X_i$  e  $Y_i$ ; en el sentido de que supuesto que las  $X_i$  y las  $Y_i$  fuesen  $N(0, \sigma)$ , sería

$$\frac{1/n \sum_{i=1}^n X_i^2}{1/m \sum_{i=1}^m Y_i^2} = \frac{1/n \sum_{i=1}^n (X_i/\sigma)^2}{1/m \sum_{i=1}^m (Y_i/\sigma)^2}$$

y la situación es idéntica a la anterior.

Según ello, puede darse la siguiente “definición”:

*Si  $X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_m$  son variables aleatorias independientes, con distribución  $N(0, \sigma)$  la distribución de*

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

*se denomina distribución  $F$  de Snedecor con  $n$  y  $m$  grados de libertad y tiene densidad*

$$\frac{\Gamma(\frac{n+m}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} \left(\frac{n}{m}\right)^{n/2} t^{(n-2)/2} \left(1 + \frac{n}{m}t\right)^{-(n+m)/2} \quad \text{para } t \geq 0.$$

Dicho de otra manera,  $F_{n,m}$  es la distribución del cociente de dos  $\chi^2$  independientes, de  $n$  y  $m$  grados de libertad respectivamente, divididas cada una de ellas por sus grados de libertad.

La representación gráfica de la densidad de la  $F_{n,m}$  tiene la misma apariencia que la de  $\chi^2$ , dependiendo su forma exacta de  $n$  y  $m$ .

El momento de orden  $r$  de  $F_{n,m}$ , que existirá si  $r < m/2$ , se calcula fácilmente mediante el cambio  $1 - y = (1 + nt/m)^{-1}$ , con el cual se obtiene:

$$\begin{aligned} & \left(\frac{n}{m}\right)^{n/2} \int_0^\infty t^{(n+2r-2)/2} \left(1 + \frac{n}{m}t\right)^{-(n+m)/2} dt \\ &= \left(\frac{m}{n}\right)^r \int_0^1 \frac{y^{(n+2r-2)/2}}{(1-y)^{(m+2r-2)/2}} (1-y)^{(n+m)/2} \frac{dy}{(1-y)^2} \\ &= \left(\frac{m}{n}\right)^r \beta\left(\frac{n}{2} + r, \frac{m}{2} - r\right), \end{aligned}$$

de forma que el momento de orden  $r$  vale

$$\frac{\Gamma(\frac{n}{2} + r)\Gamma(\frac{m}{2} - r)}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} \left(\frac{m}{n}\right)^r$$

(para  $r < m/2$ , ya que la última integral sólo converge en ese caso).

Si  $m > 2$ , la media resulta entonces

$$\frac{\Gamma(\frac{n}{2} + 1)\Gamma(\frac{m}{2} - 1)}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} \frac{m}{n} = \frac{n/2}{(m-2)/2} \frac{m}{n} = \frac{m}{m-2}$$

que, sorprendentemente, no depende del número de grados de libertad del numerador. Y, si  $m > 4$ , la varianza es

$$\begin{aligned} & \frac{\Gamma(\frac{n}{2} + 2)\Gamma(\frac{m}{2} - 2)}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} \left(\frac{m}{n}\right)^2 - \left(\frac{m}{m-2}\right)^2 = \frac{(n/2 + 1)n/2}{(m/2 - 1)(m/2 - 2)} \frac{m^2}{n^2} - \frac{m^2}{(m-2)^2} \\ &= \frac{2m^2(n+m-2)}{n(m-2)^2(m-4)}. \end{aligned}$$

La función de distribución de la  $F$  de Snedecor aparece tabulada en la tabla 6. En ella, los grados de libertad del numerador figuran en el margen superior, mientras que los del denominador aparecen en el margen izquierdo, subdivididos en varios valores de  $p$ . El cuerpo de la tabla da las abscisas  $x_p$  tales que  $1 - F(x_p) = p$ . (Véase la figura que acompaña a la tabla 6.)

Es inútil incluir en la tabla valores de  $p$  superiores a  $1/2$ , debido a la siguiente observación: Si  $T$  tiene distribución  $F_{n,m}$ , por la propia definición,

$T' = 1/T$  tiene distribución  $F_{m,n}$ ; de manera que para obtener el valor  $x_p$  tal que  $P\{T > x_p\} = p$  con  $p > 1/2$ , basta razonar:

$$p = P\{T > x_p\} = P\left\{\frac{1}{T} < \frac{1}{x_p}\right\} = P\left\{T' < \frac{1}{x_p}\right\} = 1 - P\left\{T' > \frac{1}{x_p}\right\};$$

luego, como  $1-p < 1/2$ , la tabla de  $F_{m,n}$  (la distribución de  $T'$ ) proporciona el valor  $x'_{1-p}$  tal que  $P\{T' > x'_{1-p}\} = 1-p$ . Y será  $x_p = 1/x'_{1-p}$ . Es decir que para determinar  $x_p$ , basta leer en la tabla correspondiente a  $1-p$ , con los grados de libertad permutados, e invertir el valor que proporcione la tabla.

Una vez introducida la distribución  $F$  de Snedecor, resta por verificar que resuelve el problema estadístico planteado: determinar la distribución en el muestreo del cociente  $S_1^2/S_2^2$ , de las cuasivarianzas de dos muestras aleatorias simples de sendas poblaciones normales independientes,  $N(\mu_1, \sigma_1)$  y  $N(\mu_2, \sigma_2)$ .

Pues bien, de acuerdo con el teorema de Fisher,  $(n-1)S_1^2/\sigma_1^2$  tiene distribución  $\chi_{n-1}^2$  (es decir, la misma distribución que  $\sum_{i=1}^{n-1} X_i'^2$  con  $X'_i$ ,  $i = 1, \dots, n-1$ , independientes y  $N(0, 1)$ ). Análogamente  $(m-1)S_2^2/\sigma_2^2$  tiene distribución  $\chi_{m-1}^2$  (la misma que  $\sum_{i=1}^{m-1} Y_i'^2$  siendo  $Y'_i$ ,  $i = 1, \dots, m-1$  independientes y  $N(0, 1)$ ).

Por supuesto,  $S_1^2$  y  $S_2^2$  son independientes entre sí, por provenir de poblaciones independientes. Por consiguiente, se verifica

**Teorema 3.1:** Si  $S_1^2$  y  $S_2^2$  son las cuasivarianzas de sendas muestras aleatorias simples, de dos poblaciones normales de varianzas  $\sigma_1^2$  y  $\sigma_2^2$  respectivamente, el estadístico

$$U = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$

tiene distribución  $F$  de Snedecor con  $n-1$  y  $m-1$  grados de libertad.

Ahora no importa que el estadístico  $F$  dependa de las varianzas poblacionales desconocidas. Por el contrario, la utilidad del resultado radica en que, conociendo el valor de  $S_2^2/S_1^2$  que proporciona la muestra, la distribución de  $U$  indica que valores son verosímiles para  $\sigma_2^2/\sigma_1^2 = U S_2^2/S_1^2$ . Véase ejercicio 3.4.

### 3.6. El coeficiente de correlación muestral

En el muestreo de una población normal bidimensional  $(X, Y)$ , sobre cuyas componentes no haya constancia de que sean independientes, uno de los problemas estadísticos más relevantes es comprobar, mediante una muestra aleatoria simple  $((X_1, Y_1), (X_2, Y_2) \dots, (X_n, Y_n))$ , si el coeficiente de correlación teórico,  $\rho$ , es cero y se verifica efectivamente la independencia entre  $X$  e  $Y$ .

Parece natural basar el análisis acerca del valor de  $\rho$ , en el coeficiente de correlación muestral

$$R = \frac{S_{11}}{S_1 S_2} \quad \text{donde } S_{11} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

es la cuasicovarianza muestral.

La determinación de la distribución en el muestreo del estadístico  $R$ , sin más hipótesis, resulta difícil<sup>3</sup>; pero, afortunadamente, bastará determinarla bajo la hipótesis de que  $\rho$  es cero, para que pueda servir para hacer la comprobación de que es efectivamente  $\rho = 0$ .

Supongamos, pues, que es  $\rho = 0$  y, dado que  $R$  no varía al sustituir las  $X_i$  por  $(X_i - \mu_1)/\sigma_1$  y las  $Y_i$  por  $(Y_i - \mu_2)/\sigma_2$ , se puede suponer, sin pérdida de generalidad, que es  $\mu_1 = \mu_2 = 0$  y  $\sigma_1 = \sigma_2 = 1$ .

Sean los vectores

$$\begin{aligned} c_1 &= \left( \frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}} \right) \\ c_2 &= \left( \frac{X_1 - \bar{X}}{\sqrt{n-1} S_1}, \frac{X_2 - \bar{X}}{\sqrt{n-1} S_1}, \dots, \frac{X_n - \bar{X}}{\sqrt{n-1} S_1} \right) \end{aligned}$$

que cumplen  $c_1 c_1^t = 1$ ,  $c_2 c_2^t = 1$  y  $c_1 c_2^t = 0$ . Seleccionemos  $n - 2$  vectores,  $c_3, \dots, c_n$  que junto con  $c_1$  y  $c_2$  constituyan una base ortonormal de  $\mathbb{R}^n$ , de manera que la matriz  $n \times n$ :

$$C = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} \quad \text{verificará} \quad CC^t = I.$$

Condicionado por los valores de las variables  $X_i$ , el vector  $Z^t = CY^t$  es un vector  $(n \times 1)$  de variables aleatorias normales (combinaciones lineales

---

<sup>3</sup> Puede verse, por ejemplo, en Cramer (1968).

de las  $Y_i$ ), de media  $E[Z] = E[Y]C = 0$  y matriz de covarianzas

$$E[Z^t Z] = C E[Y^t Y] C^t = C C^t = I.$$

Además  $Z_1 = \sqrt{n} \bar{Y}$  y

$$Z_2 = \frac{\sum_{i=1}^n Y_i(X_i - \bar{X})}{\sqrt{(n-1)S_1^2}} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{n-1}S_1} = \sqrt{n-1} \frac{S_{11}}{S_1}.$$

Por otra parte

$$(n-1)S_2^2 = \sum_{i=1}^n Y_i^2 - n\bar{Y}^2 = YY^t - n\bar{Y}^2 = ZZ^t - Z_1^2 = \sum_{i=2}^n Z_i^2.$$

Luego resulta

$$R^2 = \frac{S_{11}^2}{S_1^2 S_2^2} = \frac{Z_2^2}{(n-1)S_2^2} = \frac{Z_2^2}{\sum_{i=2}^n Z_i^2} = \frac{Z_2^2}{Z_2^2 + W}$$

siendo  $W = \sum_{i=3}^n Z_i^2$ .

Ahora bien, condicionado por los valores de las  $X_i$ ,  $Z_2^2$  tiene distribución  $\chi_1^2$  (puesto que es el cuadrado de una  $N(0, 1)$ ) y  $W$  tiene distribución  $\chi_{n-2}^2$ ; además ambas son independientes. Resulta entonces que la distribución de  $(Z_2, W)$ , condicionada por los valores de las  $X_i$ , no depende de ellos. Es decir que  $(Z_2, W)$  es independiente de las  $X_i$  y tiene densidad conjunta

$$\frac{1}{\sqrt{2}\Gamma(\frac{1}{2})} z^{-1/2} e^{-z/2} \frac{1}{2^{(n-2)/2}\Gamma(\frac{n-2}{2})} w^{(n-4)/2} e^{-w/2} \quad \text{para } z, v > 0.$$

Haciendo el cambio  $U = Z_2 + W$ ,  $V = Z_2/(Z_2 + W)$  se obtiene como densidad de  $(U, V)$

$$\frac{1}{2^{(n-1)/2}\Gamma(\frac{1}{2})\Gamma(\frac{n-2}{2})} u^{(n-3)/2} e^{-u/2} v^{-1/2} (1-v)^{(n-4)/2} \quad \text{en } u > 0, v \in [0, 1].$$

La densidad marginal de  $V = R^2$  es entonces

$$\frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{n-2}{2})} v^{-1/2} (1-v)^{(n-4)/2} \quad \text{para } v \in [0, 1];$$

y la de  $R$

$$\frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{n-2}{2})} (1-r^2)^{(n-4)/2} \quad \text{para } r \in [-1, 1].$$

La distribución obtenida para el coeficiente de correlación muestral (bajo la hipótesis  $\rho = 0$ ) no se emplea directamente en esta forma, ya que ello requeriría la tabulación de la función de distribución correspondiente a la densidad anterior. Puesto que hacer previsiones sobre el valor de  $R$  es equivalente a hacerlas sobre una función biyectiva de  $R$ , resulta preferible hacer el cambio

$$R^* = \sqrt{n-2} \frac{R}{\sqrt{1-R^2}}$$

que permite obtener, con facilidad, la siguiente conclusión.

**Teorema 3.2:** Si  $R$  es el coeficiente de correlación muestral de una distribución normal bidimensional, con coeficiente de correlación  $\rho = 0$ ,

$$R^* = \sqrt{n-2} \frac{R}{\sqrt{1-R^2}}$$

tiene distribución  $t$  de Student con  $n - 2$  grados de libertad.

Esta última sección suscita, a poco que se piense, más problemas de los que resuelve. En una normal bidimensional  $(X, Y)$ , no sólo queda por determinar la distribución en el muestreo de  $R$  cuando  $\rho$  no es cero; cabe también preguntarse cuál es la distribución conjunta de  $S_1^2$ ,  $S_2^2$  y  $S_{11}$ ; o, si existe un estadístico tipo Student, en el que intervengan las diferencias  $\bar{X} - \mu_1$  y  $\bar{Y} - \mu_2$ , cuya distribución en el muestreo no dependa de  $\sigma_1^2, \sigma_2^2, \sigma_{11}$ ; etc... Además, es natural que se plantee la cuestión de qué ocurre si el número de dimensiones aumenta. La respuesta a todas estas dudas suele ser el inicio de lo que se conoce como análisis estadístico multivariante. Y, para los propósitos de este curso, los resultados expuestos serán suficientes.

## Ejercicios

### Ejercicio 3.1

Dada una muestra aleatoria simple de tamaño 10 de una distribución normal  $N(\mu, 2)$ , calcular la probabilidad de que la media muestral y la poblacional difieran en más de 0'5.

¿De qué tamaño habría que seleccionar una muestra para poder afirmar, con probabilidad 0'9, que la media muestral diferirá de la poblacional en menos de 0'1?

### Ejercicio 3.2

La longitud craneal en una determinada población humana es una variable aleatoria que sigue una distribución normal con media 185'6 mm. y desviación típica 12'7 mm. ¿Cuál es la probabilidad de que una muestra aleatoria simple de tamaño 20 de esa población tenga media mayor que 190 mm?

### Ejercicio 3.3

La efectividad en días de un determinado antibiótico, sigue una distribución normal de media 14 días y desviación típica desconocida. Fue administrado a 16 enfermos, obteniéndose una cuasidesviación típica muestral de 1'4 días. Determinar la probabilidad de que la efectividad media en la muestra no supere los 13 días, que es el tiempo mínimo de efectividad requerido.

Preocupados por una posible subestimación de la varianza poblacional, que podría llevar a subestimar la probabilidad de que no se alcance la efectividad mínima, se desea determinar la probabilidad de que con una muestra de 16 enfermos se subestime la varianza en más de un 20 %. Si la muestra es de 61 pacientes, esta probabilidad ¿aumenta o disminuye?

Determinar el tamaño de muestra necesario para que la probabilidad anterior sea 0'05.

### Ejercicio 3.4

Un inversor bursátil compra o vende acciones de “Cementos S.A.” por el siguiente procedimiento: Selecciona al azar una muestra de días para los que determina el índice medio,  $\bar{x}_1$ , de la empresa “Construcciones S.A.”; selecciona también al azar otra muestra de días, para los que determina el índice medio,  $\bar{x}_2$ , de “Cementos S.A.”. Compra acciones de “Cementos S.A.” cuando  $\bar{x}_2 \leq \bar{x}_1$ , vendiendo en caso contrario.

El inversor supone que ambos índices bursátiles se distribuyen normal e independientemente con una diferencia de medias de 1'432 unidades en favor de la primera y que los índices bursátiles de ambas empresas se comportan de forma independiente todos los días.

Calcular la probabilidad de que el inversor compre en los siguientes supuestos:

- (a) Seleccionó 11 días para calcular el índice medio de “Construcciones S.A.”, obteniendo una cuasivarianza  $S_1^2 = 25$ , y 13 días para el segundo índice, en los que obtuvo una cuasivarianza de  $S_2^2 = 6$ .
- (b) Seleccionó 60 días para calcular el primer índice, y obtuvo  $S_1^2 = 23$  y 50 días para el segundo con  $S_2^2 = 7$ .

- (c) Supone que cada día existe dependencia entre los índices de ambas empresas, con lo que selecciona al azar 20 días en los que observa ambos índices, obteniendo una cuasivarianza muestral para la diferencia de índices de  $S^*{}^2 = 10'4$ .

### Ejercicio 3.5

Si  $(X_1, \dots, X_n)$  es una muestra aleatoria simple de una  $N(\mu, \sigma)$ , con  $\mu$  conocida y  $\sigma$  desconocida, comparar las distribuciones en el muestreo, la media y la varianza, de la cuasivarianza muestral  $S^2$  y de

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

### Ejercicio 3.6

Sea  $X$  una variable aleatoria con distribución  $N(\mu_1, \sigma_1)$  siendo  $\mu_1$  conocida y  $\sigma_1$  desconocida, e  $Y$  una variable aleatoria, independiente de  $X$ , con distribución  $N(\mu_2, \sigma_2)$  siendo  $\mu_2$  y  $\sigma_2$  desconocidas.

Determinar un estadístico razonable para obtener información sobre el cociente  $\sigma_1^2/\sigma_2^2$ , basado en sendas muestras aleatorias independientes de  $X$  e  $Y$ , de tamaños  $n_1$  y  $n_2$  respectivamente. Obtener su distribución en el muestreo.

### Ejercicio 3.7

El fabricante de una dieta de adelgazamiento afirma que su producto permite una reducción media del peso de 3'5 kilogramos. Con objeto de investigar su eficacia se seleccionaron al azar 40 personas, observando en ellas el peso antes de aplicar la dieta,  $X$  y el peso después de acabar el tratamiento,  $Y$ , lo que proporcionó una cuasivarianza para la diferencia de  $S_d^*{}^2 = 1/39 \sum_1^{40} [(X_i - Y_i) - (\bar{X} - \bar{Y})]^2 = 1'8$ .

Si suponemos que tanto  $X$  como  $Y$  siguen distribuciones normales, determinar la probabilidad de que en los individuos de la muestra haya una reducción media de más de tres kilogramos.

### Ejercicio 3.8

Si  $X$  es una variable aleatoria con distribución  $\gamma(n, \theta)$  siendo  $n$  un número natural, determinar una función de  $X$  que tenga distribución  $\chi^2$ .

### Ejercicio 3.9

El tiempo de fallo de ciertos sistemas tiene distribución exponencial de media  $2\theta$ . Se prueban  $n$  de ellos a la vez e independientemente, para obtener

una muestra aleatoria simple,  $X_1, \dots, X_n$  de la duración. Naturalmente, los tiempos de fallo se observan en orden creciente; primero  $X_{(1)}$ , después  $X_{(2)}$ , etc.

(a) Determinar la distribución en el muestreo de  $(X_{(1)}, \dots, X_{(r)})$ .

(b) Determinar la distribución en el muestreo de

$$\frac{1}{\theta} \left\{ \sum_{i=1}^r X_{(i)} + (n-r)X_{(r)} \right\}.$$

### Ejercicio 3.10

Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de una distribución con función de densidad

$$f(x) = e^{-x+b} \quad \text{para } x \geq b \geq 0$$

y sean  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$  los estadísticos ordenados asociados.

(a) Determinar la distribución en el muestreo de  $(X_{(1)}, \dots, X_{(n)})$ .

(b) Determinar la distribución en el muestreo de  $2 \sum_{i=1}^n [X_i - X_{(1)}]$ .

(c) Si  $b = 0$ , determinar la distribución de

$$\frac{(n-1)nX_{(1)}}{\sum_{i=1}^n [X_i - X_{(1)}]}.$$

## Solución de los ejercicios

### Ejercicio 3.1

Como la distribución poblacional es normal de desviación típica conocida,  $\sigma = 2$ , la distribución de la media muestral será  $N(\mu, 2/\sqrt{10})$ . Por tanto, si  $Z$  es  $N(0, 1)$

$$\mathcal{P}\{|\bar{X} - \mu| > 0'5\} = P\{|Z| > 0'5\sqrt{10}/2\} = 0'42927.$$

Respecto al tamaño de muestra necesario para que sea

$$\mathcal{P}\{|\bar{X} - \mu| < 0'1\} = 0'9,$$

como  $\bar{X}$  es  $N(\mu, 2/\sqrt{n})$ , debe ser  $P\{|Z| < 0'05\sqrt{n}\} = 0'9$ . Por otro lado,  $P\{|Z| < 1'645\} = 0'9$ , luego

$$0'05\sqrt{n} = 1'645;$$

es decir,  $n = 1.082'41$ . Tomando como tamaño muestral  $n = 1.083$ , se obtiene la precisión deseada.

### Ejercicio 3.2

La longitud craneal,  $X$ , tiene distribución  $N(185'6, 12'7)$ . La media muestral tendrá, por tanto, una distribución  $N(185'6, 12'7/\sqrt{20})$ , de manera que, si  $Z$  es  $N(0, 1)$ ,

$$\mathcal{P}\{\bar{X} > 190\} = P\{Z > (190 - 185'6) \frac{\sqrt{20}}{12'7}\} = P\{Z > 1'55\} = 0'0606.$$

### Ejercicio 3.3

La probabilidad es  $\mathcal{P}\{\bar{X} \leq 13\}$ . Como la población es normal, pero no se conoce la varianza poblacional, habrá que utilizar la distribución  $t$  de Student.

$$\mathcal{P}\{\bar{X} \leq 13\} = \mathcal{P}\left\{\frac{\bar{X} - \mu}{S/\sqrt{n}} \leq \frac{13 - 14}{1'4/\sqrt{16}}\right\} = P\{t_{15} \leq -2'857\} = 0'0063$$

donde el valor se obtiene, por simetría, en la distribución  $t_{15}$ .

Como  $(n - 1)S^2/\sigma^2$  tiene distribución  $\chi^2_{n-1}$ , la probabilidad de subestimación de  $\sigma^2$  en más de un 20% es

$$\mathcal{P}\{S^2 < 0'8\sigma^2\} = P\{\chi^2_{15} < 12\} = 0'3,$$

calculada a partir de las tablas por interpolación lineal.

Para una muestra de 61 individuos, utilizando la aproximación de la distribución  $\chi_n^2$ ,

$$\begin{aligned}\mathcal{P}\{S^2 < 0'8\sigma^2\} &= \mathcal{P}\{\chi_{60}^2 < 48\} \simeq \mathcal{P}\left\{\sqrt{2\chi_{60}^2} - \sqrt{119} < \sqrt{96} - \sqrt{119}\right\} \\ &= \mathcal{P}\{Z < -1'11\} = 0'1335.\end{aligned}$$

Es decir, la probabilidad disminuye al aumentar el tamaño de la muestra.

Por último, para conseguir que

$$0'05 = \mathcal{P}\{\chi_{n-1}^2 < 0'8(n-1)\} \simeq \mathcal{P}\left\{Z < \sqrt{1'6(n-1)} - \sqrt{2n-1}\right\}$$

ha de ser  $\sqrt{1'6(n-1)} - \sqrt{2n-1} = -1'645$ . De donde  $n = 117'61$ ; con tamaño muestral 118, se obtiene la precisión requerida.

### Ejercicio 3.4

(a) Para calcular  $\mathcal{P}\{\bar{X}_2 \leq \bar{X}_1\}$  es necesario determinar la distribución de  $\bar{X}_1 - \bar{X}_2$  y para ello es preciso decidir si las varianzas poblacionales  $\sigma_1^2$  y  $\sigma_2^2$  son iguales o no. Habrá que razonar, por tanto, cual de las dos opciones es la más verosímil, a la vista de las observaciones. Se sabe que

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \quad \text{tiene distribución } F_{n-1, m-1};$$

como, si  $F$  es una variable aleatoria con distribución  $F_{(10,12)}$ , es

$$\mathcal{P}\{F > 2'7534\} = 0'05 \quad \text{y} \quad \mathcal{P}\{F < 0'3433\} = 0'05,$$

si las varianzas poblacionales fuesen iguales habría probabilidad 0'9 de obtener un cociente de cuasivarianzas muestrales comprendido entre 0'3433 y 2'7534. Puesto que el cociente  $S_1^2/S_2^2$  ha sido igual a  $25/6 = 4'17$ , los datos no avalan la igualdad de varianzas poblacionales.

Habrá que usar, entonces la aproximación de Welch, según la cual

$$\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{S_1^2/n + S_2^2/m}} \quad \text{tiene distribución } t_{15},$$

ya que se obtiene  $f = 14'755$ . Por tanto,

$$\mathcal{P}\{\bar{X}_2 \leq \bar{X}_1\} = \mathcal{P}\left\{t_{15} \geq \frac{-1'432}{\sqrt{2'734}}\right\} = \mathcal{P}\{t_{15} \geq -0'866\} = 0'8.$$

(b) Como el tamaño de la muestra es suficientemente grande, puede utilizarse la aproximación normal, para obtener  $P\{\bar{X}_2 \leq \bar{X}_1\}$ :

$$P\left\{\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{S_1^2/n + S_2^2/m}} > \frac{-1'432}{\sqrt{0'523}}\right\} = P\{Z > -1'98\} = 0'9761.$$

Obsérvese que el tamaño de la muestra es lo suficientemente grande para poder utilizar la distribución normal, sin necesidad de suponer normalidad en las distribuciones poblacionales de los índices.

(c) En cuestiones bursátiles la independencia es una suposición de dudosa veracidad. Por ello, este apartado la evita, aunque solamente la independencia entre ambos índices, ya que la independencia entre las observaciones de diferentes días resulta necesaria para obtener una muestra aleatoria simple.

Ahora se trata de un caso de datos apareados, de forma que

$$\begin{aligned} P\{\bar{X}_1 - \bar{X}_2 \geq 0\} &= P\left\{\frac{\bar{X}_1 - \bar{X}_2 - 1'432}{S^*/\sqrt{n}} \geq \frac{-1'432}{\sqrt{0'52}}\right\} \\ &= P\{t_{19} > -1'986\} = 0'969. \end{aligned}$$

### Ejercicio 3.5

$(n - 1)S^2/\sigma^2$  tiene distribución  $\chi_{n-1}^2$ ; luego

$$E[S^2] = \sigma^2 \quad \text{y} \quad V(S^2) = \frac{\sigma^4}{(n-1)^2} V((n-1)S^2/\sigma^2) = \frac{2\sigma^4}{n-1}.$$

Por otro lado, como  $(X_i - \mu)/\sigma$  es  $N(0, 1)$ ,  $nT/\sigma^2$  será  $\chi_n^2$ , por ser suma de  $n$  cuadrados de  $N(0, 1)$  independientes. Así pues,

$$E[T] = \sigma^2 \quad \text{y} \quad V(T) = \frac{\sigma^4}{n^2} V(nT/\sigma^2) = \frac{2\sigma^4}{n}.$$

Ambos estadísticos se distribuyen alrededor de la media  $\sigma^2$ , y la dispersión del segundo es inferior a la del primero. Si  $n$  es pequeño,  $T$  está considerablemente más concentrado alrededor de  $\sigma^2$  que  $S^2$ ; de manera que conviene hacer uso, mediante  $T$ , del conocimiento de  $\mu$ , en lugar de ignorarlo usando  $S^2$ . Cuando  $n$  crece, la distribución de ambos estadísticos se concentra rápidamente alrededor de  $\sigma^2$  y la diferencia entre ambos se hace inapreciable. (Naturalmente, si  $\mu$  fuese desconocido, los valores de  $T$  no podrían determinarse.)

**Ejercicio 3.6**

Por las razones expuestas en el ejercicio anterior, parece razonable utilizar el estadístico  $T_1/S_2^2$ , siendo  $S_2^2$  la cuasivarianza de la muestra de  $Y_i$ , y

$$T_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} (X_i - \mu_1)^2.$$

Como  $n_1 T_1 / \sigma_1^2$  tiene distribución  $\chi_{n_1}^2$  y  $(n_2 - 1)S_2^2 / \sigma_2^2$  es  $\chi_{n_2-1}^2$ ,

$$\frac{T_1/\sigma_1^2}{S_2^2/\sigma_2^2}$$

tendrá distribución  $F_{n_1, n_2-1}$ . Por consiguiente,  $T_1/S_2^2$  es el producto del parámetro a estimar  $\sigma_1^2/\sigma_2^2$  por una variable con distribución  $F_{n_1, n_2-1}$ .

**Ejercicio 3.7**

Como las observaciones  $X_i$  e  $Y_i$  se realizan en los mismos individuos, antes y después del tratamiento, se trata de un problema de datos apareados. La diferencia de peso  $d = X - Y$  tiene media  $\mu_d = \mu_X - \mu_Y = 3'5$  y  $S_d^{*2} = 1'8$ . Por tanto

$$\mathcal{P}\{\bar{d} > 3\} = \mathcal{P}\left\{\frac{\bar{d} - \mu_d}{S_d^*/\sqrt{n}} > \frac{3 - 3'5}{\sqrt{1'8/40}}\right\} = P\{t_{39} > -2'36\} = 0'988,$$

después de interpolar en la tabla de la  $t$  de Student.

**Ejercicio 3.8**

Si  $X$  tiene densidad  $\frac{\theta^n}{\Gamma(n)} x^{n-1} e^{-\theta x}$ , para  $x > 0$ , la variable aleatoria  $Y = 2\theta X$  tendrá por densidad

$$\frac{\theta^n}{\Gamma(n)} \left(\frac{y}{2\theta}\right)^{n-1} e^{-y/2} \frac{1}{2\theta} = \frac{1}{2^n \Gamma(n)} y^{n-1} e^{-y/2} \quad \text{para } y > 0,$$

que es la densidad de una  $\chi_{2n}^2$ .

**Ejercicio 3.9**

(a) Si  $X_1, \dots, X_n$  es una muestra aleatoria simple de una distribución absolutamente continua con función de distribución  $F$  y función de densidad  $f$ , la función de densidad conjunta de  $(X_{(j_1)}, \dots, X_{(j_k)})$ , con  $1 \leq j_1 < j_2 < \dots < j_k \leq n$ , es

$$\frac{n! F(y_1)^{j_1-1} f(y_1) [F(y_2) - F(y_1)]^{j_2-j_1-1} f(y_2) \cdots [1 - F(y_k)]^{n-j_k} f(y_k)}{(j_1-1)! (j_2-j_1-1)! \cdots (n-j_k)!}$$

si  $y_1 < \dots < y_k$ .

Aquí,  $j_i = i$ , para  $i = 1, \dots, r$ , luego la densidad de  $X_{(1)}, \dots, X_{(r)}$ , es

$$\frac{n!}{(n-r)!} f(y_1)f(y_2) \cdots f(y_r)[1 - F(y_r)]^{n-r} \quad \text{para } y_1 < \dots < y_r.$$

Con la distribución exponencial de media  $2\theta$ , resulta

$$\frac{n!}{(n-r)!} \frac{1}{(2\theta)^r} \exp \left\{ -\frac{1}{2\theta} \left[ \sum_{i=1}^r y_i + (n-r)y_r \right] \right\} \quad \text{para } y_1 < \dots < y_r.$$

(b) Sean

$$\begin{aligned} Z_1 &= \frac{n}{\theta} X_{(1)}, \\ Z_i &= \frac{n-i+1}{\theta} (X_{(i)} - X_{(i-1)}) \quad \text{para } i = 2, \dots, r. \end{aligned}$$

Como

$$\sum_{i=1}^r Z_i = \frac{1}{\theta} \left( \sum_{i=1}^r X_{(i)} + (n-r)X_{(r)} \right)$$

y

$$\frac{\partial(Z_1, \dots, Z_r)}{\partial(X_{(1)}, \dots, X_{(r)})} = \frac{n!}{(n-r)!\theta^r},$$

la densidad conjunta de  $(Z_1, \dots, Z_r)$  es

$$\frac{1}{2^r} \exp \left\{ -\frac{1}{2} \sum_{i=1}^r z_i \right\} \quad \text{para } z_1, \dots, z_r > 0.$$

De forma que las  $Z_i, i = 1, \dots, r$  son independientes y tienen distribución  $\chi_2^2$ .

Por tanto

$$\frac{1}{\theta} \left( \sum_{i=1}^r X_{(i)} + (n-r)X_{(r)} \right) = \sum_{i=1}^r Z_i$$

tendrá distribución  $\chi_{2r}^2$ .

### Ejercicio 3.10

(a) La función de densidad conjunta de  $(X_{(1)}, \dots, X_{(n)})$  (obtenida en el ejercicio 1.4) es

$$n! \exp \left\{ b n - \sum_{i=1}^n y_i \right\} \quad \text{para } b < y_1 < \dots < y_n.$$

(b) Si llamamos

$$\begin{aligned} Z_1 &= nX_{(1)}, \\ Z_i &= (n - i + 1)(X_{(i)} - X_{(i-1)}) \quad \text{para } i = 2, \dots, n, \end{aligned}$$

como

$$\sum_{i=1}^n Z_i = \sum_{i=1}^n X_{(i)} \quad \text{y} \quad \frac{\partial(Z_1, \dots, Z_n)}{\partial(X_{(1)}, \dots, X_{(n)})} = n!,$$

la densidad conjunta de  $(Z_1, \dots, Z_n)$  será

$$\exp \left\{ b n - \sum_{i=1}^n z_i \right\} \quad \text{para } z_1 \geq n b, z_2, \dots, z_n \geq 0.$$

Así que  $Z_1, Z_2, \dots, Z_n$  son variables independientes y  $Z_2, \dots, Z_n$  tienen distribución exponencial de parámetro 1 o, lo que es lo mismo,  $2Z_2, \dots, 2Z_n$  tienen distribución  $\chi^2_2$ . Además

$$\sum_{i=2}^n 2Z_i = 2 \sum_{i=1}^n X_{(i)} - 2nX_{(1)} = 2 \sum_{i=1}^n [X_i - X_{(1)}],$$

luego el estadístico  $2 \sum_{i=1}^n [X_i - X_{(1)}]$  tiene distribución  $\chi^2_{2n-2}$ .

(c) En el caso  $b = 0$ ,  $2Z_1$  tiene también distribución  $\chi^2_2$ . Además

$$\frac{(n-1)nX_{(1)}}{\sum_{i=1}^n [X_i - X_{(1)}]} = \frac{Z_1}{\frac{1}{2n-2} \sum_{i=2}^n 2Z_i}$$

es el cociente de dos  $\chi^2$ , independientes, divididas por sus grados de libertad. Luego tiene distribución  $F_{2,2n-2}$ .

# Capítulo 4

## Intervalos de confianza

### 4.1. Introducción

El estudio de las distribuciones en el muestreo, abordado en los capítulos anteriores, ha puesto de relieve un gran número de casos en los que un determinado estadístico  $T$  se distribuye con escasa variabilidad alrededor de algún parámetro poblacional,  $\theta$ . Cabe esperar, por tanto, que el valor de  $T$ , observado en una muestra concreta, sea relativamente próximo a  $\theta$ ; o dicho al revés, que  $\theta$  sea relativamente próximo al valor observado de  $T$ . La distribución en el muestreo del estadístico  $T$  y su valor observado en una realización de la muestra, permiten entonces diseñar un método para determinar qué márgenes de variación son previsibles para el parámetro  $\theta$ , en el supuesto de que éste fuese desconocido. Tal método se conoce con el nombre de *estimación confidencial* o *estimación por intervalos de confianza*.

Pretender obtener información acerca del valor desconocido del parámetro  $\theta$ , es enfrentarse al problema estadístico paramétrico, descrito en el capítulo 1: La distribución teórica de una cierta población pertenece a una familia paramétrica

$$\mathcal{F} = \{F_\theta \mid \theta \in \Theta \subset \mathbb{R}^k\}.$$

Se dispone, además, de la posibilidad de extraer una muestra aleatoria (generalmente simple), para disminuir el desconocimiento sobre el valor de  $\theta$  que realmente corresponde a la población.

En el caso de la estimación confidencial, disminuir el desconocimiento acerca del valor de  $\theta$ , significa seleccionar un subconjunto  $\Theta_1$  del espacio paramétrico  $\Theta$ , en el cual puede afirmarse, con un margen de error pequeño, que se encuentra el valor de  $\theta$  que caracteriza la distribución de la población.

Por tanto el problema radica en estudiar métodos para la selección del subconjunto  $\Theta_1$ , precisando previamente qué debe entenderse por “un margen de error pequeño”.

## 4.2. Definición de intervalo de confianza

La idea básica de los intervalos de confianza es fácil de comprender en casos concretos:

### Ejemplo 4.1

Supongamos que  $(X_1, X_2, \dots, X_n)$  es una muestra aleatoria simple de una población con distribución teórica  $N(\theta, \sigma)$ , en la cual  $\sigma$  es una constante conocida, mientras que  $\theta$  tiene un valor desconocido. Se sabe entonces que el estadístico

$$T = \frac{\bar{X} - \theta}{\sigma/\sqrt{n}}$$

tiene distribución en el muestreo  $N(0, 1)$ . Por consiguiente, consultando las tablas de la  $N(0, 1)$  se obtiene, por ejemplo,

$$\mathcal{P}_\theta\{-1'96 < T < 1'96\} = 0'95$$

siendo  $\mathcal{P}_\theta$  la distribución de la muestra. La desigualdad  $-1'96 < T < 1'96$  equivale a

$$\bar{X} - 1'96 \frac{\sigma}{\sqrt{n}} < \theta < \bar{X} + 1'96 \frac{\sigma}{\sqrt{n}}$$

de manera que es

$$\mathcal{P}_\theta \left\{ \bar{X} - 1'96 \frac{\sigma}{\sqrt{n}} < \theta < \bar{X} + 1'96 \frac{\sigma}{\sqrt{n}} \right\} = 0'95$$

cualquiera que sea el valor de  $\theta$ .

La interpretación es clara: hay una probabilidad igual a 0'95, de obtener una muestra tal que el intervalo de extremos aleatorios

$$(\bar{X} - 1'96\sigma/\sqrt{n}, \bar{X} + 1'96\sigma/\sqrt{n})$$

incluya al valor de  $\theta$  con el que se han producido las observaciones muestrales; y ello independientemente de cual sea dicho valor.

Una vez utilizada la distribución en el muestreo de  $\bar{X}$  para establecer la conclusión anterior, se puede hacer efectivo el muestreo; obtener unos resultados concretos:  $x_1, x_2, \dots, x_n$ ; calcular con ellos el valor,  $\bar{x}$ , del estadístico y determinar el intervalo numérico

$$(\bar{x} - 1'96\sigma/\sqrt{n}, \bar{x} + 1'96\sigma/\sqrt{n}).$$

Desde luego, ahora no tiene sentido pensar que hay probabilidad 0'95 de que el intervalo numérico contenga al número  $\theta$  con el que se han producido las observaciones, puesto que ni uno ni otro son aleatorios. De todas formas, 0'95 expresa el *margen de confianza* con el que se puede afirmar que el valor desconocido de  $\theta$  está entre los extremos del intervalo; en el sentido de que repitiendo el muestreo un gran número de veces, se obtendrían intervalos distintos, entre los cuales, muy aproximadamente, el 95 % contendrían el valor de  $\theta$  correcto. El intervalo numérico

$$(\bar{x} - 1'96\sigma/\sqrt{n}, \bar{x} + 1'96\sigma/\sqrt{n})$$

se denomina entonces un *intervalo de confianza para  $\theta$  con nivel de confianza 0'95* (o del 95 %).

Por supuesto, el procedimiento puede repetirse con cualquier otro nivel de confianza. Por ejemplo, como  $\mathcal{P}_\theta\{-1'645 < T < 1'645\} = 0'9$ , el intervalo

$$(\bar{x} - 1'645\sigma/\sqrt{n}, \bar{x} + 1'645\sigma/\sqrt{n})$$

es un intervalo de confianza para  $\theta$  de nivel de confianza 0'9.

Análogamente, puesto que  $\mathcal{P}_\theta\{-2'576 < T < 2'576\} = 0'99$ ,

$$(\bar{x} - 2'576\sigma/\sqrt{n}, \bar{x} + 2'576\sigma/\sqrt{n})$$

será un intervalo de confianza para  $\theta$ , de nivel de confianza 0'99.

Como es lógico, a mayor nivel de confianza exigido al intervalo, mayor longitud adquiere éste; o al revés, la disminución del recorrido estimado del parámetro se paga con un mayor riesgo de error. Hay que compaginar pues ambas cosas; fijar el nivel de confianza que parezca adecuado según el contexto del problema (usualmente 0'9, 0'95, 0'975, ó 0'99) y aceptar el intervalo que resulte.

Cabe observar otro grado de libertad en la construcción de los intervalos de confianza; consultando nuevamente la tabla de la distribución  $N(0, 1)$  se obtiene, por ejemplo,  $\mathcal{P}_\theta\{-1'74 < T < 2'37\} = 0'95$ , de manera que

$$(\bar{x} - 2'37\sigma/\sqrt{n}, \bar{x} + 2'37\sigma/\sqrt{n})$$

es también un intervalo de confianza para  $\theta$  de nivel de confianza 0'95. Sin embargo, respecto al intervalo de confianza del mismo nivel construido antes, la longitud ha aumentado:  $2 \times 1'96\sigma/\sqrt{n}$  frente a  $(1'74 + 2'37)\sigma/\sqrt{n}$ , sin conseguir ninguna reducción en el riesgo de error (0'05). Resulta pues inconveniente disminuir la precisión de la estimación del parámetro, sin aumentar la confianza en la estimación. En este sentido, a la vista de la función de densidad de la  $N(0, 1)$ , es claro que la longitud,  $(b - a)\sigma/\sqrt{n}$ , siendo  $a$  y  $b$  tales que  $P\{a < T < b\} = 0'95$ , es mínima cuando  $a = -b$ <sup>1</sup>.

La técnica empleada en el ejemplo anterior puede utilizarse en muchas otras situaciones similares. Pero conviene, en primer lugar, establecer el concepto general de intervalo de confianza:

*De una población descrita por una variable aleatoria  $X$ , cuya distribución teórica  $F$  pertenece a una familia paramétrica*

$$\mathcal{F} = \{F_\theta \mid \theta \in \Theta \subset \mathbb{R}^k\}$$

*se considera una muestra aleatoria  $(X_1, X_2, \dots, X_n)$ , con distribución  $\mathcal{P}_\theta$ . Sea  $g(\theta)$  una función del parámetro, con valores reales, y  $T_1 \leq T_2$  dos estadísticos unidimensionales tales que*

$$\mathcal{P}_\theta \{ T_1(X_1, \dots, X_n) \leq g(\theta) \leq T_2(X_1, \dots, X_n) \} \geq 1 - \alpha$$

*para cada  $\theta \in \Theta$ . Entonces, para cualquier muestra concreta  $(x_1, x_2, \dots, x_n)$ , el intervalo*

$$[ T_1(x_1, \dots, x_n), T_2(x_1, \dots, x_n) ]$$

*se denomina intervalo de confianza para  $g(\theta)$ , de nivel de confianza  $1 - \alpha$ .*

La definición requiere algunas observaciones:

---

<sup>1</sup> Una comprobación formal, bastante innecesaria, consiste en minimizar  $b - a$ , sujeto a la restricción  $F(b) - F(a) = 0'95$ . Derivando respecto a  $a$  y a  $b$ , la función  $L = b - a - \lambda[F(b) - F(a) - 0'95]$ , se obtiene

$$\frac{\partial L}{\partial a} = -1 + \lambda f(a) = 0, \quad \frac{\partial L}{\partial b} = 1 - \lambda f(b) = 0$$

de donde  $f(a) = f(b)$  y, en definitiva,  $a = -b$ .

- Supuesto que se conoce la distribución conjunta de la muestra,  $\mathcal{P}_\theta$ , no es necesario suponer que el muestreo sea aleatorio simple; lo que tal hipótesis simplifica es la determinación de  $\mathcal{P}_\theta$ . Obviamente, tampoco es necesario que  $X$  sea una variable unidimensional.
- En el ejemplo de introducción se conseguía la igualdad

$$\mathcal{P}_\theta\{T_1 < \theta < T_2\} = 1 - \alpha,$$

mientras que la definición es más imprecisa, puesto que sólo exige que dicha probabilidad sea  $\geq 1 - \alpha$ . La razón es que la continuidad de la distribución  $N(0, 1)$ , utilizada en el ejemplo, permite igualar la probabilidad al nivel de confianza prefijado y hace inútil la inclusión de los extremos del intervalo  $[T_1, T_2]$ . En cambio, en general —sobre todo cuando  $X$  sea una variable discreta—, la discontinuidad de las distribuciones de  $T_1$  y  $T_2$  puede impedir alcanzar exactamente el valor  $1 - \alpha$  y aconseja incluir los extremos  $T_1$  y  $T_2$ , para aumentar la probabilidad de que  $[T_1, T_2]$  contenga a  $g(\theta)$ , sin incrementar la longitud del intervalo. Si el valor  $1 - \alpha$  no es exactamente accesible, se trata de garantizar que el intervalo  $[T_1, T_2]$  incluya al valor  $g(\theta)$  con probabilidad al menos  $1 - \alpha$ ; de ahí el signo de desigualdad.

Formalmente resulta entonces que, todo intervalo con un cierto nivel confianza (0'99, por ejemplo) es un intervalo de confianza de cualquier nivel inferior (0'95, por ejemplo). Pero, como lógicamente interesa considerar intervalos lo menos extensos posibles, sobre pasar sin necesidad el nivel de confianza especificado ha de considerarse técnicamente incorrecto.

- La definición anterior involucra una función  $g$  que reduce el parámetro  $\theta$ , en general  $k$ -dimensional, a una sola dimensión. Si  $\theta$  fuese unidimensional, nada impide considerar  $g(\theta) = \theta$ , para obtener un intervalo de confianza para  $\theta$ ; la generalidad extra que supone la función  $g$  puede ser de utilidad en muchos casos en los que el interés no se centre directamente en el parámetro, sino en una función de él.

Por supuesto, lo que no tiene sentido es considerar un intervalo de confianza (unidimensional) para un parámetro de más dimensiones. Pero la generalización del concepto de intervalo de confianza al caso  $k$ -dimensional no presenta grandes dificultades:

En las mismas condiciones anteriores, si  $S(x_1, x_2, \dots, x_n)$  es un subconjunto de  $\Theta \subset I\!\!R^k$ , dependiente de la muestra  $(x_1, x_2, \dots, x_n)$ , que verifica

$$\mathcal{P}_\theta \{S(X_1, X_2, \dots, X_n) \ni \theta\} \geq 1 - \alpha \quad \text{para cualquier } \theta \in \Theta,$$

el subconjunto  $S(x_1, x_2, \dots, x_n)$ , asociado a cualquier muestra concreta, se denomina *región de confianza* para el parámetro  $\theta$ , de *nivel de confianza*  $1 - \alpha$ .

No obstante, la dificultad de su construcción hace que las regiones de confianza se consideren menos habitualmente que los intervalos de confianza. Las próximas secciones se dedicarán al estudio de técnicas para construir intervalos de confianza.

### 4.3. Método de la cantidad pivotal para la construcción de intervalos de confianza

Los intervalos de confianza obtenidos en el ejemplo 4.1, se basan en el hecho de que

$$\frac{\bar{X} - \theta}{\sigma/\sqrt{n}}$$

tiene distribución en el muestreo independiente de  $\theta$ . La misma técnica puede utilizarse en numerosas circunstancias y suele denominarse *método de la cantidad pivotal*. La descripción general es la siguiente:

Supongamos que  $T(X_1, X_2, \dots, X_n; \theta)$  es una función real de la muestra y del parámetro, cuya distribución en el muestreo no depende de  $\theta$ . En tal caso, fijado cualquier nivel de confianza  $1 - \alpha$  entre 0 y 1, se pueden determinar constantes,  $c_1$  y  $c_2$ , (que no serán únicas) tales que

$$\mathcal{P}_\theta \{c_1 \leq T(X_1, X_2, \dots, X_n; \theta) \leq c_2\} \geq 1 - \alpha.$$

Basta, por ejemplo, que  $c_1$  y  $c_2$  sean los cuantiles de orden  $\alpha_1$  y  $1 - \alpha_2$ , con  $\alpha_1 + \alpha_2 = \alpha$ , de la distribución en el muestreo de  $T(X_1, \dots, X_n; \theta)$ .

Si es posible despejar  $g(\theta)$  en las desigualdades

$$c_1 \leq T(X_1, X_2, \dots, X_n; \theta) \quad \text{y} \quad T(X_1, X_2, \dots, X_n; \theta) \leq c_2,$$

obtendremos sendos valores  $T_1(X_1, X_2, \dots, X_n)$  y  $T_2(X_1, X_2, \dots, X_n)$  tales que, para cualquier  $\theta \in \Theta$ ,

$$\mathcal{P}_\theta \{T_1(X_1, \dots, X_n) \leq g(\theta) \leq T_2(X_1, \dots, X_n)\} \geq 1 - \alpha.$$

De manera que

$$[T_1(x_1, x_2, \dots, x_n), T_2(x_1, x_2, \dots, x_n)]$$

será un intervalo de confianza para  $g(\theta)$ , de nivel de confianza  $1 - \alpha$ , cualquiera que sea la muestra obtenida  $(x_1, x_2, \dots, x_n)$ . Ello es fácil de llevar a cabo en muchas circunstancias.

### Ejemplo 4.2

Consideremos una población con distribución uniforme en  $(0, \theta)$ , cuya distribución teórica es por tanto

$$F_\theta(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ x/\theta & \text{si } 0 < x < \theta \\ 1 & \text{si } x \geq \theta. \end{cases}$$

Dada una muestra aleatoria simple de tamaño  $n$ , el estadístico

$$T(X_1, X_2, \dots, X_n; \theta) = \frac{1}{\theta} \max_{1 \leq i \leq n} X_i = \frac{X_{(n)}}{\theta}$$

tiene distribución en el muestreo

$$\mathcal{P}_\theta \{T(X_1, X_2, \dots, X_n; \theta) \leq y\} = \mathcal{P}_\theta \{X_{(n)} \leq \theta y\} = \begin{cases} 0 & \text{si } y \leq 0 \\ y^n & \text{si } 0 < y < 1 \\ 1 & \text{si } y \geq 1 \end{cases}$$

que no depende de  $\theta$ .

Fijado entonces el nivel de confianza  $1 - \alpha$ , si  $\alpha_1, \alpha_2 > 0$  son tales que  $\alpha_1 + \alpha_2 = \alpha$  será

$$\mathcal{P}_\theta \left\{ \alpha_1^{1/n} \leq T(X_1, X_2, \dots, X_n; \theta) \leq (1 - \alpha_2)^{1/n} \right\} = 1 - \alpha_2 - \alpha_1 = 1 - \alpha.$$

Por otra parte, las ecuaciones

$$\frac{X_{(n)}}{\theta} = \alpha_1^{1/n} \quad \text{y} \quad \frac{X_{(n)}}{\theta} = (1 - \alpha_2)^{1/n}$$

tienen por soluciones

$$\theta = \alpha_1^{-1/n} X_{(n)} \quad \text{y} \quad \theta = (1 - \alpha_2)^{-1/n} X_{(n)}.$$

La segunda es menor que la primera, luego

$$\left( (1 - \alpha_2)^{-1/n} X_{(n)}, \alpha_1^{-1/n} X_{(n)} \right)$$

constituye un intervalo de confianza para  $\theta$ , de nivel de confianza  $1 - \alpha$ , cualquiera que sean  $\alpha_1, \alpha_2 > 0$ , con  $\alpha_1 + \alpha_2 = \alpha$ . Como su longitud

$$\left[ \alpha_1^{-1/n} - (1 - \alpha_2)^{-1/n} \right] x_{(n)} = \left[ (\alpha - \alpha_2)^{-1/n} - (1 - \alpha_2)^{-1/n} \right] x_{(n)}$$

es una función creciente de  $\alpha_2$ , la longitud mínima se obtiene para  $\alpha_2 = 0$ ,  $\alpha_1 = \alpha$ . Es decir que el intervalo más preciso entre los anteriores es

$$\left( x_{(n)}, \alpha^{-1/n} x_{(n)} \right).$$

La sección 4.5 contiene muchos otros ejemplos de utilización de cantidades pivotales. Por otra parte, la determinación de un estadístico que permita emplear el método de la cantidad pivotal puede llevarse a cabo, en ocasiones, gracias al hecho siguiente:

Si  $(X_1, X_2, \dots, X_n)$  es una muestra aleatoria simple de una población unidimensional cuya distribución teórica pertenece a una familia  $\mathcal{F} = \{F_\theta : \theta \in \Theta \subset \mathbb{R}\}$  cuyos elementos son funciones continuas de  $x$ , entonces el estadístico

$$T(X_1, X_2, \dots, X_n; \theta) = - \sum_{i=1}^n \log F_\theta(X_i)$$

tiene distribución en el muestreo independiente de  $\theta$ .

De hecho, para cualquier valor  $\theta \in \Theta$ ,  $-\log F(X_i)$  tiene distribución exponencial de parámetro 1, puesto que si  $y > 0$ ,

$$\mathcal{P}_\theta \{-\log F_\theta(X_i) \leq y\} = \mathcal{P}_\theta \{X_i \geq F_\theta^{-1}(e^{-y})\} = 1 - F_\theta(F_\theta^{-1}(e^{-y})) = 1 - e^{-y}.$$

Luego  $-\sum_{i=1}^n \log F_\theta(X_i)$  tiene distribución gamma  $\gamma(n, 1)$ .

Aunque la formulación de la afirmación anterior sea muy general, no siempre conduce a buenos resultados. Por ejemplo, en el contexto del ejemplo 4.1, la obtención de  $-\sum_{i=1}^n \log F_\theta(X_i)$  da lugar a cálculos inviables. En cambio, en otras ocasiones el método es operativo.

### Ejemplo 4.3

Sea  $(X_1, X_2, \dots, X_n)$  una muestra aleatoria simple de una población con distribución teórica

$$F_\theta(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ x^\theta & \text{si } 0 < x < 1 \\ 1 & \text{si } x \geq 1. \end{cases}$$

Puesto que es  $0 < X_1, X_2, \dots, X_n < 1$ , se obtiene

$$-\sum_{i=1}^n \log F_\theta(X_i) = -\sum_{i=1}^n \log X_i^\theta = -\theta \sum_{i=1}^n \log X_i.$$

Además dicho estadístico tiene distribución gamma  $\gamma(n, 1)$ ; de manera que si  $0 < c_1 < c_2$  son tales que

$$\int_{c_1}^{c_2} \frac{t^{n-1} e^{-t}}{(n-1)!} dt = 1 - \alpha$$

será, para cualquier  $\theta > 0$ ,

$$\mathcal{P}_\theta \left\{ c_1 < -\theta \sum_{i=1}^n \log X_i < c_2 \right\} = 1 - \alpha.$$

En consecuencia

$$\left( \frac{c_1}{-\sum_{i=1}^n \log x_i}, \quad \frac{c_2}{-\sum_{i=1}^n \log x_i} \right)$$

es un intervalo de confianza para  $\theta$  de nivel de confianza  $1 - \alpha$ .

Los cálculos explícitos relativos a la distribución gamma  $\gamma(n, a)$ , con  $n$  entero, pueden llevarse a cabo mediante la tabla de la distribución  $\chi^2$ . Por ejemplo, en este caso,

$$\int_{c_1}^{c_2} \frac{t^{n-1} e^{-t}}{\Gamma(n)} dt = \int_{2c_1}^{2c_2} \frac{u^{n-1} e^{-u/2}}{2^n \Gamma(n)} du$$

siendo el integrando la función de densidad de la distribución  $\chi^2_{2n}$ . Así, si el tamaño muestral fuese  $n = 7$  y se tratase de obtener el intervalo anterior correspondiente a  $1 - \alpha = 0'9$ , la tabla de la  $\chi^2_{14}$  proporciona  $c_1 = 6'571/2$ ,  $c_2 = 23'685/2$  como valores posibles.

#### Ejemplo 4.4

En la situación del ejemplo 4.2, como  $0 < x_1, \dots, x_n < \theta$ ,

$$Y = -\sum_{i=1}^n \log F_\theta(x_i) = -\sum_{i=1}^n \log \frac{x_i}{\theta} = -\sum_{i=1}^n \log x_i + n \log \theta$$

tiene distribución  $\gamma(n, 1)$ . Determinando las constantes  $c_1(\alpha)$  y  $c_2(\alpha)$  en la misma forma que en el ejemplo anterior, resulta, para cualquier  $\theta > 0$ ,

$$\mathcal{P}_\theta \left\{ c_1 < - \sum_{i=1}^n \log X_i + n \log \theta < c_2 \right\} = 1 - \alpha.$$

Luego

$$\left( e^{c_1/n} \prod_{i=1}^n x_i^{1/n}, e^{c_2/n} \prod_{i=1}^n x_i^{1/n} \right)$$

es un intervalo de confianza para  $\theta$ , de nivel de confianza  $1 - \alpha$ , basado en la media geométrica de las observaciones, en vez de en su máximo, como sucedía en el ejemplo 4.2.

Al contar con dos resultados basados en estadísticos diferentes, parece natural aplicar ambos, para una muestra concreta, y seleccionar el que produzca un intervalo de menor longitud. Sin embargo, esto no es correcto; si  $I_1$  e  $I_2$  son intervalos de extremos aleatorios, cada uno de los cuales contiene al parámetro con probabilidad  $1 - \alpha$ , el intervalo  $I^*$  más corto de ambos, coincide con  $I_1$  o con  $I_2$  para cada muestra, pero la probabilidad con que contiene al parámetro será, generalmente, inferior a  $1 - \alpha$ . El  $100\alpha\%$  de los casos en que  $I_1$  no contiene al parámetro se producirá, con mayor frecuencia, para las muestras que den una longitud pequeña de  $I_1$ ; lo mismo sucederá con  $I_2$  y la frecuencia con que el parámetro queda fuera de  $I^*$  será superior a  $\alpha$ . La manera más simple de hacerlo patente es examinar un caso concreto.

Imaginemos que se toma una muestra de tamaño 3 de una distribución uniforme en  $(\theta - 1/2, \theta + 1/2)$  y se ordena de menor a mayor; desde luego

$$\mathcal{P}_\theta \{ X_{(1)} < \theta < X_{(2)} \} = \frac{3}{8} = \mathcal{P}_\theta \{ X_{(2)} < \theta < X_{(3)} \},$$

de forma que  $(x_{(1)}, x_{(2)})$  y  $(x_{(2)}, x_{(3)})$  son intervalos de confianza para  $\theta$ , con nivel de confianza  $3/8$ . Sin embargo, es posible calcular<sup>2</sup> que

$$\mathcal{P}_\theta \{ X_{(1)} < \theta < X_{(2)} \mid X_{(2)} - X_{(1)} < X_{(3)} - X_{(2)} \} = \frac{1}{8},$$

---


$$\begin{aligned} {}^2 \quad & \mathcal{P}_\theta \{ X_{(1)} < \theta < X_{(2)}, X_{(2)} - X_{(1)} < X_{(3)} - X_{(2)} \} = \\ & = 3! \int_{\theta-1/2}^{\theta} \int_{\theta}^{\theta+1/4+x_1/2} \int_{2x_2-x_1}^{\theta+1/2} dx_3 dx_2 dx_1 = 1/16 \end{aligned}$$

puesto que el suceso se cumple si  $\theta - 1/2 < x_1 < \theta < x_2 < 2x_2 - x_1 < x_3 < \theta + 1/2$ . Además,  $\mathcal{P}_\theta \{ X_{(2)} - X_{(1)} < X_{(3)} - X_{(2)} \} = 1/2$ .

con lo cual el más corto de los intervalos  $(x_{(1)}, x_{(2)})$  y  $(x_{(2)}, x_{(3)})$  sólo tiene nivel de confianza  $1/8$ , y no  $3/8$  como cada uno de ellos.

No es lícito, por tanto, a la vista de los resultados sobre una misma muestra, seleccionar el más corto entre varios intervalos, de nivel de confianza  $1 - \alpha$ , basados en distintos estadísticos.

Lo que sí se puede intentar es comparar las longitudes esperadas, para ver si alguno de los procedimientos da, en media, mejores resultados que los demás para cualquier valor del parámetro. Así en el ejemplo 4.2, la longitud esperada del intervalo es

$$\left(\alpha^{-1/n} - 1\right) E[X_{(n)}] = \left(\alpha^{-1/n} - 1\right) \frac{n}{n+1} \theta,$$

que da  $0'207\theta$  para  $\alpha = 0'05$  y  $n = 15$ . Con las mismas especificaciones, las constantes del ejemplo 4.4 son  $e^{c_1/n} = 1'75$  y  $e^{c_2/n} = 4'787$ , de manera que la longitud esperada es

$$3'037E\left[\prod_{i=1}^n X_i^{1/n}\right] = 3'037\theta E[e^{-Y/n}] = 3'037\theta \left(1 + \frac{1}{n}\right)^{-n} = 1'154\theta,$$

puesto que  $Y$  tiene distribución  $\gamma(n, 1)$  de función característica  $(1 - it)^{-n}$ . En consecuencia, el intervalo del ejemplo 4.2 tendrá tendencia a ser más corto que el del ejemplo 4.4, cualquiera que sea  $\theta$ .

Más adelante, en relación con los contrastes de hipótesis, aparecerán otros criterios de comparación de diversos intervalos de confianza. También el estudio de los problemas de estimación puntual del parámetro servirá de orientación a la hora de seleccionar el estadístico mediante el cual realizar la estimación confidencial.

#### 4.4. Método de Neyman para la construcción de intervalos de confianza

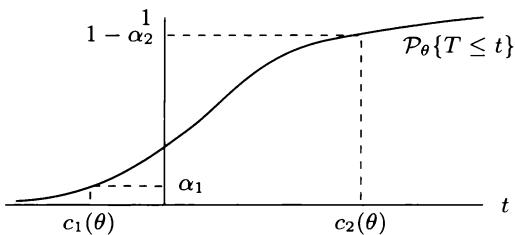
La construcción de intervalos de confianza no exige disponer de una función  $T(X_1, \dots, X_n; \theta)$  cuya distribución en el muestreo no dependa de  $\theta$ , sino que puede acometerse a partir de cualquier estadístico  $T$  unidimensional, mediante un procedimiento genérico conocido como *método de Neyman*.

De una población con distribución teórica perteneciente a una familia  $\mathcal{F} = \{F_\theta \mid \theta \in \Theta\}$ , siendo  $\Theta$  un intervalo de  $\mathbb{R}$ , se considera una muestra aleatoria  $(X_1, X_2, \dots, X_n)$ , con distribución  $\mathcal{P}_\theta$ . Sea  $T$  un estadístico unidimensional, cuyo recorrido se representará por  $L \subset \mathbb{R}$ .

Fijado un nivel de confianza  $1 - \alpha$ , para cada  $\theta \in \Theta$ , se pueden determinar dos valores  $c_1(\theta) < c_2(\theta)$  en  $L$ , tales que

$$\mathcal{P}_\theta\{T < c_1(\theta)\} \leq \alpha_1 \quad \text{y} \quad \mathcal{P}_\theta\{T > c_2(\theta)\} \leq \alpha_2,$$

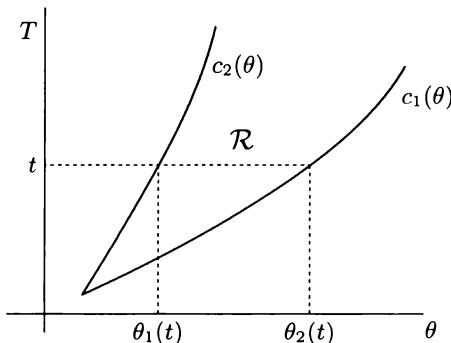
siendo  $\alpha_1, \alpha_2 > 0$  y  $\alpha_1 + \alpha_2 = \alpha$ .



Será entonces, para cada  $\theta \in \Theta$ ,

$$\mathcal{P}_\theta\{c_1(\theta) \leq T(X_1, X_2, \dots, X_n) \leq c_2(\theta)\} \geq 1 - \alpha_2 - \alpha_1 = 1 - \alpha. \quad (4.1)$$

Es decir que las desigualdades  $c_1(\theta) \leq T \leq c_2(\theta)$  determinan en el plano  $(\theta, T)$  una región  $\mathcal{R}$  cuyas secciones verticales, correspondientes a cada valor fijo de  $\theta$ , tienen probabilidad  $\mathcal{P}_\theta$  superior a  $1 - \alpha$ .



Habitualmente, para cada valor  $t$  del estadístico  $T$ , la sección horizontal de la región  $\mathcal{R}$ , por la recta  $T = t$ , será un intervalo limitado por dos abscisas  $\theta_1(t) < \theta_2(t)$ . En tal caso, se tendrá

$$\begin{aligned} & \{(x_1, x_2, \dots, x_n) \mid c_1(\theta) \leq T(x_1, x_2, \dots, x_n) \leq c_2(\theta)\} \\ &= \{(x_1, x_2, \dots, x_n) \mid (\theta, T(x_1, x_2, \dots, x_n)) \in \mathcal{R}\} \\ &= \{(x_1, x_2, \dots, x_n) \mid \theta_1(T(x_1, x_2, \dots, x_n)) \leq \theta \leq \theta_2(T(x_1, x_2, \dots, x_n))\}, \end{aligned}$$

de manera que para cada  $\theta \in \Theta$ , se cumple

$$\mathcal{P}_\theta \{ \theta_1(T(X_1, X_2, \dots, X_n)) \leq \theta \leq \theta_2(T(X_1, X_2, \dots, X_n)) \} \geq 1 - \alpha,$$

y por tanto

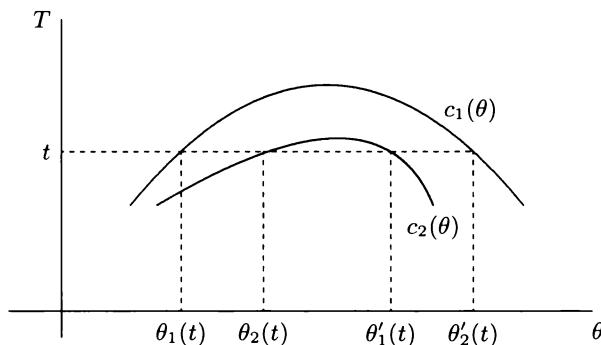
$$[ \theta_1(T(x_1, x_2, \dots, x_n)), \theta_2(T(x_1, x_2, \dots, x_n)) ]$$

es un intervalo de confianza para  $\theta$  de nivel de confianza  $1 - \alpha$ .

Acerca del método de construcción de intervalos de confianza que se acaba de describir cabe hacer dos observaciones importantes:

- a) Frecuentemente las funciones  $c_1(\theta)$  y  $c_2(\theta)$  son continuas y monótonas, en cuyo caso es  $\theta_1(t) = c_2^{-1}(t)$  y  $\theta_2(t) = c_1^{-1}(t)$ , para cada  $t \in L$ , supuesto que  $c_1$  y  $c_2$  sean crecientes, o bien  $\theta_1(t) = c_1^{-1}(t)$  y  $\theta_2(t) = c_2^{-1}(t)$  en el caso en que ambas fuesen decrecientes.

De todas maneras el método es también aplicable en el caso en que  $c_1$  y  $c_2$  no fuesen monótonas, aunque posiblemente se obtendría en este caso un conjunto de confianza compuesto por varios intervalos:



Situaciones de este tipo no son, sin embargo, habituales.

- b) Como ya se ha indicado, para no alargar innecesariamente el intervalo de confianza, es necesario conseguir que la desigualdad (4.1) sea lo más ajustada posible. En el caso de que la distribución del estadístico  $T$  sea continua, ello se logra escogiendo  $c_1(\theta)$  y  $c_2(\theta)$  de manera que

$$\mathcal{P}_\theta \{ T < c_1(\theta) \} = \alpha_1 \quad \text{y} \quad \mathcal{P}_\theta \{ T > c_2(\theta) \} = \alpha_2.$$

Sin embargo, si  $T$  tiene distribución discreta, ello no es factible en general y  $c_1(\theta), c_2(\theta)$  deben elegirse de manera que  $\alpha_1 - \mathcal{P}_\theta\{T < c_1(\theta)\}$  y  $\alpha_2 - \mathcal{P}_\theta\{T > c_2(\theta)\}$  sean positivos y tan pequeños como sea posible.

Se insistirá en ello, mediante un ejemplo, un poco más adelante. Previamente, para comparar el procedimiento anterior con el de la cantidad pivotal, se puede aplicar el método de Neyman al mismo caso del ejemplo 4.2.

### Ejemplo 4.5

Sea  $(X_1, X_2, \dots, X_n)$  una muestra aleatoria simple de una población con distribución teórica uniforme en  $(0, \theta)$ . Consideremos el estadístico

$$T(X_1, X_2, \dots, X_n) = \max_{1 \leq i \leq n} X_i = X_{(n)}$$

cuya distribución en el muestreo es

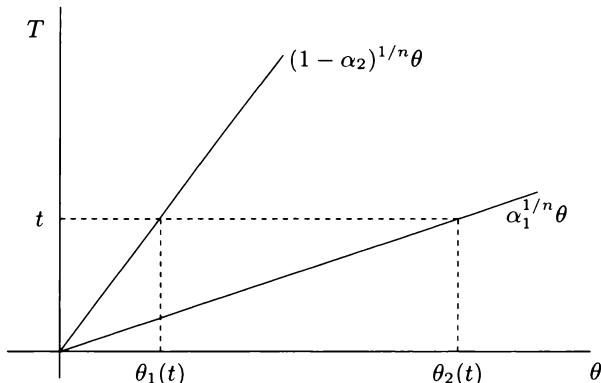
$$H_\theta(t) = \mathcal{P}_\theta \{X_{(n)} \leq t\} = \begin{cases} 0 & \text{si } t \leq 0 \\ (t/\theta)^n & \text{si } 0 < t < \theta \\ 1 & \text{si } t \geq \theta. \end{cases}$$

Fijados  $\alpha_1, \alpha_2 > 0$  tales que  $\alpha_1 + \alpha_2 = \alpha$ , para cada valor  $\theta > 0$ , las condiciones

$$H_\theta(c_1(\theta)) = \alpha_1 \quad \text{y} \quad H_\theta(c_2(\theta)) = 1 - \alpha_2$$

proporcionan los valores

$$c_1(\theta) = \alpha_1^{1/n} \theta \quad \text{y} \quad c_2(\theta) = (1 - \alpha_2)^{1/n} \theta.$$



Por tanto, será

$$\mathcal{P}_\theta \left\{ \alpha_1^{1/n} \theta < X_{(n)} < (1 - \alpha_2)^{1/n} \theta \right\} = 1 - \alpha$$

cualquiera que sea  $\theta > 0$ . Para cada valor  $t$  del estadístico, las ecuaciones

$$\alpha_1^{1/n} \theta = t \quad \text{y} \quad (1 - \alpha_2)^{1/n} \theta = t$$

tienen por soluciones

$$\theta_1(t) = (1 - \alpha_2)^{-1/n} t \quad \text{y} \quad \theta_2(t) = \alpha_1^{-1/n} t$$

Por consiguiente,

$$\mathcal{P}_\theta \left\{ (1 - \alpha_2)^{-1/n} X_{(n)} < \theta < \alpha_1^{-1/n} X_{(n)} \right\} = 1 - \alpha$$

para cada  $\theta > 0$  y resulta el mismo intervalo de confianza para  $\theta$  determinado en el ejemplo 4.2:

$$\left( (1 - \alpha_2)^{-1/n} x_{(n)}, \alpha_1^{-1/n} x_{(n)} \right).$$

### Ejemplo 4.6

Sea  $(X_1, X_2, \dots, X_n)$  una muestra aleatoria simple de una población cuya distribución teórica es

$$F_\theta(x) = \begin{cases} 0 & \text{si } x \leq \sqrt{\theta} \\ \frac{x^2 - \theta}{\theta^2 - \theta} & \text{si } \sqrt{\theta} < x < \theta \\ 1 & \text{si } x \geq \theta \end{cases}$$

siendo  $\theta$  un parámetro mayor que 1.

El estadístico

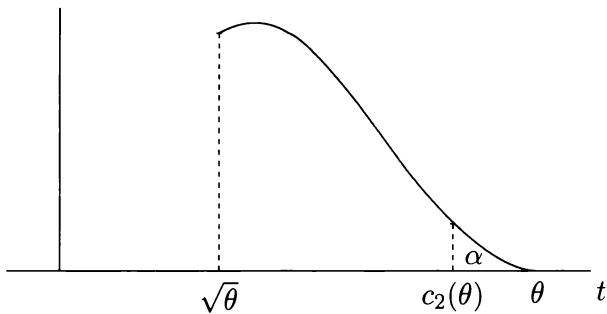
$$T(X_1, X_2, \dots, X_n) = \min_{1 \leq i \leq n} X_i = X_{(1)}$$

tiene distribución en el muestreo

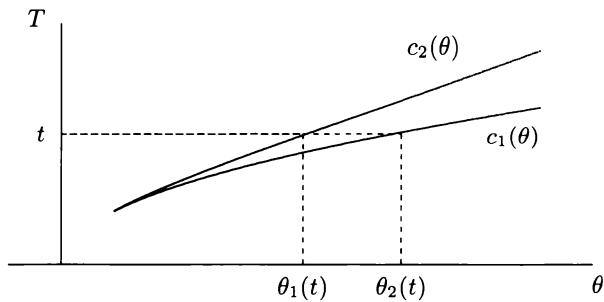
$$H_\theta(t) = \mathcal{P}_\theta \{ X_{(1)} \leq t \} = 1 - \left( 1 - \frac{t^2 - \theta}{\theta^2 - \theta} \right)^n = 1 - \left( \frac{\theta^2 - t^2}{\theta^2 - \theta} \right)^n$$

para  $\sqrt{\theta} < t < \theta$ .

La representación gráfica de su densidad:  $h_\theta(t) = \frac{2n t (\theta^2 - t^2)^{n-1}}{(\theta^2 - \theta)^n}$ ,



indica que, para obtener un intervalo de confianza, de longitud mínima, con nivel de confianza  $1 - \alpha$ , conviene tomar  $\alpha_1 = 0$  y  $\alpha_2 = \alpha$ . De forma que  $c_1(\theta) = \sqrt{\theta}$ . En cambio  $H_\theta(c_2(\theta)) = 1 - \alpha$ , da  $c_2(\theta) = \sqrt{\theta^2 - \alpha^{1/n}(\theta^2 - \theta)}$ .



Es decir, para cualquier  $\theta > 1$ ,

$$\mathcal{P}_\theta \left\{ \sqrt{\theta} < X_{(1)} < \sqrt{\theta^2 - \alpha^{1/n}(\theta^2 - \theta)} \right\} = 1 - \alpha.$$

Fijado un valor  $t$  del estadístico  $X_{(1)}$ , de la ecuación  $\sqrt{\theta} = t$  se obtiene  $\theta_2(t) = t^2$ ; mientras que  $\theta^2(1 - \alpha^{1/n}) + \theta\alpha^{1/n} = t^2$  tiene como raíz positiva

$$\theta_1(t) = \frac{\sqrt{\alpha^{2/n} + 4(1 - \alpha^{1/n})t^2} - \alpha^{1/n}}{2(1 - \alpha^{1/n})}.$$

Se obtiene pues como intervalo de confianza para  $\theta$ , de nivel de confianza  $1 - \alpha$ ,

$$\left( \frac{\sqrt{\alpha^{2/n} + 4(1 - \alpha^{1/n})x_{(1)}^2} - \alpha^{1/n}}{2(1 - \alpha^{1/n})}, x_{(1)}^2 \right).$$

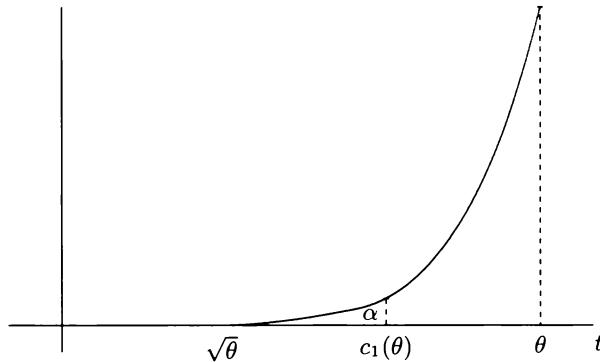
La construcción puede repetirse con el estadístico

$$T'(X_1, X_2, \dots, X_n) = \max_{1 \leq i \leq n} X_i = X_{(n)}$$

con distribución en el muestreo

$$H'_\theta(t) = \mathcal{P}_\theta \{ X_{(n)} \leq t \} = \left( \frac{t^2 - \theta}{\theta^2 - \theta} \right)^n \quad \text{para } \sqrt{\theta} < t < \theta.$$

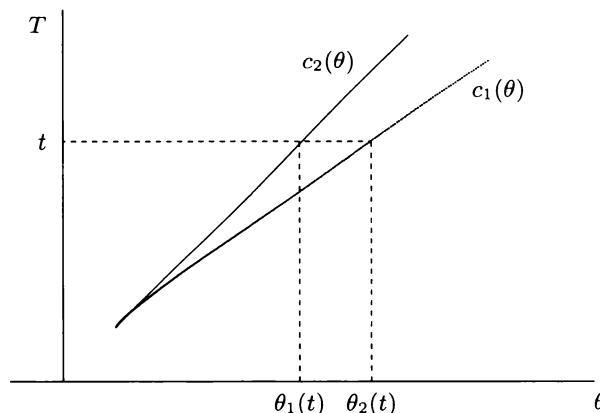
Como la densidad es ahora  $h'_\theta(t) = \frac{2n t (t^2 - \theta)^{n-1}}{(\theta^2 - \theta)^n}$ ,



para obtener un intervalo de confianza de nivel de confianza  $1 - \alpha$ , lo más corto posible, debe tomarse  $\alpha_1 = \alpha$ , y  $\alpha_2 = 0$ .

Es decir  $c_2(\theta) = \theta$ ; mientras que, a fin de que sea  $H'_\theta(c_1(\theta)) = \alpha$ , debe ser

$$c_1(\theta) = \sqrt{\theta + \alpha^{1/n}(\theta^2 - \theta)}$$



siendo, por tanto, para cualquier  $\theta > 1$ ,

$$\mathcal{P}_\theta \left\{ \sqrt{\theta + \alpha^{1/n}(\theta^2 - \theta)} < X_{(n)} < \theta \right\} = 1 - \alpha.$$

Fijado el valor  $t$  del estadístico  $X_{(n)}$ , la ecuación  $c_2(\theta) = t$  proporciona  $\theta_1(t) = t$ ; y de la ecuación  $c_1(\theta) = t$  resulta

$$\theta_2(t) = \frac{\sqrt{(1 - \alpha^{1/n})^2 + 4\alpha^{1/n}t^2} - (1 - \alpha^{1/n})}{2\alpha^{1/n}}.$$

De forma que el intervalo de confianza para  $\theta$  que se obtiene es

$$\left( x_{(n)}, \frac{\sqrt{(1 - \alpha^{1/n})^2 + 4\alpha^{1/n}x_{(n)}^2} - (1 - \alpha^{1/n})}{2\alpha^{1/n}} \right).$$

Nuevamente, se ve que la construcción de intervalos de confianza puede conducir a múltiples resultados, según el estadístico que se utilice; y la comparación de las longitudes esperadas no es, en este caso, nada simple.

Aunque conceptualmente el caso en que la distribución en el muestreo del estadístico  $T$  sea discreta no presente ninguna novedad, las dificultades técnicas que aparecen son entonces considerablemente mayores. A esta situación se dedicará el siguiente ejemplo.

### Ejemplo 4.7

Se desea construir un intervalo de confianza, de nivel de confianza  $1 - \alpha = 0'95$ , para el grado medio de radiactividad,  $\lambda$ , de una sustancia que emite partículas alfa, de forma que el número  $X$  de partículas emitidas en un minuto sigue una distribución de Poisson de parámetro  $\lambda$ . Para ello se observa el número de partículas alfa emitidas durante 10 minutos, obteniéndose una muestra aleatoria simple  $X_1, \dots, X_{10}$ .

Puesto que  $\lambda$  representa el número medio de emisiones por minuto, es razonable utilizar el estadístico

$$T = \sum_{i=1}^{10} X_i$$

directamente relacionado con la media muestral. La distribución en el muestreo de  $T$  es una distribución de Poisson de parámetro  $10\lambda$ , y por comodidad determinaremos el intervalo de colas iguales, es decir con  $\alpha_1 = \alpha_2 = 0'025$ .

Para hallar, en primer lugar, la función  $c_1(\lambda)$ , debemos buscar un valor tal que

$$\mathcal{P}_\lambda\{T < c_1(\lambda)\} \text{ sea inferior y lo más próximo posible a } 0'025.$$

La tabla de la distribución de Poisson sólo proporciona los valores de la función de probabilidad para determinados valores del parámetro, de forma que, con su ayuda, sólo podemos determinar  $c_1(\lambda)$  para algunos valores de  $\lambda$ . Por ejemplo, para  $\lambda = 0'5$ , como  $T$  tiene distribución de Poisson de parámetro 5, se obtiene

$$\mathcal{P}_\lambda\{T < 1\} = 0'0067, \quad \mathcal{P}_\lambda\{T < 2\} = 0'0404$$

luego  $c_1(0'5)$  puede ser cualquier valor en el intervalo  $(0, 1]$ ; por ejemplo puede tomarse  $c_1(0'5) = 0'5$ .

Procediendo de manera similar, con cada valor de  $\lambda$  de los que figura en la primera línea de la siguiente tabla, se obtienen los correspondientes de  $c_1(\lambda)$  que figuran en la segunda línea:

0'01	0'02	0'03	0'04	0'05	0'06	0'07	0'08	0'09	0'1	0'11	0'12
-0'5	-0'5	-0'5	-0'5	-0'5	-0'5	-0'5	-0'5	-0'5	-0'5	-0'5	-0'5
0'13	0'14	0'15	0'16	0'17	0'18	0'19	0'2	0'22	0'24	0'26	0'28
-0'5	-0'5	-0'5	-0'5	-0'5	-0'5	-0'5	-0'5	-0'5	-0'5	-0'5	-0'5
0'3	0'32	0'34	0'36	0'38	0'4	0'5	0'6	0'7	0'8	0'9	1
-0'5	-0'5	-0'5	-0'5	0'5	0'5	0'5	1'5	1'5	2'5	3'5	3'5

Por ejemplo, a partir de  $\lambda = 0'38$ , la distribución de Poisson de parámetro  $10\lambda$  asigna a  $\{0\}$  probabilidad inferior a  $0'025$ ; a partir de  $\lambda = 0'6$ , la distribución de Poisson de parámetro  $10\lambda$  asigna a  $\{0, 1\}$  probabilidad inferior a  $0'025$ , etc.

Análogamente, la determinación de  $c_2(\lambda)$  ha de realizarse intentando que

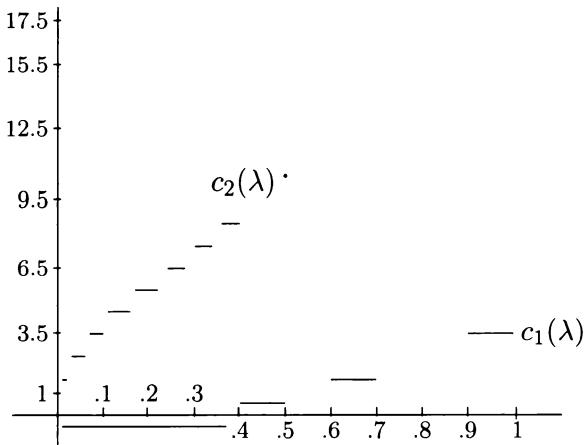
$$\mathcal{P}_\lambda\{T > c_2(\lambda)\} \text{ sea inferior y lo más próximo posible a } 0'025.$$

Si por ejemplo  $\lambda = 0'15$ , será  $\mathcal{P}_\lambda\{T > 4\} = 0'0185$  y  $\mathcal{P}_\lambda\{T > 3\} = 0'0656$ , luego  $c_2(0'15)$  debe estar comprendido en el intervalo  $[4, 5)$ ; digamos  $c_2(1'5) = 4'5$ .

Para los valores de  $\lambda$  tales que  $10\lambda$  figura en la tabla de la distribución de Poisson, se obtienen así los siguientes valores de  $c_2(\lambda)$ :

0'01	0'02	0'03	0'04	0'05	0'06	0'07	0'08	0'09	0'1	0'11	0'12
1'5	1'5	2'5	2'5	2'5	2'5	3'5	3'5	3'5	3'5	4'5	4'5
0'13	0'14	0'15	0'16	0'17	0'18	0'19	0'2	0'22	0'24	0'26	0'28
4'5	4'5	4'5	4'5	5'5	5'5	5'5	5'5	5'5	6'5	6'5	6'5
0'3	0'32	0'34	0'36	0'38	0'4	0'5	0'6	0'7	0'8	0'9	1
7'5	7'5	7'5	8'5	8'5	8'5	10'5	11'5	13'5	14'5	15'5	17'5

La representación gráfica de las funciones  $c_1$  y  $c_2$ , cuyos valores se han tabulado es la siguiente



Obsérvese que, como sólo se ha utilizado la función de probabilidad de la distribución de Poisson para determinados valores del parámetro, las funciones  $c_1$  y  $c_2$  quedan sin precisar para algunos valores de  $\lambda$ . Este problema técnico no es despreciable, puesto que el siguiente paso del método de Neyman consiste en interseccar por la recta de ordenada constante, igual al valor entero observado para el estadístico  $T$ . Dicha recta horizontal cortará a las gráficas de  $c_1$  y  $c_2$  en alguno de los escalones verticales que separan los tramos horizontales de cada una de las gráficas, de manera que interesa localizar con más exactitud los puntos en que se sitúan los saltos de las funciones  $c_1$  y  $c_2$ .

Caben entonces dos soluciones. Si no se desea mucha precisión y el valor del estadístico cae en la zona baja, donde  $c_1$  y  $c_2$  son algo más precisas, se puede prolongar los tramos horizontales de  $c_1$  hacia la derecha y los de  $c_2$

hacia la izquierda (ampliando el intervalo de confianza para no incumplir la especificación relativa al nivel de confianza), intersecando después con la recta correspondiente al valor del estadístico. Así, si por ejemplo el valor observado de  $T$  es 2, ello conduciría a dar como intervalo de confianza, de nivel de confianza 0'95, el intervalo

$$[ 0'02 , 0'8 ]$$

ya que 0'02 es el último valor de  $\lambda$  para el que consta que  $c_2$  es menor que 2 y 0'8 es el primer valor de  $\lambda$  para el que se sabe que  $c_1$  es mayor que 2.

Si se desea mayor exactitud o el valor del estadístico está en una zona en que la precisión de  $c_1$  y  $c_2$  es inaceptable, habrá que recurrir al cálculo directo para suplir las deficiencias de la tabla. Por ejemplo, si el valor observado para  $T$  es 14, una calculadora o un sencillo programa de ordenador proporcionan

$\lambda$	$P_\lambda\{T > 13\} = 1 - \sum_{k=0}^{13} e^{-10\lambda}(10\lambda)^k/k!$
0'764	0'0246734
0'765	0'0249075
0'7653	0'0249781
0'7654	0'0250016
0'766	0'0251433
0'767	0'0253807

Así que debe ser  $c_2(0'765) = 13'5$ ,  $c_2(0'7653) = 13'5$ ,  $c_2(0'7654) = 14'5$ ,  $c_2(0'766) = 14'5$ . Es decir que el salto de 13'5 a 14'5 en la función  $c_2$  se produce en el intervalo (0'765, 0'766) o, con más exactitud, en el intervalo (0'7653, 0'7654). El extremo inferior del intervalo de confianza buscado puede ser por tanto 0'765 ó 0'7653, según la precisión deseada — o, con toda exactitud,  $\sup\{\lambda \mid P_\lambda\{T \geq 14\} < 0'025\}$ .

Análogamente, un procedimiento similar da

$\lambda$	$P_\lambda\{T < 15\} = \sum_{k=0}^{14} e^{-10\lambda}(10\lambda)^k/k!$
2'347	0'0252212
2'348	0'0251083
2'3489	0'0250070
2'349	0'0249957
2'35	0'0248836

probando que  $c_1(2'348) = 13'5$ ,  $c_1(2'3489) = 13'5$ ,  $c_1(2'349) = 14'5$ . El salto de la función  $c_1$  de  $13'5$  a  $14'5$  se produce pues en el intervalo  $(2'3489, 2'349)$  y el extremo superior del intervalo buscado será  $2'349$  —o, con toda exactitud,  $\inf\{\lambda \mid \mathcal{P}_\lambda\{T \leq 14\} < 0'025\}$ .

De todas formas, dadas las características del problema, no tiene mucho sentido precisar hasta las diezmilésimas, los extremos de un intervalo de variación del parámetro que engloba desde  $0'76 \dots$  hasta  $2'35 \dots$ , pudiendo proponerse en definitiva

$$[0'765, 2'349]$$

como intervalo de confianza más razonable, al nivel de confianza  $0'95$ , para el nivel medio de emisiones por minuto.

El contexto del problema permite hacer, no obstante, una sugerencia interesante. Puesto que un nivel medio de radioactividad bajo no es alarmante, el interés debe centrarse en la cota superior de la radioactividad media. Es decir, que en lugar del intervalo de colas iguales que se ha considerado, o el intervalo de longitud mínima que podría tratar de determinarse, lo más razonable en esta situación es buscar un intervalo de confianza de la forma  $[0, \lambda_1]$  que asegure, con un nivel de confianza de  $0'95$ , que el valor medio de la radioactividad no sobrepasa  $\lambda_1$ . Ello significa trasladar toda la probabilidad de error permitida a  $\alpha_1$ , buscando solamente una función  $c'_1(\lambda)$  tal que

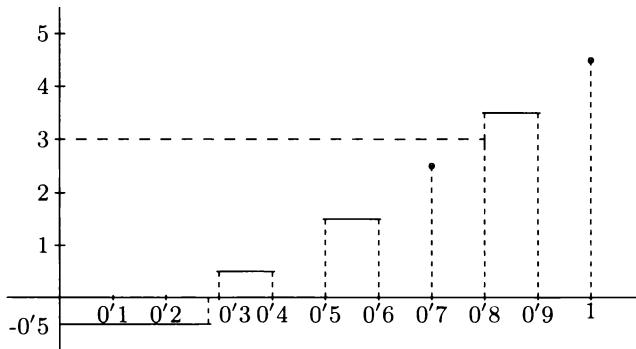
$$\mathcal{P}_\lambda\{T < c'_1(\lambda)\} \quad \text{sea inferior y lo más próximo posible a } 0'05.$$

La tabla de la distribución de Poisson proporciona nuevamente

$\lambda$	$0'01$	$\dots$	$0'28$	$0'3$	$\dots$	$0'4$	$0'5$	$0'6$	$0'7$	$0'8$	$0'9$	$1$
$c'_1(\lambda)$	$-0'5$	$\dots$	$-0'5$	$0'5$	$\dots$	$0'5$	$1'5$	$1'5$	$2'5$	$3'5$	$3'5$	$4'5$

ya que para  $\lambda \geq 0'3$  la distribución de Poisson de parámetro  $10\lambda$  asigna a  $\{0\}$  probabilidad inferior a  $0'05$ ; para  $\lambda \geq 0'5$ , asigna a  $\{0, 1\}$  probabilidad inferior a  $0'05$ , etc.

Mientras que es  $c'_2(\lambda) = \infty$ , para que  $\mathcal{P}_\lambda\{T > c'_2(\lambda)\} = 0$ , la gráfica de la función  $c'_1(\lambda)$  es de la forma



Si por ejemplo el valor observado de  $T$  es 3, en primera aproximación es  $\lambda \in (0, 0'8]$ , con nivel de confianza  $0'95$ . Si se desea más precisión como

$\lambda$	$P_\lambda\{T \leq 3\}$
0'77	0'0518188
0'775	0'0501221
0'776	0'0497889
0'78	0'0484766

es  $\lambda \in (0, 0'776]$  con nivel de confianza  $0'95$ .

Análogamente, si el valor observado de  $T$  es 37, como

$\lambda$	$P_\lambda\{T \leq 37\}$
4'8	0'0603992
4'86	0'0510813
4'87	0'0496524
4'9	0'0455681

es  $\lambda \in (0, 4'87]$  con nivel de confianza  $0'95$ .

Esta última manera de proceder, dejando toda la probabilidad  $\alpha$  en un extremo de la distribución, responde al concepto genérico de *cota de confianza*:

Si  $X$  es una variable aleatoria, cuya distribución teórica pertenece a una familia paramétrica  $\mathcal{F} = \{F_\theta \mid \theta \in \Theta \subset \mathbb{R}\}$ , de la cual se dispone de una muestra aleatoria  $(X_1, \dots, X_n)$  con distribución  $\mathcal{P}_\theta$ , se denomina

- *cota inferior de confianza* para  $\theta$ , de nivel de confianza  $1 - \alpha$ , al valor observado de cualquier estadístico  $T_1$  tal que

$$\mathcal{P}_\theta\{T_1(X_1, \dots, X_n) \leq \theta\} \geq 1 - \alpha \quad \text{para cada } \theta \in \Theta.$$

- *cota superior de confianza* para  $\theta$ , de nivel de confianza  $1 - \alpha$ , al valor observado de cualquier estadístico  $T_2$  tal que

$$\mathcal{P}_\theta\{\theta \leq T_2(X_1, \dots, X_n)\} \geq 1 - \alpha \quad \text{para cada } \theta \in \Theta.$$

Naturalmente, si la continuidad de las distribuciones de  $T_1$  y  $T_2$  lo permite, las desigualdades anteriores deben sustituirse por igualdades. En caso contrario, las probabilidades deben sobreponerse lo menos posible a  $1 - \alpha$ .

## 4.5. Intervalos de confianza para los parámetros de distribuciones normales

Mediante los resultados del último capítulo, acerca de las distribuciones en el muestreo asociadas a poblaciones normales, la construcción de intervalos de confianza para los parámetros poblacionales, bajo hipótesis de normalidad, es un simple ejercicio. Su gran importancia práctica, aconseja, sin embargo, expresarlos explícitamente.

Adoptaremos para ello la siguiente notación estándar: Si  $Z$ ,  $Y$ ,  $t$ , y  $U$  son variables aleatorias que tienen respectivamente distribución  $N(0, 1)$ ,  $\chi^2_n$ ,  $t_n$  y  $F_{n,m}$ , se designará por  $z_\alpha$ ,  $\chi^2_{n;\alpha}$ ,  $t_{n;\alpha}$  y  $F_{n,m;\alpha}$  los valores reales tales que

$$P\{Z > z_\alpha\} = \alpha, \quad P\{t > t_{n;\alpha}\} = \alpha,$$

$$P\{Y > \chi^2_{n;\alpha}\} = \alpha, \quad P\{U > F_{n,m;\alpha}\} = \alpha.$$

En primer lugar consideraremos una única población,  $N(\mu, \sigma)$ , de la cual se dispone de una muestra aleatoria simple  $(X_1, X_2, \dots, X_n)$ .

- A) *Intervalo de confianza para la media si la varianza poblacional es conocida.*

Algunos casos particulares de esta situación han sido considerados en el ejemplo 4.1; de manera general, puesto que

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad \text{tiene distribución en el muestreo } N(0, 1)$$

será

$$\mathcal{P}_\mu \left\{ -z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2} \right\} = 1 - \alpha;$$

de forma que

$$\left( \bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

es un intervalo de confianza para  $\mu$  de nivel de confianza  $1 - \alpha$ .

- B) *Intervalo de confianza para la media si la varianza poblacional es desconocida.*

Recalcando las consideraciones realizadas al introducir el estadístico de Student, puede observarse que el intervalo anterior tiene extremos indeterminados si  $\sigma$  es desconocida. Convendrá utilizar entonces que

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \quad \text{tiene distribución en el muestreo } t_{n-1}$$

para obtener

$$\mathcal{P}_{\mu,\sigma} \left\{ -t_{n-1;\alpha/2} < \frac{\bar{X} - \mu}{S/\sqrt{n}} < t_{n-1;\alpha/2} \right\} = 1 - \alpha;$$

lo cual proporciona

$$\left( \bar{x} - t_{n-1;\alpha/2} \frac{S}{\sqrt{n}}, \bar{x} + t_{n-1;\alpha/2} \frac{S}{\sqrt{n}} \right)$$

como intervalo de confianza para  $\mu$  de nivel de confianza  $1 - \alpha$ .

- C) *Intervalo de confianza para  $\sigma^2$  si la media poblacional es conocida.*

A pesar de que esta situación es poco frecuente en la práctica, la solución teórica es inmediata:

$$\sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2} \quad \text{tiene distribución en el muestreo } \chi_n^2,$$

de forma que

$$\mathcal{P}_\sigma \left\{ \chi_{n;1-\alpha/2}^2 < \sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2} < \chi_{n;\alpha/2}^2 \right\} = 1 - \alpha$$

y por tanto

$$\left( \frac{1}{\chi_{n;\alpha/2}^2} \sum_{i=1}^n (x_i - \mu)^2, \frac{1}{\chi_{n;1-\alpha/2}^2} \sum_{i=1}^n (x_i - \mu)^2 \right)$$

es un intervalo de confianza para  $\sigma^2$ , de nivel de confianza  $1 - \alpha$ .

La asimetría de la distribución  $\chi^2$  hace que el intervalo anterior, obtenido eliminando colas iguales a  $\alpha/2$  en los dos extremos de la distribución, no sea de longitud mínima. Sin embargo, la diferencia no es lo suficientemente importante (sobre todo si  $n$  no es muy pequeño) como para tomarse la molestia de buscar en la tabla los valores que minimizan la longitud.

D) *Intervalo de confianza para  $\sigma^2$ , si la media poblacional es desconocida.*

Esta situación, de mucha más utilidad que la anterior, se resuelve mediante la afirmación del teorema de Fisher:

$$\sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \quad \text{tiene distribución } \chi_{n-1}^2.$$

Según ello,

$$\mathcal{P}_{\mu,\sigma} \left\{ \chi_{n-1;1-\alpha/2}^2 < \frac{(n-1)S^2}{\sigma^2} < \chi_{n-1;\alpha/2}^2 \right\} = 1 - \alpha$$

y resulta

$$\left( \frac{(n-1)S^2}{\chi_{n-1;\alpha/2}^2}, \frac{(n-1)S^2}{\chi_{n-1;1-\alpha/2}^2} \right)$$

como intervalo de confianza para  $\sigma^2$  de nivel de confianza  $1 - \alpha$  (con la misma consideración acerca de la longitud mínima que en el caso anterior).

E) *Región de confianza para la media y la varianza poblacional.*

El teorema de Fisher, al proporcionar la distribución conjunta de los estadísticos

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad \text{y} \quad \frac{(n-1)S^2}{\sigma^2}$$

permite construir una región de confianza simultáneamente para  $\mu$  y  $\sigma^2$ . En efecto, como ambos estadísticos son independientes, y

$$\mathcal{P}_{\mu,\sigma} \left\{ -z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2} \right\} = 1 - \alpha$$

$$\mathcal{P}_{\mu,\sigma} \left\{ \chi^2_{n-1;1-\beta/2} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{n-1;\beta/2} \right\} = 1 - \beta$$

la región del espacio muestral

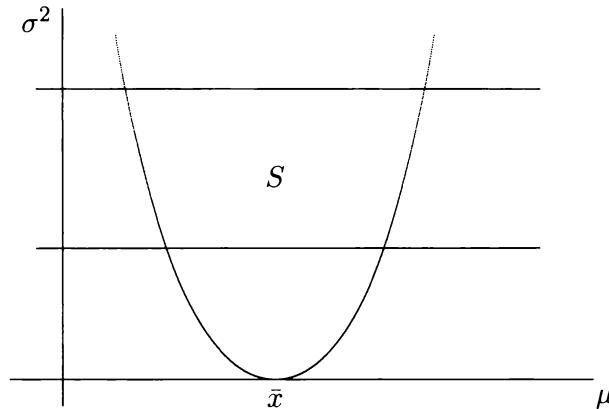
$$\left\{ -z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}, \quad \chi^2_{n-1;1-\beta/2} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{n-1;\beta/2} \right\}$$

tiene probabilidad  $(1 - \alpha)(1 - \beta)$ . Las dos desigualdades definen en el plano  $(\mu, \sigma^2)$  una región,  $S$ , limitada por la parábola

$$\frac{\sigma^2}{n} z_{\alpha/2}^2 = (\bar{X} - \mu)^2$$

y por las rectas

$$\sigma^2 = \frac{(n-1)S^2}{\chi^2_{n-1;\beta/2}}, \quad \sigma^2 = \frac{(n-1)S^2}{\chi^2_{n-1;1-\beta/2}}.$$



Entonces,  $S$  es una región de confianza para el parámetro bidimensional  $(\mu, \sigma^2)$ , de nivel de confianza  $(1 - \alpha)(1 - \beta)$ , pudiendo elegirse  $\alpha$  y  $\beta$  para obtener cualquier nivel de confianza  $(1 - \gamma)$  prefijado. Las diversas maneras de hacerlo corresponden a distintas regiones de confianza, con el mismo nivel de confianza.

Consideremos ahora la situación en la cual se dispone de dos muestras aleatorias simples  $(X_1, \dots, X_n)$  e  $(Y_1, \dots, Y_m)$  de dos poblaciones  $N(\mu_1, \sigma_1)$  y  $N(\mu_2, \sigma_2)$ , independientes.

- F) *Intervalo de confianza para la diferencia de medias con  $\sigma_1$  y  $\sigma_2$  conocidas.*

Como

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n + \sigma_2^2/m}}$$

tiene distribución en el muestreo  $N(0, 1)$ , resulta directamente que

$$\left( \bar{x} - \bar{y} - z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}, \quad \bar{x} - \bar{y} + z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} \right)$$

es un intervalo de confianza para  $\mu_1 - \mu_2$ , de nivel de confianza  $1 - \alpha$ .

Análogo resultado se obtiene, reemplazando  $\sigma_1^2$  y  $\sigma_2^2$  por  $S_1^2$  y  $S_2^2$ , en el caso en que las varianzas poblacionales sean desconocidas pero los tamaños de ambas muestras sean lo suficientemente grandes ( $n, m \geq 15$ ).

- G) *Intervalo de confianza para la diferencia de medias con varianzas desconocidas pero iguales.*

Si  $\sigma_1$  y  $\sigma_2$  tienen un valor común desconocido, el estadístico

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{ns_1^2 + ms_2^2}{n+m-2} \sqrt{\frac{1}{n} + \frac{1}{m}}}}$$

tiene distribución en el muestreo  $t_{n+m-2}$ .

De manera que el intervalo de extremos

$$\bar{x} - \bar{y} \mp t_{n+m-2; \alpha/2} \sqrt{\frac{ns_1^2 + ms_2^2}{n+m-2} \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

es un intervalo de confianza para  $\mu_1 - \mu_2$  de nivel de confianza  $1 - \alpha$ .

Análogo resultado se obtiene, con el estadístico

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{S_1^2/n + S_2^2/m}}$$

utilizando la aproximación de Welch, en el caso en que  $\sigma_1^2$  y  $\sigma_2^2$  no pudiesen suponerse iguales y alguna de las muestras fuese de pequeño tamaño. El intervalo de confianza de nivel  $1 - \alpha$  tendría entonces por extremos

$$\bar{x} - \bar{y} \mp t_{f;\alpha/2} \sqrt{S_1^2/n + S_2^2/m}$$

siendo  $f$  el valor indicado en la página 74.

H) *Intervalo de confianza para el cociente de varianzas poblacionales.*

Puesto que

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \quad \text{tiene distribución } F_{n-1,m-1},$$

será

$$\mathcal{P} \left\{ F_{n-1,m-1;1-\alpha/2} < \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} < F_{n-1,m-1;\alpha/2} \right\} = 1 - \alpha$$

y por tanto

$$\left( \frac{S_1^2/S_2^2}{F_{n-1,m-1;\alpha/2}}, \frac{S_1^2/S_2^2}{F_{n-1,m-1;1-\alpha/2}} \right)$$

es un intervalo de confianza para  $\sigma_1^2/\sigma_2^2$ , de nivel de confianza  $1 - \alpha$ .

I) *Intervalo de confianza para la diferencia de medias de dos poblaciones normales no independientes.*

En el caso en que  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$  sea una muestra aleatoria simple (de datos apareados) de una población normal bidimensional de medias  $(\mu_1, \mu_2)$  y covarianza  $\sigma_{11} \neq 0$ , la construcción de intervalos de confianza para  $\mu_1 - \mu_2$  se basa en que

$$\sqrt{n-1} \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{s^*}$$

tiene distribución en el muestreo  $t_{n-1}$  (véase el final de la sección 3.4).

Concretamente

$$\left( \bar{x} - \bar{y} - t_{n-1;\alpha/2} \frac{s^*}{\sqrt{n-1}}, \bar{x} - \bar{y} + t_{n-1;\alpha/2} \frac{s^*}{\sqrt{n-1}} \right)$$

es un intervalo de confianza para  $\mu_1 - \mu_2$ , de nivel de confianza  $1 - \alpha$ , donde  $s^*$ <sup>2</sup> representa la varianza muestral de las diferencias  $X_i - Y_i$ . Supuesto, naturalmente, que alguno de los elementos de la matriz de covarianzas poblacional es desconocido, ya que si no

$$\left( \bar{x} - \bar{y} - z_{\alpha/2} \sqrt{\frac{\sigma_1^2 + \sigma_2^2 - 2\sigma_{11}}{n}}, \bar{x} - \bar{y} + z_{\alpha/2} \sqrt{\frac{\sigma_1^2 + \sigma_2^2 - 2\sigma_{11}}{n}} \right)$$

es un intervalo de confianza más preciso, de nivel de confianza  $1 - \alpha$ , que resulta de ser  $N(0, 1)$  la distribución en el muestreo de

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\sigma_{11}}} \sqrt{n}.$$

## 4.6. Intervalos de confianza basados en distribuciones asintóticas

A lo largo del capítulo 2 se ha puesto de relieve que es a menudo difícil conocer la distribución en el muestreo de determinados estadísticos y que, en cambio, se puede conocer su distribución asintótica. Como ocurre con los cuantiles y los momentos muestrales (véase secciones 2.5 y 2.6), frecuentemente es posible disponer de una sucesión  $T_n$  de estadísticos, correspondientes a sucesivos tamaños muestrales  $n$ , tales que

$$\frac{T_n - \theta}{\sigma_n(\theta)} \xrightarrow{d} N(0, 1)$$

donde  $\theta$  representa el parámetro que caracteriza la distribución teórica y  $\sigma_n(\theta)$  depende en general de  $n$  y del parámetro poblacional.

Esta situación puede ser utilizada para obtener intervalos de confianza aproximados para el parámetro  $\theta$ . De hecho, si  $n$  es suficientemente grande, será, para cada  $\theta \in \Theta$ ,

$$\mathcal{P}_\theta \left\{ -z_{\alpha/2} < \frac{T_n - \theta}{\sigma_n(\theta)} < z_{\alpha/2} \right\} \simeq 1 - \alpha$$

de manera que si puede invertirse la desigualdad, despejando  $\theta$ , se obtendrá un intervalo de confianza para  $\theta$ , de nivel de confianza aproximado  $1 - \alpha$ .

**Ejemplo 4.8**

Como aplicación, consideremos la misma situación del ejemplo 4.7, pero observando ahora el número de emisiones  $X_1, X_2, \dots, X_n$  durante un número grande de minutos. En tal caso, según el teorema central del límite, el estadístico  $T = \sum_{i=1}^n X_i$  es tal que

$$\frac{T - n\lambda}{\sqrt{n\lambda}} = \frac{\bar{X} - \lambda}{\sqrt{\lambda/n}} \quad \text{es aproximadamente } N(0, 1).$$

Por tanto, para cualquier  $\lambda > 0$ ,

$$\mathcal{P}_\lambda \left\{ -z_{\alpha/2} < \sqrt{n} \frac{\bar{X} - \lambda}{\sqrt{\lambda}} < -z_{\alpha/2} \right\} \simeq 1 - \alpha.$$

La desigualdad se cumple cuando  $\lambda$  verifica

$$\lambda^2 - \left( 2\bar{x} + \frac{z_{\alpha/2}^2}{n} \right) \lambda + \bar{x}^2 < 0$$

es decir, cuando

$$\bar{x} + \frac{z_{\alpha/2}^2}{2n} - \frac{z_{\alpha/2}}{2} \sqrt{\frac{z_{\alpha/2}^2}{n^2} + \frac{4\bar{x}}{n}} < \lambda < \bar{x} + \frac{z_{\alpha/2}^2}{2n} + \frac{z_{\alpha/2}}{2} \sqrt{\frac{z_{\alpha/2}^2}{n^2} + \frac{4\bar{x}}{n}}$$

que es aproximadamente un intervalo de confianza para  $\lambda$ , de nivel de confianza  $1 - \alpha$ .

Puesto que  $n$  es grande y el intervalo es sólo aproximado, se pueden despreciar los términos de orden  $1/n$ , obteniendo en definitiva como intervalo

$$\bar{x} - z_{\alpha/2} \sqrt{\bar{x}/n} < \lambda < \bar{x} + z_{\alpha/2} \sqrt{\bar{x}/n};$$

que es el mismo resultado al que se llega si en la desigualdad

$$-z_{\alpha/2} < \sqrt{n} \frac{\bar{x} - \lambda}{\sqrt{\lambda}} < z_{\alpha/2}$$

se reemplaza, en el denominador, el valor de la media poblacional  $\lambda$  por la media muestral  $\bar{x}$  y se despeja el valor  $\lambda$  del numerador.

Por ejemplo, si  $n = 400$  y  $T = 960$ , el intervalo de nivel de confianza 0'95, obtenido con cada una de las expresiones (con  $z_{\alpha/2} = 1'96$  y  $\bar{x} = 2'4$ ) es

$$[ 2'253, 2'557 ] \quad \text{y} \quad [ 2'248, 2'552 ].$$

### Ejemplo 4.9

Tras realizar un gran número,  $n$ , de lanzamientos con una moneda, se desea construir un intervalo de confianza para la probabilidad  $p$  de cara. Para ello se considera el estadístico  $T$ , número total de caras obtenidas, o bien  $\bar{X} = T/n$ , número medio de caras por lanzamiento. Evidentemente la distribución en el muestreo de  $T$  es binomial  $B(n, p)$  y, según el teorema central del límite,

$$\frac{T - np}{\sqrt{np(1-p)}} = \frac{\bar{X} - p}{\sqrt{p(1-p)/n}} \quad \text{es aproximadamente } N(0, 1);$$

de manera que, para cualquier  $p \in (0, 1)$ ,

$$\mathcal{P}_p \left\{ -z_{\alpha/2} < \sqrt{n} \frac{\bar{X} - p}{\sqrt{p(1-p)}} < z_{\alpha/2} \right\} \simeq 1 - \alpha.$$

La desigualdad anterior se cumple cuando  $p$  está entre las raíces de la ecuación

$$(n + z_{\alpha/2}^2)p^2 - (2n\bar{x} + z_{\alpha/2}^2)p + n\bar{x}^2 = 0$$

es decir, cuando

$$\frac{2n\bar{x} + z^2 - z\sqrt{z^2 + 4n\bar{x} - 4n\bar{x}^2}}{2(n + z^2)} < p < \frac{2n\bar{x} + z^2 + z\sqrt{z^2 + 4n\bar{x} - 4n\bar{x}^2}}{2(n + z^2)}$$

de manera que los extremos del intervalo de confianza, de nivel aproximadamente  $1 - \alpha$ , pueden expresarse

$$\frac{\bar{x} + z_{\alpha/2}^2/(2n) \pm z_{\alpha/2} \sqrt{z_{\alpha/2}^2/(4n^2) + \bar{x}/n - \bar{x}^2/n}}{(1 + z_{\alpha/2}^2/n)}.$$

El resultado puede aproximarse prescindiendo de los términos de orden  $1/n$ , obteniéndose entonces

$$\bar{x} - \frac{z_{\alpha/2}}{\sqrt{n}} \sqrt{\bar{x}(1 - \bar{x})} < p < \bar{x} + \frac{z_{\alpha/2}}{\sqrt{n}} \sqrt{\bar{x}(1 - \bar{x})}.$$

Resultado al que se llega, también, si en la desigualdad

$$-z_{\alpha/2} < \sqrt{n} \frac{\bar{x} - p}{\sqrt{p(1-p)}} < z_{\alpha/2}$$

se reemplaza, en el denominador, el valor de la probabilidad  $p$  por la frecuencia relativa  $\bar{x}$  y se despeja el valor  $p$  del numerador.

Por ejemplo para  $n = 100$  y  $T = 64$ , los intervalos de nivel de confianza 0'95, que se obtienen con cada una de las aproximaciones propuestas (dado que  $z_{\alpha/2} = 1'96$ ,  $\bar{x} = 0'64$ ) son

$$[0'542, 0'727] \quad \text{y} \quad [0'546, 0'734].$$

## Ejercicios

### Ejercicio 4.1

Sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple de una población con función de densidad

$$f_\theta(x) = \frac{2x}{\theta^2} \quad \text{si } 0 < x < \theta.$$

Determinar, por el método de Neyman, un intervalo de confianza para  $\theta$ , con nivel de confianza  $1 - \alpha$ .

### Ejercicio 4.2

Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de una distribución uniforme  $U(\theta - 1/2, \theta + 1/2)$ . Determinar diversos intervalos de confianza para  $\theta$ , con nivel de confianza  $1 - \alpha$ . Comparar sus longitudes.

¿Qué nivel de confianza tendría el intervalo  $[x_{(r)}, x_{(s)}]$  con  $r < s$ ?

### Ejercicio 4.3

Sea  $X$  una variable aleatoria con distribución de Laplace de densidad

$$f_\theta(x) = \frac{1}{2} e^{-|x-\theta|}.$$

Determinar un intervalo de confianza para  $\theta$  a nivel de confianza  $1 - \alpha$ , basado en una mediana muestral, suponiendo que el tamaño muestral es impar.

### Ejercicio 4.4

La altura de los individuos varones de una población sigue una distribución  $N(\mu, 0'075)$ . Si en una muestra aleatoria simple, de tamaño 12, de dicha población se obtuvo una media muestral  $\bar{x} = 1'75$ , determinar un intervalo de confianza para  $\mu$  a nivel de confianza  $1 - \alpha = 0'95$ . ¿Qué tamaño

muestral sería necesario para que el intervalo de confianza del mismo nivel tuviese longitud menor que 0'01?

### Ejercicio 4.5

Se desea comparar la longitud (en mm.) del cuerpo de dos poblaciones de ranas, geográficamente aisladas. Tomada una muestra de individuos machos en cada una de las dos poblaciones se obtuvieron los siguientes datos:

$$\begin{aligned} n &= 41, & \bar{x} &= 74, & S_1^2 &= 56'25, \\ m &= 56, & \bar{y} &= 78, & S_2^2 &= 44'89. \end{aligned}$$

Determinar un intervalo de confianza para la diferencia de medias poblacionales y otro para el cociente de varianzas poblacionales, de niveles de confianza del 95 %. ¿Qué tamaños muestrales son adecuados si se pretende que el intervalo para la diferencia de medias tenga longitud inferior a 2 mm.?

### Ejercicio 4.6

Se supone que el número de erratas por página de un libro sigue una distribución de Poisson. Elegidas al azar 95 páginas se obtuvo que

había	0	1	2	3	4	5	6	erratas
en	40	30	15	7	2	1	0	páginas

Hallar un intervalo de confianza, con nivel de confianza de 0'9, para el número medio de erratas por página del libro. ¿Cuántas páginas habría que examinar para que la longitud del intervalo de confianza obtenido fuese menor que 0'2?

### Ejercicio 4.7

En una experiencia genética se eligieron con reemplazamiento 20 moscas de una caja experimental; se obtuvieron las siguientes longitudes del ala de las moscas elegidas: 93, 90, 97, 90, 93, 91, 96, 94, 91, 91, 88, 93, 95, 91, 89, 92, 87, 88, 90 y 86 (en décimas de milímetro).

Suponiendo que la longitud del ala sigue una distribución normal, hallar un intervalo de confianza al 90 % para cada uno de los dos parámetros. Determinar una región de confianza para ambos parámetros simultáneamente, con el mismo nivel de confianza.

### Ejercicio 4.8

Dada una muestra aleatoria simple de una distribución de Cauchy de densidad

$$f_\theta(x) = \frac{1}{\pi} \frac{1}{1 + (x - \theta)^2} \quad \text{para } x \in \mathbb{R},$$

determinar diversos intervalos de confianza, de nivel de confianza  $1 - \alpha$ , para  $\theta$ . Comparar sus longitudes.

**Ejercicio 4.9**

Se dispone de una muestra aleatoria simple de una distribución discreta con función de probabilidad

$$p_\theta(x) = (1-p)p^{x-\theta} \quad \text{para } x = \theta, \theta + 1, \theta + 2, \dots$$

siendo  $\theta$  un número natural desconocido y  $p$  una probabilidad conocida. Determinar un intervalo de confianza para  $\theta$ , de nivel de confianza  $1 - \alpha$ .

**Ejercicio 4.10**

Sean  $(X_1, \dots, X_n)$  e  $(Y_1, \dots, Y_m)$  dos muestras aleatorias simples de dos poblaciones independientes con distribuciones exponenciales de parámetros  $\lambda_1$  y  $\lambda_2$  respectivamente. Determinar un intervalo de confianza para el cociente  $\lambda_1/\lambda_2$ , de nivel de confianza  $1 - \alpha$ .

## Solución de los ejercicios

### Ejercicio 4.1

La función de distribución de  $X$  es

$$F_\theta(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ x^2/\theta^2 & \text{si } 0 < x < \theta \\ 1 & \text{si } x \geq \theta. \end{cases}$$

Como  $\theta$  es el extremo superior del recorrido de la variable, parece razonable considerar como estadístico a utilizar en el método de Neyman, el máximo  $X_{(n)}$ , cuya función de distribución es

$$G_\theta(y) = [F_\theta(y)]^n = \begin{cases} 0 & \text{si } y \leq 0 \\ y^{2n}/\theta^{2n} & \text{si } 0 < y < \theta \\ 1 & \text{si } y \geq \theta. \end{cases}$$

El método de Neyman requiere determinar en primer lugar dos funciones  $c_1(\theta)$  y  $c_2(\theta)$  tales que

$$\mathcal{P}\{X_{(n)} < c_1(\theta)\} = \alpha_1, \quad \mathcal{P}\{X_{(n)} > c_2(\theta)\} = \alpha_2$$

con  $\alpha_1 + \alpha_2 = \alpha$ . Es decir,

$$\frac{c_1(\theta)^{2n}}{\theta^{2n}} = \alpha_1 \quad \text{y} \quad 1 - \frac{c_2(\theta)^{2n}}{\theta^{2n}} = \alpha_2$$

de donde se obtiene

$$c_1(\theta) = \theta \alpha_1^{1/(2n)} \quad \text{y} \quad c_2(\theta) = \theta (1 - \alpha_2)^{1/(2n)}.$$

Al interseccar con la recta  $X_{(n)} = x_{(n)}$  y proyectar los puntos de intersección sobre el eje de abscisas se obtiene el intervalo de confianza

$$\left[ (1 - \alpha_2)^{-1/(2n)} x_{(n)}, \alpha_1^{-1/(2n)} x_{(n)} \right].$$

Su longitud

$$\left[ \alpha_1^{-1/(2n)} - (1 - \alpha_2)^{-1/(2n)} \right] x_{(n)} = \left[ \alpha_1^{-1/(2n)} - (1 - \alpha + \alpha_1)^{-1/(2n)} \right] x_{(n)}$$

se hace mínima para  $\alpha_1 = \alpha$  (y  $\alpha_2 = 0$ ); de manera que el intervalo más preciso entre los anteriores es

$$\left[ x_{(n)}, \alpha^{-1/(2n)} x_{(n)} \right].$$

**Ejercicio 4.2**

La función de distribución poblacional es

$$F_\theta(x) = x - \theta + 1/2 \quad \text{para } \theta - 1/2 < x < \theta + 1/2.$$

El estadístico  $T = X_{(n)} - \theta$  tiene por función de distribución

$$\mathcal{P}_\theta\{X_{(n)} - \theta \leq t\} = \mathcal{P}_\theta\{X_{(n)} \leq t + \theta\} = (t + 1/2)^n \quad \text{para } -1/2 \leq t \leq 1/2$$

que es independiente de  $\theta$ .

Dados  $\alpha_1$  y  $\alpha_2$ , con  $\alpha_1 + \alpha_2 = \alpha$ , para que

$$\mathcal{P}\{T < c_1\} = \alpha_1 \quad \text{y} \quad \mathcal{P}\{T > c_2\} = \alpha_2$$

ha de ser

$$\left(c_1 + \frac{1}{2}\right)^n = \alpha_1 \quad \text{y} \quad 1 - \left(c_2 + \frac{1}{2}\right)^n = \alpha_2$$

es decir

$$c_1 = \alpha_1^{1/n} - \frac{1}{2} \quad \text{y} \quad c_2 = (1 - \alpha_2)^{1/n} - \frac{1}{2}.$$

Entonces

$$\mathcal{P}\left\{\alpha_1^{1/n} - \frac{1}{2} \leq X_{(n)} - \theta \leq (1 - \alpha_2)^{1/n} - \frac{1}{2}\right\} = 1 - \alpha$$

o bien

$$\mathcal{P}\left\{X_{(n)} + \frac{1}{2} - (1 - \alpha_2)^{1/n} \leq \theta \leq X_{(n)} + \frac{1}{2} - \alpha_1^{1/n}\right\} = 1 - \alpha,$$

de manera que

$$\left[X_{(n)} + \frac{1}{2} - (1 - \alpha_2)^{1/n}, X_{(n)} + \frac{1}{2} - \alpha_1^{1/n}\right]$$

es un intervalo de confianza para  $\theta$  de nivel de confianza  $1 - \alpha$ . Como su longitud,  $(1 - \alpha_2)^{1/n} - \alpha_1^{1/n}$ , crece con  $\alpha_2$ ; el intervalo más preciso basado en  $x_{(n)}$  se obtiene para  $\alpha_2 = 0$  y resulta ser

$$\left[X_{(n)} - \frac{1}{2}, X_{(n)} + \frac{1}{2} - \alpha^{1/n}\right],$$

de longitud  $1 - \alpha^{1/n}$ .

La construcción se puede repetir con el estadístico  $T' = X_{(1)} - \theta$  cuya función de distribución es

$$\mathcal{P}\{X_{(1)} - \theta \leq t\} = \mathcal{P}\{X_{(1)} \leq t + \theta\} = 1 - (1/2 - t)^n \quad \text{para } -1/2 \leq t \leq 1/2.$$

Puesto que la densidad  $n(1/2 - t)^{n-1}$  es decreciente en  $(-1/2, 1/2)$ , el intervalo de confianza más corto basado en  $X_{(1)}$  se obtendrá imponiendo

$$\mathcal{P}\left\{-\frac{1}{2} \leq X_{(1)} - \theta \leq c_2\right\} = 1 - \alpha,$$

es decir

$$1 - \left(\frac{1}{2} - c_2\right)^n = 1 - \alpha$$

de donde  $c_2 = 1/2 - \alpha^{1/n}$ . Luego

$$\left[ x_{(1)} - \frac{1}{2} + \alpha^{1/n}, x_{(1)} - \frac{1}{2} \right]$$

es el intervalo de confianza para  $\theta$ , de nivel de confianza  $1 - \alpha$ , de longitud mínima entre los basados en  $x_{(1)}$ . Su longitud es  $1 - \alpha^{1/n}$  como en el caso anterior.

El estadístico  $T'' = X_{(n)} + X_{(1)} - 2\theta$  puede emplearse también de forma similar y produce como intervalo de longitud mínima, de nivel de confianza  $1 - \alpha$ ,

$$\left[ \frac{1}{2}(x_{(n)} + x_{(1)} - 1 + \alpha^{1/n}), \frac{1}{2}(x_{(n)} + x_{(1)} + 1 - \alpha^{1/n}) \right]$$

con longitud  $1 - \alpha^{1/n}$  nuevamente.

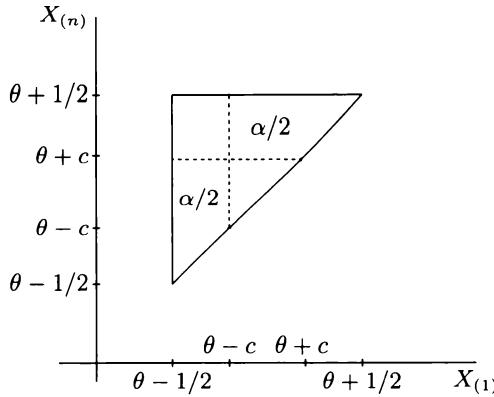
Puesto que la utilización de  $X_{(n)}$  y  $X_{(1)}$  es muy razonable, parece que el mejor resultado que puede conseguirse es un intervalo de longitud  $1 - \alpha^{1/n}$ . Sin embargo, se puede pensar en un intervalo de extremos  $X_{(1)} + c, X_{(n)} - c$ ; es decir un intervalo de confianza de la forma

$$[\min(X_{(1)} + c, X_{(n)} - c), \max(X_{(1)} + c, X_{(n)} - c)].$$

Para conseguir que  $\theta$  esté incluido entre los límites anteriores con probabilidad  $1 - \alpha$ , puede imponerse que cada una de las desigualdades

$$\theta < \min(x_{(1)} + c, x_{(n)} - c) \quad \text{y} \quad \theta > \max(x_{(1)} + c, x_{(n)} - c)$$

se produzca con probabilidad  $\alpha/2$ . Es decir, determinar  $c$  con las condiciones que indica la figura siguiente:



Puesto que la densidad conjunta de  $(X_{(1)}, X_{(n)})$  es

$$n(n-1)(y-x)^{n-2} \quad \text{para } \theta - 1/2 < x < y < \theta + 1/2$$

cada una de las dos regiones indicadas tiene probabilidad

$$\left. \begin{aligned} & \int_{\theta-1/2}^{\theta-c} \int_x^{\theta+c} n(n-1)(y-x)^{n-2} dy dx \\ & \int_{\theta+c}^{\theta+1/2} \int_{\theta-c}^y n(n-1)(y-x)^{n-2} dx dy \end{aligned} \right\} = (c+1/2)^n - (2c)^n.$$

Ha de ser pues  $(c+1/2)^n - (2c)^n = \alpha/2$ .

El estudio de la gráfica de la función  $(c+1/2)^n - (2c)^n$  muestra que la ecuación anterior (con  $\alpha < 1/2$ ) siempre tiene una solución  $c^*$  próxima a  $1/2$  (y otra menor que no interesa). Luego

$$[ \min(x_{(1)} + c^*, x_{(n)} - c^*) , \max(x_{(1)} + c^*, x_{(n)} - c^*) ]$$

es un intervalo de confianza, de nivel  $1 - \alpha$ , con longitud

$$|x_{(n)} - x_{(1)} - 2c^*|$$

Ahora la longitud es aleatoria (depende de la muestra), pero como la densidad de  $X_{(n)} - X_{(1)}$  es  $n(n-1)r^{n-2}(1-r)$  para  $0 < r < 1$ , la longitud esperada es

$$\begin{aligned} & \int_0^{2c^*} (2c^* - r)n(n-1)r^{n-2}(1-r)dr + \int_{2c^*}^1 (r - 2c^*)n(n-1)r^{n-2}(1-r)dr \\ &= 1 - 2c^* - \frac{2}{n+1} + 2(2c^*)^n - 2(2c^*)^{n+1} \frac{n-1}{n+1}. \end{aligned}$$

Por ejemplo, para  $\alpha = 0'02$  y  $n = 40$ , resulta  $c^* = 0'49974$  y la longitud esperada es  $0'04826$ . Mientras que la longitud constante de los intervalos determinados anteriormente es  $1 - \alpha^{1/n} = 0'093$ . La última solución es preferible en media, aunque hay un pequeño riesgo de que salga, por ejemplo,  $x_{(n)} - x_{(1)} < 0'85$  y la longitud sea superior a  $0'14948$ .

El nivel de confianza del intervalo  $[x_{(r)}, x_{(s)}]$  es igual a

$$\mathcal{P}_\theta \{ X_{(r)} \leq \theta \leq X_{(s)} \}.$$

Ahora bien las desigualdades anteriores equivalen a que haya, en el intervalo  $(\theta - 1/2, \theta)$ , un número de observaciones muestrales superior o igual a  $r$  e inferior a  $s$ . Como cada observación tiene probabilidad  $1/2$  de caer en  $(\theta - 1/2, \theta)$ , será

$$\mathcal{P}_\theta \{ X_{(r)} \leq \theta \leq X_{(s)} \} = \frac{1}{2^n} \sum_{i=r}^{s-1} \binom{n}{i}.$$

Por ejemplo, con  $n = 40$  y  $\alpha = 0'02$  habrá que tomar  $r = 13$  y  $s = 28$ , para que sea

$$\frac{1}{2^{40}} \sum_{i=13}^{27} \binom{40}{i} = 0'983 > 0'98.$$

La longitud esperada del intervalo  $[x_{(13)}, x_{(28)}]$  es entonces  $15/41$ ; notablemente superior a la obtenida antes.

### Ejercicio 4.3

La elección de la mediana muestral, como estadístico para la construcción de un intervalo de confianza, es viable siempre que la densidad poblacional sea de la forma  $f(x - \theta)$ , como es el caso que aquí se considera.

En general, la función de densidad de la mediana muestral,  $M$ , de una muestra de tamaño impar,  $n = 2k + 1$ , es

$$g_\theta(x) = \frac{(2k+1)!}{k!^2} [F_\theta(x)]^k [1 - F_\theta(x)]^k f_\theta(x);$$

así que, con la distribución teórica

$$F_\theta(x) = \begin{cases} \frac{1}{2} e^{x-\theta} & \text{si } x \leq \theta \\ 1 - \frac{1}{2} e^{\theta-x} & \text{si } x > \theta \end{cases}$$

resulta

$$g_\theta(x) = \begin{cases} \frac{(2k+1)!}{k!^2} \left(\frac{1}{2}e^{x-\theta}\right)^{k+1} \left(1 - \frac{1}{2}e^{x-\theta}\right)^k & \text{si } x \leq \theta \\ \frac{(2k+1)!}{k!^2} \left(1 - \frac{1}{2}e^{\theta-x}\right)^k \left(\frac{1}{2}e^{\theta-x}\right)^{k+1} & \text{si } x > \theta, \end{cases}$$

lo que convierte a  $Y = M - \theta$  en una cantidad pivotal, ya que su función de densidad

$$h(y) = \frac{(2k+1)!}{k!^2} \left(1 - \frac{1}{2}e^{-|y|}\right)^k \left(\frac{1}{2}e^{-|y|}\right)^{k+1}$$

no depende de  $\theta$ . Como la densidad es simétrica, creciente en  $(-\infty, 0)$  y decreciente en  $(0, \infty)$ , conviene buscar  $c$  tal que

$$\mathcal{P}\{Y \leq -c\} = \alpha/2 \quad \text{y} \quad \mathcal{P}\{Y \geq c\} = \alpha/2$$

es decir

$$\begin{aligned} \alpha/2 &= \int_{-\infty}^{-c} \frac{(2k+1)!}{k!^2} \left(1 - \frac{1}{2}e^y\right)^k \left(\frac{1}{2}e^y\right)^{k+1} dy \\ &= \int_0^{e^{-c}/2} \frac{\Gamma(2k+2)}{[\Gamma(k+1)]^2} (1-u)^k u^k du = \beta(e^{-c}/2, k+1, k+1) \end{aligned}$$

siendo  $\beta(x, p, q)$  la función de distribución de una beta  $\beta(p, q)$ .

Será entonces

$$\mathcal{P}\{-c < M - \theta < c\} = 1 - \alpha$$

y por tanto  $[M - c, M + c]$  un intervalo de confianza para  $\theta$ .

En la determinación de  $c$  es útil la identidad que relaciona la distribución de la beta y la binomial:

$$\sum_{i=j}^n \binom{n}{i} p^i (1-p)^{n-i} = \frac{n!}{(j-1)!(n-j)!} \int_0^p x^{j-1} (1-x)^{n-j} dx;$$

según la cual, ha de ser

$$\frac{\alpha}{2} = \beta(e^{-c}/2, k+1, k+1) = P\{W \geq k+1\}$$

siendo  $W$  una variable aleatoria con distribución binomial  $B(2k+1, e^{-c}/2)$ .

Así por ejemplo, con una muestra de tamaño  $n = 5$  y nivel de confianza 0'95, debe ser  $c = 1'226$  y el intervalo de confianza  $[M - 1'226, M + 1'226]$ .

### Ejercicio 4.4

Al ser la población normal de varianza conocida, el intervalo de confianza a utilizar es

$$\left( \bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right).$$

Como  $z_{\alpha/2} = z_{0'025} = 1'96$ , el intervalo buscado es

$$(1'7076, 1'7924).$$

Por otra parte, si ha de ser

$$2 z_{\alpha/2} \sigma / \sqrt{n} = 2 \times 1'96 \times 0'075 / \sqrt{n} < 0'01,$$

hay que tomar  $\sqrt{n} > 2 \times 1'96 \times 0'075 / 0'01 = 29'4$  o bien  $n > 864'36$ . Con tamaño muestral  $n = 865$  la longitud del intervalo de confianza será inferior a 0'01.

### Ejercicio 4.5

Las poblaciones son independientes y los tamaños muestrales son suficientemente grandes para suponer que  $S_1^2$  y  $S_2^2$  dan buenas aproximaciones de las varianzas poblacionales. Por tanto, el intervalo de confianza para  $\mu_1 - \mu_2$  buscado es

$$\left( \bar{x} - \bar{y} - z_{\alpha/2} \sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}, \bar{x} - \bar{y} + z_{\alpha/2} \sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}} \right),$$

que con los valores especificados se reduce a  $(-6'89, -1'11)$ .

Para  $\sigma_1^2 / \sigma_2^2$ , el intervalo de confianza buscado es

$$\left( \frac{S_1^2 / S_2^2}{F_{40,55;0'025}}, \frac{S_1^2 / S_2^2}{F_{40,55;0'975}} \right);$$

como  $F_{40,55;0'975} = 1'7713$  y  $F_{40,55;0'025} = 0'55$  se obtiene el intervalo

$$(0'707, 2'278).$$

Si se desea que sea  $2 \times 1'96\sqrt{S_1^2/n + S_2^2/m} < 2$ , cabe suponer que en muestras de mayor tamaño los valores de las cuasivarianzas muestrales sean similares a los obtenidos. De manera que la previsión a hacer es que sea

$$\frac{56'25}{n} + \frac{44'89}{m} < 0'26.$$

Para conseguir que se verifique esta desigualdad con  $n + m$  lo más pequeño posible, debe ser  $n = m > 389$ . Es conveniente tomar, por tanto,  $n = m = 390$ , o algo más para prevenir un ligero aumento de las cuasivarianzas de las nuevas muestras.

### Ejercicio 4.6

La variable  $X$ , número de erratas por página, sigue una distribución de Poisson de parámetro  $\lambda$ , pero el número de observaciones es elevado,  $n = 95$ , de forma que se puede utilizar el intervalo de confianza aproximado para el parámetro  $\lambda$ :

$$\left[ \bar{x} - z_{\alpha/2} \sqrt{\bar{x}/n}, \bar{x} + z_{\alpha/2} \sqrt{\bar{x}/n} \right].$$

Como la distribución de frecuencias proporciona una media muestral  $\bar{x} = 0'989$  y es  $z_{\alpha/2} = z_{0'05} = 1'645$ , se obtiene el intervalo de confianza

$$[0'821, 1'157].$$

Suponiendo que en una muestra de tamaño superior se obtendrá una media muestral similar a la anterior, el tamaño muestral debe fijarse de forma que

$$2 \times 1'645 \sqrt{0'989/n} < 0'2$$

es decir  $n > 268$ . Conviene pues tomar  $n = 269$  o algo más para prevenir un ligero aumento de la media muestral.

### Ejercicio 4.7

Como la variable en estudio sigue una distribución normal de parámetros desconocidos, el intervalo de confianza a utilizar para la media,  $\mu$ , será

$$\left( \bar{x} - t_{n-1;\alpha/2} \frac{s}{\sqrt{n-1}}, \bar{x} + t_{n-1;\alpha/2} \frac{s}{\sqrt{n-1}} \right).$$

De los datos se obtiene que  $\bar{x} = 91'25$  y  $s = 2'8614$ . Como es  $t_{n-1;\alpha/2} = t_{19;0'05} = 1'729$ , el intervalo de confianza buscado es  $(90'11, 92'38)$ .

Para la varianza, el intervalo de confianza es

$$\left( \frac{ns^2}{\chi_{n-1;\alpha/2}^2}, \frac{ns^2}{\chi_{n-1;1-\alpha/2}^2} \right) = \left( \frac{163'75}{30'14}, \frac{163'75}{10'12} \right) = (5'432, 16'09).$$

Por último, tomemos  $1 - \alpha = 1 - \beta = \sqrt{0'9} = 0'948$ . Es entonces  $z_{\alpha/2} = z_{0'026} = 1'943$ ,  $\chi_{19,0'026}^2 = 8'957$  y  $\chi_{19,0'974}^2 = 32'79$ ; de forma que la región de confianza está limitada por la parábola

$$\sigma^2 = 5'3(\mu - 91'25)^2$$

y las rectas

$$\sigma^2 = 4'99 \quad \text{y} \quad \sigma^2 = 18'28$$

Es conveniente visualizar en el plano  $(\mu, \sigma^2)$  las tres regiones que, según los sucesivos resultados, tienen probabilidad 0'9.

### Ejercicio 4.8

Si  $X_i$  representa el elemento  $i$ -ésimo de la muestra,  $Y_i = X_i - \theta$  tiene función de densidad y de distribución

$$f(y) = \frac{1}{\pi(1+y^2)}, \quad F(y) = \frac{1}{2} + \frac{1}{\pi} \arctg y$$

respectivamente y función característica  $\varphi(t) = e^{-|t|}$ .

El estadístico  $\bar{Y} = \bar{X} - \theta$  tiene entonces función característica

$$E[e^{it\bar{Y}}] = E[e^{it/n(Y_1+Y_2+\dots+Y_n)}] = E[e^{it/nY_1}]^n = e^{-|t|}$$

y, por tanto, la misma función de distribución  $F(y)$ . Como su densidad es simétrica, creciente en  $(-\infty, 0)$  y decreciente en  $(0, \infty)$ , el intervalo de confianza más corto posible, basado en  $\bar{X}$ , se obtiene observando que

$$\mathcal{P} \left\{ -\operatorname{tg}(1-\alpha)\frac{\pi}{2} < \bar{X} - \theta < \operatorname{tg}(1-\alpha)\frac{\pi}{2} \right\} = 1 - \alpha;$$

con lo cual

$$\left( \bar{x} - \operatorname{tg}(1-\alpha)\frac{\pi}{2}, \bar{x} + \operatorname{tg}(1-\alpha)\frac{\pi}{2} \right)$$

es un intervalo de confianza para  $\theta$ , de nivel de confianza  $1 - \alpha$ , para  $\theta$ .

Su longitud,  $2 \operatorname{tg} (1-\alpha)\pi/2$ , no depende de  $n$ ; de manera que la precisión obtenida es independiente del tamaño muestral: la misma que si se utilizase sólo el valor de  $X_1$  para diseñar el intervalo de confianza

$$\left( x_1 - \operatorname{tg} (1-\alpha)\frac{\pi}{2}, x_1 + \operatorname{tg} (1-\alpha)\frac{\pi}{2} \right).$$

El estadístico  $X_{(n)} - \theta = Y_{(n)}$  tiene distribución

$$\mathcal{P}\{Y_{(n)} \leq y\} = F(y)^n = \left( \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} y \right)^n$$

de densidad

$$n \left( \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} y \right)^{n-1} \frac{1}{\pi(1+y^2)}$$

que alcanza su máximo cuando

$$y \left( \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} y \right) = \frac{n-1}{2\pi};$$

es decir, en un punto  $y \geq (n-1)/2\pi$ , en el cual la densidad vale menos que

$$\frac{n}{\pi} \left[ 1 + \left( \frac{n-1}{2\pi} \right)^2 \right]^{-1}$$

La longitud de cualquier intervalo de confianza, basado en  $X_{(n)}$ , que deje una probabilidad  $1-\alpha$  sobre él y debajo de la densidad de  $X_{(n)}$ , tiene entonces que ser superior a

$$\frac{(1-\alpha)\pi}{n} \left[ 1 + \left( \frac{n-1}{2\pi} \right)^2 \right],$$

lo cual crece cuando  $n$  aumenta. Hay por tanto que desechar el uso de  $X_{(n)}$  para la construcción del intervalo de confianza requerido.

Supongamos, para evitar complicaciones, que el tamaño muestral es impar,  $n = 2k+1$ , y consideremos el estadístico basado en la mediana muestral:  $X_{(k+1)} - \theta = Y_{(k+1)}$ . Su densidad

$$h(y) = \frac{(2k+1)!}{k!^2} \left( \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} y \right)^k \left( \frac{1}{2} - \frac{1}{\pi} \operatorname{arctg} y \right)^k \frac{1}{\pi} \frac{1}{1+y^2},$$

es simétrica, creciente en  $(-\infty, 0)$  y decreciente en  $(0, \infty)$ ; así que el mejor intervalo de confianza, de nivel  $1 - \alpha$ , basado en  $X_{(k+1)}$  se obtiene determinando  $c$  con la condición

$$\int_{-\infty}^{-c} h(y) dy = \alpha/2;$$

lo cual permite afirmar que

$$\mathcal{P}\{-c < X_{(k+1)} - \theta < c\} = 1 - \alpha,$$

o bien que  $(x_{(k+1)} - c, x_{(k+1)} + c)$  es un intervalo de confianza para  $\theta$ , de nivel de confianza  $1 - \alpha$ .

La condición acerca de  $c$  equivale a

$$\begin{aligned} \alpha/2 &= \int_0^{1/2 + 1/\pi \operatorname{arctg} c} \frac{(2k+1)!}{k!^2} u^k (1-u)^k du \\ &= \beta(1/2 + 1/\pi \operatorname{arctg} c, k+1, k+1) = P\{W \geq k+1\} \end{aligned}$$

siendo  $\beta(x, p, q)$  la función de distribución de una  $\beta(p, q)$  y  $W$  una variable con distribución binomial  $B(2k+1, 1/2 + 1/\pi \operatorname{arctg} c)$ .

Al margen de la determinación de  $c$ , para  $n$  y  $\alpha$  dados, cabe observar que la densidad de la  $\beta(k+1, k+1)$  es cada vez más concentrada alrededor de  $1/2$ , a medida que  $n$  crece; por consiguiente,  $1/2 + 1/\pi \operatorname{arctg} c$  debe acercarse a  $1/2$ , o bien  $c$  acercarse a cero. Así que la precisión es buena si el tamaño muestral es grande y el intervalo basado en  $X_{(k+1)}$  es preferible a los anteriores.

### Ejercicio 4.9

Puesto que  $\theta$  es el menor valor observable, la menor de las observaciones muestrales,  $X_{(1)}$ , es un estadístico adecuado para la construcción del intervalo de confianza.

La función de distribución de la población es

$$F_\theta(x) = (1-p) \sum_{i=\theta}^{[x]} p^{i-\theta} = 1 - p^{[x]-\theta+1} \quad \text{para } x \geq \theta$$

siendo  $[x]$  la parte entera de  $x$ . Si  $n$  representa el tamaño muestral, la distribución de  $X_{(1)}$  será

$$\mathcal{P}\{X_{(1)} \leq t\} = 1 - [1 - F(t)]^n = 1 - p^{n([t]-\theta+1)} \quad \text{para } t \geq \theta;$$

con lo cual  $X_{(1)} - \theta$  tiene función de distribución

$$\mathcal{P}\{X_{(1)} - \theta \leq z\} = 1 - p^{n([z]+1)} \quad \text{si } z \geq 0,$$

que es independiente de  $\theta$ .

Si  $c_1 \leq c_2$  son tales que

$$\mathcal{P}\{c_1 \leq X_{(1)} - \theta \leq c_2\} = p^{n[c_1]} - p^{n([c_2]+1)} \geq 1 - \alpha$$

entonces  $[x_{(1)} - c_2, x_{(1)} - c_1]$  es un intervalo de confianza para  $\theta$ , de nivel de confianza  $1 - \alpha$ .

Como  $\mathcal{P}\{X_{(1)} - \theta = k\} = (1 - p^n)p^{nk}$  decrece con  $k$ , el intervalo anterior será de longitud mínima si  $c_1 = 0$  y  $c_2$  es lo más pequeño posible entre los valores que verifican  $p^{n(c_2+1)} \leq \alpha$ ; es decir,

$$[c_2] \geq \frac{\log \alpha}{n \log p} - 1.$$

Por tanto, debe ser  $c_2 = [\log \alpha / n \log p]$  y el intervalo de confianza resulta

$$\left[ x_{(1)} - \left[ \frac{\log \alpha}{n \log p} \right], x_{(1)} \right];$$

que se reduce al punto  $x_{(1)}$  si  $\alpha > p^n$ .

### Ejercicio 4.10

$2\lambda_1 X_1, \dots, 2\lambda_1 X_n; 2\lambda_2 Y_1, \dots, 2\lambda_2 Y_m$  son variables aleatorias independientes, con distribución exponencial de parámetro  $1/2$ . Por tanto

$$2\lambda_1 n \bar{X} = \sum_{i=1}^n 2\lambda_1 X_i \quad \text{tiene distribución } \gamma(n, 1/2) = \chi_{2n}^2,$$

$$2\lambda_2 m \bar{Y} = \sum_{i=1}^m 2\lambda_2 Y_i \quad \text{tiene distribución } \gamma(m, 1/2) = \chi_{2m}^2;$$

con lo cual

$$\frac{\lambda_1 \bar{X}}{\lambda_2 \bar{Y}} \quad \text{tiene distribución } F_{2n, 2m}.$$

Por consiguiente

$$\mathcal{P} \left\{ F_{2n, 2m; 1-\alpha/2} < \frac{\lambda_1 \bar{X}}{\lambda_2 \bar{Y}} < F_{2n, 2m; \alpha/2} \right\} = 1 - \alpha$$

y resulta como intervalo de confianza para  $\lambda_1/\lambda_2$ , de nivel de confianza  $1-\alpha$ , el intervalo

$$\left( \frac{\bar{Y}}{\bar{X}} F_{2n, 2m; 1-\alpha/2}, \frac{\bar{Y}}{\bar{X}} F_{2n, 2m; \alpha/2} \right).$$

# Capítulo 5

## Estimación puntual

### 5.1. Introducción

Ante el problema estadístico de evaluar el parámetro que identifica la distribución de una población dentro de una familia paramétrica  $\mathcal{F} = \{F_\theta | \theta \in \Theta\}$ , en muchas ocasiones no es suficiente determinar un intervalo en el que se encuentre, sino que es preciso seleccionar un único valor de  $\Theta$  que constituya un pronóstico individual sobre el parámetro. Se habla entonces de *estimación puntual*, puesto que la disminución del desconocimiento acerca del valor paramétrico se pretende que conduzca a un único punto.

Por ejemplo es difícil imaginar que el Instituto de Estadística emitiese, al final de un semestre, la información de que el coste de la vida ha subido durante ese período un porcentaje comprendido entre 2'6 % y 3'1 %. Aunque tal fuese la conclusión inicial a la que condujesen sus datos, sin duda tendría que arbitrar un procedimiento para delimitar una única cifra (2'9 %, por ejemplo) que pudiera ser utilizada unívocamente en todas las cuestiones económicas en las que repercute el I.P.C.

Sin embargo es necesario tener en cuenta que, si no hay seguridad absoluta, sino sólo un nivel de confianza elevado, de que el valor del parámetro pertenezca a un intervalo de confianza diseñado para él, menos se puede pretender acertar con el valor paramétrico exacto mediante un único número. Así, si tras lanzar 100 veces un dado, se da 0'17 como estimación puntual de la probabilidad de obtener 4, no está excluido que el valor real pueda ser 0'165, 0'175 ó 1/6. El número 0'17 debe entenderse como resultado del compromiso de suministrar una única conjectura estadística acerca de dicha probabilidad, e indicará —si la estimación es “buena”— que su valor es “próximo” a 0'17... Mejor sería decir, en primera aproximación, que la

estimación puntual será tanto mejor cuanto más próxima sea al valor desconocido del parámetro.

En este orden de ideas, lo razonable en la práctica es adjuntar a la estimación puntual, un intervalo de confianza que mida el margen de error que, de acuerdo con las observaciones, puede tener aquella. Ello llevaría, en la situación anterior, a afirmar que la probabilidad buscada es  $0'17 \pm 0'074$ , con nivel de confianza  $0'95$ , por ejemplo. Afirmaciones de este tipo son frecuentes (a menudo en letra pequeña) en las buenas encuestas sociológicas que publica la prensa. Sin embargo la preocupación en este capítulo debe ser ante todo diseñar los métodos viables para obtener estimaciones puntuales, puesto que la construcción de intervalos de confianza ya ha sido considerada anteriormente.

La manera de proceder en un problema estadístico para obtener una estimación puntual del valor paramétrico es sumamente sencilla: consiste simplemente en seleccionar un estadístico  $T(X_1, X_2, \dots, X_n)$ , función únicamente de las observaciones muestrales y tomar, como estimación, el valor de  $T$  calculado a partir de la muestra obtenida. Ello supone aceptar exclusivamente que la estimación puntual del parámetro debe ser función de la muestra observada.

De manera genérica, los estadísticos  $T(X_1, X_2, \dots, X_n)$ , independientes del parámetro, cuyos valores se utilizan para obtener su estimación puntual, se denominan estimadores. Normalmente su recorrido debería estar contenido en el espacio paramétrico  $\Theta$ , es decir que no debería proporcionar estimaciones fuera del rango del parámetro.

Así pues la obtención de una estimación puntual, una vez dado el estimador, no plantea ninguna dificultad —a diferencia de lo que ocurría en relación con los intervalos de confianza. Aquí, por el contrario, la cuestión se centra en la necesidad de elegir cuidadosamente el estimador, en función del modelo estadístico que se maneje. Algunas situaciones concretas pueden mostrarlo con claridad.

Por ejemplo, para la estimación de la probabilidad de obtener 4 con un cierto dado, después de observados los resultados  $x_1, x_2, \dots, x_n$  en  $n$  lanzamientos independientes, es razonable considerar el estimador

$$T_4(X_1, X_2, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n I_{\{X_i=4\}}$$

que expresa la frecuencia observada de cuatros en la muestra. En cambio, carece de toda lógica utilizar, con el mismo fin, la frecuencia observada de cincos. Si el objetivo fuese estimar las probabilidades de las diversas caras:  $\theta = (p_1, p_2, \dots, p_6)$ , las proporciones muestrales de cada resultado constituirían un estimador razonable de dimensión 6:  $T = (T_1, T_2, \dots, T_6)$ , indudablemente mejor que  $(T_4, T_1, T_3, T_5, T_6, T_2)$ .

De manera similar, para estimar la estatura media de los individuos de una población, supuesto que se distribuye con distribución  $N(\mu, \sigma)$ , parece sensato calcular la media muestral  $\bar{x}$  de las estaturas contenidas en la muestra y tomarla como estimación puntual de  $\mu$ . En cambio, si se trata de estimar el número  $N$  de bolas contenidas en una urna y numeradas correlativamente a partir de 1:  $\{1, 2, \dots, N\}$ , a nadie se le ocurriría usar la media muestral como estimador, ya que el parámetro a estimar supera a la mayor de las observaciones efectuadas.

En el caso de una distribución exponencial de parámetro desconocido  $\theta$ , puesto que  $1/\theta$  es la media de la distribución y  $1/\theta^2$  su varianza, las primeras propuestas para la estimación puntual de  $\theta$  pueden ser  $1/\bar{X}$  ó  $1/s$ . Pero nada permite excluir otros estimadores, como  $(nX_{(1)})^{-1}$  o bien el inverso de la mediana muestral, mientras no se analicen los criterios para comparar unos estimadores con otros.

Estas primeras situaciones esbozadas, parecen ratificar lo que ya se ha indicado: todo el problema radica en la selección del estimador adecuado a las características del modelo estadístico en estudio. Pero ello requiere saber juzgar lo que es adecuado y lo que no lo es, según el modelo. La idea inicial de que la estimación puntual sea próxima al valor real del parámetro, que no es conocido, es demasiado vaga para que pueda ser usada en bruto; habrá, por tanto, que precisar criterios para comparar diversos resultados obtenidos, estudiar qué propiedades son deseables para que un estadístico produzca buenas estimaciones puntuales, diseñar métodos para obtener estimadores razonables, etc. Éstos son los aspectos fundamentales que se desarrollarán en las próximas secciones y en los dos capítulos siguientes.

Muchas de estas cuestiones de comparación, selección y cualidades de los estimadores, tendrán que basarse, no en los valores particulares que se obtengan con ellos para muestras determinadas, sino en la distribución probabilística de las estimaciones que produzcan al elegir al azar la muestra. Ello equivale a decir que la distribución en el muestreo de los estimadores seguirá jugando un papel esencial en el análisis de su calidad relativa.

Por otra parte, los resultados obtenidos en el diseño de buenos estimadores puntuales resultarán también útiles a la hora de emplear otras técnicas estadísticas: por ejemplo, en relación con los intervalos de confianza, no se ha hecho mucho hincapié en la selección del estadístico con el cual llevar a cabo la construcción. En este sentido, la estimación puntual puede considerarse la técnica más primaria ya que, tratando de encontrar estadísticos que proporcionen resultados próximos al valor paramétrico (o con otras cualidades alternativas en esa dirección), se obtendrán sin duda estadísticos adecuados para la construcción de intervalos de confianza.

## 5.2. Comparación de estimadores: Riesgo y error cuadrático medio

Para enfrentarse con el problema de la estimación puntual de un parámetro  $\theta$  del que depende la distribución de una población, a partir de las observaciones  $(x_1, x_2, \dots, x_n)$  de una muestra aleatoria, se cuenta en principio con la posibilidad de utilizar cualquier estimador:

$$T : \mathcal{X} \rightarrow \Theta$$

función (medible) del espacio muestral en el espacio paramétrico. Ante la amplia gama de posibilidades, cabe intentar introducir una preferencia entre ellas, precisando algún criterio que permita comparar unos estimadores con otros. Naturalmente dichos criterios pueden depender del contexto del problema y, consecuentemente, deben ser añadidos en el planteamiento mediante algún nuevo elemento.

Puesto que en definitiva se trata de valorar la calidad de las estimaciones, lo más razonable es suponer que se cuenta con un dato más en el problema: una función de pérdida

$$L : \Theta \times \Theta \rightarrow \mathbb{R}$$

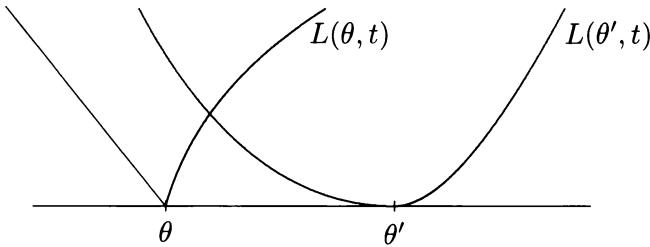
cuyos valores  $L(\theta, t)$  indican el coste en que se incurre si se da como estimación el valor  $t$ , cuando realmente el parámetro que identifica la distribución teórica vale  $\theta$ .

Una función de pérdida coherente debe satisfacer las propiedades:

- i)  $L(\theta, t) \geq 0$  para cualesquiera  $\theta, t \in \Theta$ .
- ii)  $L(\theta, t) = 0$ , si  $t = \theta$ .

- iii)  $L(\theta, t) \leq L(\theta, t')$ , si la distancia de  $t$  a  $\theta$  es menor que la distancia de  $t'$  a  $\theta$ .

Así, en el caso en que  $\Theta$  sea unidimensional, la representación gráfica de los valores  $L(\theta, t)$ , para cada valor fijo de  $\theta$ , serán curvas del tipo:



Siempre en el caso unidimensional, ejemplos habituales de funciones de pérdida son:

$L_1(\theta, t) = |\theta - t|$ , que mide directamente el error en la estimación.

$L_2(\theta, t) = (\theta - t)^2$ , error cuadrático de estimación.

$L_3(\theta, t) = \frac{|\theta - t|}{\theta}$ , error relativo en la estimación.

$L_4(\theta, t) = \begin{cases} c & \text{si } |\theta - t| > \varepsilon \\ 0 & \text{si } |\theta - t| \leq \varepsilon \end{cases}$ , que penaliza con un coste  $c$  los errores mayores que  $\varepsilon$ , etc.

La valoración de las estimaciones en relación con el valor del parámetro, que introduce la función de pérdida, ha de traducirse en una valoración de los estimadores: Si se decide hacer uso del estimador  $T(X_1, X_2, \dots, X_n)$ , la estimación que se obtendrá es una variable aleatoria (puesto que, antes de extraer la muestra, el resultado de las observaciones es aleatorio); lo mismo ocurrirá entonces con la pérdida en que se incurre,  $L(\theta, T(X_1, X_2, \dots, X_n))$ , según cuál sea el valor del parámetro. Puede considerarse por tanto la pérdida media:

$$R_T(\theta) = E_\theta[L(\theta, T(X_1, X_2, \dots, X_n))] = \int_{\mathbb{R}} L(\theta, t) H_\theta(dt)$$

donde  $H_\theta(t)$  representa la distribución en el muestreo del estadístico  $T$ , correspondiente al valor  $\theta$  del parámetro.

Dicha pérdida media mide el *riesgo del estimador*  $T$  y es, evidentemente, una función positiva de  $\theta$ , que puede expresarse equivalentemente en la forma

$$R_T(\theta) = \int_{\mathcal{X}} L(\theta, T(x_1, x_2, \dots, x_n)) F_\theta(dx_1) F_\theta(dx_2) \cdots F_\theta(dx_n)$$

en términos de la distribución de la muestra, supuesto que es aleatoria simple. Así pues, el riesgo  $R_T(\theta)$  depende, además de  $T$  y de  $\theta$ , de la función de pérdida introducida y de la distribución teórica  $F_\theta$  del modelo.

El concepto de riesgo proporciona un criterio para la comparación de estimadores; de hecho, un estimador  $T_1$  será *preferible* a otro estimador  $T_2$  si

$$\begin{aligned} R_{T_1}(\theta) &\leq R_{T_2}(\theta) && \text{para cualquier } \theta \in \Theta \quad \text{y} \\ R_{T_1}(\theta) &< R_{T_2}(\theta) && \text{para algún } \theta \in \Theta. \end{aligned}$$

El cálculo del riesgo no suele ser sencillo y depende además, obviamente, de la forma funcional del estimador; de manera que el criterio anterior sólo puede ser utilizado para realizar comparaciones dos a dos o dentro de una familia de estimadores con la misma expresión funcional. Veamos como:

### Ejemplo 5.1

Considérese una población con distribución uniforme en el intervalo  $(0, \theta)$  siendo  $\theta > 0$  un parámetro desconocido. Puesto que  $\theta$  es el máximo de los valores observables en la población, para dar una estimación puntual de  $\theta$  a partir de una muestra aleatoria simple  $(X_1, X_2, \dots, X_n)$  parece conveniente considerar el estadístico

$$T_1 = \max_{1 \leq i \leq n} X_i$$

o bien algún múltiplo suyo:  $T_k = kT_1$ , con  $k > 0$ .

La distribución en el muestreo de  $T_1$  es

$$\mathcal{P}_\theta \left\{ \max_{1 \leq i \leq n} X_i \leq t \right\} = \left( \frac{t}{\theta} \right)^n \quad \text{si } t \in (0, \theta),$$

de densidad  $\frac{n}{\theta} \left( \frac{t}{\theta} \right)^{n-1}$  para  $t \in (0, \theta)$ ; de manera que

$$\mathbb{E}_\theta[T_1] = \int_0^\theta t \frac{n}{\theta} \left( \frac{t}{\theta} \right)^{n-1} dt = \frac{n}{n+1} \theta$$

y

$$E_\theta[T_1^2] = \int_0^\theta t^2 \frac{n}{\theta} \left(\frac{t}{\theta}\right)^{n-1} dt = \frac{n}{n+2} \theta^2.$$

Luego, si se considera como función de pérdida  $L(\theta, t) = (\theta - t)^2/\theta^2$  (error relativo en la estimación al cuadrado), el riesgo del estimador  $T_k$  será

$$\begin{aligned} R_{T_k}(\theta) &= E_\theta \left[ \frac{(\theta - T_k)^2}{\theta^2} \right] = 1 - \frac{2}{\theta} E_\theta[T_k] + \frac{1}{\theta^2} E_\theta[T_k^2] \\ &= 1 - \frac{2}{\theta} E_\theta[kT_1] + \frac{1}{\theta^2} E_\theta[k^2T_1^2] = 1 - \frac{2k}{\theta} \frac{n}{n+1} \theta + \frac{k^2}{\theta^2} \frac{n}{n+2} \theta^2 \\ &= 1 - \frac{2kn}{n+1} + \frac{k^2 n}{n+2}. \end{aligned}$$

En este caso, el riesgo no depende de  $\theta$ , y, entre los estimadores  $T_k$ , será preferible aquel que haga mínimo el riesgo. Como el mínimo de  $R_{T_k}(\theta)$  se alcanza para  $k = (n+2)/(n+1)$ , el mejor estimador (en el sentido de la pérdida introducida) entre los múltiplos de  $T_1$  es

$$\frac{n+2}{n+1} \max_{1 \leq i \leq n} X_i.$$

Con función de pérdida  $L(\theta, t) = (\theta - t)^2$ , el riego se multiplica por  $\theta^2$  y, por tanto, depende de  $\theta$ . Sin embargo, el mismo estimador sigue dando el mínimo riesgo, cualquiera que sea el valor de  $\theta$ .

Puede observarse que el procedimiento empleado no indica nada acerca de otros posibles estimadores y, sólo ante cada propuesta concreta de un nuevo estimador  $T'$ , podría intentarse comparar su riesgo con el obtenido. Además la situación anterior es poco representativa, en el sentido de que no suele ocurrir que un mismo estimador minimice el riesgo para todos los valores de  $\theta$  simultáneamente. He aquí una situación similar pero de resultado más habitual:

### Ejemplo 5.2

Ciertas piezas tienen una duración mínima  $\theta > 0$  y la duración extra tiene distribución exponencial de parámetro 1; de manera que la distribución de las duraciones de la población de piezas tiene densidad

$$f_\theta(x) = e^{\theta-x} \quad \text{si } x > \theta.$$

La estimación de la duración mínima  $\theta$  a partir de una muestra aleatoria simple  $(X_1, X_2, \dots, X_n)$ , puede hacerse mediante el estimador

$$T_1 = \min_{1 \leq i \leq n} X_i$$

ó algún múltiplo suyo:  $T_k = kT_1$ , con  $k > 0$ .

La distribución en el muestreo de  $T_1$  es

$$\mathcal{P}_\theta \left\{ \min_{1 \leq i \leq n} X_i \leq t \right\} = 1 - P\{X_i > t\}^n = 1 - e^{(\theta-t)n} \quad \text{si } t > \theta,$$

cuya densidad es  $ne^{(\theta-t)n}$  para  $t > \theta$ .

Supongamos que la función de pérdida es  $L(\theta, t) = |\theta - t|$ , entonces para  $k \geq 1$

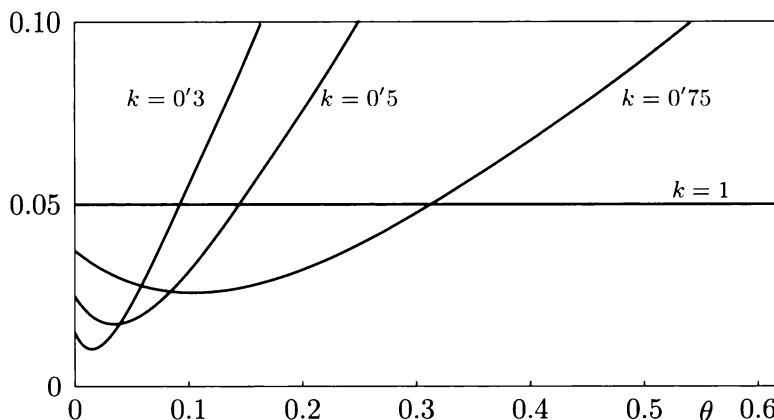
$$R_{T_k}(\theta) = E_\theta[|\theta - kT_1|] = \int_\theta^\infty (kt - \theta) ne^{n(\theta-t)} dt = (k-1)\theta + \frac{k}{n}.$$

Por tanto, todos los estimadores  $T_k$  con  $k > 1$  pueden ser eliminados frente al estimador  $T_1$ , ya que  $R_{T_k}(\theta) > R_{T_1}(\theta)$  si  $k > 1$ .

Sin embargo, para  $k < 1$ ,

$$\begin{aligned} R_{T_k}(\theta) &= \int_\theta^{\theta/k} (\theta - kt) ne^{n(\theta-t)} dt + \int_{\theta/k}^\infty (kt - \theta) ne^{n(\theta-t)} dt \\ &= n\theta e^{n\theta} \left\{ \int_\theta^{\theta/k} e^{-nt} dt - \int_{\theta/k}^\infty e^{-nt} dt \right\} + kn e^{n\theta} \left\{ \int_{\theta/k}^\infty te^{-nt} dt - \int_\theta^{\theta/k} te^{-nt} dt \right\} \\ &= \theta e^{n\theta} \left\{ e^{-n\theta} - 2e^{-n\theta/k} \right\} + ke^{n\theta} \left\{ \frac{2\theta}{k} e^{-n\theta/k} + \frac{2}{n} e^{-n\theta/k} - \theta e^{-n\theta} - \frac{1}{n} e^{-n\theta} \right\} \\ &= \theta + \frac{2k}{n} e^{n\theta(1-1/k)} - k\theta - \frac{k}{n}. \end{aligned}$$

Como función de  $\theta$ ,  $R_{T_k}(\theta)$  vale  $k/n$  en el origen y tiene por asíntota  $(1-k)\theta - k/n$  cuando  $\theta$  tiende a infinito. La representación gráfica de las funciones  $R_{T_k}(\theta)$  (para  $n = 20$ ), que aparece en la figura siguiente, muestra que ningún estimador  $T_k$  es preferible a ningún otro  $T_{k'}$  (si  $k, k' < 1$ ), pues el riesgo de  $T_k$  es menor que el de  $T_{k'}$  para unos valores de  $\theta$  y mayor para otros.

Figura 5.1: Riesgo de  $T_k$ 

En resumen, el criterio de preferencia establecido elimina, en este caso, los estimadores  $T_k$  con  $k > 1$  pero no permite seleccionar ninguno entre los estimadores  $T_k$  con  $k \leq 1$ . Lógicamente, cuanto más pequeño es  $k$  menos riesgo tiene el estimador  $T_k$  frente a pequeños valores de  $\theta$ .

Antes de discutir la situación planteada por el ejemplo anterior, hay que señalar que, muy a menudo, el problema estadístico no especifica una función de pérdida determinada, para valorar los errores de estimación, ni tampoco el contexto del problema permite adoptar una especial. En tal caso lo usual es escoger

$$L(\theta, t) = (\theta - t)^2$$

(o bien  $\sum_{i=1}^m (\theta_i - t_i)^2$  en el caso de un parámetro  $m$ -dimensional). Tanto es así que el riesgo recibe entonces el nombre específico de *error cuadrático medio del estimador*  $T$  en cuestión y se representa por  $ECM_T(\theta)$ :

$$ECM_T(\theta) = E_\theta[|\theta - T|^2]$$

Las razones para tal elección son dos: en primer lugar, la función  $|\theta - t|^2$  se presta adecuadamente al cálculo explícito, creando menos dificultades que  $|\theta - t|$  u otra pérdida más compleja. Por otra parte, el error cuadrático medio supone una medida de la variabilidad del estimador  $T$  alrededor del valor de  $\theta$ , similar a la medida de dispersión alrededor de la media que proporciona la varianza. Así, el que un estimador tenga un error cuadrático medio pequeño, informa de que su distribución está muy concentrada alrededor de

la cantidad a estimar. Si las condiciones del problema no imponen otra función de pérdida, es muy razonable tomar, como grado de preferencia entre estimadores, la medida de lo concentrada que es su distribución alrededor del valor del parámetro.

Sin embargo, cualitativamente la situación con el error cuadrático medio es la misma que con cualquier otra función de riesgo. Ello queda patente en el ejemplo siguiente:

### Ejemplo 5.3

Una máquina produce piezas, cada una de las cuales independientemente puede ser defectuosa con probabilidad  $\theta \in (0, 1)$  desconocida. A fin de estimar  $\theta$ , se pueden calibrar  $n$  piezas anotando

$$X_i = \begin{cases} 1 & \text{si la pieza } i\text{-ésima es defectuosa,} \\ 0 & \text{en caso contrario.} \end{cases}$$

Bajo la hipótesis de que la pérdida de estimación es  $L(\theta, t) = (\theta - t)^2$ , se desea comparar el estimador  $T_1 = \bar{X}$  (que proporciona la frecuencia de piezas defectuosas en la muestra) con sus múltiplos:  $T_k = k\bar{X}$ , con  $k > 0$ .

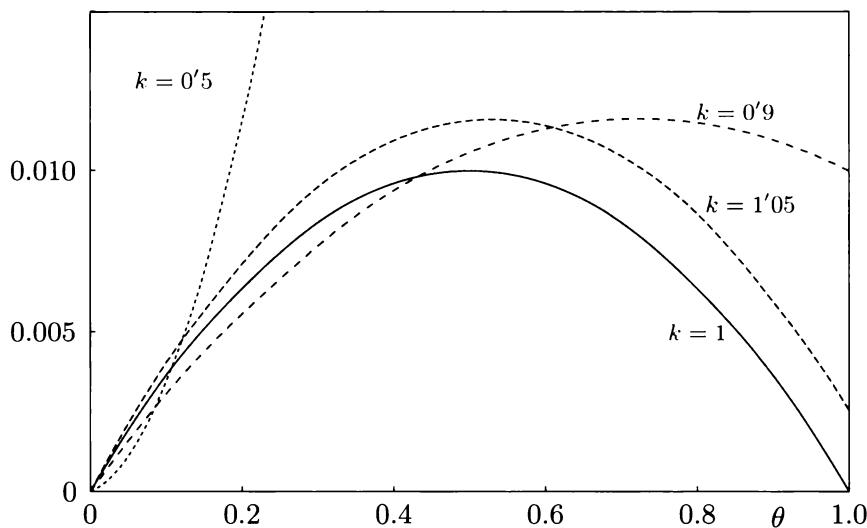
Como  $n\bar{X}$  tiene distribución binomial  $B(n, \theta)$ , será

$$\mathbb{E}_\theta[\bar{X}] = \theta \quad \text{y} \quad \mathbb{E}_\theta[(\bar{X} - \theta)^2] = \frac{\theta(1 - \theta)}{n}.$$

Por tanto, el error cuadrático medio del estimador  $T_k$  es

$$\begin{aligned} ECM_{T_k}(\theta) &= \mathbb{E}_\theta[(k\bar{X} - \theta)^2] = \mathbb{E}_\theta[(k\bar{X} - k\theta + k\theta - \theta)^2] \\ &= k^2\mathbb{E}_\theta[(\bar{X} - \theta)^2] + 2k\mathbb{E}_\theta[\bar{X} - \theta](k - 1)\theta + (k - 1)^2\theta^2 \\ &= k^2 \frac{\theta(1 - \theta)}{n} + (k - 1)^2\theta^2 \\ &= \left[(1 - k)^2 - \frac{k^2}{n}\right]\theta^2 + \frac{k^2}{n}\theta. \end{aligned}$$

Como función de  $\theta$ , el riesgo  $ECM_{T_k}(\theta)$  es una parábola cuya pendiente en el origen,  $k^2/n$ , crece con  $k$  y que vale  $(1 - k)^2$  para  $\theta = 1$ . Esquemáticamente, para  $n = 25$ :

Figura 5.2: Riesgo de  $T_k$ 

Como en el ejemplo anterior el estimador  $T_1$  es preferible a todos los estimadores  $T_k$  con  $k > 1$  (que además pueden producir estimaciones superiores a 1); pero entre los estimadores  $T_k$  con  $k \in [0, 1]$  ninguno es preferible a los demás.

Ello puede verse de otra forma: el valor mínimo de  $ECM_{T_k}(\theta)$  se alcanza para  $k = n\theta/(n\theta + 1 - \theta)$ , que es un valor de  $[0, 1]$  que depende de  $\theta$ ; por tanto el estimador  $T_k$  que minimiza el error cuadrático medio es uno distinto para cada valor de  $\theta$ .

A primera vista, los dos últimos ejemplos parecen indicar que el criterio de preferencia establecido no es muy útil: permite eliminar algunos estimadores —llamados por ello *inadmisibles*—, pero gran número de estimadores resultan incomparables; de manera que no se puede seleccionar uno preferido a todos los demás. De hecho se podía haber previsto la dificultad, cayendo en la cuenta de que ningún estimador podrá dar menos riesgo, para el valor  $\theta_0$  del parámetro, que el estimador constante  $T = \theta_0$ ; el cual es ciertamente indeseable para valores del parámetro lejanos de  $\theta_0$ . No tiene pues sentido buscar un estimador preferible a todos los demás, ya que ninguno podrá ser preferido a los pésimos estimadores constantes.

Puestas así las cosas, el criterio de preferencia a través del riesgo ha de

ser complementado de alguna forma, si se quiere que proporcione una regla de conducta para la selección de estimadores. Las soluciones sólo pueden ser de dos tipos:

- (1) Restringir la clase de los estimadores considerados a aquellos que cumplan alguna propiedad adicional de interés, tratando con ello de eliminar los estimadores indeseables y examinando si el criterio del mínimo riesgo selecciona entre ellos uno preferible a todos los demás.

En la próxima sección se examinarán las propiedades deseables para un buen estimador (*ausencia de sesgo, consistencia, invarianza*) y en el capítulo siguiente se analizará cuáles combinan adecuadamente con el criterio del mínimo riesgo. En esta dirección, aunque no sea directamente una de las propiedades deseable para un estimador, el concepto de *suficiencia*, que se analiza en la sección 5.4, juega un papel importante.

- (2) Reforzar el criterio de preferencia de estimadores, mediante el procedimiento de reducir toda la función de riesgo  $R_T(\theta)$  a un único número representativo  $r_T$ , que permita ordenar linealmente todos los estimadores, para tratar de seleccionar el mejor.

En esta línea destacan principalmente los criterios *Bayes* y *minimax*, que serán estudiados en la sección 5.5.

Cabría en principio una tercera posibilidad: hallar el subconjunto de todos los *estimadores admisibles* (aquellos para los cuales no existe otro estimador preferible a él) y seleccionar entre ellos aquél que tuviese propiedades adicionales deseables. Sin embargo, este procedimiento no es operativo, debido a la dificultad de enumerar todos los estimadores admisibles. Sólo se cita para señalar que no produce en general el mismo resultado que la solución (1), como habrá ocasión de ver en algún ejemplo (cf. Ejemplos 5.4 y 6.8).

Hasta ahora, hemos pensado en estimar el propio parámetro  $\theta$  con el que se identifica la distribución poblacional. Con frecuencia, sin embargo, el propósito no será estimar directamente  $\theta$ , sino una cierta característica poblacional (media, varianza, mediana o cualquier otra cosa) que se expresa mediante una determinada función  $g(\theta)$ . A menudo,  $g(\theta)$  no será una función biyectiva de  $\theta$ , de forma que no es posible reparametrizar el problema en términos del parámetro  $\omega = g(\theta)$  (puesto que éste no identifica la distribución teórica). Tampoco hay que precipitarse a pensar que la mejor

estimación posible de  $g(\theta)$  se obtendrá, siempre, sustituyendo en  $g$  la mejor estimación posible de  $\theta$ . Así que no queda más remedio que generalizar ligeramente las ideas anteriores.

Por ejemplo, el error cuadrático medio de  $T$  como estimador de  $g(\theta)$  será

$$E_\theta[|g(\theta) - T(X_1, \dots, X_n)|^2]$$

que mide la dispersión de  $T$  alrededor de la cantidad a estimar  $g(\theta)$ , de la misma manera que el error cuadrático medio (sin más precisión) mide la dispersión alrededor de  $\theta$ . La misma extensión puede realizarse para cualquier otra función de pérdida que proporcione el coste  $L(g(\theta), t)$  de dar  $t$  como estimación de  $g(\theta)$ .

Aunque el estimador  $T$  minimizase  $E_\theta[|\theta - T|^2]$ , no hay ninguna garantía de que el mínimo de  $E_\theta[|g(\theta) - T'|^2]$  se alcance con  $T' = g(T)$ . De forma que los criterios relacionados con el riesgo no hacen que el problema de estimación sea invariante por transformaciones del parámetro y, al utilizarlos, el dato de la función  $g(\theta)$  que se desea estimar resulta fundamental.

## 5.3. Propiedades deseables de los estimadores

Al margen de las relaciones de preferencia que pueda haber entre los estimadores, existen propiedades globales que pueden resultar convenientes para un “buen” estimador.

### 5.3.1. Estimadores insesgados

Puesto que la esperanza de una variable aleatoria es una medida de la posición de su distribución, una manera simple de juzgar si la distribución en el muestreo de un estimador  $T(X_1, X_2, \dots, X_n)$  es “próxima” a la función del parámetro que se desea estimar, es comparar su media con dicha función. Concretamente:

*Si  $E_\theta[T]$  representa la media de la distribución en el muestreo del estadístico  $T$ , la diferencia*

$$b_T(\theta) = E_\theta[T] - g(\theta)$$

*se denomina sesgo del estimador  $T$  como estimador de  $g(\theta)$ . En caso de que el sesgo sea idénticamente nulo, es decir si*

$$E_\theta[T] = g(\theta) \quad \text{para cada } \theta \in \Theta,$$

*el estimador  $T$  se dice insesgado o centrado para  $g(\theta)$ .*

Por ejemplo en la sección 2.4 se ha observado que la media de la distribución en el muestreo del estadístico  $\bar{X}$  —media muestral— coincide siempre con la media poblacional; en la nueva terminología, ello significa que  $\bar{X}$  es siempre un estimador insesgado de la media poblacional.

En cambio, también resultó allí que la varianza muestral  $s^2$ , como estimador de la varianza poblacional  $\sigma^2$ , tiene como sesgo:

$$b_{s^2} = E[s^2] - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{1}{n}\sigma^2,$$

razón por la cual se introdujo la cuasivarianza muestral  $S^2 = \frac{ns^2}{n-1}$  que es un estimador centrado de  $\sigma^2$ .

Más ilustrativas son las situaciones siguientes:

#### Ejemplo 5.4

En el ejemplo 5.1 se ha considerado, como estimador natural del parámetro  $\theta$  de una población uniforme en  $(0, \theta)$ , el estadístico  $T_1 = \max_{1 \leq i \leq n} X_i$ . Sin embargo,  $T_1 < \theta$ , así que siempre dará una estimación por defecto de  $\theta$ ; en particular

$$E_\theta[T_1] = \frac{n}{n+1} \theta,$$

de forma que  $T_1$  tiene sesgo:

$$b_{T_1}(\theta) = \frac{n}{n+1} \theta - \theta = -\frac{1}{n+1} \theta.$$

Parece entonces lógico, corregir al alza el estimador  $T_1$  y una posibilidad es considerar  $T'_1 = (n+1)T_1/n$  que verifica

$$E_\theta[T'_1] = \frac{n+1}{n} E_\theta[T_1] = \theta$$

y es, por tanto, un estimador insesgado de  $\theta$ .

Obsérvese que  $T'_1$  no es admisible en el sentido del error cuadrático medio, puesto que es preferible el estimador de mínimo error cuadrático medio determinado en el ejemplo 5.1:  $T''_1 = (n+2)T_1/(n+1)$ . Este, a su vez, no es centrado, pues

$$E_\theta[T''_1] = \frac{n+2}{n+1} E_\theta[T_1] = \frac{(n+2)n}{(n+1)^2} \theta,$$

de forma que  $b_{T''_1}(\theta) = -\frac{1}{(n+1)^2} \theta$ .

### Ejemplo 5.5

Una población con distribución exponencial de parámetro  $\theta$  tiene como media poblacional  $1/\theta$ . Para estimar el valor de  $\theta$ , a partir de una muestra aleatoria simple  $(X_1, X_2, \dots, X_n)$ , puede considerarse entonces el estimador  $T = 1/\bar{X}$ .

La distribución en el muestreo de  $n\bar{X} = X_1 + X_2 + \dots + X_n$  es una distribución gamma  $\gamma(n, \theta)$  de densidad  $\theta^n z^{n-1} e^{-\theta z}/(n-1)!$  para  $z > 0$ . Luego

$$\begin{aligned}\mathrm{E}_\theta[T] &= E_\theta\left[\frac{1}{\bar{X}}\right] = n\mathrm{E}_\theta\left[\frac{1}{n\bar{X}}\right] \\ &= \frac{n}{(n-1)!} \int_0^\infty \frac{1}{z} \theta^n z^{n-1} e^{-\theta z} dz \\ &= \frac{n\theta^n}{(n-1)!} \frac{(n-2)!}{\theta^{n-1}} = \frac{n}{n-1} \theta\end{aligned}$$

de forma que, para estimar  $\theta$ ,  $T$  tiene sesgo

$$b_T(\theta) = \frac{n}{n-1} \theta - \theta = \frac{1}{n-1} \theta.$$

En cambio

$$T' = \frac{n-1}{n} T = \frac{n-1}{n\bar{X}}$$

es un estimador insesgado de  $\theta$  pues

$$E_\theta[T'] = \frac{n-1}{n} E_\theta[T] = \theta.$$

La utilización de estimadores insesgados, o al menos de pequeño sesgo, es muy conveniente en el caso de muestras de tamaño elevado. La razón es que, con gran frecuencia, la varianza  $V_\theta(T)$  de un estimador  $T$ , basado en un gran número de observaciones, es muy pequeña (véase Cap. 2), y por tanto su distribución en el muestreo está muy concentrada alrededor de su media  $\mathrm{E}_\theta[T]$ . Si ésta fuese  $g(\theta) + b_T(\theta)$  en lugar de  $g(\theta)$ , habría probabilidad cercana a uno de obtener un valor del estimador próximo a  $g(\theta) + b_T(\theta)$  en vez de a  $g(\theta)$ .

Desde el punto de vista del error cuadrático medio, se puede dar una explicación más detallada: Dado que

$$ECM_T(\theta) = \mathrm{E}_\theta[(T - \mathrm{E}_\theta[T])^2] + (\mathrm{E}_\theta[T] - g(\theta))^2 = V_\theta(T) + b_T^2(\theta) \quad (5.1)$$

lo ideal, para estimar  $g(\theta)$ , sería minimizar a la vez la varianza y el sesgo del estimador (para todos los valores de  $\theta$  simultáneamente). Puede ocurrir, como en el ejemplo 5.4, que sea mejor elevar ligeramente el sesgo para obtener una mayor disminución de la varianza; sin embargo, en general, no es mala estrategia anular el sesgo —es decir, limitarse a considerar estimadores centrados— y determinar entre ellos el de mínima varianza. Sobre todo porque el problema de determinar el estimador centrado de mínima varianza (uniformemente en  $\theta$ ) admite una solución constructiva, bastante completa y satisfactoria, que se expondrá en el próximo capítulo.

En el caso de muestras de tamaño pequeño, los estimadores insesgados pierden gran parte de su justificación, ya que lo habitual es entonces que la varianza de los estimadores sea grande y el sesgo sea despreciable frente a ella, en lugar de tener un papel predominante como ocurre cuando la varianza es pequeña.

Además, la propiedad de ausencia de sesgo no está exenta de anomalías; puede haber ocasiones en que no exista ningún estimador centrado, como ocurre en la situación siguiente.

### Ejemplo 5.6

Se ofrece un premio por obtener dos caras, al lanzar dos veces una moneda con probabilidad de cara,  $p$ , desconocida; el jugador está entonces interesado en estimar  $p^2$ , para saber cuánto debe estar dispuesto a pagar por participar en el juego. Sin embargo sólo se le permite realizar un lanzamiento de prueba, antes de aceptar las condiciones impuestas.

Sólo podrá entonces observar el valor de

$$X_1 = \begin{cases} 1 & \text{si obtiene cara} \\ 0 & \text{si obtiene cruz} \end{cases}$$

y cualquier estimador  $T$ , que piense en utilizar, quedará determinado por los valores  $T(1), T(0) \in (0, 1)$  para cada una de las dos únicas muestras posibles. Entonces

$$\mathbf{E}_p[T] = T(1)p + T(0)(1 - p)$$

lo cual es imposible que coincida con  $p^2$ . No hay pues ningún estimador centrado de  $p^2$ .

Por supuesto,  $T = X_1$  es un estimador insesgado de  $p$ , pero eso no resuelve el problema de dar una estimación insesgada de  $p^2$ , sino que pone de relieve otro inconveniente de los estimadores centrados:  $T$  puede ser estimador centrado de  $\theta$  sin que  $g(T)$  lo sea de  $g(\theta)$ . Hecho que ya se podía haber

observado en el ejemplo 5.5, en el cual  $1/\bar{X}$  no es estimador insesgado de  $\theta$  aunque  $\bar{X}$  lo sea de la media poblacional  $1/\theta$ .

He aquí otra situación clásica en la cual el empeño de utilizar estimadores centrados conduce a resultados absurdos.

### Ejemplo 5.7

El número de llamadas telefónicas a una centralita durante una hora, sigue una distribución de Poisson, de parámetro  $\theta > 0$  desconocido. Después de observar el número  $X_1$  de llamadas recibidas durante una única hora, se desea estimar:  $e^{-3\theta}$ .

Dado cualquier estimador  $T(X_1)$ , será

$$E_\theta[T] = e^{-\theta} \sum_{n=0}^{\infty} \frac{\theta^n}{n!} T(n)$$

y si  $T$  ha de ser centrado para  $e^{-3\theta}$ , debe ser

$$\sum_{n=0}^{\infty} \frac{\theta^n}{n!} T(n) = e^{-2\theta} = \sum_{n=0}^{\infty} (-2)^n \frac{\theta^n}{n!}$$

lo cual obliga a tomar  $T(X_1) = (-2)^{X_1}$ , que es negativo en el caso de que  $X_1$  fuese impar. Habría que concluir por tanto que no hay ningún estimador insesgado o dar para  $e^{-3\theta} \in (0, 1)$  estimaciones fuera de  $(0, 1)$ .

### 5.3.2. Estimadores consistentes

Con frecuencia los problemas de estimación no precisan un tamaño muestral determinado, sino que han de ser resueltos en función de  $n$ , proponiendo una sucesión de estimadores  $T_n$ , formalmente distintos puesto que se aplican en espacios muestrales de distinta dimensión, pero comúnmente con la misma expresión funcional:

$$\frac{1}{n} \sum_1^n X_i, \quad \max_{1 \leq i \leq n} X_i, \quad \text{etc.}$$

La consistencia de los estimadores hace referencia a su comportamiento cuando el tamaño muestral  $n$  crece hacia infinito. Concretamente:

Una sucesión  $T_n$  de estimadores, asociada a los sucesivos tamaños muestrales  $n$ , se denomina consistente para estimar una función  $g(\theta)$  del parámetro poblacional si, cuando  $n \rightarrow \infty$ , se verifica

$$T_n \xrightarrow{\mathcal{P}_\theta} g(\theta)$$

cualquiera que sea  $\theta \in \Theta$ . O bien explícitamente, si

$$\mathcal{P}_\theta\{|T_n - g(\theta)| > \varepsilon\} \rightarrow 0$$

para todo  $\varepsilon > 0$  y todo  $\theta \in \Theta$ .

El concepto de consistencia admite versiones más estrictas si se sustituye la convergencia en probabilidad ( $\mathcal{P}_\theta$ ) utilizada, por convergencia casi segura o en media cuadrática:

$$T_n \xrightarrow{\mathcal{P}_\theta-c.s.} g(\theta) \quad \text{y} \quad \mathbb{E}_\theta[|T_n - g(\theta)|^2] \rightarrow 0$$

respectivamente<sup>1</sup>. Como se sabe, la consistencia casi segura ó en media cuadrática implican la consistencia en el sentido inicial.

La consistencia en media cuadrática significa que, para estimar  $g(\theta)$ ,  $ECM_{T_n}(\theta) \rightarrow 0$  para cualquier  $\theta \in \Theta$ ; o bien, según (5.1), para cada  $\theta \in \Theta$

$$V_\theta(T_n) \rightarrow 0 \quad \text{y} \quad b_{T_n}(\theta) \rightarrow 0.$$

Así que, desde el punto de vista del error cuadrático medio, es ciertamente deseable, si el tamaño muestral es grande, emplear un estimador consistente en media cuadrática.

El capítulo 2 contiene múltiples ejemplos de sucesiones consistentes de estimadores para distintas características poblacionales (que serán determinadas funciones del parámetro  $\theta$  que identifica la distribución teórica). Así, con probabilidad 1:

$$a_k(n) \rightarrow \alpha_k(\theta), \quad b_k(n) \rightarrow \mu_k(\theta), \quad F_n^*(x) \rightarrow F_\theta(x), \quad \text{etc.}$$

lo cual garantiza la consistencia (casi segura) de cada momento muestral, como estimador del correspondiente momento poblacional, o la consistencia

---

<sup>1</sup>En el primer caso,  $\mathcal{P}_\theta$  representa la distribución de la sucesión muestral:  $X_1, X_2, \dots, X_n, \dots$ , y no la de una muestra de tamaño  $n$ ; ya que la convergencia casi segura no sólo depende de la distribución de cada  $T_n$ , sino de la distribución conjunta de todos ellos.

(casi segura) de los valores de la función de distribución muestral, como estimador del correspondiente valor de la función de distribución poblacional.

Veamos una situación diferente:

### Ejemplo 5.8

La sucesión de estimadores  $X_{(n)} = \max_{1 \leq i \leq n} X_i$  es consistente para estimar el parámetro  $\theta$  de una distribución uniforme en  $(0, \theta)$ , ya que

$$\mathcal{P}_\theta \left\{ \left| \max_{1 \leq i \leq n} X_i - \theta \right| > \varepsilon \right\} = \mathcal{P}_\theta \left\{ \max_{1 \leq i \leq n} X_i < \theta - \varepsilon \right\} = \left(1 - \frac{\varepsilon}{\theta}\right)^n \rightarrow 0$$

cuando  $n \rightarrow \infty$ , para cualquier  $\varepsilon < \theta$ . (Y si  $\varepsilon \geq \theta$ , la probabilidad es cero.)

Por tanto

$$\max_{1 \leq i \leq n} X_i \xrightarrow{\mathcal{P}_\theta} \theta.$$

Ahora bien, como  $X_{(n)}$  es una sucesión creciente con  $n$ , se tiene de hecho

$$X_{(n)} \xrightarrow{\mathcal{P}_\theta-c.s.} \theta.$$

Por otra parte, según los resultados del ejemplo 5.1,

$$\mathbb{E}_\theta \left[ (\theta - X_{(n)})^2 \right] = \left( 1 - \frac{2n}{n+1} + \frac{n}{n+2} \right) \theta^2$$

converge a cero, cuando  $n$  tiende a infinito, de forma que  $X_{(n)}$  es consistente en media cuadrática para estimar  $\theta$ .

La consistencia de un estimador (dependiente del tamaño muestral) es un requisito que no presta mucha ayuda en la selección de buenos estimadores, puesto que se trata de una propiedad asintótica que no depende de sus cualidades para ningún  $n$  fijo; sin embargo generalmente se considera una propiedad indispensable para cualquier método de estimación satisfactorio. De hecho, si el teorema de Glivenko-Cantelli garantiza la convergencia uniforme casi segura de la distribución muestral a la distribución teórica, no hay razón para admitir que la estimación de un parámetro no sea tan precisa como se deseé para tamaños muestrales suficientemente grandes.

### 5.3.3. Estimadores invariantes

Las ideas de invariancia pueden entenderse mejor examinando algunas situaciones concretas.

**(a)** Una empresa llena los envases de su producto con un peso  $\theta$ , más una cantidad adicional cuyo peso se elige uniformemente entre 0 y 100 gramos; de forma que el contenido tiene un peso uniformemente distribuido en el intervalo  $(\theta, \theta + 100)$ . Para comprobar las especificaciones del fabricante acerca del peso mínimo contenido en los envases, se ha pesado el contenido de  $n$  de ellos y los resultados  $(x_1, x_2, \dots, x_n)$  han conducido a dar, como estimación puntual de  $\theta$ , el valor  $\theta_1$  gramos. Sin embargo, poco después se descubre que la balanza utilizada para el estudio pesaba sistemáticamente  $c$  gramos de menos. Parece obligatorio entonces corregir la estimación de  $\theta$  al valor  $\theta_1 + c$ ; pero también cabe considerar la muestra corregida  $(x_1 + c, x_2 + c, \dots, x_n + c)$  y volver a aplicar el método de estimación utilizado. Ambos resultados serán coherentes si se ha usado un *estimador invariante por translaciones*, en el sentido de que

$$T(x_1 + c, x_2 + c, \dots, x_n + c) = T(x_1, x_2, \dots, x_n) + c$$

para toda muestra  $(x_1, x_2, \dots, x_n)$  y cualquier  $c \in \mathbb{R}$ . Tal es el caso de los estimadores

$$X_{(1)} = \min_{1 \leq i \leq n} X_i, \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{o} \quad \frac{1}{2} (X_{(1)} + X_{(n)})$$

pero la condición se incumple con la media geométrica:  $(\prod_{i=1}^n X_i)^{1/n}$ , o con la media armónica:  $n / \sum_{i=1}^n X_i^{-1}$ .

**(b)** Una persona tiene que esperar diariamente su autobús durante un tiempo uniformemente distribuido en el intervalo  $(0, \theta)$ . Después de cronometrar su tiempo de espera durante  $n$  días, dispone de una muestra aleatoria simple  $(x_1, x_2, \dots, x_n)$ , expresada en segundos, con la cual estima el valor de  $\theta$  en  $\theta_1$  segundos. Si desea expresar su estimación en otras unidades (minutos, horas, etc...), sólo tiene que multiplicar  $\theta_1$  por el factor de conversión adecuado,  $b > 0$ . Sin embargo, puede traducir la muestra a las nuevas unidades:  $(bx_1, bx_2, \dots, bx_n)$  y aplicar de nuevo el método de estimación. Para que ambos resultados sean coherentes, el estimador utilizado deber ser *invariante por cambios de escala*:

$$T(bx_1, bx_2, \dots, bx_n) = bT(x_1, x_2, \dots, x_n)$$

para cualquier muestra  $(x_1, x_2, \dots, x_n)$  y cualquier  $b > 0$ . Tal es el caso de  $\bar{X}$ ,  $X_{(n)} = \max_{1 \leq i \leq n} X_i$ , etc., pero no del estimador  $\log(\frac{1}{n} \sum_{i=1}^n e^{X_i})$  ni del estimador  $X_{(n)}/X_{(1)}$ .

Los ejemplos anteriores muestran la conveniencia, en determinadas circunstancias, de utilizar estimadores invariantes frente a ciertos grupos de transformaciones de la muestra. Es más, es inherente a cada situación, el tipo de invariancia que le resulta adecuado; así en el supuesto (a), la escala está fijada por la amplitud constante (100 gr.) de la distribución y un cambio de unidades de medida afectaría a la estructura del modelo, cambiando las distribuciones uniformes en  $(\theta, \theta + 100)$  por las distribuciones uniformes en  $(\theta, \theta + 100b)$ . En cambio, los errores sistemáticos de la balanza, sólo afectan al valor de  $\theta$ , trasladando la distribución al intervalo  $(\theta', \theta' + 100)$ , con  $\theta' = \theta + c$ .

Inversamente, en el supuesto (b), el origen está fijado en el tiempo de espera cero y, un cambio de unidades afecta exclusivamente al parámetro  $\theta$ , sustituyendo el intervalo  $(0, \theta)$  por el  $(0, \theta')$ , con  $\theta' = b\theta$ . En cambio una translación modifica el modelo introduciendo la distribución uniforme en  $(c, \theta + c)$ .

De manera general, en un problema estadístico caracterizado por la familia de distribuciones continuas y unidimensionales de la población  $\mathcal{F} = \{F_\theta | \theta \in \Theta\}$ , para que el uso de estimadores invariantes por translaciones sea un requisito exigible, debe ser  $\Theta = \mathbb{R}$  y la propia familia  $\mathcal{F}$  ha de ser *invariante por translaciones*, en el sentido de que

$$F_\theta(x) = F_{\theta+c}(x+c) \quad \text{para cada } c \in \mathbb{R},$$

o bien, en términos de las funciones de densidad,

$$f_\theta(x) = f_{\theta+c}(x+c) \quad \text{para cada } c \in \mathbb{R}.$$

Más claramente, las condiciones anteriores se pueden expresar (haciendo  $c = -\theta$ ) en la forma

$$F_\theta(x) = F_0(x - \theta),$$

lo cual indica que la variable aleatoria, que describe la población, se puede escribir:

$$X = \theta + Z \text{ siendo } Z \text{ una variable aleatoria con distribución fija } F_0.$$

Un cambio de origen:  $X' = c + X$ , sólo supone entonces una variación del parámetro, sin afectar a la “forma” de la distribución, que determina  $F_0$ . Se dice por tanto que  $\{F_\theta | \theta \in \Theta\}$  depende del *parámetro de posición*  $\theta$ .

En cambio los estimadores invariantes por cambios de escala son adecuados cuando  $\Theta = (0, \infty)$  y la propia familia  $\mathcal{F} = \{F_\theta | \theta \in \Theta\}$  es *invariante*

por cambios de escala; esto es

$$F_\theta(x) = F_{b\theta}(bx) \quad \text{para cada } b > 0,$$

lo que se traduce para la densidad en

$$f_\theta(x) = b f_{b\theta}(bx) \quad \text{para cada } b > 0.$$

Haciendo  $b = 1/\theta$ , las condiciones anteriores resultan equivalentes a

$$F_\theta(x) = F_1\left(\frac{x}{\theta}\right) \quad \text{o} \quad f_\theta(x) = \frac{1}{\theta} f_1\left(\frac{x}{\theta}\right),$$

lo cual significa que la variable aleatoria  $X$ , que describe la población, es de la forma

$$X = \theta Z \quad \text{siendo } Z \text{ una variable con distribución fija } F_1.$$

Ahora es un cambio de unidades:  $X' = bX$ , el que es absorbido por el parámetro  $\theta$ , sin influir en  $F_1$ . Se dice entonces que  $\mathcal{F}$  depende del parámetro de escala  $\theta$ .

Muy frecuentemente se presentan situaciones mixtas: familias

$$\{F_{\theta_1, \theta_2} \mid \theta_1 \in \mathbb{R}, \theta_2 > 0\}$$

cuyos parámetros absorben conjuntamente el efecto de cualquier cambio de origen y de escala. Concretamente, si la variable poblacional  $X$  se puede expresar en la forma:

$$X = \theta_1 + \theta_2 Z \quad \text{siendo } Z \text{ una variable con distribución fija } F,$$

o, dicho de otra manera, si la familia de distribuciones verifica:

$$F_{\theta_1, \theta_2}(x) = F\left(\frac{x - \theta_1}{\theta_2}\right) \quad \text{o bien} \quad f_{\theta_1, \theta_2}(x) = \frac{1}{\theta_2} f\left(\frac{x - \theta_1}{\theta_2}\right),$$

entonces cualquier transformación lineal de las observaciones,  $X' = c + bX$ , repercute sólo en el cambio de  $\theta_2$  por  $b\theta_2$  y de  $\theta_1$  por  $c + b\theta_1$ .

La familia  $\{F_{\theta_1, \theta_2}(x) \mid \theta_1 \in \mathbb{R}, \theta_2 > 0\}$  se dice entonces *invariante por cambios de origen y de escala* o bien, dependiente de sendos parámetros de posición ( $\theta_1$ ) y de escala ( $\theta_2$ ). (Véase ejercicio 5.3.)

En tales situaciones, el *principio de invariancia* aconseja que las estimaciones de  $\theta_1$  y  $\theta_2$  se transformen coherentemente al cambiar el sistema

de referencia en el que se expresan los elementos de la muestra; es decir, emplear en la estimación de  $\theta_2$  estimadores que verifiquen

$$T_2(c + bx_1, c + bx_2, \dots, c + bx_n) = bT_2(x_1, x_2, \dots, x_n)$$

y en la estimación de  $\theta_1$ , estimadores con la condición

$$T_1(c + bx_1, c + bx_2, \dots, c + bx_n) = c + bT_1(x_1, x_2, \dots, x_n).$$

Bajo condiciones de invariancia, veremos en el próximo capítulo que el criterio del mínimo riesgo es capaz de seleccionar un estimador invariante preferible a todos los demás, supuesto que la función de pérdida se elige adecuadamente para no destruir la invariancia del problema.

## 5.4. Estadísticos suficientes

En cualquier problema de Inferencia paramétrica, al observar los valores de una muestra aleatoria, puede ocurrir que el resultado contenga sistemáticamente información superflua, no relevante a la hora de disminuir el desconocimiento del parámetro. Sería importante entonces, saber eliminar la información superflua, para poder simplificar el proceso de selección del estimador puntual adecuado (o del estadístico conveniente para construir un intervalo de confianza), limitándolo al conjunto de los que utilizan únicamente la información relevante.

La idea puede quedar más clara en una situación concreta.

### Ejemplo 5.9

Una urna contiene una proporción desconocida,  $\theta$ , de bolas marcadas con 1 y el resto marcadas con 0. Para estimar  $\theta$  pueden extraerse  $n$  bolas con reemplazamiento, obteniéndose una muestra aleatoria simple  $(X_1, X_2, \dots, X_n)$  de la población:

$$X = \begin{cases} 1 & \text{con probabilidad } \theta, \\ 0 & \text{con probabilidad } 1 - \theta. \end{cases}$$

Intuitivamente es claro que lo único que interesa de la muestra, para tratar de precisar el valor de  $\theta$ , es el número total de unos y ceros obtenidos; o sea, el valor del estadístico  $T = \sum_{i=1}^n X_i$ . Por el contrario, no parece lógico, cuando se conoce  $T$ , preocuparse por el resultado que se obtuvo en la

5<sup>a</sup> extracción, ni de si se obtuvieron alguna vez 3 unos consecutivos. Es decir, no parecen sensatos procedimientos de estimación que tengan en cuenta la posición de los unos y los ceros dentro de la muestra; y los únicos estimadores que no caen en tal defecto son los que dependen de  $T$  exclusivamente:  $T/n$ , por ejemplo.

Tratando de justificar esta peculiaridad, intuitivamente clara, del problema, puede observarse que todas las muestras de tamaño  $n$ , con el mismo número  $t$  de unos, tienen la misma probabilidad. Concretamente, la función de probabilidad de la muestra es

$$f_\theta(x_1, x_2, \dots, x_n) = \theta^t(1 - \theta)^{n-t}$$

cualquiera que sean  $x_1, x_2, \dots, x_n \in \{0, 1\}$ , siempre que  $\sum_{i=1}^n x_i = t$ , e independientemente por tanto del orden de los ceros y los unos en la muestra.

Por consiguiente, dado que  $T$  tiene distribución binomial  $B(n, \theta)$ ,

$$\begin{aligned} \mathcal{P}_\theta\{X_1 = x_1, \dots, X_n = x_n \mid T = t\} &= \\ &= \begin{cases} \frac{\theta^t(1 - \theta)^{n-t}}{\binom{n}{t} \theta^t(1 - \theta)^{n-t}} = \binom{n}{t}^{-1} & \text{si } \sum_{i=1}^n x_i = t, \\ 0 & \text{si } \sum_{i=1}^n x_i \neq t. \end{cases} \end{aligned}$$

lo cual indica que, sabiendo que  $T = t$ , la muestra obtenida tiene la misma probabilidad  $\binom{n}{t}^{-1}$ , de coincidir con cada una de las  $\binom{n}{t}$  muestras que constan de  $t$  unos y  $n - t$  ceros.

Se puede imaginar entonces la muestra construida en dos fases: primero se elige el valor  $t$  del estadístico  $T$ , con distribución  $B(n, \theta)$  y después se sitúan al azar  $t$  unos y  $n - t$  ceros en las  $n$  posiciones de la muestra. Obviamente el valor  $\theta$  influye en la realización de la primera fase, pero no afecta al resultado de la segunda: de manera que observar el reparto de los unos y los ceros no indicará nada acerca de  $\theta$ .

Lo más característico de la situación anterior es el hecho de que la distribución de la muestra condicionada por el valor del estadístico, sea independiente de  $\theta$ . Con poblaciones discretas, siempre se puede descomponer la función de probabilidad de la muestra en la forma:

$$\mathcal{P}_\theta\{X_1 = x_1, \dots, X_n = x_n\} = \mathcal{P}_\theta\{T = t\}\mathcal{P}_\theta\{X_1 = x_1, \dots, X_n = x_n \mid T = t\}$$

cualquiera que sea  $T$  y supuesto que  $T(x_1, \dots, x_n) = t$ . El primer factor rige la obtención del valor de  $T$  y el segundo, la elección de una muestra concreta que proporcione dicho valor a  $T$ . Cuando el estadístico  $T$  logre que el segundo factor no dependa de  $\theta$ , la observación adicional de la muestra concreta es inútil, pues no puede conseguirse información sobre  $\theta$  observando algo independiente de  $\theta$ . Se dice entonces que el estadístico  $T$  es *suficiente* (para la estimación), como indica la siguiente definición:

*Dada una muestra aleatoria  $(X_1, X_2, \dots, X_n)$  de una población con distribución teórica perteneciente a una familia paramétrica  $\{F_\theta \mid \theta \in \Theta\}$ , un estadístico  $T(X_1, X_2, \dots, X_n)$  se denomina suficiente si la distribución de la muestra, condicionada por el valor del estadístico  $T$ , no depende de  $\theta$ .*

Por supuesto, la definición no requiere que las distribuciones  $F_\theta$  sean discretas (a diferencia de las consideraciones introductorias). Tampoco es necesario que la muestra sea aleatoria simple (aunque nuestro principal interés se centre en este caso). Pero la diferencia más sustancial con el ejemplo de introducción es que, en general, el estadístico  $T$  puede ser  $k$ -dimensional:  $T = (T_1, \dots, T_k)$ . Un caso no trivial, en que es así, es el siguiente:

### Ejemplo 5.10

Una urna contiene  $N$  bolas numeradas correlativamente de  $N + 1$  a  $2N$ ; como se desconoce el valor de  $N$ , para estimarlo se pueden extraer  $n$  bolas, con reemplazamiento, que darán lugar a una muestra aleatoria simple  $(X_1, X_2, \dots, X_n)$  de la distribución uniforme en  $\{N + 1, \dots, 2N\}$ . La función de probabilidad de la muestra será:

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

cualquiera que sean  $x_1, \dots, x_n \in \{N + 1, \dots, 2N\}$ .

Intuitivamente es claro que, si se conocen el menor y el mayor de los valores observados, los valores intermedios que contenga la muestra no suministran ninguna información sobre  $N$ . Para tratar de justificar esta impresión, analicemos las distribuciones de

$$X_{(1)} = \min_{1 \leq i \leq n} X_i \quad \text{y} \quad X_{(n)} = \max_{1 \leq i \leq n} X_i.$$

Por un lado,

$$\begin{aligned}\mathcal{P}_N\{X_{(1)} = y\} &= \mathcal{P}_N\{X_{(1)} \geq y\} - \mathcal{P}_N\{X_{(1)} \geq y + 1\} \\ &= \left(\frac{2N-y+1}{N}\right)^n - \left(\frac{2N-y}{N}\right)^n\end{aligned}$$

para  $y = N+1, N+2, \dots, 2N$ . Análogamente

$$\begin{aligned}\mathcal{P}_N\{X_{(n)} = z\} &= \mathcal{P}_N\{X_{(n)} \leq z\} - \mathcal{P}_N\{X_{(n)} \leq z-1\} \\ &= \left(\frac{z-N}{N}\right)^n - \left(\frac{z-N-1}{N}\right)^n\end{aligned}$$

para  $y = N+1, N+2, \dots, 2N$ . Y la distribución conjunta de  $X_{(1)}$  y  $X_{(n)}$ , es:

$$\begin{aligned}\mathcal{P}_N\{X_{(1)} = y, X_{(n)} = z\} &= \\ &= \mathcal{P}_N\{X_{(1)} \geq y, X_{(n)} \leq z\} - \mathcal{P}_N\{X_{(1)} \geq y+1, X_{(n)} \leq z\} \\ &\quad - \mathcal{P}_N\{X_{(1)} \geq y, X_{(n)} \leq z-1\} + \mathcal{P}_N\{X_{(1)} \geq y+1, X_{(n)} \leq z-1\} \\ &= \left(\frac{z-y+1}{N}\right)^n - 2\left(\frac{z-y}{N}\right)^n + \left(\frac{z-y-1}{N}\right)^n\end{aligned}$$

para  $y < z \in \{N+1, N+2, \dots, 2N\}$  (y  $1/N^n$  en el caso  $y = z$ ). Como

$$\mathcal{P}_N\{X_1 = x_1, \dots, X_n = x_n, X_{(1)} = y, X_{(n)} = z\} = \frac{1}{N^n},$$

siempre que  $y \leq x_1, \dots, x_n \leq z \in \{N+1, \dots, 2N\}$  y haya algún  $x_i$  que coincida con  $y$  y algún  $x_j$  que coincida con  $z$ , se obtiene

$$\begin{aligned}\mathcal{P}_N\{X_1 = x_1, \dots, X_n = x_n \mid X_{(1)} = y, X_{(n)} = z\} &= \\ &= \frac{1/N^n}{((z-y+1)/N)^n - 2((z-y)/N)^n + ((z-y-1)/N)^n} \\ &= \frac{1}{(z-y+1)^n - 2(z-y)^n + (z-y-1)^n}\end{aligned}$$

para cualesquiera  $x_1, \dots, x_n \in \{y, y+1, \dots, z\}$ , con algún  $x_i = y$  y algún  $x_j = z$ . El resultado no depende de  $N$ , lo cual prueba que  $T = (X_{(1)}, X_{(n)})$  es suficiente. En cambio

$$\mathcal{P}_N\{X_1 = x_1, \dots, X_n = x_n \mid X_{(1)} = y\} = \frac{1}{(2N-y+1)^n - (2N-y)^n}$$

siempre que  $x_1, \dots, x_n \in \{y, y+1, \dots, 2N\}$  y algún  $x_i$  coincide con  $y$ ; lo cual pone en evidencia que  $X_{(1)}$  no es suficiente. Lo mismo ocurre con  $X_{(n)}$ .

Como se ve, en este caso los estadísticos suficientes son de dimensión por lo menos 2. Podrían obtenerse estadísticos suficientes de mayor dimensión, pero no hay ninguna ventaja en añadir más información ya que, una vez que se conoce  $(X_{(1)}, X_{(n)})$  cualquier otra información es superflua para la estimación de  $N$ .

En general interesa naturalmente considerar estadísticos suficientes  $T$  de dimensión lo menor posible; porque cuanta menor dimensión se pueda conseguir, más cantidad de información superflua se habrá eliminado y más reducida será la clase de los estimadores que actúan únicamente en función de  $T$ . De no ser así, siempre podría adoptarse la solución trivial de considerar la propia muestra  $(X_1, X_2, \dots, X_n)$  como un estadístico  $n$ -dimensional, que evidentemente es suficiente, pero no reduce en absoluto el ámbito de búsqueda de estimadores adecuados.

En definitiva, en un problema de estimación, lo que el *principio de suficiencia* aconseja es condensar al máximo la información relevante, en un estadístico suficiente “mínimo”  $T$ , y seleccionar un estimador puntual  $T'$  entre los estadísticos que son función de la muestra a través de  $T$ :

$$T'(X_1, X_2, \dots, X_n) = \varphi(T(X_1, X_2, \dots, X_n)).$$

La noción de minimalidad se precisará más adelante, después de examinar más detenidamente la idea de suficiencia.

En los dos ejemplos anteriores la distribución de la muestra, condicionada por el estadístico suficiente, ha resultado uniforme (independiente de  $(x_1, x_2, \dots, x_n)$ ) entre las muestras posibles; puede verse que no es forzosamente así.

### Ejemplo 5.11

Sea  $(X_1, X_2, \dots, X_n)$  una muestra aleatoria simple del número de accidentes diarios en una localidad, el cual se supone distribuido con distribución de Poisson de parámetro  $\lambda$  desconocido. La función de probabilidad de la muestra es entonces

$$\mathcal{P}_\lambda\{X_1 = x_1, \dots, X_n = x_n\} = \frac{e^{-n\lambda} \lambda^{x_1 + \dots + x_n}}{x_1! x_2! \dots x_n!}$$

para  $x_1, x_2, \dots, x_n = 0, 1, 2, \dots$

Como  $T = \sum_{i=1}^n X_i$  tiene distribución de Poisson de parámetro  $n\lambda$ , será

$$\begin{aligned}\mathcal{P}_\lambda\{X_1 = x_1, \dots, X_n = x_n \mid T = t\} &= \frac{e^{-n\lambda} \lambda^{x_1 + \dots + x_n} / x_1! x_2! \cdots x_n!}{e^{-n\lambda} (n\lambda)^t / t!} \\ &= \frac{t!}{x_1! x_2! \cdots x_n!} \left(\frac{1}{n}\right)^t\end{aligned}$$

supuesto que  $\sum_1^n x_i = t$  (y 0 en otro caso).

Es decir que la muestra, condicionada por su suma, tiene distribución multinomial (igual que si el número total  $t$  de accidentes se distribuyese al azar entre  $n$  días) y no depende de  $\lambda$ . El estadístico  $T = \sum_1^n X_i$  es por tanto suficiente, aunque, condicionado por el valor de  $T$ , no todas las muestras tengan la misma probabilidad.

Todos los ejemplos anteriores consideran el caso de poblaciones descriptas por variables aleatorias discretas; la razón es que el estudio directo de estadísticos suficientes para familias de distribuciones continuas entraña mayores dificultades. Cabe el recurso de emplear el artificio siguiente:

### Ejemplo 5.12

Sea  $(X_1, X_2, \dots, X_n)$  una muestra aleatoria simple de una población con distribución exponencial, de densidad

$$f_\theta(x) = \theta e^{-\theta x} \quad \text{para } x > 0.$$

La densidad  $n$ -dimensional de la muestra es entonces

$$f_\theta(x_1, \dots, x_n) = \theta^n e^{-\theta(x_1 + \dots + x_n)} \quad \text{para } x_1, x_2, \dots, x_n > 0.$$

La prueba de que  $T = \sum_1^n X_i$  es un estadístico suficiente para la familia de distribuciones de la población, puede darse mediante la transformación biyectiva

$$\begin{cases} Y_1 &= \sum_1^n X_i \\ Y_2 &= X_2 \\ &\vdots \\ Y_n &= X_n \end{cases}$$

cuyo jacobiano es 1, de manera que  $(Y_1, Y_2, \dots, Y_n)$  tendrá densidad conjunta

$$g_\theta(y_1, y_2, \dots, y_n) = \theta^n e^{-\theta y_1} \quad \text{para } y_2, \dots, y_n > 0 \text{ e } y_1 > y_2 + \dots + y_n.$$

Como la distribución marginal de  $Y_1$  es  $\gamma(n, \theta)$ , de densidad

$$\frac{\theta^n y_1^{n-1} e^{-\theta y_1}}{(n-1)!} \quad \text{para } y_1 > 0$$

resulta que la densidad de  $(Y_2, \dots, Y_n)$ , condicionada por  $Y_1 = y_1$ , es

$$\frac{\theta^n e^{-\theta y_1}}{\theta^n y_1^{n-1} e^{-\theta y_1} / (n-1)!} = \frac{(n-1)!}{y_1^{n-1}} \quad \text{para } y_2, \dots, y_n > 0 \text{ e } y_2 + \dots + y_n < y_1$$

que no depende de  $\theta$ . Quiere decirse que, si se conoce el valor  $y_1$  de  $T$ , entonces  $X_2, X_3, \dots, X_n$  pueden elegirse con distribución independiente de  $\theta$  y el valor de  $X_1 = T - X_2 - X_3 - \dots - X_n$  queda determinado. O sea, que la distribución de  $(X_1, X_2, \dots, X_n)$  condicionada por  $T = y_1$  (concentrada en el hiperplano  $x_1 + x_2 + \dots + x_n = y_1$ ) no depende de  $\theta$ .

Obsérvese que la dificultad del caso continuo obedece a que la distribución de la muestra, condicionada por el valor del estadístico, será casi siempre singular (concentrada sobre una superficie del espacio muestral), y el concepto de condicionamiento estudiado hasta ahora no provee de buenas herramientas para este caso. Afortunadamente existe un método general, y mucho más cómodo que los cálculos de los ejemplos anteriores, para comprobar que un estadístico es suficiente; además, con frecuencia, sugiere cuál es estadístico suficiente lo más reducido posible.

**Teorema de factorización:** Sea  $(X_1, X_2, \dots, X_n)$  una muestra aleatoria simple<sup>2</sup> de una población cuya distribución pertenece a  $\mathcal{F} = \{F_\theta | \theta \in \Theta\}$  y representemos por  $f_\theta(x_1, x_2, \dots, x_n)$  la función de probabilidad o la función de densidad de la muestra, según que la población sea discreta o continua.

Un estadístico  $T(X_1, X_2, \dots, X_n)$  es suficiente si y sólo si

$$f_\theta(x_1, x_2, \dots, x_n) = g_\theta(T(x_1, x_2, \dots, x_n))h(x_1, x_2, \dots, x_n)$$

siendo  $g_\theta$  una función que sólo depende de  $x_1, \dots, x_n$  a través del valor  $T(x_1, \dots, x_n)$  del estadístico y  $h(x_1, \dots, x_n)$  una función<sup>3</sup> que no depende de  $\theta$ .

La aplicación del teorema es muy simple y sus ventajas pueden apreciarse resolviendo en una línea cada uno de los ejemplos anteriores; así en el

---

<sup>2</sup>La hipótesis de que la muestra aleatoria es simple no es en realidad necesaria.

<sup>3</sup>Técnicamente,  $g_\theta$  y  $h$  tienen que ser funciones medibles, lo cual no supone ninguna limitación en la práctica.

ejemplo 5.9:

$$f_\theta(x_1, x_2, \dots, x_n) = \theta^{x_1 + \dots + x_n} (1 - \theta)^{n - (x_1 + \dots + x_n)} = g_\theta \left( \sum_1^n x_i \right)$$

lo cual, tomando  $h(x_1, \dots, x_n) = 1$ , prueba que  $\sum_1^n X_i$  es suficiente.

En el ejemplo 5.11,

$$\begin{aligned} f_\lambda(x_1, x_2, \dots, x_n) &= e^{-n\lambda} \lambda^{x_1 + \dots + x_n} \cdot (x_1! x_2! \cdots x_n!)^{-1} \\ &= g_\lambda \left( \sum_1^n x_i \right) h(x_1, x_2, \dots, x_n) \end{aligned}$$

siendo  $g_\lambda(t) = e^{-n\lambda} \lambda^t$  y  $h(x_1, \dots, x_n) = (x_1! x_2! \cdots x_n!)^{-1}$ ; luego también  $\sum_1^n X_i$  es suficiente en este caso.

Para el ejemplo 5.12,

$$f_\theta(x_1, x_2, \dots, x_n) = \theta^n e^{-\theta(x_1 + \dots + x_n)} = g_\theta \left( \sum_1^n x_i \right)$$

de manera que, con  $h(x_1, x_2, \dots, x_n) = 1$  se obtiene el resultado.

En cuanto al ejemplo 5.10, que ha quedado para el final por ser ligeramente más complicado,

$$f_\theta(x_1, x_2, \dots, x_n) = \frac{1}{N^n} I_{\{x_{(1)} \geq N+1\}} I_{\{x_{(n)} \leq 2N\}} = g_N(x_{(1)}, x_{(n)})$$

de forma que  $(X_{(1)}, X_{(n)})$  es suficiente; además, como no hay manera de factorizar  $f_\theta$  en la forma requerida, usando sólo  $X_{(1)}$  ó  $X_{(n)}$ ; ni uno ni otro son individualmente suficientes.

Debido a las limitaciones en el concepto de distribuciones condicionadas a las que nos hemos referido en el ejemplo 5.12, sólo podemos efectuar la demostración del teorema de factorización en el caso de poblaciones discretas. Bajo esa hipótesis, la cuestión es sencilla:

Si  $T = T(X_1, X_2, \dots, X_n)$  es un estadístico suficiente para la familia de distribuciones  $\{F_\theta | \theta \in \Theta\}$ , la función de probabilidad de la muestra, condicionada por el valor de  $T$ :  $f(x_1, \dots, x_n | T = t)$ , no depende de  $\theta$ . Como

$$f_\theta(x_1, \dots, x_n) = \mathcal{P}_\theta\{T = T(x_1, \dots, x_n)\} f(x_1, \dots, x_n | T = T(x_1, \dots, x_n)),$$

basta tomar

$$\begin{cases} g_\theta(t) = \mathcal{P}_\theta\{T(X_1, X_2, \dots, X_n) = t\} \\ h(x_1, x_2, \dots, x_n) = f(x_1, \dots, x_n \mid T = T(x_1, \dots, x_n)) \end{cases}$$

para obtener el resultado. Recíprocamente, si

$$f_\theta(x_1, \dots, x_n) = g_\theta(T(x_1, \dots, x_n)) h(x_1, \dots, x_n)$$

llamando  $A_t = \{(x_1, \dots, x_n) \in \mathcal{X} \mid T(x_1, \dots, x_n) = t\}$ , será

$$\begin{aligned} \mathcal{P}_\theta\{T(X_1, X_2, \dots, X_n) = t\} &= \sum_{A_t} g_\theta(T(x_1, \dots, x_n)) h(x_1, \dots, x_n) \\ &= g_\theta(t) \sum_{A_t} h(x_1, \dots, x_n); \end{aligned}$$

con lo cual

$$\begin{aligned} \mathcal{P}_\theta\{X_1 = x_1, \dots, X_n = x_n \mid T = t\} &= \\ &= \begin{cases} \frac{h(x_1, \dots, x_n)}{\sum_{A_t} h(x_1, \dots, x_n)} & \text{si } T(x_1, \dots, x_n) = t, \\ 0 & \text{si } T(x_1, \dots, x_n) \neq t. \end{cases} \end{aligned}$$

De manera que la distribución de  $(X_1, \dots, X_n)$ , condicionada por el valor de  $T$ , no depende de  $\theta$  y  $T$  es suficiente.

Cuando la distribución teórica dependa de un parámetro multidimensional el estadístico suficiente varía mucho según la información de que se disponga acerca del parámetro:

### Ejemplo 5.13

Si  $(X_1, X_2, \dots, X_n)$  es una muestra aleatoria simple de una distribución uniforme en el intervalo  $(\theta_1, \theta_2)$  con  $\theta_1 < \theta_2$ , su función de densidad es

$$\begin{aligned} f_{\theta_1, \theta_2}(x_1, \dots, x_n) &= \frac{1}{(\theta_2 - \theta_1)^n} \quad \text{para } \theta_1 < x_1, x_2, \dots, x_n < \theta_2 \\ &= \frac{1}{(\theta_2 - \theta_1)^n} I_{\{x_{(1)} > \theta_1\}} I_{\{x_{(n)} < \theta_2\}}; \end{aligned}$$

de forma que  $(X_{(1)}, X_{(n)})$  es un estadístico suficiente, (tomando en la factorización  $h \equiv 1$ ).

Sin embargo, si  $\theta_1$  fuese conocido, se puede tomar  $h = I_{\{x_{(1)} < \theta_1\}}$  obteniéndose que  $X_{(n)}$  es un estadístico suficiente. Y al revés, si  $\theta_2$  es conocido,  $X_{(1)}$  es suficiente ya que se puede tomar  $h = I_{\{X_{(n)} < \theta_2\}}$

Si se sabe que  $\theta_1 = -\theta_2$ , expresando

$$f_{\theta_1, \theta_2}(x_1, \dots, x_n) = \frac{1}{(2\theta_2)^n} I_{\{\max_{1 \leq i \leq n} |x_i| < \theta_2\}}$$

se obtiene que  $\max_{1 \leq i \leq n} |X_i|$  es un estadístico suficiente.

En cambio, si  $\theta_2 = 3\theta_1$  (lo cual no es más que una traslación del caso anterior), no hay ninguna simplificación posible del estadístico suficiente  $(X_{(1)}, X_{(n)})$ .

### Ejemplo 5.14

Si  $(X_1, X_2, \dots, X_n)$  es una muestra aleatoria simple de una población con distribución  $N(\mu, \sigma)$ , la función de densidad de la muestra es

$$\begin{aligned} f_{\mu, \sigma^2}(x_1, x_2, \dots, x_n) &= \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_1^n (x_i - \mu)^2 \right\} = \\ &= \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_1^n (x_i - \bar{x})^2 \right\} \exp \left\{ -\frac{n}{2\sigma^2} (\bar{x} - \mu)^2 \right\} \\ &= \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp \left\{ -\frac{ns^2}{2\sigma^2} \right\} \exp \left\{ -\frac{n}{2\sigma^2} (\bar{x} - \mu)^2 \right\}. \end{aligned}$$

Si  $\mu$  es una constante conocida, de la primera expresión se obtiene que  $\sum_1^n (X_i - \mu)^2$  es un estadístico suficiente (para  $\sigma^2$ ).

Si  $\sigma^2$  es una constante conocida, la segunda expresión prueba que  $\bar{X}$  es un estadístico suficiente (para  $\mu$ ).

En el caso en que ambos parámetros fuesen desconocidos,  $(\bar{X}, s^2)$  es un estadístico suficiente (para  $\mu$  y  $\sigma^2$ ), según muestra la última expresión.

#### 5.4.1. Estadísticos suficientes minimales

Ya se ha señalado la conveniencia de disponer de un estadístico suficiente de dimensión lo menor posible, para eliminar el máximo de información superflua, no relevante a la hora de seleccionar un estimador del parámetro. La formulación a través de la dimensión no es, sin embargo,

la formulación precisa. Por ejemplo, como en el muestreo aleatorio simple  $f_\theta(x_1, \dots, x_n) = f_\theta(x_1) \cdots f_\theta(x_n)$ , no depende del orden de las variables, la muestra ordenada  $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$  es siempre un estadístico suficiente que, a pesar de tener la misma dimensión que la muestra original, ha eliminado la información superflua relativa a la posición en que se produjo cada observación de la muestra.

De hecho, independientemente de que la dimensión se reduzca o no, cualquier transformación, mediante una función (medible) no inyectiva  $\varphi$ , a la que se someta un estadístico  $T$ , produce un estadístico  $T' = \varphi(T)$  con menos información que  $T$  (ya que  $T'$  puede obtenerse a partir de  $T$ , pero  $T$  es irrecuperable a partir de  $T'$ ). Además, en virtud del teorema de factorización, si  $T'$  fuese suficiente,  $T$  lo sería también; de manera que  $T$  podría ser desecharo y sustituido por  $T'$ .

La expresión precisa del deseo de condensar al máximo la información relevante para la estimación del parámetro, queda entonces recogida en la siguiente definición:

*Dada una muestra  $(X_1, X_2, \dots, X_n)$  de una población con distribución perteneciente a  $\{F_\theta \mid \theta \in \Theta\}$ , un estadístico suficiente  $T$  se denomina minimal suficiente si, para cualquier otro estadístico suficiente  $T'$ , existe una función (medible)  $\varphi$  tal que  $T = \varphi(T')$ .*

Aunque la determinación de estadísticos suficientes por aplicación del teorema de factorización, suele conducir a estadísticos minimales suficientes, ello depende obviamente de la habilidad de la factorización y es conveniente contar con un método que permita construir sistemáticamente estadísticos minimales suficientes:

**Teorema 5.2:** *Con la misma notación del teorema de factorización,  $T$  es un estadístico minimal suficiente si y sólo si*

$$\left\{ \begin{array}{ll} T(x_1, \dots, x_n) = T(x'_1, \dots, x'_n) & \text{si } \frac{f_\theta(x_1, \dots, x_n)}{f_\theta(x'_1, \dots, x'_n)} \text{ no depende de } \theta, \\ T(x_1, \dots, x_n) \neq T(x'_1, \dots, x'_n) & \text{si } \frac{f_\theta(x_1, \dots, x_n)}{f_\theta(x'_1, \dots, x'_n)} \text{ depende de } \theta. \end{array} \right.$$

Sólo probaremos el resultado para poblaciones discretas. Supongamos

que  $T$  cumple las condiciones del teorema, entonces si  $T(x'_1, \dots, x'_n) = t$ ,

$$\begin{aligned}\mathcal{P}_\theta\{X_1 = x'_1, \dots, X_n = x'_n \mid T = t\} &= \frac{\mathcal{P}_\theta\{X_1 = x'_1, \dots, X_n = x'_n\}}{\mathcal{P}_\theta\{T = t\}} \\ &= \frac{f_\theta(x'_1, \dots, x'_n)}{\sum_{A_t} f_\theta(x_1, \dots, x_n)} = \left[ \sum_{A_t} \frac{f_\theta(x_1, \dots, x_n)}{f_\theta(x'_1, \dots, x'_n)} \right]^{-1}\end{aligned}$$

y, como  $f_\theta(x_1, \dots, x_n) / f_\theta(x'_1, \dots, x'_n)$  es independiente de  $\theta$  en  $A_t$ , resulta que

$$\mathcal{P}_\theta\{X_1 = x'_1, X_2 = x'_2, \dots, X_n = x'_n \mid T = t\}$$

no depende de  $\theta$  y  $T$  es suficiente.

Si  $T'$  es otro estadístico suficiente y  $T'(x_1, \dots, x_n) = T'(x'_1, \dots, x'_n) = t'$ ,

$$\mathcal{P}_\theta\{X_1 = x_1, \dots, X_n = x_n \mid T' = t'\} = \frac{f_\theta(x_1, \dots, x_n)}{\mathcal{P}_\theta\{T' = t'\}}$$

será independiente de  $\theta$ , lo mismo que

$$\mathcal{P}_\theta\{X_1 = x'_1, \dots, X_n = x'_n \mid T' = t'\} = \frac{f_\theta(x'_1, \dots, x'_n)}{\mathcal{P}_\theta\{T' = t'\}}.$$

Dividiendo término a término resulta que  $f_\theta(x_1, \dots, x_n) / f_\theta(x'_1, \dots, x'_n)$  no depende de  $\theta$ ; por consiguiente  $T(x_1, \dots, x_n) = T(x'_1, \dots, x'_n)$ , lo cual indica que  $T$  es función de  $T'$  (es decir, para cada valor  $t'$  de  $T'$ , existe un único valor  $\varphi(t')$  tal que  $T(x_1, \dots, x_n) = \varphi(t')$  si  $T'(x_1, \dots, x_n) = t'$ ). En definitiva  $T$  es minimal suficiente.

Para establecer el recíproco, supongamos que  $T'$  es minimal suficiente. Su suficiencia indica que

$$\begin{cases} f_\theta(x_1, \dots, x_n) = g_\theta(T'(x_1, \dots, x_n))h(x_1, \dots, x_n), \\ f_\theta(x'_1, \dots, x'_n) = g_\theta(T'(x'_1, \dots, x'_n))h(x'_1, \dots, x'_n). \end{cases}$$

Así que, si  $f_\theta(x_1, \dots, x_n) / f_\theta(x'_1, \dots, x'_n)$  depende de  $\theta$ , también dependerá de  $\theta$  el cociente  $g_\theta(T'(x_1, \dots, x_n)) / g_\theta(T'(x'_1, \dots, x'_n))$ ; lo cual obliga a que sea  $T'(x_1, \dots, x_n) \neq T'(x'_1, \dots, x'_n)$ .

Además, si  $T$  es un estadístico que cumple las condiciones del teorema, tiene que ser  $T' = \varphi(T)$ . Luego, si  $f_\theta(x_1, \dots, x_n) / f_\theta(x'_1, \dots, x'_n)$  no depende de  $\theta$ , será  $T(x_1, \dots, x_n) = T(x'_1, \dots, x'_n)$  y, por consiguiente  $T'(x_1, \dots, x_n) = T'(x'_1, \dots, x'_n)$ .

El teorema anterior asegura que el estadístico minimal suficiente es único salvo transformaciones biunívocas, puesto que no se especifica el valor común de  $T$  en cada conjunto en el cual  $f_\theta(x_1, \dots, x_n) / f_\theta(x'_1, \dots, x'_n)$  no depende de  $\theta$ . Sólo se indica que ha de ser el mismo en cada uno de ellos y diferente para cada dos distintos.

Además el resultado tiene la ventaja de que permite comprobar si un determinado estadístico no es suficiente, lo cual puede no ser tan simple a partir del teorema de factorización, debido a la posibilidad de que existe una factorización “oculta”. Ahora, basta comprobar que el estadístico minimal suficiente no puede expresarse en función de él.

Nuevamente la aplicación del teorema anterior es muy simple.

### Ejemplo 5.15

En la situación del ejemplo 5.9, como

$$\frac{f_\theta(x_1, \dots, x_n)}{f_\theta(x'_1, \dots, x'_n)} = \left[ \frac{\theta}{1-\theta} \right]^{\sum_1^n x_i - \sum_1^n x'_i}$$

el cociente será independiente de  $\theta$  si y sólo si  $\sum_1^n x_i = \sum_1^n x'_i$ ; de manera que el estadístico minimal suficiente es  $T = \sum_1^n X_i$  (o cualquier transformación biyectiva suya). Aunque poca duda podía caber de ello a la vista de la factorización de  $f_\theta$ .

### Ejemplo 5.16

En el ejemplo 5.14,

$$\frac{f_{\mu, \sigma^2}(x_1, \dots, x_n)}{f_{\mu, \sigma^2}(x'_1, \dots, x'_n)} = \exp \left\{ -\frac{1}{2\sigma^2} \left[ \sum_1^n x_i^2 - 2\mu \sum_1^n x_i - \sum_1^n x'^2_i + 2\mu \sum_1^n x'_i \right] \right\}.$$

Si  $\mu$  es conocido, el cociente no dependerá de  $\sigma^2$  cuando

$$\sum_1^n x_i^2 - 2\mu \sum_1^n x_i = \sum_1^n x'^2_i - 2\mu \sum_1^n x'_i,$$

lo cual da como estadístico minimal suficiente

$$\sum_1^n X_i^2 - 2\mu \sum_1^n X_i,$$

ó  $\sum_1^n (X_i - \mu)^2$  que está en correspondencia biunívoca con el anterior.

Si  $\sigma^2$  es conocido, el cociente será independiente de  $\mu$  cuando

$$\sum_1^n x_i = \sum_1^n x'_i$$

de manera que  $\sum_1^n X_i$ , o bien  $\bar{X}$ , es el estadístico minimal suficiente.

Para  $\mu$  y  $\sigma^2$  desconocidos, la independencia de  $\mu$  y  $\sigma^2$  del cociente equivale a la anulación del exponente para cualquier valor de  $\mu$ ; es decir

$$\sum_1^n x_i^2 = \sum_1^n x'^2_i \quad \text{y} \quad \sum_1^n x_i = \sum_1^n x'_i,$$

con lo cual el estadístico minimal suficiente es  $(\sum_1^n X_i, \sum_1^n X_i^2)$  ó  $(\bar{X}, s^2)$ .

### Ejemplo 5.17

Muchas de las familias paramétricas más usuales (Bernouilli, Poisson, exponencial, normal, etc.) tienen una función de probabilidad o de densidad que responde a la expresión general

$$f_\theta(x) = c(\theta)h(x) \exp \left\{ \sum_{j=1}^p q_j(\theta)T_j(x) \right\}$$

(sin que existan relaciones lineales entre las  $q_j(\theta)$ , ni entre las  $T_j(x)$ , que permitan simplificar el exponente). Tales familias se denominan por ello *de tipo exponencial*. (Véanse los ejercicios 5.6 y 5.7).

Para cualquier familia de tipo exponencial, la distribución de la muestra tiene función de probabilidad o de densidad:

$$f_\theta(x_1, x_2, \dots, x_n) = c(\theta)^n h(x_1)h(x_2)\cdots h(x_n) \exp \left\{ \sum_{j=1}^p q_j(\theta) \sum_{i=1}^n T_j(x_i) \right\},$$

de manera que el teorema de factorización garantiza que

$$\left( \sum_{i=1}^n T_1(X_i), \sum_{i=1}^n T_2(X_i), \dots, \sum_{i=1}^n T_p(X_i) \right)$$

es un estadístico suficiente, pues basta tomar

$$\begin{cases} g_\theta(t_1, t_2, \dots, t_p) = c(\theta)^n \exp \left\{ \sum_1^p q_j(\theta)t_j \right\}, \\ h(x_1, \dots, x_n) = h(x_1)h(x_2)\cdots h(x_n). \end{cases}$$

Para ser más precisos, se puede aplicar el teorema anterior, formando el cociente

$$\begin{aligned} \frac{f_\theta(x_1, \dots, x_n)}{f_\theta(x'_1, \dots, x'_2)} &= \\ &= \frac{h(x_1)h(x_2)\cdots h(x_n)}{h(x'_1)h(x'_2)\cdots h(x'_n)} \exp \left\{ \sum_{j=1}^p q_j(\theta) \left[ \sum_{i=1}^n T_j(x_i) - \sum_{i=1}^n T_j(x'_i) \right] \right\}, \end{aligned}$$

que será independiente de  $\theta$  si y sólo si

$$\sum_{j=1}^p [q_j(\theta) - q_j(\theta_0)] \left\{ \sum_{i=1}^n T_j(x_i) - \sum_{i=1}^n T_j(x'_i) \right\} = 0$$

para todo  $\theta, \theta_0 \in \Theta$ .

Puesto que se supone que existen  $p$  valores de  $\theta$  para los cuales los correspondientes vectores  $(q_1(\theta) - q_1(\theta_0), \dots, q_p(\theta) - q_p(\theta_0))$  son linealmente independientes, tiene que ser

$$\sum_{i=1}^n T_j(x_i) = \sum_{i=1}^n T_j(x'_i)$$

para cada  $j = 1, \dots, p$ , de manera que

$$\left( \sum_1^n T_1(X_i), \sum_1^n T_2(X_i), \dots, \sum_1^n T_p(X_i) \right)$$

es minimal suficiente.

## 5.5. Criterios de selección de estimadores

Volviendo a la situación tal y como quedó planteada al final de la sección 2, una vez atribuido a cada estimador su riesgo  $R_T(\theta)$ , si no es posible elegir un estimador que minimice simultáneamente  $R_T(\theta)$  para todos los valores de  $\theta$ , en lugar de restringir la clase de los estimadores considerados —como se ha sugerido en las secciones anteriores—, se puede valorar de alguna manera toda la función  $R_T(\theta)$  para asignar a cada estimador un único índice de calidad independiente de  $\theta$ . Dos son las maneras principales de lograr este objetivo, las cuales serán consideradas a continuación.

### 5.5.1. Estimadores minimax

Puesto que el riesgo  $R_T(\theta)$  expresa la pérdida esperada, con el estimador  $T$ , cuando el valor del parámetro es  $\theta$ , la manera de asegurar una pérdida esperada pequeña, cualquiera que sea  $\theta$ , consiste en controlar el

$$\max_{\theta \in \Theta} R_T(\theta)$$

y preferir un estimador  $T_1$  a otro  $T_2$  si

$$\max_{\theta \in \Theta} R_{T_1}(\theta) < \max_{\theta \in \Theta} R_{T_2}(\theta).$$

Con ello, cualquier par de estimadores resulta comparable (aunque dos estimadores pueden resultar indiferentes si ambos máximos coinciden), de forma que sería óptimo, con este criterio, cualquier estimador que alcance el valor

$$\min_T \max_{\theta \in \Theta} R_T(\theta).$$

Consecuentemente, los estimadores  $T^*$  tales que

$$\max_{\theta \in \Theta} R_{T^*}(\theta) = \min_T \max_{\theta \in \Theta} R_T(\theta)$$

se denominan *estimadores minimax* y por extensión el criterio de comparar los estimadores por el máximo de su riesgo se denomina *criterio minimax*.

En estricto rigor, sería necesario reemplazar los máximos y los mínimos de las expresiones anteriores por supremos e ínfimos, puesto que no hay garantía de que sean accesibles. En ese sentido no puede asegurarse que existan siempre estimadores minimax y por ello se definen los *estimadores  $\varepsilon$ -minimax* por la condición:

$$\sup_{\theta \in \Theta} R_{T^*}(\theta) \leq \inf_T \sup_{\theta \in \Theta} R_T(\theta) + \varepsilon$$

como manera de asegurar que  $T^*$  es óptimo a menos de  $\varepsilon > 0$ .

La determinación de los estimadores minimax es realmente difícil, puesto que se trata de resolver un problema de mínimo, en el espacio de todas las funciones de la muestra y no precisamente de un funcional simple, sino del máximo riesgo asociado a cada función.

Un ejemplo irreal por lo reducido del tamaño muestral puede, de todas maneras, aclarar el significado del método.

**Ejemplo 5.18**

Una urna contiene una proporción  $\theta$  de bolas marcadas con 1 y el resto con 0; se extrae una única bola y se pretende hallar el estimador minimax de  $\theta$ , considerando como función de pérdida  $L(\theta, t) = (\theta - t)^2$ .

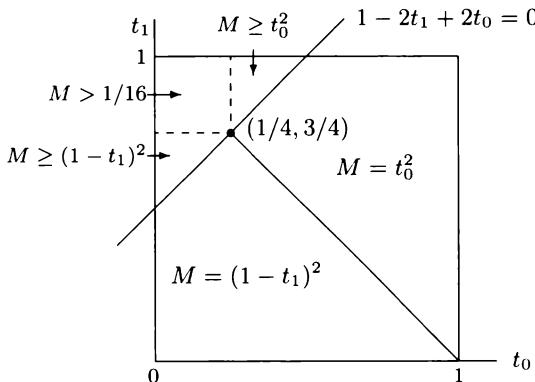
Para ello obsérvese que cualquier estimador  $T$  queda caracterizado por los valores de  $t_0 = T(0)$  y  $t_1 = T(1)$ , asignados a las dos únicas muestras posibles. El riesgo es entonces

$$R_T(\theta) = \theta(t_1 - \theta)^2 + (1 - \theta)(t_0 - \theta)^2 = \theta^2(1 - 2t_1 + 2t_0) + \theta(t_1^2 - 2t_0 - t_0^2) + t_0^2.$$

Si  $1 - 2t_1 + 2t_0 > 0$ ,  $R_T(\theta)$  es una parábola convexa, cuyo máximo se alcanza en  $\theta = 0$  o bien en  $\theta = 1$ . Como  $R_T(0) = t_0^2$  y  $R_T(1) = (1 - t_1)^2$ , es

$$M = \max_{\theta \in [0,1]} R_T(\theta) = \begin{cases} t_0^2 & \text{si } t_0 \geq 1 - t_1, \\ (1 - t_1)^2 & \text{si } t_0 < 1 - t_1. \end{cases}$$

Para  $1 - 2t_1 + 2t_0 < 0$ , el cálculo exacto es más molesto; sin embargo, como  $R_T(\theta)$  es una parábola cóncava,  $M = \max_{\theta \in [0,1]} R_T(\theta)$  es superior a  $R_T(0)$  y  $R_T(1)$ .



Por otra parte, para  $t_0 \in [0, 1/4]$  y  $t_1 \in (3/4, 1]$  es

$$\begin{aligned} \max_{\theta \in [0,1]} R_T(\theta) &\geq R_T\left(\frac{1}{2}\right) = \frac{1}{4} - \frac{t_1}{2} - \frac{t_0}{2} + \frac{t_1^2}{2} + \frac{t_0^2}{2} \\ &= \frac{1}{16}\{4 - 8t_0(1 - t_0) - 8t_1(1 - t_1)\} > \frac{1}{16}. \end{aligned}$$

Luego,  $\max_{\theta \in [0,1]} R_T(\theta)$  se hace mínimo para  $t_0 = \frac{1}{4}, t_1 = \frac{3}{4}$ . Así que el estimador minimax da la estimación  $\hat{\theta} = \frac{1}{4}$ , si se observa una bola con un 0, y  $\hat{\theta} = \frac{3}{4}$  si se observa una bola con un 1.

El criterio minimax se puede utilizar dentro de una clase reducida de estimadores, para seleccionar, con tal criterio, el mejor de entre ellos; aunque, por supuesto, ello no establece ninguna comparación con estimadores externos a la clase. Esto es lo que se hace en el ejemplo siguiente:

### Ejemplo 5.19

En la misma situación anterior, después de obtener una muestra aleatoria simple de tamaño  $n$ , se quiere determinar el estimador minimax entre los de la forma  $\alpha\bar{X}$ , con  $\alpha > 0$ .

Según los resultados del ejemplo 5.3,

$$R_{\alpha\bar{X}}(\theta) = \left[ (1 - \alpha)^2 - \frac{\alpha^2}{n} \right] \theta^2 + \frac{\alpha^2}{n} \theta$$

y se pueden ignorar los estimadores con  $\alpha > 1$ , para los cuales se verifica  $R_{\alpha\bar{X}}(\theta) > R_{\bar{X}}(\theta)$  para todo  $\theta \in [0, 1]$ .

Si  $(1 - \alpha)^2 - \alpha^2/n > 0$ , lo cual equivale a  $\alpha < \frac{\sqrt{n}}{\sqrt{n} + 1}$ , el máximo de  $R_T(\theta)$  se alcanza para  $\theta = 1$  y es

$$\max_{\theta \in [0, 1]} R_{\alpha\bar{X}}(\theta) = R_T(1) = (1 - \alpha)^2.$$

Si  $(1 - \alpha)^2 - \alpha^2/n < 0$ , la parábola  $R_{\alpha\bar{X}}(\theta)$  tiene su vértice en el punto  $\theta = \frac{\alpha^2/n}{2[\alpha^2/n - (1 - \alpha)^2]}$ , que cae en  $[0, 1]$  para  $\alpha \geq \frac{\sqrt{2n}}{\sqrt{2n} + 1}$ . Por debajo de ese punto  $\max_{\theta \in [0, 1]} R_{\alpha\bar{X}}(\theta)$  sigue valiendo  $(1 - \alpha)^2$ ; mientras que por encima es

$$\max_{\theta \in [0, 1]} R_{\alpha\bar{X}}(\theta) = \frac{\alpha^4/n^2}{4[\alpha^2/n - (1 - \alpha)^2]}.$$

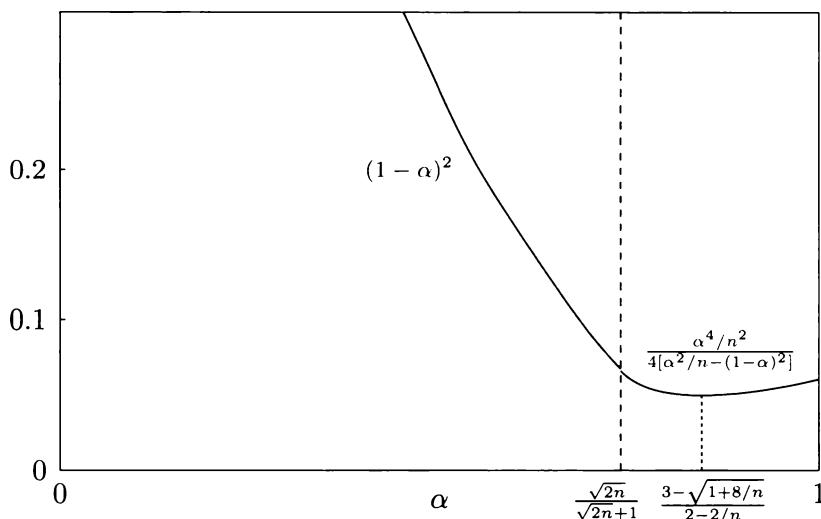
La representación de  $\max_{\theta \in [0, 1]} R_{\alpha\bar{X}}(\theta)$  en función de  $\alpha$  aparece en el gráfico siguiente. Es fácil ver que el mínimo se alcanza para

$$\alpha = \frac{3 - \sqrt{1 + 8/n}}{2 - 2/n},$$

de manera que entre los múltiplos de  $\bar{X}$ , el estimador minimax para la proporción  $\theta$  es

$$\frac{3 - \sqrt{1 + 8/n}}{2 - 2/n} \bar{X}$$

lo cual es muy poco inferior a  $\bar{X}$  si  $n$  es grande.

Figura 5.3: Máximo riesgo de  $T_{\alpha\bar{X}}$ 

El ejemplo 5.18 muestra que el criterio minimax no siempre conduce a estimadores centrados y tampoco puede asegurarse que dé estimadores basados en un estadístico suficiente minimal. De todas maneras, la crítica principal al criterio minimax es que supone un comportamiento demasiado pesimista: un estimador con muy poco riesgo para “casi todos” los valores de  $\theta$ , pero un riesgo alto,  $R$ , para algunos de ellos, es desecharlo frente a un estimador de riesgo constante, igual a  $R - \varepsilon$ . De hecho, el criterio del minimax proviene de la Teoría de juegos, la cual, en un contexto similar, considera sin embargo que el valor de  $\theta$  es elegido por otro jugador, que intenta obtener el máximo beneficio (mientras el primer jugador trata de elegir  $T$  para minimizar su pérdida). Desde ese punto de vista, sí está justificado pensar que el valor de  $\theta$  será todo lo desfavorable que pueda ser.

### 5.5.2. Estimadores Bayes

El defecto que se acaba de achacar a los estimadores minimax, podría ser evitado si, en lugar de usar el máximo de  $R_T(\theta)$  como piedra de toque para contrastar la calidad de un estadístico, se sustituyese por un valor promedio de los riesgos  $R_T(\theta)$  para los diversos valores de  $\theta$ . La cuestión radica en saber con respecto a qué distribución sobre el espacio paramétrico efectuar

el promedio, porque indudablemente el resultado se verá afectado por ella.

La utilización del criterio Bayes requiere, por tanto, comportarse como si el parámetro  $\theta$  fuese una variable aleatoria de la que se conoce su *distribución a priori*,  $\pi(\theta)$ , ley de probabilidad con la que ha resultado elegido el valor del parámetro. Con ello, el riesgo Bayes de un estimador  $T$  frente a la distribución a priori  $\pi$  es el promedio

$$r_T(\pi) = \int_{\Theta} R_T(\theta)\pi(d\theta)$$

y un estimador  $T_1$  es preferido a  $T_2$  si  $r_{T_1}(\pi) < r_{T_2}(\pi)$ . Lo ideal es entonces encontrar un estimador  $T^*$  que proporcione el mínimo riesgo Bayes, es decir, tal que

$$r_{T^*}(\pi) = \inf_T r_T(\pi)$$

y que, caso de existir, se denomina *estimador Bayes frente a la distribución a priori*  $\pi$ .

La determinación de estimadores Bayes puede llevarse a cabo de una forma mucho más sistemática que la de los estimadores minimax. En primer lugar, los elementos  $F_\theta(x)$  de la familia paramétrica de distribuciones teóricas, pueden interpretarse como la distribución de la variable poblacional  $X$  condicionada por el valor  $\theta$  del parámetro aleatorio. Por consiguiente, puede considerarse la distribución conjunta de  $\theta$  y  $X$ ; formada a partir de la distribución marginal  $\pi$  y de la distribución condicionada  $F_\theta(x)$ .

Siendo obvios los cambios de interpretación necesarios en el caso en que  $\theta$  ó  $X$  fuesen discretas, supondremos que  $\pi(\theta)$  es la densidad de  $\theta$  y  $f_\theta(x)$  la densidad correspondiente a  $F_\theta$ , en cuyo caso:

$$f(\theta, x) = \pi(\theta)f_\theta(x)$$

es la función de densidad conjunta de  $\theta$  y  $X$ . Más en general,

$$f(\theta, x_1, x_2, \dots, x_n) = \pi(\theta)f_\theta(x_1, x_2, \dots, x_n)$$

es la densidad conjunta de  $\theta$  y los elementos  $X_1, X_2, \dots, X_n$  de una muestra aleatoria simple de la población<sup>4</sup>. La densidad marginal de  $(X_1, X_2, \dots, X_n)$  será entonces

$$f(x_1, x_2, \dots, x_n) = \int_{\Theta} f_\theta(x_1, x_2, \dots, x_n)\pi(\theta)d\theta$$

---

<sup>4</sup> La hipótesis de que la muestra es aleatoria simple no es en realidad necesaria, salvo para que sea  $f_\theta(x_1, x_2, \dots, x_n) = f_\theta(x_1)f_\theta(x_2) \cdots f_\theta(x_n)$ .

y por tanto la distribución de  $\theta$  condicionada por  $(X_1, X_2, \dots, X_n)$  tendrá por densidad

$$\begin{aligned}\pi(\theta | x_1, x_2, \dots, x_n) &= \frac{\pi(\theta) f_\theta(x_1, x_2, \dots, x_n)}{f(x_1, x_2, \dots, x_n)} \\ &= \frac{\pi(\theta) f_\theta(x_1, x_2, \dots, x_n)}{\int_{\Theta} f_\theta(x_1, x_2, \dots, x_n) \pi(\theta) d\theta}.\end{aligned}$$

Dicha distribución se denomina *distribución a posteriori* del parámetro  $\theta$ , puesto que expresa la probabilidad (o densidad de probabilidad) de que se haya producido cada posible valor del parámetro, una vez que se han realizado las observaciones de la población. En ese sentido, como es habitual en Cálculo de Probabilidades, la observación de la muestra altera la ley de probabilidad que rige la situación del parámetro.

Por otra parte, el riesgo Bayes de un estimador se puede expresar

$$\begin{aligned}r_T(\pi) &= \int_{\Theta} \int_{\mathcal{X}} L(\theta, T(x_1, \dots, x_n)) f_\theta(x_1, \dots, x_n) dx_1 \cdots dx_n \pi(\theta) d\theta \\ &= \int_{\mathcal{X}} \int_{\Theta} L(\theta, T(x_1, \dots, x_n)) \pi(\theta | x_1, \dots, x_n) d\theta f(x_1, \dots, x_n) dx_1 \cdots dx_n.\end{aligned}$$

Con lo cual la búsqueda del estimador Bayes frente a la distribución a priori  $\pi$ , se reduce a determinar, para cada muestra  $(x_1, x_2, \dots, x_n)$  el valor  $t$  que minimice

$$\int_{\Theta} L(\theta, t) \pi(\theta | x_1, x_2, \dots, x_n) d\theta.$$

Por ejemplo, en el caso de pérdida cuadrática  $L(\theta, t) = (\theta - t)^2$ , es sabido que el valor  $t$  que minimiza  $\int_{\Theta} (\theta - t)^2 dG(\theta)$  es la media de la distribución  $G$ , cualquiera que ésta sea. Resulta pues que, con pérdida cuadrática, el estimador Bayes frente a la distribución a priori  $\pi$  es la media de la distribución a posteriori  $\pi(\theta | x_1, x_2, \dots, x_n)$ .

Análogamente, como  $\int_{\Theta} |\theta - t| dG(\theta)$  se hace mínimo cuando  $t$  coincide con la mediana de  $G$ , con función de pérdida  $L(\theta, t) = |\theta - t|$  el estimador Bayes frente a la distribución a priori  $\pi$  es la mediana de la distribución a posteriori  $\pi(\theta | x_1, x_2, \dots, x_n)$ .

Veamos como se opera en situaciones concretas.

**Ejemplo 5.20**

Un número real  $\theta$  se escoge con densidad  $\pi(\theta) = \frac{k-1}{\theta^k}$  en el intervalo  $(1, \infty)$ , siendo  $k > 1$  un número conocido. No se permite observar el resultado, pero para estimarlo se pueden observar  $n$  valores  $X_1, X_2, \dots, X_n$  elegidos independientemente, al azar entre 0 y  $\theta$ ; de manera que  $(X_1, X_2, \dots, X_n)$  constituye una muestra aleatoria simple de la distribución uniforme en  $(0, \theta)$ .

Nótese en primer lugar que la situación es diferente a las de los ejemplos 5.1, 5.4 y 5.8, por cuanto aquí se conoce la ley de probabilidad con la que se ha elegido  $\theta$  y no sería razonable despreciar la información que ello proporciona, sino que lo interesante es combinarla con la información que proporciona la muestra. En ese sentido, la densidad conjunta de  $\theta$  y  $(X_1, X_2, \dots, X_n)$  es

$$f(\theta, x_1, x_2, \dots, x_n) = \frac{k-1}{\theta^k} I_{\{\theta>1\}} \frac{1}{\theta^n} I_{\{\max_{1 \leq i \leq n} x_i < \theta\}} = \frac{k-1}{\theta^{n+k}} I_{\{\tau < \theta\}}$$

donde se ha tomado  $x_0 = 1$  y  $\tau = \max_{0 \leq i \leq n} x_i$ , para simplificar la notación.

La densidad marginal de  $(X_1, X_2, \dots, X_n)$  es por tanto

$$f(x_1, x_2, \dots, x_n) = \int_{\tau}^{\infty} \frac{(k-1)}{\theta^{n+k}} d\theta = \frac{(k-1)}{(n+k-1)\tau^{n+k-1}}$$

para  $x_1, x_2, \dots, x_n > 0$ ; y la densidad a posteriori de  $\theta$ , una vez observada la muestra, resulta

$$\pi(\theta | x_1, x_2, \dots, x_n) = \frac{(n+k-1)\tau^{n+k-1}}{\theta^{n+k}} \quad \text{para } \theta > \tau;$$

(nótese que, si se observa algún valor  $x_i < 1$ , ya no hay ninguna posibilidad de que  $\theta$  pertenezca al intervalo  $(1, x_i)$ ).

Si se considera  $(\theta - t)^2$  como función de pérdida, la media a posteriori

$$E[\theta | x_1, x_2, \dots, x_n] = (n+k-1)\tau^{n+k-1} \int_{\tau}^{\infty} \frac{\theta}{\theta^{n+k}} d\theta = \frac{n+k-1}{n+k-2} \tau$$

constituye el estimador Bayes frente a la distribución a priori  $\pi$ , expresado en términos del estadístico  $\tau$ . Nótese que, como la distribución a priori depende de la constante  $k$ , se obtiene en realidad un estimador Bayes diferente para cada una de las distribuciones a priori consideradas.

Tomando como función de pérdida  $|\theta - t|$ , puesto que la función de distribución a posteriori de  $\theta$  es

$$1 - \left(\frac{\tau}{\theta}\right)^{n+k-1} \quad \text{para } \theta > \tau,$$

la mediana a posteriori es

$$2^{1/(n+k-1)}\tau$$

y proporciona el estimador Bayes frente a  $\pi$  con la nueva función de pérdida.

En el caso de la función de pérdida atípica:

$$L(\theta, t) = \begin{cases} 1 - \frac{t}{\theta} & \text{si } t < \theta \\ 1 - \frac{\theta}{t} & \text{si } t > \theta, \end{cases}$$

la pérdida media a posteriori si se da la estimación  $t > \tau$  es

$$\begin{aligned} \int_{\Theta} L(\theta, t) \pi(\theta | x_1, \dots, x_n) d\theta &= \\ &= \left[ \int_{\tau}^t \left(1 - \frac{\theta}{t}\right) \frac{d\theta}{\theta^{n+k}} + \int_t^{\infty} \left(1 - \frac{t}{\theta}\right) \frac{d\theta}{\theta^{n+k}} \right] (n+k-1)\tau^{n+k-1} \\ &= 1 - \left[ \frac{1}{t} \int_{\tau}^t \frac{d\theta}{\theta^{n+k-1}} - t \int_t^{\infty} \frac{d\theta}{\theta^{n+k+1}} \right] (n+k-1)\tau^{n+k-1}. \end{aligned}$$

Prescindiendo del factor independiente de  $t$ , la derivada respecto a  $t$  es

$$\begin{aligned} \frac{1}{t^2} \int_{\tau}^t \frac{d\theta}{\theta^{n+k-1}} - \frac{1}{t^{n+k}} - \int_t^{\infty} \frac{d\theta}{\theta^{n+k+1}} + \frac{1}{t^{n+k}} &= \\ &= \frac{1}{t^2} \frac{1}{n+k-2} \left[ \frac{1}{\tau^{n+k-2}} - \frac{1}{t^{n+k-2}} \right] - \frac{1}{n+k} \frac{1}{t^{n+k}} \\ &= \frac{1}{(n+k-2) \tau^{n+k-2} t^2} - \frac{1}{t^{n+k}} \left[ \frac{1}{n+k-2} + \frac{1}{n+k} \right] \end{aligned}$$

que se anula para

$$t = \left( \frac{2(n+k-1)}{n+k} \right)^{1/(n+k-2)} \tau,$$

proporcionando el estimador Bayes frente a  $\pi$  con la función de pérdida considerada.

### Ejemplo 5.21

Entre las piezas producidas por una máquina hay una proporción  $\theta$  de defectuosas; sin embargo  $\theta$  depende del estado de puesta a punto de la máquina, de manera que puede suponerse que  $\theta$  tiene una distribución a

priori beta  $\beta(p, q)$ , siendo  $p$  y  $q$  conocidos. Para comprobar el estado de la máquina en un cierto momento, se observan  $n$  piezas producidas por ella, anotando los defectos:  $X_i = 1$  ó  $0$ , según que la pieza  $i$ -ésima sea o no defectuosa.

La función de probabilidad de la muestra, condicionada por el valor  $\theta$  del parámetro es entonces

$$f_{\theta}(x_1, x_2, \dots, x_n) = \theta^{x_1 + \dots + x_n} (1 - \theta)^{n - (x_1 + \dots + x_n)}$$

para  $x_1, \dots, x_n \in \{0, 1\}$ ; de manera que la distribución conjunta de  $\theta$  y de la muestra queda caracterizada por

$$f(\theta, x_1, x_2, \dots, x_n) = K \theta^{p-1} (1 - \theta)^{q-1} \theta^{x_1 + \dots + x_n} (1 - \theta)^{n - (x_1 + \dots + x_n)}$$

para  $x_1, \dots, x_n \in \{0, 1\}$  y  $\theta \in (0, 1)$ <sup>5</sup>, siendo  $K = \Gamma(p+q)/(\Gamma(p)\Gamma(q))$ . Llamando  $\tau = \sum_{i=1}^n x_i$ , la distribución marginal de  $(X_1, X_2, \dots, X_n)$  tiene por función de probabilidad

$$f(x_1, x_2, \dots, x_n) = K \int_0^1 \theta^{p-1+\tau} (1-\theta)^{n+q-1-\tau} d\theta = K \frac{\Gamma(p+\tau)\Gamma(n+q-\tau)}{\Gamma(n+p+q)}$$

para  $x_1, \dots, x_n \in \{0, 1\}$ . Y la densidad de la distribución a posteriori de  $\theta$  resulta

$$\pi(\theta | x_1, \dots, x_n) = \frac{\Gamma(n+p+q)}{\Gamma(p+\tau)\Gamma(n+q-\tau)} \theta^{p-1+\tau} (1-\theta)^{n+q-1-\tau}$$

para  $\theta \in (0, 1)$ . La media a posteriori es entonces

$$\begin{aligned} E[\theta | x_1, \dots, x_n] &= \int_0^1 \theta \pi(\theta | x_1, \dots, x_n) d\theta \\ &= \frac{\Gamma(n+p+q)}{\Gamma(p+\tau)\Gamma(n+q-\tau)} \int_0^1 \theta^{p-1+\tau} (1-\theta)^{n+q-1-\tau} d\theta \\ &= \frac{\Gamma(n+p+q)}{\Gamma(p+\tau)\Gamma(n+q-\tau)} \frac{\Gamma(p+1+\tau)\Gamma(n+q-\tau)}{\Gamma(n+p+q+1)} \\ &= \frac{p+\tau}{n+p+q}. \end{aligned}$$

---

<sup>5</sup> En el sentido de que, para  $a < b \in [0, 1]$ ,  $x_1, \dots, x_n \in \{0, 1\}$ .

$$P\{\theta \in (a, b), X_1 = x_1, \dots, X_n = x_n\} = \int_a^b K \theta^{p-1} (1-\theta)^{q-1} \theta^{x_1 + \dots + x_n} (1-\theta)^{n - (x_1 + \dots + x_n)} d\theta$$

De forma que, con función de pérdida  $L(\theta, t) = (\theta - t)^2$  el estimador Bayes de  $\theta$  frente a la distribución a priori beta( $p, q$ ) es

$$\frac{\bar{X} + p/n}{1 + p/n + q/n}.$$

Puede interpretarse que la información a priori (representada en la fórmula por  $p$  y  $q$ ) influye en el estimador natural  $\bar{X}$  para hacerle preocuparse más de los valores de  $\theta$  con alta probabilidad a priori y despreciar relativamente los valores de probabilidad a priori baja. A medida que  $n$  crece, incrementando la evidencia aportada por la muestra, el efecto de la distribución a priori pierde importancia y el estimador se approxima a  $\bar{X}$ .

### Ejemplo 5.22

Se elige un número  $\theta$  con distribución geométrica:  $(1-p)p^{\theta-1}$  para  $\theta = 1, 2, 3, \dots$ ; a continuación se observa una muestra aleatoria simple  $(X_1, X_2, \dots, X_n)$  de la población con función de densidad

$$f_\theta(x) = x^{\theta-1} \quad \text{para } x \in [0, \theta^{1/\theta}].$$

La distribución conjunta de  $\theta$  y  $(X_1, X_2, \dots, X_n)$  viene dada entonces por:

$$f(\theta, x_1, x_2, \dots, x_n) = (1-p)p^{\theta-1}(x_1 x_2 \cdots x_n)^{\theta-1}$$

para  $\theta = 1, 2, \dots$  y  $0 < x_{(n)} = \max_{1 \leq i \leq n} x_i \leq \theta^{1/\theta}$  (que expresa la probabilidad del valor  $\theta$  y la densidad de probabilidad en el punto  $(x_1, x_2, \dots, x_n)$ ).

Llamemos  $\tau = \prod_{i=1}^n x_i$  y nótese que es  $1 \leq \theta^{1/\theta} \leq 3^{1/3}$  para cualquier  $\theta$  natural. Entonces, fijada una muestra cualquiera  $(x_1, x_2, \dots, x_n)$  que verifique  $x_{(n)} \leq 1$ , la densidad marginal de  $(X_1, X_2, \dots, X_n)$  en dicho punto, es

$$f(x_1, x_2, \dots, x_n) = \sum_{\theta=1}^{\infty} (1-p)(p\tau)^{\theta-1} = \frac{1-p}{1-p\tau}.$$

En cambio, si  $x_{(n)} \in (1, 3^{1/3}]$ , la desigualdad  $\theta^{1/\theta} \geq x_{(n)}$  se verifica en un intervalo de números naturales  $[\theta_1, \theta_2]$  (siendo  $\theta_1$  y  $\theta_2$  las raíces enteras, por exceso y por defecto respectivamente, de la ecuación  $\theta^{1/\theta} = x_{(n)}$ ). De manera que la densidad marginal de  $(X_1, X_2, \dots, X_n)$  en dichos puntos, es

$$f(x_1, x_2, \dots, x_n) = \sum_{\theta=\theta_1}^{\theta_2} (1-p)(p\tau)^{\theta-1} = \frac{(1-p)(p\tau)^{\theta_1-1} [1 - (p\tau)^{\theta_2-\theta_1+1}]}{1-p\tau}.$$

La función de probabilidad de la distribución a posteriori de  $\theta$  es pues

$$\pi(\theta | x_1, x_2, \dots, x_n) = \begin{cases} (1 - p\tau)(p\tau)^{\theta-1} & \text{para } \theta = 1, 2, 3, \dots \\ \frac{(1 - p\tau)(p\tau)^{\theta-\theta_1}}{1 - (p\tau)^{\theta_2-\theta_1+1}} & \text{para } \theta = \theta_1, \theta_1 + 1, \dots, \theta_2, \end{cases}$$

según que sea  $x_{(n)} \leq 1$  ó  $x_{(n)} > 1$ . La media a posteriori resulta

$$E[\theta | X_1, X_2, \dots, X_n] =$$

$$= \begin{cases} \sum_{\theta=1}^{\infty} \theta(1 - p\tau)(p\tau)^{\theta-1} = \frac{1}{1 - p\tau}, \\ \sum_{\theta=\theta_1}^{\theta_2} \theta \frac{(1 - p\tau)}{1 - (p\tau)^{\theta_2-\theta_1+1}} (p\tau)^{\theta-\theta_1} = \frac{p\tau}{1 - p\tau} + \frac{\theta_1 - (\theta_2 + 1)(p\tau)^{\theta_2-\theta_1+1}}{1 - (p\tau)^{\theta_2-\theta_1+1}}, \end{cases}$$

según que sea  $x_{(n)} \leq 1$  ó  $x_{(n)} > 1$ . Tal es pues el estimador Bayes, con pérdida cuadrática, frente a la distribución geométrica (de parámetro  $p$ ), en términos de los estadísticos  $x_{(n)}, \tau, \theta_1$  y  $\theta_2$ .

La determinación de la distribución a posteriori de  $\theta$  y de su media es, a veces, inviable explícitamente y habría que emplear integración numérica para obtener  $\int_{\Theta} f_{\theta}(x_1, x_2, \dots, x_n) \pi(\theta) d\theta$  y  $\int_{\Theta} \theta \pi(\theta | x_1, x_2, \dots, x_n) d\theta$ .

Puede observarse en los ejemplos anteriores que el método Bayes no produce, en general, estimadores insesgados. Más aún, en condiciones muy frecuentes, no puede producirlos<sup>6</sup>.

---

<sup>6</sup> Concretamente, si  $T$  es la media a posteriori,  $T$  no puede ser estimador centrado de  $\theta$ , a no ser que la distribución a posteriori sea causal (en cuyo caso  $V_{\theta}(T) = 0$ ), o bien  $T$  ó  $\theta$  tengan varianza infinita. En efecto, para cualquier estimador  $T$ , se tiene

$$E[(E_{\theta}[T] - E[T])(T - E_{\theta}[T])] = E[(E_{\theta}[T] - E[T])E_{\theta}[T - E_{\theta}[T]]] = 0$$

y por tanto

$$\begin{aligned} V(T) &= E[(T - E_{\theta}[T] + E_{\theta}[T] - E[T])^2] \\ &= E[(T - E_{\theta}[T])^2] + E[(E_{\theta}[T] - E[T])^2] = E[V_{\theta}(T)] + V(E_{\theta}[T]). \end{aligned}$$

En el caso en que  $T$  fuese centrado, resulta  $V(T) = E[V_{\theta}(T)] + V(\theta)$ . De forma similar,

$$V(\theta) = E[V(\theta | X_1, \dots, X_n)] + V(E[\theta | X_1, \dots, X_n]).$$

En cambio criterio Bayes siempre da lugar a estimadores *admisibles*, porque si  $T'$  fuese un estimador preferible al estimador Bayes  $T$  (frente a una cierta distribución a priori  $\pi$ ) sería

$$R_{T'}(\theta) \leq R_T(\theta) \quad \text{para cualquier } \theta \in \Theta,$$

con lo cual  $r_{T'}(\pi) \leq r_T(\pi)$  y  $T$  no sería el estimador Bayes frente a  $\pi$ . (A no ser que  $r'_T(\pi) = r_T(\pi)$  y ambos fuesen estimadores Bayes frente a  $\pi$ ; en cuyo caso, el estimador Bayes  $T$  no sería admisible, pero sí el estimador Bayes  $T'$ ).

También es notable que, la distribución a posteriori es función del estadístico minimal suficiente. De hecho, si  $T$  es cualquier estadístico suficiente, en virtud del teorema de factorización será

$$f_\theta(x_1, x_2, \dots, x_n) = g_\theta(T(x_1, x_2, \dots, x_n))h(x_1, x_2, \dots, x_n),$$

con lo cual

$$f(x_1, x_2, \dots, x_n) = h(x_1, x_2, \dots, x_n) \int_{\Theta} g_\theta(T(x_1, x_2, \dots, x_n))\pi(\theta)d\theta$$

y

$$\pi(\theta | x_1, x_2, \dots, x_n) = \frac{\pi(\theta)g_\theta(T(x_1, x_2, \dots, x_n))}{\int_{\Theta} g_\theta(T(x_1, x_2, \dots, x_n))\pi(\theta)d\theta}.$$

Por tanto, el valor  $t$  que minimiza  $\int_{\Theta} L(\theta, t)\pi(\theta | x_1, x_2, \dots, x_n)d\theta$  coincide para todas las muestras que asignan el mismo valor a  $T$ . Así que, *cualquier estimador Bayes es función únicamente del estadístico minimal suficiente*.

Una última observación interesante es la siguiente: *si  $T$  es el estimador Bayes, frente a una cierta distribución a priori  $\pi_0$  y su riesgo  $R_T(\theta)$  es independiente de  $\theta$ , entonces  $T$  es también el estimador minimax*. De hecho, por ser  $R_T(\theta)$  constante,

$$r_T(\pi_0) = \int_{\Theta} R_T(\theta)\pi_0(\theta)d\theta = \max_{\theta \in \Theta} R_T(\theta)$$

---

Luego si  $T$  es la media a posteriori y fuese centrada, sería

$$V(T) = E[V_\theta(T)] + E[V(\theta|X_1, \dots, X_n)] + V(T).$$

de donde  $E[V_\theta(T)] + E[V(\theta|X_1, \dots, X_n)] = 0$ ; y, como  $V_\theta(T)$  y  $V(\theta|X_1, \dots, X_n)$  son no negativos, se concluye que  $V(\theta|X_1, \dots, X_n) = 0$  —o sea que la distribución a posteriori es causal.

y para cualquier otro estimador  $T'$

$$r_{T'}(\pi_0) = \int_{\Theta} R_{T'}(\theta) \pi_0(\theta) d\theta \leq \max_{\theta \in \Theta} R_{T'}(\theta).$$

Como  $r_T(\pi_0) \leq r_{T'}(\pi_0)$ , se obtiene

$$\max_{\theta \in \Theta} R_T(\theta) \leq \max_{\theta \in \Theta} R_{T'}(\theta).$$

Luego  $T$  proporciona el mínimo valor de  $\max_{\theta} R_{T'}(\theta)$ , siendo así el estimador minimax.

Se cuenta por tanto con un nuevo procedimiento para determinar el estimador minimax.

### Ejemplo 5.23

En la situación del ejemplo 5.21, hemos visto que el estimador Bayes frente a la distribución beta  $\beta(p, q)$ , considerando como pérdida  $(\theta - t)^2$ , es  $(p + \sum_1^n X_i) / (n + p + q)$ . Además el riesgo de cualquier estimador  $T$  de la forma  $\alpha \sum_1^n X_i + \beta$  es

$$\begin{aligned} R_T(\theta) &= E_{\theta}[(\theta - \alpha n \bar{X} - \beta)^2] = E_{\theta} \left[ (\alpha n(\theta - \bar{X}) + (1 - \alpha n)\theta - \beta)^2 \right] \\ &= \alpha^2 n^2 \frac{\theta(1 - \theta)}{n} + (1 - \alpha n)^2 \theta^2 + \beta^2 - 2(1 - \alpha n)\theta\beta \\ &= [(1 - \alpha n)^2 - \alpha^2 n]\theta^2 + [\alpha^2 n - 2(1 - \alpha n)\beta]\theta + \beta^2. \end{aligned}$$

De manera que se puede conseguir que el riesgo  $R_T(\theta)$  sea constante tomando  $\alpha$  y  $\beta$  como soluciones de las ecuaciones

$$(1 - \alpha n)^2 - \alpha^2 n = 0 \quad \text{y} \quad \alpha^2 n - 2(1 - \alpha n)\beta = 0$$

es decir

$$\alpha = \frac{1}{\sqrt{n}(1 + \sqrt{n})} \quad \text{y} \quad \beta = \frac{1}{2(1 + \sqrt{n})}$$

El estimador obtenido:

$$\frac{\sqrt{n}\bar{X} + 1/2}{1 + \sqrt{n}}$$

tiene riesgo independiente de  $\theta$  y es el estimador Bayes frente a la distribución  $\beta(\sqrt{n}/2, \sqrt{n}/2)$ . Es, por tanto, el estimador minimax de  $\theta$ ; no dentro de una clase reducida de estimadores, como en el ejemplo 5.19, sino en general. Por supuesto, para  $n = 1$  se obtiene el resultado hallado en el ejemplo 5.18.

Naturalmente, no siempre es fácil encontrar un estimador Bayes, frente a cierta distribución a priori, cuyo riego sea constante. En este sentido, normalmente, el criterio minimax sigue siendo difícil de aplicar, además de ser —como ya se ha apuntado— excesivamente conservador en el contexto de los problemas de inferencia.

En cambio, el criterio Bayes está perfectamente adecuado a la situación en que se aplica y debe ser utilizado, por tanto, siempre que haya una manera razonable de contar con una distribución a priori para el parámetro.

## Ejercicios

### Ejercicio 5.1

El número diario de ingresos en un servicio de urgencia sigue una distribución de Poisson de parámetro  $\theta > 0$ , con independencia entre un día y otro. Se van a observar los números de ingresos,  $X_1, \dots, X_n$ , durante  $n$  días, a fin de estimar  $\theta$  para hacer las previsiones de mantenimiento del servicio. Se supone que los errores en la estimación de  $\theta$  producen una pérdida cuadrática y se sabe que  $T = \sum_{i=1}^n X_i$  es un estadístico suficiente (cf. ejemplo 5.11).

- a) Determinar entre los estimadores  $T_k = k T$  con  $k > 0$ , cuáles son admisibles y cuáles no lo son.
- b) Determinar  $k$  para que el estimador  $T_k$  sea centrado.
- c) Si se puede hacer crecer el tamaño muestral  $n$ , ¿qué condición ha de cumplir la sucesión de constantes  $k_n$  para obtener una sucesión consistente de estimadores?
- d) Determinar entre los estimadores  $T_k$  el estimador Bayes si la distribución a priori es  $\gamma(p, a)$ .
- e) Hallar el estimador Bayes, frente a dicha distribución, en el conjunto de todos los estimadores.

### Ejercicio 5.2

La distribución de una población tiene media y varianza  $(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$ . Para estimar  $\mu$ , con pérdida cuadrática, se consideran los estimadores

lineales, basados en una muestra aleatoria simple de tamaño  $n$ :

$$T = \sum_{i=1}^n \lambda_i X_i$$

- a) Determinar cuáles de tales estimadores son insesgados.
- b) Comprobar que sólo son admisibles aquellos estimadores tales que, para cada  $i = 1, \dots, n$ , es  $\lambda_i = \lambda$  siendo  $\lambda \in [0, 1/n]$ . Determinar los estimadores admisibles si se sabe que  $\sigma^2 > \varepsilon\mu^2$ .
- c) Cuando  $n$  crece, hallar la condición que deben cumplir los coeficientes  $\lambda_i(n)$  para que  $T$  sea consistente en media cuadrática.

### Ejercicio 5.3

Estudiar si los parámetros ( $a \in \mathbb{R}$ ,  $b > 0$ ) de las siguientes familias de densidades son de posición o escala.

(a)  $N(a, b)$ .

(b) Uniforme en  $(a - b, a + b)$ .

(c)  $f_{a,b}(x) = \frac{1}{2b} e^{-|x-a|/b} \quad (x \in \mathbb{R})$ .

(d)  $f_{a,b}(x) = \frac{b}{\pi(b^2 + (x - a)^2)} \quad (x \in \mathbb{R})$ .

(e)  $f_{a,b}(x) = \frac{be^{b(a-x)}}{(1 + e^{b(a-x)})^2} \quad (x \in \mathbb{R})$ .

(f)  $f_{a,b}(x) = (b+1)(x-a)^b \quad \text{para } a < x < a+1$ ,

### Ejercicio 5.4

Se dispone de una muestra aleatoria simple de tamaño  $n$  de una distribución uniforme en  $(\theta - 1, \theta + 1)$ .

- a) Comprobar que  $T = ((X_{(n)} + X_{(1)})/2, X_{(n)} - X_{(1)})$  es un estadístico suficiente minimal.
- b) Determinar la distribución en el muestreo de  $T$  y deducir que la distribución de  $X_{(n)} - X_{(1)}$  no depende de  $\theta$ .

c) ¿No es paradójico que un estadístico minimal suficiente tenga una componente cuya distribución no depende del parámetro, de cuya observación no puede obtenerse información sobre él?

### Ejercicio 5.5

Determinar un estadístico suficiente minimal, de una muestra aleatoria simple de tamaño  $n$ , para cada una de las poblaciones con densidad:

$$(a) f_{\theta}(x) = \left(\frac{\theta}{x}\right)^{\theta+1} \quad \text{para } x > \theta.$$

$$(b) f_{\theta}(x) = \frac{e^{\theta-x}}{[1 + e^{\theta-x}]^2} \quad \text{para } x \in \mathbb{R}.$$

### Ejercicio 5.6

Comprobar que son familias de tipo exponencial cada una de las siguientes familias de distribuciones y determinar en cada caso un estadístico minimal suficiente para una muestra aleatoria simple de tamaño  $n$ .

- (a) Binomial  $B(k, \theta)$  ( $k$  fijo).
- (b) Binomial negativa  $BN(k, \theta)$  ( $k$  fijo).
- (c) Gamma  $\gamma(\theta_1, \theta_2)$ .
- (d) Beta  $\beta(\theta_1, \theta_2)$ .
- (e) Distribución de Rayleigh, de densidad

$$f_{\theta}(x) = \frac{x}{\theta^2} \exp\left\{-\frac{x^2}{2\theta^2}\right\} \quad \text{para } x > 0.$$

### Ejercicio 5.7

Comprobar que son familias de tipo exponencial cada una de las siguientes familias de distribuciones y determinar en cada caso un estadístico minimal suficiente para una muestra aleatoria simple de tamaño  $n$ .

- (a) La distribución de dos variables independientes  $X$  e  $Y$  con distribuciones  $\gamma(\theta_1 + \theta_2, \theta_1)$  y  $\gamma(\theta_1 + \theta_2, \theta_2)$ , respectivamente.

- (b) La distribución trinomial de parámetros  $k, \theta_1, \theta_2$ .
- (c) La distribución de densidad  $f_\theta(x) = 2\Gamma(1/4)^{-1}e^{-(x-\theta)^4}$  para  $x \in \mathbb{R}$ .

### Ejercicio 5.8

La duración de las reparaciones efectuadas en un taller se supone exponencial de parámetro  $\theta$  desconocido. La experiencia con otros talleres del mismo tipo indica que  $\theta$  tiene distribución a priori  $\gamma(p, a)$ , siendo  $p$  y  $a$  conocidos. Se observan las duraciones,  $X_i$ , de  $n$  reparaciones efectuadas en el taller.

- Determinar la distribución a posteriori de  $\theta$ .
- Determinar la distribución marginal del estadístico  $T = \sum_{i=1}^n X_i$ .
- Determinar el estimador Bayes para  $\theta$  suponiendo función de pérdida cuadrática. Calcular su riesgo Bayes.
- Determinar el estimador Bayes para  $1/\theta$ , si la función de pérdida es cuadrática.

### Ejercicio 5.9

Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución normal  $N(\theta, \sigma)$  con  $\theta$  desconocido y  $\sigma$  conocido. Determinar el estimador Bayes para  $\theta$ , con pérdida cuadrática, si la distribución a priori de  $\theta$  es  $N(\mu, r)$ . Calcular el riesgo Bayes de dicho estimador.

Deducir de lo anterior que  $\bar{X}$  es el estimador minimax de  $\theta$  si la función de pérdida es cuadrática.

### Ejercicio 5.10

De una urna que contiene una proporción  $\theta$  de bolas blancas, se realizan  $n$  extracciones con reemplazamiento.

- Hallar el estimador Bayes de  $\theta$ , frente a la distribución a priori  $\beta(p, q)$ , si la función de pérdida es

$$L(\theta, t) = \frac{(\theta - t)^2}{\theta(1 - \theta)}.$$

- Deducir cual es el estimador minimax de  $\theta$  con la función de pérdida anterior.

c) Determinar el estimador minimax de  $\theta$  cuando la pérdida es

$$L(\theta, t) = \frac{(\theta - t)^2}{\theta^2}.$$

## Solución de los ejercicios

### Ejercicio 5.1

a)  $T$  tiene distribución de Poisson de parámetro  $n\theta$ , luego

$$\mathbb{E}_\theta[T] = n\theta, \quad V_\theta(T) = n\theta \quad \text{y} \quad \mathbb{E}_\theta[T^2] = n\theta + n^2\theta^2.$$

El riesgo del estimador  $T_k = kT$  es entonces

$$R_{T_k}(\theta) = \mathbb{E}_\theta[(\theta - kT)^2] = \theta^2 + k^2\mathbb{E}_\theta[T^2] - 2\theta k\mathbb{E}_\theta[T] = (nk - 1)^2\theta^2 + k^2n\theta.$$

Para  $k = 1/n$  se obtiene el estimador  $T_{1/n} = \bar{X}$ , cuyo riesgo se reduce a  $\theta/n$ . Por otra parte, si  $k > 1/n$ , el segundo sumando es superior a  $\theta/n$  (y el primero positivo), de forma que  $R_{T_k}(\theta) > R_{\bar{X}}(\theta)$ , para todo  $\theta > 0$ , y el estimador  $\bar{X}$  es preferible a cualquier  $T_k$  con  $k > 1/n$ . Dichos estimadores son, por tanto, inadmisibles.

En cambio, la derivada de  $R_{T_k}(\theta)$  respecto a  $k$ :  $2n\theta[\theta(nk - 1) + k]$ , se anula para  $k = \theta/(n\theta + 1)$  (más exactamente es negativa para  $k$  inferior a dicho valor y positiva en caso contrario). Luego, fijado un valor de  $\theta$ , el valor de  $k$  que minimiza  $R_{T_k}(\theta)$  es  $k = \theta/(n\theta + 1)$ , que es siempre inferior a  $1/n$ . Los estimadores  $T_k$  con  $0 < k < 1/n$  son, por consiguiente, admisibles.

b)  $T_k$  es insesgado si  $\mathbb{E}_\theta[T_k] = k\mathbb{E}_\theta[T] = kn\theta$  coincide con  $\theta$ . Para ello es inmediato que ha de ser  $k = 1/n$ ; de manera que la media muestral,  $\bar{X}$ , es el único estimador de la forma  $kT$  que es insesgado.

c) Cuando  $n \rightarrow \infty$ , según la ley de los grandes números, se verifica  $\bar{X} \rightarrow \theta$  con probabilidad uno. Así pues,

$$T_{k_n} = k_n n\bar{X} \longrightarrow \theta \quad \text{casi seguro}$$

si y sólo si  $k_n n \rightarrow 1$ . Dicho de otra manera, la condición para que la sucesión de estimadores  $T_{k_n}$  sea consistente es que sea  $k_n \sim 1/n$ .

d) Frente a la distribución a priori  $\gamma(p, a)$ , cuya densidad representaremos por  $\pi(\theta)$ , el riesgo Bayes del estimador  $T_k$  es

$$r_{T_k}(\pi) = \int_0^\infty [\theta^2(nk - 1)^2 + k^2n\theta]\pi(\theta) d\theta = (nk - 1)^2 \left( \frac{p}{a^2} + \frac{p^2}{a^2} \right) + k^2n\frac{p}{a}$$

La derivada de  $r_{T_k}(\pi)$  respecto a  $k$  es entonces

$$2n\frac{p}{a} \left[ (nk - 1)\frac{p+1}{a} + k \right],$$

de forma que el mínimo riesgo Bayes (entre los estimadores  $T_k$ ) se alcanza para

$$k = \frac{p+1}{n(p+1)+a}.$$

e) Como la densidad a priori es

$$\pi(\theta) = \frac{a^p}{\Gamma(p)} \theta^{p-1} e^{-a\theta}$$

y la función de probabilidad de las observaciones viene dada por

$$f_\theta(x_1, \dots, x_n) = e^{-n\theta} \frac{\theta^{T(x_1, \dots, x_n)}}{x_1! \cdots x_n!},$$

la densidad a posteriori de  $\theta$  será

$$\pi(\theta | x_1, \dots, x_n) = \frac{\theta^{p-1} e^{-a\theta} e^{-n\theta} \theta^T}{\int_0^\infty \theta^{p-1} e^{-a\theta} e^{-n\theta} \theta^T d\theta} = C \theta^{T+p-1} e^{-(a+n)\theta}$$

donde  $C$  es una constante independiente de  $\theta$ . Se trata pues de una distribución  $\gamma(T+p, a+n)$  (y, por tanto,  $C = (a+n)^{T+p}/\Gamma(T+p)$ ). Como la pérdida es cuadrática, el estimador Bayes es la media de la distribución a posteriori

$$\mathbb{E}[\theta | x_1, \dots, x_n] = \frac{T+p}{a+n}.$$

### Ejercicio 5.2

a) Desde luego  $\mathbb{E}[X_i] = \mu$ , así que

$$\mathbb{E}[T] = \sum_{i=1}^n \lambda_i \mathbb{E}[X_i] = \mu \sum_{i=1}^n \lambda_i$$

y el estimador es insesgado si y sólo si  $\sum_{i=1}^n \lambda_i = 1$ .

b) El riesgo de  $T$  es

$$\begin{aligned} \mathbb{E}[(T - \mu)^2] &= \mathbb{E} \left[ \left( \sum_{i=1}^n \lambda_i (X_i - \mu) - \left( 1 - \sum_{i=1}^n \lambda_i \right) \mu \right)^2 \right] \\ &= \sigma^2 \sum_{i=1}^n \lambda_i^2 + \left( 1 - \sum_{i=1}^n \lambda_i \right)^2 \mu^2 \end{aligned}$$

(habiéndola cuenta que  $E[(X_i - \mu)(X_j - \mu)] = 0$ , si  $i \neq j$ , y  $E[X_i - \mu] = 0$ ).

Si se toma  $\lambda'_i = 1/n \sum_{i=1}^n \lambda_i$ , para todo  $i$ , se obtiene como riesgo

$$n\sigma^2 \left( \frac{1}{n} \sum_{i=1}^n \lambda_i \right)^2 + \left( 1 - \sum_{i=1}^n \lambda_i \right)^2 \mu^2,$$

que resulta menor que el anterior puesto que  $1/n \sum_{i=1}^n \lambda_i^2 > (1/n \sum_{i=1}^n \lambda_i)^2$ . Por tanto son inadmisibles los estimadores que no verifiquen que todos los  $\lambda_i$  tienen el mismo valor  $\lambda$ .

Para estos últimos, el riesgo se reduce a

$$n\sigma^2\lambda^2 + (1 - n\lambda)^2\mu^2$$

y, en particular,  $\sigma^2/n$  en el caso  $\lambda = 1/n$ . La expresión anterior es superior a  $\sigma^2/n$  siempre que sea  $|\lambda| > 1/n$ . Por otra parte, el riesgo asociado al valor negativo,  $-\lambda$ , es superior al riesgo asociado a  $\lambda$ . Luego, son inadmisibles todos los estimadores  $\lambda \sum_{i=1}^n X_i$  con  $\lambda \notin [0, 1/n]$ .

En general, los valores de  $\lambda \in [0, 1/n]$  son admisibles. En efecto, puesto que la derivada respecto a  $\lambda$  de la función de riesgo es  $2n[(\sigma^2 + n\mu^2)\lambda - \mu^2]$ , cada valor  $\lambda = 1/(k+n)$ , con  $k > 0$ , proporciona el mínimo riesgo en el caso en que fuese  $\sigma^2 = k\mu^2$ .

Si la distribución de la población verifica  $\sigma^2 > \varepsilon\mu^2$ , la derivada respecto a  $\lambda$  del riesgo verifica

$$2n[(\sigma^2 + n\mu^2)\lambda - \mu^2] > 2n[(n + \varepsilon)\lambda - 1]\mu^2$$

y es positiva para  $\lambda > 1/(n + \varepsilon)$ . Los estimadores correspondientes a tales valores de  $\lambda$  —y, en particular, la media muestral— son, pues, inadmisibles como estimadores de  $\mu$ , con pérdida cuadrática, para toda distribución en la que se verifique  $\sigma^2 > \varepsilon\mu^2$ . Por ejemplo, para la distribución exponencial, como  $\sigma^2 = \mu^2$ , el único estimador lineal admissible es  $1/(n+1) \sum_{i=1}^n X_i$ .

c) Puesto que

$$E[(T_n - \mu)^2] = \sigma^2 \sum_{i=1}^n \lambda_i^2(n) + \left( 1 - \sum_{i=1}^n \lambda_i(n) \right)^2 \mu^2,$$

la consistencia en media cuadrática equivale a que ambos sumandos tiendan a cero al crecer  $n$ . Es decir

$$\sum_{i=1}^n \lambda_i(n) \rightarrow 1 \quad \text{y} \quad \sum_{i=1}^n \lambda_i^2(n) \rightarrow 0.$$

Ahora bien, como

$$\left( \max_{i=1,\dots,n} \lambda_i(n) \right)^2 \leq \sum_{i=1}^n \lambda_i^2(n) \leq \max_{i=1,\dots,n} \lambda_i(n) \sum_{i=1}^n \lambda_i(n),$$

las condiciones anteriores equivalen a

$$\sum_{i=1}^n \lambda_i(n) \rightarrow 1 \quad \text{y} \quad \max_{i=1,\dots,n} \lambda_i(n) \rightarrow 0.$$

Los estimadores de la forma  $\lambda(n) \sum_{i=1}^n X_i$ , con  $n\lambda(n) \rightarrow 1$  son consistentes en media cuadrática y, también, casi seguramente (en virtud de la ley de los grandes números). Hay muchos otros coeficientes que dan un estimador consistente de  $\mu$ ; por ejemplo:  $\lambda_i(n) = [(n+i) \log 2]^{-1}$ .

### Ejercicio 5.3

a) Si  $Z$  es una variable con distribución  $N(0, 1)$ , la variable  $X = bZ + a$  tiene distribución  $N(a, b)$ . Por consiguiente, en dichas distribuciones  $a$  es un parámetro de posición y  $b$  un parámetro de escala.

b) La distribución uniforme en  $(a - b, a + b)$  tiene densidad

$$f_{a,b}(x) = \frac{1}{2b} I_{\{a-b < x < a+b\}} = \frac{1}{b} f_{0,1}\left(\frac{x-a}{b}\right)$$

en términos de la densidad  $f_{0,1}$ . Luego  $a$  es un parámetro de posición y  $b$  un parámetro de escala.

c) Si  $Z$  es una variable con densidad

$$f(z) = \frac{1}{2} e^{-|z|} \quad \text{para } z \in \mathbb{R},$$

la densidad de  $X = a + bZ$  es

$$\frac{1}{b} f\left(\frac{x-a}{b}\right) = \frac{1}{2b} e^{-|x-a|/b},$$

luego  $a$  actúa como parámetro de posición y  $b$  como parámetro de escala.

d) Puesto que

$$f_{a,b}(x) = \frac{1}{b} \frac{1}{\pi(1+z^2)} \quad \text{con } z = \frac{x-a}{b},$$

la familia de densidades se obtiene de la distribución de Cauchy, introduciendo el factor de escala  $b$  y el parámetro de posición  $a$ .

e) Como

$$f_{0,1}(z) = \frac{e^{-z}}{(1 + e^{-z})^2},$$

si  $Z$  tiene densidad  $f_{0,1}$ , la densidad de  $X = bZ + a$  es

$$\frac{e^{(a-x)/b}}{b(1 + e^{(a-x)/b})^2} = f_{a,1/b}(x);$$

luego  $a$  es un parámetro de posición, mientras que el parámetro de escala es  $1/b$ .

f) Como

$$f_{a+c,b}(x+c) = (b+1)(x-a)^b \quad \text{para } a+c < x+c < a+c+1$$

coincide con  $f_{a,b}(x)$ ,  $a$  es un parámetro de posición. Sin embargo

$$f_{a,b}(x) = (b+1)(x-a)^b \neq \frac{1}{b} f_{a,1}(x) = \frac{1}{b} 2(x-a)$$

luego  $b$  no es un parámetro de escala.

#### Ejercicio 5.4

a) La densidad teórica es

$$f_\theta(x) = \frac{1}{2} I_{\{\theta-1 < x < \theta+1\}};$$

así que la de una muestra aleatoria simple de tamaño  $n$  será

$$f_\theta(x_1, \dots, x_n) = \frac{1}{2^n} I_{\{\theta-1 < x_{(1)} \leq x_{(n)} < \theta+1\}}$$

que sólo depende de  $x_{(1)}$  y  $x_{(n)}$ . Por tanto  $(X_{(1)}, X_{(n)})$  es un estadístico suficiente.

Además, dadas dos muestras  $(x_1, \dots, x_n)$  y  $(x'_1, \dots, x'_n)$ , el cociente

$$\frac{f_\theta(x_1, \dots, x_n)}{f_\theta(x'_1, \dots, x'_n)} = \frac{I_{\{x_{(n)}-1 < \theta < x_{(1)}+1\}}}{I_{\{x'_{(n)}-1 < \theta < x'_{(1)}+1\}}}$$

tiene un valor constante, independiente de  $\theta$ , si y sólo si

$$I_{\{x_{(n)}-1 < \theta < x_{(1)}+1\}} = I_{\{x'_{(n)}-1 < \theta < x'_{(1)}+1\}};$$

lo cual ocurre cuando  $x_{(1)} = x'_{(1)}$  y  $x_{(n)} = x'_{(n)}$ , y sólo en este caso. Por consiguiente, según el teorema 5.2,  $(X_{(1)}, X_{(n)})$  es un estadístico minimal suficiente.

Evidentemente, existe una correspondencia biunívoca entre los valores de  $(X_{(1)}, X_{(n)})$  y los de  $T = ((X_{(n)} + X_{(1)})/2, X_{(n)} - X_{(1)})$ . Así que,  $T$  es también minimal suficiente.

b) La densidad de  $(X_{(1)}, X_{(n)})$  es

$$g_\theta(x, y) = \frac{n(n-1)}{2^n} (y-x)^{n-2} \quad \text{para } \theta-1 < x \leq y \leq \theta+1;$$

luego  $U = (X_{(n)} + X_{(1)})/2$  y  $V = X_{(n)} - X_{(1)}$  (mediante el cambio  $x = u-v/2$ ,  $y = u+v/2$ ) tienen densidad conjunta

$$\frac{n(n-1)}{2^n} v^{n-2} \quad \text{para } 0 < v < 2, |u-\theta| < 1-v/2.$$

La distribución marginal de  $V$  tiene entonces densidad

$$\frac{n(n-1)}{2^n} v^{n-2}(2-v) \quad \text{para } 0 < v < 2,$$

que no depende de  $\theta$ . Tales estadísticos, cuya distribución es independiente del parámetro poblacional, se denominan *anciliares* porque su observación (aislada) no aporta ninguna información acerca del parámetro.

c) Aunque la distribución marginal del estadístico  $V$  no depende de  $\theta$ , su distribución condicionada por  $U$  si depende de  $\theta$ . Concretamente, puesto que la distribución marginal de  $U$  tiene densidad

$$\frac{n}{2} (1+|u-\theta|)^{n-1} \quad \text{para } \theta-1 < u < \theta+1,$$

la densidad condicionada de  $V$  por  $U = u \in (\theta-1, \theta+1)$  es

$$\frac{n-1}{2^{n-1}} \frac{v^{n-2}}{(1+|u-\theta|)^{n-1}} \quad \text{para } 0 < v < 2(1+|u-\theta|),$$

que sí depende de  $\theta$ . Por tanto, después de observar  $U$ , la observación de  $V$  aporta información acerca de  $\theta$ .

Visto al revés, la observación inicial de  $V$  no revela nada sobre la posición de  $\theta$ , pero al observar posteriormente  $U$ , se sabrá que  $|U-\theta| < 1-V/2$ ; mientras que, si no se observa  $V$ , sólo se conocerá que  $|U-\theta| < 1$ .

**Ejercicio 5.5**

- a) La densidad de una muestra aleatoria simple de tamaño  $n$  es

$$f_\theta(x_1, \dots, x_n) = \frac{\theta^{n(\theta+1)}}{(x_1 \cdots x_n)^{\theta+1}} I_{\{x_{(1)} > \theta\}}$$

siendo  $x_{(1)}$  la más pequeña de las  $x_i$ . Según el teorema de factorización, el estadístico

$$T = \left( X_{(1)}, \prod_{i=1}^n X_i \right)$$

es suficiente puesto que  $f_\theta$  sólo depende de las  $x_i$  a través de  $T$ . Parece claro que es imposible factorizar  $f_\theta$  en términos de un estadístico que contenga menos información que  $T$ ; de forma que  $T$  debe ser minimal suficiente. Para comprobarlo formalmente se puede utilizar el teorema 5.2, formando el cociente

$$\frac{f_\theta(x_1, \dots, x_n)}{f_\theta(x'_1, \dots, x'_n)} = \left( \frac{x'_1 \cdots x'_n}{x_1 \cdots x_n} \right)^{\theta+1} \frac{I_{\{x_{(1)} > \theta\}}}{I_{\{x'_{(1)} > \theta\}}}.$$

Para que dicho cociente no dependa de  $\theta$  es necesario y suficiente que sea

$$\prod_{i=1}^n x_i = \prod_{i=1}^n x'_i \quad \text{y} \quad x_{(1)} = x'_{(1)}.$$

- b) La densidad de la muestra es

$$f_\theta(x_1, \dots, x_n) = \exp \left\{ - \sum_{i=1}^n (x_i - \theta) \right\} \prod_{i=1}^n [1 + e^{\theta - x_i}]^{-2}$$

cuyo valor no depende del orden de las  $x_i$ , lo cual asegura que la muestra ordenada  $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$  es un estadístico suficiente (como ocurre siempre que el muestreo es aleatorio simple).

No parece posible factorizar la densidad en términos de una función de las  $x_i$  de dimensión menor que  $n$ . Para comprobarlo —es decir, para ver que la muestra ordenada es un estadístico minimal suficiente— obsérvese que el cociente

$$\frac{f_\theta(x_1, \dots, x_n)}{f_\theta(y_1, \dots, y_n)} = \frac{\exp \left\{ \sum_{i=1}^n y_i \right\} \prod_{i=1}^n [1 + e^{\theta - y_i}]^2}{\exp \left\{ \sum_{i=1}^n x_i \right\} \prod_{i=1}^n [1 + e^{\theta - x_i}]^2} = \frac{\prod_{i=1}^n (e^{y_i} + e^\theta)^2}{\prod_{i=1}^n (e^{x_i} + e^\theta)^2}$$

será independiente de  $\theta$  si los polinomios  $\prod_{i=1}^n(\lambda + u_i)$  y  $\prod_{i=1}^n(\lambda + v_i)$  (donde  $\lambda = e^\theta$ ,  $u_i = e^{y_i}$  y  $v_i = e^{x_i}$ ) tienen las mismas raíces. Ello equivale a que  $\{x_1, \dots, x_n\} = \{y_1, \dots, y_n\}$  o, dicho de otra manera, que las dos muestras, una vez ordenadas, coincidan. La muestra ordenada es pues, según el teorema 5.2, el estadístico minimal suficiente.

### Ejercicio 5.6

a) Para expresar la función de probabilidad en la forma

$$f_\theta(x) = c(\theta)h(x) \exp \left\{ \sum_{j=1}^p q_j(\theta)T_j(x) \right\},$$

basta observar que

$$p_\theta(x) = \binom{k}{x} \theta^x (1-\theta)^{k-x} = (1-\theta)^k \binom{k}{x} \exp \left\{ x \log \frac{\theta}{1-\theta} \right\}$$

luego  $c(\theta) = (1-\theta)^k$ ,  $h(x) = \binom{k}{x}$ ,  $p = 1$ ,  $q_1(\theta) = \log \theta / (1-\theta)$  y  $T_1(x) = x$ .

El estadístico minimal suficiente, correspondiente a una muestra de tamaño  $n$  es entonces  $\sum_{i=1}^n X_i$  o, equivalentemente,  $\bar{X}$ .

b) De la misma manera, la binomial negativa tiene función de probabilidad

$$p_\theta(x) = \binom{k+x-1}{x} \theta^k (1-\theta)^x = \theta^k \binom{k+x-1}{x} \exp\{x \log(1-\theta)\},$$

con lo cual  $p = 1$ ,  $q_1(\theta) = \log(1-\theta)$  y  $T_1(x) = x$ .

El estadístico minimal suficiente, de una muestra de tamaño  $n$ , vuelve a ser  $\sum_{i=1}^n X_i$ , o bien  $\bar{X}$ .

c) La densidad de la  $\gamma(\theta_1, \theta_2)$  se expresa

$$f_{\theta_1, \theta_2}(x) = \frac{\theta_2^{\theta_1}}{\Gamma(\theta_1)} x^{\theta_1-1} e^{-\theta_2 x} = \frac{\theta_2^{\theta_1}}{\Gamma(\theta_1)} \exp\{(\theta_1-1) \log x - \theta_2 x\}$$

luego ahora es  $p = 2$ ,  $q_1(\theta) = \theta_1 - 1$ ,  $q_2(\theta) = -\theta_2$ ,  $T_1(x) = \log x$  y  $T_2(x) = x$ .

Ni que decir tiene que entre los vectores  $\theta = (\theta_1, \theta_2) \in \mathbb{R}_+^2$ , pueden hallarse 2 linealmente independientes. Con lo cual el estadístico minimal suficiente es

$$\left( \sum_{i=1}^n \log X_i, \sum_{i=1}^n X_i \right)$$

o cualquier función biunívoca de él.

Si alguno de los dos parámetros fuese conocido se obtendría una familia exponencial de un sólo parámetro (pasando el término conocido a  $h(x)$ ). El estadístico minimal suficiente unidimensional sería entonces  $\sum_{i=1}^n \log X_i$  si  $\theta_2$  es conocido y  $\sum_{i=1}^n X_i$  si  $\theta_1$  es conocido.

d) La densidad de la  $\beta(\theta_1, \theta_2)$  es

$$\frac{1}{\beta(\theta_1, \theta_2)} x^{\theta_1-1} (1-x)^{\theta_2-1} = \frac{1}{\beta(\theta_1, \theta_2)} e^{(\theta_1-1)\log x + (\theta_2-1)\log(1-x)}$$

que es de tipo exponencial con  $p = 2$ ,  $q_1(\theta) = \theta_1 - 1$ ,  $q_2(\theta) = \theta_2 - 1$ ,  $T_1(x) = \log x$  y  $T_2(x) = \log(1-x)$ . Además,  $\{(\theta_1, \theta_2) \mid \theta_1, \theta_2 > 0\}$  tiene dimensión 2; luego el estadístico minimal suficiente es

$$\left( \sum_{i=1}^n \log X_i, \sum_{i=1}^n \log(1-X_i) \right).$$

e) La densidad de Rayleigh

$$f_\theta(x) = \frac{1}{\theta^2} x \exp \left\{ -\frac{1}{2\theta^2} x^2 \right\}$$

es exponencial con  $p = 1$ ,  $q_1(\theta) = -1/2\theta^2$  y  $T_1(x) = x^2$ . De forma que el estadístico  $\sum_{i=1}^n X_i^2$  es minimal suficiente.

### Ejercicio 5.7

a) La densidad de la variable poblacional  $(X, Y)$  es

$$\begin{aligned} f_{\theta_1, \theta_2}(x, y) &= \frac{(\theta_1 \theta_2)^{\theta_1 + \theta_2}}{\Gamma(\theta_1 + \theta_2)^2} (xy)^{\theta_1 + \theta_2 - 1} e^{-\theta_1 x} e^{-\theta_2 y} \\ &= \frac{(\theta_1 \theta_2)^{\theta_1 + \theta_2}}{\Gamma(\theta_1 + \theta_2)^2} e^{-\theta_1 x - \theta_2 y + (\theta_1 + \theta_2 - 1) \log xy} \end{aligned}$$

que aparece puesta en forma exponencial con  $p = 3$ . Sin embargo,

$$\left( \sum_{i=1}^n X_i, \sum_{i=1}^n Y_i, \sum_{i=1}^n \log X_i Y_i \right)$$

no es un estadístico minimal suficiente, debido a la relación lineal que existe entre las tres funciones de  $\theta$  del exponente. Lo adecuado es expresar

$$f_{\theta_1, \theta_2}(x, y) = \frac{(\theta_1 \theta_2)^{\theta_1 + \theta_2}}{\Gamma(\theta_1 + \theta_2)^2} (xy)^{-1} e^{-\theta_1(x - \log xy) - \theta_2(y - \log xy)}$$

que prueba que el estadístico minimal suficiente es

$$\left( \sum_{i=1}^n X_i - \sum_{i=1}^n \log X_i Y_i, \sum_{i=1}^n Y_i - \sum_{i=1}^n \log X_i Y_i \right).$$

b) La función de probabilidad bidimensional considerada es

$$f_{\theta_1, \theta_2}(x, y) = \frac{k!}{x!y!(k-x-y)!} \theta_1^x \theta_2^y (1-\theta_1-\theta_2)^{k-x-y}$$

$$= \frac{k!}{x!y!(k-x-y)!} \exp\{x \log \theta_1 + y \log \theta_2 + (k-x-y) \log(1-\theta_1-\theta_2)\}$$

que tiene forma exponencial con  $p = 3$ . Sin embargo, existe la relación lineal entre los estadísticos  $T_1, T_2, T_3$ :  $T_1 + T_2 + T_3 = k$ . Luego, evidentemente, no se obtiene de ella un estadístico minimal suficiente. La representación de  $f_{\theta_1, \theta_2}(x, y)$  en la forma

$$(1-\theta_1-\theta_2)^k \frac{k!}{x!y!(k-x-y)!} \exp\left\{x \log \frac{\theta_1}{1-\theta_1-\theta_2} + y \log \frac{\theta_2}{1-\theta_1-\theta_2}\right\}$$

prueba que el estadístico minimal suficiente es

$$\left( \sum_{i=1}^n X_i, \sum_{i=1}^n Y_i \right).$$

c) La densidad puede expresarse

$$f_\theta(x) = 2\Gamma(1/4)^{-1} e^{-\theta^4} e^{-x^4} e^{4\theta^3 x - 6\theta^2 x^2 + 4\theta x}$$

en forma exponencial con  $p = 3$ . Sin duda existen tres vectores  $(\theta^3, \theta^2, \theta)$  linealmente independientes; luego el estadístico minimal suficiente es

$$\left( \sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2, \sum_{i=1}^n X_i^3 \right).$$

### Ejercicio 5.8

a) La densidad de la muestra es

$$f_\theta(x_1, \dots, x_n) = \theta^n e^{-\theta T} \quad (x_1, \dots, x_n > 0)$$

donde  $T = \sum_{i=1}^n x_i$ . Por otra parte la distribución a priori tiene densidad

$$\pi(\theta) = \frac{a^p}{\Gamma(p)} \theta^{p-1} e^{-a\theta} \quad (\theta > 0);$$

luego la densidad de  $\theta$  una vez observada la muestra es, para  $\theta > 0$ ,

$$\pi(\theta | x_1, \dots, x_n) = \frac{\theta^{p-1} e^{-a\theta} \theta^n e^{-\theta T}}{\int_0^\infty \theta^{p-1} e^{-a\theta} \theta^n e^{-\theta T} d\theta} = k \theta^{n+p-1} e^{-(T+a)\theta}$$

siendo  $k$  independiente de  $\theta$ . Ello indica que  $\theta$  tiene distribución a posteriori  $\gamma(n+p, T+a)$  y, por tanto, que  $k$  vale  $(T+a)^{n+p}/\Gamma(n+p)$ .

b) Condicionado por el valor de  $\theta$ ,  $T$  tiene distribución  $\gamma(n, \theta)$  de densidad

$$\frac{\theta^n}{\Gamma(n)} t^{n-1} e^{-\theta t} \quad (t > 0).$$

Puesto que  $\theta$  se supone elegido con densidad  $\gamma(p, a)$ , la densidad marginal de  $T$ , para  $t > 0$ , resulta

$$\begin{aligned} g(t) &= \int_0^\infty \frac{\theta^n}{\Gamma(n)} t^{n-1} e^{-\theta t} \frac{a^p}{\Gamma(p)} \theta^{p-1} e^{-a\theta} d\theta \\ &= \frac{a^p t^{n-1}}{\Gamma(n)\Gamma(p)} \int_0^\infty \theta^{n+p-1} e^{-(t+a)\theta} d\theta = \frac{\Gamma(n+p)}{\Gamma(n)\Gamma(p)} \frac{a^p t^{n-1}}{(t+a)^{n+p}}. \end{aligned}$$

c) Con función de pérdida cuadrática el estimador Bayes es la media de la distribución a posteriori de  $\theta$ . A la vista del resultado obtenido en (a), el estimador Bayes es

$$T^* = E[\theta | x_1, \dots, x_n] = \frac{n+p}{T+a}.$$

El riesgo a posteriori —una vez obtenida la muestra— se calcula fácilmente puesto que  $T^*$  es la media de la distribución  $\pi(\theta | x_1, \dots, x_n)$ . Concretamente:

$$E[(T^* - \theta)^2 | x_1, \dots, x_n] = V(\theta | x_1, \dots, x_n) = \frac{n+p}{(T+a)^2}$$

y depende lógicamente de la muestra (a través de  $T$ , en este caso). El riesgo Bayes es entonces

$$r_{T^*}(\pi) = \int_{\mathcal{X}} \frac{n+p}{(T+a)^2} f(x_1, \dots, x_n) dx_1 \dots dx_n$$

que se calcula mejor mediante la distribución marginal de  $T$ :

$$\begin{aligned} r_{T^*}(\pi) &= \int_0^\infty \frac{n+p}{(t+a)^2} g(t) dt = \frac{\Gamma(n+p)}{\Gamma(n)\Gamma(p)} \int_0^\infty \frac{a^p t^{n-1}}{(t+a)^{n+p+2}} dt \\ &= \frac{\Gamma(n+p)}{\Gamma(n)\Gamma(p)} \int_0^1 a^p \left(\frac{y}{a}\right)^{n+p+2} \left(a \frac{1-y}{y}\right)^{n-1} \frac{a dy}{y^2} \\ &= \frac{\Gamma(n+p)}{\Gamma(n)\Gamma(p)} a^{-2} \int_0^1 y^{p+1} (1-y)^{n-1} dy \\ &= \frac{\Gamma(n+p)}{\Gamma(n)\Gamma(p)} a^{-2} \frac{\Gamma(p+2)\Gamma(n)}{\Gamma(n+p+2)} = \frac{(p+1)p}{(n+p+1)(n+p)} a^{-2} \end{aligned}$$

(mediante el cambio  $y = a/(t+a)$ ).

d) Naturalmente la distribución  $\pi(\theta | x_1, \dots, x_n)$  determina la distribución a posteriori de  $1/\theta$ ; pero no es preciso calcularla explícitamente, puesto que sólo interesa obtener su media. El estimador Bayes de  $1/\theta$  es

$$\begin{aligned} T^{**} &= E[1/\theta | x_1, \dots, x_n] = \int_0^\infty \frac{1}{\theta} \frac{(T+a)^{n+p}}{\Gamma(n+p)} \theta^{n+p-1} e^{-(T+a)\theta} d\theta \\ &= \frac{(T+a)^{n+p}}{\Gamma(n+p)} \frac{\Gamma(n+p-1)}{(T+a)^{n+p-1}} = \frac{T+a}{n+p-1}. \end{aligned}$$

### Ejercicio 5.9

La densidad de la muestra se puede expresar

$$\begin{aligned} f_\theta(x_1, \dots, x_n) &= \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2 \right\} \\ &= \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{x})^2 \right\} \exp \left\{ -\frac{n}{2\sigma^2} (\theta - \bar{x})^2 \right\}. \end{aligned}$$

Por otra parte, la distribución a priori tiene densidad

$$\pi(\theta) = \frac{1}{\sqrt{2\pi}r} \exp \left\{ -\frac{1}{2r^2} (\theta - \mu)^2 \right\}.$$

La distribución a posteriori será entonces de la forma

$$\pi(\theta | x_1, \dots, x_n) = k \exp \left\{ -\frac{1}{2r^2} (\theta - \mu)^2 \right\} \exp \left\{ -\frac{n}{2\sigma^2} (\theta - \bar{x})^2 \right\}$$

donde todos los factores independientes de  $\theta$  (incluido el denominador) han sido englobados en la constante  $k$  puesto que se trata de obtener una función de  $\theta$ . Entonces

$$\begin{aligned}\pi(\theta | x_1, \dots, x_n) &= k_1 \exp \left\{ -\frac{1}{2r^2}(\theta^2 - 2\mu\theta) - \frac{n}{2\sigma^2}(\theta^2 - 2\bar{x}\theta) \right\} \\ &= k_1 \exp \left\{ -\frac{1}{2} \left[ \theta^2 \left( \frac{1}{r^2} + \frac{n}{\sigma^2} \right) - 2\theta \left( \frac{\mu}{r^2} + \frac{n\bar{x}}{\sigma^2} \right) \right] \right\} \\ &= k_1 \exp \left\{ -\frac{1}{2} \left( \frac{1}{r^2} + \frac{n}{\sigma^2} \right) [\theta^2 - 2\mu^*\theta] \right\} \\ &= k_2 \exp \left\{ -\frac{1}{2} \left( \frac{1}{r^2} + \frac{n}{\sigma^2} \right) [\theta - \mu^*]^2 \right\}\end{aligned}$$

siendo

$$\mu^* = \left( \frac{\mu}{r^2} + \frac{n\bar{x}}{\sigma^2} \right) \left( \frac{1}{r^2} + \frac{n}{\sigma^2} \right)^{-1}$$

Lo cual muestra que la distribución a posteriori de  $\theta$  es una distribución  $N(\mu^*, (1/r^2 + n/\sigma^2)^{-1/2})$ .

Con pérdida  $(t - \theta)^2$  (y también para  $|t - \theta|$ ), el estimador Bayes resulta ser la media (o la mediana) de la distribución a posteriori:

$$T = \mu^* = \frac{\rho}{\rho + n\tau} \mu + \frac{n\tau}{\rho + n\tau} \bar{x}$$

donde  $\rho = 1/r^2$  y  $\tau = 1/\sigma^2$  son los inversos de las varianzas de la distribución a priori y de la distribución de la población, respectivamente. Así pues, el estimador Bayes es un promedio entre la media,  $\mu$ , de la distribución a priori y la media muestral,  $\bar{x}$ , de las observaciones; el peso de la media a priori es inversamente proporcional a la varianza a priori y pierde relevancia relativa cuando el tamaño muestral,  $n$ , aumenta.

Con función de pérdida cuadrática, el riesgo a posteriori es

$$\int_{\Theta} (\theta - \mu^*)^2 \pi(\theta | x_1, \dots, x_n) d\theta = V(\theta | x_1, \dots, x_n) = \left( \frac{1}{r^2} + \frac{n}{\sigma^2} \right)^{-1}$$

que no depende de  $x_1, \dots, x_n$ . Por consiguiente, el riesgo Bayes es

$$r_{\mu^*}(\pi) = \left( \frac{1}{r^2} + \frac{n}{\sigma^2} \right)^{-1}$$

$\bar{X}$  es un estimador insesgado de  $\theta$ , con lo cual no puede ser estimador Bayes de  $\theta$  frente a ninguna distribución a priori. Sin embargo, su función

de riesgo es

$$R_{\bar{X}}(\theta) = \text{E}_\theta[(\bar{X} - \theta)^2] = V_\theta(\bar{X}) = \frac{\sigma^2}{n}.$$

Ahora bien, para cualquier estimador  $T$  de  $\theta$ , será

$$\max_{\theta} R_T(\theta) \geq \int_{\Theta} R_T(\theta) \pi(\theta) d\theta \geq \left( \frac{1}{r^2} + \frac{n}{\sigma^2} \right)^{-1}$$

y, como la desigualdad es cierta cualquiera que sea el valor de  $r$ , resulta

$$\max_{\theta} R_T(\theta) \geq \lim_{r \rightarrow \infty} \left( \frac{1}{r^2} + \frac{n}{\sigma^2} \right)^{-1} = \frac{\sigma^2}{n}$$

lo cual prueba que  $\bar{X}$  proporciona el menor valor posible de  $\max_{\theta} R_T(\theta)$  y es, por tanto, el estimador minimax de  $\theta$ .

### Ejercicio 5.10

a) La distribución a posteriori de  $\theta$ , determinada en el ejemplo 5.21, es una distribución  $\beta(p + \tau, n + q - \tau)$ , donde  $\tau$  es el número total de bolas blancas obtenidas en las  $n$  extracciones. Con la función de pérdida indicada, el riesgo a posteriori es

$$\begin{aligned} & \int_0^1 L(\theta, t) \pi(\theta | x_1, \dots, x_n) = \\ &= \frac{\Gamma(n + p + q)}{\Gamma(p + \tau)\Gamma(n + q - \tau)} \int_0^1 \frac{(\theta - t)^2}{\theta(1 - \theta)} \theta^{p-1+\tau} (1 - \theta)^{n+q-1-\tau} d\theta \\ &= \frac{\Gamma(n + p + q)}{\Gamma(p + \tau)\Gamma(n + q - \tau)} \int_0^1 (\theta^2 + t^2 - 2t\theta) \theta^{p-2+\tau} (1 - \theta)^{n+q-2-\tau} d\theta \\ &= \frac{p + \tau}{n + q - \tau - 1} + t^2 \frac{(n + p + q - 1)(n + p + q - 2)}{(p + \tau - 1)(n + q - \tau - 1)} - 2t \frac{n + p + q - 1}{n + q - \tau - 1}. \end{aligned}$$

El valor de  $t$  que minimiza este riesgo a posteriori, proporciona el estimador Bayes buscado; concretamente

$$T^* = \frac{p + \tau - 1}{n + p + q - 2}$$

que resulta ligeramente distinto del hallado en el ejemplo 5.21, debido a la modificación de la función de pérdida.

b) Para  $p = q = 1$  (es decir frente a la distribución uniforme en  $(0, 1)$ ) el estimador Bayes anterior se reduce a

$$T^* = \frac{\tau}{n} = \bar{X}$$

que expresa la proporción de bolas blancas observadas en la muestra. Su función de riesgo es

$$R_{T^*}(\theta) = \frac{E[(\bar{X} - \theta)^2]}{\theta(1 - \theta)} = \frac{1}{n},$$

ya que  $\bar{X}$  tiene media  $\theta$  y varianza  $\theta(1 - \theta)/n$ . Puesto que el riesgo es constante, la proporción muestral de bolas blancas es el estimador minimax de  $\theta$ , con la función de pérdida considerada.

c) El estimador  $T_0 = 0$  tiene riesgo

$$R_{T_0}(\theta) = E_\theta[L(\theta, 0)] = E_\theta(1) = 1.$$

Para cualquier otro estimador  $T$ , con valores en  $[0, 1]$ , no será vacío el conjunto de muestras:  $B = \{(x_1, \dots, x_n) \mid T(x_1, \dots, x_n) > 0\}$  y su función de riesgo vendrá dada por

$$R_T(\theta) = \sum_{B^c} p_\theta(x_1, \dots, x_n) + \sum_B \frac{(T(x_1, \dots, x_n) - \theta)^2}{\theta^2} p_\theta(x_1, \dots, x_n)$$

donde  $p_\theta(x_1, \dots, x_n)$  representa la probabilidad de cada muestra.

Si  $c(T) = \min_B T(x_1, \dots, x_n)$ , siempre que sea  $\theta < c(T)/3$  se verifica

$$\frac{(T(x_1, \dots, x_n) - \theta)^2}{\theta^2} > 4$$

para cualquier  $(x_1, \dots, x_n) \in B$ . Por consiguiente,

$$\max_{\theta \in (0, 1)} R_T(\theta) > \sum_{B^c} p_\theta(x_1, \dots, x_n) + 4 \sum_B p_\theta(x_1, \dots, x_n) > 1.$$

Lo cual establece que es minimax el estimador  $T_0$ , que ignora los resultados de las observaciones y estima, siempre,  $\theta = 0$ .

Nótese que el resultado es válido para cualquier distribución poblacional que dependa de un parámetro que puede hacerse arbitrariamente próximo a cero, siempre que la segunda suma no se anule al hacer  $\theta$  muy pequeño. Según esto, aunque el criterio minimax proporciona resultados sensatos en algunas circunstancias —por ejemplo, las del apartado (b)— hay funciones de pérdida con las que da resultados poco razonables, al menos para gran número de distribuciones poblacionales. No es, por tanto, un criterio fiable.

# Capítulo 6

## Estimadores de mínimo riesgo

### 6.1. Introducción

Como se ha visto en el capítulo anterior, los criterios más precisos de comparación de estimadores corresponden a elegir una función de pérdida  $L(\theta, t)$  y calcular el riesgo de cada estimador:

$$R_T(\theta) = \mathbb{E}_\theta[L(\theta, T(X_1, \dots, X_n))].$$

Puesto que el riesgo es función de  $\theta$ , en general dicha valoración sólo ordena parcialmente los estimadores, y no permite ir más allá de clasificarlos en *admisibles* —los que ocupan un lugar minimal en la ordenación parcial— e *inadmisibles* —término al que hay que dar un valor relativo, pese a lo drástico de la denominación. En concreto, los pésimos estimadores constantes, de la forma  $T = \theta_0$ , son siempre admisibles, puesto que ningún estimador puede dar menos riesgo que él para  $\theta = \theta_0$ .

Consecuentemente, la búsqueda de un estimador  $T$  tal que

$$R_T(\theta) \leq R_{T'}(\theta)$$

para cualquier otro estimador  $T'$  y para todos los valores del parámetro  $\theta$  simultáneamente, está condenada al fracaso, si no se restringe la clase de los estimadores considerados mediante algún requisito adicional, que excluya, entre otros, a los estimadores constantes.

En esta línea se ha apuntado, en la sección 5.3, la ausencia de sesgo y las propiedades de invariancia como exigencias razonables —al menos en determinadas circunstancias— que permiten descartar de entrada un buen número de estimadores. En este capítulo se analizará si, con este tipo de

restricciones, el criterio del riesgo selecciona un estimador “uniformemente” mejor que los demás; es decir, mejor para cualquier posible valor del parámetro.

En la próxima sección se aborda el problema, con la imposición de que los estimadores sean insesgados y eligiendo la función de pérdida más usual:  $L(\theta, t) = |\theta - t|^2$ . La sección 3, se ocupa de la cuestión, relacionada con el mismo enfoque, de hallar cotas inferiores para el error cuadrático medio de los estimadores.

En cambio, en la sección 4, se imponen condiciones de invariancia para la estimación de parámetros de posición y de escala, eligiendo funciones de pérdida que respeten la estructura del problema.

## 6.2. Estimadores centrados de mínima varianza

Como ya se ha indicado, aunque  $T$  fuese el mejor estimador centrado de  $\theta$ ,  $g(T)$  no tiene por qué ser ni centrado, ni, aunque lo fuese, el mejor estimador de  $g(\theta)$ . Por tanto, habrá que proponerse como objetivo la estimación de una cierta función real,  $g(\theta)$ , del parámetro  $\theta$  que identifica la distribución teórica de la población (y que puede ser pluridimensional).

Para ello, se considera la clase  $C(g)$  de los *estimadores centrados para  $g(\theta)$* , con momentos de segundo orden finito; es decir aquellos que cumplan

$$\mathbb{E}_\theta[T] = g(\theta) \quad \text{y} \quad \mathbb{E}_\theta[T^2] < \infty \quad \text{para todo } \theta \in \Theta.$$

Para  $T \in C(g)$ , eligiendo como pérdida  $L(\theta, t) = (\theta - t)^2$ , el riesgo o error cuadrático medio en la estimación de  $g(\theta)$ , coincide con la varianza de  $T$ :

$$\mathbb{E}_\theta[(g(\theta) - T)^2] = V_\theta(T)$$

Una vez decidido limitarse a estimadores insesgados, evidentemente interesan aquellos que tengan la menor dispersión posible alrededor de la función a estimar. De manera que el propósito es buscar *estimadores centrados uniformemente de mínima varianza* (ECUMV) en la estimación de  $g(\theta)$ . Esto es, aquellos estimadores  $T \in C(g)$  que verifiquen

$$V_\theta(T) \leq V_\theta(T') \quad \text{para cualquier } \theta \in \Theta \text{ y cualquier } T' \in C(g).$$

Lo primero que cabe observar es que, cuando exista, el estimador centrado uniformemente de mínima varianza es esencialmente único. En efecto

si  $T$  es ECUMV para  $g(\theta)$  y  $T'$  es otro estimador cualquiera de  $C(g)$ , como  $T + \lambda(T' - T) \in C(g)$ , será

$$V_\theta(T) \leq V_\theta(T + \lambda(T' - T)) = V_\theta(T) + \lambda^2 V_\theta(T' - T) + 2\lambda E_\theta[T(T' - T)]$$

cualquiera que sea  $\lambda \in \mathbb{R}$ . Por tanto, tiene que ser

$$E_\theta[T(T' - T)] = 0 \quad \text{o bien} \quad E_\theta[TT'] = E_\theta[T^2].$$

Luego, si  $T'$  fuese también ECUMV para  $g(\theta)$ , resultaría

$$E_\theta[(T - T')^2] = E_\theta[T^2] + E_\theta[T'^2] - 2E_\theta[TT'] = 2E_\theta[T^2] - 2E_\theta[TT'] = 0,$$

con lo cual  $T = T'$ ,  $\mathcal{P}_\theta$ -casi seguro, cualquiera que sea  $\theta$ .

Por supuesto, ello no significa que siempre exista ECUMV, e, incluso, algunos ejemplos de la subsección 5.3.1, prueban que puede ser  $C(g) = \emptyset$ . La cuestión se centra pues en encontrar un procedimiento que permita hallar el ECUMV, en caso de que exista. En este sentido, el paso fundamental consiste en concretar la idea, ya mencionada, de que la búsqueda puede limitarse a aquellos estimadores que actúan únicamente en función de un estadístico suficiente —o más exactamente del estadístico suficiente minimal— sin tener en cuenta la información superflua que contenga la muestra. Para dar forma explícita a ese argumento, habrá que comprobar que cualquier estimador  $T$ , centrado para  $g(\theta)$ , puede mejorarse (en el sentido de disminuir su varianza), si se modifica para que sólo dependa del valor de un estadístico suficiente  $S$ . Tal es el contenido del siguiente resultado:

**Teorema de Rao-Blackwell:** Sea  $S$  un estadístico suficiente para la familia  $\{F_\theta | \theta \in \Theta\}$  de distribuciones de la población. Para cualquier estimador  $T$ , centrado para  $g(\theta)$ , el estimador  $T' = E_\theta[T | S]$  verifica

- (a) Es independiente de  $\theta$  (y por tanto puede utilizarse como estimador de  $g(\theta)$ ).
- (b) Es un estimador centrado de  $g(\theta)$ :  $E_\theta[T'] = g(\theta)$ , para todo  $\theta \in \Theta$ .
- (c)  $V_\theta(T') \leq V_\theta(T)$ , para todo  $\theta \in \Theta$ .

Además  $V_\theta(T') = V_\theta(T)$  si y sólo si  $T' = T$ ,  $\mathcal{P}_\theta$ -casi seguro, para cada  $\theta \in \Theta$ .

Demostración: La suficiencia de  $S$  indica que la distribución de la muestra  $(X_1, X_2, \dots, X_n)$  condicionada por  $S$  es independiente de  $\theta$ . Por consiguiente, lo mismo ocurrirá con la distribución de  $T$  condicionada por  $S$ :

$H(t | S = s)$ , y con su media:

$$E[T | S = s] = \int_{\mathbb{R}} t dH(t | S = s)$$

(en cuyas notaciones se ha omitido por tanto el subíndice  $\theta$ ).

Por otra parte si  $H_\theta(t, s)$  representa la distribución conjunta de  $(T, S)$  (en  $\mathbb{R} \times \mathbb{R}^k$ ) y  $G_\theta(s)$  la distribución marginal de  $S$  (en  $\mathbb{R}^k$ ), es

$$\begin{aligned} E_\theta [E[T | S]] &= \int_{\mathbb{R}^k} \int_{\mathbb{R}} t H(dt | S = s) G_\theta(ds) \\ &= \int_{\mathbb{R} \times \mathbb{R}^k} t H_\theta(dt, ds) = E_\theta[T] = g(\theta). \end{aligned}$$

Por último, entre el momento de segundo orden de  $H(t | S = s)$  y el cuadrado de su media, se verifica:  $E[T^2 | S = s] \geq E[T | S = s]^2$ ; es decir  $T'^2 \leq E[T^2 | S]$ . Por tanto:

$$E_\theta[T'^2] \leq E_\theta[T^2] \quad \text{o bien} \quad V_\theta(T') \leq V_\theta(T) \text{ para todo } \theta \in \Theta.$$

Naturalmente  $E[T | S]$  depende de la muestra a través del estadístico  $S$ , de manera que se sigue del teorema anterior que, si existe, el ECUMV para cualquier función del parámetro, es función de cualquier estadístico suficiente y, en particular, del estadístico minimal suficiente.

Es interesante señalar que, una ligera modificación del razonamiento anterior, permite generalizar el resultado al caso de una pérdida convexa arbitraria:  $L(\theta, t)$ . Es decir que, si la pérdida es convexa,  $E[T | S]$  tiene siempre riesgo inferior o igual a  $T$ ; lo cual da una justificación precisa del hecho intuitivo de que, desde el punto de vista del riesgo, la elección del estimador puede limitarse a aquellos que se basan en la información suficiente.

Pero, en este momento, mejor que generalizar el teorema anterior, es ver cómo actúa en casos concretos.

### Ejemplo 6.1

A partir de una muestra aleatoria simple  $(X_1, X_2, \dots, X_n)$ , de tamaño  $n > 1$ , de una población con distribución de Poisson de parámetro  $\theta$ , se desea estimar la probabilidad que la distribución teórica asigna al valor cero:  $g(\theta) = e^{-\theta}$ .

De entrada puede escogerse un estimador muy malo de  $e^{-\theta}$ , por ejemplo:

$$T = \begin{cases} 1 & \text{si } X_1 = 0, \\ 0 & \text{si } X_1 \neq 0, \end{cases}$$

cuya única ventaja es que es centrado para  $e^{-\theta}$  (puesto que  $\mathcal{P}_\theta\{X_1 = 0\} = e^{-\theta}$ ), aunque sólo proporcione las estimaciones 0 y 1 y desprecie toda la información suministrada por  $X_2, \dots, X_n$ .

Sin embargo,  $S = \sum_{i=1}^n X_i$  es un estadístico suficiente para la familia de distribuciones considerada, de manera que podemos mejorar el estimador  $T$  calculando  $E[T | S]$ . Para ello, ha de obtenerse

$$\begin{aligned}\mathcal{P}_\theta \left\{ X_1 = 0 \mid \sum_{i=1}^n X_i = s \right\} &= \frac{\mathcal{P}_\theta\{X_1 = 0, \sum_{i=2}^n X_i = s\}}{\mathcal{P}_\theta\{\sum_{i=1}^n X_i = s\}} \\ &= \frac{e^{-\theta} e^{-(n-1)\theta} (n-1)^s \theta^s / s!}{e^{-n\theta} n^s \theta^s / s!} \\ &= \left( \frac{n-1}{n} \right)^s\end{aligned}$$

(puesto que  $\sum_{i=1}^n X_i$  y  $\sum_{i=2}^n X_i$  tienen distribución de Poisson de parámetros  $n\theta$  y  $(n-1)\theta$  respectivamente).

Como era previsible, el resultado es independiente de  $\theta$  y proporciona

$$E[T | S] = \left( \frac{n-1}{n} \right)^S = \left( \frac{n-1}{n} \right)^{\sum_{i=1}^n X_i}$$

Aunque no sea necesario, puede verificarse directamente que  $E[T | S]$  es insesgado para  $e^{-\theta}$ :

$$\sum_{k=0}^{\infty} \left( \frac{n-1}{n} \right)^k \frac{e^{-n\theta} (n\theta)^k}{k!} = e^{-n\theta} \sum_{k=0}^{\infty} \frac{(n-1)^k \theta^k}{k!} = e^{-\theta}.$$

Además, mientras  $V_\theta(T) = e^{-\theta} - e^{-2\theta} = e^{-\theta}(1 - e^{-\theta})$ , se tiene

$$\begin{aligned}V_\theta(E[T | S]) &= \sum_{k=0}^{\infty} \left( \frac{n-1}{n} \right)^{2k} \frac{e^{-n\theta} (n\theta)^k}{k!} - e^{-2\theta} \\ &= e^{-n\theta} \sum_{k=0}^{\infty} \frac{1}{k!} \left( \frac{(n-1)^2 \theta}{n} \right)^k - e^{-2\theta} \\ &= e^{-n\theta} e^{(n-1)^2 \theta / n} - e^{-2\theta} \\ &= e^{-2\theta} (e^{\theta/n} - 1);\end{aligned}$$

de manera que

$$\frac{V_\theta(E[T | S])}{V_\theta(T)} = \frac{e^{\theta/n} - 1}{e^\theta - 1},$$

poniendo de relieve el ahorro en varianza conseguido.

De hecho,  $((n-1)/n)^S$  constituye el ECUMV para  $e^{-\theta}$ . Porque los estimadores  $f(S)$ , función de  $S$ , centrados para  $e^{-\theta}$  deben cumplir

$$\sum_{k=0}^{\infty} f(k) \frac{e^{-n\theta} (n\theta)^k}{k!} = e^{-\theta};$$

es decir

$$\sum_{k=0}^{\infty} f(k) \frac{n^k \theta^k}{k!} = e^{(n-1)\theta} = \sum_{k=0}^{\infty} \frac{(n-1)^k \theta^k}{k!}$$

y para que ambas series de potencias de  $\theta$  coincidan, cualquiera que sea  $\theta > 0$ , debe ser  $f(k) = ((n-1)/n)^k$ . Es decir que  $((n-1)/n)^S$  es la única función de  $S$  que proporciona un estimador insesgado de  $e^{-\theta}$  y el ECUMV para  $e^{-\theta}$  (que ha de ser función de  $S$ ) tiene que coincidir con ella.

## Ejemplo 6.2

A partir de una muestra aleatoria simple  $(X_1, X_2, \dots, X_n)$  de una población con distribución uniforme en  $(0, \theta)$ , para estimar el valor de  $\theta$ , se cuenta en principio con el estimador centrado:  $T = 2X_1$ , que sólo tiene en cuenta la primera de las observaciones.

Como  $X_{(n)} = \max_{i \leq i \leq n} X_i$  es un estadístico suficiente,  $E[T | X_{(n)}]$  será un estimador centrado de menor varianza que  $T$ . Para obtener su expresión explícita, obsérvese que

$$\begin{aligned} \mathcal{P}_{\theta} \{ X_1 \leq x, X_{(n)} \leq y \} &= \mathcal{P}_{\theta} \{ X_{(n)} \leq y \} - \mathcal{P}_{\theta} \{ X_{(n)} \leq y, X_1 > x \} = \\ &= \begin{cases} \left(\frac{y}{\theta}\right)^n - \left(\frac{y}{\theta}\right)^{n-1} \frac{y-x}{\theta} = \frac{y^{n-1}x}{\theta^n} & \text{si } 0 \leq x < y < \theta, \\ \left(\frac{y}{\theta}\right)^n & \text{si } x \geq y \text{ e } y < \theta; \end{cases} \end{aligned}$$

de donde, si  $y < \theta$ , se obtiene

$$\begin{aligned} \mathcal{P}_{\theta} \{ X_1 \leq x \mid X_{(n)} \in (y, y+h) \} &= \\ &= \begin{cases} \frac{(y+h)^{n-1}x/\theta^n - y^{n-1}x/\theta^n}{(y+h)^n/\theta^n - y^n/\theta^n} = x \frac{(y+h)^{n-1} - y^{n-1}}{(y+h)^n - y^n} & \text{si } 0 \leq x < y, \\ 1 & \text{si } x \geq y. \end{cases} \end{aligned}$$

Por tanto,

$$\mathcal{P}_\theta \{ X_1 \leq x \mid X_{(n)} = y \} = \begin{cases} \frac{n-1}{n} \frac{x}{y} & \text{si } 0 \leq x < y, \\ 1 & \text{si } x \geq y. \end{cases}$$

Es decir que, condicionado por  $X_{(n)} = y$ ,  $X_1$  tiene probabilidad  $1/n$  de coincidir con  $y$  y, con probabilidad  $1 - \frac{1}{n}$ , está uniformemente distribuido en  $(0, y)$ . Por tanto

$$\mathbb{E}[T \mid X_{(n)} = y] = \frac{n-1}{ny} \int_0^y 2x dx + \frac{2y}{n} = \frac{n+1}{n} y.$$

Resulta así,

$$\mathbb{E}[T \mid X_{(n)}] = \frac{n+1}{n} X_{(n)}$$

que coincide con el estimador centrado que ya se consideró en el ejemplo 5.4.

Mientras que  $V_\theta(T) = \theta^2/3$ , según las conclusiones del ejemplo 5.1, es

$$V_\theta(\mathbb{E}[T \mid X_{(n)}]) = \frac{\theta^2}{n(n+2)}$$

que da un resultado notablemente menor, si  $n$  es grande.

Además, para que un estimador  $f(X_{(n)})$ , función de  $X_{(n)}$ , sea insesgado para  $\theta$ , ha de ser

$$\int_0^\theta f(y) n \frac{y^{n-1}}{\theta^n} dy = \theta \quad \text{o bien} \quad \int_0^\theta f(y) y^{n-1} dy = \frac{\theta^{n+1}}{n};$$

de donde, derivando respecto a  $\theta$ ,

$$f(\theta) \theta^{n-1} = \frac{n+1}{n} \theta^n \quad \text{o bien} \quad f(\theta) = \frac{n+1}{n} \theta \quad \text{para todo } \theta > 0.$$

Luego,  $(n+1)X_{(n)}/n$  es el único estimador centrado de  $\theta$ , función de  $X_{(n)}$ , y tiene que ser el ECUMV para  $\theta$ .

Los dos ejemplos anteriores muestran que el método empleado en el teorema de Rao-Blackwell —condicionar un estimador centrado cualquiera por un estadístico suficiente  $S$ — aparte de mejorar el estimador original, conduce con frecuencia a la obtención del ECUMV. Para garantizarlo, en ambos casos se ha usado el mismo argumento: *si existe una única función*

$f(S)$ , que proporcione un estimador insesgado de  $g(\theta)$ , necesariamente  $f(S)$  es el ECUMV para  $g(\theta)$ .

El que exista un único estimador centrado, función de  $S$ , está directamente relacionado con la noción de *completitud* siguiente:

*Dada una muestra aleatoria  $(X_1, X_2, \dots, X_n)$  de una población con distribución teórica perteneciente a una familia paramétrica  $\{F_\theta | \theta \in \Theta\}$ , un estadístico  $S(X_1, \dots, X_n)$  se dice completo si*

$$E_\theta[f(S)] = 0 \quad \text{para todo } \theta \in \Theta,$$

implica  $f(s) = 0$ ,  $\mathcal{P}_\theta$ -casi seguro<sup>1</sup>, para todo  $\theta \in \Theta$ .

Dicho de otra manera,  $S$  es completo si el único estimador insesgado de 0, función de  $S$ , es el estadístico que vale cero, con probabilidad uno. La manera de proceder, para verificar que un estadístico es completo, es la empleada al final de los ejemplos anteriores.

### Ejemplo 6.3

Sea  $(X_1, X_2, \dots, X_n)$  una muestra aleatoria simple de la población

$$X = \begin{cases} 1 & \text{con probabilidad } \theta, \\ 0 & \text{con probabilidad } 1 - \theta. \end{cases}$$

El estadístico  $S = \sum_{i=1}^n X_i$  es completo. En efecto, su distribución es binomial  $B(n, \theta)$ , luego

$$E_\theta[f(S)] = \sum_{k=0}^n f(k) \binom{n}{k} \theta^k (1-\theta)^{n-k} = (1-\theta)^n \sum_{k=0}^n f(k) \binom{n}{k} \left(\frac{\theta}{1-\theta}\right)^k$$

y si, para todo  $\theta \in (0, 1)$ , ha de ser

$$\sum_{k=0}^n f(k) \binom{n}{k} \left(\frac{\theta}{1-\theta}\right)^k = 0$$

el primer miembro es un polinomio de grado  $n$  en la variable  $x = \theta/(1-\theta)$ , que sólo puede ser idénticamente nulo si  $f(k) = 0$  para  $k = 0, 1, \dots, n$ .

---

<sup>1</sup> Siempre que  $f$  sea medible y no dependa de  $\theta$ .

Naturalmente, también es completo el estadístico  $X_1$ ; pues

$$E_\theta[f(X_1)] = f(1)\theta + f(0)(1 - \theta) = 0 \quad \text{para todo } \theta \in (0, 1)$$

implica  $f(0) = f(1) = 0$ .

En cambio  $(X_1, X_2)$  no es un estadístico completo, pues  $E_\theta[X_1 - X_2] = 0$ , sin que sea  $X_1 - X_2 = 0$  con probabilidad uno.

### Ejemplo 6.4

Dada una muestra aleatoria simple de una distribución exponencial de parámetro  $\theta$ , el estadístico  $S = \sum_{i=1}^n X_i$  es completo. En efecto, su distribución es gamma  $\gamma(n, \theta)$ , luego

$$E_\theta[f(S)] = \int_0^\infty f(x) \frac{\theta^n x^{n-1} e^{-\theta x}}{(n-1)!} dx.$$

Por tanto,  $E_\theta[f(S)] = 0$  para cada  $\theta > 0$ , significa que la transformada de Laplace:  $\int_0^\infty f(x)x^{n-1}e^{-\theta x} dx$ , de la función  $f(x)x^{n-1}$  se anula. Según el teorema de unicidad de la transformada de Laplace<sup>2</sup>, ello sólo es posible si  $f(x)x^{n-1} = 0$  o bien  $f(x) = 0$  para todo  $x > 0$ .

En cambio, como antes,  $X_1 - X_2$  no es completo. Ni tampoco lo es  $X_1^2 X_2 - X_3^2 X_4$ .

### Ejemplo 6.5

Si  $(X_1, X_2, \dots, X_n)$  es una muestra aleatoria simple de una población con función de densidad

$$f_\theta(x) = \frac{k x^{k-1}}{\theta^k} \quad \text{para } x \in (0, \theta),$$

el estadístico  $X_{(n)} = \max_{1 \leq i \leq n} X_i$  es completo. Efectivamente

$$\mathcal{P}_\theta \{ X_{(n)} \leq y \} = \left( \frac{y^k}{\theta^k} \right)^n \quad \text{para } y < \theta;$$

luego

$$E_\theta[f(X_{(n)})] = \int_0^\theta f(y) \frac{n k y^{nk-1}}{\theta^{nk}} dy$$

es idénticamente nula si  $\int_0^\theta f(y)y^{nk-1}dy = 0$  para todo  $\theta > 0$ . Derivando respecto a  $\theta$ , resulta  $f(\theta)\theta^{nk-1} = 0$ ; es decir  $f(\theta) = 0$  para todo  $\theta > 0$ .

---

<sup>2</sup>Véase M.R. Spielgel *Transformada de Laplace*, Schaum.

### Ejemplo 6.6

Si  $(X_1, X_2, \dots, X_n)$  es una muestra aleatoria simple de una población con distribución de Poisson de parámetro  $\theta$ , el estadístico  $S = \sum_1^n X_i$  es completo; pues si, para cualquier  $\theta > 0$ , es

$$E_\theta[f(S)] = \sum_{k=1}^{\infty} f(k) e^{-n\theta} \frac{(\theta n)^k}{k!} = 0 \quad \text{será} \quad \sum_{k=1}^{\infty} \frac{f(k)}{k!} n^k \theta^k = 0$$

y, para que esta serie de potencias sea idénticamente nula, tiene que ser  $f(k) = 0$  para todo  $k$ . El razonamiento coincide básicamente con el realizado al final del ejemplo 6.1.

La utilización del concepto de completitud en la determinación de estimadores centrados, uniformemente de mínima varianza, es casi inmediata; pero se conoce como

**Teorema de Lehmann-Scheffé:** *Sea  $S$  un estadístico suficiente y completo para la familia  $\{F_\theta | \theta \in \Theta\}$ . Si existe un estimador centrado de  $g(\theta)$ :  $T \in C(g)$ , entonces  $E[T | S]$  es (salvo cambios en conjuntos de probabilidad cero) el único estimador centrado de  $g(\theta)$ , función de  $S$ . Es, por tanto, el ECUMV para  $g(\theta)$ .*

Demostración: Como  $E[T | S]$  es una función,  $f(S)$ , del estadístico  $S$ , si existiese otro estimador  $f'(S)$ , centrado para  $g(\theta)$ , sería

$$E_\theta[f(S) - f'(S)] = 0 \quad \text{para cualquier } \theta \in \Theta.$$

En virtud de la completitud de  $S$ , tendría que ser  $f'(S) = f(S)$ ,  $\mathcal{P}_\theta$ -casi seguro para todo  $\theta \in \Theta$ .

Puesto que el ECUMV de  $g(\theta)$  ha de ser función de  $S$ , tiene que coincidir con  $E[T | S]$ ,  $\mathcal{P}_\theta$ -c.s. para cada  $\theta \in \Theta$ .

Los resultados anteriores proporcionan diversas maneras de proceder en los casos concretos.

### Ejemplo 6.7

Con una muestra  $(X_1, X_2, \dots, X_n)$  de una distribución exponencial de parámetro  $\theta$ , se sabe que  $S = \sum_{i=1}^n X_i$  es un estadístico suficiente y completo. (Ejemplos 5.12 y 6.4).

Puesto que  $E[\sum_{i=1}^n X_i] = n/\theta$  es inmediato que  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  es el ECUMV para la media poblacional  $1/\theta$ , puesto que no puede haber otro estimador insesgado de  $1/\theta$ , función de  $S$ .

Si lo que se desea estimar es  $\theta$ , cabe pensar en un estimador de la forma  $c/S$ , que será insesgado para  $\theta$  si su media vale  $\theta$ . Como

$$\begin{aligned} E_\theta \left[ \frac{c}{S} \right] &= c \int_0^\infty \frac{1}{y} \frac{\theta^n y^{n-1} e^{-\theta y}}{(n-1)!} dy \\ &= \frac{c\theta}{(n-1)} \int_0^\infty \frac{\theta^{n-1} y^{n-2} e^{-\theta y}}{(n-2)!} dy = \frac{c\theta}{n-1}, \end{aligned}$$

basta tomar  $c = n - 1$  y resulta que

$$\frac{n-1}{\sum_{i=1}^n X_i}$$

es el único estimador centrado de  $\theta$ , función de  $S$  y, por consiguiente, su ECUMV.

Para la estimación de  $e^{-k\theta}$ , probabilidad de que la variable poblacional supere el valor de  $k > 0$ , es más difícil prever el resultado; de manera que se puede elegir un estimador insesgado sencillo, por ejemplo:

$$T = \begin{cases} 1 & \text{si } X_1 > k \\ 0 & \text{si } X_1 \leq k, \end{cases}$$

y calcular  $E[T | S]$  que, independientemente del estimador  $T$  elegido, ha de dar la única función de  $S$  que estime  $e^{-k\theta}$  sin sesgo y constituya su ECUMV.

Como  $\sum_{i=2}^n X_i$  tiene distribución gamma  $\gamma(n-1, \theta)$ , independiente de  $X_1$ , será

$$\mathcal{P}_\theta \{ S \leq y | X_1 = x \} = \mathcal{P}_\theta \left\{ \sum_{i=2}^n X_i \leq y - x \right\};$$

de forma que  $S$ , condicionado por  $X_1 = x$ , tiene densidad

$$\frac{\theta^{n-1} (y-x)^{n-2} e^{-\theta(y-x)}}{(n-2)!} \quad \text{para } y > x.$$

Luego  $X_1$ , condicionado por  $S = y$ , tiene densidad

$$\frac{\theta e^{-\theta x} \cdot \theta^{n-1} (y-x)^{n-2} e^{-\theta(y-x)} / (n-2)!}{\theta^n y^{n-1} e^{-\theta y} / (n-1)!} = (n-1) \frac{(y-x)^{n-2}}{y^{n-1}}$$

para  $x \in (0, y)$ . Resulta entonces

$$\begin{aligned} E[T | S = y] &= \mathcal{P}\{X_1 > k | S = y\} \\ &= \begin{cases} 0 & \text{si } y < k, \\ \int_k^y (n-1) \frac{(y-x)^{n-2}}{y^{n-1}} dx = \left( \frac{y-k}{y} \right)^{n-1} & \text{si } y \geq k. \end{cases} \end{aligned}$$

Es decir que  $E[T | S] = \left(1 - \frac{k}{S}\right)^{n-1} I_{\{S \geq k\}}$  es el ECUMV para  $e^{-k\theta}$ .

### Ejemplo 6.8

En el caso de una población  $N(\mu, \sigma^2)$ , con ambos parámetros desconocidos, el estadístico  $(\bar{X}, s^2)$  es suficiente (véase ejemplo 5.14). Además, según el teorema de Fisher, su densidad conjunta es

$$K e^{-n(x-\mu)^2/2\sigma^2} \left(\frac{ny}{\sigma^2}\right)^{\frac{n-3}{2}} e^{-ny/2\sigma^2} \quad \text{para } x \in \mathbb{R}, \quad y > 0,$$

siendo  $K$  una constante que no juega ningún papel en la comprobación de su completitud. De hecho, para comprobar que  $(\bar{X}, s^2)$  es completo, hay que partir de la hipótesis:

$$\int_{-\infty}^{\infty} \int_0^{\infty} f(x, y) e^{-n(x-\mu)^2/2\sigma^2} \left(\frac{ny}{\sigma^2}\right)^{\frac{n-3}{2}} e^{-ny/2\sigma^2} dy dx = 0$$

de la cual se pueden eliminar los factores constantes (en  $x$  e  $y$ ).

Llamando  $\theta_1 = n/(2\sigma^2)$ ,  $\theta_2 = n\mu/\sigma^2$  y  $h(x, y) = f(x, y)y^{\frac{n-3}{2}}$ , ello se expresa

$$\int_{-\infty}^{\infty} \int_0^{\infty} h(x, y) e^{-\theta_1 x^2} e^{\theta_2 x} e^{-\theta_1 y} dy dx = 0 \quad \text{para cada } \theta_1 > 0, \theta_2 \in \mathbb{R};$$

es decir, haciendo el cambio  $u = x, v = x^2 + y$ ,

$$\int_0^{\infty} \int_{-\sqrt{v}}^{\sqrt{v}} h(u, v - u^2) e^{\theta_2 u} du e^{-\theta_1 v} dv = 0.$$

Según el teorema de unicidad de la transformada de Laplace, ha de ser entonces

$$\int_{\sqrt{v}}^{-\sqrt{v}} h(u, v - u^2) e^{\theta_2 u} du = 0 \quad \text{para cada } \theta_2 \in \mathbb{R} \quad \text{y } v > 0.$$

Derivando  $r$  veces con respecto a  $\theta_2$  y haciendo  $\theta_2 = 0$  se obtiene

$$\int_{\sqrt{v}}^{-\sqrt{v}} u^r h(u, v - u^2) du = 0 \quad \text{para cada } r \in \mathbb{N} \quad \text{y } v > 0;$$

de donde

$$\int_{-\sqrt{v}}^{\sqrt{v}} e^{itu} h(u, v - u^2) du = \sum_{r=0}^{\infty} \frac{(it)^r}{r!} \int_{-\sqrt{v}}^{\sqrt{v}} u^r h(u, v - u^2) du = 0.$$

Por tanto,  $h(u, v - u^2) = 0$  para cada  $u \in \mathbb{R}, v > 0$ ; o bien  $h(x, y) = 0$  para cada  $x \in \mathbb{R}, y > 0$ .

Una vez comprobado que  $(\bar{X}, S^2)$  es un estadístico suficiente y completo, es inmediato que  $\bar{X}$  es ECUMV para  $\mu$  y  $(n - 1)s^2/n = S^2$  es ECUMV para  $\sigma^2$ .

Recuérdese que  $V(S^2) = 2\sigma^4/(n - 1)$ . Sin embargo,

$$\begin{aligned} E[(kS^2 - \sigma^2)^2] &= E\left[\left(k(S^2 - \sigma^2) + (k - 1)\sigma^2\right)^2\right] \\ &= k^2V(S^2) + (k - 1)^2\sigma^4 = \left[\frac{2k^2}{n - 1} + (k - 1)^2\right]\sigma^4 \end{aligned}$$

se hace mínimo para  $k = (n - 1)/(n + 1)$ ; de forma que  $(n - 1)S^2/(n + 1)$  es un estimador con menos error cuadrático medio que  $S^2$ , aunque sesgado.

Así pues, el ECUMV para  $\sigma^2$  no es admisible; lo cual muestra que en la determinación de estimadores centrados de mínima varianza, la condición de centrados prima sobre la condición de mínimo error cuadrático medio.

Por otra parte, como  $ns^2/\sigma^2$  tiene distribución  $\chi_{n-1}^2$ ,

$$\begin{aligned} E[s] &= \frac{\sigma}{\sqrt{n}}E\left[\frac{\sqrt{ns}}{\sigma}\right] = \frac{\sigma}{\sqrt{n}} \int_0^\infty y^{\frac{1}{2}}y^{\frac{n-3}{2}}e^{-\frac{y}{2}}dy \frac{1}{2^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})} \\ &= \frac{\sigma}{\sqrt{n}} \frac{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}{2^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})} = \sigma \frac{\sqrt{2}\Gamma(\frac{n}{2})}{\sqrt{n}\Gamma(\frac{n-1}{2})} \end{aligned}$$

de forma que el ECUMV para  $\sigma$  es  $\frac{\sqrt{n}\Gamma(\frac{n-1}{2})}{\sqrt{2}\Gamma(\frac{n}{2})}s$ .

En resumidas cuentas, lo que asegura el teorema de Lehmann-Sheffé es que cada función de un estadístico suficiente y completo, es ECUMV para su esperanza. En aquellas situaciones en las que haya un estadístico suficiente y completo, la cuestión de la existencia de estimadores centrados uniformemente de mínima varianza queda perfectamente perfilada: las funciones del parámetro, para las que exista algún estimador centrado, tendrán ECUMV; la demás, obviamente, no lo tendrán.

En ausencia de estadísticos suficientes y completos, la situación es más compleja: puede haber funciones del parámetro que posean ECUMV; y funciones del parámetro para las que no haya ECUMV, pese a existir estimadores centrados para ella. He aquí uno de estos últimos casos:

### Ejemplo 6.9

Para una muestra aleatoria simple  $(X_1, \dots, X_n)$  de una distribución uniforme en  $(\theta, \theta + 1)$ , el estadístico  $(X_{(1)}, X_{(n)})$  es suficiente, puesto que la densidad de la muestra es

$$f_\theta(x_1, x_2, \dots, x_n) = I_{\{x_{(n)} - 1 < \theta < x_{(1)}\}}.$$

Sin embargo no es completo, pues

$$\mathbb{E}_\theta[X_{(n)}] = \int_\theta^{\theta+1} xn(x - \theta)^{n-1} dx = \theta + 1 - \frac{1}{n+1}$$

y

$$\mathbb{E}_\theta[X_{(1)}] = \int_\theta^{\theta+1} xn(\theta + 1 - x)^{n-1} dx = \theta + \frac{1}{n+1}.$$

De manera que, tanto  $X_{(1)} - 1/(n+1)$  como  $X_{(n)} - 1 + 1/(n+1)$  son estimadores insesgados de  $\theta$ , basados en el estadístico suficiente. (O, dicho de otro modo,  $X_{(1)} - X_{(n)} + (n-1)/(n+1)$  es una función, no nula, del estadístico suficiente, cuya esperanza es cero).

Aún cabría pensar que entre las funciones de  $X_{(1)}$  y  $X_{(n)}$ , que proporcionen un estimador insesgado de  $\theta$  (por ejemplo  $(X_{(1)} + X_{(n)} - 1)/2$ ), podría existir una con menor varianza que las demás, que constituyese el ECUMV para  $\theta$ . Tal posibilidad queda descartada mediante el siguiente argumento:

Supongamos que conociésemos la parte decimal  $\alpha$  de  $\theta$  (es decir  $\theta = k + \alpha$  con  $\alpha \in (0, 1)$  y  $k$  natural). En tal caso, la primera observación,  $X_1$ , bastaría para dar una estimación sin error de  $\theta$ , puesto que  $\theta$  tendría que coincidir con

$$T_\alpha(X_1) = \text{mayor número, con parte decimal } \alpha, \text{ menor que } X_1.$$

Además, cualquiera que fuese el valor de  $\theta$ , en el intervalo  $(\theta, \theta + 1)$  hay un único número de la forma  $k + \alpha$  y sería

$$\begin{aligned} \mathbb{E}_\theta[T_\alpha] &= \int_\theta^{k+\alpha} (k - 1 + \alpha) dx_1 + \int_{k+\alpha}^{\theta+1} (k + \alpha) dx_1 \\ &= (k - 1 + \alpha)(k + \alpha - \theta) + (k + \alpha)(\theta + 1 - k - \alpha) = \theta. \end{aligned}$$

Luego  $T_\alpha$  es un estimador insesgado de  $\theta$ , que tiene varianza cero para todos los valores de  $\theta$  de la forma  $k + \alpha$ , con  $k$  natural. Obviamente si existiese ECUMV para  $\theta$ , tendría que tener varianza idénticamente nula

(para mejorar simultáneamente a todos los  $T_\alpha$ ). Esto es imposible, pues supondría que  $\theta$  puede determinarse, sin error, a partir de cualquier muestra. Y, por el contrario, está claro que una muestra, con  $x_{(n)} - x_{(1)} < 1$ , es compatible con diversos valores de  $\theta$ .

La ausencia de cualquier referencia a los estadísticos minimales suficientes, en el teorema de Lehman-Scheffé, podría causar cierta inquietud: Si  $S$  es un estadístico suficiente y completo, no minimal, formando la esperanza de las funciones de  $S$ , condicionada por el estadístico suficiente minimal  $S'$ , disminuiríamos la varianza de los ECUMV; lo cual es una contradicción. De hecho, no hay tal problema, ya que *un estadístico suficiente que no sea minimal, no puede ser completo*.

En efecto,  $E[S | S']$  es una función de  $S'$ :  $h(S')$ ; pero  $S'$  es función de  $S$ :  $S' = \varphi(S)$ , por ser  $S'$  minimal. Luego,  $E[S | S'] = h(\varphi(S))$ , es función de  $S$  y proporciona un estimador insesgado de  $E_\theta[S]$ . Como  $S$  es completo, ha de ser  $S = h(\varphi(S)) = h(S')$ ; de forma que  $S$  y  $S'$  están en correspondencia biúnica y  $S$  es también minimal suficiente.

### 6.3. Acotaciones de la varianza de un estimador

En los casos en que resulta posible determinar el ECUMV de una determinada función  $g(\theta)$  del parámetro, su varianza supone, por definición, la más pequeña varianza que puede lograrse mediante estimadores insesgados de  $g(\theta)$ .

Aún sin conocer el ECUMV, es posible, en muchos casos, determinar funciones  $c(\theta)$  que acotan inferiormente la varianza de cualquier estimador centrado de  $g(\theta)$ ; de manera que se verifica

$$c(\theta) \leq V_\theta(T)$$

para cualquier estimador  $T$  insesgado de  $g(\theta)$ .

En tales casos, verificar que  $V_\theta(\tilde{T}) = c(\theta)$ , supone un método rápido para comprobar que un determinado estimador  $\tilde{T}$  es ECUMV de  $g(\theta)$ . Además, siempre puede resultar provechosa la comparación entre  $V_\theta(T)$  y  $c(\theta)$  para decidir si merece la pena sustituir  $T$  por otro estimador de menor varianza, o si la proximidad de  $V_\theta(T)$  a  $c(\theta)$  hace poco rentable el esfuerzo.

Para concretar el problema de determinación de tales cotas,  $c(\theta)$ , consideremos una muestra aleatoria simple,  $(X_1, X_2, \dots, X_n)$ , de una población

cuya distribución pertenece a  $\mathcal{F} = \{F_\theta | \theta \in \Theta\}$  y sea  $f_\theta(x_1, \dots, x_n)$  la función de probabilidad o de densidad de la muestra, según que la distribución teórica  $F_\theta$  sea discreta o continua<sup>3</sup>. Prestaremos atención, sobre todo, al segundo caso, puesto que no hay gran diferencia con el primero. Admitiremos además que se verifican las condiciones de regularidad siguientes:

- (i)  $\Theta$  es un intervalo abierto de  $\mathbb{R}$ .
- (ii)  $\{(x_1, \dots, x_n) \in \mathcal{X} | f_\theta(x_1, \dots, x_n) > 0\}$  no depende  $\theta$ .
- (iii) Existe  $\frac{\partial}{\partial \theta} f_\theta(x_1, \dots, x_n)$  para cada  $(x_1, \dots, x_n) \in \mathcal{X}$  y  $\theta \in \Theta$ .
- (iv)  $\int_{\mathcal{X}} \frac{\partial}{\partial \theta} f_\theta(x_1, \dots, x_n) dx_1 \cdots dx_n = 0$   
(o bien,  $\sum_{\mathcal{X}} \frac{\partial}{\partial \theta} f_\theta(x_1, \dots, x_n) = 0$ , en caso de ser  $\mathcal{X}$  numerable).

Excepto (ii), que no se verifica para aquellas familias de distribuciones cuyo recorrido dependa de  $\theta$  —la uniforme en  $(0, \theta)$ , por ejemplo—, el resto de las condiciones no supone una limitación práctica importante acerca de  $\{F_\theta | \theta \in \Theta\}$ <sup>4</sup>. En particular, (iv) sólo supone que la identidad:  $\int_{\mathcal{X}} f_\theta(x_1, \dots, x_n) dx_1 \dots dx_n = 1$ , puede derivarse bajo el signo de integración.

Nótese que, bajo la hipótesis (iii),

$$\frac{\frac{\partial}{\partial \theta} f_\theta(x_1, \dots, x_n)}{f_\theta(x_1, \dots, x_n)} = \frac{\partial}{\partial \theta} \log f_\theta(x_1, \dots, x_n)$$

representa la tasa de variación relativa de  $f_\theta$ , al variar ligeramente  $\theta$ , si la muestra observada ha sido  $(x_1, \dots, x_n)$ . Promediando su cuadrado, se obtiene la *cantidad de información de Fisher*, de la muestra aleatoria acerca

<sup>3</sup> La hipótesis de que el muestreo aleatorio es simple no es realmente necesaria, razón por la cual no expresamos  $f_\theta(x_1, \dots, x_n)$  en términos de la función de probabilidad o de densidad de  $F_\theta$ .

<sup>4</sup> Existe una acotación alternativa, válida también para aquellas familias de distribuciones cuyo recorrido dependa del parámetro, debida a Chapman, Robbins y Kiefer, que no se expondrá aquí. Véase por ejemplo Rohatgi (1976), pág. 365.

de  $\theta$ :

$$\begin{aligned} I(\theta) &= \mathbb{E}_\theta \left[ \left( \frac{\partial}{\partial \theta} \log f_\theta(X_1, \dots, X_n) \right)^2 \right] \\ &= \int_{\mathcal{X}} \left( \frac{\partial}{\partial \theta} \log f_\theta(x_1, \dots, x_n) \right)^2 f_\theta(x_1, \dots, x_n) dx_1 \dots dx_n \end{aligned}$$

que es un índice de la ayuda que proporciona la muestra aleatoria, para discernir cada valor de  $\theta$  de los valores próximos. Así pues, es razonable que  $I(\theta)$  intervenga en el análisis de la precisión con que puede estimarse cualquier función  $g(\theta)$ . Exactamente, se verifica la siguiente

**Desigualdad de Fréchet-Cramer-Rao:** *Bajo las condiciones (i), (ii), (iii) y (iv), si  $T$  es un estadístico unidimensional, con*

$$\mathbb{E}_\theta[T^2] < \infty, \quad \mathbb{E}_\theta[T] = g(\theta)$$

y tal que

$$g'(\theta) = \int_{\mathcal{X}} T(x_1, \dots, x_n) \frac{\partial}{\partial \theta} f_\theta(x_1, \dots, x_n) dx_1 \dots dx_n$$

(o bien  $g'(\theta) = \sum_{\mathcal{X}} T(x_1, \dots, x_n) \frac{\partial}{\partial \theta} f_\theta(x_1, \dots, x_n)$ , en caso de ser discreta la distribución teórica), se verifica la desigualdad

$$g'(\theta)^2 \leq V_\theta(T)I(\theta) \quad \text{para todo } \theta \in \Theta.$$

Por tanto,  $V_\theta(T) \geq \frac{g'(\theta)^2}{I(\theta)}$ , a no ser que  $I(\theta) = 0$ .

La igualdad  $g'(\theta)^2 = V_\theta(T)I(\theta)$  se produce si y sólo si existe una función  $k(\theta)$  tal que  $T(x_1, \dots, x_n) = g(\theta) + k(\theta) \frac{\partial}{\partial \theta} \log f_\theta(x_1, \dots, x_n)$ ,  $\mathcal{P}_\theta$ -casi seguro para cada  $\theta \in \Theta$ .

Demostración: Sea  $U = \frac{\partial}{\partial \theta} \log f_\theta(X_1, \dots, X_n)$ ; según (iv) es  $\mathbb{E}_\theta[U] = 0$ , así que

$$\text{Cov}_\theta(T, U) = \mathbb{E}_\theta[TU] = \int_{\mathcal{X}} T(x_1, \dots, x_n) \frac{\partial}{\partial \theta} f_\theta(x_1, \dots, x_n) dx_1 \dots dx_n = g'(\theta)$$

Por consiguiente,

$$g'(\theta)^2 \leq V_\theta(T)V_\theta(U) = V_\theta(T)\mathbb{E}_\theta[U^2] = V_\theta(T)I(\theta).$$

La igualdad  $g'(\theta)^2 = V_\theta(T)I(\theta)$ , significa que el coeficiente de correlación entre  $T$  y  $U$  es  $\pm 1$ , para cada  $\theta \in \Theta$ . Es decir que  $T - g(\theta)$  y  $U$  son (con probabilidad 1) proporcionales<sup>5</sup>.

Como se ve el resultado anterior no afirma nada para aquellos estadísticos anómalos  $T$ , cuya media:  $g(\theta) = \int_{\chi} T(x_1, \dots, x_n) f_\theta(x_1, \dots, x_n) dx_1 \dots dx_n$ , no pueda derivarse bajo el signo de integración. Ello supone, en primer lugar, que la desigualdad es cierta si se estiman funciones  $g(\theta)$  derivables — lo cual no es grave — y, en segundo lugar, que la desigualdad se convierte en una acotación valiosa en aquellos casos en que no existan tales estadísticos anómalos<sup>6</sup>.

Generalmente no es fácil comprobar que la desigualdad se verifica para cualquier estadístico  $T$ , pero ello es cierto en las situaciones usuales, de manera que, muy a menudo, se omite la comprobación. Veamos como habría que proceder en algunas situaciones concretas y qué conclusiones pueden obtenerse en ellas.

### Ejemplo 6.10

Sea  $(X_1, X_2, \dots, X_n)$  una muestra aleatoria simple de una distribución de Poisson de parámetro  $\theta > 0$  desconocido. La función de probabilidad de la muestra es entonces

$$f_\theta(x_1, \dots, x_n) = e^{-n\theta} \frac{\theta^{x_1 + \dots + x_n}}{x_1! \dots x_n!} \quad \text{para } x_1, \dots, x_n = 0, 1, 2, \dots$$

con lo cual

$$\log f_\theta(x_1, \dots, x_n) = -n\theta + \sum_{i=1}^n x_i \log \theta - \sum_{i=1}^n \log x_i!$$

y

$$\frac{\partial}{\partial \theta} \log f_\theta(x_1, \dots, x_n) = -n + \frac{1}{\theta} \sum_{i=1}^n x_i.$$

<sup>5</sup> La constante de proporcionalidad,  $k(\theta)$ , es el coeficiente de regresión de  $T$  sobre  $U$ :

$$k(\theta) = \frac{\text{Cov}(T, U)}{V_\theta(U)} = \frac{g'(\theta)}{I(\theta)}$$

<sup>6</sup> Pues la diferencia entre las afirmaciones “ $V_\theta(T) \geq c(\theta)$  para cualquier  $T$ ” y “ $V_\theta(T) \geq c(\theta)$  para la mayoría de estimadores  $T$ ” es muy considerable.

De forma que

$$\begin{aligned} I(\theta) &= \mathbb{E}_\theta \left[ \frac{1}{\theta^2} \left( \sum_{i=1}^n X_i \right)^2 - \frac{2n}{\theta} \sum_{i=1}^n X_i + n^2 \right] \\ &= \frac{n\theta + n^2\theta^2}{\theta^2} - \frac{2n}{\theta}n\theta + n^2 = \frac{n}{\theta} \end{aligned}$$

(habida cuenta que  $\sum_{i=1}^n X_i$  tiene distribución de Poisson de parámetro  $n\theta$ ).

Para comprobar que se puede derivar bajo el signo de sumación, consideremos cualquier estadístico  $T(x_1, \dots, x_n)$  con media  $g(\theta)$ , derivable respecto a  $\theta$ ; es decir:

$$\sum_{x_1, \dots, x_n=0}^{\infty} T(x_1, \dots, x_n) e^{-n\theta} \frac{\theta^{x_1+\dots+x_n}}{x_1! \cdots x_n!} = g(\theta).$$

Llamando  $B(s) = \sum_{x_1+\dots+x_n=s} \frac{T(x_1, \dots, x_n)}{x_1! \cdots x_n!}$  y  $h(\theta) = e^{n\theta} g(\theta)$ , resulta

$$\sum_{s=0}^{\infty} B(s)\theta^s = h(\theta) \quad \text{para todo } \theta > 0.$$

Como una serie de potencias convergente puede ser derivada término a término<sup>7</sup>, se obtiene

$$\sum_{s=0}^{\infty} B(s)s\theta^{s-1} = e^{n\theta} \{ng(\theta) + g'(\theta)\};$$

es decir

$$\begin{aligned} g'(\theta) &= \sum_{s=0}^{\infty} B(s)s\theta^{s-1}e^{-n\theta} - ng(\theta) \\ &= \sum_{x_1, \dots, x_n=0}^{\infty} T(x_1, \dots, x_n)(x_1 + \dots + x_n) \frac{\theta^{x_1+\dots+x_n-1}}{x_1! \cdots x_n!} e^{-n\theta} \\ &\quad - n \sum_{x_1, \dots, x_n=0}^{\infty} T(x_1, \dots, x_n) \frac{\theta^{x_1+\dots+x_n}}{x_1! \cdots x_n!} e^{-n\theta} \\ &= \sum_{x_1, \dots, x_n=0}^{\infty} T(x_1, \dots, x_n) \frac{\partial}{\partial \theta} \left( e^{-n\theta} \frac{\theta^{x_1+\dots+x_n}}{x_1! \cdots x_n!} \right). \end{aligned}$$

---

<sup>7</sup> Véase, T.M. Apostol (1976), *Análisis Matemático*, pág. 287.

En particular, para  $T \equiv 1$ , se obtiene (iv). Y, en general, el razonamiento anterior prueba que cualquier estadístico verifica la desigualdad de Fréchet-Cramer-Rao (correspondiente a su media).

Si el propósito es estimar  $\theta$ , será entonces

$$V_\theta(T) \geq \frac{1}{I(\theta)} = \frac{\theta}{n} \quad \text{para cualquier estimador } T \text{ con } E_\theta[T] = \theta.$$

Nótese que  $E_\theta[\bar{X}] = \theta$  y  $V_\theta(\bar{X}) = \theta/n$ ; de manera que  $\bar{X}$  es un estimador centrado de  $\theta$  que alcanza la cota de Fréchet-Cramer-Rao. Tiene que ser por tanto el ECUMV para  $\theta$ ; resultado que ya conocíamos, pues, en virtud del ejemplo 6.6,  $\bar{X}$  es suficiente y completo.

En cambio, si el objetivo fuese estimar  $g(\theta) = e^{-\theta}$ , se puede concluir que

$$V_\theta(T) \geq e^{-2\theta} \frac{\theta}{n} \quad \text{para cualquier estimador } T \text{ centrado de } e^{-\theta}.$$

Sin embargo, el ECUMV para  $e^{-\theta}$  es  $((n-1)/n)^{n\bar{X}}$  (cf. ejemplo 6.1) que tiene por varianza  $e^{-2\theta}(e^{\theta/n}-1)$ . De manera que, la cota de Fréchet-Cramer-Rao para estimadores insesgados de  $e^{-\theta}$  resulta inaccesible: queda estrictamente por debajo de la varianza de cualquier estimador centrado de  $e^{-\theta}$ .

Aquellos estimadores  $T$ , cuya varianza coincide con la cota de Fréchet-Cramer-Rao, para estimadores centrados de  $g(\theta) = E_\theta[T]$ , se denominan *estimadores eficientes* de  $g(\theta)$ . Así, en el ejemplo anterior,  $\bar{X}$  es eficiente para  $\theta$ , pero  $((n-1)/n)^{n\bar{X}}$  no lo es para  $e^{-\theta}$ ; a pesar de que ambos son ECUMV de su respectiva media. Naturalmente<sup>8</sup>, si  $\tilde{T}$  es estimador eficiente de  $g(\theta)$ , también es ECUMV de  $g(\theta)$ .

Estudiemos otro ejemplo.

### Ejemplo 6.11

Sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple de una distribución exponencial de parámetro  $\theta > 0$ . Su densidad es

$$f_\theta(x_1, \dots, x_n) = \theta^n e^{-\theta(x_1 + \dots + x_n)} \quad \text{para } x_1, \dots, x_n > 0.$$

---

<sup>8</sup> Supuesto que se pueda asegurar que cualquier estimador centrado de  $g(\theta)$  cumple la desigualdad de Fréchet-Cramer-Rao.

Luego  $\log f_\theta(x_1, \dots, x_n) = n \log \theta - \theta \sum_{i=1}^n x_i$  y

$$\frac{\partial}{\partial \theta} \log f_\theta(x_1, \dots, x_n) = \frac{n}{\theta} - \sum_{i=1}^n x_i;$$

con lo cual

$$I(\theta) = E_\theta \left[ \left( \sum_{i=1}^n X_i \right)^2 - \frac{2n}{\theta} \sum_{i=1}^n X_i + \frac{n^2}{\theta^2} \right] = \frac{n+n^2}{\theta^2} - \frac{2n^2}{\theta^2} + \frac{n^2}{\theta^2} = \frac{n}{\theta^2},$$

puesto que  $E_\theta \left[ \sum_{i=1}^n X_i \right] = \frac{n}{\theta}$  y  $V_\theta \left( \sum_{i=1}^n X_i \right) = \frac{n}{\theta^2}$ .

Para comprobar que cualquier estadístico  $T$  verifica la desigualdad de Fréchet-Cramer-Rao, supongamos que su media es una función derivable  $g(\theta)$ , es decir

$$\int_{[0,\infty)^n} T(x_1, \dots, x_n) e^{-\theta(x_1 + \dots + x_n)} dx_1 \dots dx_n = g(\theta) \theta^{-n}.$$

Derivando<sup>9</sup> respecto a  $\theta$ , se obtiene

$$-\int_{[0,\infty)^n} T(x_1, \dots, x_n) e^{-\theta \sum_{i=1}^n x_i} \sum_{i=1}^n x_i dx_1 \dots dx_n = g'(\theta) \theta^{-n} - n g(\theta) \theta^{-n-1}$$

es decir

$$\begin{aligned} g'(\theta) &= \int_{[0,\infty)^n} T(x_1, \dots, x_n) \left\{ n - \theta \sum_{i=1}^n x_i \right\} \theta^{n-1} e^{-\theta \sum_{i=1}^n x_i} dx_1 \dots dx_n \\ &= \int_{[0,\infty)^n} T(x_1, \dots, x_n) \frac{\partial}{\partial \theta} \left( \theta^n e^{-\theta \sum_{i=1}^n x_i} \right) dx_1 \dots dx_n. \end{aligned}$$

<sup>9</sup> La derivación bajo el signo integral está justificada en cada intervalo  $(\theta_1, \theta_2)$  por ser

$$|T(x_1, \dots, x_n) e^{-\theta \sum_{i=1}^n x_i} \sum_{i=1}^n x_i| \leq |T(x_1, \dots, x_n)| e^{-\theta_1/2 \sum_{i=1}^n x_i} \cdot \sum_{i=1}^n x_i e^{-\theta_1/2 \sum_{i=1}^n x_i}$$

siendo el primer factor integrable y el segundo acotado por una constante. Véase T.M. Apostol (1976): *Análisis Matemático*, pág. 345.

Si el propósito es estimar  $\theta$ , la acotación de Fréchet-Cramer-Rao para estimadores centrados de  $\theta$  asegura

$$V_\theta(T) \geq \frac{1}{I(\theta)} = \frac{\theta^2}{n}$$

para cualquier estimador centrado de  $\theta$ . Sin embargo, el ECUMV para  $\theta$  es  $(n - 1) / \sum_{i=1}^n X_i$  (cf. ejemplo 6.7), cuya varianza es

$$\begin{aligned} V_\theta\left(\frac{n-1}{\sum_{i=1}^n X_i}\right) &= E\left[\left(\frac{n-1}{\sum_{i=1}^n X_i}\right)^2\right] - \theta^2 \\ &= \int_0^\infty \frac{(n-1)^2}{y^2} \frac{\theta^n y^{n-1} e^{-\theta y}}{(n-1)!} dy - \theta^2 \\ &= \frac{n-1}{n-2} \theta^2 - \theta^2 = \frac{\theta^2}{n-2}, \end{aligned}$$

puesto que  $\sum_{i=1}^n X_i$  tiene distribución gamma  $\gamma(n, \theta)$ . Luego, a pesar de ser el ECUMV de su media,  $(n - 1) / \sum_{i=1}^n X_i$  no es eficiente.

En cambio, para estimar  $1/\theta$ , el ECUMV es  $\bar{X}$ , cuya varianza  $V_\theta(\bar{X}) = 1/n\theta^2$  coincide con la cota de Fréchet-Cramer-Rao para estimadores insesgados de  $1/\theta$ :

$$\frac{(-1/\theta^2)^2}{I(\theta)} = \frac{1}{n\theta^2}$$

De manera que  $\bar{X}$  es estimador eficiente de  $1/\theta$ .

Hay dos observaciones que facilitan el cálculo de la cota de Fréchet-Cramer-Rao:

(a) Si el muestreo es efectivamente aleatorio simple, será

$$f_\theta(x_1, \dots, x_n) = f_\theta(x_1)f_\theta(x_2) \cdots f_\theta(x_n)$$

siendo  $f_\theta(x)$  la función de densidad o de probabilidad correspondiente la distribución teórica  $F_\theta$ . Por tanto,  $\frac{\partial}{\partial \theta} \log f_\theta(x_1, \dots, x_n) = \sum_{i=1}^n \frac{\partial}{\partial \theta} \log f_\theta(x_i)$ .

Como  $\frac{\partial}{\partial \theta} \log f_\theta(X_i)$  ( $i = 1, \dots, n$ ), son variables independientes e igualmente distribuidas, con

$$E_\theta \left[ \frac{\partial}{\partial \theta} \log f_\theta(X_i) \right] = \frac{1}{n} E_\theta \left[ \frac{\partial}{\partial \theta} \log f_\theta(X_1, \dots, X_n) \right] = 0,$$

resulta

$$\begin{aligned} I(\theta) &= \mathbb{E}_\theta \left[ \left( \frac{\partial}{\partial \theta} \log f_\theta(X_1, \dots, X_n) \right)^2 \right] = V_\theta \left( \frac{\partial}{\partial \theta} \log f_\theta(X_1, \dots, X_n) \right) \\ &= \sum_{i=1}^n V_\theta \left( \frac{\partial}{\partial \theta} \log f_\theta(X_i) \right) = n \mathbb{E}_\theta \left[ \left( \frac{\partial}{\partial \theta} \log f_\theta(X_1) \right)^2 \right] = n i(\theta) \end{aligned}$$

siendo  $i(\theta)$  la información de Fisher de una muestra de tamaño 1.

(b) Si la igualdad  $\int_X \frac{\partial}{\partial \theta} f_\theta(x_1, \dots, x_n) dx_1 \dots dx_n = 0$ , puede derivarse bajo el signo de integración, se cumple  $\int_X \frac{\partial^2}{\partial \theta^2} f_\theta(x_1, \dots, x_n) dx_1 \dots dx_n = 0$ . Pero,

$$-\frac{\partial^2}{\partial \theta^2} \log f_\theta(x_1, \dots, x_n) = \left( \frac{\partial}{\partial \theta} \log f_\theta(x_1, \dots, x_n) \right)^2 - \frac{\frac{\partial^2}{\partial \theta^2} f_\theta(x_1, \dots, x_n)}{f_\theta(x_1, \dots, x_n)},$$

así que se obtiene

$$\mathbb{E}_\theta \left[ -\frac{\partial^2}{\partial \theta^2} \log f_\theta(X_1, \dots, X_n) \right] = \mathbb{E}_\theta \left[ \left( \frac{\partial}{\partial \theta} \log f_\theta(X_1, \dots, X_n) \right)^2 \right] = I(\theta).$$

Con ambas simplificaciones, en el ejemplo 6.10, basta calcular

$$f_\theta(x) = e^{-\theta} \frac{\theta^x}{x!}; \quad \log f_\theta(x) = -\theta + x \log \theta - \log x!; \quad -\frac{\partial^2}{\partial \theta^2} \log f_\theta(x) = \frac{x}{\theta^2};$$

para obtener  $i(\theta) = \frac{\mathbb{E}[X]}{\theta^2} = \frac{1}{\theta}$ , e  $I(\theta) = \frac{n}{\theta^2}$ .

En cuanto al ejemplo 6.11, como

$$f_\theta(x) = \theta e^{-\theta x}; \quad \log f_\theta(x) = \log \theta - \theta x; \quad -\frac{\partial^2}{\partial \theta^2} \log f_\theta(x) = \frac{1}{\theta^2};$$

se obtiene  $i(\theta) = \frac{1}{\theta^2}$ , e  $I(\theta) = \frac{n}{\theta^2}$ .

Estimadores eficientes sólo hay en un número muy limitado de casos. Concretamente, bajo las condiciones de regularidad impuestas para la acotación de Fréchet-Cramer-Rao, la igualdad de  $V_\theta(T)$  con su cota, se produce cuando

$$\frac{\partial}{\partial \theta} \log f_\theta(x_1, \dots, x_n) = \frac{1}{k(\theta)} [T(x_1, \dots, x_n) - g(\theta)];$$

es decir, si

$$\log f_\theta(x_1, \dots, x_n) = K(\theta)T(x_1, \dots, x_n) + M(\theta) + R(x_1, \dots, x_n)$$

(siendo  $K'(\theta) = 1/k(\theta)$  y  $M'(\theta) = -g(\theta)/k(\theta)$ ); lo cual equivale a

$$f_\theta(x_1, \dots, x_n) = a(\theta)b(x_1, \dots, x_n)e^{q(\theta)T(x_1, \dots, x_n)}.$$

Ello corresponde a una familia exponencial (véase ejemplo 5.17);  $T$  tiene que ser el estadístico que figura en su exponente y  $g(\theta)$  su media. Si la distribución de la muestra no queda caracterizada por una función  $f_\theta(x_1, \dots, x_n)$  de la forma anterior, o la función  $g(\theta)$  a estimar no es la esperanza del estimador del exponente, no existe estimador eficiente para  $g(\theta)$ .

Pese a la limitación anterior, la cota de Fréchet-Cramer-Rao, se usa frecuentemente como instrumento de medida de la calidad de un estimador. De hecho, se denomina *eficiencia de un estimador  $T$*  al cociente entre su cota y su varianza:

$$\text{eficiencia}_T(\theta) = \frac{g'(\theta)^2/I(\theta)}{V_\theta(T)} \quad \text{siendo } g(\theta) = E_\theta[T];$$

lo cual siempre da un número menor que 1, excepto para estimadores eficientes cuya eficiencia es 1.

La acotación de Fréchet-Cramer-Rao puede generalizarse al caso en que la distribución poblacional depende de un parámetro  $\theta = (\theta_1, \dots, \theta_k)$ , que es  $k$ -dimensional. Naturalmente, las condiciones de regularidad han de ser extendidas a esta nueva situación, suponiendo:

- (i)  $\Theta$  es un abierto conexo de  $I\!\!R^k$ .
- (ii)  $\{(x_1, \dots, x_n) \in \mathcal{X} \mid f_\theta(x_1, \dots, x_n) > 0\}$  no depende de  $\theta$ .
- (iii) Existe  $\frac{\partial}{\partial \theta_j} f_\theta(x_1, \dots, x_n)$  para cada  $j = 1, 2, \dots, k$ ,  $(x_1, \dots, x_n) \in \mathcal{X}$  y  $\theta \in \Theta$ .
- (iv)  $\int_{\mathcal{X}} \frac{\partial}{\partial \theta_j} f_\theta(x_1, \dots, x_n) dx_1 \cdots dx_n = 0$  para cada  $j = 1, 2, \dots, k$   
(o bien,  $\sum_{\mathcal{X}} \frac{\partial}{\partial \theta_j} f_\theta(x_1, \dots, x_n) = 0$ , en caso de ser  $\mathcal{X}$  numerable).

La información de Fisher,  $I(\theta)$ , es, en este caso, una matriz simétrica,  $k \times k$ , cuyo elemento  $(i, j)$  es

$$E_{\theta} \left[ \frac{\partial}{\partial \theta_i} \log f_{\theta}(X_1, \dots, X_n) \cdot \frac{\partial}{\partial \theta_j} \log f_{\theta}(X_1, \dots, X_n) \right]$$

de manera que los términos de la diagonal,  $I(\theta)_{ii}$ , son las cantidades de información, definidas anteriormente, acerca de cada parámetro individual  $\theta_i$ .

Si  $T = (T_1, \dots, T_r)$  es un estimador  $r$ -dimensional, que estima sin sesgo una función paramétrica de dimensión  $r$ :  $g(\theta) = (g_1(\theta), \dots, g_r(\theta))$ , la matriz de covarianzas de los errores de estimación:

$$V_{\theta}(T) = E_{\theta} [(T - g(\theta))^t \cdot (T - g(\theta))],$$

puede usarse como criterio de calidad del estimador  $T$ , puesto que mide la dispersión, en el espacio  $\mathbb{R}^r$ , de la distribución del vector  $T = (T_1, \dots, T_r)$ , alrededor del punto  $g(\theta) = (g_1(\theta), \dots, g_r(\theta))$ .

Se verifica entonces

**Teorema:** *Bajo las condiciones (i), (ii), (iii) y (iv), si  $T = (T_1, \dots, T_r)$  es un estadístico  $r$ -dimensional, con*

$$E_{\theta}[T_i^2] < \infty, \quad E_{\theta}[T_i] = g_i(\theta) \quad \text{para cada } i = 1, \dots, r$$

y tal que, para cada  $i = 1, \dots, r$  y  $j = 1, \dots, k$ :

$$\frac{\partial}{\partial \theta_j} g_i(\theta) = \int_{\mathcal{X}} T_i(x_1, \dots, x_n) \frac{\partial}{\partial \theta_j} f_{\theta}(x_1, \dots, x_n) dx_1 \dots dx_n$$

(o bien  $\frac{\partial}{\partial \theta_j} g_i(\theta) = \sum_{\mathcal{X}} T_i(x_1, \dots, x_n) \frac{\partial}{\partial \theta_j} f_{\theta}(x_1, \dots, x_n)$ , en caso de ser discreta la distribución teórica), se verifica que

$$V_{\theta}(T) - \Delta I(\theta)^{-1} \Delta^t$$

es una matriz definida no negativa, supuesto que  $I(\theta)$  es definida positiva y siendo  $\Delta$  la matriz  $r \times k$  cuyo término  $(i, j)$  es  $\frac{\partial}{\partial \theta_j} g_i(\theta)$ .

Demostración: Sea

$$U = \left( \frac{\partial}{\partial \theta_1} \log f_{\theta}(X_1, \dots, X_n), \dots, \frac{\partial}{\partial \theta_k} \log f_{\theta}(X_1, \dots, X_n) \right).$$

Según (iv), es  $E_\theta[U] = 0$  y, por otra parte,

$$E_\theta[T_i \cdot U_j] = \int_{\mathcal{X}} T_i(x_1, \dots, x_n) \frac{\partial}{\partial \theta_j} f_\theta(x_1, \dots, x_n) dx_1 \dots dx_n = \frac{\partial}{\partial \theta_j} g_i(\theta);$$

es decir,  $E_\theta[T^t \cdot U] = \Delta$ . Por tanto la matriz de covarianzas del vector  $(T, U)$  es

$$V_\theta(T, U) = \begin{pmatrix} V_\theta(T) & \Delta \\ \Delta^t & I(\theta) \end{pmatrix},$$

que necesariamente será definida no negativa. De ahí que  $MV_\theta(T, U)M'$  sea definida no negativa para cualquier matriz  $M$  (con  $r + k$  columnas). Y, en particular, es definida no negativa

$$\begin{pmatrix} I & -\Delta I(\theta)^{-1} \\ 0 & I(\theta)^{-1} \end{pmatrix} \begin{pmatrix} V_\theta(T) & \Delta \\ \Delta^t & I(\theta) \end{pmatrix} \begin{pmatrix} I & 0 \\ -I(\theta)^{-1}\Delta^t & I(\theta)^{-1} \end{pmatrix} = \\ = \begin{pmatrix} V_\theta(T) - \Delta I(\theta)^{-1}\Delta^t & 0 \\ 0 & I(\theta)^{-1} \end{pmatrix};$$

lo cual establece el resultado.

Salvo en el caso  $r = 1$ , la afirmación del teorema no es propiamente una desigualdad, pero puede transformarse fácilmente en varias de ellas:

(1) Considerando un sólo término de la diagonal, se obtiene

$$V_\theta(T_i) \geq (\Delta I(\theta)^{-1}\Delta^t)_{ii}$$

que acota la varianza de los estimadores insesgados de una función real cualquiera  $g_i(\theta)$  de los parámetros. (Sólo interviene en esta desigualdad, la fila  $i$ -ésima de  $\Delta$ , constituida por las derivadas de  $g_i(\theta)$  respecto a los diversos parámetros). Se trata, por tanto, de una generalización directa de la acotación de Fréchet-Cramer-Rao original, al caso de la estimación de una única función de varios parámetros.

En particular, con  $g_i(\theta) = \theta_i$ , se obtiene<sup>10</sup>

$$V_\theta(T_i) \geq (I(\theta)^{-1})_{ii}$$

como acotación de la varianza de los estimadores insesgados de  $\theta_i$ , en presencia de los demás parámetros desconocidos de los que depende la población.

---

<sup>10</sup> La fila  $i$ -ésima de  $\Delta$  es entonces  $(0, \dots, 0, 1, 0, \dots, 0)$ .

Obsérvese que, si se descompone:

$$I(\theta) = \begin{pmatrix} a_{11} & b \\ b' & A_{11} \end{pmatrix} \quad \text{con} \quad a_{11} = I(\theta)_{11},$$

puesto que  $|A_{11}| > 0$  (en el supuesto que  $I(\theta)$  es definida positiva), existe un vector  $x$  tal que  $A_{11}x^t = b^t$  y, por consiguiente,

$$|I(\theta)| = \begin{vmatrix} a_{11} - bx^t & b \\ 0 & A_{11} \end{vmatrix} = (a_{11} - bx^t)|A_{11}| = (a_{11} - xA_{11}x^t)|A_{11}|$$

o bien

$$a_{11} \frac{|A_{11}|}{|I(\theta)|} = 1 + xA_{11}x^t \frac{|A_{11}|}{|I(\theta)|};$$

con lo cual  $\frac{|A_{11}|}{|I(\theta)|} \geq \frac{1}{a_{11}}$ .

Ello prueba<sup>11</sup> que  $(I(\theta)^{-1})_{ii} \geq 1/I(\theta)_{ii}$ . El segundo miembro es la cota de Fréchet-Cramer-Rao unidimensional, para la varianza de los estimadores insesgados de  $\theta_i$ , cuando se suponen conocidos todos los demás parámetros. En cambio,  $(I(\theta)^{-1})_{ii}$  es una cota más precisa, que tiene en cuenta el aumento de varianza, en la estimación de  $\theta_i$ , debido a la presencia de los otros parámetros desconocidos.

$$(2) \quad \sum_{i=1}^r V_\theta(T_i) = \text{Traza}[V_\theta(T)] \geq \text{Traza}[\Delta I(\theta)^{-1} \Delta^t]$$

acotación para la suma de las varianzas de  $T_1, \dots, T_r$ , que representa el riesgo del estimador  $r$ -dimensional  $T$ , centrado para  $g(\theta)$ , si se considera como función de pérdida  $L(g(\theta), t) = \sum_{i=1}^r (t_i - g_i(\theta))^2$ .

$$(3) \quad |V_\theta(T)| \geq |\Delta I(\theta)^{-1} \Delta'|$$

que, en el caso  $r = k$ , se expresa  $|V_\theta(T)| \geq \frac{|\Delta|^2}{|I(\theta)|}$ , con una apariencia similar a la del caso unidimensional. En este contexto, el determinante  $|V_\theta(T)|$  se denomina *varianza generalizada* del estimador  $T = (T_1, \dots, T_r)$ .

---

<sup>11</sup> Para  $i = 1$ ; pero el razonamiento sería el mismo para cualquier  $i$ .

### Ejemplo 6.12

Los individuos de una población pueden ser clasificados, según su grupo sanguíneo, en cuatro categorías: 1, 2, 3 y 4, cuyas proporciones poblacionales:  $\theta_1, \theta_2, \theta_3$  y  $\theta_4$  son desconocidas. Consideraremos  $\theta = (\theta_1, \theta_2, \theta_3)$  como parámetro tridimensional, mientras que  $\theta_4 = 1 - \theta_1 - \theta_2 - \theta_3$  es una función del parámetro. Para efectuar estimaciones sobre  $\theta$ , se observan y clasifican  $n$  individuos; de forma que la probabilidad de obtener la muestra  $x_1, x_2, \dots, x_n \in \{1, 2, 3, 4\}$ , es

$$f_{\theta}(x_1, \dots, x_n) = \theta_1^{n_1} \theta_2^{n_2} \theta_3^{n_3} \theta_4^{n_4}$$

siempre que  $(x_1, \dots, x_n)$  contenga  $n_1$  unos,  $n_2$  doses,  $n_3$  treses y  $n_4 = n - n_1 - n_2 - n_3$  cuatros. Como

$$\log f_{\theta}(X_1, \dots, X_n) = n_1 \log \theta_1 + n_2 \log \theta_2 + n_3 \log \theta_3 + n_4 \log \theta_4,$$

$$\text{será } \frac{\partial}{\partial \theta_j} f_{\theta}(x_1, \dots, x_n) = \frac{n_j}{\theta_j} - \frac{n_4}{\theta_4} \text{ para } j = 1, 2, 3.$$

Ahora bien, la variable aleatoria  $N_j$  cuyo resultado es  $n_j$  —número de individuos, entre los  $n$ , clasificados en la categoría  $j$ — tiene distribución binomial  $B(n, \theta_j)$ ; con lo cual  $E[N_j] = n\theta_j$  y  $V_{\theta}(N_j) = n\theta_j(1 - \theta_j)$ . Además, expresando  $N_j = \sum_{i=1}^n I_{\{X_i=j\}}$ , si  $j \neq l$  se obtiene

$$E_{\theta}[N_j \cdot N_l] = E_{\theta} \left[ \sum_{i \neq k} I_{\{X_i=j\}} I_{\{X_k=l\}} \right] = n(n-1)\theta_j\theta_l.$$

Por tanto, para  $j = 1, 2, 3$ ,

$$E_{\theta} \left[ \left( \frac{\partial}{\partial \theta_j} \log f_{\theta}(X_1, \dots, X_n) \right)^2 \right] = E_{\theta} \left[ \frac{N_j^2}{\theta_j^2} + \frac{N_4^2}{\theta_4^2} - 2 \frac{N_j N_4}{\theta_j \theta_4} \right]$$

$$= n \frac{1 - \theta_j}{\theta_j} + n^2 + n \frac{1 - \theta_4}{\theta_4} + n^2 - 2n(n-1) = n \left( \frac{1}{\theta_j} + \frac{1}{\theta_4} \right)$$

$$\text{y } E_{\theta} \left[ \frac{\partial}{\partial \theta_j} \log f_{\theta}(X_1, \dots, X_n) \cdot \frac{\partial}{\partial \theta_l} \log f_{\theta}(X_1, \dots, X_n) \right] =$$

$$= E_{\theta} \left[ \left( \frac{N_j}{\theta_j} - \frac{N_4}{\theta_4} \right) \left( \frac{N_l}{\theta_l} - \frac{N_4}{\theta_4} \right) \right] = \frac{n}{\theta_4}.$$

Con lo cual<sup>12</sup>,

$$I(\theta) = \frac{n}{\theta_4} \begin{pmatrix} 1 + \theta_4/\theta_1 & 1 & 1 \\ 1 & 1 + \theta_4/\theta_2 & 1 \\ 1 & 1 & 1 + \theta_4/\theta_3 \end{pmatrix}$$

y resulta

$$I(\theta)^{-1} = \frac{1}{n} \begin{pmatrix} \theta_1(1 - \theta_1) & -\theta_1\theta_2 & -\theta_1\theta_3 \\ -\theta_1\theta_2 & \theta_2(1 - \theta_2) & -\theta_2\theta_3 \\ -\theta_1\theta_3 & -\theta_2\theta_3 & \theta_3(1 - \theta_3) \end{pmatrix}$$

Como estaba previsto, la cota  $\theta_1(1 - \theta_1)/n$  para la varianza de cualquier estimador insesgado de  $\theta_1$ , es superior a la cota  $[n(1/\theta_1 + 1/\theta_4)]^{-1}$ , que se obtendría en el caso de que  $\theta_2$  y  $\theta_3$  fuesen conocidos.

Si el objetivo consiste en estimar  $\theta = (\theta_1, \theta_2, \theta_3)$ , será  $\Delta = I$ , con lo cual  $\Delta I(\theta)^{-1}\Delta' = I(\theta)^{-1}$ . Como el estimador  $T = (N_1/n, N_2/n, N_3/n)$ , estima sin sesgo  $(\theta_1, \theta_2, \theta_3)$  y tiene matriz de covarianzas  $I(\theta)^{-1}$ ,  $T$  consigue la igualdad en todas las desigualdades obtenidas; en este sentido, es eficiente. Ello no es sorprendente puesto que  $T$  es un estadístico suficiente y completo<sup>13</sup>.

Si se trata de estimar las probabilidades de que una pareja de individuos pertenezcan ambos al primer o al segundo grupo:  $g_1(\theta) = \theta_1^2$ ,  $g_2(\theta) = \theta_2^2$ , será

$$\Delta = 2 \begin{pmatrix} \theta_1 & 0 & 0 \\ 0 & \theta_2 & 0 \end{pmatrix} \quad \text{y} \quad \Delta I(\theta)^{-1}\Delta' = \frac{4}{n} \begin{pmatrix} \theta_1^3(1 - \theta_1) & -\theta_1^2\theta_2^2 \\ -\theta_1^2\theta_2^2 & \theta_2^3(1 - \theta_2) \end{pmatrix}.$$

En este caso los ECUMV de  $\theta_1^2$  y  $\theta_2^2$  son  $\frac{N_1(N_1 - 1)}{n(n - 1)}$  y  $\frac{N_2(N_2 - 1)}{n(n - 1)}$  respectivamente. La matriz de covarianzas (bastante pesada de calcular) de

<sup>12</sup> El cálculo anterior hubiese resultado más sencillo, utilizando la observación (b):

$$E_{\theta} \left[ \left( \frac{\partial}{\partial \theta_j} \log f_{\theta}(X_1, \dots, X_n) \right)^2 \right] = E_{\theta} \left[ -\frac{\partial^2}{\partial \theta_j^2} \log f_{\theta}(X_1, \dots, X_n) \right]$$

y razonando de manera similar que

$$E_{\theta} \left[ \frac{\partial}{\partial \theta_j} \log f_{\theta}(X_1, \dots, X_n) \cdot \frac{\partial}{\partial \theta_l} \log f_{\theta}(X_1, \dots, X_n) \right] = E_{\theta} \left[ -\frac{\partial^2}{\partial \theta_j \partial \theta_l} \log f_{\theta}(X_1, \dots, X_n) \right]$$

siempre que la igualdad  $\int_{\mathcal{X}} \frac{\partial}{\partial \theta_j} f_{\theta}(x_1, \dots, x_n) dx_1 \dots dx_n = 0$  pueda derivarse, respecto a  $\theta_l$ , bajo el signo de integración.

<sup>13</sup> La completitud se establece de manera similar a la utilizada en el ejemplo 6.3.

$$T' = \left( \frac{N_1(N_1 - 1)}{n(n-1)}, \frac{N_2(N_2 - 1)}{n(n-1)} \right) \text{ es}$$

$$\frac{2}{n(n-1)} \begin{pmatrix} \theta_1^2(1-\theta_1)[1+(2n-3)\theta_1^2] & -(2n-3)\theta_1^2\theta_2^2 \\ -(2n-3)\theta_1^2\theta_2^2 & \theta_2^2(1-\theta_2)[1+(2n-3)\theta_2^2] \end{pmatrix}.$$

Su comparación con  $\Delta I(\theta)^{-1}\Delta'$  ilustra las tres desigualdades enunciadas.

## 6.4. Estimación de parámetros de posición y escala

### 6.4.1. Parámetros de posición

El primer objetivo, en esta sección, será la estimación invariante óptima, en familias de distribuciones que dependan únicamente de un parámetro de posición.

Se supone pues, por una parte, que se dispone de una muestra aleatoria  $(X_1, \dots, X_n)$ , de una población cuya distribución es unidimensional y verifica

$$F_\theta(x) = F_0(x - \theta)$$

siendo  $F_0$  una distribución conocida.

Por otra parte, y sobre todo, se abandona el empeño en conseguir estimadores centrados de  $\theta$ , para buscar estimadores invariantes por translaciones; es decir, que verifiquen

$$T(x_1 + c, \dots, x_n + c) = T(x_1, \dots, x_n) + c \quad \text{para cualquier } c \in \mathbb{R}.$$

Como ya se indicó en la sección 5.3, la justificación para ello es que, como la forma de la distribución poblacional no varía, sino que sólo se traslada al cambiar el valor de  $\theta$ , la evidencia que proporciona la muestra  $(x_1 + c, \dots, x_n + c)$  en favor de la distribución  $F_{\theta_1+c}$ , es idéntica a la que proporciona la muestra  $(x_1, \dots, x_n)$  en favor de  $F_{\theta_1}$ .

Está implícito, sin embargo, en el argumento anterior que la pérdida en que se incurre al dar la estimación  $t_1$  cuando el valor del parámetro es  $\theta_1$ , coincide con la de dar la estimación  $t_1 + c$  cuando el valor paramétrico es  $\theta_1 + c$ . Pues, si la pérdida no se mantuviese constante bajo translaciones a lo largo del eje, la misma información, en distintas posiciones, habría de ser valorada de manera diferente, pudiendo dar lugar a estimaciones que no fuesen una translación de la otra. Así pues, para que el problema de estimación de un

parámetro de posición sea realmente invariante por translaciones, es preciso que la función de pérdida considerada cumpla:

$$L(\theta + c, t + c) = L(\theta, t) \quad \text{para cada } c \in \mathbb{R},$$

lo cual (haciendo  $c = -\theta$ ) es equivalente a

$$L(\theta, t) = l(t - \theta) \quad (\text{siendo } l(t - \theta) = L(0, t - \theta))$$

que indica que la pérdida ha de ser función, únicamente de  $(t - \theta)$ . Desde luego, la pérdida cuadrática:  $L(\theta, t) = (\theta - t)^2$ , verifica tal propiedad.

En la situación así planteada, es importante observar que *los estimadores invariantes por translaciones tienen riesgo constante —que no depende de  $\theta$ .* En efecto, como la distribución de  $X_i + c$  cuando el valor del parámetro es  $\theta$ , coincide<sup>14</sup> con la distribución de  $X_i$  cuando es  $\theta + c$ , será

$$\begin{aligned} R_T(\theta + c) &= E_{\theta+c}[L(\theta + c, T(X_1, \dots, X_n))] \\ &= E_\theta[L(\theta + c, T(X_1 + c, \dots, X_n + c))] \\ &= E_\theta[L(\theta + c, T(X_1, \dots, X_n) + c)] \\ &= E_\theta[L(\theta, T(X_1, \dots, X_n))] = R_T(\theta). \end{aligned}$$

Ello significa que el riesgo de cada estimador, queda caracterizado por una única constante y que, por tanto, tales estimadores resultan totalmente ordenados mediante su riesgo. Es fácil así que exista uno preferible a todos los demás.

Para hallarlo, examinemos primero el caso en que el tamaño muestral sea 1; es decir, aquel en que se observa una única variable aleatoria  $X_1$ . En tal caso, como ha de ser

$$T(x_1 + c) = T(x_1) + c \quad \text{para cada } c \in \mathbb{R},$$

resulta  $T(c) = T(0) + c$ , lo cual significa que cada estimador invariante queda determinado por una constante:  $a = -T(0)$  y es de la forma

$$T(X_1) = X_1 - a.$$

El riesgo, independiente de  $\theta$ , puede ser calculado con  $\theta = 0$  y vale

$$R_T(0) = \int_{\mathbb{R}} l(x_1 - a) F_0(dx_1);$$

---

<sup>14</sup>  $\mathcal{P}_\theta\{X_i + c \leq x\} = \mathcal{P}_\theta\{X_i \leq x - c\} = F_\theta(x - c) = F_0(x - c - \theta) = \mathcal{P}_{\theta+c}\{X_i \leq x\}$

de forma que el mejor estimador invariante corresponde a elegir, como valor de  $a$ , aquel  $a^*$  (supuesto que exista) que minimice la integral anterior. Por ejemplo, con pérdida cuadrática,  $l(u) = u^2$ ,  $a^*$  es la media de la distribución  $F_0$ . En cambio, para  $l(u) = |u|$ ,  $a^*$  es la mediana de  $F_0$ .

En el caso general, con tamaño muestral arbitrario, es preciso disponer de alguna expresión genérica de los estimadores invariantes, antes de proceder a la elección del mejor. Una posibilidad es la siguiente:  $T(x_1, \dots, x_n)$  es invariante si y sólo si

$$T(x_1, \dots, x_n) = x_n - a(x_1 - x_n, \dots, x_{n-1} - x_n)$$

siendo  $a$  una función arbitraria de  $\mathbb{R}^{n-1}$  en  $\mathbb{R}$ . En efecto, está claro que cualquier estimador con la expresión anterior es invariante, pues

$$T(x_1 + c, \dots, x_n + c) = x_n + c - a(x_1 - x_n, \dots, x_{n-1} - x_n) = T(x_1, \dots, x_n) + c.$$

Recíprocamente, si  $T$  es invariante, tomando  $c = -x_n$ , resulta

$$T(x_1 - x_n, \dots, x_{n-1} - x_n, 0) = T(x_1, \dots, x_n) - x_n$$

con lo cual  $T(x_1, \dots, x_n) = x_n + T(x_1 - x_n, \dots, x_{n-1} - x_n, 0)$ .

Así pues, ahora, cada estimador invariante queda determinado por una función arbitraria,  $a$ , de  $n - 1$  variables. Su riesgo —independiente de  $\theta$ — se puede expresar entonces:

$$\begin{aligned} R_T(0) &= \int_{\mathbb{R}^n} l(x_n - a(x_1 - x_n, \dots, x_{n-1} - x_n)) F_0(dx_1) \cdots F_0(dx_n) \\ &= \int_{\mathbb{R}^{n-1}} \int_{\mathbb{R}} l(x_n - a(y_1, \dots, y_{n-1})) F_0(dx_n \mid y_1, \dots, y_{n-1}) G_0(dy_1, \dots, y_{n-1}), \end{aligned}$$

siendo  $G_0$  la distribución de  $(X_1 - X_n, \dots, X_{n-1} - X_n)$  y  $F_0(x \mid y_1, \dots, y_{n-1})$  la distribución de  $X_n$  condicionada por  $X_1 - X_n = y_1, \dots, X_{n-1} - X_n = y_{n-1}$  (calculadas ambas de acuerdo con la distribución  $F_0$ , asociada al valor paramétrico  $\theta = 0$ ).

Puesto que se trata de escoger la función  $a$  de manera a hacer mínimo  $R_T(0)$ , para cada  $(y_1, \dots, y_{n-1})$ , ha de elegirse  $a(y_1, \dots, y_{n-1})$  que minimice

$$\int_{\mathbb{R}} l(x_n - a(y_1, \dots, y_{n-1})) F_0(dx_n \mid y_1, \dots, y_{n-1}).$$

Es decir,  $a(y_1, \dots, y_{n-1})$  ha de ser el valor  $a^*$  al que nos referímos en el caso  $n = 1$ , pero determinado a partir de la distribución  $F_0(x \mid y_1, \dots, y_{n-1})$ , en lugar de  $F_0(x)$ .

En el caso concreto de pérdida cuadrática:  $l(u) = u^2$ , será

$$a^*(y_1, \dots, y_{n-1}) = E_0[X_n | X_1 - X_n = y_1, \dots, X_{n-1} - X_n = y_{n-1}]$$

con lo que el estimador invariante de mínimo error cuadrático medio es

$$T^*(x_1, \dots, x_n) = x_n - E_0[X_n | X_1 - X_n = x_1 - x_n, \dots, X_{n-1} - X_n = x_{n-1} - x_n]$$

que se denomina *estimador de Pitman del parámetro de posición*  $\theta$ .

Por supuesto, la expresión de  $T^*$  no es única; por ejemplo, el papel de  $X_n$  se puede intercambiar con el de cualquier otra  $X_i$ . En el caso en que la población tenga densidad  $f_\theta(x) = f_0(x - \theta)$ , se puede dar una expresión más explícita y simétrica de  $T^*$ . De hecho, la densidad de la muestra cumple

$$f_\theta(x_1, \dots, x_n) = f_0(x_1 - \theta) \cdots f_0(x_n - \theta) = f_0(x_1 - \theta, \dots, x_n - \theta).$$

Además, para  $\theta = 0$ ,  $Y_1 = X_1 - X_n, \dots, Y_{n-1} = X_{n-1} - X_n$  y  $X_n$  tendrán densidad conjunta  $f_0(y_1 + x_n, \dots, y_{n-1} + x_n, x_n)$ ; de forma que la densidad de  $X_n$ , condicionada por  $Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}$ , será

$$\frac{f_0(y_1 + x_n, \dots, y_{n-1} + x_n, x_n)}{\int_{\mathbb{R}} f_0(y_1 + x, \dots, y_{n-1} + x, x) dx}$$

y

$$E_0[X_n | Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}] = \frac{\int_{\mathbb{R}} x f_0(y_1 + x, \dots, y_n + x, x) dx}{\int_{\mathbb{R}} f_0(y_1 + x, \dots, y_n + x, x) dx}.$$

Resulta pues

$$T^*(x_1, \dots, x_n) = x_n - \frac{\int_{\mathbb{R}} x f_0(x_1 - x_n + x, \dots, x_{n-1} - x_n + x, x) dx}{\int_{\mathbb{R}} f_0(x_1 - x_n + x, \dots, x_{n-1} - x_n + x, x) dx}$$

o, haciendo el cambio  $x = x_n - z$ ,

$$T^*(x_1, \dots, x_n) = \frac{\int_{\mathbb{R}} z f_z(x_1, \dots, x_n) dz}{\int_{\mathbb{R}} f_z(x_1, \dots, x_n) dz}.$$

Algunos ejemplos pueden ilustrar la manera de proceder en situaciones concretas.

### Ejemplo 6.13

Sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple de una distribución uniforme en  $(\theta, \theta + k)$ , siendo  $k$  conocido. Para  $\theta = 0$ , la densidad de la muestra es

$$f_0(x_1, \dots, x_n) = \frac{1}{k^n} I_{\{0 < x_{(1)} \leq x_{(n)} < k\}}.$$

Todo radica entonces en calcular la distribución de  $X_n$ , condicionada por  $Y_1 = X_1 - X_n, \dots, Y_{n-1} = X_{n-1} - X_n$ . Ahora bien  $(Y_1, \dots, Y_{n-1}, X_n)$  tiene densidad

$$\frac{1}{k^n} I_{\{0 < x_n + y_{(1)} \leq x_n + y_{(n)} < k\}},$$

donde se ha tomado  $y_n = 0$ ,  $y_{(1)} = \min_{1 \leq i \leq n} y_i$  e  $y_{(n)} = \max_{1 \leq i \leq n} y_i$ , para simplificar la notación. La función de densidad de  $X_n$ , condicionada por  $Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}$  será, pues, uniforme en el intervalo  $(-y_{(1)}, k - y_{(n)})$ .

Para dicha distribución, el punto medio del recorrido:  $\frac{1}{2}(k - y_{(1)} - y_{(n)})$ , constituye la media y la mediana. Por consiguiente, tanto si  $l(u) = u^2$ , como si  $l(u) = |u|$ , el mejor estimador invariante de  $\theta$  es

$$T^* = X_n + \frac{Y_{(1)} + Y_{(n)} - k}{2} = \frac{X_{(1)} + X_{(n)} - k}{2}$$

La fórmula explícita para el estimador de Pitman produce, más rápidamente, el mismo resultado, pero no permite saber que  $T^*$  también es óptimo (entre los estimadores invariantes) cuando la pérdida se mide mediante el valor absoluto.

### Ejemplo 6.14

Sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple de una población con densidad

$$f_\theta(x) = e^{-(x-\theta)} \quad \text{para } x > \theta,$$

para la cual  $\theta$  es evidentemente un parámetro de posición. Si  $\theta = 0$ , la densidad de la muestra es

$$f_0(x_1, \dots, x_n) = e^{-(x_1 + \dots + x_n)} I_{\{x_{(1)} > 0\}}.$$

La densidad de  $(Y_1, \dots, Y_{n-1}, X_n)$  resulta ser

$$e^{-(y_1 + \dots + y_{n-1})} e^{-nx_n} I_{\{x_n > -y_{(1)}\}}$$

tomando, para simplificar,  $y_n = 0$  e  $y_{(1)} = \min_{1 \leq i \leq n} y_i$ . La densidad de  $X_n$ , condicionada por  $Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}$ , es entonces

$$\frac{e^{-nx_n} I_{\{x_n > y_{(1)}\}}}{\int_{-y_{(1)}}^{\infty} e^{-nx_n} dx_n} = ne^{-n(x_n + y_{(1)})} I_{\{x_n > -y_{(1)}\}},$$

cuya media es  $n^{-1} - y_{(1)}$ , mientras que la mediana es  $n^{-1} \log 2 - y_{(1)}$ . Por tanto, el estimador de Pitman es

$$T^* = X_n - \frac{1}{n} + Y_{(1)} = X_{(1)} - \frac{1}{n},$$

que coincide con el ECUMV para  $\theta$ . En cambio, con pérdida  $l(u) = |u|$ , el estimador invariante óptimo es  $X_{(1)} - \log 2/n$ .

En ambos ejemplos ha resultado que *el estimador de Pitman es función del estadístico suficiente minimal*. Se trata de un hecho general, al menos si existe densidad poblacional; ya que si  $S$  es un estadístico suficiente,  $f_\theta(x_1, \dots, x_n) = g_\theta(S(x_1, \dots, x_n))h(x_1, \dots, x_n)$ , de manera que

$$T^*(x_1, \dots, x_n) = \frac{\int_{\mathbb{R}} z g_z(S(x_1, \dots, x_n)) dz}{\int_{\mathbb{R}} g_z(S(x_1, \dots, x_n)) dz}$$

es función de cualquier estadístico suficiente.

Otra propiedad general es que *el estimador de Pitman es siempre insesgado*, como puede apreciarse en los dos ejemplos previos. La razón es que el sesgo de cualquier estimador invariante es constante:

$$\begin{aligned} b_T(\theta + c) &= E_{\theta+c}[T(X_1, \dots, X_n)] - \theta - c \\ &= E_\theta[T(X_1 + c, \dots, X_n + c)] - \theta - c \\ &= E_\theta[T(X_1, \dots, X_n) + c] - \theta - c = b_T(\theta); \end{aligned}$$

de forma que, si  $T^*$  tuviese sesgo  $b_{T^*}$ , sería

$$E[(T^* - \theta)^2] = E[(T^* - b_{T^*} - \theta + b_{T^*})^2] = E[(T^* - b_{T^*} - \theta)^2] + b_{T^*}^2$$

y  $T^* - b_{T^*}$  sería un estimador invariante con menor error cuadrático medio que  $T^*$ .

Como consecuencia de la propiedad anterior: *Si el ECUMV de un parámetro de posición es invariante, coincide con el estimador de Pitman.* Ya que el estimador de Pitman (por ser centrado) tendría más varianza que el ECUMV; éste a su vez (por ser invariante) tendría más varianza que aquél. Tienen pues la misma varianza, ambos son ECUMV y, por tanto, coinciden.

Por ejemplo, en el caso de una población normal de varianza conocida, esta observación permite concluir, sin ningún cálculo, que  $\bar{X}$  es el estimador de Pitman de la media poblacional.

#### 6.4.2. Parámetros de escala

Consideremos ahora una familia de distribuciones  $\{F_\theta \mid \theta > 0\}$ , que depende sólo de un parámetro de escala  $\theta$ ; es decir  $F_\theta(x) = F_1(x/\theta)$ , siendo  $F_1$  una distribución unidimensional dada. Supondremos que  $F_1$  tiene densidad  $f_1$ , de manera que se cumplirá

$$f_\theta(x) = \frac{1}{\theta} f_1\left(\frac{x}{\theta}\right).$$

Más en general, la densidad de una muestra aleatoria simple, de tamaño  $n$ , verificará

$$f_\theta(x_1, \dots, x_n) = \frac{1}{\theta^n} f_1\left(\frac{x_1}{\theta}, \dots, \frac{x_n}{\theta}\right)$$

y los resultados de este apartado serán aplicables, aunque la muestra no sea aleatoria simple, siempre que su densidad verifique esta condición.

Para la estimación del parámetro de escala  $\theta$ , la búsqueda de estimadores invariantes por cambios de escala, tales que

$$T(bx_1, \dots, bx_n) = bT(x_1, \dots, x_n) \quad \text{para cada } b > 0,$$

quedará plenamente justificada si se adopta una función de pérdida que no se altere al cambiar las unidades de medida; es decir que verifique la condición:

$$L(b\theta, bt) = L(\theta, t) \quad \text{para cada } b > 0;$$

lo cual puede expresarse (con  $b = 1/\theta$ ) en la forma equivalente

$$L(\theta, t) = l(t/\theta) \quad \text{siendo } l(t/\theta) = L(1, t/\theta).$$

El ejemplo más típico de tales pérdidas es el cuadrado del error relativo de estimación:  $L(\theta, t) = \left(\frac{\theta - t}{\theta}\right)^2 = (1 - t/\theta)^2$ .

En tales condiciones, el riesgo de los estimadores invariantes por cambios de escala no depende de  $\theta$ ; pues<sup>15</sup>

$$\begin{aligned} R_T(\theta) &= \text{E}_\theta [L(\theta, T(X_1, \dots, X_n))] = \text{E}_1 [L(\theta, T(\theta X_1, \dots, \theta X_n))] \\ &= \text{E}_1 [L(\theta, \theta T(X_1, \dots, X_n))] = \text{E}_1 [L(1, T(X_1, \dots, X_n))] = R_T(1). \end{aligned}$$

Por otra parte, un estimador  $T$  es invariante por cambios de escala si y sólo si

$$T(x_1, \dots, x_n) = \begin{cases} x_n \varphi(x_1 x_n^{-1}, \dots, x_{n-1} x_n^{-1}) & \text{si } x_n > 0, \\ -x_n \psi(x_1 x_n^{-1}, \dots, x_{n-1} x_n^{-1}) & \text{si } x_n < 0, \end{cases}$$

siendo  $\varphi$  y  $\psi$  funciones arbitrarias de  $\mathbb{R}^{n-1}$  en  $\mathbb{R}$ . En efecto, si  $T$  es invariante y  $x_n > 0$ ,

$$T(x_1, \dots, x_n) = x_n T(x_1 x_n^{-1}, \dots, x_{n-1} x_n^{-1}, 1)$$

y basta definir  $\varphi(y_1, \dots, y_{n-1}) = T(y_1, \dots, y_{n-1}, 1)$ . De manera análoga, si  $x_n < 0$ ,

$$T(x_1, \dots, x_n) = -x_n T(-x_1 x_n^{-1}, \dots, -x_{n-1} x_n^{-1}, -1)$$

y basta hacer  $\psi(y_1, \dots, y_{n-1}) = T(-y_1, \dots, -y_{n-1}, -1)$ . El recíproco es inmediato.

Con tal expresión de  $T$ , su riesgo —independiente de  $\theta$ — adopta la forma

$$\begin{aligned} R_T(1) &= \int_{x_n > 0} l(x_n \varphi(x_1 x_n^{-1}, \dots, x_{n-1} x_n^{-1})) f_1(x_1, \dots, x_n) dx_1 \dots dx_n \\ &\quad + \int_{x_n < 0} l(-x_n \psi(x_1 x_n^{-1}, \dots, x_{n-1} x_n^{-1})) f_1(x_1, \dots, x_n) dx_1 \dots dx_n. \end{aligned}$$

Si  $g_1$  es la densidad de  $(X_1 X_n^{-1}, \dots, X_{n-1} X_n^{-1})$  y  $h_1(x \mid y_1, \dots, y_{n-1})$  es la densidad de  $X_n$ , condicionada por  $X_1 X_n^{-1} = y_1, \dots, X_{n-1} X_n^{-1} = y_{n-1}$  (correspondientes ambas a  $\theta = 1$ ), resulta que  $R_T(1)$  se obtiene integrando, respecto a  $g_1$ , la función:

$$\begin{aligned} &\int_0^\infty l(x_n \varphi(y_1, \dots, y_{n-1})) h_1(x_n \mid y_1, \dots, y_{n-1}) dx_n + \\ &\quad + \int_{-\infty}^0 l(-x_n \psi(y_1, \dots, y_{n-1})) h_1(x_n \mid y_1, \dots, y_{n-1}) dx_n. \end{aligned}$$

---

<sup>15</sup> dado que  $\mathcal{P}_\theta\{X_i \leq x\} = F_\theta(x) = F_1(x/\theta) = \mathcal{P}_1\{\theta X_i \leq x\}$ .

Por consiguiente, la elección de  $\varphi$  y  $\psi$  ha de hacerse de manera a minimizar dicha función, para cada  $(y_1, \dots, y_{n-1})$  fijos.

En el caso de la pérdida  $l(u) = (1 - u)^2$ , el problema se reduce, dada cualquier densidad  $h$  sobre  $\mathbb{R}$ , a determinar las constantes  $\varphi$  y  $\psi$  que minimizan

$$\begin{aligned} & \int_0^\infty (1 - x\varphi)^2 h(x) dx + \int_{-\infty}^0 (1 + x\psi)^2 h(x) dx = \\ & = 1 + \varphi^2 \int_0^\infty x^2 h(x) dx - 2\varphi \int_0^\infty x h(x) dx + \psi^2 \int_{-\infty}^0 x^2 h(x) dx + 2\psi \int_{-\infty}^0 x h(x) dx, \end{aligned}$$

lo que se consigue con

$$\varphi = \frac{\int_0^\infty x h(x) dx}{\int_0^\infty x^2 h(x) dx} \quad \text{y} \quad \psi = -\frac{\int_{-\infty}^0 x h(x) dx}{\int_{-\infty}^0 x^2 h(x) dx}.$$

Resulta entonces, como estimador invariante por cambios de escala de mínimo riesgo (para la pérdida  $L(\theta, t) = (1 - t/\theta)^2$ ):

$$T^*(x_1, \dots, x_n) = \begin{cases} x_n \frac{\int_0^\infty x h_1(x | x_1 x_n^{-1}, \dots, x_{n-1} x_n^{-1}) dx}{\int_0^\infty x^2 h_1(x | x_1 x_n^{-1}, \dots, x_{n-1} x_n^{-1}) dx} & \text{si } x_n > 0 \\ x_n \frac{\int_{-\infty}^0 x h_1(x | x_1 x_n^{-1}, \dots, x_{n-1} x_n^{-1}) dx}{\int_{-\infty}^0 x^2 h_1(x | x_1 x_n^{-1}, \dots, x_{n-1} x_n^{-1}) dx} & \text{si } x_n < 0 \end{cases}$$

que se denomina *estimador de Pitman del parámetro de escala  $\theta$* .

Se puede obtener una expresión más explícita del estimador anterior, si se observa que  $Y_1 = X_1 X_n^{-1}, \dots, Y_{n-1} = X_{n-1} X_n^{-1}$  y  $X_n$  tienen como densidad conjunta (para  $\theta = 1$ ):  $f_1(y_1 x_n, \dots, y_{n-1} x_n, x_n) |x_n|^{n-1}$ . Luego, si  $x_n > 0$ , es

$$T^*(x_1, \dots, x_n) = x_n \frac{\int_0^\infty x^n f_1(x_1 x_n^{-1} x, \dots, x_{n-1} x_n^{-1} x, x) dx}{\int_0^\infty x^{n+1} f_1(x_1 x_n^{-1} x, \dots, x_{n-1} x_n^{-1} x, x) dx}$$

que, con el cambio  $x = x_n/u$ , se expresa

$$T^*(x_1, \dots, x_n) = \frac{\int_0^\infty u^{-2} f_u(x_1, \dots, x_n) du}{\int_0^\infty u^{-3} f_u(x_1, \dots, x_n) du}.$$

Exactamente el mismo resultado se obtiene para  $x_n < 0$ .

### Ejemplo 6.15

Supongamos que se desea estimar, de forma invariante por cambios de unidades, el parámetro  $\theta$  de una distribución uniforme en  $(0, \theta)$ , a partir de una muestra aleatoria simple  $(X_1, \dots, X_n)$ .

Para  $\theta = 1$ , la densidad de la muestra es

$$f_1(x_1, \dots, x_n) = \prod_{i=1}^n I_{\{0 < x_i < 1\}},$$

de forma que la densidad conjunta de  $Y_1 = X_1 X_n^{-1}, \dots, Y_{n-1} = X_{n-1} X_n^{-1}$  y  $X_n$  es

$$x_n^{n-1} I_{\{0 < x_n < y_{(n)}^{-1}\}} \prod_{i=1}^{n-1} I_{\{y_i > 0\}}$$

(tomando  $y_n = 1$  e  $y_{(n)} = \max_{1 \leq i \leq n} y_i$ ). La densidad de  $Y_1, \dots, Y_{n-1}$  resulta entonces

$$\frac{1}{n y_{(n)}^n} \prod_{i=1}^{n-1} I_{\{y_i > 0\}}$$

y la densidad de  $X_n$ , condicionada por  $Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}$ ,

$$n y_{(n)}^n x_n^{n-1} I_{\{0 < x_n < y_{(n)}^{-1}\}},$$

supuesto que  $y_1, \dots, y_{n-1} > 0$ . Como

$$\int_0^\infty x^k n y_{(n)}^n x^{n-1} I_{\{0 < x < y_{(n)}^{-1}\}} dx = \frac{n}{n+k} y_{(n)}^{-k},$$

el estimador de Pitman de  $\theta$  es

$$T^*(X_1, \dots, X_n) = X_n \frac{n+2}{n+1} Y_{(n)} = \frac{n+2}{n+1} X_{(n)}.$$

Recuérdese (cf. ejemplo 6.2) que el ECUMV para  $\theta$  es  $\frac{n+1}{n}X_{(n)}$ , de manera que el estimador de Pitman para parámetros de escala no es, en general, insesgado. El estimador  $T^*$  había sido hallado, en el ejemplo 5.1, buscando el estimador de mínimo riesgo entre los múltiplos de  $X_{(n)}$ ; ahora sabemos que es óptimo (con la pérdida considerada), no sólo entre estos, sino entre todos los estimadores invariantes por cambios de escala.

En caso de que la pérdida  $l(t/\theta)$ , en vez de ser  $(1-t/\theta)^2$ , fuese  $|1-t/\theta|$ , la determinación del estimador invariante de mínimo riesgo exigiría hallar el valor de  $\varphi$  que minimiza

$$\begin{aligned} \int_0^\infty |1-x\varphi| n y_{(n)}^n x^{n-1} I_{\{0 < x < y_{(n)}^{-1}\}} dx &= \\ &= ny_{(n)}^n \left[ \int_0^{1/\varphi} (1-x\varphi)x^{n-1} dx + \int_{1/\varphi}^{1/y_{(n)}} (x\varphi - 1)x^{n-1} dx \right] \\ &= \frac{2 y_{(n)}^n}{(n+1)\varphi^n} + \frac{n\varphi}{(n+1)y_{(n)}} - 1, \end{aligned}$$

lo que se consigue con  $\varphi = 2^{1/(n+1)}y_{(n)}$ ; de manera que el estimador invariante de mínimo riesgo es ahora

$$T^*(X_1, \dots, X_n) = X_n 2^{1/(n+1)} Y_{(n)} = 2^{1/(n+1)} X_{(n)}.$$

### Ejemplo 6.16

Para una distribución exponencial de densidad

$$f_\theta(x) = \frac{1}{\theta} e^{-x/\theta} \quad \text{para } x > 0,$$

su media,  $\theta$ , es un parámetro de escala. La aplicación de la fórmula del estimador de Pitman requiere calcular

$$\begin{aligned} \int_0^\infty u^{-k} \frac{1}{u} e^{-x_1/u} \cdots \frac{1}{u} e^{-x_n/u} du &= \int_0^\infty \frac{1}{u^{n+k}} e^{-(x_1+\cdots+x_n)/u} du = \\ &= \int_0^\infty v^{n+k-2} e^{-v(x_1+\cdots+x_n)} dv = \frac{(n+k-2)!}{(\sum_{i=1}^n x_i)^{n+k-1}}. \end{aligned}$$

De donde se obtiene

$$T^*(X_1, \dots, X_n) = \frac{n!}{(\sum_{i=1}^n X_i)^{n+1}} \frac{(\sum_{i=1}^n X_i)^{n+2}}{(n+1)!} = \frac{1}{n+1} \sum_{i=1}^n X_i$$

que evidentemente no es un estimador centrado.

### Ejemplo 6.17

En el caso de una distribución  $N(0, \sigma)$ , la desviación típica  $\sigma$  es un parámetro de escala. Como

$$\begin{aligned} I_k &= \int_0^\infty \frac{1}{(2\pi)^{n/2}\sigma^k} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2\right\} d\sigma \\ &= \frac{1}{(2\pi)^{n/2}} \int_0^\infty v^{k/2} \exp\left\{-\frac{v}{2} \sum_{i=1}^n x_i^2\right\} \frac{dv}{2v^{3/2}} \\ &= \frac{\Gamma(\frac{k-1}{2})}{2(2\pi)^{n/2}} \left(\frac{1}{2} \sum_{i=1}^n x_i^2\right)^{-(k-1)/2}, \end{aligned}$$

el estimador de Pitman de  $\sigma$  es

$$T^*(X_1, \dots, X_n) = \frac{I_{n+2}}{I_{n+3}} = \frac{\sqrt{n} \Gamma(\frac{n+1}{2})}{\sqrt{2} \Gamma(\frac{n+2}{2})} \left(\frac{1}{n} \sum_{i=1}^n X_i^2\right)^{1/2}$$

que recuerda la expresión del ECUMV para  $\sigma$ , obtenido en el ejemplo 6.8.

Al igual que en el caso de los parámetros de posición, un razonamiento similar establece que *el estimador de Pitman de un parámetro de escala es función del estadístico minimal suficiente*.

#### 6.4.3. Parámetros de posición y escala

Hay que analizar, en último lugar, el caso en que la distribución de la población depende a la vez de un parámetro de posición y otro de escala; esto es

$$F_{\theta_1, \theta_2}(x) = F\left(\frac{x - \theta_1}{\theta_2}\right) \quad \text{con } \theta_1 \in \mathbb{R} \text{ y } \theta_2 > 0,$$

siendo  $F$  una distribución fija. Supondremos que  $F$  tiene densidad  $f$ ; de manera que se verifica

$$f_{\theta_1, \theta_2}(x) = \frac{1}{\theta_2} f\left(\frac{x - \theta_1}{\theta_2}\right),$$

lo cual hace que la densidad de una muestra aleatoria simple, de tamaño  $n$ , sea de la forma

$$f_{\theta_1, \theta_2}(x_1, \dots, x_n) = \frac{1}{\theta_2^n} f\left(\frac{x_1 - \theta_1}{\theta_2}, \dots, \frac{x_n - \theta_1}{\theta_2}\right).$$

En la estimación del parámetro de escala,  $\theta_2$ , es lógico (cf. §5.3) emplear estimadores  $T_2$  que satisfagan

$$T_2(c + bx_1, \dots, c + bx_n) = bT_2(x_1, \dots, x_n) \quad \text{para cada } c \in \mathbb{R} \text{ y } b > 0,$$

supuesto que la función de pérdida es de la forma:  $L(\theta_1, \theta_2, t_2) = l(t_2/\theta_2)$ . Seguiremos considerando, pues,  $(1 - t_2/\theta_2)^2$  como pérdida más representativa, en la estimación de  $\theta_2$ .

La condición de invariancia de  $T_2$  indica, en particular, que ha de ser  $T_2(c + x_1, \dots, c + x_n) = T_2(x_1, \dots, x_n)$  para cada  $c \in \mathbb{R}$ ; así que  $T_2$  se puede expresar en función de las diferencias  $y_1 = x_1 - x_n, \dots, y_{n-1} = x_{n-1} - x_n$  (sin más que hacer  $c = -x_n$ ). Ahora bien, como la densidad conjunta de  $Y_1 = X_1 - X_n, \dots, Y_{n-1} = X_{n-1} - X_n$  es

$$\begin{aligned} g_{\theta_2}(y_1, \dots, y_{n-1}) &= \int_{-\infty}^{\infty} f_{\theta_1, \theta_2}(y_1 + x_n, \dots, y_{n-1} + x_n, x_n) dx_n \\ &= \int_{-\infty}^{\infty} f_{0, \theta_2}(y_1 + x, \dots, y_{n-1} + x, x) dx, \end{aligned} \quad (6.1)$$

(haciendo el cambio  $x = x_n - \theta_1$ ) resulta que  $g_{\theta_2}$  no depende de  $\theta_1$  y verifica la condición

$$g_{\theta_2}(y_1, \dots, y_{n-1}) = \frac{1}{\theta_2^{n-1}} g_1 \left( \frac{y_1}{\theta_2}, \dots, \frac{y_{n-1}}{\theta_2} \right),$$

que constituía el punto de partida de los resultados del apartado anterior.

Así pues, se puede utilizar el resultado obtenido allí, para la muestra (no aleatoria simple)  $Y_1, \dots, Y_{n-1}$ , y concluir que *el estimador invariante por cambios lineales, de mínimo riesgo en la estimación del parámetro de escala  $\theta_2$  (con pérdida  $(1 - t_2/\theta_2)^2$ )* es

$$T_2^*(x_1, \dots, x_n) = \frac{\int_0^{\infty} v^{-2} g_v(x_1 - x_n, \dots, x_{n-1} - x_n) dv}{\int_0^{\infty} v^{-3} g_v(x_1 - x_n, \dots, x_{n-1} - x_n) dv}.$$

Sustituyendo  $g_v$  mediante (6.1) y haciendo el cambio  $u = x_n - x$ , se obtiene

$$T_2^*(x_1, \dots, x_n) = \frac{\int_0^{\infty} \int_{-\infty}^{\infty} v^{-2} f_{u,v}(x_1, \dots, x_n) du dv}{\int_0^{\infty} \int_{-\infty}^{\infty} v^{-3} f_{u,v}(x_1, \dots, x_n) du dv}$$

como expresión explícita del *estimador de Pitman del parámetro de escala*  $\theta_2$ , en presencia del parámetro de posición  $\theta_1$ .

Para la estimación del parámetro de posición  $\theta_1$ , debemos buscar estimadores  $T_1$  que verifiquen

$$T_1(c + bx_1, \dots, c + bx_n) = c + bT_1(x_1, \dots, x_n) \quad \text{para cada } c \in \mathbb{R} \text{ y } b > 0.$$

Ha de suponerse, ahora, que la pérdida es de la forma

$$L(\theta_1, \theta_2, t_1) = l((t_1 - \theta_1)/\theta_2)$$

cuyo ejemplo más típico es  $\left(\frac{t_1 - \theta_1}{\theta_2}\right)^2$ .

Como ya viene siendo habitual, el considerar estimadores y pérdidas invariantes tiene por efecto que el riesgo no dependa del parámetro:

$$\begin{aligned} R_{T_1}(\theta_1, \theta_2) &= E_{\theta_1, \theta_2}[L(\theta_1, \theta_2, T_1(X_1, \dots, X_n))] \\ &= E_{0,1}[L(\theta_1, \theta_2, T_1(\theta_1 + \theta_2 X_1, \dots, \theta_1 + \theta_2 X_n))] \\ &= E_{0,1}[L(\theta_1, \theta_2, \theta_1 + \theta_2 T_1(X_1, \dots, X_n))] \\ &= E_{0,1}[L(0, 1, T_1(X_1, \dots, X_n))] = R_{T_1}(0, 1), \end{aligned}$$

lo cual permite fijar, en la determinación del riesgo,  $\theta_1 = 0$  y  $\theta_2 = 1$ .

Ahora bien, los estimadores  $T_1$  con la condición de invariancia requerida, deben satisfacer (haciendo  $c = -bx_n$ ):

$$T_1(x_1, \dots, x_n) = x_n + \frac{1}{b} T_1(b(x_1 - x_n), \dots, b(x_{n-1} - x_n), 0)$$

y, por tanto, serán de la forma<sup>16</sup>:

$$T_1(x_1, \dots, x_n) = \begin{cases} x_n + (x_{n-1} - x_n)\varphi(z_1, \dots, z_{n-2}) & \text{si } x_{n-1} > x_n, \\ x_n + (x_n - x_{n-1})\psi(z_1, \dots, z_{n-2}) & \text{si } x_{n-1} < x_n, \end{cases}$$

donde  $z_i = (x_i - x_n)/(x_{n-1} - x_n)$  para  $i = 1, 2, \dots, n-2$ .

Su riesgo —que no depende de los parámetros— se expresa

$$\begin{aligned} &\int_{x_{n-1} > x_n} l(x_n + (x_{n-1} - x_n)\varphi(z_1, \dots, z_{n-2})) f(x_1, \dots, x_n) dx_1 \dots dx_n \\ &+ \int_{x_{n-1} < x_n} l(x_n + (x_n - x_{n-1})\psi(z_1, \dots, z_{n-2})) f(x_1, \dots, x_n) dx_1 \dots dx_n. \end{aligned}$$

---

<sup>16</sup> que se obtiene tomando  $b = 1/|x_{n-1} - x_n|$ ,  $\varphi(z_1, \dots, z_{n-2}) = T_1(z_1, \dots, z_{n-2}, 1, 0)$  y  $\psi(z_1, \dots, z_{n-2}) = T_1(-z_1, \dots, -z_{n-2}, -1, 0)$ .

Si  $g$  es la densidad de  $(Z_1, \dots, Z_{n-2})$  y  $h(x, x' | z_1, \dots, z_{n-2})$  es la densidad de  $(X_{n-1}, X_n)$ , condicionada por  $Z_1 = z_1, \dots, Z_{n-2} = z_{n-2}$ , el riesgo  $R_{T_1}(0, 1)$  se obtiene integrando, respecto a  $g$ , la función

$$\int_{x>x'} l(x' + (x - x')\varphi(z_1, \dots, z_{n-2})) h(x, x' | z_1, \dots, z_{n-2}) dx dx' \\ + \int_{x< x'} l(x' + (x' - x)\psi(z_1, \dots, z_{n-2})) h(x, x' | z_1, \dots, z_{n-2}) dx dx'.$$

De forma que la determinación del estimador  $T_1$  de mínimo riesgo exige elegir, para cada  $z_1, \dots, z_{n-2}$  fijos,  $\varphi$  y  $\psi$  que minimicen dicha función.

Con pérdida cuadrática,  $l(u) = u^2$ , la cuestión estriba en hallar, para cualquier densidad  $h$  en  $\mathbb{R}^2$ ,  $\varphi$  y  $\psi$  que hagan mínimo

$$\int_{x>x'} [x' + (x - x')\varphi]^2 h(x, x') dx dx' + \int_{x< x'} [x' + (x' - x)\psi]^2 h(x, x') dx dx'.$$

La solución es inmediata; debe ser

$$\left\{ \begin{array}{l} \varphi = \frac{\int_{x>x'} x'(x' - x)h(x, x') dx dx'}{\int_{x>x'} (x' - x)^2 h(x, x') dx dx'}, \\ \psi = \frac{\int_{x< x'} x'(x - x')h(x, x') dx dx'}{\int_{x< x'} (x - x')^2 h(x, x') dx dx'}. \end{array} \right.$$

Por consiguiente, si  $x_{n-1} > x_n$ , el estimador de  $\theta_1$ , invariante por cambios lineales, con mínimo riesgo (asociado a la pérdida  $(t_1 - \theta_1)^2/\theta_2^2$ ) es

$$T_1^*(x_1, \dots, x_n) = x_n + (x_{n-1} - x_n) \frac{\int_{x>x'} x'(x' - x)h(x, x' | z_1, \dots, z_{n-2}) dx dx'}{\int_{x>x'} (x - x')^2 h(x, x' | z_1, \dots, z_{n-2}) dx dx'}.$$

Como, para  $\theta_1 = 0$  y  $\theta_2 = 1$ , la densidad conjunta de  $Z_1, \dots, Z_{n-2}, X_{n-1}, X_n$  es

$$f(x_n + z_1(x_{n-1} - x_n), \dots, x_n + z_{n-2}(x_{n-1} - x_n), x_{n-1}, x_n) |x_{n-1} - x_n|^{n-2},$$

el último factor en  $T_1^*$  se expresa

$$-\frac{\int_{x>x'} x'(x-x')^{n-1} f(x'+z_1(x-x'), \dots, x'+z_{n-2}(x-x'), x, x') dx dx'}{\int_{x>x'} (x-x')^n f(x'+z_1(x-x'), \dots, x'+z_{n-2}(x-x'), x, x') dx dx'}$$

o bien, haciendo el cambio  $x = (x_{n-1} - u)v^{-1}$ ,  $x' = (x_n - u)v^{-1}$ ,

$$-\frac{1}{x_{n-1} - x_n} \frac{\int_0^\infty \int_{-\infty}^\infty (x_n - u)v^{-3} f_{u,v}(x_1, \dots, x_n) du dv}{\int_0^\infty \int_{-\infty}^\infty v^{-3} f_{u,v}(x_1, \dots, x_n) du dv};$$

con lo cual resulta

$$T_1^*(x_1, \dots, x_n) = \frac{\int_0^\infty \int_{-\infty}^\infty uv^{-3} f_{u,v}(x_1, \dots, x_n) du dv}{\int_0^\infty \int_{-\infty}^\infty v^{-3} f_{u,v}(x_1, \dots, x_n) du dv}.$$

El resultado es exactamente el mismo en el caso  $x_{n-1} < x_n$  y supone, pues, una expresión explícita para el estimador de Pitman del parámetro de posición  $\theta_1$ , en presencia del parámetro de escala  $\theta_2$ .

Como en los apartados anteriores, resulta claro que  $T_1^*$  y  $T_2^*$  son siempre funciones del estadístico minimal suficiente. Ello puede verse en los ejemplos de aplicación siguientes.

### Ejemplo 6.18

Para una distribución uniforme en  $(\theta_1, \theta_1 + \theta_2)$ ,  $\theta_1$  es un parámetro de posición y  $\theta_2$  un parámetro de escala. La densidad de una muestra aleatoria simple, de tamaño  $n$ , es

$$f_{\theta_1, \theta_2}(x_1, \dots, x_n) = \frac{1}{\theta_2^n} I_{\{\theta_1 < x_{(1)} < x_{(n)} < \theta_1 + \theta_2\}}.$$

Entonces

$$\begin{aligned} \int_0^\infty \int_{-\infty}^\infty v^{-k} I_{\{u < x_{(1)} < x_{(n)} < u+v\}} du dv &= \int_{x_{(n)} - x_{(1)}}^\infty v^{-k} \int_{x_{(n)} - v}^{x_{(1)}} du dv \\ &= \frac{1}{(k-1)(k-2)(x_{(n)} - x_{(1)})^{k-2}}. \end{aligned}$$

Luego, con función de pérdida  $(1-t_2/\theta_2)^2$ , el estimador invariante de mínimo riesgo para  $\theta_2$  es

$$T_2^*(X_1, \dots, X_n) = \frac{n+2}{n}(X_{(n)} - X_{(1)})$$

Además,

$$\begin{aligned} \int_0^\infty \int_{-\infty}^\infty uv^{-n-3} I_{\{u < x_{(1)} < x_{(n)} < u+v\}} du dv &= \int_{x_{(n)}-x_{(1)}}^\infty v^{-n-3} \int_{x_{(n)}-v}^{x_{(1)}} u du dv \\ &= \frac{(n+1)x_{(1)} - x_{(n)}}{(n+2)(n+1)n(x_{(n)} - x_{(1)})^{n+1}}. \end{aligned}$$

Luego, con función de pérdida  $(t_1 - \theta_1)^2 / \theta_2^2$ , el estimador invariante de mínimo riesgo para  $\theta_1$  es

$$T_1^*(X_1, \dots, X_n) = \frac{(n+1)X_{(1)} - X_{(n)}}{n}.$$

### Ejemplo 6.19

Para la distribución de densidad

$$f_{\theta_1, \theta_2}(x) = \frac{1}{\theta_2} e^{-(x-\theta_1)/\theta_2} I_{\{x > \theta_1\}},$$

$\theta_1$  es un parámetro de posición y  $\theta_2$  un parámetro de escala. Una muestra aleatoria simple, de tamaño  $n$ , tiene densidad

$$f_{\theta_1, \theta_2}(x_1, \dots, x_n) = \frac{1}{\theta_2^n} e^{-n(\bar{x}-\theta_1)/\theta_2} I_{\{\theta_1 < x_{(1)}\}}$$

Entonces

$$\begin{aligned} \int_0^\infty \int_{-\infty}^{x_{(1)}} v^{-n-k} e^{-n(\bar{x}-u)/v} du dv &= \int_0^\infty v^{-n-k} \frac{v}{n} e^{-n(\bar{x}-x_{(1)})/v} dv \\ &= \frac{1}{n} \int_0^\infty w^{n+k-3} e^{-n(\bar{x}-x_{(1)})w} dw = \frac{1}{n^{n+k-1}} \frac{(n+k-3)!}{(\bar{x}-x_{(1)})^{n+k-2}}. \end{aligned}$$

Luego

$$T_2^*(X_1, \dots, X_n) = \bar{X} - X_{(1)}.$$

Además,

$$\int_0^\infty \int_{-\infty}^{x_{(1)}} uv^{-n-3} e^{-n(\bar{x}-u)/v} du dv = \int_0^\infty v^{-n-3} \left( \frac{v}{n} x_{(1)} - \frac{v^2}{n^2} \right) e^{-n(\bar{x}-x_{(1)})/v} dv$$

$$= \frac{x_{(1)}}{n^{n+2}} \frac{n!}{(\bar{x} - x_{(1)})^{n+1}} - \frac{1}{n^{n+2}} \frac{(n-1)!}{(\bar{x} - x_{(1)})^n};$$

con lo cual

$$T_1^*(x_1, \dots, x_n) = X_{(1)} - \frac{1}{n}(\bar{X} - X_{(1)}) = \frac{n+1}{n}X_{(1)} - \frac{\bar{X}}{n}.$$

## Ejercicios

### Ejercicio 6.1

Se dispone de una muestra aleatoria simple de tamaño  $n$  de la densidad

$$f_\theta(x) = \frac{\theta}{(1+x)^{1+\theta}} \quad \text{para } x > 0,$$

siendo  $\theta > 0$ .

- a) Determinar un estadístico suficiente y completo.
- b) Determinar el ECUMV de  $\theta$  y de  $1/\theta$ .
- c) Determinar la cota de Fréchet-Cramer-Rao para estimadores insesgados de  $1/\theta$  y analizar si su ECUMV es eficiente.

### Ejercicio 6.2

Se lanza  $n$  veces una moneda de probabilidad de cara desconocida,  $\theta$ , observando en cada lanzamiento el resultado obtenido.

- a) Determinar, si existe, el ECUMV para cada una de las siguientes funciones del parámetro:  $g_1(\theta) = \theta$ ,  $g_2(\theta) = \theta(1-\theta)$ ,  $g_3(\theta) = \theta^m$  con  $m \leq n$ ,  $g_4(\theta) = 1/\theta$ .
- b) Determinar la forma que debe tener una función  $g(\theta)$  para que exista su ECUMV. Expresar el ECUMV de tales funciones.
- c) Determinar la cota de Fréchet-Cramer-Rao para estimadores insesgados de  $\theta$ . Hallar un estimador eficiente para  $\theta$ .

### Ejercicio 6.3

En la fabricación de cable eléctrico, el número,  $X$ , de fallos por kilómetro producido sigue una distribución de Poisson de parámetro  $\theta$  desconocido. Se controla el número de fallos existentes en  $n$  kilómetros.

- a) Determinar, si existe, el ECUMV para cada una de las siguientes funciones del parámetro:  $g_1(\theta) = \theta^r$  con  $r$  natural,  $g_2(\theta) = P_\theta\{X = r\}$  con  $r$  natural,  $g_3(\theta) = 1/\theta$ .
- b) Determinar la forma que debe tener una función  $g(\theta)$  para que tenga ECUMV. Expresar el ECUMV de tales funciones.
- c) Si no se tienen en cuenta las observaciones muestrales en las cuales no aparezcan fallos y se prosigue la observación hasta disponer de  $n$  observaciones  $X_i > 0$ , hallar la función de  $\theta$  para la cual existe un estimador eficiente.

#### Ejercicio 6.4

Se han realizado  $n$  observaciones independientes de una población con función de distribución

$$F_\theta(x) = \left(\frac{x}{\theta}\right)^k \quad \text{para } x \in (0, \theta),$$

donde  $k \geq 1$  es una constante conocida y  $\theta$  es un parámetro desconocido.

- a) Hallar el ECUMV de  $\theta$ .
- b) Probar que cualquier función derivable,  $g(\theta)$ , tiene ECUMV y determinarlo.
- c) Obtener el estimador de Pitman de  $\theta$ .

#### Ejercicio 6.5

De una variable aleatoria con función de probabilidad

$$f_\theta(x) = \begin{cases} (1-\theta)^2\theta^x & \text{para } x = 0, 1, 2, \dots \\ \theta & \text{para } x = -1, \end{cases}$$

siendo  $\theta \in [0, 1]$  desconocido, se obtiene una muestra aleatoria simple de tamaño  $n$ .

- a) Determinar un estadístico suficiente minimal y analizar si es completo.
- b) Determinar diversos estimadores centrados de  $\theta$ , función del estadístico suficiente minimal y comparar sus varianzas.

- c) Supuesto que  $n = 1$ , hallar todas las funciones de  $X_1$  cuya esperanza es nula. Deducir que no existe ECUMV para  $\theta$ , pero que  $(1 - \theta)^2$  si tiene ECUMV.

### Ejercicio 6.6

$n$  números han sido elegidos al azar e independientemente entre  $a$  y  $b$  ( $0 < a < b$ ), pero se ha perdido la información acerca del valor de  $a$  y de  $b$ .

- Comprobar que el estadístico suficiente minimal es completo.
- Hallar el ECUMV de  $a$ , de  $b$  y de  $b - a$ .
- De pronto se recuerda que  $b = 2a > 0$ . Comprobar que  $a$  es entonces un parámetro de escala y obtener el estimador de Pitman.

### Ejercicio 6.7

Una muestra aleatoria simple de tamaño  $n$  proviene de una población con distribución  $\gamma(p, \theta)$  en la que ambos parámetros son desconocidos.

- Obtener la cota para la varianza de los estimadores centrados de  $p$ , para estimadores centrados de  $\theta$  y la cota para la varianza generalizada de pares de estimadores centrados de  $p$  y  $\theta$ .
- Comprobar que  $\bar{X}$  y  $n^{-1} \sum_{i=1}^n \log X_i$  son estimadores eficientes de sus medias.
- Razonar que ambos estimadores constituyen un estadístico completo, y son, por tanto ECUMV de sus medias.

### Ejercicio 6.8

Sea  $X_1, \dots, X_n$  una muestra de una distribución con densidad

$$f_\theta(x) = \sqrt{\frac{2}{\pi}} e^{-(x-\theta)^2/2} \quad \text{para } x > \theta.$$

Determinar el estimador de Pitman del parámetro de posición  $\theta$ .

### Ejercicio 6.9

En una población  $N(0, \sigma^2)$  se desea estimar  $\sigma^2$ , mediante un estimador que verifique

$$T(bx_1, \dots, bx_n) = b^2 T(x_1, \dots, x_n) \quad \text{para cada } b > 0$$

considerando como función de pérdida  $L(\sigma^2, t) = (\sigma^2 - t)^2/\sigma^4$ . Hallar el estimador del tipo anterior que minimice el riesgo.

**Ejercicio 6.10**

Hallar los estimadores de Pitman para la media y la desviación típica de una población  $N(\mu, \sigma)$ , basados en una muestra aleatoria simple de tamaño  $n$ .

## Solución de los ejercicios

### Ejercicio 6.1

a) La función de densidad de la muestra es

$$f_\theta(x_1, \dots, x_n) = \theta^n \left( \prod_{i=1}^n (1 + x_i) \right)^{-(1+\theta)}$$

de manera que  $S = \sum_{i=1}^n \log(1 + X_i)$  es un estadístico suficiente que, desde luego, es minimal.

Su distribución en el muestreo es  $\gamma(n, \theta)$  (véase ejercicio 1.7), de forma que si, para cualquier  $\theta > 0$ , es

$$0 = E[f(S)] = \int_0^\infty f(s) \frac{\theta^n s^{n-1} e^{-\theta s}}{(n-1)!} ds = \frac{\theta^n}{(n-1)!} \int_0^\infty h(s) e^{-\theta s} ds$$

(donde  $h(s) = f(s)s^{n-1}$ ) ha de ser  $h(s) = 0$ , en virtud de la unicidad de la transformada de Laplace, y por tanto,  $f(s) = 0$  para cada  $s > 0$ . Luego  $S$  es un estadístico completo.

b) Como  $E_\theta[S] = n/\theta$ , resulta inmediato que el ECUMV de  $1/\theta$  es

$$\frac{S}{n} = \frac{1}{n} \sum_{i=1}^n \log(1 + X_i).$$

Por otra parte, cabe pensar que el ECUMV de  $\theta$  pueda ser de la forma  $c/S$ . De hecho

$$E_\theta \left[ \frac{c}{S} \right] = c \int_0^\infty s^{-1} \frac{\theta^n}{(n-1)!} s^{n-1} e^{-\theta s} ds = \frac{c\theta}{n-1},$$

de forma que  $(n-1)/S$  es estimador insesgado de  $\theta$  y, como es función de  $S$ , es su ECUMV.

c) Como

$$\log f_\theta(x_1, \dots, x_n) = n \log \theta - (1 + \theta)S,$$

será

$$\frac{\partial}{\partial \theta} \log f_\theta(x_1, \dots, x_n) = \frac{n}{\theta} - S;$$

de manera que

$$\begin{aligned} I(\theta) &= E_\theta \left[ \left( \frac{n}{\theta} - S \right)^2 \right] = \frac{n^2}{\theta^2} - 2 \frac{n}{\theta} E_\theta[S] + E_\theta[S^2] \\ &= \frac{n^2}{\theta^2} - 2 \frac{n^2}{\theta^2} + \frac{n}{\theta^2} + \frac{n^2}{\theta^2} = \frac{n}{\theta^2}. \end{aligned}$$

Así que la cota de Fréchet-Cramer-Rao, para estimadores insesgados de  $1/\theta$ , afirma que

$$V_\theta(T) \geq \frac{[-1/\theta^2]^2}{n/\theta^2} = \frac{1}{n\theta^2}.$$

Puesto que el ECUMV de  $1/\theta$  tiene varianza

$$V_\theta(S/n) = \frac{1}{n^2} \frac{n}{\theta^2} = \frac{1}{n\theta^2},$$

igual a la cota, es estimador eficiente de  $1/\theta$ .

### Ejercicio 6.2

a) Segundo los ejemplos 5.9 y 6.3, el número total de caras obtenidas,  $S = \sum_{i=1}^n X_i$  (con  $X_i = 1$  ó  $0$ , según que se obtenga cara o cruz en el lanzamiento  $i$ -ésimo) es un estadístico suficiente y completo, cuya distribución es binomial de parámetros  $n$  y  $\theta$ .

Como  $\bar{X} = S/n$  es una función de  $S$  que es, además, estimador insesgado de  $\theta$ ,  $S/n$  es el ECUMV para  $\theta$ .

$g_2(\theta) = \theta(1 - \theta)$  es la varianza del número de caras obtenidas en un lanzamiento, es decir la varianza de la población; la cuasivarianza muestral

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

será entonces un estimador centrado de  $g_2(\theta)$ , que es su ECUMV, puesto que es una función de  $S$ .

Dado que

$$\theta^m = \mathcal{P}_\theta\{X_1 = 1, X_2 = 1, \dots, X_m = 1\},$$

un estimador insesgado de  $\theta^m$  es

$$T_1 = \begin{cases} 1 & \text{si } X_1 = X_2 = \dots = X_m = 1, \\ 0 & \text{en caso contrario.} \end{cases}$$

Pero

$$\begin{aligned} E[T_1 | S = s] &= P\{X_1 = 1, X_2 = 1, \dots, X_m = 1 | S = s\} \\ &= \frac{P\{X_1 = 1, X_2 = 1, \dots, X_m = 1, S = s\}}{P\{S = s\}} \\ &= \begin{cases} 0 & \text{si } m > s \\ \frac{\theta^m \binom{n-m}{s-m} \theta^{s-m} (1-\theta)^{n-s}}{\binom{n}{s} \theta^s (1-\theta)^{n-s}} = \frac{(n-m)! s!}{(s-m)! n!} & \text{si } m \leq s; \end{cases} \end{aligned}$$

de forma que el ECUMV para  $\theta^m$  es

$$E[T_1 \mid S] = \begin{cases} 0 & \text{si } S < m, \\ \frac{S! (n-m)!}{n! (S-m)!} & \text{si } S \geq m. \end{cases}$$

La función  $g_4(\theta) = 1/\theta$  no posee ningún estimador centrado, ya que, en caso contrario, habría un estimador centrado función de  $S$ ; para lo cual tendría que ser

$$\sum_{s=0}^n f(s) \binom{n}{s} \theta^s (1-\theta)^{n-s} = \frac{1}{\theta}.$$

Pero tal igualdad es imposible como se deduce, por ejemplo, del hecho de que el primer miembro tienda a  $f(0)$ , cuando  $\theta \rightarrow 0$ , mientras el segundo tiende a  $\infty$ .

b) Para que exista una función  $f(S)$  que sea estimador centrado de  $g(\theta)$ , debe ser

$$\begin{aligned} g(\theta) &= E[f(S)] = \sum_{s=0}^n f(s) \binom{n}{s} \theta^s (1-\theta)^{n-s} \\ &= \sum_{s=0}^n f(s) \binom{n}{s} \theta^s \sum_{j=0}^{n-s} (-1)^j \binom{n-s}{j} \theta^j \\ &= \sum_{i=0}^n \theta^i \sum_{s=0}^i f(s) \binom{n}{s} (-1)^{i-s} \binom{n-s}{i-s} \\ &= \sum_{i=0}^n c_i \theta^i \end{aligned}$$

(donde  $c_i$  es el valor de la suma anterior). Por tanto, sólo los polinomios de grado inferior o igual a  $n$  tienen ECUMV. Las demás funciones de  $\theta$  (como, por ejemplo,  $1/\theta$  ó  $\theta/(1-\theta)$ ) no poseen ECUMV, ni ningún estimador centrado.

Puesto que el ECUMV de  $\theta^i$  ha sido determinado en (a), el ECUMV de  $\sum_{i=0}^n c_i \theta^i$  será

$$\frac{S!}{n!} \sum_{i=0}^S c_i \frac{(n-i)!}{(S-i)!}.$$

De hecho

$$\begin{aligned}
 E \left[ \frac{S!}{n!} \sum_{i=0}^S c_i \frac{(n-i)!}{(S-i)!} \right] &= \sum_{s=0}^n \frac{s!}{n!} \sum_{i=0}^s c_i \frac{(n-i)!}{(s-i)!} \binom{n}{s} \theta^s (1-\theta)^{n-s} \\
 &= \sum_{i=0}^n c_i \frac{(n-i)!}{n!} \sum_{s=i}^n \frac{s!}{(s-i)!} \binom{n}{s} \theta^s (1-\theta)^{n-s} \\
 &= \sum_{i=0}^n c_i \theta^i \sum_{s=i}^n \binom{n-i}{s-i} \theta^{s-i} (1-\theta)^{n-s} \\
 &= \sum_{i=0}^n c_i \theta^i \sum_{j=0}^{n-i} \binom{n-i}{j} \theta^j (1-\theta)^{n-i-j} \\
 &= \sum_{i=0}^n c_i \theta^i.
 \end{aligned}$$

c) La función de probabilidad de cada observación es

$$f_\theta(x) = \theta^x (1-\theta)^{1-x} \quad \text{para } x = 1, 0,$$

cuyo logaritmo

$$\log f_\theta(x) = x \log \theta + (1-x) \log(1-\theta)$$

tiene por derivada

$$\frac{\partial}{\partial \theta} \log f_\theta(x) = \frac{x}{\theta} - \frac{1-x}{1-\theta} = \frac{x-\theta}{\theta(1-\theta)}.$$

La información de Fisher contenida en cada observación muestral es entonces

$$i(\theta) = E_\theta \left[ \frac{(X-\theta)^2}{\theta^2(1-\theta)^2} \right] = \frac{V_\theta(X)}{\theta^2(1-\theta)^2} = \frac{1}{\theta(1-\theta)},$$

con lo cual la información total de la muestra de tamaño  $n$  es  $I(\theta) = n i(\theta)$  y la cota de Fréchet-Cramer-Rao para la varianza de los estimadores insesgados de  $\theta$  resulta

$$V_\theta(T) \geq \frac{1}{I(\theta)} = \frac{\theta(1-\theta)}{n}.$$

Como  $\bar{X}$  es un estimador centrado de  $\theta$ , cuya varianza:  $V_\theta(\bar{X}) = \theta(1-\theta)/n$ , coincide con la cota,  $\bar{X}$  es un estimador eficiente de  $\theta$ . El resultado era previsible puesto que la densidad de la muestra:

$$f_\theta(x_1, \dots, x_n) = \theta^S (1-\theta)^{n-S} = (1-\theta)^n e^{n\bar{X}[\log \theta - \log(1-\theta)]}$$

es de tipo exponencial y  $E_\theta[\bar{X}] = \theta$ .

### Ejercicio 6.3

a) Según los ejemplos 5.11 y 6.6, el número total de fallos observados en los  $n$  kilómetros,  $S = \sum_{i=1}^n X_i$ , es un estadístico suficiente y completo, que tiene distribución de Poisson de parámetro  $\lambda = n\theta$ . Obsérvese que

$$\begin{aligned} E_\theta[S(S-1)(S-2)\cdots(S-j)] &= \sum_{s=0}^{\infty} s(s-1)(s-2)\cdots(s-j) \frac{e^{-\lambda}\lambda^s}{s!} \\ &= e^{-\lambda}\lambda^{j+1} \sum_{s=0}^{\infty} \frac{d^{j+1}}{d\lambda^{j+1}} \frac{\lambda^s}{s!} \\ &= e^{-\lambda}\lambda^{j+1} \frac{d^{j+1}}{d\lambda^{j+1}} e^\lambda \\ &= \lambda^{j+1} = n^{j+1}\theta^{j+1}. \end{aligned}$$

Luego

$$T = \frac{S!}{n^r(S-r)!}$$

es un estimador insesgado de  $\theta^r$ . Como  $T$  es función de  $S$ , será el ECUMV de  $\theta^r$ .

Para  $g_2(\theta) = P_\theta\{X = r\} = e^{-\theta}\theta^r/r!$ , es insesgado el estimador, basado en la primera observación,

$$T_1 = \begin{cases} 1 & \text{si } X_1 = r \\ 0 & \text{si } X_1 \neq r \end{cases}$$

Pero

$$\begin{aligned} E[T_1 | S = s] &= P\{X_1 = r | S = s\} = \frac{P\{X_1 = r\}P\{S = s | X_1 = r\}}{P\{S = s\}} = \\ &= \begin{cases} 0 & \text{si } s < r, \\ \frac{e^{-\theta}\theta^r}{r!} \frac{e^{-(n-1)\theta}[(n-1)\theta]^{s-r}}{(s-r)!} \frac{s!}{e^{-n\theta}[n\theta]^s} = \binom{s}{r} \frac{(n-1)^{s-r}}{n^s} & \text{si } s \geq r; \end{cases} \end{aligned}$$

luego el ECUMV para  $g_2(\theta)$  es

$$E[T_1 | S] = \begin{cases} 0 & \text{si } S < r, \\ \binom{S}{r} \frac{(n-1)^{S-r}}{n^S} & \text{si } S \geq r. \end{cases}$$

No existe ningún estimador centrado de  $g_3(\theta) = 1/\theta$ , pues, para que una función de  $S$  fuese estimador insesgado de  $1/\theta$ , habría de ser

$$\sum_{s=0}^{\infty} f(s) e^{-n\theta} \frac{(n\theta)^s}{s!} = \frac{1}{\theta};$$

lo cual es imposible, ya que el primer miembro converge a  $f(0)$ , cuando  $\theta \rightarrow 0$ , y el segundo miembro tiende a  $\infty$ .

b) Si  $f(S)$  es estimador centrado de  $g(\theta)$ , será

$$\begin{aligned} g(\theta) &= e^{-n\theta} \sum_{s=0}^{\infty} f(s) \frac{(n\theta)^s}{s!} = \sum_{j=0}^{\infty} (-1)^j \frac{(n\theta)^j}{j!} \sum_{s=0}^{\infty} f(s) \frac{(n\theta)^s}{s!} \\ &= \sum_{i=0}^{\infty} (n\theta)^i \sum_{j=0}^i \frac{(-1)^j}{j!} \frac{f(i-j)}{(i-j)!} = \sum_{i=0}^{\infty} c_i \theta^i. \end{aligned}$$

De forma que son estimables, mediante estimadores insesgados, todas las funciones de  $\theta$  que puedan expresarse como una serie de potencias, para los valores de  $\theta$  comprendidos en el espacio paramétrico ( $\mathbb{R}^+$  si no se tiene más información sobre  $\theta$ ). Por ejemplo,  $(1-\theta)^{-1}$  no admite ningún estimador centrado, a no ser que se sepa que  $\theta < 1$ .

A partir del ECUMV de  $\theta^i$  el ECUMV de  $g(\theta) = \sum_{i=0}^{\infty} c_i \theta^i$  resulta:

$$S! \sum_{i=0}^S \frac{c_i}{n^i (S-i)!}$$

que, desde luego, es estimador centrado de  $g(\theta)$  puesto que

$$\begin{aligned} E_{\theta} \left[ S! \sum_{i=0}^S \frac{c_i}{n^i (S-i)!} \right] &= \sum_{s=0}^{\infty} e^{-n\theta} \frac{(n\theta)^s}{s!} s! \sum_{i=0}^s \frac{c_i}{n^i (s-i)!} \\ &= e^{-n\theta} \sum_{i=0}^{\infty} c_i \theta^i \sum_{s=i}^{\infty} \frac{(n\theta)^{s-i}}{(s-i)!} = \sum_{i=0}^{\infty} c_i \theta^i. \end{aligned}$$

c) La función de probabilidad de cada observación  $X_i$  es

$$P_{\theta}\{X = x \mid X > 0\} = \frac{P_{\theta}\{X = x\}}{1 - P_{\theta}\{X = 0\}} = \frac{e^{-\theta} \theta^x}{x!(1 - e^{-\theta})} \quad \text{para } x = 1, 2, 3 \dots$$

con lo cual la función de probabilidad de la muestra es

$$f_{\theta}(x_1, \dots, x_n) = \frac{e^{-n\theta}}{(1 - e^{-\theta})^n} \frac{\theta^S}{\prod_{i=0}^n x_i!}$$

y

$$\frac{\partial}{\partial \theta} \log f_\theta(x_1, \dots, x_n) = \frac{S}{\theta} - \frac{n}{1 - e^{-\theta}} = \frac{n}{\theta} \left[ \frac{S}{n} - \frac{\theta}{1 - e^{-\theta}} \right].$$

Comparado con la última expresión del enunciado de la cota de Fréchet-Cramer-Rao, resulta que  $g(\theta) = \theta/(1-e^{-\theta})$  es la única función de  $\theta$  que tiene estimador insesgado eficiente. Dicho estimador es  $S/n = \bar{X}$  y su varianza es  $V_\theta(\bar{X}) = g'(\theta)\theta/n$ .

### Ejercicio 6.4

a) La densidad de una muestra aleatoria simple de tamaño  $n$  es

$$f_\theta(x_1, \dots, x_n) = \frac{k^n (x_1 \cdots x_n)^{k-1}}{\theta^{nk}} I_{\{x_{(n)} < \theta\}} \quad \text{para } x_1, \dots, x_n > 0;$$

por consiguiente,  $X_{(n)}$  es un estadístico suficiente. Su función de distribución ha sido determinada en el ejemplo 6.5, donde también se ha comprobado que es un estadístico completo.

Como

$$E_\theta[X_{(n)}] = \frac{nk}{\theta^{nk}} \int_0^\theta y^{nk} dy = \frac{nk}{nk+1} \theta,$$

el estadístico

$$T = \frac{nk+1}{nk} X_{(n)}$$

es el único estimador centrado de  $\theta$ , función de  $X_{(n)}$ . Luego  $T$  es el ECUMV de  $\theta$ .

b) Para que una función  $g(\theta)$  tenga ECUMV, tiene que haber una función  $\varphi(X_{(n)})$  cuya esperanza coincida con  $g(\theta)$ ; es decir

$$\frac{nk}{\theta^{nk}} \int_0^\theta \varphi(y) y^{nk-1} dy = g(\theta)$$

o bien, si  $g(\theta)$  es derivable,

$$nk \varphi(\theta) \theta^{nk-1} = nk \theta^{nk-1} g(\theta) + \theta^{nk} g'(\theta);$$

de donde se obtiene

$$\varphi(\theta) = g(\theta) + \frac{\theta}{nk} g'(\theta).$$

Así pues, cualquier función derivable,  $g(\theta)$ , admite ECUMV que viene dado por

$$T_g(X_{(n)}) = g(X_{(n)}) + \frac{X_{(n)}}{nk} g'(X_{(n)}).$$

Por ejemplo, el ECUMV de  $\sin \theta$  es  $\sin X_{(n)} + X_{(n)}/(nk) \cos X_{(n)}$ .

c)  $\theta$  es un parámetro de escala, puesto que, si  $Z$  tiene distribución  $x^k$  en  $(0, 1)$ ,  $\theta Z$  tiene la distribución considerada. El estimador de Pitman corresponde a suponer que la función de pérdida en la estimación es  $L(t, \theta) = (\theta - t)^2/\theta^2$  y obtener el estimador invariante por cambios de escala de mínimo riesgo.

Para calcularlo, nótese que

$$\begin{aligned} I_r &= \int_0^\infty \theta^{-r} f_\theta(x_1, \dots, x_n) d\theta = k^n (x_1 \cdots x_n)^{k-1} \int_{x_{(n)}}^\infty \frac{d\theta}{\theta^{nk+r}} \\ &= \frac{k^n (x_1 \cdots x_n)^{k-1}}{(nk + r - 1)x_{(n)}^{nk+r-1}}, \end{aligned}$$

de forma que el estimador de Pitman es

$$T^* = \frac{I_2}{I_3} = \frac{nk + 2}{nk + 1} X_{(n)}$$

que tiene sesgo  $-\theta/(nk + 1)^2$ .

### Ejercicio 6.5

a) A partir de la muestra aleatoria  $(X_1, \dots, X_n)$ , conviene considerar los estadísticos

$$\begin{aligned} N &= \text{Número de } X_i \text{ con el valor } -1, \\ T &= \sum_{i=1}^n |X_i|, \end{aligned}$$

en función de los cuales la función de probabilidad de la muestra se expresa

$$f_\theta(x_1, \dots, x_n) = (1 - \theta)^{2(n - N)} \theta^T.$$

No cabe duda entonces de que  $(N, T)$  es un estadístico suficiente minimal; al igual que cualquier función biunívoca suya, por ejemplo:  $(P, S)$ , siendo

$$\begin{aligned} P &= \text{Número de } X_i \geq 0 \\ S &= \text{Suma de todos los } X_i \geq 0 \end{aligned}$$

(de hecho  $P = n - N$  y  $S = T - N$ ).

Aunque  $(N, T)$  es bidimensional, nótese que, en el caso  $n = 2$ , la muestra ordenada:  $(X_{(1)}, X_{(2)})$ , no es un estadístico suficiente minimal, pues saber

cuántas son no negativas y la suma de ambas no permite determinarlas (en el caso en que ninguna valga  $-1$ ).

Por otra parte,

$$E_\theta[X_i] = (1 - \theta)^2 \sum_{j=0}^{\infty} j\theta^j - \theta = 0,$$

de manera que  $E_\theta[\sum_{i=1}^n X_i] = 0$ . Sin embargo,

$$\sum_{i=1}^n X_i = T - 2N$$

es una función del estadístico minimal suficiente y, desde luego, no es casi seguramente nula (antes bien, es muy poco probable que sea  $T = 2N$ ). Luego el estadístico minimal suficiente no es completo.

b) Según lo anterior, deben existir diversas funciones de  $(N, T)$  que estimen  $\theta$  sin sesgo. No es difícil encontrar un par de ellas:

▷ Como  $N$  tiene distribución binomial de parámetros  $n$  y  $\theta$ , será

$$E_\theta\left[\frac{N}{n}\right] = \theta \quad \text{y} \quad V_\theta\left(\frac{N}{n}\right) = \frac{\theta(1 - \theta)}{n};$$

luego  $N/n$  es un estimador centrado de  $\theta$  con la varianza indicada.

▷ Como

$$E_\theta[|X_i|] = 2\theta \quad \text{y} \quad E_\theta[X_i^2] = \theta + (1 - \theta)^2 \sum_{j=1}^{\infty} j^2 \theta^j = \frac{2\theta}{1 - \theta},$$

el estadístico  $T/2n$  es un estimador centrado de  $\theta$  de varianza

$$V_\theta\left(\frac{T}{2n}\right) = \frac{1}{4n} V_\theta(|X_i|) = \frac{1}{4n} \left( \frac{2\theta}{1 - \theta} - 4\theta^2 \right) = \frac{\theta(1 - \theta)}{2n} + \frac{\theta^3}{2n(1 - \theta)}.$$

▷ Cualquier combinación lineal de la forma  $\lambda N/n + (1 - \lambda)T/2n$  será también un estimador insesgado de  $\theta$ , de varianza

$$\lambda^2 \frac{\theta(1 - \theta)}{n} + (1 - \lambda)^2 \left[ \frac{\theta(1 - \theta)}{2n} + \frac{\theta^3}{2n(1 - \theta)} \right] + 2\lambda(1 - \lambda) \frac{\theta(1 - 2\theta)}{2n}$$

ya que

$$\begin{aligned} E_\theta[N T] &= E_\theta \left[ \sum_{i=1}^n I_{\{X_i=-1\}} \sum_{j=1}^n |X_j| \right] \\ &= nE_\theta[|X_1|I_{\{X_1=-1\}}] + n(n-1)E_\theta[|X_2|]E_\theta[I_{\{X_2=-1\}}] \\ &= n\theta + n(n-1)2\theta^2, \end{aligned}$$

con lo cual

$$\text{Cov}_\theta \left( \frac{N}{n}, \frac{T}{2n} \right) = \frac{\theta}{2n} + \frac{n-1}{n}\theta^2 - \theta^2 = \frac{\theta}{2n} - \frac{\theta^2}{n}.$$

Se puede ver que las funciones de  $\theta$  correspondientes a los distintos valores de  $\lambda$  se entrecruzan, si  $0 \leq \lambda \leq 1$ , de manera que, entre los estimadores asociados a tales valores de  $\lambda$ , no hay ninguno uniformemente de menos varianza que los demás. (La única varianza que permanece acotada, al variar  $\theta$ , es la correspondiente a  $\lambda = 1$ ; pero es la más grande si  $\theta$  es próximo a 0, así que sólo sería preferible desde el punto de vista minimax).

Es posible que la situación sea similar si se consideran funciones no lineales de  $n$  y  $T$  y que, por tanto, no exista ECUMV para  $\theta$ . De hecho, es lo que ocurre si  $n = 1$ , como muestra el apartado siguiente.

c) Para que una función  $\varphi(X_1)$  tenga esperanza cero, para cualquier valor de  $\theta$ , debe ser

$$\theta\varphi(-1) + (1-\theta)^2 \sum_{j=0}^{\infty} \theta^j \varphi(j) = 0,$$

o bien

$$\sum_{j=0}^{\infty} \varphi(j)\theta^j = -\varphi(-1) \frac{\theta}{(1-\theta)^2} = -\varphi(-1) \sum_{j=0}^{\infty} j\theta^j;$$

de donde resulta  $\varphi(j) = -j\varphi(-1)$  para cada  $j \geq 0$ ; es decir  $\varphi(j) = cj$  para cada  $j \geq -1$  (siendo  $c = -\varphi(-1)$ ). Las funciones  $cX_1$  son, por tanto, las únicas de esperanza cero.

Puesto que  $T_1 = I_{\{X_1=-1\}}$  es un estimador insesgado de  $\theta$ , cualquier otro estimador insesgado se obtendrá sumándole un estadístico de esperanza nula. Es decir, todos los estimadores insesgados de  $\theta$  son de la forma  $T_1 + cX_1$ . Sus varianzas son

$$V_\theta(T_1 + cX_1) = V_\theta(T_1) + c^2 V_\theta(X_1) + 2c \text{Cov}_\theta(T_1, X_1) = \theta(1-\theta) + c^2 \frac{2\theta}{1-\theta} - 2c\theta.$$

Entre ellas es mínima, en el punto  $\theta$ , la correspondiente a  $c = (1 - \theta)/2$ ; como dicho valor depende de  $\theta$ , no hay ningún estimador centrado de  $\theta$  que tenga menor varianza que los demás, simultáneamente para todos los valores de  $\theta$ .

En cambio, para  $(1 - \theta)^2$ , se cuenta con el estimador insesgado  $T_2 = I_{\{X_1=0\}}$ , y  $T_2 + cX_1$  es la familia de todos sus estimadores insesgados. Dado que  $\text{Cov}_\theta(T_2, X_1) = 0$ , la varianza resulta

$$V_\theta(T_2 + cX_1) = V_\theta(T_2) + c^2 V_\theta(X_1) = (1 - \theta)^2 - (1 - \theta)^4 + c^2 \frac{2\theta}{1 - \theta},$$

que es mínima para  $c = 0$ , independientemente del valor de  $\theta$ . Así que  $T_2$  es el ECUMV de  $(1 - \theta)^2$ .

### Ejercicio 6.6

a) La función de densidad de las  $n$  observaciones es

$$f_{a,b}(x_1, \dots, x_n) = \frac{1}{(b-a)^n} I_{\{a < x_{(1)} \leq x_{(n)} < b\}},$$

de forma que  $(X_{(1)}, X_{(n)})$  es el estadístico suficiente minimal. Su densidad es

$$g_{a,b}(y, z) = n(n-1) \frac{(z-y)^{n-2}}{(b-a)^n} I_{\{a < y \leq z < b\}}.$$

Si  $\varphi(y, z)$  es una función, definida para  $0 < y \leq z$ , que cumple

$$0 = E_{a,b}[\varphi(X_{(1)}, X_{(n)})] = \frac{n(n-1)}{(b-a)^n} \int_a^b \int_a^z \varphi(y, z)(z-y)^{n-2} dy dz$$

para cualesquiera  $a < b$ , eliminando el primer factor y derivando respecto a  $b$ , se obtiene

$$\int_a^b \varphi(y, b)(b-y)^{n-2} dy = 0$$

y, derivando ahora respecto a  $a$ ,

$$-\varphi(a, b)(b-a)^{n-2} = 0;$$

lo cual prueba que  $\varphi$  es nula; es decir, que  $(X_{(1)}, X_{(n)})$  es un estadístico completo.

b) La densidad marginal de  $X_{(n)}$  es  $n(z-a)^{n-1}/(b-a)^n I_{\{a < z < b\}}$ , de manera que

$$E_{a,b}[X_{(n)}] = \frac{n}{(b-a)^n} \int_a^b z(z-a)^{n-1} dz = \frac{nb}{n+1} + \frac{a}{n+1}.$$

Análogamente,  $n(b-y)^{n-1}/(b-a)^n I_{\{a < y < b\}}$  es la densidad marginal de  $X_{(1)}$ , con lo cual

$$\text{E}_{a,b}[X_{(1)}] = \frac{n}{(b-a)^n} \int_a^b y(b-y)^{n-1} dy = \frac{b}{n+1} + \frac{na}{n+1}.$$

Entonces

$$T_1 = \frac{nX_{(1)} - X_{(n)}}{n-1} \quad \text{y} \quad T_2 = \frac{nX_{(n)} - X_{(1)}}{n-1}$$

tienen media

$$\begin{aligned} \text{E}_{a,b}[T_1] &= \frac{n}{n-1} \left( \frac{b}{n+1} + \frac{na}{n+1} \right) - \frac{1}{n-1} \left( \frac{nb}{n+1} + \frac{a}{n+1} \right) = a, \\ \text{E}_{a,b}[T_2] &= \frac{n}{n-1} \left( \frac{nb}{n+1} + \frac{a}{n+1} \right) - \frac{1}{n-1} \left( \frac{b}{n+1} + \frac{na}{n+1} \right) = b. \end{aligned}$$

Son, por tanto, los ECUMV de  $a$  y  $b$  respectivamente. Por otra parte,

$$\text{E}_{a,b} \left[ \frac{n+1}{n-1} (X_{(n)} - X_{(1)}) \right] = b - a;$$

de forma que  $(n+1)/(n-1)$   $(X_{(n)} - X_{(1)})$  es el ECUMV del recorrido. Nótese que los resultados son similares, pero no coinciden exactamente, con los estimadores invariantes de mínimo riesgo, determinados en el ejemplo 6.18.

c) Evidentemente, si  $Z$  tiene distribución uniforme en  $(1, 2)$ ,  $aZ$  tiene distribución uniforme en  $(a, 2a)$ . Luego  $a$  es un parámetro de escala. Puesto que la densidad de la muestra es

$$f_a(x_1, \dots, x_n) = \frac{1}{a^n} I_{\{a < x_{(1)} \leq x_{(n)} < 2a\}},$$

será

$$\begin{aligned} I_k &= \int_0^\infty a^{-k} f_a(x_1, \dots, x_n) da = \int_{x_{(n)}/2}^{x_{(1)}} a^{-n-k} da \\ &= \frac{1}{n+k-1} \left[ (x_{(n)}/2)^{-n-k+1} - x_{(1)}^{-n-k+1} \right]; \end{aligned}$$

con lo cual, el estimador de Pitman resulta

$$\begin{aligned} T^* &= \frac{I_2}{I_3} = \frac{n+2}{n+1} \frac{(X_{(n)}/2)^{-n-1} - X_{(1)}^{-n-1}}{(X_{(n)}/2)^{-n-2} - X_{(1)}^{-n-2}} \\ &= \frac{n+2}{n+1} X_{(1)} X_{(n)}/2 \frac{X_{(1)}^{n+1} - (X_{(n)}/2)^{n+1}}{X_{(1)}^{n+2} - (X_{(n)}/2)^{n+2}}. \end{aligned}$$

**Ejercicio 6.7**

a) La densidad de la muestra es

$$f_{p,\theta}(x_1, \dots, x_n) = \frac{\theta^{np}}{\Gamma(p)^n} (x_1 \cdots x_n)^{p-1} \exp\left\{-\theta \sum_{i=1}^n x_i\right\},$$

cuyo logaritmo,

$$\log f_{p,\theta} = np \log \theta - n \log \Gamma(p) + (p-1) \sum_{i=1}^n \log x_i - \theta \sum_{i=1}^n x_i$$

tiene derivadas parciales

$$\frac{\partial}{\partial p} \log f_{p,\theta} = n \log \theta - n \frac{\Gamma'(p)}{\Gamma(p)} + \sum_{i=1}^n \log x_i,$$

$$\frac{\partial}{\partial \theta} \log f_{p,\theta} = \frac{np}{\theta} - \sum_{i=1}^n x_i,$$

y derivadas segundas

$$\begin{aligned} \frac{\partial^2}{\partial p^2} \log f_{p,\theta} &= -n \frac{\Gamma''(p)\Gamma(p) - \Gamma'(p)^2}{\Gamma(p)^2} = -n\psi(p), \\ \frac{\partial^2}{\partial \theta^2} \log f_{p,\theta} &= -\frac{np}{\theta^2}, \\ \frac{\partial^2}{\partial p \partial \theta} \log f_{p,\theta} &= \frac{n}{\theta}. \end{aligned}$$

La matriz de información es entonces

$$I(p, \theta) = n \begin{pmatrix} \psi(p) & -1/\theta \\ -1/\theta & p/\theta^2 \end{pmatrix},$$

de manera que

$$I(p, \theta)^{-1} = \frac{1}{n} \begin{pmatrix} p & \theta \\ \theta & \theta^2 \psi(p) \end{pmatrix} \frac{1}{p\psi(p) - 1}.$$

Por tanto, las cotas para la varianza de estimadores centrados de  $p$  y  $\theta$  son respectivamente:

$$\frac{p}{n[p\psi(p) - 1]} \quad \text{y} \quad \frac{\theta^2 \psi(p)}{n[p\psi(p) - 1]}.$$

(Mientras que, si  $\theta$  fuese conocido, la varianza de los estimadores centrados de  $p$  estaría acotada por  $1/[n\psi(p)]$  y, si  $p$  fuese conocido, la varianza de los estimadores centrados de  $\theta$  estaría acotada por  $\theta^2/(np)$ .)

Si  $T = (T_1, T_2)$  es un estimador insesgado de  $(p, \theta)$ , su varianza generalizada,  $|V_{p,\theta}(T)|$  está acotada por

$$|\Delta I(p, \theta)^{-1} \Delta^t| = |I(p, \theta)^{-1}| = \frac{\theta^2}{n}.$$

b) La media poblacional es  $p/\theta$ , de forma que

$$\mathbb{E}_{p,\theta}[\bar{X}] = p/\theta.$$

Por otra parte,

$$\begin{aligned}\mathbb{E}_{p,\theta}[\log X_i] &= \frac{\theta^p}{\Gamma(p)} \int_0^\infty \log x \ x^{p-1} e^{-\theta x} dx \\ &= \frac{1}{\Gamma(p)} \int_0^\infty (\log y - \log \theta) y^{p-1} e^{-y} dy = \frac{\Gamma'(p)}{\Gamma(p)} - \log \theta;\end{aligned}$$

con lo cual  $\mathbb{E}_{p,\theta}[n^{-1} \sum_{i=1}^n \log X_i] = \Gamma'(p)/\Gamma(p) - \log \theta$ . Para las funciones de  $p$  y  $\theta$  obtenidas es

$$\Delta = \begin{pmatrix} 1/\theta & -p/\theta^2 \\ \psi(p) & -1/\theta \end{pmatrix}$$

y

$$\Delta I(p, \theta)^{-1} \Delta^t = \frac{1}{n} \begin{pmatrix} p/\theta^2 & 1/\theta \\ 1/\theta & \psi(p) \end{pmatrix};$$

de forma que  $p/(n\theta^2)$  y  $\psi(p)/n$  son las cotas para las varianzas de los estimadores insesgados de  $p/\theta$  y  $\Gamma'(p)/\Gamma(p) - \log \theta$ , respectivamente.

Desde luego,  $V_{p,\theta}(\bar{X}) = p/(n\theta^2)$ ; por otra parte

$$V_{p,\theta} \left( \frac{1}{n} \sum_{i=1}^n \log X_i \right) = \frac{\psi(p)}{n},$$

puesto que

$$\begin{aligned}\mathbb{E}_{p,\theta}[(\log X_i)^2] &= \frac{1}{\Gamma(p)} \int_0^\infty (\log y - \log \theta)^2 y^{p-1} e^{-y} dy \\ &= \frac{\Gamma''(p)}{\Gamma(p)} - 2 \log \theta \frac{\Gamma'(p)}{\Gamma(p)} + \log^2 \theta;\end{aligned}$$

con lo cual  $V_{p,\theta}(\log X_i) = \psi(p)$ . Resulta así que  $\bar{X}$  y  $n^{-1} \sum_{i=1}^n \log X_i$  son estimadores eficientes de sus medias.

c) Para probar la completitud de  $Y = \sum_{i=1}^n X_i$  y  $Z = -\sum_{i=1}^n \log X_i$ , obsérvese en primer lugar, que, fijado  $Y$ , el menor valor de  $Z$  se alcanza cuando todas las  $X_i$  son iguales a  $Y/n$ . Supongamos, entonces, que existe una función  $\varphi(y, z)$ , definida para  $z > -n \log(y/n)$ , tal que

$$\int_{(0,\infty)^n} \varphi(y, z) e^{-z(p-1)} e^{-\theta y} dx_1 \dots dx_n = 0$$

(donde se ha eliminado la constante  $\theta^{np}/\Gamma(p)^n$ ). Al hacer el cambio de la variable de integración a  $x_1, \dots, x_{n-2}, y, z$  e integrar, primero, respecto a  $x_1, \dots, x_{n-2}$ , se obtendrá una función  $f(y, z)$ , difícil de determinar explícitamente, pero que no depende de los parámetros, no se anula (para  $z > -n \log(y/n)$ ) y puede ser englobada junto a  $\varphi(y, z)$  y  $e^z$ . La igualdad anterior adoptará entonces la forma

$$\int_0^\infty \int_{-n \log(y/n)}^\infty \psi(y, z) e^{-pz} dz e^{-\theta y} dy = 0.$$

Aplicando la unicidad de la transformada de Laplace, resulta

$$\int_{-n \log(y/n)}^\infty \psi(y, z) e^{-pz} dz = 0$$

para cada  $y$ , o bien

$$e^{pn \log(y/n)} \int_0^\infty \psi(y, u - n \log(y/n)) e^{-pu} du = 0$$

y, nuevamente en virtud de la unicidad de la transformada de Laplace, resulta  $\psi(y, z) = 0$ , o bien  $\varphi(y, z) = 0$ .

Establecida la completitud de  $(Y, Z)$ , el teorema de Lehman-Scheffé asegura que  $\bar{X}$  y  $-Z/n$  son ECUMV de sus medias; lo cual podía haberse deducido de su eficiencia. Existe un resultado genérico<sup>17</sup> que garantiza la completitud del estadístico minimal suficiente correspondiente a una población de tipo exponencial (véase ejemplo 5.17) siempre que la imagen de la función  $(q_1(\theta), q_2(\theta), \dots, q_p(\theta))$  contenga un abierto de  $\mathbb{R}^p$ . Su utilización evita el trabajo anterior.

---

<sup>17</sup> Cf. Lehman [1986] pág. 142.

**Ejercicio 6.8**

La función de densidad de la muestra aleatoria simple de tamaño  $n$  es

$$f_\theta(x_1, \dots, x_n) = \left(\frac{2}{\pi}\right)^{n/2} e^{-1/2 \sum_{i=1}^n (x_i - \theta)^2} I_{\{x_{(1)} > \theta\}};$$

de forma que el estimador de Pitman es

$$\begin{aligned} T^*(x_1, \dots, x_n) &= \frac{\int_{-\infty}^{x_{(1)}} z e^{-1/2 \sum_{i=1}^n (x_i - z)^2} dz}{\int_{-\infty}^{x_{(1)}} e^{-1/2 \sum_{i=1}^n (x_i - z)^2} dz} = \frac{\int_{-\infty}^{x_{(1)}} z e^{-n/2(z - \bar{x})^2} dz}{\int_{-\infty}^{x_{(1)}} e^{-n/2(z - \bar{x})^2} dz} \\ &= \frac{\int_{-\infty}^{\sqrt{n}(x_{(1)} - \bar{x})} (\bar{x} + y/\sqrt{n}) e^{-y^2/2} dy}{\int_{-\infty}^{\sqrt{n}(x_{(1)} - \bar{x})} e^{-y^2/2} dy} \\ &= \bar{x} + \frac{\int_{-\infty}^{\sqrt{n}(x_{(1)} - \bar{x})} y e^{-y^2/2} dy}{\sqrt{n} \int_{-\infty}^{\sqrt{n}(x_{(1)} - \bar{x})} e^{-y^2/2} dy} \\ &= \bar{x} - \frac{e^{-n(x_{(1)} - \bar{x})^2/2}}{\sqrt{2\pi n} \phi(\sqrt{n}(x_{(1)} - \bar{x}))} \end{aligned}$$

donde  $\phi$  representa la función de distribución  $N(0, 1)$ .

**Ejercicio 6.9**

La estimación del parámetro de escala,  $\sigma$ , ha sido considerada en el ejemplo 6.17; ahora se trata de estimar  $\sigma^2$ . Los estimadores considerados son los cuadrados de estimadores invariantes de  $\sigma$  y la función de pérdida verifica

$$L(b^2\sigma^2, b^2t) = L(\sigma^2, t),$$

de forma que la pérdida no depende de las unidades de medida utilizadas. Por consiguiente, el riesgo de los estimadores utilizados no dependerá del valor del parámetro:

$$\begin{aligned} R_T(\sigma) &= E_\sigma[L(\sigma^2, T(X_1, \dots, X_n))] = E_1[L(\sigma^2, T(\sigma X_1, \dots, \sigma X_n))] \\ &= E_1[L(\sigma^2, \sigma^2 T(X_1, \dots, X_n))] = E_1[L(1, T(X_1, \dots, X_n))] = R_T(1). \end{aligned}$$

Además, los estimadores propuestos son de la forma

$$T(x_1, \dots, x_n) = x_n^2 \varphi(x_1 x_n^{-1}, \dots, x_{n-1} x_n^{-1})$$

(siendo  $\varphi(y_1, \dots, y_{n-1}) = T(y_1, \dots, y_{n-1}, \pm 1)$  según que  $x_n > 0$  ó  $x_n < 0$ ).

La densidad de  $X_1, \dots, X_n$ , asociada al valor  $\sigma = 1$  es

$$f(x_1, \dots, x_n) = (2\pi)^{-n/2} e^{-1/2 \sum_{i=1}^n x_i^2}$$

de forma que,  $X_1 X_n^{-1}, \dots, X_{n-1} X_n^{-1}$  y  $X_n$  tienen densidad conjunta

$$g(y_1, \dots, y_{n-1}, x_n) = (2\pi)^{-n/2} e^{-x_n^2 S/2} |x_n|^{n-1}$$

donde  $S = 1 + \sum_{i=1}^{n-1} y_i^2$ . La densidad marginal de  $X_1 X_n^{-1}, \dots, X_{n-1} X_n^{-1}$  es pues

$$\begin{aligned} g(y_1, \dots, y_{n-1}) &= 2(2\pi)^{-n/2} \int_0^\infty x_n^{n-1} e^{-x_n^2 S/2} dx_n \\ &= (2\pi)^{-n/2} \int_0^\infty \left(\frac{2u}{S}\right)^{(n-1)/2} e^{-u} \left(\frac{2}{S}\right)^{1/2} u^{-1/2} du \\ &= (\pi S)^{-n/2} \Gamma\left(\frac{n}{2}\right) \end{aligned}$$

y la densidad de  $X_n$  condicionada por  $X_1 X_n^{-1}, \dots, X_{n-1} X_n^{-1}$  resulta

$$h(x_n \mid y_1, \dots, y_{n-1}) = \Gamma\left(\frac{n}{2}\right)^{-1} \left(\frac{S}{2}\right)^{n/2} e^{-x_n^2 S/2} |x_n|^{n-1}.$$

Como el riesgo  $R_T(1)$ , independiente de  $\sigma$ , de cada estimador vale

$$\iint_{\mathbb{R}} (1 - x_n^2 \varphi(y_1, \dots, y_{n-1}))^2 h(x_n \mid y_1, \dots, y_{n-1}) dx_n g(y_1, \dots, y_{n-1}) dy_1 \dots dy_{n-1}$$

habrá de escogerse  $\varphi(y_1, \dots, y_{n-1})$  que minimice

$$2 \int_0^\infty (1 - x_n^2 \varphi(y_1, \dots, y_{n-1}))^2 h(x_n \mid y_1, \dots, y_{n-1}) dx_n;$$

es decir

$$\varphi(y_1, \dots, y_{n-1}) = \frac{\int_0^\infty x_n^2 h(x_n \mid y_1, \dots, y_{n-1}) dx_n}{\int_0^\infty x_n^4 h(x_n \mid y_1, \dots, y_{n-1}) dx_n}.$$

Ahora bien

$$\begin{aligned}
 I_{2k} &= \int_0^\infty x_n^{2k} h(x_n | y_1, \dots, y_{n-1}) dx_n \\
 &= \Gamma\left(\frac{n}{2}\right)^{-1} \left(\frac{S}{2}\right)^{n/2} \int_0^\infty x_n^{2k+n-1} e^{-x_n^2 S/2} dx_n \\
 &= \Gamma\left(\frac{n}{2}\right)^{-1} \left(\frac{S}{2}\right)^{-k} \frac{1}{2} \int_0^\infty u^{(2k+n-2)/2} e^{-u} du \\
 &= \Gamma\left(\frac{n}{2}\right)^{-1} \left(\frac{S}{2}\right)^{-k} \frac{1}{2} \Gamma\left(\frac{n+2k}{2}\right),
 \end{aligned}$$

con lo cual

$$\varphi(y_1, \dots, y_{n-1}) = \frac{I_2}{I_4} = \frac{S}{2} \frac{\Gamma\left(\frac{n+2}{2}\right)}{\Gamma\left(\frac{n+4}{2}\right)} = \frac{S}{n+2} = \frac{1}{n+2} \left(1 + \sum_{i=1}^{n-1} y_i^2\right)$$

y el estimador buscado es

$$T^* = x_n^2 \varphi(x_1 x_n^{-1}, \dots, x_{n-1} x_n^{-1}) = \frac{x_n^2}{n+2} \left(1 + \sum_{i=1}^{n-1} x_i^2 x_n^{-2}\right) = \frac{1}{n+2} \sum_{i=1}^n x_i^2.$$

Nótese que, aunque ambos sean función de  $\sum_{i=1}^n x_i^2$ , el estimador invariante óptimo de  $\sigma^2$  no es el cuadrado del estimador invariante óptimo de  $\sigma$ , obtenido en el ejemplo 6.17.

### Ejercicio 6.10

La densidad de una muestra aleatoria simple de tamaño  $n$  es

$$f_{\mu, \sigma}(x_1, \dots, x_n) = \left(\sqrt{2\pi}\sigma\right)^{-n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right\}.$$

Entonces, si  $s^2$  es la varianza muestral,

$$\begin{aligned}
 I_k &= \int_0^\infty \int_{-\infty}^\infty v^{-k} f_{u,v}(x_1, \dots, x_n) du dv \\
 &= \left(\sqrt{2\pi}\right)^{-n} \int_0^\infty v^{-n-k} e^{-ns^2/(2v^2)} \int_{-\infty}^\infty e^{-n(u-\bar{x})^2/(2v^2)} du dv \\
 &= \left(\sqrt{2\pi}\right)^{-n} \int_0^\infty v^{-n-k} e^{-ns^2/(2v^2)} \sqrt{\frac{2\pi}{n}} v dv
 \end{aligned}$$

$$\begin{aligned}
&= \left(\sqrt{2\pi}\right)^{-n} \sqrt{\frac{2\pi}{n}} \int_0^\infty \left(\frac{2y}{ns^2}\right)^{(n+k-1)/2} e^{-y} \frac{1}{2} \left(\frac{ns^2}{2}\right)^{1/2} y^{-3/2} dy \\
&= \left(\sqrt{2\pi}\right)^{-n} \sqrt{\frac{\pi}{2n}} \left(\frac{2}{ns^2}\right)^{(n+k-2)/2} \int_0^\infty y^{(n+k-4)/2} e^{-y} dy \\
&= \left(\sqrt{2\pi}\right)^{-n} \sqrt{\pi} \frac{2^{(n+k-3)/2}}{n^{(n+k-1)/2} s^{n+k-2}} \Gamma\left(\frac{n+k-2}{2}\right);
\end{aligned}$$

con lo cual el estimador de Pitman de  $\sigma$  resulta

$$T_2^* = \frac{I_2}{I_3} = \frac{\sqrt{n}}{\sqrt{2}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n+1}{2})} s,$$

que no coincide con el estimador hallado en el ejemplo 6.17, debido a la presencia del parámetro de posición desconocido.

Para determinar el estimador de Pitman de  $\mu$  hay que calcular

$$\begin{aligned}
J &= \int_0^\infty \int_{-\infty}^\infty uv^{-3} f_{u,v}(x_1, \dots, x_n) du dv \\
&= \left(\sqrt{2\pi}\right)^{-n} \int_0^\infty v^{-n-3} e^{-ns^2/(2v^2)} \int_{-\infty}^\infty ue^{n(u-\bar{x})^2} du dv \\
&= \left(\sqrt{2\pi}\right)^{-n} \int_0^\infty v^{-n-3} e^{-ns^2/(2v^2)} \sqrt{\frac{2\pi}{n}} v \bar{x} dv \\
&= \left(\sqrt{2\pi}\right)^{-n} \sqrt{\frac{2\pi}{n}} \bar{x} \int_0^\infty \left(\frac{2y}{ns^2}\right)^{(n+2)/2} e^{-y} \frac{1}{2} \left(\frac{ns^2}{2}\right)^{1/2} y^{-3/2} dy \\
&= \left(\sqrt{2\pi}\right)^{-n} \frac{\sqrt{\pi}}{\sqrt{2n}} \bar{x} \left(\frac{2}{ns^2}\right)^{(n+1)/2} \int_0^\infty y^{(n-1)/2} e^{-y} dy \\
&= \left(\sqrt{2\pi}\right)^{-n} \sqrt{\pi} \frac{2^{n/2}}{n^{(n+2)/2} s^{n+1}} \Gamma\left(\frac{n+1}{2}\right) \bar{x};
\end{aligned}$$

de forma que se obtiene como estimador de Pitman para  $\mu$ :

$$T_1^* = \frac{J}{I_3} = \bar{x}.$$



## Capítulo 7

# Métodos de estimación

### 7.1. Introducción

Los estimadores puntuales obtenidos mediante criterios ligados al concepto de riesgo, de los que nos hemos ocupado en los últimos capítulos, no agotan las posibilidades de construcción de estimadores. Al contrario, algunas de las técnicas de estimación más clásicas son completamente ajenas a la noción de riesgo, e incluso a cualquier tipo de optimalidad. Sin embargo, tienen la ventaja de permitir con comodidad la construcción de estimadores, cuyas cualidades pueden analizarse más tarde —estudiando si son centrados, invariantes, eficientes, etc. Entre tales técnicas destacan el *método de los momentos* y, sobre todo, el *método de la máxima verosimilitud*, que posee una mejor justificación teórica y propiedades más destacables.

Se añaden, en las últimas secciones, dos métodos de estimación más específicos; bien sea por el tipo de información que utilizan, como es el caso de la *estimación bayesiana*; bien sea por el tipo de problemas en que se emplean, tal como sucede con la *estimación mínimo cuadrática*.

### 7.2. El método de los momentos

Consideremos una población cuya distribución teórica depende de  $k$  parámetros desconocidos que se desean estimar:  $\theta_1, \theta_2, \dots, \theta_k$ . En el caso en que existan, los momentos poblacionales serán, en general, función de ellos; es decir

$$E_{\theta}[X^r] = \alpha_r(\theta_1, \theta_2, \dots, \theta_k)$$

para cada  $r = 1, 2, 3, \dots$ . La idea del método de los momentos consiste en igualar los primeros momentos poblacionales, que no sean constantes, a los correspondientes momentos muestrales:

$$a_r = \frac{1}{n} \sum_{i=1}^n X_i^r,$$

en número suficiente para obtener de tales ecuaciones una solución para  $\theta_1, \dots, \theta_k$  en función de los  $a_r$  y, por tanto, en función de  $X_1, X_2, \dots, X_n$ .

Así, si el sistema de ecuaciones

$$\begin{cases} \alpha_r(\theta_1, \theta_2, \dots, \theta_k) = a_r \\ r = 1, 2, \dots, k \end{cases}$$

tiene alguna solución:  $\tilde{\theta}_1(X_1, \dots, X_n), \dots, \tilde{\theta}_k(X_1, \dots, X_n)$ , se denomina a  $\tilde{\theta}_j(X_1, \dots, X_n)$  estimador de  $\theta_j$  por el método de los momentos. Ocasionalmente puede ocurrir que los momentos que haya que igualar no sean los  $k$  primeros.

A menudo los resultados que se obtienen son razonables:

### Ejemplo 7.1

En el caso de una población  $N(\mu, \sigma^2)$ , de la que se dispone de una muestra aleatoria simple de tamaño  $n$ , para obtener los estimadores de  $\mu$  y  $\sigma^2$  basta resolver el sistema

$$\begin{cases} \mu = \bar{X} \\ \sigma^2 + \mu^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 \end{cases}$$

cuya solución es  $\tilde{\mu} = \bar{X}$  y  $\tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 = s^2$ .

Lo cual muestra que los estimadores obtenidos pueden no ser centrados.

### Ejemplo 7.2

Si  $X_1, \dots, X_n$  es una muestra aleatoria simple de una distribución de Poisson de parámetro  $\lambda$ , el método de los momentos consiste en una única ecuación

$$\lambda = \bar{X}$$

de manera que  $\tilde{\lambda} = \bar{X}$  es el estimador de  $\lambda$  por el método de los momentos, coincidiendo con el ECUMV para  $\lambda$  (véase ejemplo 6.5).

En cambio en otras ocasiones los resultados del método son difícilmente aceptables:

### Ejemplo 7.3

Si  $X_1, \dots, X_n$  es una muestra aleatoria simple de una distribución uniforme en  $(0, \theta)$ , el estimador de  $\theta$  por el método de los momentos, solución de la ecuación  $\theta/2 = \bar{X}$ , es

$$\tilde{\theta} = 2\bar{X},$$

que es centrado, pero podría dar una estimación de  $\theta$  menor que algunas de las observaciones muestrales.

Si se tratase de una población uniforme en  $(-\theta, \theta)$ , como su media es cero, habría que igualar su momento de segundo orden,  $\theta^2/3$ , con el momento muestral  $a_2$ , para obtener el estimador

$$\tilde{\theta} = \sqrt{\frac{3}{n} \sum_{i=1}^n X_i^2},$$

que no es función del estadístico suficiente  $\max_{1 \leq i \leq n} |X_i|$ .

En resumidas cuentas, el defecto del método de los momentos es que carece de una justificación seria —ligada a consideraciones de suficiencia u optimalidad. Si en ocasiones conduce a buenos resultados es casi por casualidad.

La consistencia es la única propiedad de que gozan, con cierta generalidad, los estimadores construidos por el método de los momentos. En efecto, si es posible expresar  $\theta_i = g_i(\alpha_1, \dots, \alpha_k)$ , mediante una función  $g_i$  continua, el estimador de  $\theta_i$  por el método de los momentos será

$$\tilde{\theta}_i = g_i(a_1, \dots, a_k)$$

que, cuando  $n$  crece, converge casi seguro a  $\theta_i = g_i(\alpha_1, \dots, \alpha_k)$ , puesto que  $a_r$  converge casi seguro a  $\alpha_r$  (véase Cap. 2 §6).

## 7.3. El método de la máxima verosimilitud

El método de la máxima verosimilitud está basado en un principio natural que puede ilustrarse inicialmente con un ejemplo.

Imaginemos que una urna contiene 6 bolas, entre blancas y negras, no todas del mismo color, pero se ignora cuantas hay de cada uno. En tal caso la proporción de bolas blancas puede ser

$$p = \frac{1}{6}, \frac{2}{6}, \frac{3}{6}, \frac{4}{6} \text{ ó } \frac{5}{6}.$$

Se permite extraer dos bolas con reemplazamiento, para tratar de adivinar la composición de la urna. Evidentemente el número de bolas blancas obtenidas puede ser  $T = 0, 1$  ó  $2$  y la siguiente tabla proporciona la probabilidad de cada resultado, en función del valor del parámetro  $p$ :

$p$	$T = 0$	$T = 1$	$T = 2$
$\frac{1}{6}$	$(\frac{5}{6})^2 = 0'694$	$2\frac{1}{6}\frac{5}{6} = 0'277$	$(\frac{1}{6})^2 = 0'027$
$\frac{2}{6}$	$(\frac{4}{6})^2 = 0'444$	$2\frac{2}{6}\frac{4}{6} = 0'444$	$(\frac{2}{6})^2 = 0'111$
$\frac{3}{6}$	$(\frac{1}{2})^2 = 0'25$	$2\frac{1}{2}\frac{1}{2} = 0'5$	$(\frac{3}{6})^2 = 0'25$
$\frac{4}{6}$	$(\frac{2}{6})^2 = 0'111$	$2\frac{4}{6}\frac{2}{6} = 0'444$	$(\frac{4}{6})^2 = 0'444$
$\frac{5}{6}$	$(\frac{1}{6})^2 = 0'027$	$2\frac{5}{6}\frac{1}{6} = 0'277$	$(\frac{5}{6})^2 = 0'694$

Si resultan  $T = 0$  bolas blancas, podría ser que fuese  $p = 4/6$  y que se haya producido un suceso de probabilidad  $0'111$ ; pero es más verosímil que sea  $p = 2/6$  y que haya ocurrido un suceso de probabilidad  $0'444$ ; y todavía más verosímil que sea  $p = 1/6$  y que se haya producido un suceso de probabilidad  $0'694$ . Si hubiese que apostar acerca de cuántas bolas blancas hay en la urna, lo razonable es hacerlo a favor de que hay una.

De la misma manera, si se obtiene  $T = 1$  —una bola blanca en la muestra— el valor de  $p$  que otorga mayor probabilidad al suceso observado es  $p = 1/2$ . Y si el número de bolas blancas obtenidas es  $T = 2$ , es el valor de  $p = 5/6$  el que hace más verosímil el resultado observado.

En definitiva, es razonable utilizar como estimador del valor de  $p$  la función:

$$\hat{p}(T) = \begin{cases} 1/6 & \text{si } T = 0, \\ 1/2 & \text{si } T = 1, \\ 5/6 & \text{si } T = 2. \end{cases}$$

Generalizando ligeramente la situación anterior, imaginemos que realizamos 25 observaciones, en cada una de las cuales puede suceder un suceso  $A$  de probabilidad  $p \in [0, 1]$ , desconocida. Si  $A$  ocurre, por ejemplo, 6 veces, se ha producido un suceso de probabilidad

$$\binom{25}{6} p^6 (1-p)^{19}.$$

Un pequeño análisis de esta función de  $p$  muestra que tiene un máximo para  $p = 6/25$ . De manera que el resultado obtenido es tanto más probable cuanto más próximo sea  $p$  al valor  $6/25$ . Es razonable, por tanto, tomar  $\hat{p} = 6/25$  como estimación de  $p$ .

La misma idea puede emplearse en múltiples situaciones y se conoce como *principio de la máxima verosimilitud*. En primera aproximación, podría enunciarse en la forma: Si al observar una muestra aleatoria simple de una población con distribución teórica  $F_\theta$ , se obtienen valores  $(x_1, x_2, \dots, x_n)$  cuya probabilidad es

$$\mathcal{P}_\theta\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\}$$

debe estimarse  $\theta$  mediante el valor que maximiza dicha probabilidad.

Evidentemente el enunciado anterior carece de sentido en el caso en que la distribución poblacional  $F_\theta$  sea continua, ya que será

$$\mathcal{P}_\theta(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = 0$$

para cualesquiera valores de  $(x_1, x_2, \dots, x_n)$ . Pero es entonces natural reemplazar la función de probabilidad por la función de densidad y dar la siguiente formulación general:

*Sea  $(X_1, X_2, \dots, X_n)$  una muestra aleatoria simple<sup>1</sup> de una población cuya distribución pertenece a  $\mathcal{F} = \{F_\theta | \theta \in \Theta\}$ .*

*Según que la población sea discreta o continua, la función de probabilidad o la función de densidad de la muestra:  $f_\theta(x_1, \dots, x_n)$ , se denomina función de verosimilitud, al considerarla como función de  $\theta$  para valores fijos de la muestra.*

*Un estimador  $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$  se denomina estimador de máxima verosimilitud de  $\theta$  si*

$$f_{\hat{\theta}}(x_1, x_2, \dots, x_n) = \max_{\theta \in \Theta} f_\theta(x_1, x_2, \dots, x_n)$$

*para cada  $(x_1, x_2, \dots, x_n) \in \mathcal{X}$ .*

---

<sup>1</sup> La hipótesis de que el muestreo aleatorio es simple no es en realidad necesaria.

Puesto que la función logaritmo es monótona creciente, resulta a menudo más cómodo trabajar con  $\log f_\theta$  en lugar de  $f_\theta$ . Es decir buscar el estimador de máxima verosimilitud para que verifique

$$\log f_\theta(x_1, x_2, \dots, x_n) = \max_{\theta \in \Theta} \log f_\theta(x_1, x_2, \dots, x_n).$$

Si  $f_\theta$  es una función derivable respecto a  $\theta$  en el interior del espacio paramétrico  $\Theta$ , la forma usual de determinar el estimador de máxima verosimilitud es examinar primero los máximos relativos de  $\log f_\theta$ , para compararlos, después, con los valores sobre la frontera de  $\Theta$ . Ello conduce a resolver las ecuaciones de verosimilitud:

$$\begin{cases} \frac{\partial}{\partial \theta_j} \log f_\theta(x_1, x_2, \dots, x_n) = 0 \\ j = 1, 2, \dots, k \end{cases}$$

(en el supuesto de que  $\theta = (\theta_1, \theta_2, \dots, \theta_k)$  sea un parámetro  $k$ -dimensional), seleccionando las soluciones correspondientes a un máximo de  $\log f_\theta$ .

El procedimiento es sencillo de aplicar en muchas situaciones:

#### Ejemplo 7.4

La función de verosimilitud de una muestra aleatoria simple, de tamaño  $n$ , de una población  $N(\mu, \sigma)$ , es

$$f_{\mu, \sigma}(x_1, x_2, \dots, x_n) = \left( \frac{1}{2\pi} \right)^{n/2} \frac{1}{\sigma^n} e^{-\sum_{i=1}^n (x_i - \mu)^2 / 2\sigma^2},$$

con lo cual

$$\log f_{\mu, \sigma}(x_1, x_2, \dots, x_n) = -\frac{n}{2} \log(2\pi) - n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2.$$

Las ecuaciones de verosimilitud son entonces

$$\begin{cases} \frac{\partial}{\partial \mu} \log f_{\mu, \sigma}(x_1, x_2, \dots, x_n) = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0, \\ \frac{\partial}{\partial \sigma} \log f_{\mu, \sigma}(x_1, x_2, \dots, x_n) = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - \mu)^2 = 0, \end{cases}$$

cuya solución es

$$\left\{ \begin{array}{l} \hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}, \\ \hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} = s. \end{array} \right.$$

Fácilmente se comprueba que dicha solución corresponde a un máximo relativo. Además, en las proximidades de la frontera  $\sigma = 0$  del espacio paramétrico,  $f_{\mu, \sigma}$  se acerca a 0. Luego los estimadores de máxima verosimilitud de  $\mu$  y  $\sigma$  son  $\hat{\mu} = \bar{X}$  y  $\hat{\sigma} = s$ .

### Ejemplo 7.5

La situación considerada en la introducción de esta sección es aquella en que se dispone de una muestra aleatoria simple de tamaño  $n$  de la población

$$X = \begin{cases} 1 & \text{con probabilidad } p, \\ 0 & \text{con probabilidad } 1 - p, \end{cases}$$

cuya función de probabilidad se expresa

$$f_p(x) = p^x (1 - p)^{1-x} \quad \text{para } x = 0, 1,$$

siendo  $p \in [0, 1]$  desconocido.

Por consiguiente la función de verosimilitud es

$$f_p(x_1, x_2, \dots, x_n) = p^{x_1 + \dots + x_n} (1 - p)^{n - (x_1 + \dots + x_n)}$$

de forma que

$$\log f_p(x_1, x_2, \dots, x_n) = \sum_{i=1}^n x_i \log p + \left( n - \sum_{i=1}^n x_i \right) \log(1 - p).$$

La ecuación de verosimilitud es entonces

$$\frac{\partial}{\partial p} \log f_p(x_1, x_2, \dots, x_n) = \frac{1}{p} \sum_{i=1}^n x_i - \frac{1}{1-p} \left( n - \sum_{i=1}^n x_i \right) = 0,$$

cuya solución es

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x},$$

que expresa la proporción de unos observados en la muestra y proporciona el estimador de máxima verosimilitud de  $p$ , puesto que en la frontera del espacio paramétrico,  $p = 0$  ó  $p = 1$ , es  $f_p(x_1, \dots, x_n) = 0$ .

### Ejemplo 7.6

Generalizando la situación anterior, supongamos que los elementos de una población pueden ser clasificados, según cierta característica, en una de  $k$  categorías:  $1, 2, \dots, k$ . Las proporciones  $p_1, p_2, \dots, p_k$  de cada grupo son desconocidas, pero (si abarcan todas las posibilidades) cumplirán  $p_k = 1 - p_1 - p_2 - \dots - p_{k-1}$ .

Al clasificar  $n$  elementos de la población, tomados al azar e independientemente, la probabilidad de obtener los resultados  $(x_1, x_2, \dots, x_n)$ , siendo cada  $x_i \in \{1, 2, \dots, k\}$ , es

$$f_{p_1, p_2, \dots, p_{k-1}}(x_1, x_2, \dots, x_n) = p_1^{n_1} p_2^{n_2} \cdots p_{k-1}^{n_{k-1}} (1 - p_1 - p_2 - \cdots - p_{k-1})^{n_k},$$

donde  $n_j$  es el número de elementos clasificados en el grupo  $j$  (es decir el número de  $x_i = j$ ). Por consiguiente,

$$\log f_{p_1, p_2, \dots, p_{k-1}}(x_1, x_2, \dots, x_n) = \sum_{i=1}^{k-1} n_i \log p_i + n_k \log(1 - p_1 - p_2 - \cdots - p_{k-1}).$$

Las ecuaciones de verosimilitud son entonces:

$$\begin{cases} \frac{n_j}{p_j} - \frac{n_k}{1 - p_1 - p_2 - \cdots - p_{k-1}} = 0, \\ j = 1, 2, \dots, k-1; \end{cases}$$

cuya solución verifica

$$\frac{n_k}{1 - \hat{p}_1 - \hat{p}_2 - \cdots - \hat{p}_{k-1}} = \frac{n_1}{\hat{p}_1} = \frac{n_2}{\hat{p}_2} = \cdots = \frac{n_{k-1}}{\hat{p}_{k-1}} = n$$

(el último término se obtiene sumando los numeradores y los denominadores de los anteriores). La solución es por tanto

$$\hat{p}_j = \frac{n_j}{n},$$

que representa la proporción muestral de cada grupo.

La matriz de las derivadas segundas:

$$-\begin{pmatrix} n_k/p_k^2 + n_1/p_1^2 & n_k/p_k^2 & n_k/p_k^2 \\ n_k/p_k^2 & n_k/p_k^2 + n_2/p_2^2 & n_k/p_k^2 \\ \vdots & \vdots & \vdots \\ n_k/p_k^2 & n_k/p_k^2 & n_k/p_k^2 + n_{k-1}/p_{k-1}^2 \end{pmatrix}$$

es indudablemente definida negativa. La solución hallada corresponde, pues, a un máximo relativo; que es también máximo absoluto puesto que, sobre la frontera del espacio paramétrico

$$\Theta = \{(p_1, p_2, \dots, p_k) \mid p_j \geq 0, \sum_{j=1}^k p_j = 1\}$$

alguno de los  $p_j$  se anula y el logaritmo de la verosimilitud se acerca a  $-\infty$ .

En ocasiones (y principalmente cuando  $\{x \mid f_\theta(x) > 0\}$  depende de  $\theta$ )  $f_\theta(x_1, x_2, \dots, x_n)$  no es derivable y hay que recurrir entonces a un análisis directo, sin el recurso de las ecuaciones de verosimilitud.

### Ejemplo 7.7

Consideremos una población con distribución uniforme en  $[0, \theta]$  cuya función de densidad es

$$f_\theta(x) = \frac{1}{\theta} I_{\{0 \leq x \leq \theta\}},$$

siendo  $\theta > 0$  un parámetro desconocido<sup>2</sup>.

La función de densidad de una muestra aleatoria simple de tamaño  $n$  es entonces

$$f_\theta(x_1, x_2, \dots, x_n) = \frac{1}{\theta^n} I_{\{x_{(1)} \geq 0\}} I_{\{x_{(n)} \leq \theta\}}$$

que no es derivable respecto a  $\theta$ . Sin embargo, es claro que se anula para  $\theta < x_{(n)}$  y decrece a medida que  $\theta$  avanza por el intervalo  $[x_{(n)}, \infty)$ . El valor máximo de  $f_\theta(x_1, x_2, \dots, x_n)$  se alcanza pues en  $\hat{\theta} = x_{(n)}$ . Es decir que el estimador de máxima verosimilitud de  $\theta$  es

$$\hat{\theta} = \max_{1 \leq i \leq n} X_i.$$

---

<sup>2</sup> La población uniforme en  $[0, \theta]$  es idéntica a la uniforme en  $(0, \theta)$ , puesto que sus funciones de distribución coinciden; aunque el intervalo se toma cerrado para evitar anomalías en el análisis, los resultados son los mismos en ambos casos.

Recuérdese que  $E_\theta[X_{(n)}] = \frac{n}{n+1}\theta$ , lo cual pone de relieve que el estimador de máxima verosimilitud no tiene por qué ser centrado.

Un ejemplo similar muestra que el estimador de máxima verosimilitud puede no ser único.

### Ejemplo 7.8

En el caso de una población uniforme en  $[\theta - 1, \theta + 1]$  con función de densidad

$$f_\theta(x) = \frac{1}{2} I_{\{x \in [\theta-1, \theta+1]\}},$$

la función de densidad de una muestra aleatoria simple es

$$f_\theta(x_1, x_2, \dots, x_n) = \frac{1}{2^n} I_{\{x_{(1)} \geq \theta - 1\}} I_{\{x_{(n)} \leq \theta + 1\}};$$

es decir,

$$f_\theta(x_1, x_2, \dots, x_n) = \frac{1}{2^n} I_{\{x_{(n)} - 1 \leq \theta \leq x_{(1)} + 1\}}.$$

Resulta evidente entonces que es de máxima verosimilitud cualquier estimador  $T$  que verifique  $X_{(n)} - 1 < T < X_{(1)} + 1$ ; por ejemplo

$$T = \alpha(X_{(n)} - 1) + (1 - \alpha)(X_{(1)} + 1)$$

para cualquier  $\alpha \in [0, 1]$ . La selección del estadístico tendrá entonces que recurrir a otro criterio, puesto que el principio de máxima verosimilitud no precisa el resultado.

No es difícil darse cuenta de que *el estimador de máxima verosimilitud es función de cualquier estadístico suficiente* (a no ser que no fuese único, en cuyo caso se puede escoger uno que lo sea).

De hecho, si  $T(x_1, x_2, \dots, x_n)$  es un estadístico suficiente, según el teorema de factorización, se podrá expresar:

$$f_\theta(x_1, x_2, \dots, x_n) = g_\theta(T(x_1, x_2, \dots, x_n)) h(x_1, x_2, \dots, x_n).$$

Por consiguiente, si dos muestras  $(x_1, x_2, \dots, x_n)$  y  $(x'_1, x'_2, \dots, x'_n)$  dan el mismo valor  $t$  del estadístico  $T$ , los valores de  $\theta$  que maximizan

$$g_\theta(t)h(x_1, x_2, \dots, x_n) \quad \text{y} \quad g_\theta(t)h(x'_1, x'_2, \dots, x'_n)$$

son el mismo (o se pueden escoger iguales, si hay varios) y no depende más que de  $t$ .

Ello no significa que el estimador de máxima verosimilitud tenga que ser suficiente; he aquí la prueba:

### Ejemplo 7.9

Para una población uniforme en  $[\theta, 2\theta]$ , de densidad

$$f_\theta(x) = \frac{1}{\theta} I_{\{\theta \leq x \leq 2\theta\}},$$

la densidad de la muestra aleatoria simple será

$$f_\theta(x_1, x_2, \dots, x_n) = \frac{1}{\theta^n} I_{\{x_{(1)} \geq \theta\}} I_{\{x_{(n)} \leq 2\theta\}}$$

o bien

$$f_\theta(x_1, x_2, \dots, x_n) = \frac{1}{\theta^n} I_{\{x_{(n)}/2 \leq \theta \leq x_{(1)}\}};$$

lo cual muestra que el estadístico suficiente minimal es

$$(X_{(1)}, X_{(n)}).$$

Sin embargo,  $f_\theta(x_1, x_2, \dots, x_n)$  sólo es no nula cuando  $\theta \in [x_{(n)}/2, x_{(1)}]$  y decrece en dicho intervalo. Luego el estimador de máxima verosimilitud es

$$\hat{\theta} = \frac{1}{2} X_{(n)},$$

que no es suficiente.

En el caso de que  $\theta$  sea unidimensional, existe también una relación clara de los estimadores de máxima verosimilitud con los estimadores eficientes. Concretamente, supongamos que se cumplen las condiciones de regularidad introducidas para la acotación de Frechet-Cramér-Rao y que  $T$  es un estadístico centrado y eficiente para  $\theta$ . Vimos entonces que ha de ser ( $\mathcal{P}_\theta$  casi seguro)

$$T(x_1, x_2, \dots, x_n) = \theta + \frac{1}{I(\theta)} \frac{\partial}{\partial \theta} \log f_\theta(x_1, x_2, \dots, x_n)$$

(siendo  $I(\theta) > 0$  la cantidad de información de Fisher). Es decir

$$\frac{\partial}{\partial \theta} \log f_\theta(x_1, x_2, \dots, x_n) = I(\theta) (T(x_1, x_2, \dots, x_n) - \theta)$$

de manera que la ecuación de verosimilitud tiene como única solución

$$\hat{\theta} = T(x_1, x_2, \dots, x_n),$$

que corresponde a un máximo absoluto, puesto que  $\frac{\partial}{\partial \theta} \log f_\theta(x_1, \dots, x_n)$  es positiva para  $\theta < \hat{\theta}$  y negativa para  $\theta > \hat{\theta}$ . Así pues bajo las condiciones de regularidad impuestas, *si existe un estimador centrado y eficiente para  $\theta$ , es el único estimador de máxima verosimilitud.*

Esta confluencia (parcial) de los estimadores centrados de mínimo *ECM* con los estimadores de máxima verosimilitud, resulta tanto más curiosa si se observa que ambos criterios son radicalmente diferentes, en el siguiente sentido: En el primer caso, al anular el sesgo e intentar minimizar la dispersión, ambas condiciones expresan buenas propiedades “promedio” del estimador, al variar la muestra entre todas las muestras posibles. En cambio, el principio de máxima verosimilitud actúa sobre cada muestra concreta, sin preocuparse, en cada caso, de cuál sería la estimación de  $\theta$  si la muestra hubiese sido otra. Resulta agradable que, al deducir el estimador con mejores propiedades “promedio”, se obtenga (en buenas condiciones) el estimador que opta por la solución “más verosímil” para cada muestra concreta. Esta idea quedará reforzada, en la próxima sección, al poner de manifiesto que los estimadores de máxima verosimilitud tienen una alta eficiencia para tamaños muestrales grandes.

Una propiedad adicional de los estimadores de máxima verosimilitud es su invariancia respecto a transformaciones biunívocas del parámetro. Esto es, si  $\omega = h(\theta)$ , siendo  $h$  una función biyectiva, los estimadores de máxima verosimilitud de  $\omega$  y  $\theta$  están relacionados por la ecuación

$$\hat{\omega} = h(\hat{\theta}).$$

De hecho,

$$g_\omega(x_1, \dots, x_n) = f_{h^{-1}(\omega)}(x_1, \dots, x_n)$$

representa la función de probabilidad o de densidad de la muestra, expresada en términos del parámetro  $\omega$  en lugar del parámetro  $\theta$ . Si  $\hat{\theta}$  es el estimador de máxima verosimilitud de  $\theta$  se cumplirá para cualquier valor de  $\theta$

$$f_{\hat{\theta}}(x_1, \dots, x_n) \geq f_\theta(x_1, \dots, x_n)$$

y, por consiguiente,

$$g_{h(\hat{\theta})}(x_1, \dots, x_n) \geq g_{h(\theta)}(x_1, \dots, x_n).$$

Luego

$$g_{h(\hat{\theta})}(x_1, \dots, x_n) \geq g_\omega(x_1, \dots, x_n)$$

para cualquier  $\omega$ ; lo cual establece que  $\hat{\omega} = h(\hat{\theta})$  es el estimador de máxima verosimilitud para  $\omega$ .

Tomando como base la propiedad anterior, el principio de máxima verosimilitud se prolonga adoptando  $\hat{\omega} = h(\hat{\theta})$  como estimación de cualquier función  $\omega = h(\theta)$  del parámetro —aún cuando no sea biyectiva. Ello equivale a buscar el valor  $\hat{\omega}$  para el cual el conjunto  $h^{-1}(\omega)$  contiene al valor de  $\theta$  que hace más verosímil la muestra observada (a pesar de que dicho conjunto pudiera contener valores de  $\theta$  que la hiciesen muy poco verosímil).

### Ejemplo 7.10

La duración de cierto tipo de lámparas es exponencial de media  $\theta$  desconocida. Después de observar la vida de  $n$  de ellas, se desea estimar la probabilidad de que la duración de una de tales lámparas sea superior a 500 horas; es decir el valor de  $\omega = e^{-500/\theta}$ .

Naturalmente, la función de densidad de la muestra es

$$f_\theta(x_1, x_2, \dots, x_n) = \frac{1}{\theta^n} \exp -\frac{1}{\theta} \sum_{i=1}^n x_i;$$

con lo cual

$$\log f_\theta(x_1, x_2, \dots, x_n) = -n \log \theta - \frac{1}{\theta} \sum_{i=1}^n x_i$$

y

$$\frac{\partial}{\partial \theta} \log f_\theta(x_1, x_2, \dots, x_n) = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^n x_i.$$

Igualando a cero, se obtiene por solución:

$$\hat{\theta} = \sum_{i=1}^n x_i / n = \bar{x},$$

que constituye el estimador de máxima verosimilitud de  $\theta$ , puesto que la derivada anterior es positiva si  $\theta \in (0, \hat{\theta})$  y negativa si  $\theta > \hat{\theta}$ .

La estimación de máxima verosimilitud de  $\omega = e^{-500/\theta}$  es entonces

$$\hat{\omega} = e^{-500/\hat{\theta}} = e^{-500/\bar{x}},$$

como podría comprobarse repitiendo los cálculos con la densidad poblacional reparametrizada

$$g_\omega(x) = \frac{-\log \omega}{500} \omega^{x/500}.$$

Si lo que se desea estimar fuese  $\omega = [\theta]$  —la parte entera de la vida media de las lámparas— ya no es posible reparametrizar la densidad poblacional en términos de  $\omega$ , pero parece forzoso dar como estimador de máxima verosimilitud:  $\hat{\omega} = [\hat{\theta}]$ .

## 7.4. Propiedades asintóticas de los estimadores de máxima verosimilitud

Los estimadores de máxima verosimilitud, basados en una muestra aleatoria simple de tamaño  $n$ , tienen al crecer  $n$  algunas propiedades importantes.

Para mostrarlo, consideraremos una población con función de densidad (o de probabilidad)  $f_\theta(x)$ , de la cual se obtiene una muestra aleatoria simple de tamaño  $n$ , cada vez mayor.  $\mathcal{P}_\theta$  representará entonces la distribución de la sucesión muestral:  $X_1, X_2, \dots, X_n, \dots$ , cuyas  $n$  primeras variables tienen función de densidad (o de probabilidad):

$$f_\theta(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f_\theta(x_i).$$

Supondremos además, explícitamente, que el valor del parámetro identifica la distribución; es decir que no existen  $\theta$  y  $\theta'$ , tales que  $f_\theta$  y  $f_{\theta'}$  coincidan. Ello es necesario, por ejemplo, para obtener resultados acerca de la consistencia de la sucesión de estimadores de máxima verosimilitud, del tipo siguiente:

**Teorema 7.1:** *Sea el espacio paramétrico  $\Theta$  un intervalo abierto de  $\mathbb{R}$ , en el cual  $\log f_\theta(x)$  es derivable con respecto a  $\theta$ . Para cualquier  $\theta \in \Theta$ , con probabilidad  $\mathcal{P}_\theta$  igual a uno, existe una sucesión,  $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$  de raíces de la ecuación de verosimilitud:*

$$\sum_{i=1}^n \frac{\partial}{\partial \theta} \log f_\theta(x_i) = 0,$$

tal que

$$\hat{\theta}_n \rightarrow \theta \quad \text{cuando } n \rightarrow \infty.$$

En efecto, fijado  $\theta \in \Theta$ , es

$$\mathbb{E}_\theta \left[ \frac{f_{\theta'}(X)}{f_\theta(X)} \right] = \int_{\mathbb{R}} f_\theta(x) \frac{f_{\theta'}(x)}{f_\theta(x)} dx = 1.$$

Como

$$\log \frac{f_{\theta'}(x)}{f_\theta(x)} < \frac{f_{\theta'}(x)}{f_\theta(x)} - 1,$$

salvo si  $f_{\theta'}(x) = f_\theta(x)$ , será

$$\mathbb{E}_\theta \left[ \log \frac{f_{\theta'}(X)}{f_\theta(X)} \right] < \mathbb{E}_\theta \left[ \frac{f_{\theta'}(X)}{f_\theta(X)} - 1 \right] = 0.$$

La ley fuerte de los grandes números aplicada a la sucesión de variables aleatorias  $\log(f_{\theta'}(X_i)/f_\theta(X_i))$ , asegura que con probabilidad  $\mathcal{P}_\theta$  igual a 1, a partir de un  $n$  (que puede depender de  $\theta'$ ) en adelante, se cumplirá

$$\frac{1}{n} \sum_{i=1}^n \log \frac{f_{\theta'}(X_i)}{f_\theta(X_i)} < 0,$$

con lo cual

$$\sum_{i=1}^n \log f_\theta(X_i) > \sum_{i=1}^n \log f_{\theta'}(X_i).$$

Tomemos en particular,  $\theta' = \theta \pm \delta$ . Puesto que  $\log f_\theta(x)$  es una función continua, se concluye que  $\sum_{i=1}^n \log f_\theta(X_i)$  tiene un máximo relativo, en el intervalo  $(\theta - \delta, \theta + \delta)$  y, por tanto, existe en dicho intervalo una solución,  $\hat{\theta}_n$ , de la ecuación de verosimilitud. Dicho explícitamente, para todo  $\delta > 0$ ,

$$\mathcal{P}_\theta \{ \text{Existe } n_\delta \text{ tal que, si } n \geq n_\delta \text{ hay una raíz } \hat{\theta}_n \text{ con } |\hat{\theta}_n - \theta| < \delta \} = 1.$$

El teorema anterior tiene interés exclusivamente en el caso en que la ecuación de verosimilitud tenga una raíz única, que ha de converger entonces a  $\theta$ ,  $\mathcal{P}_\theta$  casi seguro, para cada  $\theta \in \Theta$ . En caso contrario, la elección a realizar de  $\hat{\theta}_n$  entre las distintas raíces podría depender de  $\theta$  y no se conseguiría una sucesión de estimadores que sea consistente para  $\theta$ . La consistencia del estimador de máxima verosimilitud suele ser sencilla de probar, una vez que se ha conseguido determinarlo en cada situación concreta. Por ejemplo, ello puede establecerse fácilmente en los ejemplos 7.4, 7.5, 7.6 y 7.7. El resultado anterior no es por tanto excesivamente valioso<sup>3</sup>; pero sirve de prólogo al

---

<sup>3</sup> Resultados más precisos pueden verse en Schmetterer (1974) Cap. V, §3.

siguiente teorema que proporciona la distribución límite del estimador de máxima verosimilitud, bajo ciertas condiciones de regularidad:

**Teorema 7.2:** Si  $\Theta$  es un intervalo abierto de  $\mathbb{R}$  y la función de densidad (o de probabilidad)  $f_\theta(x)$  de la población verifica:

- 1) Existe  $\frac{\partial^3}{\partial \theta^3} \log f_\theta(x)$ , cuyo valor absoluto está acotado por una función  $K(x)$  tal que  $E_\theta[K(X)] \leq k$ .
- 2)  $E_\theta \left[ \frac{\partial}{\partial \theta} \log f_\theta(X) \right] = 0, \quad E_\theta \left[ \frac{1}{f_\theta(X)} \frac{\partial^2}{\partial \theta^2} f_\theta(X) \right] = 0.$
- 3)  $i(\theta) = E_\theta \left[ \left( \frac{\partial}{\partial \theta} \log f_\theta(X) \right)^2 \right] > 0.$

Entonces, cualquier sucesión  $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \dots, X_n)$  de raíces de la ecuación de verosimilitud, que sea consistente para  $\theta$ , verifica que

$$\sqrt{n}(\hat{\theta}_n - \theta) \text{ converge en distribución a una } N(0, \sqrt{i(\theta)^{-1}}),$$

supuesto que los elementos muestrales provienen de la distribución poblacional, asociada al valor  $\theta$  del parámetro, y cualquiera que sea  $\theta \in \Theta$ .

Como observaciones previas a la demostración, nótese que  $i(\theta)$  es la información de Fisher, introducida en relación con la acotación de Frechet-Cramér-Rao, correspondiente a una muestra de tamaño 1. En virtud de las observaciones allí realizadas (cf. pág. 231), es

$$E_\theta \left[ \frac{\partial^2}{\partial \theta^2} \log f_\theta(X) \right] = -i(\theta),$$

supuesto que se cumple la hipótesis (2). Acerca de esta hipótesis cabe señalar que se verifica, siempre que la identidad  $\int_{\mathbb{R}} f_\theta(x) dx = 1$  pueda derivarse dos veces bajo el signo de integración; de manera que se cumplirá en situaciones suficientemente regulares (concretamente siempre que las dos primeras derivadas respecto a  $\theta$  de  $f_\theta$  estén acotadas por sendas funciones integrables).

Por su parte, la conclusión del teorema es que la sucesión de estimadores de máxima verosimilitud, supuesto que se obtienen como raíz única de la ecuación de verosimilitud (y son por tanto una sucesión consistente), tienen asintóticamente distribución  $N(\theta, \sqrt{(ni(\theta))^{-1}})$  siendo  $\theta$  el valor del

parámetro —sea cual sea— del que proviene la muestra. En este sentido, la distribución de  $\hat{\theta}_n$  es, para  $n$  grande, aproximadamente normal y cada vez más concentrada alrededor del valor del parámetro con el que se está produciendo la muestra.

La prueba del teorema es la siguiente.

Demostración: Representemos abreviadamente

$$u(\theta) = \frac{\partial}{\partial \theta} \log f_\theta(X_1, X_2, \dots, X_n) = \sum_{i=1}^n \frac{\partial}{\partial \theta} \log f_\theta(X_i).$$

Supuesto que la sucesión muestral tiene distribución  $\mathcal{P}_\theta$ , asociada al valor  $\theta$  del parámetro, la aplicación del teorema central del límite a la suma del último miembro (habida cuenta que  $E[\partial \log f_\theta(X)/\partial \theta] = 0$ ) prueba que

$$\frac{1}{\sqrt{n}} u(\theta) \xrightarrow{d_\theta} N(0, \sqrt{i(\theta)}) \quad \text{cuando } n \rightarrow \infty,$$

siendo la convergencia en distribución (que se denota por  $d_\theta$  para recordar que corresponde a suponer que las  $X_i$  provienen de la distribución asociada al valor  $\theta$  del parámetro).

Ahora bien, puesto que  $\hat{\theta}_n$  es solución de la ecuación de verosimilitud será

$$0 = u(\hat{\theta}_n) = u(\theta) + u'(\theta)(\hat{\theta}_n - \theta) + u''(\theta_1) \frac{(\hat{\theta}_n - \theta)^2}{2},$$

haciendo el desarrollo de Taylor de  $u(\theta)$  hasta el segundo término.

De ahí que

$$\sqrt{n}(\hat{\theta}_n - \theta) = \frac{-u(\theta)/\sqrt{n}}{\left[u'(\theta) + u''(\theta_1)(\hat{\theta}_n - \theta)/2\right]/n}$$

o bien

$$\sqrt{n}(\hat{\theta}_n - \theta)i(\theta) - \frac{1}{\sqrt{n}}u(\theta) = -\frac{u(\theta)}{\sqrt{n}}(\xi_n + 1),$$

siendo

$$\xi_n = \frac{i(\theta)}{\left[u'(\theta) + u''(\theta_1)(\hat{\theta}_n - \theta)/2\right]/n}.$$

En virtud de la ley fuerte de los grandes números,

$$\frac{u'(\theta)}{n} = \frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} \log f_\theta(X_i) \longrightarrow -i(\theta) \quad \mathcal{P}_\theta \text{ casi seguro};$$

mientras que, por hipótesis,

$$\hat{\theta}_n - \theta \xrightarrow{\mathcal{P}_{\theta}} 0.$$

Así que, utilizando de nuevo la ley de los grandes números,

$$\frac{|u''(\theta_1)|}{n} \leq \frac{1}{n} \sum_{i=1}^n \left| \frac{\partial^3}{\partial \theta^3} \log f_{\theta}(X_i) \right|_{\theta=\theta_1} \leq \frac{1}{n} \sum_{i=1}^n K(X_i) < k + \varepsilon,$$

a partir de un  $n$  en adelante, con probabilidad  $\mathcal{P}_{\theta}$  igual a 1.

En definitiva, el denominador de  $\xi_n$  converge en probabilidad ( $\mathcal{P}_{\theta}$ ) a  $-i(\theta)$ , de forma que  $\xi_n + 1$  lo hace hacia 0 y se concluye que

$$\sqrt{n}(\hat{\theta}_n - \theta)i(\theta) - \frac{1}{\sqrt{n}}u(\theta) \xrightarrow{\mathcal{P}_{\theta}} 0.$$

La distribución límite de  $\sqrt{n}(\hat{\theta}_n - \theta)I(\theta)$  y de  $u(\theta)/\sqrt{n}$  es, por consiguiente la misma. Es decir,

$$\sqrt{n}(\hat{\theta}_n - \theta)i(\theta) \xrightarrow{d_{\theta}} N\left(0, \sqrt{i(\theta)}\right),$$

o bien

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d_{\theta}} N\left(0, \sqrt{i(\theta)^{-1}}\right).$$

### Ejemplo 7.11

Se dispone de una muestra aleatoria simple, de tamaño  $n$ , de una población  $\gamma(k, \theta)$ , siendo  $k > 0$  una constante conocida y  $\theta$  un parámetro desconocido. La densidad de la muestra es entonces

$$f_{\theta}(x_1, \dots, x_n) = \frac{\theta^{nk}}{\Gamma(k)^n} (x_1 \cdots x_n)^{k-1} e^{-\theta(x_1 + \cdots + x_n)}$$

para  $x_1, \dots, x_n > 0$ . Por tanto

$$\log f_{\theta}(x_1, \dots, x_n) = nk \log \theta - n \log \Gamma(k) + (k-1) \sum_{i=1}^n \log x_i - \theta \sum_{i=1}^n x_i.$$

Como

$$\frac{\partial}{\partial \theta} \log f_{\theta}(x_1, \dots, x_n) = \frac{nk}{\theta} - \sum_{i=1}^n x_i$$

es positivo para  $\theta \in (0, k/\bar{x})$  y negativo para  $\theta \in (k/\bar{x}, \infty)$ , el estimador de máxima verosimilitud de  $\theta$  es

$$\hat{\theta} = \frac{k}{\bar{X}}.$$

La hipótesis (2) se verifica, ya que

$$\frac{\partial}{\partial \theta} \log f_\theta(X) = \frac{k}{\theta} - X$$

tiene esperanza 0; al igual que

$$\begin{aligned} \frac{1}{f_\theta(X)} \frac{\partial^2}{\partial \theta^2} f_\theta(X) &= \frac{\partial^2}{\partial \theta^2} \log f_\theta(X) + \left( \frac{\partial}{\partial \theta} \log f_\theta(X) \right)^2 \\ &= -\frac{k}{\theta^2} + \left( \frac{k}{\theta} - X \right)^2. \end{aligned}$$

Además,

$$i(\theta) = \frac{k}{\theta^2},$$

de forma que se cumple (3).

Si se sabe que es  $\theta > 1$  —por ejemplo—

$$\frac{\partial^3}{\partial \theta^3} \log f_\theta(x) = \frac{2k}{\theta^3}$$

está acotado por  $2k$ , se cumple (1) y el teorema permite concluir que

$$\sqrt{n} \left( \frac{k}{\bar{X}} - \theta \right) \xrightarrow{d_\theta} N \left( 0, \frac{\theta}{\sqrt{k}} \right);$$

o, dicho de otra manera,  $\hat{\theta}$  es asintóticamente  $N \left( \theta, \theta / \sqrt{nk} \right)$ .

En cambio, la hipótesis (1) no se verifica en el caso en que el espacio paramétrico es  $(0, \infty)$ ; pero un razonamiento directo permite, de todas maneras, establecer la misma conclusión. En efecto, según el teorema central del límite,

$$\sqrt{n} \left( \bar{X} - \frac{k}{\theta} \right) \xrightarrow{d_\theta} N \left( 0, \frac{\sqrt{k}}{\theta} \right).$$

Pero

$$\sqrt{n} \left( \frac{k}{\bar{X}} - \theta \right) = -\sqrt{n} \left( \bar{X} - \frac{k}{\theta} \right) \cdot \frac{\theta}{\bar{X}},$$

y mientras el primer factor converge en distribución a  $N(0, \sqrt{k}/\theta)$ , el segundo factor converge casi seguro a  $\theta^2/k$ . Luego

$$\sqrt{n} \left( \frac{k}{\bar{X}} - \theta \right) \xrightarrow{d_\theta} N \left( 0, \frac{\theta}{\sqrt{k}} \right).$$

Como se ve, la validez de la conclusión del teorema es más amplia de lo que indican sus hipótesis. Con gran frecuencia, es más fácil y preciso razonar directamente, en cada caso concreto, que verificar las hipótesis del enunciado (que son suficientes pero, en modo alguno, necesarias).

De todas maneras, algún tipo de regularidad es necesaria para la normalidad asintótica de los estimadores de máxima verosimilitud:

### Ejemplo 7.12

Según el ejemplo 7.7, el estimador de máxima verosimilitud para el parámetro de una uniforme en  $[0, \theta]$  es  $\hat{\theta} = \max_{1 \leq i \leq n} X_i$ , cuya distribución en el muestreo es

$$P\{\hat{\theta}_n \leq x\} = \begin{cases} 0 & \text{si } x < 0, \\ (x/\theta)^n & \text{si } 0 \leq x < \theta, \\ 1 & \text{si } x \geq \theta. \end{cases}$$

Es claro entonces que

$$P\{n(\hat{\theta}_n - \theta) \leq x\} = P\left\{\hat{\theta}_n \leq \theta + \frac{x}{n}\right\} = \begin{cases} 0 & \text{si } x < -n\theta \\ \left(1 + \frac{x}{n\theta}\right)^n & \text{si } -n\theta \leq x < \theta \\ 1 & \text{si } x \geq 0 \end{cases}$$

converge hacia la distribución

$$\begin{cases} e^{x/\theta} & \text{si } x < 0, \\ 1 & \text{si } x \geq 0. \end{cases}$$

Ello indica que, en este caso,  $n(\hat{\theta}_n - \theta)$  es asintóticamente el opuesto de una variable con distribución exponencial de parámetro  $1/\theta$ . Lo cual no deja ninguna oportunidad para que  $\sqrt{n}(\hat{\theta}_n - \theta)$  pueda ser asintóticamente normal (sino que convergerá a 0 casi seguro).

En aquellos casos en que la distribución exacta de  $\hat{\theta}_n$  sea difícil de determinar, conocer su distribución asintótica permite, para valores grandes de  $n$ ,

utilizarlos en la determinación aproximada de intervalos de confianza para el parámetro, analizar sus propiedades como estimadores puntuales (sesgo, error cuadrático medio, ...), etc.

En este último sentido, el que los estimadores de máxima verosimilitud tengan, con mucha frecuencia, distribución asintóticamente normal, de varianza  $1/(ni(\theta))$ , ha inducido a dar la siguiente definición general:

*Una sucesión de estimadores  $T_n$ , asociada a los sucesivos tamaños muestrales  $n$ , se denomina asintóticamente normal y consistente para  $\theta$  si existe una sucesión  $\sigma_n(\theta) > 0$ , convergente a 0, cuando  $n \rightarrow \infty$ , tal que*

$$\frac{T_n - \theta}{\sigma_n(\theta)} \xrightarrow{d_\theta} N(0, 1);$$

*o dicho de otra manera  $T_n \simeq N(\theta, \sigma_n(\theta))$  para  $n$  grande.*

Es inmediato deducir de la definición anterior la consistencia de  $T_n$ . En cambio, no es cierto en general que haya de ser:  $E_\theta[T_n] \rightarrow \theta$ , ni tampoco  $V_\theta(T_n)/\sigma_n(\theta) \rightarrow 1$ . Podrían incluso diseñarse situaciones en que sea  $V_\theta(T_n) = \infty$ , pero  $\sigma_n(\theta)$  converja rápidamente a cero.

Debido a ello, a la hora de medir la eficiencia asintótica de una sucesión  $T_n$  asintóticamente normal y consistente para  $\theta$ , se utiliza el comportamiento límite de

$$\frac{1/(ni(\theta))}{\sigma_n^2(\theta)},$$

en lugar del límite de

$$\frac{\left(\frac{\partial}{\partial \theta} E_\theta[T_n]\right)^2 / (ni(\theta))}{V_\theta(T_n)}.$$

Ello corresponde a sustituir en la definición de eficiencia (cf. §6.3), los momentos de  $T_n$  por los momentos de su distribución límite<sup>4</sup>.

Puesto que en el caso de  $\hat{\theta}_n$ , bajo condiciones de regularidad, se verifica  $\sigma_n^2(\theta) = 1/(ni(\theta))$ , resulta que los estimadores de máxima verosimilitud tienen eficiencia asintótica 1.

Durante mucho tiempo se pensó que no podían existir sucesiones de estimadores asintóticamente normales y consistentes, con eficiencia asintótica

---

<sup>4</sup> La cantidad de información de la muestra es  $I(\theta) = ni(\theta)$ , puesto que suponemos que el muestreo es aleatorio simple.

superior a 1 —conjetura enunciada por primera vez por Fisher en 1925— lo cual representaba una importante propiedad de optimalidad de los estimadores máximo verosímiles, para tamaños muestrales grandes. Sin embargo, más adelante, se comprobó que sucesiones de estimadores muy ineficientes para tamaños muestrales pequeños, pueden tener eficiencia asintótica superior a uno, para algunos valores del parámetro —fenómeno que se conoce como *supereficiencia*. De hecho la eficiencia asintótica de cualquier sucesión asintóticamente normal y consistente de estimadores puede mejorarse (por un procedimiento estandar<sup>5</sup>) para cualquier valor concreto del parámetro, de manera a obtener eficiencia mayor que 1. Ello no tiene ninguna importancia práctica (puesto que se trata de valores paramétricos individuales) pero muestra que no tiene sentido buscar una sucesión de estimadores asintóticamente más eficiente que cualquier otra.

El teorema 2 puede generalizarse (mediante un razonamiento análogo) al caso de parámetros  $k$ -dimensionales:  $\theta = (\theta_1, \dots, \theta_k)$ . Bajo condiciones de regularidad semejantes, cualquier sucesión consistente de soluciones de las ecuaciones de verosimilitud, verifica:

$$\sqrt{n} (\hat{\theta}_n - \theta) \xrightarrow{d_\theta} N_k (0, i(\theta)^{-1}).$$

Así que el vector  $\hat{\theta}_n$  es asintóticamente como una normal ( $k$ -dimensional) de vector de medias  $\theta$  y matriz de covarianzas  $i(\theta)^{-1}$ ; siendo  $i(\theta)$  la matriz de información de Fisher correspondiente a una muestra de tamaño 1.

### Ejemplo 7.13

Imaginemos que para calibrar un aparato que mide simultáneamente el ancho y el largo de ciertas piezas, se lleva a cabo  $n$  veces la medida de una pieza patrón. Los errores  $(X, Y)$  cometidos tienen una distribución normal bidimensional de medias cero, cuya matriz de covarianzas se desea estimar a partir de las  $n$  observaciones realizadas.

Para abreviar, utilicemos la notación

$$a_{20} = \frac{1}{n} \sum_{i=1}^n x_i^2, \quad a_{02} = \frac{1}{n} \sum_{i=1}^n y_i^2, \quad a_{11} = \frac{1}{n} \sum_{i=1}^n x_i y_i.$$

En función de las desviaciones típicas  $\sigma_1$  y  $\sigma_2$  de  $X$  e  $Y$  y del coeficiente de

---

<sup>5</sup> Cf. Schmetterer (1974) Cap. V, §4.

correlación  $\rho$  entre ambas, la función de verosimilitud es

$$f = \frac{1}{(2\pi\sigma_1\sigma_2\sqrt{1-\rho^2})^n} \exp \left\{ -\frac{n}{2(1-\rho^2)} \left[ \frac{a_{20}}{\sigma_1^2} - \frac{2\rho a_{11}}{\sigma_1\sigma_2} + \frac{a_{02}}{\sigma_2^2} \right] \right\}.$$

Con lo cual

$$\log f = -n \log(2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}) - \frac{n}{2(1-\rho^2)} \left[ \frac{a_{20}}{\sigma_1^2} - \frac{2\rho a_{11}}{\sigma_1\sigma_2} + \frac{a_{02}}{\sigma_2^2} \right].$$

Las ecuaciones de verosimilitud son entonces

$$\begin{cases} \frac{\partial}{\partial\sigma_1} f = -\frac{n}{\sigma_1} + \frac{na_{20}}{(1-\rho^2)\sigma_1^3} - \frac{n\rho a_{11}}{(1-\rho^2)\sigma_1^2\sigma_2} = 0, \\ \frac{\partial}{\partial\sigma_2} f = -\frac{n}{\sigma_2} + \frac{na_{02}}{(1-\rho^2)\sigma_2^3} - \frac{n\rho a_{11}}{(1-\rho^2)\sigma_1\sigma_2^2} = 0, \\ \frac{\partial}{\partial\rho} f = \frac{n\rho}{(1-\rho^2)} - \frac{na_{20}\rho}{(1-\rho^2)^2\sigma_1^2} - \frac{na_{02}\rho}{(1-\rho^2)^2\sigma_2^2} + \frac{na_{11}(1+\rho^2)}{(1-\rho^2)^2\sigma_1\sigma_2} = 0. \end{cases}$$

Las dos primeras proporcionan

$$\frac{a_{20}}{\sigma_1^2} = 1 - \rho^2 + \rho \frac{a_{11}}{\sigma_1\sigma_2}, \quad \frac{a_{02}}{\sigma_2^2} = 1 - \rho^2 + \rho \frac{a_{11}}{\sigma_1\sigma_2},$$

lo cual sustituido en la tercera ecuación, puesta en la forma

$$(1 - \rho^2)\rho - \rho \frac{a_{20}}{\sigma_1^2} - \rho \frac{a_{02}}{\sigma_2^2} + (1 + \rho^2) \frac{a_{11}}{\sigma_1\sigma_2} = 0,$$

da lugar a

$$(1 - \rho^2) \left( \frac{a_{11}}{\sigma_1\sigma_2} - \rho \right) = 0$$

y únicamente la raíz  $\rho = a_{11}/\sigma_1\sigma_2$  corresponde a un máximo. Con dicho valor de  $\rho$  se obtiene

$$\hat{\sigma}_1 = \sqrt{a_{20}}, \quad \hat{\sigma}_2 = \sqrt{a_{02}}, \quad \text{y} \quad \hat{\rho} = \frac{a_{11}}{\sqrt{a_{20}a_{02}}},$$

como estimadores de máxima verosimilitud de los tres parámetros.

Hay que calcular ahora la matriz de información para una muestra de tamaño  $n = 1$ :

$$-\frac{\partial^2}{\partial\sigma_1^2} f = -\frac{1}{\sigma_1^2} + \frac{3X^2}{(1-\rho^2)\sigma_1^4} - \frac{2\rho XY}{(1-\rho^2)\sigma_1^3\sigma_2}$$

cuya esperanza es

$$i_{11} = -\frac{1}{\sigma_1^2} + \frac{3}{(1-\rho^2)\sigma_1^2} - \frac{2\rho^2}{(1-\rho^2)\sigma_1^2} = \frac{2-\rho^2}{(1-\rho^2)\sigma_1^2}.$$

Simétricamente

$$i_{22} = \frac{2-\rho^2}{(1-\rho^2)\sigma_2^2}.$$

De la misma manera,

$$-\frac{\partial^2}{\partial\rho^2}f = -\frac{1+\rho^2}{(1-\rho^2)^2} + \left(\frac{X^2}{\sigma_1^2} + \frac{Y^2}{\sigma_2^2}\right) \frac{1+2\rho^2-3\rho^4}{(1-\rho^2)^4} - \frac{\rho XY}{\sigma_1\sigma_2} \frac{6-4\rho^2-2\rho^4}{(1-\rho^2)^4},$$

con lo cual

$$i_{33} = -\frac{1+\rho^2}{(1-\rho^2)^2} + 2 \frac{1+2\rho^2-3\rho^4}{(1-\rho^2)^4} - \rho^2 \frac{6-4\rho^2-2\rho^4}{(1-\rho^2)^4} = \frac{1+\rho^2}{(1-\rho^2)^2}.$$

Por otra parte,

$$-\frac{\partial^2}{\partial\sigma_1\partial\sigma_2}f = -\frac{\rho XY}{(1-\rho^2)\sigma_1^2\sigma_2^2},$$

cuya esperanza es

$$i_{12} = -\frac{\rho^2}{(1-\rho^2)\sigma_1\sigma_2}.$$

Por último

$$-\frac{\partial^2}{\partial\sigma_1\partial\rho}f = -\frac{X^2}{\sigma_1^3} \frac{2\rho}{(1-\rho^2)^2} + \frac{XY}{\sigma_1^2\sigma_2} \frac{1+\rho^2}{(1-\rho^2)^2},$$

da lugar a

$$i_{13} = -\frac{2\rho}{(1-\rho^2)^2\sigma_1} + \frac{\rho}{\sigma_1} \frac{1+\rho^2}{(1-\rho^2)^2} = \frac{-\rho}{(1-\rho^2)\sigma_1}$$

y simétricamente

$$i_{23} = \frac{-\rho}{(1-\rho^2)\sigma_2}.$$

En definitiva la matriz de información es

$$i = \frac{1}{1-\rho^2} \begin{pmatrix} (2-\rho^2)/\sigma_1^2 & -\rho^2/\sigma_1\sigma_2 & -\rho/\sigma_1 \\ -\rho^2/\sigma_1\sigma_2 & (2-\rho^2)/\sigma_2^2 & -\rho/\sigma_2 \\ -\rho/\sigma_1 & -\rho/\sigma_2 & (1+\rho^2)/(1-\rho^2) \end{pmatrix}.$$

Con un esfuerzo más, puede comprobarse que

$$i^{-1} = \frac{1}{2} \begin{pmatrix} \sigma_1^2 & \rho^2\sigma_1\sigma_2 & \rho(1-\rho^2)\sigma_1 \\ \rho^2\sigma_1\sigma_2 & \sigma_2^2 & \rho(1-\rho^2)\sigma_2 \\ \rho(1-\rho^2)\sigma_1 & \rho(1-\rho^2)\sigma_2 & 2(1-\rho^2)^2 \end{pmatrix},$$

lo cual proporciona la matriz de covarianzas de la distribución asintótica, normal tridimensional, de  $(\hat{\sigma}_1, \hat{\sigma}_2, \hat{\rho})$ .

## 7.5. Estimación Bayesiana

En el capítulo 5, el criterio Bayes apareció como un método para resumir el riesgo de cada estimador, asociado a una función de pérdida dada, en una única cantidad, obtenida promediando respecto a una distribución *a priori*  $\pi(\theta)$ , de manera a poder ordenar los estimadores y seleccionar el mejor. Sin embargo, la metodología Bayesiana puede utilizarse también sin considerar ni función de pérdida ni, por tanto, el riesgo asociado a cada estimador.

Lo distintivo del método es, entonces, la existencia de una distribución *a priori* del parámetro, que consideraremos caracterizada mediante su densidad  $\pi(\theta)$ , y que en combinación con la función de densidad (o de probabilidad)  $f_\theta(x)$  de la población, permite determinar la distribución *a posteriori* del parámetro, una vez conocida la muestra:

$$\pi(\theta | x_1, x_2, \dots, x_n) = \frac{\pi(\theta) f_\theta(x_1, x_2, \dots, x_n)}{\int_{\Theta} f_\theta(x_1, x_2, \dots, x_n) \pi(\theta) d\theta},$$

tal y como se indicó en la sección 5.5.2.

Obviamente, con este planteamiento, la distribución *a posteriori* contiene toda la información disponible sobre el parámetro, una vez que se ha incorporado a la información *a priori*, la evidencia proporcionada por la muestra. Cada posible valor  $\theta$  del parámetro tiene una probabilidad, de ser aquél con el cual se ha producido la muestra, perfectamente calibrada mediante  $\pi(\theta | x_1, x_2, \dots, x_n)$ .

A la vista de la distribución *a posteriori*, si hay obligación de dar una estimación puntual del valor del parámetro, el único trámite a realizar es elegir una única característica numérica de dicha distribución que nos parezca la más representativa de toda ella. La elección puede recaer sobre la media *a posteriori* (y estaríamos entonces minimizando implícitamente el riesgo a

posteriori asociado a la pérdida  $(\theta - t)^2$ . O puede recaer sobre la mediana a posteriori (que proporciona el mínimo riesgo asociado a la pérdida  $|\theta - t|$ ).

Sin embargo, también es razonable elegir como pronóstico acerca del valor del parámetro, la *moda de la distribución a posteriori*. Es decir aquel valor  $\theta^*$  tal que

$$\pi(\theta^* | x_1, x_2, \dots, x_n) = \sup_{\theta \in \Theta} \pi(\theta | x_1, x_2, \dots, x_n)$$

Aunque no minimice, en principio, el riesgo asociado a ninguna función de pérdida, al fin y al cabo,  $\theta^*$  es el valor del parámetro que más probabilidad tiene una vez observada la muestra. En este sentido, tomar como estimador de  $\theta$  la moda de la distribución a posteriori, tiene en el contexto Bayesiano una justificación del mismo tipo que la utilización del estimador de máxima verosimilitud en el contexto no Bayesiano.

De hecho, en ausencia de información a priori más precisa, una distribución a priori uniforme ( $\pi(\theta) = \text{cte}$ ) sobre los valores del parámetro, hace que la distribución a posteriori sea proporcional a la función de densidad (o de probabilidad) de la muestra. En cuyo caso, la moda a posteriori coincidirá con el estimador de máxima verosimilitud.

### Ejemplo 7.14

En la situación del ejemplo 5.20, en la que se observa una muestra aleatoria simple de la distribución uniforme en  $(0, \theta)$ , siendo  $\theta$  un número elegido con densidad  $\pi(\theta) = (k - 1)/\theta^k$  en el intervalo  $(1, \infty)$ , se obtenía como densidad a posteriori de  $\theta$ :

$$\pi(\theta | x_1, x_2, \dots, x_n) = \frac{(n + k - 1)x_{(n)}^{n+k-1}}{\theta^{n+k}} I_{\{\theta > x_{(n)}\}}.$$

La densidad a posteriori se anula para  $\theta < x_{(n)}$  y es decreciente en el intervalo  $(x_{(n)}, \infty)$ . La estimación bayesiana de  $\theta$ , es decir la moda de la distribución a posteriori, es entonces:

$$\theta^* = x_{(n)}$$

coincidiendo con el estimador de máxima verosimilitud y ligeramente diferente de los estimadores Bayes obtenidos con las funciones de pérdida utilizadas en el ejemplo 5.20.

### Ejemplo 7.15

De una población con distribución de Poisson de parámetro  $\theta$  desconocido, se sabe que el parámetro  $\theta$  se ha elegido con distribución  $\gamma(p, \lambda)$ , y se desea estimar  $\theta$  a partir de una muestra aleatoria simple de tamaño  $n$ .

La función de probabilidad de la muestra es naturalmente

$$f_\theta(x_1, x_2, \dots, x_n) = e^{-n\theta} \frac{\theta^{n\bar{x}}}{x_1! \cdots x_n!},$$

con  $x_i = 0, 1, 2, \dots$ ; y la densidad a priori de  $\theta$  es

$$\pi(\theta) = \frac{\lambda^p}{\Gamma(p)} \theta^{p-1} e^{-\lambda\theta}$$

para  $\theta > 0$ . Por tanto,

$$\begin{aligned} f(\theta, x_1, x_2, \dots, x_n) &= \frac{\lambda^p}{\Gamma(p)} \theta^{p-1} e^{-\lambda\theta} e^{-n\theta} \frac{\theta^{n\bar{x}}}{x_1! \cdots x_n!} \\ &= \frac{\lambda^p}{\Gamma(p)} \theta^{n\bar{x}+p-1} e^{-(\lambda+n)\theta} \frac{1}{x_1! \cdots x_n!} \end{aligned}$$

expresa conjuntamente la densidad de probabilidad de cada valor  $\theta > 0$  y la probabilidad de cada muestra posible de números naturales.

La función de probabilidad marginal de la muestra será entonces

$$\begin{aligned} f(x_1, x_2, \dots, x_n) &= \frac{\lambda^p}{\Gamma(p)} \frac{1}{x_1! \cdots x_n!} \int_0^\infty \theta^{n\bar{x}+p-1} e^{-(\lambda+n)\theta} d\theta \\ &= \frac{\lambda^p}{\Gamma(p)} \frac{1}{x_1! \cdots x_n!} \frac{\Gamma(n\bar{x} + p)}{(\lambda + n)^{n\bar{x}+p}}, \end{aligned}$$

con lo cual la densidad a posteriori de  $\theta$  resulta

$$\pi(\theta | x_1, x_2, \dots, x_n) = \frac{f(\theta, x_1, x_2, \dots, x_n)}{f(x_1, x_2, \dots, x_n)} = \frac{(\lambda + n)^{n\bar{x}+p}}{\Gamma(n\bar{x} + p)} \theta^{n\bar{x}+p-1} e^{-(\lambda+n)\theta}$$

para  $\theta > 0$ , que corresponde a una distribución  $\gamma(n\bar{x} + p, \lambda + n)$ .

La media a posteriori, estimador Bayes con pérdida cuadrática, es entonces

$$E[\theta | X_1, X_2, \dots, X_n] = \frac{n\bar{X} + p}{\lambda + n} = \frac{\bar{X} + p/n}{1 + \lambda/n}.$$

Mientras que la moda a posteriori —el valor más probable con la densidad  $\gamma(n\bar{x} + p, \lambda + n)$ — es el valor de  $\theta$  que maximiza  $(n\bar{X} + p - 1) \log \theta - (\lambda + n)\theta$ ; es decir

$$\theta^* = \frac{n\bar{X} + p - 1}{\lambda + n} = \frac{\bar{X} + (p - 1)/n}{1 + \lambda/n}.$$

Para  $n$  grande, ambos estimadores son sensiblemente iguales entre sí; y muy similares al estimador de máxima verosimilitud  $\hat{\theta} = \bar{X}$ .

## 7.6. Estimación mínimo cuadrática

Con gran frecuencia se presentan, en muchos campos científicos o técnicos, situaciones en las que se observa una magnitud que depende de ciertas condiciones experimentales y de ciertos parámetros que se desean estimar. Concretamente:

$$X = \varphi(t, \theta)$$

donde  $t$  representa las diversas condiciones experimentales,  $\theta \in I\!\!R^k$  representa los parámetros desconocidos,  $\varphi$  es una función dada y  $X$  es el efecto que se observa.

Si las observaciones se producen sin error, una vez realizado un número suficiente de observaciones de  $X$ , la determinación de los parámetros conduce al problema algebraico de resolver un sistema de ecuaciones, en las incógnitas:  $\theta_1, \theta_2, \dots, \theta_k$ .

Sin embargo, usualmente se cometerán errores aleatorios de medida, de forma que, con más precisión, el resultado de la medición será

$$X = \varphi(t, \theta) + \varepsilon,$$

siendo  $\varepsilon$  una variable aleatoria sobre la cual —al menos en primera aproximación— puede suponerse que tiene una distribución que no depende de  $t$  ni de  $\theta$ . La ausencia de errores sistemáticos, más fáciles de corregir en la práctica, permite suponer que los errores no tienen sesgo; es decir que  $E[\varepsilon] = 0$ ; así como que los errores en diferentes medidas son independientes.

En estas condiciones el problema adquiere, además de su componente algebraica, una clara componente estadística. Así, después de realizar un número suficientemente grande de observaciones,  $X_i$ , en condiciones experimentales prefijadas,  $t_i$ , podríamos dar a los parámetros  $\theta_j$  valores arbitrarios y echarle la culpa a los errores,  $\varepsilon_i$ , cometidos en cada observación, de las discrepancias entre las observaciones y los valores  $\varphi(t_i, \theta)$ .

Por supuesto, ello no es en absoluto razonable. Lo sensato es buscar aquellos valores de los parámetros que hagan que el vector

$$(\varphi(t_1, \theta), \varphi(t_2, \theta), \dots, \varphi(t_n, \theta))$$

sea lo más próximo posible al vector de las observaciones

$$(X_1, X_2, \dots, X_n),$$

de forma que el primero constituya la mejor previsión posible del segundo.

La distancia entre dos vectores suele medirse por su norma euclídea y, así, la estimación *mínimo cuadrática* de los parámetros consiste en determinarlos a fin de hacer mínimo:

$$\sum_{i=1}^n [X_i - \varphi(t_i, \theta)]^2.$$

Si la función  $\varphi(t, \theta)$  es derivable respecto a  $\theta$ , ello conduce a resolver el sistema de ecuaciones:

$$\left\{ \begin{array}{l} \sum_{i=1}^n [X_i - \varphi(t_i, \theta)] \frac{\partial}{\partial \theta_j} \varphi(t_i, \theta) = 0 \\ j = 1, 2, \dots, k. \end{array} \right.$$

### Ejemplo 7.16

Supongamos que se desea estimar la aceleración ( $\theta$ ) de la gravedad en una determinada localidad. Para ello se deja caer un objeto, durante intervalos de tiempo prefijados  $t_1, t_2, \dots, t_n$ , midiéndose los espacios recorridos  $X_1, X_2, \dots, X_n$ , con errores de medida independientes y de media 0:  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ . Es decir:

$$X_i = \frac{1}{2} \theta t_i^2 + \varepsilon_i.$$

El estimador *mínimo cuadrático* de  $\theta$  es entonces aquel valor que minimice

$$\sum_{i=1}^n [X_i - \frac{1}{2} \theta t_i^2]^2,$$

que se obtendrá cuando sea

$$\sum_{i=1}^n [X_i - \frac{1}{2} \theta t_i^2] t_i^2 = 0.$$

Por consiguiente el estimador *mínimo cuadrático* es

$$\check{\theta} = \frac{2 \sum_{i=1}^n X_i t_i^2}{\sum_{i=1}^n t_i^4}.$$

Muy a menudo, la distribución de los errores podrá suponerse  $N(0, \sigma)$ . De hecho, la distribución normal fue introducida por primera vez por Gauss en problemas relativos a errores de medida. En tal caso, como cada  $X_i$

tendrá distribución  $N(\varphi(t_i, \theta), \sigma)$ , la función de densidad conjunta de las observaciones será

$$f_\theta(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n [X_i - \varphi(t_i, \theta)]^2 \right\},$$

de manera que para aplicar el principio de máxima verosimilitud, habrá que maximizar

$$\log f_\theta(x_1, x_2, \dots, x_n) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n [X_i - \varphi(t_i, \theta)]^2.$$

Queda claro entonces que, *bajo hipótesis de normalidad de los errores, los estimadores de máxima verosimilitud coinciden con los estimadores mínimo cuadráticos.*

## 7.7. Conclusión

A lo largo de los tres últimos capítulos se han introducido un buen número de criterios, enfoques y métodos para la selección de estimadores puntuales de los parámetros desconocidos de una población.

Sin duda no puede afirmarse categóricamente que ningún procedimiento tenga una validez universal y sea siempre preferible a los demás. El método a emplear depende de demasiados factores, en alguna medida subjetivos, para que pueda adoptarse una posición dogmática acerca de los problemas de inferencia.

Por ejemplo, no es lo mismo hacer la estimación de un parámetro una vez en la vida, que montar un proceso de estimación para que una empresa lo utilice todos los días. En un caso el riesgo puede tener poca importancia y en el otro ser un factor capital, por idéntica razón a la que hace que el beneficio esperado de un juego no tenga demasiada relevancia si sólo se ha de jugar una partida, pero es fundamental si se han de jugar un gran número de ellas.

En presencia de una función de pérdida explícita hay que atenerse a ella, lo cual obliga a emplear criterios de riesgo; aunque, como hemos visto, estos no proporcionan una regla de conducta única, sino que admiten diversas variantes: ECUMV, Pitman; minimax, Bayes. Además, la realidad raramente impone una función de pérdida inapelable.

Si puede disponerse, de manera razonable, de información a priori sobre el parámetro, sería una estupidez no incorporarla, mediante una técnica bayesiana. Pero en determinados contextos puede no ser prudente forzar su aparición. ¿Qué opinaríamos, por ejemplo, si en la estimación del IPC interviniese la opinión a priori del Ministerio de Economía?

Las consideraciones ligadas a la suficiencia son, en cambio, muy asépticas; pero por si solas, no dan una solución al problema. La invariancia puede ser muy relevante en unos casos y muy poco en otros. Etc.

De todas maneras, no conviene considerar los diversos métodos como alternativas que compiten entre sí, sino como técnicas complementarias. A menudo, sobre todo con muestras de gran tamaño, darán resultados muy parecidos, o incluso idénticos; de forma que cada uno estará reforzando a los demás y no desautorizándolos. Sólo si varios de ellos ofreciesen conclusiones muy dispares, sería necesario optar por el que mejor se adapte a las condiciones reales del problema.

## Ejercicios

### Ejercicio 7.1

Sea  $X_1, \dots, X_n$  una muestra aleatoria obtenida de una distribución binomial de parámetros  $k$  y  $p$  desconocidos. Determinar los estimadores de  $k$  y  $p$  por el método de los momentos. Si una muestra de tamaño 3 ha proporcionado los resultados  $x_1 = 6$ ,  $x_2 = 8$  y  $x_3 = 4$ , hallar las estimaciones de máxima verosimilitud de  $k$  y  $p$ .

### Ejercicio 7.2

A partir de una muestra aleatoria simple, de tamaño  $n$ , de una distribución uniforme en el intervalo  $(a, a + b)$ , determinar los estimadores de  $a$  y  $b$ , (a) por el método de los momentos; (b) por el método de la máxima verosimilitud.

### Ejercicio 7.3

Se sabe que el número de descendientes de las hembras de cierta especie de insectos sigue (aproximadamente) una distribución lognormal, con función de densidad

$$f_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left\{-\frac{1}{2\sigma^2}(\log x - \mu)^2\right\} \quad \text{para } x > 0$$

Si 20 hembras, elegidas al azar, tuvieron como descendencia: 47, 34, 28,

44, 23, 32, 39, 27, 36, 33, 28, 32, 29, 35, 30, 37, 41, 26, 52, 31, determinar una estimación de  $\mu$  y  $\sigma$ , (a) por el método de los momentos, (b) por el método de la máxima verosimilitud.

### Ejercicio 7.4

Con objeto de estimar el número desconocido,  $N$ , de peces de un lago, se capturan  $M$  peces, se marcan y se devuelven al lago. A continuación, se capturan  $n$  peces, entre los que aparecen  $m$  marcados. Calcular el estimador de máxima verosimilitud de  $N$ , distinguiendo el caso en que las segundas capturas se hagan (a) con reemplazamiento y (b) sin reemplazamiento.

### Ejercicio 7.5

Se lanza una moneda de probabilidad de cara desconocida,  $\theta$ , hasta que aparece la primera cara. La experiencia se repite  $n$  veces anotando en cada una el lanzamiento en que apareció la primera cara. Basándose en tales resultados:

- Obtener el estimador de  $\theta$  mediante el método de los momentos.
- Obtener un estimador de máxima verosimilitud de  $\theta$  y de la media de la distribución observada.
- Determinar la distribución asintótica del estimador de máxima verosimilitud.
- Supuesto que  $\theta$  se elige con densidad  $a\theta^{a-1}$  en el intervalo  $(0,1)$ , con  $a > 0$  conocido, calcular el estimador bayesiano de  $\theta$ .

### Ejercicio 7.6

Dada una muestra aleatoria simple de tamaño  $n$  de una población con densidad

$$f_\theta(x) = \frac{\theta}{(1+x)^{1+\theta}} \quad \text{para } x > 0.$$

- Determinar el estimador de  $\theta$  por el método de los momentos, supuesto que es  $\theta > 1$ . Obtener su distribución asintótica en el supuesto de que es  $\theta > 2$ .
- Determinar el estimador de máxima verosimilitud de  $\theta$ . Calcular su distribución en el muestreo y obtener su distribución asintótica.
- Comparar la eficiencia asintótica de ambos estimadores.

**Ejercicio 7.7**

Probar que para estimar, a partir de una muestra aleatoria simple de tamaño  $n$ , los parámetros  $\alpha_j$  de cualquier densidad de la forma

$$f(x) = c(\alpha_1, \dots, \alpha_k) e^{-\sum_{j=1}^k \alpha_j x^j}$$

siendo  $\alpha_j > 0$ , el método de máxima verosimilitud coincide con el método de los momentos.

Aplicarlo en la estimación de los parámetros de una distribución normal de media  $-\alpha_1/(2\alpha_2)$  y varianza  $1/(2\alpha_2)$ .

**Ejercicio 7.8**

Se dispone de una muestra aleatoria simple de tamaño  $n$  de la densidad

$$2\theta x e^{-\theta x^2} \quad \text{para } x > 0,$$

donde  $\theta > 0$  es un parámetro desconocido.

- a) Determinar el estimador de  $\theta$  por el método de los momentos y obtener su distribución asintótica.
- b) Determinar el estimador de máxima verosimilitud de  $\theta$  y hallar su distribución asintótica.
- c) Comparar la eficiencia asintótica de ambos estimadores.

**Ejercicio 7.9**

Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de una distribución de densidad

$$f_{\lambda, \theta}(x) = \lambda e^{\lambda(\theta-x)} \quad \text{para } x > \theta.$$

Se supone que  $\lambda$  y  $\theta$  han sido elegidos independientemente con densidades  $\lambda^{-2}$  ( $\lambda > 1$ ) y  $\pi^{-1}(1 + \theta^2)^{-1}$  respectivamente. Determinar las estimaciones bayesianas de  $\theta$  y  $\lambda$ . Compararlas con las estimaciones de máxima verosimilitud de  $\theta$  y  $\lambda$ .

**Ejercicio 7.10**

La cantidad de un producto obtenido en una reacción química viene dada teóricamente por la fórmula

$$c = \beta_0 + \beta_1 x + \beta_2 y,$$

donde  $c$  es la masa obtenida,  $x$  el tiempo empleado e  $y$  la temperatura del medio.

El tiempo y la temperatura pueden fijarse sin error, pero la medida de  $c$  viene afectada de un error  $e$  que es una variable aleatoria de media 0.

Experimentalmente se obtuvieron los datos

$c :$	6	13	13	29'2	33'1	32	46'2	117'5
$x :$	1	2	3	4	5	6	8	20
$y :$	10	10	12	11	14	15	18	30

Determinar la estimación de mínimos cuadrados de  $\beta_0$ ,  $\beta_1$ , y  $\beta_2$ .

Supuesto que  $e$  tiene distribución  $N(0, \sigma)$ , hallar la estimación de máxima verosimilitud de  $\sigma$ .

## Solución de los ejercicios

### Ejercicio 7.1

Si  $X$  tiene distribución binomial  $B(k, p)$  es

$$\mathbb{E}[X] = kp \quad \text{y} \quad \mathbb{E}[X^2] = kp(1 - p) + k^2p^2.$$

El método de los momentos indica que los estimadores de  $k$  y  $p$  se obtienen del sistema de ecuaciones:

$$\begin{cases} kp = \bar{x}, \\ kp(1 - p) + k^2p^2 = \frac{1}{n} \sum_{i=1}^n x_i^2. \end{cases}$$

Sustituyendo la primera ecuación en la segunda, resulta

$$\bar{x}(1 - p) + \bar{x}^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 \quad \text{o bien} \quad \bar{x}(1 - p) = s^2;$$

de donde

$$\tilde{p} = 1 - \frac{s^2}{\bar{x}} \quad \text{y} \quad \tilde{k} = \frac{\bar{x}}{\tilde{p}} = \frac{\bar{x}^2}{\bar{x} - s^2}.$$

Con tamaño muestral  $n = 3$ , la función de verosimilitud de la muestra es

$$f_{k,p}(x_1, x_2, x_3) = \binom{k}{x_1} \binom{k}{x_2} \binom{k}{x_3} p^t (1 - p)^{3k - t}$$

donde  $t = x_1 + x_2 + x_3$ . El valor de  $p$  que maximiza la verosimilitud es fácil de hallar, puesto que

$$\frac{\partial}{\partial p} \log f_{k,p} = \frac{t}{p} - \frac{3k - t}{1 - p}$$

es positiva para  $p < t/(3k) = \bar{x}/k$  y negativa en caso contrario; luego  $\hat{p} = \frac{\bar{x}}{k}$ . Pero la determinación del valor de  $k$  que proporciona el máximo ha de hacerse numéricamente, en cada caso. Con los datos del enunciado,  $\bar{x} = 6$ , así que, después de reemplazar  $\hat{p}$  en  $f_{k,p}$  queda

$$\binom{k}{6} \binom{k}{8} \binom{k}{4} \left(\frac{6}{k}\right)^{18} \left(1 - \frac{6}{k}\right)^{3k-18}$$

que toma los valores:

$k = 8$	$k = 9$	$k = 10$	$k = 11$	$k = 12$
0'00270	0'00327	0'00338	0'00335	0'00329

y decrece después monótonamente (hacia 0'00223 aproximadamente). Por tanto el estimador de máxima verosimilitud de  $k$  vale  $\hat{k} = 10$  y  $\hat{p} = 0'6$ . Naturalmente, no es posible dar una expresión explícita de  $\hat{k}$  en función de las observaciones muestrales.

Como la varianza muestral es  $s^2 = 8/3$ , los estimadores de los momentos valen  $\tilde{p} = 0'555$  y  $\tilde{k} = 10'8$ , para la muestra obtenida.

### Ejercicio 7.2

a) En función de los parámetros indicados, los primeros momentos de la distribución uniforme son

$$\begin{aligned}\mathrm{E}[X] &= \frac{1}{b} \int_a^{a+b} x dx = a + \frac{b}{2}, \\ \mathrm{E}[X^2] &= \frac{1}{b} \int_a^{a+b} x^2 dx = a^2 + ab + \frac{b^2}{3};\end{aligned}$$

de manera que los estimadores de  $a$  y  $b$  por el método de los momentos se obtienen del sistema de ecuaciones:

$$\left\{ \begin{array}{l} a + \frac{b}{2} = \bar{x}, \\ a^2 + ab + \frac{b^2}{3} = \frac{1}{n} \sum_{i=1}^n x_i^2. \end{array} \right.$$

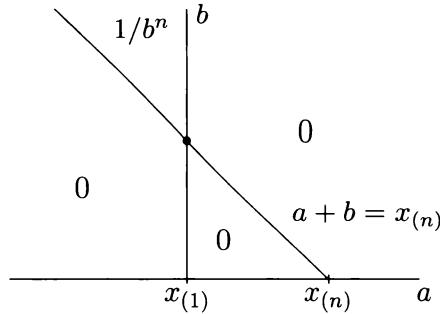
Si se resta el cuadrado de la primera ecuación a la segunda, resulta

$$\frac{b^2}{12} = s^2 \quad \text{de donde} \quad \tilde{b} = 2\sqrt{3}s \quad \text{y} \quad \tilde{a} = \bar{x} - \sqrt{3}s.$$

b) La función de verosimilitud de una muestra de tamaño  $n$  es

$$f_{a,b}(x_1, \dots, x_n) = \frac{1}{b^n} I_{\{a < x_{(1)} \leq x_{(n)} < a+b\}}.$$

El máximo de  $f_{a,b}$  se obtiene, dentro de la región en la que no se anula, tomando  $b$  lo más pequeño posible; es decir, en el punto marcado en el siguiente gráfico:



Los estimadores de máxima verosimilitud de  $a$  y  $b$  son pues

$$\hat{a} = x_{(1)} \quad \text{y} \quad \hat{b} = x_{(n)} - x_{(1)}.$$

### Ejercicio 7.3

a) Puesto que la distribución lognormal es la distribución de  $e^Y$  supuesto que  $Y$  es  $N(\mu, \sigma)$ , la media y el momento de segundo orden de la distribución lognormal se pueden obtener de la función característica  $E[e^{itY}] = \varphi(t) = e^{it\mu - \sigma^2 t^2/2}$  de la distribución  $N(\mu, \sigma)$ . En concreto,

$$E[e^Y] = \varphi\left(\frac{1}{i}\right) = e^{\mu + \sigma^2/2} \quad \text{y} \quad E[e^{2Y}] = \varphi\left(\frac{2}{i}\right) = e^{2\mu + 2\sigma^2}.$$

Luego las ecuaciones del método de los momentos:

$$\begin{cases} e^{\mu + \sigma^2/2} = \bar{x}, \\ e^{2\mu + 2\sigma^2} = s^2 + \bar{x}^2, \end{cases}$$

conducen a los estimadores

$$\tilde{\mu} = 2 \log \bar{x} - \frac{1}{2} \log(s^2 + \bar{x}^2), \quad \tilde{\sigma} = \sqrt{\log(s^2 + \bar{x}^2) - \log \bar{x}^2}.$$

Para la muestra obtenida es  $\bar{x} = 34'2$  y  $s^2 = 52'26$ ; luego los estimadores valen  $\tilde{\mu} = 3'51$  y  $\tilde{\sigma} = 0'21$ .

b) La función de verosimilitud de una muestra de tamaño  $n$

$$f_{\mu, \sigma}(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2} \sigma^n} \prod_{i=1}^n x_i \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (\log x_i - \mu)^2\right\}$$

tiene por derivadas de su logaritmo

$$\begin{cases} \frac{\partial}{\partial \mu} \log f_{\mu, \sigma}(x_1, \dots, x_n) = \frac{1}{\sigma^2} \sum_{i=1}^n (\log x_i - \mu), \\ \frac{\partial}{\partial \sigma} \log f_{\mu, \sigma}(x_1, \dots, x_n) = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (\log x_i - \mu)^2, \end{cases}$$

que, igualadas a cero, proporcionan los estimadores

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \log X_i, \quad \hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^n \left( \log X_i - \frac{1}{n} \sum_{i=1}^n \log X_i \right)^2}.$$

Para la muestra obtenida resultan las estimaciones  $\hat{\mu} = 3'51$  y  $\hat{\sigma} = 0'203$ ; similares a las obtenidas por el método de los momentos.

### Ejercicio 7.4

a) Si las segundas capturas se hacen con reemplazamiento, cada uno de los peces obtenidos tiene probabilidad  $M/N$  de estar marcado y probabilidad  $1 - M/N$  de no estarlo. La verosimilitud de cualquier muestra  $(x_1, x_2, \dots, x_n)$ , con  $x_i = 1$  ó  $0$  según que el  $i$ -ésimo pez esté marcado o no, es

$$f_N(x_1, \dots, x_n) = \binom{n}{m} \left(\frac{M}{N}\right)^m \left(1 - \frac{M}{N}\right)^{n-m},$$

donde  $m = \sum_{i=1}^n x_i$  es el número de peces marcados de la muestra. Por tanto

$$\log f_N = \log \binom{n}{m} + m [\log M - \log N] + (n - m) [\log(N - M) - \log N]$$

y

$$\frac{\partial}{\partial N} \log f_N = -\frac{m}{N} + (n - m) \left[ \frac{1}{N - M} - \frac{1}{N} \right] = -\frac{m}{N} + (n - m) \frac{M}{N(N - M)},$$

que es positiva si  $N < nM/m$  y negativa si  $N > nM/m$ . Así pues la estimación de máxima verosimilitud de  $N$  es aproximadamente

$$\hat{N} = \frac{nM}{m}.$$

Más exactamente, puesto que  $N$  ha de ser un número entero,  $\hat{N}$  puede ser el entero anterior o el entero siguiente a  $nM/m$ , según cual dé un valor mayor a  $(M/N)^m(1 - M/N)^{n-m}$ .

b) Cuando las segundas capturas se hacen sin reemplazamiento, la muestra obtenida no es aleatoria simple (puesto que cada resultado afecta a los siguientes). Sin embargo el número de peces marcados contenidos en la muestra tiene distribución hipergeométrica, de forma que la verosimilitud de cualquier muestra que contenga  $m$  marcados es

$$f_N(m) = \frac{\binom{M}{m} \binom{N-M}{n-m}}{\binom{N}{n}}.$$

Ahora bien, la desigualdad  $f_N(m) > f_{N+1}(m)$  equivale a

$$(N+1)(N+1-M-n+m) > (N+1-M)(N+1-n) \quad \text{o bien} \quad N > \frac{nM}{m} - 1;$$

de forma que  $f_N(m)$  alcanza su máximo en el primer entero que sobrepase el valor  $nM/m - 1$ ; es decir

$$\hat{N} = \left[ \frac{nM}{m} \right],$$

salvo que  $nM/m$  sea entero en cuyo caso  $nM/m$  y  $nM/m - 1$  dan el mismo valor a la verosimilitud.

Como se ve, no hay gran diferencia entre tomar la muestra con o sin reemplazamiento.

### Ejercicio 7.5

a) El lanzamiento en que aparece la primera cara tiene distribución geométrica de función de probabilidad

$$f_\theta(x) = \theta(1 - \theta)^{x-1} \quad \text{para } x = 1, 2, 3, \dots$$

Su media es  $1/\theta$  y, al igualarla con la media muestral  $\bar{x}$ , se obtiene  $\tilde{\theta} = 1/\bar{x}$  como estimador de  $\theta$  por el método de los momentos.

b) La función de verosimilitud de la muestra es

$$f_\theta(x_1, \dots, x_n) = \theta^n (1 - \theta)^{x_1 + \dots + x_n - n}$$

para  $x_1, x_2, \dots, x_n \in \mathbb{N}$ . De forma que

$$\frac{\partial}{\partial \theta} \log f_\theta(x_1, \dots, x_n) = \frac{n}{\theta} - \left( \sum_{i=1}^n x_i - n \right) \frac{1}{1-\theta}$$

es positiva si  $\theta < 1/\bar{x}$  y negativa en caso contrario. Luego  $\hat{\theta} = 1/\bar{x}$  y el estimador de máxima verosimilitud coincide con el obtenido por el método de los momentos.

Para la media de la distribución  $\mu = 1/\theta$ , el estimador de máxima verosimilitud es  $\hat{\mu} = 1/\hat{\theta} = \bar{x}$ .

c) La distribución asintótica de  $\hat{\theta} = 1/\bar{x}$  se puede obtener mediante un razonamiento directo (inspirado en el lema 2.1):

$$\sqrt{n} \left( \frac{1}{\bar{x}} - \theta \right) = \sqrt{n} \left( \bar{x} - \frac{1}{\theta} \right) \cdot \left( -\frac{\theta}{\bar{x}} \right).$$

Como la varianza de la distribución geométrica es  $(1-\theta)/\theta^2$ , según el teorema central del límite, el primer factor converge en distribución a una  $N(0, \sqrt{1-\theta}/\theta)$ ; mientras que el segundo factor converge casi seguro a  $-\theta^2$ , puesto que  $\bar{x}$  converge casi seguro a la media,  $1/\theta$ . Consecuentemente, el primer miembro tiende en distribución a una  $N(0, \theta\sqrt{1-\theta})$  o, dicho de otro modo,

$$\frac{1}{\bar{x}} \quad \text{es asintóticamente } N \left( \theta, \theta \sqrt{\frac{1-\theta}{n}} \right).$$

El razonamiento anterior evita utilizar el teorema 7.2, el cual proporciona también el resultado correcto. En efecto,

$$-\frac{\partial^2}{\partial \theta^2} \log f_\theta(x) = \frac{1}{\theta^2} + \frac{x-1}{(1-\theta)^2};$$

luego

$$i(\theta) = \frac{1}{\theta^2} + \frac{1/\theta - 1}{(1-\theta)^2} = \frac{1}{\theta^2(1-\theta)}$$

y el teorema 7.2 da como distribución asintótica de  $\hat{\theta}$  la distribución normal de media  $\theta$  y varianza  $\theta^2(1-\theta)/n$ .

d) Con la verosimilitud y la distribución a priori indicadas, la densidad marginal de la muestra es

$$f(x_1, \dots, x_n) = \int_0^1 a\theta^{a-1}\theta^n(1-\theta)^{S-n}d\theta = a \frac{\Gamma(n+a)\Gamma(S-n+1)}{\Gamma(S+a+1)};$$

de forma que la densidad a posteriori de  $\theta$

$$\pi(\theta | x_1, \dots, x_n) = \theta^{n+a-1} (1-\theta)^{S-n} \frac{\Gamma(S+a+1)}{\Gamma(n+a)\Gamma(S-n+1)}$$

tiene por logaritmo

$$(n+a-1) \log \theta + (S-n) \log(1-\theta) + \log \frac{\Gamma(S+a+1)}{\Gamma(n+a)\Gamma(S-n+1)},$$

cuya derivada respecto a  $\theta$

$$\frac{\partial}{\partial \theta} \log \pi(\theta | x_1, \dots, x_n) = \frac{n+a-1}{\theta} - \frac{S-n}{1-\theta}$$

es positiva si  $\theta < (n+a-1)/(S+a-1)$  y negativa en caso contrario. La moda de la distribución a posteriori es pues

$$\theta^* = \frac{n+a-1}{S+a-1}.$$

### Ejercicio 7.6

a) Si  $\theta > 1$ , la media poblacional es

$$E_\theta[X] = \int_0^\infty x \theta (1+x)^{-\theta-1} dx = \int_0^\infty (1+x)^{-\theta} dx = \frac{1}{\theta-1}$$

que, igualada a la media muestral,  $\bar{x}$ , da como estimador de  $\theta$  por el método de los momentos:

$$\tilde{\theta} = \frac{1}{\bar{x}} + 1.$$

(Mientras que, para  $\theta < 1$ , la media es infinita y no se puede aplicar el método de los momentos.)

Supuesto que  $\theta > 2$ , la población tiene momento de segundo orden

$$E_\theta[X^2] = \int_0^\infty x^2 \theta (1+x)^{-\theta-1} dx = \frac{2}{(\theta-1)(\theta-2)}$$

y varianza

$$V_\theta(X) = \frac{\theta}{(\theta-1)^2(\theta-2)}.$$

El teorema central del límite asegura entonces que

$$\sqrt{n} \left( \bar{x} - \frac{1}{\theta-1} \right) \xrightarrow{d} N \left( 0, \frac{1}{\theta-1} \sqrt{\frac{\theta}{\theta-2}} \right).$$

Ahora bien

$$\sqrt{n} \left( \frac{1}{\bar{x}} - \theta + 1 \right) = \sqrt{n} \left( \bar{x} - \frac{1}{\theta-1} \right) \left( -\frac{\theta-1}{\bar{x}} \right),$$

donde el segundo factor converge casi seguro hacia  $-(\theta-1)^2$ ; luego

$$\sqrt{n} \left( \frac{1}{\bar{x}} - \theta + 1 \right) \xrightarrow{d} N \left( 0, (\theta-1) \sqrt{\frac{\theta}{\theta-2}} \right);$$

es decir que  $\tilde{\theta}$  tiene distribución asintótica  $N \left( \theta, (\theta-1) \sqrt{\frac{\theta}{n(\theta-2)}} \right)$ .

b) La función de verosimilitud es

$$f_{\theta}(x_1, x_2, \dots, x_n) = \theta^n (1+x_1)^{-\theta-1} (1+x_2)^{-\theta-1} \cdots (1+x_n)^{-\theta-1}$$

para  $x_1, x_2, \dots, x_n > 0$ . Luego

$$\log f_{\theta}(x_1, \dots, x_n) = n \log \theta - (\theta+1) \sum_{i=1}^n \log(1+x_i)$$

y

$$\frac{\partial}{\partial \theta} f_{\theta}(x_1, \dots, x_n) = \frac{n}{\theta} - \sum_{i=1}^n \log(1+x_i);$$

de manera que el estimador de máxima verosimilitud de  $\theta$  es

$$\hat{\theta} = \frac{n}{\sum_{i=1}^n \log(1+X_i)}.$$

Cada  $\log(1+X_i)$  tiene distribución exponencial de parámetro  $\theta$ , luego  $Y = \sum_{i=1}^n \log(1+X_i)$  tiene distribución  $\gamma(n, \theta)$ , de densidad

$$\frac{\theta^n}{(n-1)!} y^{n-1} e^{-\theta y} \quad \text{para } y > 0.$$

Por tanto,  $\hat{\theta} = n/Y$  tiene densidad

$$\frac{n^n \theta^n}{(n-1)!} \frac{e^{-n\theta/z}}{z^{n+1}} \quad \text{para } z > 0.$$

Habida cuenta que  $\log(1+X_i)$  tiene media  $1/\theta$  y varianza  $1/\theta^2$ , el teorema central del límite asegura que

$$\frac{\sum_{i=1}^n \log(1+X_i) - n/\theta}{\sqrt{n}/\theta} \xrightarrow{d} N(0, 1).$$

Pero el primer miembro de la expresión anterior se puede expresar

$$\frac{\sqrt{n}(\theta - \hat{\theta})}{\hat{\theta}}$$

y, según la ley de los grandes números, es

$$\frac{1}{n} \sum_{i=1}^n \log(1 + X_i) \xrightarrow{c.s.} \frac{1}{\theta} \quad \text{o bien} \quad \hat{\theta} \xrightarrow{c.s.} \theta.$$

Luego  $\sqrt{n}(\theta - \hat{\theta})$  converge a una  $N(0, \theta)$ ; es decir que  $\hat{\theta}$  tiene distribución asintótica  $N(\theta, \theta/\sqrt{n})$ .

c) Puesto que la información de Fisher contenida en la muestra es

$$ni(\theta) = -\frac{\partial^2}{\partial \theta^2} \log f_\theta(x_1, \dots, x_n) = \frac{n}{\theta^2},$$

la eficiencia asintótica del estimador obtenido por el método de los momentos es

$$\frac{\theta^2/n}{(\theta - 1)^2 \theta/n(\theta - 2)} = \frac{\theta(\theta - 2)}{(\theta - 1)^2} < 1.$$

Mientras que la eficiencia asintótica del estimador de máxima verosimilitud es 1.

### Ejercicio 7.7

Desde luego  $c(\alpha_1, \dots, \alpha_k) = \left( \int_{\mathbb{R}} e^{-\sum_{j=1}^k \alpha_j x^j} dx \right)^{-1}$ , así que la verosimilitud es

$$f_{\alpha_1, \dots, \alpha_k}(x_1, \dots, x_n) = e^{-\sum_{j=1}^k \alpha_j \sum_{i=1}^n x_i^j} \left( \int_{\mathbb{R}} e^{-\sum_{j=1}^k \alpha_j x^j} dx \right)^{-n};$$

de logaritmo,

$$\log f_{\alpha_1, \dots, \alpha_k}(x_1, \dots, x_n) = - \sum_{j=1}^k \alpha_j \sum_{i=1}^n x_i^j - n \log \int_{\mathbb{R}} e^{-\sum_{j=1}^k \alpha_j x^j} dx.$$

Las ecuaciones de verosimilitud,  $\frac{\partial}{\partial \alpha_j} \log f_{\alpha_1, \dots, \alpha_k} = 0$ , tienen entonces la forma

$$-\sum_{i=1}^n x_i^j + n \frac{\int_{\mathbb{R}} e^{-\sum_{j=1}^k \alpha_j x^j} x^j dx}{\int_{\mathbb{R}} e^{-\sum_{j=1}^k \alpha_j x^j} dx} = 0;$$

es decir,

$$-\sum_{i=1}^n x_i^j + nE[X^j] = 0,$$

o bien  $E[X^j] = \sum_{i=1}^n x_i^j/n$ . Esta última ecuación, iguala el momento de orden  $j = 1, 2, \dots, k$  de la población, con el momento muestral de orden  $j$ , tal y como exige el método de los momentos.

La función de densidad  $N(-\alpha_1/(2\alpha_2), 1/\sqrt{2\alpha_2})$  es

$$f(x) = \frac{\sqrt{\alpha_2}}{\sqrt{\pi}} \exp\left\{-\frac{\alpha_1^2}{4\alpha_2}\right\} \exp\left\{-\alpha_1 x - \alpha_2 x^2\right\},$$

de forma que la verosimilitud se expresa

$$f(x_1, \dots, x_n) = \frac{\alpha_2^{n/2}}{\pi^{n/2}} \exp\left\{-\frac{n\alpha_1^2}{4\alpha_2}\right\} \exp\left\{-\alpha_1 \sum_1^n x_i - \alpha_2 \sum_1^n x_i^2\right\}$$

de logaritmo

$$\frac{n}{2} \log \alpha_2 - \frac{n}{2} \log \pi - \frac{n\alpha_1^2}{4\alpha_2} - \alpha_1 \sum_1^n x_i - \alpha_2 \sum_1^n x_i^2.$$

Las ecuaciones de verosimilitud son entonces

$$\begin{cases} -\frac{2n\alpha_1}{4\alpha_2} - \sum_{i=1}^n x_i = 0 \\ \frac{n}{2\alpha_2} + \frac{n\alpha_1^2}{4\alpha_2^2} - \sum_{i=1}^n x_i^2 = 0 \end{cases} \quad \text{o bien} \quad \begin{cases} -\frac{\alpha_1}{2\alpha_2} = \frac{1}{n} \sum_{i=1}^n x_i \\ \frac{1}{2\alpha_2} + \frac{\alpha_1^2}{4\alpha_2^2} = \frac{1}{n} \sum_{i=1}^n x_i^2 \end{cases}$$

como resulta directamente de igualar los dos primeros momentos respecto al origen, poblacionales y muestrales. En términos de la media y la varianza muestrales, la solución es

$$\hat{\alpha}_2 = \frac{1}{2s^2} \quad \text{y} \quad \hat{\alpha}_1 = -\frac{\bar{x}}{s^2}.$$

### Ejercicio 7.8

a) La media poblacional:

$$E_\theta[X] = 2\theta \int_0^\infty x^2 e^{-\theta x^2} dx = \frac{1}{\sqrt{\theta}} \int_0^\infty y^{1/2} e^{-y} dy = \frac{\sqrt{\pi}}{2\sqrt{\theta}},$$

igualada a la media muestral,  $\bar{x}$ , da como estimador por el método de los momentos:

$$\tilde{\theta} = \frac{\pi}{4\bar{x}^2}.$$

Además

$$E_\theta[X^2] = 2\theta \int_0^\infty x^3 e^{-\theta x^2} dx = \frac{1}{\theta} \quad \text{y} \quad V_\theta(X) = \frac{4 - \pi}{4\theta}.$$

Luego, en virtud del teorema central del límite,

$$\sqrt{n} \left( \bar{x} - \frac{\sqrt{\pi}}{2\sqrt{\theta}} \right) \xrightarrow{d} N \left( 0, \frac{\sqrt{4 - \pi}}{2\sqrt{\theta}} \right).$$

Según el lema 2.1 (con  $g(x) = \pi/(4x^2)$ ) será entonces

$$\sqrt{n} \left( \frac{\pi}{4\bar{x}^2} - \theta \right) \xrightarrow{d} N \left( 0, \frac{\sqrt{4 - \pi}}{2\sqrt{\theta}} \frac{\pi}{2(\sqrt{\pi}/2\sqrt{\theta})^3} \right) = N \left( 0, 2\theta \frac{\sqrt{4 - \pi}}{\sqrt{\pi}} \right),$$

de forma que  $\tilde{\theta}$  es asintóticamente normal de media  $\theta$  y desviación típica  $2\theta\sqrt{(4 - \pi)/(n\pi)}$ .

b) Para  $x_1, \dots, x_n > 0$ , la función de verosimilitud es

$$f_\theta(x_1, \dots, x_n) = 2^n \theta^n x_1 \cdots x_n \exp \left\{ -\theta \sum_{i=1}^n x_i^2 \right\},$$

con lo cual

$$\frac{\partial}{\partial \theta} \log f_\theta(x_1, \dots, x_n) = \frac{n}{\theta} - \sum_{i=1}^n x_i^2$$

y el estimador de máxima verosimilitud de  $\theta$  es

$$\hat{\theta} = \frac{n}{\sum_{i=1}^n X_i^2}.$$

Ahora bien

$$E_\theta[X^4] = 2\theta \int_0^\infty x^5 e^{-\theta x^2} dx = \frac{2}{\theta^2} \quad \text{y} \quad V_\theta(X^2) = \frac{1}{\theta^2};$$

así que el teorema central del límite asegura que es asintóticamente  $N(0, 1)$  el estadístico

$$\frac{\sum_{i=1}^n X_i^2 - n/\theta}{\sqrt{n}/\theta} = \sqrt{n} \frac{\theta - \hat{\theta}}{\hat{\theta}}.$$

Como  $\hat{\theta} \xrightarrow{c.s.} \theta$  (puesto que  $1/n \sum_{i=1}^n X_i^2 \xrightarrow{c.s.} 1/\theta$ ) resulta que  $\hat{\theta}$  es asintóticamente  $N(\theta, \theta/\sqrt{n})$ .

c)

$$-\frac{\partial^2}{\partial \theta^2} \log f_\theta(x_1, \dots, x_n) = \frac{n}{\theta^2};$$

de forma que la información de Fisher de la muestra es  $ni(\theta) = n/\theta^2$ . Por tanto, la eficiencia asintótica de  $\tilde{\theta}$  es

$$\frac{\theta^2/n}{4\theta^2(4-\pi)/(n\pi)} = \frac{\pi}{4(4-\pi)} = 0'915$$

Mientras que la eficiencia asintótica de  $\hat{\theta}$  es 1.

### Ejercicio 7.9

La función de verosimilitud es

$$f_{\lambda, \theta}(x_1, \dots, x_n) = \lambda^n e^{n\lambda(\theta - \bar{x})} \quad \text{si } x_{(1)} > \theta.$$

Sea

$$f(x_1, \dots, x_n) = \int_1^\infty \int_{-\infty}^{x_{(1)}} \lambda^n e^{n\lambda(\theta - \bar{x})} \frac{d\theta}{\pi(1 + \theta^2)} \frac{d\lambda}{\lambda^2}$$

la densidad marginal de  $(X_1, \dots, X_n)$ , que no depende de  $\lambda$  ni de  $\theta$ . La densidad a posteriori de  $\lambda$  y  $\theta$  será entonces

$$\pi(\lambda, \theta | x_1, \dots, x_n) = \frac{1}{\pi(1 + \theta^2)} \frac{\lambda^{n-2} e^{n\lambda(\theta - \bar{x})}}{f(x_1, \dots, x_n)} \quad \text{para } \theta < x_{(1)} \text{ y } \lambda > 1,$$

cuyo logaritmo, salvo términos independientes de  $\lambda$  y  $\theta$  vale

$$(n-2) \log \lambda + n\lambda(\theta - \bar{x}) - \log(1 + \theta^2).$$

La derivada respecto a  $\theta$

$$n\lambda - \frac{2\theta}{1 + \theta^2}$$

es siempre positiva, puesto que  $n\lambda > 1 > 2\theta/(1 + \theta^2)$ . Luego el valor de  $\theta$  que hace máxima la densidad a posteriori es

$$\theta^* = x_{(1)}.$$

Con tal valor de  $\theta$ , la derivada respecto a  $\lambda$  del logaritmo de la densidad a posteriori resulta

$$\frac{n-2}{\lambda} + n(x_{(1)} - \bar{x});$$

de manera que el valor de  $\lambda$  que hace máxima la densidad a posteriori es

$$\lambda^* = \begin{cases} \frac{n-2}{n(\bar{x} - x_{(1)})} & \text{si } \bar{x} - x_{(1)} < \frac{n-2}{n}, \\ 1 & \text{si } \bar{x} - x_{(1)} \geq \frac{n-2}{n}. \end{cases}$$

Como el logaritmo de la verosimilitud:  $n \log \lambda + n\lambda(\theta - \bar{x})$ , es creciente con  $\theta$ , el estimador de máxima verosimilitud de  $\theta$  es  $\hat{\theta} = x_{(1)}$ . En cambio, una vez hallado el máximo en  $\theta$ , puesto que la derivada respecto a  $\lambda$  del logaritmo de la función de verosimilitud es

$$\frac{n}{\lambda} + n(x_{(1)} - \bar{x}),$$

el estimador de máxima verosimilitud de  $\lambda$  es

$$\hat{\lambda} = \frac{1}{\bar{x} - x_{(1)}},$$

o bien

$$\hat{\lambda} = \begin{cases} \frac{1}{\bar{x} - x_{(1)}} & \text{si } \bar{x} - x_{(1)} < 1 \\ 1 & \text{si } \bar{x} - x_{(1)} \geq 1 \end{cases}$$

en el caso en que se considere como espacio paramétrico  $\lambda > 1$ .

### Ejercicio 7.10

El estimador de mínimos cuadrados de  $(\beta_0, \beta_1, \beta_2)$  será el que minimice

$$\sum_{i=1}^n [c_i - \beta_0 - \beta_1 x_i - \beta_2 y_i]^2$$

Las derivadas parciales respecto a  $\beta_0, \beta_1$ , y  $\beta_2$ , igualadas a cero, proporcionan el sistema de ecuaciones

$$\begin{cases} n\beta_0 + \beta_1 \sum_{i=1}^n x_i + \beta_2 \sum_{i=1}^n y_i &= \sum_{i=1}^n c_i, \\ \beta_0 \sum_{i=1}^n x_i + \beta_1 \sum_{i=1}^n x_i^2 + \beta_2 \sum_{i=1}^n x_i y_i &= \sum_{i=1}^n x_i c_i, \\ \beta_0 \sum_{i=1}^n y_i + \beta_1 \sum_{i=1}^n x_i y_i + \beta_2 \sum_{i=1}^n y_i^2 &= \sum_{i=1}^n y_i c_i, \end{cases}$$

que, con los datos obtenidos, se reduce a

$$\begin{cases} 8\beta_0 + 49\beta_1 + 120\beta_2 & = 290 \\ 49\beta_0 + 555\beta_1 + 1.014\beta_2 & = 3264'9 \\ 120\beta_0 + 1.014\beta_1 + 2.110\beta_2 & = 5967'2 \end{cases}$$

cuya solución proporciona las estimaciones

$$\hat{\beta}_0 = 22'8, \quad \hat{\beta}_1 = 8'787, \quad \hat{\beta}_2 = -2'691.$$

Si  $e$  es  $N(0, \sigma)$ , la función de verosimilitud es

$$\left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (c_i - \beta_0 - \beta_1 x_i - \beta_2 y_i)^2\right\};$$

de logaritmo

$$-\frac{n}{2} \log(2\pi) - n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (c_i - \beta_0 - \beta_1 x_i - \beta_2 y_i)^2.$$

Los valores de  $\beta_0, \beta_1, \beta_2$  que minimizan tal expresión son los mismos determinados anteriormente. Por otra parte, la derivada respecto a  $\sigma$  es

$$-\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (c_i - \beta_0 - \beta_1 x_i - \beta_2 y_i)^2,$$

que es positiva si  $\sigma^2 < \frac{1}{n} \sum_{i=1}^n (c_i - \beta_0 - \beta_1 x_i - \beta_2 y_i)^2$  y negativa en caso contrario. La estimación de máxima verosimilitud de  $\sigma^2$  es pues

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (c_i - \beta_0 - \beta_1 x_i - \beta_2 y_i)^2 = 5'802,$$

una vez sustituidos los valores de  $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$  y evaluada la suma con los datos obtenidos. Es decir,  $\hat{\sigma} = 2'4$ .

## Capítulo 8

# Contraste de hipótesis

### 8.1. Introducción

La finalidad del contraste de hipótesis es decidir si una determinada hipótesis, afirmación o conjetura, sobre la distribución poblacional estudiada, es confirmada o invalidada estadísticamente a partir de las observaciones contenidas en una muestra. Tales situaciones se plantean con mucha frecuencia, en casos tan diversos como decidir si un determinado tratamiento influye favorablemente en la curación de una enfermedad o, simplemente, si un dado está cargado o no.

Lo distintivo, por tanto, del contraste de hipótesis, frente a las técnicas de inferencia anteriores, es la existencia de una hipótesis acerca de la distribución poblacional, que se denomina genéricamente *hipótesis nula* y se designa por  $H_0$ , la cual se pretende contrastar frente a una segunda *hipótesis alternativa*,  $H_1$ , que agrupa todas aquellas distribuciones poblacionales posibles en las que  $H_0$  no es cierta.

Es preciso dejar claro desde el principio, que el contraste de hipótesis no tiene forzosamente un comportamiento imparcial frente a  $H_0$  y  $H_1$ , en el sentido de que no se trata de juzgar cual de ellas es más verosímil a la vista de los datos, sino decidir si la muestra proporciona, o no, evidencia suficiente para descartar la hipótesis  $H_0$ ; para lo cual, desde luego, habrá que tener en consideración la alternativa  $H_1$  que se corrobora cuando se descarta  $H_0$ .

La asimetría entre las hipótesis permite un tratamiento matizado de la decisión entre ambas, en la cual pueden influir, además de la evidencia empírica, el grado de confirmación anterior de que disponga la hipótesis nula, las pérdidas en que se incurra rechazando una hipótesis correcta o aceptando una hipótesis falsa, etc.

Aquellas hipótesis que especifican totalmente la distribución poblacional se denominan *hipótesis simples*, constituidas por una única función de distribución; mientras que aquellas que comprenden varias distribuciones poblacionales se califican de *hipótesis compuestas*.

En los problemas estadísticos paramétricos, en los que la distribución teórica pertenece a una familia  $\mathcal{F} = \{F_\theta \mid \theta \in \Theta\}$  —que se estudian en éste y en el próximo capítulo—, tanto la hipótesis nula como la hipótesis alternativa serán especificadas mediante subconjuntos disjuntos  $\Theta_0$  y  $\Theta_1$  de  $\Theta$ , tales que  $\Theta_0 \cup \Theta_1 = \Theta$ . De forma que será  $H_0 : \theta \in \Theta_0$  y  $H_1 : \theta \in \Theta_1$ .

Cada una de las dos hipótesis será simple si es  $\Theta_0 = \{\theta_0\}$  y  $\Theta_1 = \{\theta_1\}$  respectivamente. Ni que decir tiene, que ambas condiciones simultáneamente suponen el caso más sencillo posible, que puede ser resuelto a plena satisfacción gracias al lema de Neyman-Pearson; este resultado se expone en la sección 8.3, después de haber planteado, con más precisión, los principios básicos generales.

Los contrastes de hipótesis *simple frente a simple* constituyen la pieza básica para abordar, en la sección 8.4, situaciones paramétricas más complejas, en las que intervienen hipótesis compuestas. En estos casos no siempre existe una solución óptima, de manera que el capítulo 9 describirá una forma sistemática —aunque no óptima, en general— de resolver problemas de contraste: los *test de razón de verosimilitudes*, inspirados también en la solución propuesta por el lema de Neyman-Pearson.

En cambio, el capítulo 10, se ocupará de situaciones no paramétricas, en las cuales la hipótesis a contrastar sobre la distribución poblacional no es el valor de alguna de sus características, sino una propiedad global de la distribución.

## 8.2. Planteamiento general de los contrastes de hipótesis

Imaginemos que surge la sospecha de que un determinado fármaco, empleado habitualmente en el tratamiento de ciertas afecciones, tiene como efecto secundario un aumento de la tensión ocular de su nivel normal, 15, a 18; efecto insensible para los pacientes pero que aumenta, a la larga, el riesgo de glaucoma.

Para ser más precisos, los estudios médicos han establecido que en el conjunto de la población, la característica tensión ocular sigue una distribución normal  $N(\theta_0, 1)$ , con  $\theta_0 = 15$ . Así mismo, la sospecha se concreta

en que, entre los consumidores del citado fármaco, la tensión ocular tiene distribución  $N(\theta_1, 1)$ , con  $\theta_1 = 18$ .

La comisión encargada de la distribución de productos farmacéuticos tiene entonces el problema de decidir entre la hipótesis tradicional de ausencia de efectos secundarios entre los consumidores,  $H_0 : \theta = 15$ , frente a la alternativa,  $H_1 : \theta = 18$ , que traduce la presencia de efectos secundarios. Naturalmente no quiere correr el riesgo de retirar del mercado un producto de probada utilidad, sin que haya una seguridad razonable de que existen efectos perjudiciales; ni tampoco dejar a la venta un producto peligroso.

Tendrá, por tanto, que emprender un estudio, basado en la observación de un cierto número de pacientes, para acabar tomando una decisión acerca de  $H_0$  frente a la sospecha  $H_1$ , en la cual se valoren ambos riesgos.

Supongamos que observa 4 pacientes y calcula la tensión ocular media de la muestra obtenida,  $\bar{x}$ . Si  $H_0$  es correcta, como  $\bar{X}$  tiene distribución  $N(15, 1/2)$ , puede hacerse la previsión de que  $\bar{X}$  debería ser, con probabilidad 0'95, inferior a  $15 + 1'645/2$ . La obtención de un valor de  $\bar{x}$  superior a este límite, admite dos interpretaciones:

- ▷ o bien  $H_0$  es cierta pero el azar ha producido una muestra poco representativa, de manera que ha ocurrido un suceso de probabilidad 0'05,
- ▷ o bien realmente el fármaco aumenta la tensión ocular, lo cual explica el valor obtenido.

Siempre caben ambas posibilidades; puesto que un suceso de probabilidad 0'05 no es imposible. Pero si se disminuye progresivamente esta probabilidad —aumentando consecuentemente el límite de los valores previsibles de  $\bar{x}$ — llegará un momento en que ya no estemos dispuestos a aceptar que se ha producido un suceso tan raro.

¿Dónde debe fijarse el límite de las probabilidades por debajo del cual no admitimos explicaciones basadas en la casualidad y rechazamos la posibilidad de que  $H_0$  sea correcta?... Eso depende de la hipótesis  $H_1$ ; no vaya a ser que se rechace  $H_0$  porque hace muy inverosímil el resultado obtenido y se caiga en aceptar  $H_1$  que lo hace más inverosímil todavía.

De momento la conclusión ingenua es que la decisión a favor de  $H_0$  parece clara si  $\bar{x} < 15 + 1'645/2$ , mientras que si  $\bar{x} > 18 - 1'645/2$  es preferible inclinarse por  $H_1$ . Aunque alguna de estas desigualdades se producirá para 95 de cada 100 muestras de tamaño 4 (sea cual sea la hipótesis correcta), queda entre ellas un terreno de nadie en el que la decisión es más comprometida. Si hubiese sido  $H_1 : \theta = 16'5$ , la situación sería peor, porque las

dos zonas determinadas se solaparían y, ante observaciones de  $\bar{x}$  comunes a ambas, no se sabría que decisión tomar. Habrá, por tanto, que aquilatar más para zanjar de alguna forma las ambigüedades.

Procesos de decisión similares al descrito se plantean con gran frecuencia en muy diversos ámbitos y se denominan técnicamente *contrastos o tests de hipótesis*. Sus elementos constituyentes siempre consisten en una variable poblacional  $X$ , y dos hipótesis  $H_0$  y  $H_1$  —no intercambiables— acerca de su distribución, entre las cuales se trata de analizar, mediante una muestra aleatoria (simple) de  $X$ , si  $H_0$  debe ser rechazada en favor de  $H_1$  o bien, por el contrario, no hay razones suficientes para descartar  $H_0$ .

En un problema de contraste de hipótesis, los dos únicos resultados posibles consisten, pues, en *aceptar* o *rechazar*  $H_0$ . Ello deberá hacerse, evidentemente, a la vista de los resultados incluidos en la muestra; de forma que habrá que dividir el espacio muestral  $\mathcal{X}$  en dos regiones disjuntas:  $C$  y  $C^c = \mathcal{X} - C$ , tales que

- $C$  se denomina la *región crítica* y, si la muestra pertenece a  $C$ , se rechaza  $H_0$  para inclinarse por  $H_1$ .
- $C^c$  es la *región de aceptación* y  $H_0$  es aceptada en caso de que la muestra pertenezca a ella.

No se admite la posibilidad de no tomar la decisión, por lo cual la unión de ambas regiones debe ser  $\mathcal{X}$ <sup>1</sup>. Así pues,

*Un test de hipótesis es cualquier partición del espacio muestral  $\mathcal{X}$  en una región crítica  $C$ , en la cual se rechaza la hipótesis nula, y una región de aceptación  $C^c$ , en la cual la hipótesis nula no es rechazada. Tales tests reciben el calificativo de no aleatorizados.*

*En cambio, un test aleatorizado es cualquier función (medible)*

$$\varphi : \mathcal{X} \longrightarrow [0, 1]$$

*llamada función crítica del test, tal que  $\varphi(x_1, \dots, x_n)$  expresa la probabilidad de rechazar la hipótesis nula, cuando se observa la muestra  $(x_1, \dots, x_n) \in \mathcal{X}$ .*

---

<sup>1</sup> De hecho, en la realidad, si los resultados son poco concluyentes, siempre cabe la posibilidad de realizar más observaciones y volver a plantear el problema con mayor tamaño muestral.

Como su propio nombre indica, en un test aleatorizado, tras la observación de la muestra obtenida,  $(x_1, \dots, x_n)$ , ha de efectuarse un sorteo, con probabilidad  $\varphi(x_1, \dots, x_n)$  de rechazar  $H_0$  y probabilidad  $1 - \varphi(x_1, \dots, x_n)$  de aceptarla. En principio, por tanto, mientras los tests no aleatorizados adoptan una decisión tajante a la vista de la muestra, los tests aleatorizados encomiendan parte de la decisión al azar. A primera vista, puede dudarse de la eficacia de esta regla de conducta no determinística; sin embargo, los tests no aleatorizados son un caso particular de ella, con

$$\varphi(x) = I_C(x) = \begin{cases} 1 & \text{si } (x_1, \dots, x_n) \in C, \\ 0 & \text{si } (x_1, \dots, x_n) \in C^c, \end{cases}$$

valores que hacen inútil el sorteo, puesto que dan probabilidad uno de rechazar  $H_0$ , si  $(x_1, \dots, x_n) \in C$ , y, a la vez, probabilidad cero de hacerlo si  $(x_1, \dots, x_n) \in C^c$ .

Evidentemente, el concepto de test de hipótesis sólo sirve como enumeración de todas las posibles reglas de comportamiento que pueden adoptarse frente a un problema de contraste. Entre los tests, habrá muchos muy malos y algunos más precisos y recomendables, para cada situación concreta. Se trata de un concepto genérico —como el de estimador puntual— que no prejuzga los criterios con los cuales debe elegirse el test a aplicar en cada problema. Nada se pierde pues por aumentar el conjunto de soluciones susceptibles de ser utilizadas, incluyendo entre ellas los tests aleatorizados.

La elección del test adecuado será función de sus efectos al aplicarlo. Tal y como está planteada la cuestión, existen dos disyuntivas: la verdad o falsedad de la hipótesis nula y, por otra parte, aceptarla o rechazarla. Cuatro son, por tanto, las alternativas en un contraste:

	$H_0$ es cierta	$H_0$ es falsa
Rechazar $H_0$	Error de tipo I	decisión correcta
Aceptar $H_0$	decisión correcta	Error de tipo II

Rechazar la hipótesis nula cuando es cierta, recibe el nombre de *error de tipo I*, mientras que no descartar  $H_0$  siendo falsa se denomina *error de tipo II*. Ambos tipos de errores deben constituir la principal preocupación ante un contraste de hipótesis.

Lo idóneo sería conseguir un test que hiciese mínima las probabilidades de ambos errores. Desgraciadamente ello no será habitualmente posible, ya

que (salvo en casos excepcionales y triviales) la reducción de la probabilidad de un tipo de error tendrá que hacerse a costa de aumentar la probabilidad de cometer el otro. En consecuencia, el criterio tradicional en la teoría de contrastes para el diseño de un test consiste en:

- 1) Fijar cuidadosamente, en función de las hipótesis y del contexto del problema, una cota para la probabilidad de cometer el error de tipo I, denominada *nivel de significación*,  $\alpha$ , del contraste.
- 2) Excluir, de acuerdo con lo anterior, todos aquellos tests de región crítica  $C$  o de función crítica  $\varphi$  para los que no se verifique la condición:

$$\mathcal{P}_{H_0}\{C\} \leq \alpha \quad \text{o bien} \quad \mathbb{E}_{H_0}[\varphi(X_1, \dots, X_n)] \leq \alpha$$

que impone que la probabilidad de rechazar  $H_0$ , cuando es cierta, no supere el valor  $\alpha$  del nivel de significación.

- 3) Por último, entre los tests no excluidos por la condición anterior, tratar de minimizar la probabilidad de error de tipo II; es decir intentar obtener la región crítica  $C$  o la función crítica  $\varphi$  que hagan mínimo

$$\mathcal{P}_{H_1}(C^c) \quad \text{o bien} \quad \mathbb{E}_{H_1}[1 - \varphi(X_1, \dots, X_n)]$$

que representa la probabilidad de aceptar  $H_0$  siendo cierta  $H_1$ .

Tal criterio otorga, en principio, prioridad a rebajar el riesgo del error de tipo I por debajo del nivel de significación; de ahí que el tratamiento que reciben ambas hipótesis sea asimétrico y éstas no sean intercambiables. Ciertamente, la asimetría queda matizada por el valor que se elija como nivel de significación y por la probabilidad de error de tipo II que resulte una vez diseñado el test; podría ocurrir que ésta fuese, en ocasiones, menor que aquél.

Sin embargo, en el contraste de hipótesis tiende a considerarse que  $H_0$  es la hipótesis preestablecida, contra la cual es necesario esgrimir una gran evidencia, para poder invalidarla. Por consiguiente, los test de hipótesis se emplean con un carácter conservador, a favor de la hipótesis  $H_0$ : el nivel de significación que se fija, intenta garantizar que sea muy infrecuente rechazar una hipótesis nula correcta, y la preocupación por dejar vigente una hipótesis nula falsa es menor, por lo cual pueden aceptarse en ello riesgos más altos. En este sentido, si el resultado de un contraste de hipótesis es aceptar la hipótesis nula, debe interpretarse que las observaciones no han aportado evidencia suficiente para descartarla; mientras que, si se rechaza, es porque

se está razonablemente seguro de que  $H_0$  es falsa y, por consiguiente,  $H_1$  verdadera.

En sí mismos, los tests de hipótesis no llevan incorporado este modo de ver las cosas; lo adquieren, en parte, a través del criterio con el que se escogen y, sobre todo, mediante la elección del nivel de significación que han de satisfacer. Así, en los problemas prácticos, el rango de valores de  $\alpha$  que se utiliza suele ser  $\alpha = 0'1, 0'05, 0'025, 0'01, \dots$ ; de manera que, una vez diseñado un test cuya región crítica tenga un tal nivel de significación, sea “muy raro” obtener una muestra en ella cuando la hipótesis nula sea cierta.

De todas maneras, el rango de valores de  $\alpha$  citado está adaptado al tipo de contrastes que suelen presentarse con más frecuencia en las aplicaciones. No conviene transcribirlo automáticamente a cualquier situación, pues podría en ocasiones dar lugar a resultados poco razonables (en los contrastes de hipótesis simple frente a simple, por ejemplo). En definitiva, la elección del nivel de significación es una cuestión delicada, a la que a menudo se presta menos atención de la debida. Volveremos sobre ella al final de este apartado.

En el marco de los problemas paramétricos, el criterio descrito para la selección de tests, puede formularse más explícitamente, mediante la siguiente definición.

*Si la distribución teórica depende de un parámetro  $\theta \in \Theta$ , acerca del cual hay que contrastar la hipótesis nula  $H_0 : \theta \in \Theta_0$ , frente a la alternativa  $H_1 : \theta \in \Theta_1$ , se denomina función de potencia,  $\beta(\theta)$ , de un test a la probabilidad de rechazar  $H_0$  cuando el valor del parámetro es  $\theta$ . Es decir, si se trata de un test no aleatorizado de región crítica  $C$ , entonces*

$$\beta(\theta) = \mathcal{P}_\theta(C);$$

*mientras que, en general, para un test de función crítica  $\varphi$ , es*

$$\beta(\theta) = E_\theta[\varphi(X_1, \dots, X_n)].$$

*Un test tiene nivel de significación  $\alpha$  si es*

$$\beta(\theta) \leq \alpha \quad \text{para cada } \theta \in \Theta_0;$$

*y se denomina tamaño del test al número*

$$\sup_{\theta \in \Theta_0} \beta(\theta).$$

En estos términos, el criterio enunciado consiste en limitarse a considerar tests de nivel de significación  $\alpha$  —o, lo que es lo mismo, de tamaño menor o igual que  $\alpha$ — siendo  $\alpha$  un valor pequeño, prefijado. A continuación, puesto que, cuando el valor del parámetro es  $\theta \in \Theta_1$ , la probabilidad de error de tipo II es  $1 - \beta(\theta)$ , ha de intentarse maximizar la potencia,  $\beta(\theta)$ , para todos los valores  $\theta$  de  $\Theta_1$  simultáneamente.

Ambos requisitos son claramente contrapuestos. Un nivel de significación pequeño se consigue reduciendo la región crítica  $C$  (o la función crítica  $\varphi$ ), de manera a disminuir la potencia  $\beta(\theta)$  para  $\theta \in \Theta_0$ . En cambio, una potencia  $\beta(\theta)$  alta, para  $\theta \in \Theta_1$ , exige aumentar  $C$  (o  $\varphi$ ) todo lo posible. Por consiguiente, una vez que se ha fijado el nivel de significación  $\alpha$  que se considera adecuado, no conviene extralimitarse y diseñar un test de tamaño menor que  $\alpha$ , porque ello se traducirá normalmente en una pérdida de potencia  $\beta(\theta)$  para  $\theta \in \Theta_1$ ; es decir, es deseable que tamaño y nivel de significación coincidan, verificándose

$$\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha.$$

Para distribuciones teóricas de tipo continuo, cada muestra tiene probabilidad cero y  $\mathcal{P}_\theta(C)$  crece continuamente al incrementar la región crítica  $C$ ; de forma que la igualdad anterior puede conseguirse mediante tests no aleatorizados. En cambio, para distribuciones poblacionales discretas, cada muestra tiene una probabilidad positiva y puede ocurrir que no haya manera de encontrar un conjunto  $C$  de muestras para las que se verifique la igualdad anterior; en tal caso, será conveniente utilizar tests aleatorizados en los cuales se puede hacer variar  $\varphi$  de forma continua. De hecho la estructura de las funciones críticas de los tests aleatorizados que realmente se usan, suele ser

$$\varphi(x_1, \dots, x_n) = \begin{cases} 1 & \text{si } (x_1, \dots, x_n) \in C \\ \gamma & \text{si } (x_1, \dots, x_n) \in C' \\ 0 & \text{si } (x_1, \dots, x_n) \in \mathcal{X} - C - C' \end{cases}$$

con lo cual

$$\mathbb{E}_\theta[\varphi(X_1, \dots, X_n)] = \mathcal{P}_\theta(C) + \gamma \mathcal{P}_\theta(C')$$

y, supuesto que  $\sup_{\theta \in \Theta_0} \mathcal{P}_\theta(C) < \alpha$  y  $\sup_{\theta \in \Theta_0} \mathcal{P}_\theta(C \cup C') > \alpha$ , se puede ajustar  $\gamma$  para que sea

$$\sup_{\theta \in \Theta_0} \mathbb{E}_\theta[\varphi] = \sup_{\theta \in \Theta_0} [\mathcal{P}_\theta(C) + \gamma \mathcal{P}_\theta(C')] = \alpha.$$

Los tests aleatorizados garantizan pues, en cualquier situación, que puede lograrse un tamaño igual al nivel de significación prefijado, cumpliendo con exactitud la limitación impuesta a la probabilidad de error de tipo I. Esto deja todo el margen de maniobra posible para tratar de minimizar la probabilidad de error de tipo II o, dicho de otra manera, hacer máxima la potencia  $\beta(\theta)$  para  $\theta \in \Theta_1$ .

A este respecto, salvo en el caso en que  $\Theta_1$  sea simple, la dificultad está en querer maximizar varios valores a la vez: si se consigue hacer máximo  $\beta(\theta_1)$ , puede que  $\beta(\theta'_1)$  sea pequeño y viceversa. Únicamente en circunstancias favorables (de las que nos ocuparemos en las secciones 8.3 y 8.4), existirá un test uniformemente de máxima potencia (*UMP*), que tenga más potencia que cualquier otro test del mismo nivel de significación, para todos los valores  $\theta \in \Theta_1$  simultáneamente.

En general, sólo será posible una ordenación parcial de los tests, en el siguiente sentido:

*Dados dos tests de funciones críticas  $\varphi$  y  $\varphi'$ , con nivel de significación  $\alpha$ , para contrastar  $H_0 : \theta \in \Theta_0$  frente a  $H_1 : \theta \in \Theta_1$ , el primero se denomina uniformemente más potente que el segundo si se cumple*

$$\beta_\varphi(\theta) \geq \beta_{\varphi'}(\theta) \quad \text{para todo } \theta \in \Theta_1$$

Sin duda puede ocurrir que  $\varphi$  no sea uniformemente más potente que  $\varphi'$ , ni  $\varphi'$  uniformemente más potente que  $\varphi$ .

*En los términos anteriores, un test de función crítica  $\varphi$  es uniformemente de máxima potencia, dentro de una familia  $\mathcal{K}$  de tests, si es uniformemente más potente que cualquier otro test de  $\mathcal{K}$ , con el mismo nivel de significación.*

Describa la idea para actuar ante un problema de contraste de hipótesis, se comprende fácilmente la influencia que el nivel de significación elegido tiene en el test que resultará diseñado y en la decisión final que se adopte acerca de las hipótesis. Desde el punto de vista tradicional, ello significa que la elección de  $\alpha$  debe hacerse “a conciencia”, valorando todas las componentes del problema y su contexto. Ello es a menudo ambiguo —en situaciones como la descrita al principio de esta sección— y, además, resulta poco informativo, pues oculta el grado de evidencia conseguido con la muestra: sólo indica si es más o menos que el patrón fijado.

La tendencia actual es adoptar un punto de vista menos comprometido. El diseño del test puede hacerse en función de  $\alpha$ , sin prefijar su valor. A continuación, lo ideal sería obtener la función de potencia,  $\beta_\alpha(\theta)$ , de cada uno de los tests resultantes. Puesto que la familia de curvas de potencia, asociadas a cada valor de  $\alpha$ , da una información completa de cada uno de los riesgos en que se está incurriendo al elegir cada test, el decisor dispone con ello de todos los elementos de juicio para adoptar el valor de  $\alpha$  que le parezca preferible.

En realidad, el programa anterior resulta, a menudo, demasiado ambicioso debido a las dificultades que comporta el cálculo de la potencia. Como solución de compromiso, se puede proceder al diseño en función de  $\alpha$  y, una vez obtenida la muestra concreta, observar el nivel de significación más pequeño para el que tal muestra obliga a rechazar la hipótesis nula. Este número  $\alpha(x_1, \dots, x_n)$  se denomina *nivel crítico*<sup>2</sup> asociado a la muestra obtenida e indica el apoyo que la hipótesis nula recibe de las observaciones; cuanto más grande sea su valor, dentro del intervalo  $[0, 1]$ , más confirmada queda la hipótesis  $H_0$ .

Según el espíritu conservador antes apuntado, un nivel crítico superior a 0'1 se considera un apoyo suficiente para mantener la hipótesis nula. Por otra parte, cuanto más próximo a cero sea el nivel crítico más confiadamente podemos rechazar la hipótesis nula; aunque convendrá compararlo con la probabilidad de error de tipo II... no vaya a ser que se sustituya una hipótesis con muy poco apoyo por otra con menos todavía!

### 8.3. Contraste de hipótesis simple frente a simple

La filosofía de los contrastes, expuesta en la sección anterior, no resuelve de manera automática los problemas concretos, como el de diseñar un test adecuado para dilucidar si la tensión ocular,  $X$ , de los consumidores de un fármaco, tiene distribución  $N(15, 1)$  ( $H_0$ ) o distribución  $N(18, 1)$  ( $H_1$ ).

En este caso particular, es natural calcular la media de las observaciones,  $\bar{X}$ ; se dispone con ello de un buen estimador puntual de la media poblacional,  $\theta$ , cuya distribución en el muestreo,  $N(\theta, 1/\sqrt{n})$ , está concentrada alrededor del valor real de  $\theta$  (15 ó 18).

Ahora que está más preciso lo que se trata de hacer, las consideraciones intuitivas del principio de la sección anterior parecen aconsejar un test que

---

<sup>2</sup> *p*-valor en la terminología anglo-sajona.

tenga una región crítica de la forma

$$C = \{\bar{X} > c\},$$

donde  $c$  es un punto crítico a determinar, en función del riesgo de error de tipo I que se admite y del tamaño muestral (del que depende lo concentrada que sea la distribución de  $\bar{X}$  alrededor de  $\theta$ ).

Concretamente, si el nivel de significación es  $\alpha$ , habrá de ser

$$\mathcal{P}_{\theta=15}\{\bar{X} > c\} = \alpha,$$

lo cual da como resultado

$$c = 15 + \frac{z_\alpha}{\sqrt{n}}$$

donde  $z_\alpha$  es el valor tal que  $P\{Z > z_\alpha\} = \alpha$  si  $Z$  tiene distribución  $N(0, 1)$ .

En la siguiente figura se han representado (para un  $n$  muy pequeño) las distribuciones de  $\bar{X}$  asociadas a ambas hipótesis. Aparece claramente la representación del tamaño ( $\alpha$ ) y de la probabilidad de error de tipo II ( $1 - \beta$ ), correspondientes a un cierto valor del punto crítico  $c$ .

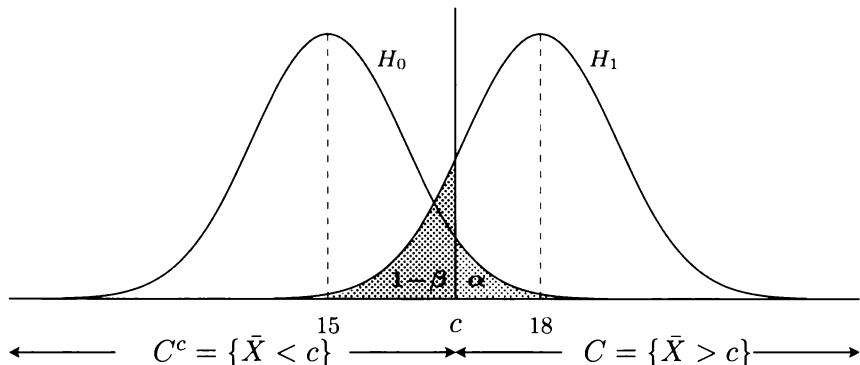


Figura 8.1: Test de hipótesis simple frente a simple

Desplazar el punto crítico  $c$  hacia la derecha disminuye la probabilidad de error de tipo I ( $\alpha$ ), a costa de aumentar la probabilidad de error de tipo II ( $1 - \beta$ ). Inversamente, cuanto mayor nivel de significación se admite más próximo a  $\theta = 15$  estará el punto  $c$  y ello redundará en una menor probabilidad de error de tipo II. Después de obtener una muestra concreta, situando

el punto crítico en  $\bar{x}$ , se podrá observar cual es el nivel crítico,  $\alpha(\bar{x})$ , que exigen los datos obtenidos para rechazar la hipótesis nula y determinar la correspondiente probabilidad de aceptarla cuando es falsa; se dispondrá así de todos los elementos de juicio necesarios para tomar la decisión entre  $H_0$  y  $H_1$ .

En principio, podría pensarse en repetir las consideraciones anteriores sustituyendo  $\bar{X}$  por otro estadístico —como la mediana muestral o el estadístico de Student, por ejemplo. El resultado que se expone a continuación demostrará, sin embargo, que el sentido común ha conducido al test óptimo o, para ser más precisos, al test de máxima potencia para contrastar  $H_0 : \theta = 15$  frente a  $H_1 : \theta = 18$ . De hecho, el siguiente lema proporciona un método sistemático para determinar el test de máxima potencia en cualquier contraste de hipótesis simple frente a simple. Conviene analizar, en primer lugar, al caso de tests no aleatorizados, ya que los razonamientos son entonces más simples y suficientes para resolver múltiples casos.

**Lema de Neyman-Pearson (tests no aleatorizados):** *Dada una muestra aleatoria con distribución  $\mathcal{P}_\theta$ , sea  $f_\theta(x_1, \dots, x_n)$  su función de densidad o de probabilidad (según que la población sea continua o discreta).*

*Si  $C^* \subset \mathcal{X}$  es tal que*

$$\{(x_1, \dots, x_n) \in \mathcal{X} \mid f_{\theta_1} > k f_{\theta_0}\} \subset C^* \subset \{(x_1, \dots, x_n) \in \mathcal{X} \mid f_{\theta_1} \geq k f_{\theta_0}\}$$

*para alguna constante  $k > 0$ , y es  $\mathcal{P}_{\theta_0}(C^*) = \alpha$ , entonces  $C^*$  es la región crítica de un test de nivel de significación  $\alpha$ , para contrastar  $H_0 : \theta = \theta_0$  frente a  $H_1 : \theta = \theta_1$ , de máxima potencia<sup>3</sup> dentro de la familia de tests no aleatorizados.*

El resultado informa de que, para contrastar con máxima potencia entre las hipótesis indicadas, debe rechazarse  $H_0$  siempre que  $f_{\theta_1} > k f_{\theta_0}$  y no puede rechazarse nunca cuando  $f_{\theta_1} < k f_{\theta_0}$ ; en caso de igualdad,  $f_{\theta_1} = k f_{\theta_0}$  puede adoptarse cualquier decisión. La constante  $k$  ha de determinarse, intentando conseguir que  $\mathcal{P}_{\theta_0}(C^*)$  sea igual al nivel de significación  $\alpha$  que se desea conseguir. No se garantiza que se pueda lograr para cualquier  $\alpha$ ; sino que, para los  $\alpha$  para los que ello sea posible,  $C^*$  expresa la región crítica de un test más potente que cualquier otro test no aleatorizado de nivel de significación  $\alpha$ .

---

<sup>3</sup> En esta sección se suprime el adverbio “uniformemente”, puesto que sólo existe un valor  $\theta_1$  como alternativa.

Demostración: Supongamos que  $C$  es la región crítica de otro test no aleatorizado, con nivel de significación  $\alpha$ ; es decir  $\mathcal{P}_{\theta_0}(C) \leq \alpha$ . Hay que comprobar que  $\mathcal{P}_{\theta_1}(C^*) \geq \mathcal{P}_{\theta_1}(C)$ .

Como  $C^* = (C^* \cap C) \cup (C^* \cap C^c)$  y  $C = (C \cap C^*) \cup (C \cap C^{*c})$ , será

$$\mathcal{P}_{\theta}(C^*) - \mathcal{P}_{\theta}(C) = \mathcal{P}_{\theta}(C^* \cap C^c) - \mathcal{P}_{\theta}(C \cap C^{*c})$$

para ambos valores de  $\theta$ . Pero, la condición que verifica  $C^*$  asegura

$$\mathcal{P}_{\theta_1}(C^* \cap C^c) \geq k \mathcal{P}_{\theta_0}(C^* \cap C^c),$$

$$\mathcal{P}_{\theta_1}(C \cap C^{*c}) \leq k \mathcal{P}_{\theta_0}(C \cap C^{*c}).$$

Por tanto

$$\begin{aligned} \mathcal{P}_{\theta_1}(C^*) - \mathcal{P}_{\theta_1}(C) &= \mathcal{P}_{\theta_1}(C^* \cap C^c) - \mathcal{P}_{\theta_1}(C \cap C^{*c}) \\ &\geq k [\mathcal{P}_{\theta_0}(C^* \cap C^c) - \mathcal{P}_{\theta_0}(C \cap C^{*c})] \\ &= k [\mathcal{P}_{\theta_0}(C^*) - \mathcal{P}_{\theta_0}(C)] \end{aligned}$$

y este último término es no negativo por ser  $\mathcal{P}_{\theta_0}(C^*) = \alpha \geq \mathcal{P}_{\theta_0}(C)$ . Así que  $\mathcal{P}_{\theta_1}(C^*) \geq \mathcal{P}_{\theta_1}(C)$ .

El resultado anterior es suficiente para diseñar los contrastes óptimos en muchos casos. Por ejemplo, el que se ha venido utilizando como ilustración:

### Ejemplo 8.1

Se dispone de una muestra aleatoria simple de tamaño  $n$  de una población  $N(\theta, \sigma)$ , donde  $\sigma$  es una constante conocida y  $\theta$  puede valer  $\theta_0$  ó  $\theta_1$ , con  $\theta_0 < \theta_1$ . La densidad de la muestra es entonces

$$f_{\theta}(x_1, \dots, x_n) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2 \right\}.$$

Para contrastar  $H_0 : \theta = \theta_0$  frente a  $H_1 : \theta = \theta_1$ , con máxima potencia, el lema de Neyman-Pearson indica que hay que efectuar comparaciones del tipo

$$\frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta_1)^2} > k \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta_0)^2}$$

que equivalen a

$$-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta_1)^2 > \log k - \frac{1}{2} \sum_{i=1}^n (x_i - \theta_0)^2$$

o bien

$$(\theta_1 - \theta_0) \sum_{i=1}^n x_i + n(\theta_0^2 - \theta_1^2) > 2\sigma^2 \log k.$$

Puesto que  $\theta_1 - \theta_0 > 0$ , ello se reduce a

$$\bar{x} > c,$$

siendo  $c$  una constante (que depende de  $k$ ,  $\sigma^2$ ,  $\theta_0$ ,  $\theta_1$  y  $n$ ). La región crítica de máxima potencia debe satisfacer por tanto

$$\{\bar{x} > c\} \subset C^* \subset \{\bar{x} \geq c\},$$

lo cual sólo deja como duda irrelevante qué hacer en el caso (de probabilidad cero) en que  $\bar{x} = c$ . Puede tomarse pues, indiferentemente,  $C^* = \{\bar{x} > c\}$  o bien  $C^* = \{\bar{x} \geq c\}$ .

Queda por ajustar el valor del punto crítico  $c$ . Ahora bien, para un nivel de significación  $\alpha$ , debe ser

$$\alpha = \mathcal{P}_{\theta_0}(C^*) = P \left\{ Z > \frac{c - \theta_0}{\sigma/\sqrt{n}} \right\},$$

siendo  $Z$  una variable  $N(0, 1)$  (puesto que  $\bar{X}$  tiene distribución  $N(\theta, \sigma/\sqrt{n})$ ). Se obtiene pues

$$c = \theta_0 + z_\alpha \frac{\sigma}{\sqrt{n}}.$$

Y la probabilidad de aceptar  $H_0$  siendo falsa resulta

$$1 - \beta = \mathcal{P}_{\theta_1}(C^{*c}) = P \left\{ Z > \frac{c - \theta_1}{\sigma/\sqrt{n}} \right\} = P \left\{ Z > z_\alpha + \frac{\theta_0 - \theta_1}{\sigma/\sqrt{n}} \right\}.$$

Por ejemplo, si  $\theta_0 = 15$ ,  $n = 25$  y  $\sigma = 1$ , se obtiene

$$c = 15 + 0'2z_\alpha = \begin{cases} 15'329 & \text{si } \alpha = 0'05 \\ 15'464 & \text{si } \alpha = 0'01 \\ 15'618 & \text{si } \alpha = 0'001 \\ 15'744 & \text{si } \alpha = 10^{-4} \\ 16'033 & \text{si } \alpha = 10^{-7} \end{cases}$$

Si  $\theta_1 = 18$ , las correspondientes probabilidades de error de tipo II son cero (no exactamente, pero sí a cualquier efecto práctico).

Hay dos conclusiones que sacar de estos resultados numéricos. Primero que una muestra de tamaño  $n = 25$  permite discernir “perfectamente” entre la  $N(15, 1)$  y la  $N(18, 1)$ ; con probabilidades de cometer alguno de los errores inferiores a una diezmillonésima, para ser más precisos. Segundo que no tiene ningún sentido, en un problema como este, usar un nivel de significación “estándar”, como  $0'05$  ó  $0'01$ , pues estaríamos aceptando un riesgo de error de tipo I del 5% o el 1%, cuando se puede rebajar hasta  $10^{-5}\%$  sin que el riesgo de error de tipo II deje de ser cero (a cualquier efecto práctico).

En definitiva, la comisión de vigilancia de productos farmacéuticos, no tiene ningún problema si examina al menos 25 pacientes. Rechazando que el producto es inócuo, si observa una media muestral superior a  $16'033$  corre un riesgo de  $10^{-7}$  de prescindir de un producto adecuado y un riesgo inferior de dejar a la venta un producto dañino.

En el ejemplo anterior, la continuidad de la distribución muestral ha permitido diseñar un test de tamaño exactamente igual a  $\alpha$  (cualquiera que éste sea). Cabe la duda de si podría existir un test aleatorizado de mayor potencia que el test no aleatorizado obtenido. De hecho no es así; lo que se gana considerando tests aleatorizados es la capacidad de alcanzar cualquier tamaño dado y no mejorar los tests no aleatorizados dentro de su tamaño. El siguiente ejemplo muestra cual es la dificultad en el caso de poblaciones discretas.

### Ejemplo 8.2

Una empresa produce lotes de piezas de dos calidades, con proporciones de piezas correctas  $\theta_0$  y  $\theta_1$  respectivamente, siendo  $\theta_0 > \theta_1$ . Compramos un lote de la primera calidad y debemos tratar de determinar si nos han engañado, dandonos uno de baja calidad. Como la comprobación de una pieza es cara, decidimos comprobar únicamente  $n$  piezas elegidas al azar y con reemplazamiento.

Se trata pues de efectuar el contraste de  $H_0 : \theta = \theta_0$  frente a  $H_1 : \theta = \theta_1$ , a partir de una muestra aleatoria simple de función de probabilidad

$$f_\theta(x_1, \dots, x_n) = \theta^{x_1 + \dots + x_n} (1 - \theta)^{n - (x_1 + \dots + x_n)} \quad \text{para } x_1, \dots, x_n = 0, 1$$

( $x_i = 1$  significa que la pieza  $i$ -ésima es correcta).

El test de máxima potencia se basa en comparaciones del tipo

$$\theta_1^{x_1 + \dots + x_n} (1 - \theta_1)^{n - (x_1 + \dots + x_n)} > k \theta_0^{x_1 + \dots + x_n} (1 - \theta_0)^{n - (x_1 + \dots + x_n)}$$

que, llamando  $T$  al estadístico  $\sum_{i=1}^n x_i$  (número total de piezas correctas observadas), equivalen a

$$T \log \theta_1 + (n - T) \log(1 - \theta_1) > \log k + T \log \theta_0 + (n - T) \log(1 - \theta_0),$$

o bien

$$\left[ \log \frac{\theta_1}{\theta_0} + \log \frac{1 - \theta_0}{1 - \theta_1} \right] T > \log k + n \left[ \log \frac{1 - \theta_0}{1 - \theta_1} \right].$$

Puesto que  $\theta_0 > \theta_1$ , ello se reduce a

$$T < c \quad \text{con} \quad T = \sum_{i=1}^n x_i,$$

donde  $c$  no depende de la muestra (sino sólo de  $k$ ,  $\theta_0$ ,  $\theta_1$  y  $n$ ). La región crítica del test de máxima potencia cumplirá entonces

$$\{T < c\} \subset C^* \subset \{T \leq c\}.$$

En este caso,  $T = c$  no tiene probabilidad nula y el resultado indica que cualquier región crítica de la forma

$$C^* = \{T < c\} \cup A \quad \text{siendo } A \text{ un subconjunto de } \{T = c\}$$

es de máxima potencia entre los tests de nivel de significación  $\mathcal{P}_{\theta_0}(C^*)$ .

Veamos cómo varía  $\mathcal{P}_{\theta_0}\{T < c\}$  al variar  $c$ . Por ejemplo, si  $n = 10$  y  $\theta_0 = 3/4$ , será

$$\mathcal{P}_{\theta_0}\{T < c\} = \sum_{j=0}^{c-1} \binom{10}{j} \left(\frac{3}{4}\right)^j \left(\frac{1}{4}\right)^{10-j}$$

que vale 0'0035 para  $c \in [4, 5)$ , 0'0197 para  $c \in [5, 6)$ , 0'0781 para  $c \in [6, 7)$ , 0'2241 para  $c \in [7, 8)$ , etc.

Si decidimos usar  $\alpha = 0'0197$  como nivel de significación, el test de región crítica  $\{T < 5\}$  será de máxima potencia, no sólo entre los no aleatorizados (aunque ello sea lo único que nos conste todavía), sino entre todos los tests de nivel de significación 0'0197.

En cambio, si fijamos  $\alpha = 0'1$ , habría que tratar de añadir a  $\{T < 6\}$  un subconjunto de  $\{T = 6\}$ , de probabilidad 0'0219, para conseguir una región crítica  $C^*$  de tamaño 0'1. Porque  $\{T < 6\}$  tiene tamaño 0'0781 y, por tanto, cumple la especificación sobre el nivel de significación; pero la holgura entre

el tamaño y el nivel de significación hace que este test no pueda competir en potencia —como test de nivel 0'1— con  $C^*$ .

Ahora bien, en vez de investigar si las muestras comprendidas en  $\{T = 6\}$  se pueden agrupar (de manera arbitraria) para conseguir un subconjunto  $A$  de probabilidad 0'0219<sup>4</sup>, es más sencillo y con garantías de éxito, usar el test aleatorizado de función crítica

$$\varphi(x_1, \dots, x_n) = \begin{cases} 1 & \text{si } T \leq 5 \\ \gamma & \text{si } T = 6 \\ 0 & \text{si } T > 6 \end{cases}$$

donde debe ser  $\gamma = 0'15$ , para que sea

$$0'1 = E_{\theta_0}[\varphi] = P_{\theta_0}\{T < 6\} + \gamma P_{\theta_0}\{T = 6\} = 0'0781 + \gamma 0'146.$$

Se habrá conseguido así un test de tamaño 0'1, cuya potencia en  $\theta_1 = 1/2$  —por ejemplo— es

$$E_{\theta_1}[\varphi] = 0'623 + \gamma \binom{10}{6} \frac{1}{2^{10}} = 0'656,$$

mientras que la potencia del  $\{T < 6\}$  es

$$P_{\theta_1}\{T < 6\} = \sum_{j=0}^5 \binom{10}{j} \frac{1}{2^{10}} = 0'623;$$

de forma que, al ajustar el tamaño al nivel de significación, se consigue disminuir la probabilidad de error de tipo II de 0'377 a 0'344.

El ejemplo anterior muestra la necesidad, en ciertos casos, de considerar tests aleatorizados que permitan igualar el tamaño al nivel de significación, para poder sacar de ello un mejor partido en cuanto a potencia en la hipótesis alternativa. Se hace necesario así ampliar a este caso el lema de Neyman-Pearson.

**Lema de Neyman-Pearson:** *En las mismas condiciones del resultado anterior:*

(a) *Cualquier test con función crítica de la forma*

$$\varphi(x_1, \dots, x_n) = \begin{cases} 1 & \text{si } f_{\theta_1}(x_1, \dots, x_n) > k f_{\theta_0}(x_1, \dots, x_n) \\ \gamma(x_1, \dots, x_n) & \text{si } f_{\theta_1}(x_1, \dots, x_n) = k f_{\theta_0}(x_1, \dots, x_n) \\ 0 & \text{si } f_{\theta_1}(x_1, \dots, x_n) < k f_{\theta_0}(x_1, \dots, x_n) \end{cases}$$

<sup>4</sup>Ello es imposible, pues cada muestra de  $\{T = 6\}$  tiene probabilidad  $(\frac{3}{4})^6 (\frac{1}{4})^4$  y harían falta 31'5 muestras cualesquiera para lograrlo.

con  $k \geq 0$  y  $\gamma(x_1, \dots, x_n) \in [0, 1]$ , es de máxima potencia para contrastar  $H_0 : \theta = \theta_0$  frente a  $H_1 : \theta = \theta_1$ .

(b) Para cada  $\alpha \in (0, 1)$ , existe un test de la forma anterior, en el cual es constante  $\gamma(x_1, \dots, x_n)$  y cuyo tamaño es  $\alpha$ .

(c) Si  $\varphi$  es la función crítica de un test de máxima potencia para contrastar  $H_0 : \theta = \theta_0$  frente a  $H_1 : \theta = \theta_1$ , es de la forma descrita en (a) (excepto quizás en un subconjunto de  $\mathcal{X}$  que tiene probabilidad cero con  $\mathcal{P}_{\theta_0}$  y  $\mathcal{P}_{\theta_1}$ ).

Las afirmaciones (a) y (c) conjuntamente expresan la condición suficiente y necesaria para que un test  $\varphi$  sea de máxima potencia; esto es, más potente que cualquier otro test de nivel de significación  $\alpha$ , igual al tamaño de  $\varphi$ . Por su parte (b) confirma que es posible conseguir cualquier tamaño, tomando  $\gamma$  constante.

Demostración: Consideraremos el caso continuo, en que  $f_{\theta_0}$  y  $f_{\theta_1}$  son las funciones de densidad de la muestra, ya que en el caso discreto basta reemplazar las integrales por sumas y el razonamiento es el mismo.

(a) El tamaño de  $\varphi$  es  $E_{\theta_0}[\varphi]$ , así que consideremos cualquier otro test  $\varphi'$  con tamaño  $E_{\theta_0}[\varphi'] \leq E_{\theta_0}[\varphi]$ . Hay que probar que  $E_{\theta_1}[\varphi'] \leq E_{\theta_1}[\varphi]$ . Pero

$$\int_{\mathcal{X}} [\varphi - \varphi'] [f_{\theta_1} - kf_{\theta_0}] dx_1 \dots dx_n = \int_{\{f_{\theta_1} < kf_{\theta_0}\}} [-\varphi'] [f_{\theta_1} - kf_{\theta_0}] dx_1 \dots dx_n + \int_{\{f_{\theta_1} > kf_{\theta_0}\}} [1 - \varphi'] [f_{\theta_1} - kf_{\theta_0}] dx_1 \dots dx_n$$

da un resultado no negativo, puesto que los integrandos son no negativos. Por tanto

$$E_{\theta_1}[\varphi] - E_{\theta_1}[\varphi'] \geq k (E_{\theta_0}[\varphi] - E_{\theta_0}[\varphi']) \geq 0.$$

(b) Sea  $Y$  la variable aleatoria  $f_{\theta_1}(X_1, \dots, X_n)/f_{\theta_0}(X_1, \dots, X_n)$  y  $F$  su función de distribución cuando  $\theta = \theta_0$  (puede darse a  $Y$  un valor arbitrario en  $\{f_{\theta_0} = 0\}$  y ello no altera la función  $F$ , puesto que dicho conjunto tiene probabilidad  $\mathcal{P}_{\theta_0}$  nula). Si  $\gamma(x_1, \dots, x_n)$  es una constante  $\gamma$ , será

$$\begin{aligned} E_{\theta_0}[\varphi] &= \mathcal{P}_{\theta_0}\{f_{\theta_1} > kf_{\theta_0}\} + \gamma \mathcal{P}_{\theta_0}\{f_{\theta_1} = kf_{\theta_0}\} \\ &= 1 - \mathcal{P}_{\theta_0}\{Y \leq k\} + \gamma \mathcal{P}_{\theta_0}\{Y = k\} \\ &= 1 - F(k) + \gamma[F(k) - F(k^-)]. \end{aligned}$$

Si existe  $k_0$  con  $F(k_0) = 1 - \alpha$ , basta tomar  $k = k_0$  y  $\gamma = 0$  para tener  $E_{\theta_0}[\varphi] = \alpha$ . El test obtenido será entonces un test no aleatorizado.

En caso contrario, existirá  $k_0$  tal que  $F(k_0^-) \leq 1 - \alpha < F(k_0)$ ; basta tomar entonces  $k = k_0$  y

$$\gamma = \frac{F(k_0) - (1 - \alpha)}{F(k_0) - F(k_0^-)}$$

para obtener  $E_{\theta_0}[\varphi] = \alpha$ .

(c) Sea  $\varphi'$  la función crítica de un test de máxima potencia y nivel de significación  $\alpha \in (0, 1)$ , para contrastar  $H_0$  frente a  $H_1$ . Según (b), se puede obtener un test  $\varphi$  de la forma descrita en (a) —para un cierto  $k$ — con tamaño igual a  $\alpha$ . Si fuese

$$\int_{\mathcal{X}} [\varphi - \varphi'][f_{\theta_1} - kf_{\theta_0}] dx_1 \dots dx_n > 0,$$

resultaría

$$E_{\theta_1}[\varphi] - E_{\theta_1}[\varphi'] > k(E_{\theta_0}[\varphi] - E_{\theta_0}[\varphi']) \geq 0$$

y  $\varphi$  tendría mayor potencia que  $\varphi'$ . Por consiguiente, la integral anterior ha de ser igual a 0 y, como el integrando es no negativo (pues vale  $1 - \varphi'$  si  $f_{\theta_1} > kf_{\theta_0}$  y  $-\varphi'$  si  $f_{\theta_1} < kf_{\theta_0}$ ) debe ser  $\varphi' = \varphi$  siempre que  $f_{\theta_1} \neq kf_{\theta_0}$ , excepto a lo sumo en un subconjunto de longitud cero, que tendrá, por tanto, probabilidad cero bajo  $\mathcal{P}_{\theta_0}$  y  $\mathcal{P}_{\theta_1}$ . En definitiva,  $\varphi'$  tiene la forma descrita en (a), casi seguro respecto a ambas hipótesis.

Veamos un ejemplo de utilización de este último resultado.

### Ejemplo 8.3

El número de emisiones de partículas radiactivas por segundo, de una determinada sustancia, sigue una distribución de Poisson de parámetro  $\theta$ . Se desea contrastar  $H_0 : \theta = \theta_0$  frente a  $H_1 : \theta = \theta_1$ , siendo  $\theta_0 < \theta_1$ , mediante la observación del número de partículas emitidas durante  $n$  segundos (entre los cuales se supone que hay independencia).

La función de probabilidad de la muestra es

$$f_{\theta}(x_1, \dots, x_n) = \frac{e^{-n\theta} \theta^{x_1 + \dots + x_n}}{x_1! \dots x_n!}.$$

De manera que, llamando  $T = \sum_{i=1}^n x_i$ , la comparación  $f_{\theta_1} > kf_{\theta_0}$  es equivalente a

$$T \log \theta_1 - n\theta_1 > \log k + T \log \theta_0 - n\theta_0,$$

lo cual, habida cuenta de que  $\theta_1 > \theta_0$ , equivale a

$$T > c,$$

siendo  $c = (\log k - n\theta_1 - n\theta_0) / (\log \theta_1 - \log \theta_0)$ .

La función crítica de los tests de máxima potencia para contrastar  $H_0$  frente a  $H_1$  será pues de la forma

$$\varphi(x_1, \dots, x_n) = \begin{cases} 1 & \text{si } T > c \\ \gamma & \text{si } T = c \\ 0 & \text{si } T < c \end{cases}$$

donde  $c$  y  $\gamma$  han de determinarse para conseguir el nivel de significación deseado. Concretamente, como  $T$  tiene distribución de Poisson de parámetro  $n\theta$ , ha de ser

$$\begin{aligned} \alpha &= E_{\theta_0}[\varphi] = P_{\theta_0}\{T > c\} + \gamma P_{\theta_0}\{T = c\} \\ &= \gamma e^{-n\theta_0} \frac{(n\theta_0)^c}{c!} + e^{-n\theta_0} \sum_{j=c+1}^{\infty} \frac{(n\theta_0)^j}{j!} \end{aligned}$$

o bien

$$1 - \alpha = e^{-n\theta_0} \sum_{j=0}^{c-1} \frac{(n\theta_0)^j}{j!} + (1 - \gamma) e^{-n\theta_0} \frac{(n\theta_0)^c}{c!}$$

lo cual permite determinar  $c$  y  $\gamma$ .

Por ejemplo, si  $\theta_0 = 0'8$ ,  $n = 10$  y  $\alpha = 0'02$ , resulta

$$c = 14 \quad \text{y} \quad \gamma = 0'16,$$

pues  $P_{\theta_0}\{T < 14\} = 0'9658$  y  $P_{\theta_0}\{T = 14\} = 0'0169$ .

Para  $\theta_1 = 1'1$ , la potencia del contraste es

$$\begin{aligned} E_{\theta_1}[\varphi] &= P_{\theta_1}\{T > 14\} + 0'16 P_{\theta_1}\{T = 14\} \\ &= e^{-11} \sum_{j=15}^{\infty} \frac{11^j}{j!} + 0'84 e^{-11} \frac{11^{14}}{14!} = 0'1576, \end{aligned}$$

lo cual supone una probabilidad de error de tipo II de  $0'8424$ .

Con nivel de significación  $\alpha = 1 - 0'9658 = 0'0342$ , habría resultado el test no aleatorizado de región crítica  $C^* = \{T \geq 14\}$ . Lo que la segunda versión del Lema de Neyman-Pearson permite —además de construir tests

aleatorizados óptimos de cualquier nivel— es asegurar que el test de región crítica  $C^* = \{T \geq 14\}$  es más potente que cualquier test aleatorizado de nivel de significación 0'0342.

Puede observarse, en todos los ejemplos anteriores, que el resultado de todos los tests diseñados se basa en un estadístico suficiente para la familia de distribuciones de la población. No es por casualidad:

*Si  $T$  es un estadístico suficiente, dado cualquier test  $\varphi$ , existe otro test  $\varphi'$ , función exclusivamente de  $T$ , que tiene la misma función de potencia que  $\varphi$ .* La razón es clara,  $\varphi' = E[\varphi | T]$  no depende de  $\theta$  y sólo depende de  $T$ ; además

$$E_\theta[\varphi'] = E_\theta[E_\theta[\varphi | T]] = E_\theta[\varphi].$$

Según ello, siempre que se pueda determinar un test  $\varphi$  uniformemente de máxima potencia,  $\varphi$  será función del estadístico suficiente minimal  $T$ . En el caso de que  $T$  sea unidimensional, ello supone una ayuda útil para averiguar la forma final de  $\varphi$ .

## 8.4. Contraste de hipótesis unilaterales y bilaterales

En las aplicaciones reales raramente se plantean contrastes de hipótesis simple frente a simple, por la sencilla razón de que las conjeturas alternativas nunca suelen ser tan precisas que estén definidas por un único valor del parámetro. Por otra parte, tampoco es posible analizar el contraste óptimo de  $H_0 : \theta \in \Theta_0$  frente a  $H_1 : \theta \in \Theta_1$ , siendo  $\Theta_0$  y  $\Theta_1$  conjuntos arbitrarios. Con tal grado de generalidad, poco se puede decir, excepto que normalmente no existirá test uniformemente de máxima potencia.

Afortunadamente muchas de las situaciones que se plantean en la práctica corresponden a parámetros unidimensionales con hipótesis de la forma

$$H_0 : \theta \leq \theta_0 \quad \text{frente a} \quad H_1 : \theta > \theta_0,$$

o bien

$$H_0 : \theta \geq \theta_0 \quad \text{frente a} \quad H_1 : \theta < \theta_0,$$

que se denominan *hipótesis unilaterales*.

En el ejemplo 8.3, cabe observar que el test obtenido, para cualquier valor de  $\alpha$ , no depende del valor de  $\theta_1$ , siempre que se mantenga la desigualdad  $\theta_1 > \theta_0$ . Ello significa que el test de máxima potencia para contrastar  $H_0 : \theta = \theta_0$  frente a  $H_1 : \theta = \theta_1$ , a nivel de significación  $\alpha$ , es también de máxima potencia si la alternativa es  $H_1 : \theta = \theta'_1$  (supuesto que  $\theta_1, \theta'_1 > \theta_0$ ) y, por consiguiente, se trata del test uniformemente de máxima potencia en el caso de que fuese  $H_1 : \theta > \theta_0$ .

Además es fácil ver que la función de potencia  $\beta(\theta) = E_\theta[\varphi]$  es creciente con  $\theta$ ; de manera que, como se cumple  $\beta(\theta_0) = \alpha$ , con más razón se cumplirá  $\beta(\theta) \leq \alpha$  para cada  $\theta < \theta_0$ . Ello indica que se trata de un test de nivel de significación  $\alpha$  para la hipótesis nula  $H_0 : \theta \leq \theta_0$ .

Luego resulta, en definitiva, que el test obtenido en el ejemplo 8.3 es un test de nivel significación  $\alpha$  y uniformemente de máxima potencia para contrastar  $H_0 : \theta \leq \theta_0$  frente a  $H_1 : \theta > \theta_0$ .

Puesto que una conclusión similar es válida para los ejemplos 8.1 y 8.2, puede pensarse que el test de máxima potencia y tamaño  $\alpha$  para contrastar  $\theta = \theta_0$  frente a  $\theta = \theta_1$  (con  $\theta_0 < \theta_1$ ), tiene las mismas propiedades en el contraste de  $\theta \leq \theta_0$  frente a  $\theta > \theta_0$ . En general, ello no es cierto; la validez de esta afirmación para los ejemplos de la sección anterior se basa en que las distribuciones utilizadas en ellos tienen la siguiente propiedad:

*La distribución  $\mathcal{P}_\theta$  de una muestra, dependiente de un parámetro unidimensional  $\theta$ , es de razón de verosimilitud monótona si existe un estadístico unidimensional  $T$  tal que, para cada  $\theta < \theta'$  fijos, el cociente de densidades o funciones de probabilidad*

$$\frac{f_{\theta'}(x_1, \dots, x_n)}{f_\theta(x_1, \dots, x_n)}$$

*(definido como  $+\infty$  si  $f_\theta = 0$  y  $f_{\theta'} > 0$ ) es una función creciente de  $T(x_1, \dots, x_n)$ .*

En realidad se podría decir “... función monótona de  $T$ ”; pero como si es función decreciente de  $T$ , es función creciente de  $-T$ , la definición se concreta en los términos anteriores, que permiten una expresión más explícita de los resultados. La condición anterior se verifica para muchas distribuciones usuales, como las de los ejemplos 8.1, 8.2 y 8.3. De hecho:

### Ejemplo 8.4

Si la distribución de la muestra es de tipo exponencial uniparamétrico, con función de densidad o probabilidad de la forma

$$f_\theta(x_1, \dots, x_n) = c(\theta)h(x_1, \dots, x_n)e^{q(\theta)T(x_1, \dots, x_n)}$$

y  $q(\theta)$  es una función monótona, el cociente

$$\frac{f_{\theta'}(x_1, \dots, x_n)}{f_\theta(x_1, \dots, x_n)} = \frac{c(\theta')}{c(\theta)}e^{[q(\theta')-q(\theta)]T(x_1, \dots, x_n)}$$

es una función creciente de  $T$  o de  $-T$ , según que  $q(\theta)$  sea creciente o decreciente.

Bajo la condición anterior no hay ninguna dificultad en obtener tests uniformemente de máxima potencia para contrastar hipótesis unilaterales.

**Teorema de Karlin-Rubin:** Si la distribución  $\mathcal{P}_\theta$  de la muestra tiene razón de verosimilitud monótona en el estadístico  $T(x_1, \dots, x_n)$ , entonces:

(a) Para cada  $\alpha \in (0, 1)$ , existe un test  $\varphi$  de tamaño  $\alpha$  y uniformemente de máxima potencia para contrastar  $H_0 : \theta \leq \theta_0$  frente a  $H_1 : \theta > \theta_0$ , que es de la forma

$$\varphi(x_1, \dots, x_n) = \begin{cases} 1 & \text{si } T(x_1, \dots, x_n) > c \\ \gamma & \text{si } T(x_1, \dots, x_n) = c \\ 0 & \text{si } T(x_1, \dots, x_n) < c. \end{cases}$$

(b) Todo test con función crítica de la forma anterior tiene función de potencia  $\beta(\theta)$  estrictamente creciente, mientras sea  $0 < \beta(\theta) < 1$ .

(c) Para cualquier test  $\varphi'$  con  $E_{\theta_0}[\varphi'] = \alpha$  se cumple  $E_\theta[\varphi] \leq E_\theta[\varphi']$  para cualquier  $\theta \leq \theta_0$ .

La conclusión (c) afirma que, una vez fijado el tamaño, el test  $\varphi$  no sólo consigue la menor probabilidad de error de tipo II posible, sino también la menor probabilidad posible de error de tipo I (entre los tests con dicho tamaño), para todo  $\theta \leq \theta_0$ .

El resultado anterior tiene su contrapartida simétrica, para el contraste  $H_0 : \theta \geq \theta_0$  frente a  $H_1 : \theta < \theta_0$ ; los tests uniformemente de máxima potencia tienen, en este caso, la misma expresión, con las desigualdades invertidas, y la función de potencia es decreciente en lugar de creciente.

Demostración: Conviene empezar demostrando la afirmación (b). Fijados  $\theta_1 < \theta_2$ , el lema de Neyman-Pearson asegura que cualquier test con función crítica

$$\varphi = \begin{cases} 1 & \text{si } f_{\theta_2} > k f_{\theta_1} \\ \gamma & \text{si } f_{\theta_2} = k f_{\theta_1} \\ 0 & \text{si } f_{\theta_2} < k f_{\theta_1} \end{cases}$$

es de máxima potencia para contrastar  $\theta = \theta_1$  frente a  $\theta = \theta_2$  y tiene tamaño  $E_{\theta_1}[\varphi]$ . Como la razón de verosimilitudes,  $f_{\theta_2}/f_{\theta_1}$  es creciente con  $T$ , se puede expresar

$$\varphi = \begin{cases} 1 & \text{si } T > c \\ \gamma & \text{si } T = c \\ 0 & \text{si } T < c \end{cases}$$

que no depende de  $\theta_2$  (aunque dependerá de  $\theta_1$ , a través de  $c$  y  $\gamma$ , en cuanto se fije su tamaño).

Como el test trivial de función crítica  $\varphi'$  constante, igual a  $E_{\theta_1}[\varphi]$ , tiene que tener menos potencia que  $\varphi$ , es

$$E_{\theta_2}[\varphi'] \leq E_{\theta_2}[\varphi]$$

o bien, puesto que  $E_{\theta_2}[\varphi'] = E_{\theta_1}[\varphi]$ ,

$$E_{\theta_1}[\varphi] \leq E_{\theta_2}[\varphi].$$

La última desigualdad no puede ser una igualdad, a no ser que  $\varphi'$  tenga la misma potencia que  $\varphi$  y sea por tanto de la forma indicada por el lema de Neyman-Pearson; lo cual obliga a que sea  $E_{\theta_1}[\varphi] = 0$  ó 1. Luego  $\beta(\theta) = E_{\theta}[\varphi]$  es estrictamente creciente en todo el intervalo en el que no sea 0, ni 1.

(a) Dado  $\alpha \in (0, 1)$ , el lema de Neyman-Pearson garantiza que se pueden determinar  $c$  y  $\gamma$  para que sea  $E_{\theta_0}[\varphi] = \alpha$ . Como la función de potencia es estrictamente creciente, será  $E_{\theta}[\varphi] < \alpha$  para todo  $\theta < \theta_0$ ; lo cual indica que  $\varphi$  es un test de tamaño  $\alpha$  para contrastar  $\theta \leq \theta_0$  frente a  $\theta > \theta_0$ .

Por otra parte, si  $\varphi'$  es cualquier test con  $E_{\theta}[\varphi'] \leq \alpha$  para  $\theta \leq \theta_0$ , será, en particular,  $E_{\theta_0}[\varphi'] \leq \alpha$ . Pero  $\varphi$  es de máxima potencia para contrastar  $\theta = \theta_0$  frente a  $\theta = \theta_1$  cualquiera que sea  $\theta_1 > \theta_0$ ; debe ser pues

$$E_{\theta_1}[\varphi'] \leq E_{\theta_1}[\varphi] \quad \text{para todo } \theta_1 > \theta_0$$

Según ello,  $\varphi$  es uniformemente de máxima potencia para el contraste de  $\theta \leq \theta_0$  frente a  $\theta > \theta_0$ .

(c) Para contrastar  $\theta = \theta_0$  frente a  $\theta = \theta_1$ , con  $\theta_1 < \theta_0$ , el lema de Neyman-Pearson indica que el test de función crítica  $1 - \varphi$  es de máxima potencia, siendo su tamaño  $1 - \alpha$ . Puesto que no depende de  $\theta_1$ ,  $1 - \varphi$  será también uniformemente de máxima potencia para  $H_0 : \theta = \theta_0$  frente a  $H_1 : \theta < \theta_0$ . Ello significa que, si  $\varphi''$  es cualquier test con  $E_{\theta_0}[\varphi''] \leq 1 - \alpha$ , se verifica

$$E_{\theta}[\varphi''] \leq E_{\theta}[1 - \varphi] \quad \text{para cualquier } \theta < \theta_0.$$

Por consiguiente, si  $\varphi'$  es cualquier test con  $E_{\theta_0}[\varphi'] = \alpha$ , como  $E_{\theta_0}[1 - \varphi'] = 1 - \alpha$ , tiene que ser

$$E_{\theta}[1 - \varphi'] \leq E_{\theta}[1 - \varphi] \quad \text{para cualquier } \theta < \theta_0,$$

o bien

$$E_{\theta}[\varphi'] \geq E_{\theta}[\varphi] \quad \text{para cualquier } \theta < \theta_0.$$

En las condiciones del teorema anterior, el procedimiento de construcción de tests uniformemente de máxima potencia para hipótesis unilaterales es similar al método del lema de Neyman-Pearson; incluso más simple porque la comprobación de las hipótesis indica directamente el estadístico en función del que se ha de expresar el test.

### Ejemplo 8.5

En la situación del ejemplo 8.3, la distribución de la muestra es de razón de verosimilitud monótona, puesto que si  $\theta_1 < \theta_2$

$$\frac{f_{\theta_2}(x_1, \dots, x_n)}{f_{\theta_1}(x_1, \dots, x_n)} = \frac{e^{-n\theta_2}\theta_2^T}{e^{-n\theta_1}\theta_1^T} = e^{n(\theta_1 - \theta_2)} \left( \frac{\theta_2}{\theta_1} \right)^T$$

es una función creciente de  $T = \sum_{i=1}^n x_i$ . Los tests de máxima potencia para contrastar  $H_0 : \theta \leq \theta_0$  frente a  $H_1 : \theta > \theta_0$  son entonces de la forma

$$\varphi(x_1, \dots, x_n) = \begin{cases} 1 & \text{si } T > c \\ \gamma & \text{si } T = c \\ 0 & \text{si } T < c \end{cases}$$

(que ya había proporcionado el lema de Neyman-Pearson para el contraste de  $\theta = \theta_0$  frente a  $\theta = \theta_1$ , con  $\theta_1 > \theta_0$ ). Puesto que la función de potencia es creciente, la determinación de  $c$  y  $\gamma$  para conseguir tamaño  $\alpha$  se consigue imponiendo

$$E_{\theta_0}[\varphi] = \alpha,$$

exactamente igual que en el ejemplo 8.3.

En cambio, para contrastar  $H_0 : \theta \geq \theta_0$  frente a  $H_1 : \theta < \theta_0$ , el teorema anterior (o su simétrico, mejor dicho) prescribe tests de la forma

$$\varphi(x_1, \dots, x_n) = \begin{cases} 1 & \text{si } T < c \\ \gamma & \text{si } T = c \\ 0 & \text{si } T > c, \end{cases}$$

igual que haría el lema de Neyman-Pearson para el contraste de  $\theta = \theta_0$  frente a  $\theta = \theta_1$ , con  $\theta_1 < \theta_0$ . Para obtener tamaño  $\alpha$  hay que imponer

$$\begin{aligned} \alpha &= E_{\theta_0}[\varphi] = P_{\theta_0}\{T < c\} + \gamma P_{\theta_0}\{T = c\} \\ &= e^{-n\theta_0} \sum_{j=1}^{c-1} \frac{(n\theta_0)^j}{j!} + \gamma e^{-n\theta_0} \frac{(n\theta_0)^c}{c!}. \end{aligned}$$

Por ejemplo, con  $\theta_0 = 0'8$ ,  $n = 10$  y  $\alpha = 0'02$ , se obtiene

$$c = 3 \quad \text{y} \quad \gamma = 0'22,$$

pues  $P_{\theta_0}\{T < 3\} = 0'0137$  y  $P_{\theta_0}\{T = 3\} = 0'0286$ .

La función de potencia del contraste resulta

$$\beta(\theta) = P_{\theta}\{T < 3\} + \gamma P_{\theta}\{T = 3\},$$

que evidentemente decrece con  $\theta$ , como afirma el resultado simétrico al enunciado.

### Ejemplo 8.6

Se dispone de una muestra aleatoria simple, de tamaño  $n$ , de una población con función de densidad

$$f_{\theta}(x) = e^{\theta-x} \quad \text{para } x \geq \theta.$$

La densidad de la muestra es entonces

$$f_{\theta}(x_1, \dots, x_n) = e^{n\theta - (x_1 + \dots + x_n)} I_{\{x_{(1)} \geq \theta\}}$$

y, para  $\theta < \theta'$ , la razón de verosimilitudes

$$\frac{f_{\theta'}(x_1, \dots, x_n)}{f_{\theta}(x_1, \dots, x_n)} = e^{n(\theta' - \theta)} \frac{I_{\{x_{(1)} \geq \theta'\}}}{I_{\{x_{(1)} \geq \theta\}}}$$

sólo está definida para  $x_{(1)} \geq \theta$ ; pero es función creciente de  $x_{(1)}$ , puesto que se anula si  $x_{(1)} \in [\theta, \theta')$  y vale  $e^{n(\theta' - \theta)}$  si  $x_{(1)} \in [\theta', \infty)$ .

La distribución de  $X_{(1)}$  es continua, con función de distribución

$$G_\theta(y) = 1 - [1 - F_\theta(y)]^n = \begin{cases} 0 & \text{si } y < \theta \\ 1 - e^{n(\theta-y)} & \text{si } y \geq \theta \end{cases}$$

Para el contraste de  $H_0 : \theta \leq \theta_0$  frente a  $H_1 : \theta > \theta_0$ , son uniformemente de máxima potencia los tests con función crítica de la forma

$$\varphi(x_1, \dots, x_n) = \begin{cases} 1 & \text{si } x_{(1)} > c \\ 0 & \text{si } x_{(1)} < c \end{cases}$$

en los cuales se puede incorporar  $x_{(1)} = c$  a cualquiera de los dos lados, puesto que hay probabilidad cero de que se produzca tal igualdad. Se trata pues de tests no aleatorizados, con región crítica de la forma  $\{x_{(1)} > c\}$  (la misma que se obtiene para el contraste de máxima potencia de hipótesis simples, cuando  $\theta_0 < \theta_1$ ).

En el gráfico siguiente se ha dibujado la densidad de  $X_{(1)}$ , para  $\theta_0$  y para otro valor de  $\theta$  (dentro de la hipótesis alternativa), permitiendo la representación del nivel de significación y de la probabilidad de error de tipo II.

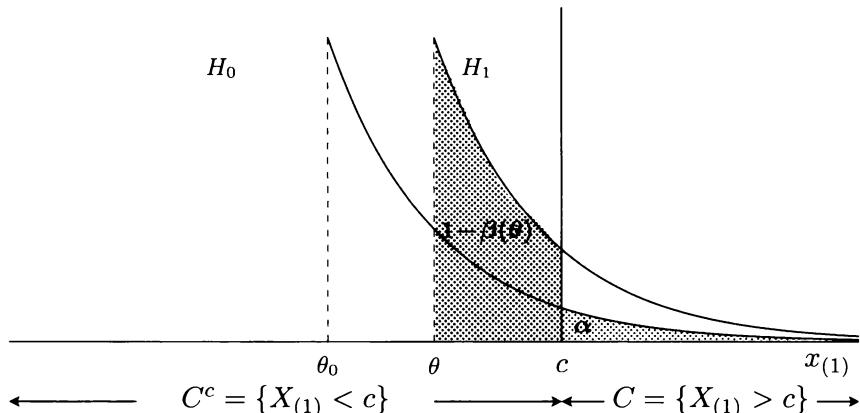


Figura 8.2: Test de hipótesis unilaterales

Al crecer  $\theta$  se observa claramente que la potencia aumenta; luego para conseguir el test de tamaño  $\alpha$ , ha de imponerse

$$\alpha = \mathcal{P}_{\theta_0} \{X_{(1)} > c\} = e^{n(\theta_0 - c)},$$

de manera que resulta  $c = \theta_0 - (\log \alpha)/n$  (que es mayor que  $\theta_0$  y se aleja de él al disminuir  $\alpha$ ).

La función de potencia vale concretamente

$$\beta(\theta) = \mathcal{P}_{\theta}\{X_{(1)} > c\} = \begin{cases} \alpha e^{n(\theta-\theta_0)} & \text{si } \theta \leq \theta_0 - (\log \alpha)/n \\ 1 & \text{si } \theta > \theta_0 - (\log \alpha)/n \end{cases}$$

que, como estaba previsto, es función creciente de  $\theta$ :

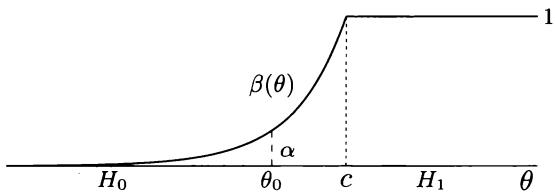


Figura 8.3: Función de potencia

El rango de valores de la probabilidad de error de tipo I es  $(0, \alpha]$ , mientras que las probabilidades de error de tipo II,  $1 - \beta(\theta)$ , varían en  $[0, 1 - \alpha]$ .

También puede expresarse de forma explícita el nivel crítico asociado a una muestra  $(x_1, \dots, x_n)$ :

$$\alpha(x_1, \dots, x_n) = \mathcal{P}_{\theta_0}\{X_{(1)} > x_{(1)}\} = \begin{cases} 1 & \text{si } x_{(1)} \leq \theta_0 \\ e^{n(\theta_0 - x_{(1)})} & \text{si } x_{(1)} > \theta_0, \end{cases}$$

que indica que, cuando  $x_{(1)} \leq \theta_0$ , no es posible rechazar  $H_0$ ; mientras que, a medida que  $x_{(1)}$  se aleja hacia la derecha de  $\theta_0$ , el apoyo que recibe  $H_0$  de las observaciones decrece rápidamente.

Supongamos ahora que las hipótesis del contraste son  $H_0 : \theta \geq \theta_0$  y  $H_1 : \theta < \theta_0$ ; el test uniformemente de máxima potencia y tamaño  $\alpha$  tiene entonces función crítica

$$\varphi(x_1, \dots, x_n) = \begin{cases} 1 & \text{si } x_{(1)} < c \\ 0 & \text{si } x_{(1)} \geq c, \end{cases}$$

donde, a fin de que sea  $\alpha = \mathcal{P}_{\theta_0}\{X_{(1)} < c\}$ , debe ser  $c = \theta_0 - n^{-1} \log(1 - \alpha)$  (que es mayor que  $\theta_0$ , pero se acerca hacia él al disminuir  $\alpha$ ). La interpretación gráfica puede observarse sobre la siguiente representación de las densidades de  $X_{(1)}$  correspondientes a  $\theta_0$  y a  $\theta$ .

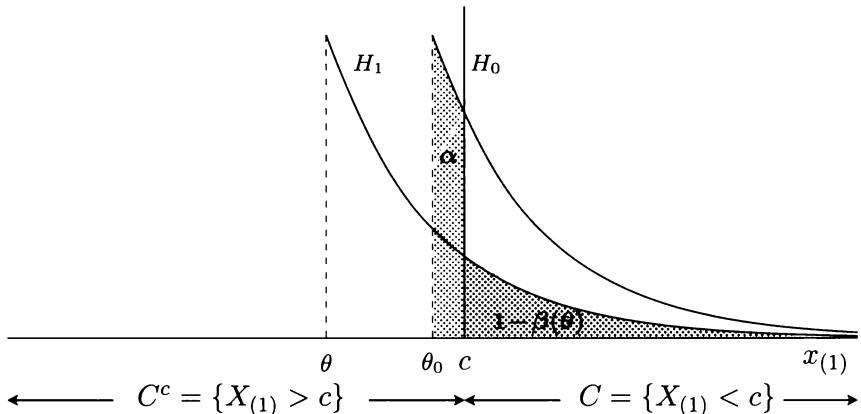


Figura 8.4: Test de hipótesis unilaterales

En este caso, la función de potencia es decreciente con  $\theta$ ; de hecho vale

$$\beta(\theta) = \mathcal{P}_\theta \{X_{(1)} < c\} = \begin{cases} 1 - (1 - \alpha)e^{n(\theta - \theta_0)} & \text{si } \theta \leq \theta_0 - n^{-1} \log(1 - \alpha) \\ 0 & \text{si } \theta > \theta_0 - n^{-1} \log(1 - \alpha) \end{cases}$$

cuya representación gráfica:

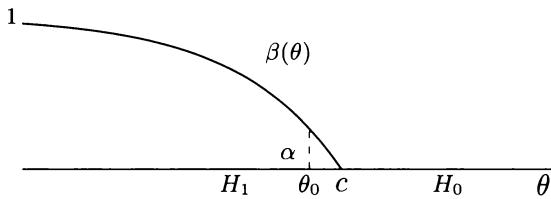


Figura 8.5: Función de potencia

muestra unas probabilidades de error de tipo I que varían en  $[0, \alpha]$  y unas probabilidades de error de tipo II que recorren el intervalo  $(0, 1 - \alpha)$ .

El nivel crítico asociado a cada muestra es ahora

$$\alpha(x_1, \dots, x_n) = \mathcal{P}_{\theta_0} \{X_{(1)} < x_{(1)}\} = \begin{cases} 0 & \text{si } x_{(1)} \leq \theta_0 \\ 1 - e^{n(\theta_0 - x_{(1)})} & \text{si } x_{(1)} > \theta_0, \end{cases}$$

que indica que se puede rechazar, con seguridad, si  $x_{(1)} \leq \theta_0$  (como es evidente); pero el riesgo de error de tipo I se aproxima a 1 en cuanto  $x_{(1)}$  se separa sensiblemente a la derecha de  $\theta_0$ .

Los resultados de ambos casos ponen de manifiesto que, en los contrastes frente a hipótesis alternativas compuestas, la comparación del nivel de

significación con la probabilidad de error de tipo II pierde la relevancia que tenía en el caso de alternativa simple. La amplitud de la hipótesis alternativa tiene por efecto que sea muy amplio el rango de riesgos de error de tipo II (desde 0 a  $1 - \alpha$ ), de manera que es inevitable que haya riesgos altos (si  $\alpha$  es pequeño) y muy bajos. El nivel de significación —o el nivel crítico— cobra así un papel predominante.

## Hipótesis bilaterales

Además de las hipótesis unilaterales, tienen una gran importancia práctica contrastes de hipótesis de la forma

$$H_0 : \theta = \theta_0 \quad \text{frente a} \quad H_1 : \theta \neq \theta_0,$$

o bien

$$H_0 : \theta \in [\theta_1, \theta_2] \quad \text{frente a} \quad H_1 : \theta \notin [\theta_1, \theta_2],$$

que se denominan *hipótesis bilaterales*. La situación es más delicada en estos casos, como puede ponerse de relieve mediante un ejemplo.

### Ejemplo 8.7

A partir de una muestra aleatoria simple de tamaño  $n$ , de una distribución  $N(\theta, \sigma^2)$ , con  $\sigma$  conocido, se desea contrastar  $H_0 : \theta = 0$  frente a  $H_1 : \theta \neq 0$ . La densidad de la muestra es de la forma

$$c(\theta)h(x_1, \dots, x_n)e^{q(\theta)T},$$

siendo  $c(\theta) = (\sqrt{2\pi}\sigma)^{-n}e^{-n\theta^2/(2\sigma^2)}$ ,  $h(x_1, \dots, x_n) = e^{-(x_1^2 + \dots + x_n^2)/(2\sigma^2)}$ ,  $q(\theta) = \theta/\sigma^2$  y  $T = \sum_{i=1}^n x_i$ ; luego tiene razón de verosimilitud monótona en  $T$ . Por tanto, para contrastar  $\theta = 0$  frente a  $\theta = \theta_1 > 0$  son uniformemente de máxima potencia los contrastes de la forma

$$\varphi_1 = \begin{cases} 1 & \text{si } T > c \\ 0 & \text{si } T < c, \end{cases}$$

donde hay que elegir  $c = z_{\alpha}\sigma\sqrt{n}$ , si se quiere que el contraste tenga el nivel de significación  $\alpha$  (habida cuenta que  $T$  tiene distribución  $N(0, \sigma\sqrt{n})$  si  $H_0$  es cierta). Y el contraste obtenido no depende de  $\theta_1$ .

Análogamente, para contrastar  $\theta = 0$  frente a  $\theta = \theta_2 < 0$  es uniformemente de máxima potencia, a nivel  $\alpha$ , el test

$$\varphi_2 = \begin{cases} 1 & \text{si } T < z_{1-\alpha}\sigma\sqrt{n} \\ 0 & \text{si } T > z_{1-\alpha}\sigma\sqrt{n} \end{cases}$$

que no depende de  $\theta_2$ .

Ahora bien, un test  $\varphi^*$  de nivel  $\alpha$  y uniformemente de máxima potencia para contrastar  $\theta = \theta_0$  frente a  $\theta \neq \theta_0$ , no puede tener menos potencia que  $\varphi_1$  para ningún valor de  $\theta \neq 0$ . Luego, para cualquier  $\theta_1 > 0$ , tendría que ser

$$E_{\theta_1}[\varphi^*] = E_{\theta_1}[\varphi_1]$$

(puesto que  $\varphi_1$  es de máxima potencia en  $\theta_1$ ). Por idéntica razón, para cualquier  $\theta_2 < 0$ , habría de ser

$$E_{\theta_2}[\varphi^*] = E_{\theta_2}[\varphi_2].$$

Según la afirmación (c) del lema de Neyman-Pearson, la primera igualdad obliga a que  $\varphi^*$  coincida con  $\varphi_1$  (salvo en conjuntos de probabilidad cero); y la segunda igualdad obliga a que  $\varphi^*$  coincida (casi seguro) con  $\varphi_2$ . Como ambas cosas son incompatibles,  $\varphi^*$  no existe.

Según esto, para las hipótesis bilaterales consideradas no existe en general test uniformemente de máxima potencia. Cabe entonces el recurso de excluir un gran número de tests, que no cumplan alguna propiedad adicional deseable, y restringir la búsqueda de tests óptimos a una clase más reducida, dentro de la cual pueda existir solución.

No entraremos aquí en detalle en esta cuestión, omitiendo por ejemplo cualquier restricción de la clase de tests por razones de invariancia. Sólo citaremos, para dar solución a algunas situaciones importantes, el concepto siguiente:

*Un test  $\varphi$  de tamaño  $\alpha$ , para contrastar  $H_0 : \theta \in \Theta_0$  frente a  $H_1 : \theta \in \Theta_1$ , se denomina insesgado si se cumple*

$$\beta(\theta) = E_\theta[\varphi] \geq \alpha \quad \text{para cualquier } \theta \in \Theta_1$$

*de manera que la probabilidad de rechazar  $H_0$  cuando es falsa es mayor que la de rechazarla cuando es cierta.*

Nótese que cualquier test  $\varphi$  uniformemente de máxima potencia es insesgado, pues, si su tamaño es  $\alpha$ , basta compararlo con el test de función crítica idénticamente igual a  $\alpha$ , para concluir

$$E_\theta[\varphi] \geq \alpha \quad \text{para todo } \theta \in \Theta_1.$$

Según ello, al restringirse al conjunto  $\mathcal{I}$  de los tests insesgados no es necesario revisar ninguna de las conclusiones anteriores: los tests óptimos en el conjunto de todos los tests, son los mejores dentro de  $\mathcal{I}$ .

En  $\mathcal{I}$  tienen solución muchos contrastes bilaterales del tipo propuesto. Concretamente

**Teorema de Lehmann<sup>5</sup>:** Si la distribución  $\mathcal{P}_\theta$  de una muestra es de tipo exponencial uniparamétrico, es decir con función de densidad o de probabilidad de la forma

$$f_\theta(x_1, \dots, x_n) = c(\theta)h(x_1, \dots, x_n)e^{q(\theta)T(x_1, \dots, x_n)}$$

para cada  $\alpha \in (0, 1)$ , existe un test de tamaño  $\alpha$ , uniformemente de máxima potencia dentro de la familia  $\mathcal{I}$  de tests insesgados, para el contraste de hipótesis bilaterales. Su función crítica es de la forma

$$\varphi(x_1, \dots, x_n) = \begin{cases} 1 & \text{si } T(x_1, \dots, x_n) < c_1 \text{ ó } T(x_1, \dots, x_n) > c_2 \\ \gamma_1 & \text{si } T(x_1, \dots, x_n) = c_1 \\ \gamma_2 & \text{si } T(x_1, \dots, x_n) = c_2 \\ 0 & \text{si } c_1 < T(x_1, \dots, x_n) < c_2 \end{cases}$$

donde, si el tamaño es  $\alpha$ , las constantes  $c_1, c_2, \gamma_1$  y  $\gamma_2$  han de verificar las condiciones:

$$E_{\theta_1}[\varphi] = \alpha \quad \text{y} \quad E_{\theta_2}[\varphi] = \alpha$$

en caso de ser  $H_0 : \theta \in [\theta_1, \theta_2]$ ; o bien

$$E_{\theta_0}[\varphi] = \alpha \quad \text{y} \quad E_{\theta_0}[\varphi T] = \alpha E_{\theta_0}[T]^6$$

en el caso  $H_0 : \theta = \theta_0$ . Todos los tests de la forma indicada, que verifiquen las condiciones anteriores, tienen la misma función de potencia.

<sup>5</sup> Cf. Lehmann [1986] (pág. 135 y siguientes) y Ferguson [1967] (pág. 215 y siguientes).

<sup>6</sup> Esta última condición resulta de imponer que la función de potencia de cualquier test insesgado  $\varphi$  para contrastar  $\theta = \theta_0$ , ha de tener un mínimo en  $\theta_0$ . Como

$$\frac{d}{d\theta} E_\theta[\varphi] = \int_{\mathcal{X}} \varphi \frac{\partial}{\partial \theta} \left[ c(\theta)h(x_1, \dots, x_n)e^{q(\theta)T} \right] dx_1 \dots dx_n = q'(\theta)E_\theta[\varphi T] + \frac{c'(\theta)}{c(\theta)}E_\theta[\varphi]$$

y en particular, para  $\varphi$  idénticamente igual a  $\alpha$ , se obtiene  $0 = q'(\theta_0)E_{\theta_0}[\varphi T] + c'(\theta_0)/c(\theta_0)$ , tiene que ser

$$0 = q'(\theta_0)E_{\theta_0}[\varphi T] + \frac{c'(\theta_0)}{c(\theta_0)}E_{\theta_0}[\varphi] = q'(\theta_0)(E_{\theta_0}[\varphi T] - E_{\theta_0}[T]\alpha)$$

Si la distribución de  $T$  es continua, no hay necesidad de incluir las alternativas correspondientes a los sucesos de probabilidad cero:  $T = c_1$  y  $T = c_2$ . Pero, aún en caso contrario, la aplicación sólo supone dificultades de cálculo.

### Ejemplo 8.8

Se afirma que los portadores de un virus desarrollan, en un año, síntomas de la enfermedad en una proporción  $\theta_0 = 1/4$  de los casos. Para comprobar esta afirmación, un equipo médico ha seleccionado al azar 20 de sus pacientes portadores del virus, para contrastar, al nivel de significación  $\alpha = 0'1$ , la hipótesis  $H_0 : \theta = \theta_0$  frente a la alternativa  $H_1 : \theta \neq \theta_0$ .

Valorando como 1 los casos afirmativos y 0 los negativos, la función de probabilidad del resultado de  $n$  observaciones es

$$f_\theta(x_1, \dots, x_n) = \theta^T(1 - \theta)^{n-T} = c(\theta)e^{q(\theta)T} \quad \text{para } x_1, \dots, x_n = 0, 1,$$

siendo  $T = \sum_{i=1}^n x_i$ ,  $c(\theta) = (1 - \theta)^n$  y  $q(\theta) = \log \theta - \log(1 - \theta)$ .

El teorema anterior proporciona la forma del test insesgado, uniformemente de máxima potencia. Su resultado no puede ser más intuitivo; como  $T$  representa el número total de enfermos, hay que rechazar  $\theta = 1/4$ , si son muy pocos (menos de  $c_1$ ) o muchos (más de  $c_2$ ) y aceptarlo cuando  $T$  está (estrictamente) entre dichos límites. En los casos de igualdad, hay que aleatorizar, con probabilidades  $\gamma_1, \gamma_2$ , para poder alcanzar el tamaño deseado.

Supuesto que  $\theta = 1/4$ ,  $T$  tiene distribución binomial  $B(20, 1/4)$ ; así que

$$\mathcal{P}_{1/4}\{T = k\} = \binom{20}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{n-k} \quad \text{y} \quad E_{1/4}[T] = 5.$$

Las condiciones indicadas en el teorema anterior se traducen pues en

$$\left\{ \begin{array}{ll} \sum_{k < c_1, k > c_2} \binom{20}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{20-k} & + \gamma_1 \binom{20}{c_1} \left(\frac{1}{4}\right)^{c_1} \left(\frac{3}{4}\right)^{20-c_1} \\ & + \gamma_2 \binom{20}{c_2} \left(\frac{1}{4}\right)^{c_2} \left(\frac{3}{4}\right)^{20-c_2} = 0'1, \\ \sum_{k < c_1, k > c_2} k \binom{20}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{20-k} & + \gamma_1 c_1 \binom{20}{c_1} \left(\frac{1}{4}\right)^{c_1} \left(\frac{3}{4}\right)^{20-c_1} \\ & + \gamma_2 c_2 \binom{20}{c_2} \left(\frac{1}{4}\right)^{c_2} \left(\frac{3}{4}\right)^{20-c_2} = 0'5. \end{array} \right.$$

Cualesquiera valores que verifiquen estas condiciones, proporcionan un test insesgado uniformemente de máxima potencia (entre los insesgados).

Como

$$\sum_{k < c_1} \binom{20}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{20-k} = \begin{cases} 0'0243 & \text{para } c_1 = 2 \\ 0'0913 & \text{para } c_1 = 3, \end{cases}$$

$$\sum_{k > c_2} \binom{20}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{20-k} = \begin{cases} 0'0409 & \text{para } c_2 = 8 \\ 0'1018 & \text{para } c_2 = 7, \end{cases}$$

puede elegirse  $c_1 = 2$  y  $c_2 = 8$ . Entonces,

$$\sum_{k < c_1, k > c_2} k \binom{20}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{20-k} = 0'4084,$$

con lo cual resulta el sistema de ecuaciones

$$\begin{cases} 0'0669\gamma_1 + 0'0609\gamma_2 = 0'0348 \\ 0'1339\gamma_1 + 0'4871\gamma_2 = 0'0916, \end{cases}$$

que da  $\gamma_1 = 0'4643$  y  $\gamma_2 = 0'0603$ . En definitiva, el test obtenido es

$$\varphi = \begin{cases} 1 & \text{si } T < 2 \text{ ó } T > 8 \\ 0'4643 & \text{si } T = 2 \\ 0'0603 & \text{si } T = 8 \\ 0 & \text{si } 2 < T < 8. \end{cases}$$

Si  $n$  fuese más grande (superior a 30), además de disponer de más información, se podría emplear un método más simple y sin mucho error. En efecto, la distribución de  $T$  sería próxima a una  $N(n\theta, \sqrt{n\theta(1-\theta)})$ ; de manera que se consigue una buena aproximación con el test

$$\varphi = \begin{cases} 1 & \text{si } T < c_1 \text{ ó } T > c_2 \\ 0 & \text{si } c_1 < T < c_2, \end{cases}$$

donde  $c_1$  y  $c_2$  deben cumplir

$$\begin{cases} \mathcal{P}_{\theta_0}\{T < c_1\} + \mathcal{P}_{\theta_0}\{T > c_2\} = \alpha, \\ E_{\theta_0} \cdot [TI_{\{T < c_1\} \cup \{T > c_2\}}] = \alpha n \theta_0 \end{cases}$$

La simetría de la distribución de  $T$  alrededor de  $n\theta_0$  sugiere que se puede tomar  $c_1 = n\theta_0 - c$  y  $c_2 = n\theta_0 + c$ ; en tal caso, como

$$E_{\theta_0} [(T - n\theta_0)I_{\{T < n\theta_0 - c\} \cup \{T > n\theta_0 + c\}}] = 0,$$

la segunda condición se reduce a

$$n\theta_0 \mathcal{P}_{\theta_0} (\{T < n\theta_0 - c\} \cup \{T > n\theta_0 + c\}) = \alpha n \theta_0,$$

que coincide con la primera condición. La solución es, por tanto,

$$c_1 = n\theta_0 - z_{\alpha/2} \sqrt{n\theta_0(1 - \theta_0)} \quad \text{y} \quad c_2 = n\theta_0 + z_{\alpha/2} \sqrt{n\theta_0(1 - \theta_0)}.$$

Puede observarse que los resultados expuestos en esta sección sólo se ocupan de distribuciones poblacionales que dependan de un único parámetro. Las dificultades aumentan notablemente incluso en el caso de familias exponenciales multiparamétricas, con función de densidad o de probabilidad de la forma

$$c(\theta) h(x_1, \dots, x_n) e^{\theta_1 T_1(x_1, \dots, x_n) + \dots + \theta_p T_p(x_1, \dots, x_n)}$$

sobre las que se desea efectuar un contraste de hipótesis (unilaterales o bilaterales) acerca de uno sólo de los parámetros, sin involucrar a los restantes. Es posible, en esta situación, construir tests insesgados uniformemente de máxima potencia, pero a costa de un trabajo teórico considerable, necesario aunque sólo sea para describir las conclusiones.

Una excepción, de gran interés práctico, la constituyen las poblaciones con distribución  $N(\mu, \sigma)$ , en las que ambos parámetros son desconocidos. Para ellas, al menos, los tests insesgados uniformemente de máxima potencia son sencillos de describir:

- ▷ Para  $H_0 : \mu = \mu_0$  frente a  $H_1 : \mu \neq \mu_0$ , la región crítica insesgada óptima es

$$C^* = \left\{ \sqrt{n-1} \frac{|\bar{x} - \mu_0|}{s} > k \right\},$$

donde, con nivel de significación  $\alpha$ , el valor crítico debe ser  $k = t_{n-1, \alpha/2}$  (puesto que el estadístico de Student tiene distribución  $t_{n-1}$ ).

- ▷ Para  $H_0 : \mu \leq \mu_0$  frente a  $H_1 : \mu > \mu_0$  tiene las mismas propiedades el test

$$C^* = \left\{ \sqrt{n-1} \frac{\bar{x} - \mu_0}{s} > t_{n-1; \alpha} \right\}$$

o el simétrico cuando se invierten las hipótesis.

- ▷ Para  $H_0 : \sigma = \sigma_0$  frente a  $H_1 : \sigma \neq \sigma_0$ , es insesgado óptimo el test

$$C^* = \left\{ \frac{(n-1)s^2}{\sigma_0^2} < c_1 \right\} \cup \left\{ \frac{(n-1)s^2}{\sigma_0^2} > c_2 \right\}$$

donde, para alcanzar tamaño  $\alpha$ ,  $c_1$  y  $c_2$  deben cumplir

$$\int_{c_1}^{c_2} f_{n-1}(y) dy = 1 - \alpha \quad \text{y} \quad \int_{c_1}^{c_2} y f_{n-1}(y) dy = (n-1)(1-\alpha),$$

siendo  $f_{n-1}$  la densidad de la distribución  $\chi_{n-1}^2$  (correspondiente a  $(n-1)s^2/\sigma_0^2$ ). La proximidad de la solución de estas ecuaciones a  $c_1 = \chi_{n-1;1-\alpha/2}^2$  y  $c_2 = \chi_{n-1;\alpha/2}^2$ , hace que estos últimos valores se empleen corrientemente en las aplicaciones.

- ▷ Para  $H_0 : \sigma \leq \sigma_0$  frente a  $H_1 : \sigma > \sigma_0$  resulta

$$C^* = \left\{ \frac{(n-1)s^2}{\sigma_0^2} > \chi_{n-1;\alpha}^2 \right\}$$

y el test simétrico para las hipótesis invertidas.

Como se ha indicado, la deducción de que los tests anteriores son uniformemente de máxima potencia entre los tests insesgados, comporta dificultades que aconsejan evitarla en este primer estudio de la teoría. Sobre todo teniendo en cuenta que los métodos del próximo capítulo para diseñar tests razonables, cualquiera que sean las circunstancias, darán soluciones muy similares a estos problemas.

## Ejercicios

### Ejercicio 8.1

Un fabricante de máquinas-herramienta afirma que sus productos tienen una probabilidad  $\theta \leq 0'1$  de fallar durante cada jornada de trabajo (y en tal caso necesitan un ajuste que les impide funcionar hasta el día siguiente). Sin embargo, la competencia afirma que es  $\theta = 0'6$ . Una empresa que ha adquirido una de tales máquinas ha observado que los números de días de funcionamiento satisfactorio entre los 8 primeras fallos han sido: 2, 5, 0, 3, 4, 4, 1, 3.

- a) Determinar el test de máxima potencia para contrastar  $\theta = 0'1$  frente a  $\theta = 0'6$ , de tamaño 0'01. ¿Es sensato utilizar dicho test? Determinar el test de máxima potencia que haga mínimo  $2\alpha + 1 - \beta$  siendo  $\alpha$  y  $1 - \beta$  las probabilidades de error de tipo I y II respectivamente.
- b) Contrastar la afirmación del fabricante frente a su contraria. Determinar el nivel crítico asociado a la muestra obtenida.

### Ejercicio 8.2

La empresa del ejercicio anterior, desechada la información del fabricante y basándose en su propia experiencia, decide comunicar a sus filiales que  $\theta$  está comprendida entre 0'3 y 0'4. Una de ellas pide una máquina a prueba y observa que los 5 primeros fallos se producen los días: 4, 10, 13, 18, 21. Con nivel de significación 0'1, ¿puede comunicar a la empresa que su estimación es errónea?

### Ejercicio 8.3

El error de medida de un amperímetro tiene distribución normal de media cero y desviación típica  $\theta_0$ , según afirmación del fabricante. Para comprobar la calibración, se han tomado  $n$  medidas independientes de una corriente patrón, lo que permite conocer los errores cometidos:  $x_1, \dots, x_n$ .

- a) Contrastar  $H_0 : \theta = \theta_0$  frente a  $H_1 : \theta = \theta_1$ , con nivel de significación  $\alpha$ , distinguiendo el caso  $\theta_1 < \theta_0$  y  $\theta_1 > \theta_0$ .
- b) Contrastar  $H_0 : \theta \leq \theta_0$  frente a  $H_1 : \theta > \theta_0$ , con nivel de significación  $\alpha$ . Hallar la función de potencia del test.

### Ejercicio 8.4

El número diario de llamadas atendidas por una centralita telefónica tiene distribución de Poisson de parámetro  $\lambda$ . Se considera que la centralita no es rentable y debe ser suprimida si  $\lambda < \lambda_0$ . A fin de estudiar la conveniencia de cerrarla se registra el numero de llamadas durante  $n$  días.

- a) Obtener el test de tamaño  $\alpha$  y uniformemente de máxima potencia para contrastar  $H_0 : \lambda \geq \lambda_0$  frente a  $H_1 : \lambda < \lambda_0$ . Aplicarlo al caso:  $\lambda_0 = 50$ ,  $n = 12$  y  $\alpha = 0'025$ . Con tales especificaciones, ¿qué decisión debe adoptarse si se ha obtenido la muestra: 32, 63, 39, 31, 42, 73, 37, 58, 46, 45, 30, 44? ¿Cuál es el nivel crítico asociado a ella?

- b) El riesgo de cerrar una centralita claramente rentable:  $\lambda \geq 55$ , se desea que sea menor que  $0'001$  y el de mantener abierta una centralita muy poco rentable:  $\lambda \leq 45$ , se quiere que sea menor que  $0'005$ . ¿Cuál es el tamaño de la muestra necesario para que el test diseñado cumpla tales especificaciones?

### Ejercicio 8.5

Los tiempos de vida de  $n$  válvulas siguen una distribución exponencial de parámetro  $\theta$ . La observación de los instantes de fallo se produce, por supuesto, en orden creciente. Cuando se han observado los  $r$  primeros:

- a) Determinar la región crítica de un test de nivel de significación  $\alpha$  y máxima potencia para contrastar  $H_0 : \theta = \theta_0$  frente a  $H_1 : \theta = \theta_1$ , distinguiendo el caso  $\theta_1 < \theta_0$  y el caso  $\theta_1 > \theta_0$ .
- b) Si  $r = 9$  y  $\alpha = 0'05$  determinar el test insesgado uniformemente de máxima potencia para contrastar  $H_0 : \theta = \theta_0$  frente a  $H_1 : \theta \neq \theta_0$ .

### Ejercicio 8.6

Se desea contrastar, con nivel de significación  $0'05$ , si una muestra aleatoria simple de tamaño  $n$  proviene de una densidad uniforme en  $(0, 1)$ , frente a la alternativa de que la densidad es  $2x$  para  $x \in (0, 1)$ . Determinar el contraste de máxima potencia y calcular su potencia en los casos: a)  $n = 1$ ; b)  $n = 2$ ; c)  $n = 3$ ; d)  $n$  grande.

### Ejercicio 8.7

Cierto programa de ordenador se ha preparado para producir números al azar en  $(0, \theta_0)$ . Después de realizadas  $n$  repeticiones del programa, se sospecha que, por error, los resultados son uniformes en  $(0, \theta_1)$ , donde es  $\theta_1 < \theta_0$ . Sea  $\theta$  el verdadero extremo de la distribución uniforme de los números producidos.

- a) Determinar un test de máxima potencia, de nivel de significación  $\alpha$ , para contrastar  $H_0 : \theta = \theta_0$  frente a  $H_1 : \theta = \theta_1$ .
- b) Determinar un test uniformemente de máxima potencia, de nivel de significación  $\alpha$ , para contrastar  $H_0 : \theta \geq \theta_0$  frente a  $H_1 : \theta < \theta_0$ .
- c) Determinar un test uniformemente de máxima potencia, de nivel de significación  $\alpha$ , para contrastar  $H_0 : \theta = \theta_0$  frente a  $H_1 : \theta \neq \theta_0$ .

**Ejercicio 8.8**

Para una población con densidad

$$f_\theta(x) = e^{\theta-x} \quad \text{si } x \geq \theta,$$

se desea contrastar  $H_0 : \theta = \theta_0$  frente a  $H_1 : \theta \neq \theta_0$ , mediante una muestra aleatoria simple de tamaño  $n$ . Determinar un test uniformemente de máxima potencia y nivel de significación  $\alpha$ .

**Ejercicio 8.9**

Se dispone de una muestra aleatoria simple de una población con función de densidad

$$f_\theta(x) = \frac{2}{(x-\theta)^3} \quad \text{para } x \geq \theta + 1.$$

- a) Diseñar un test de máxima potencia y nivel de significación  $\alpha$ , para contrastar  $H_0 : \theta = \theta_0$  frente a  $H_1 : \theta = \theta_1$ , siendo  $\theta_1 > \theta_0$ . ¿De qué tamaño conviene tomar la muestra?
- b) Deducir si existen tests uniformemente de máxima potencia para el contraste de  $H_0 : \theta \leq \theta_0$  frente a  $\theta > \theta_0$ .

**Ejercicio 8.10**

Se dispone de una muestra aleatoria simple de tamaño  $n$  de una población con densidad

$$f_{\theta,\lambda}(x) = \frac{1}{\lambda} e^{(\theta-x)/\lambda} \quad \text{para } x \geq \theta,$$

siendo  $\lambda > 0$ . Sea  $\theta_1 < \theta_0$ . Determinar el test de nivel de significación  $\alpha$  y máxima potencia para el contraste de  $H_0 : \theta = \theta_0, \lambda = \lambda_0$  frente a  $H_1 : \theta = \theta_1, \lambda = \lambda_1$ , distinguiendo los casos: a)  $\lambda_1 > \lambda_0$  y b)  $\lambda_1 < \lambda_0$ . Deducir si existe test uniformemente de máxima potencia para contrastar

- i)  $H_0 : \theta = \theta_0, \lambda = \lambda_0$  frente a  $H_1 : \theta < \theta_0, \lambda > \lambda_0$ .
- ii)  $H_0 : \theta = \theta_0, \lambda = \lambda_0$  frente a  $H_1 : \theta < \theta_0, \lambda < \lambda_0$ .
- iii)  $H_0 : \theta = \theta_0, \lambda = \lambda_0$  frente a  $H_1 : \theta = \theta_1$ .
- iv)  $H_0 : \theta \geq \theta_0, \lambda \leq \lambda_0$  frente a  $H_1 : \theta = \theta_1, \lambda = \lambda_1$ , siendo  $\lambda_1 > \lambda_0$ .

## Solución de los ejercicios

### Ejercicio 8.1

a) Suponiendo que la máquina una vez reparada queda como nueva, los lapsos (en número de días) entre dos averías son independientes y tienen distribución geométrica, de función de probabilidad

$$f_\theta(x) = (1 - \theta)^x \theta \quad \text{para } x = 0, 1, 2, \dots$$

La función de probabilidad de las observaciones es por tanto

$$f_\theta(x_1, \dots, x_n) = \theta^n (1 - \theta)^{n\bar{x}}$$

y los tests de máxima potencia para contrastar  $\theta = 0'1$  frente a  $\theta = 0'6$ , tienen función crítica

$$\varphi = \begin{cases} 1 & \text{si } 0'6^n 0'4^{n\bar{x}} > k 0'1^n 0'9^{n\bar{x}} \\ \gamma & \text{si } 0'6^n 0'4^{n\bar{x}} = k 0'1^n 0'9^{n\bar{x}} \\ 0 & \text{si } 0'6^n 0'4^{n\bar{x}} < k 0'1^n 0'9^{n\bar{x}} \end{cases} = \begin{cases} 1 & \text{si } n\bar{x} < c \\ \gamma & \text{si } n\bar{x} = c \\ 0 & \text{si } n\bar{x} > c. \end{cases}$$

El total de la muestra,  $n\bar{x}$ , tiene distribución binomial negativa de parámetros  $n$  y  $\theta$ ; es decir

$$\mathcal{P}_\theta\{n\bar{X} = j\} = \binom{n+j-1}{j} (1 - \theta)^j \theta^n \quad \text{para } j = 0, 1, 2, \dots$$

Para que el test tenga tamaño  $\alpha$ , debe ser pues

$$\sum_{j=0}^{c-1} \mathcal{P}_{0'1}\{n\bar{X} = j\} + \gamma \mathcal{P}_{0'1}\{n\bar{X} = c\} = \alpha.$$

Para  $\alpha = 0'01$ , como  $n = 8$  y

$$\sum_{j=0}^{c-1} \mathcal{P}_{0'1}\{n\bar{X} = j\} = \begin{cases} 0'00959 & \text{si } c = 24 \\ 0'0117 & \text{si } c = 25, \end{cases}$$

ha de ser  $c = 24$  y  $\gamma = 0'194$ . La muestra obtenida da un total  $n\bar{x} = 22$ , luego puede rechazarse  $H_0 : \theta = 0'1$  al nivel de significación  $\alpha = 0'01$ .

Sin embargo, usar este nivel de significación para el contraste en cuestión no tiene mucho sentido, puesto que la probabilidad de error de tipo II es entonces

$$1 - \mathcal{P}_{0'6}\{n\bar{X} < 24\} - 0'194 \mathcal{P}_{0'6}\{n\bar{X} = 24\} = 0'000024.$$

Por tanto, se puede disminuir sensiblemente el nivel de significación sin que la probabilidad de error de tipo II pase de límites aceptables.

Una manera de equilibrar ambos riesgos es la que propone el enunciado: si la pérdida asociada al error de tipo I es el doble que la del error de tipo II, en vez de fijar  $\alpha$ , se puede determinar su valor para que  $2\alpha + 1 - \beta$  sea mínimo. Concretamente,  $2\alpha - \beta$  vale

$$2\mathcal{P}_{0'1}\{n\bar{X} < c\} + 2\gamma\mathcal{P}_{0'1}\{n\bar{X} = c\} - \mathcal{P}_{0'6}\{n\bar{X} < c\} - \gamma\mathcal{P}_{0'6}\{n\bar{X} = c\}$$

$$= \sum_{j=0}^{c-1} \binom{j+7}{j} [2 \cdot 1^8 \cdot 9^j - 6^8 \cdot 4^j] + \gamma \binom{c+7}{c} [2 \cdot 1^8 \cdot 9^c - 6^8 \cdot 4^c].$$

Los sumandos son negativos para  $j \leq 16$  y positivos para  $j \geq 17$ ; el mínimo valor de  $2\alpha - \beta$  corresponde pues a  $c = 17$  y  $\gamma = 0$ . Ello supone un tamaño

$$\alpha = \sum_{j=0}^{16} \binom{j+7}{j} 1^8 9^j = 0'00168$$

y un error de tipo II

$$1 - \beta = 1 - \sum_{j=0}^{16} \binom{j+7}{j} 6^8 4^j = 0'0022.$$

Respecto al test anterior el riesgo de error de tipo I ha disminuido del 1% al 0'17% sin aumentar el riesgo de error de tipo II más que al 0'22%. La muestra obtenida, con  $n\bar{x} = 22$  no permite rechazar la hipótesis nula con tal nivel de significación.

b) Eliminando ahora la exageración de la competencia, se trata ahora de averiguar si los datos son significativos de que es falsa  $H_0 : \theta \leq 0'1$  frente a  $H_1 : \theta > 0'1$ .

La distribución de la muestra tiene razón de verosimilitud monótona en  $-n\bar{x}$ , puesto que es de la forma

$$f_\theta(x_1, \dots, x_n) = \theta^n e^{q(\theta) n\bar{x}},$$

siendo  $q(\theta) = \log(1 - \theta)$ , función decreciente de  $\theta \in (0, 1)$ . Según el teorema de Karlin-Rubin, los tests uniformemente de máxima potencia para el contraste propuesto tienen entonces función crítica de la forma

$$\varphi = \begin{cases} 1 & \text{si } n\bar{x} < c \\ \gamma & \text{si } n\bar{x} = c \\ 0 & \text{si } n\bar{x} > c, \end{cases}$$

que lógicamente es la misma determinada en el apartado anterior para cualquier contraste de  $\theta = \theta_0$  frente a  $\theta = \theta_1$ , con  $\theta_1 > \theta_0$ .

Ya sabemos que, con nivel de significación  $\alpha = 0'01$ , resulta  $c = 24$  y la muestra permite rechazar  $\theta \leq 0'1$  y concluir  $\theta > 0'1$ .

Para que la muestra obtenida obligue a rechazar la hipótesis nula, ha de ser  $c = 23$  y  $\gamma = 0$ ; el nivel crítico es entonces

$$\alpha(2, 5, 0, 3, 4, 4, 1, 3) = \sum_{j=0}^{22} \binom{j+7}{j} 0'1^8 0'9^j = 0'00778,$$

de manera que los datos son fuertemente significativos de que  $\theta > 0'1$ .

### Ejercicio 8.2

El contraste es ahora de la hipótesis nula  $H_0 : \theta \in [0'3, 0'4]$  frente a la alternativa  $H_1 : \theta \in [0, 1] - [0'3, 0'4]$ ; se trata pues de un contraste bilateral.

Si los datos se transforman para hacer aparecer los números de días sin fallos entre dos fallos consecutivos, se obtiene  $x_1 = 3, x_2 = 5, x_3 = 2, x_4 = 4$  y  $x_5 = 2$ ; mientras que la función de probabilidad que rige la aparición de  $n$  de tales observaciones sigue siendo:

$$f_\theta(x_1, \dots, x_n) = \theta^n e^{n\bar{x} \log(1-\theta)}.$$

Según el teorema de Lehmann, un test insesgado uniformemente de máxima potencia para contrastar  $H_0$  frente a  $H_1$ , tiene función crítica de la forma

$$\varphi = \begin{cases} 1 & \text{si } n\bar{x} < c_1 \text{ ó } n\bar{x} > c_2 \\ \gamma_1 & \text{si } n\bar{x} = c_1 \\ \gamma_2 & \text{si } n\bar{x} = c_2 \\ 0 & \text{si } c_1 < n\bar{x} < c_2, \end{cases}$$

donde, para que el test tenga nivel de significación  $\alpha = 0'1$ , debe ser

$$E_\theta[\varphi] = 0'1 \quad \text{para } \theta = 0'3, 0'4;$$

es decir

$$\mathcal{P}_\theta\{n\bar{X} < c_1\} + \gamma_1 \mathcal{P}_\theta\{n\bar{X} = c_1\} + \gamma_2 \mathcal{P}_\theta\{n\bar{X} = c_2\} + \mathcal{P}_\theta\{n\bar{X} > c_2\} = 0'1.$$

Las distribuciones binomiales negativas de parámetros  $n = 5, \theta = 0'3$  y  $\theta = 0'4$  respectivamente, proporcionan:

$c$	$\mathcal{P}_{0'3}\{n\bar{X} < c\}$	$\mathcal{P}_{0'3}\{n\bar{X} = c\}$	$\mathcal{P}_{0'3}\{n\bar{X} > c\}$
2	0'010935	0'017860	
3	0'028790	0'029172	
4	0'057962	0'040841	
$\vdots$	$\vdots$	$\vdots$	$\vdots$
20		0'020603	0'090472
21		0'017169	0'073302
22		0'014204	0'059098
23		0'011672	0'047426
24		0'009532	0'037895

$c$	$\mathcal{P}_{0'4}\{n\bar{X} < c\}$	$\mathcal{P}_{0'4}\{n\bar{X} = c\}$	$\mathcal{P}_{0'4}\{n\bar{X} > c\}$
2	0'040960	0'005529	
3	0'096256	0'077414	
4	0'173670	0'092897	
$\vdots$	$\vdots$	$\vdots$	$\vdots$
20		0'003978	0'009471
21		0'002842	0'006629
22		0'002015	0'004614
23		0'001419	0'000993
24		0'003195	0'002201

Después de algunas pruebas con  $c_1 = 2$  ó  $c_1 = 3$  se encuentra que  $c_1 = 2$ ,  $c_2 = 22$  dan como condiciones

$$0'010935 + 0'01786\gamma_1 + 0'014204\gamma_2 + 0'059098 = 0'1,$$

$$0'04096 + 0'0055296\gamma_1 + 0'002015\gamma_2 + 0'004614 = 0'1,$$

que tienen por solución

$$\gamma_1 = 0'950956 \quad \text{y} \quad \gamma_2 = 0'914029.$$

Hay que rechazar pues si  $n\bar{x} < 2$  ó  $n\bar{x} > 22$  y con alta probabilidad si  $n\bar{x} = 2$  ó  $n\bar{x} = 22$ . Como la muestra obtenida da  $n\bar{x} = 16$ , no puede rechazarse la hipótesis nula, al nivel de significación 0'1.

Nótese la dificultad de cálculo que supone, en este caso, la determinación del nivel crítico. Habría que ir aumentando el valor de  $\alpha$  ( $0'2, 0'3, \dots$ ) hasta que  $c_2$  descienda por debajo del valor observado:  $n\bar{x} = 16$  (sólo para conocer su primera cifra decimal).

### Ejercicio 8.3

La densidad de la muestra es

$$f_\theta(x_1, \dots, x_n) = (2\pi\theta^2)^{-n/2} e^{-ns^2/(2\theta^2)},$$

siendo  $s^2 = 1/n \sum_{i=1}^n x_i^2$  la varianza muestral. Luego el test de máxima potencia para el contraste de  $\theta = \theta_0$  frente a  $\theta = \theta_1$  tiene función crítica

$$\begin{aligned} \varphi &= \begin{cases} 1 & \text{si } (2\pi\theta_1^2)^{-n/2} e^{-ns^2/(2\theta_1^2)} > k(2\pi\theta_0^2)^{-n/2} e^{-ns^2/(2\theta_0^2)} \\ 0 & \text{si } (2\pi\theta_1^2)^{-n/2} e^{-ns^2/(2\theta_1^2)} < k(2\pi\theta_0^2)^{-n/2} e^{-ns^2/(2\theta_0^2)} \end{cases} \\ &= \begin{cases} 1 & \text{si } (\theta_0^{-2} - \theta_1^{-2})s^2 > k' \\ 0 & \text{si } (\theta_0^{-2} - \theta_1^{-2})s^2 < k'. \end{cases} \end{aligned}$$

Si es  $\theta_1 < \theta_0$ ,  $\theta_0^{-2} - \theta_1^{-2}$  es negativo, luego resulta

$$\varphi = \begin{cases} 1 & \text{si } s^2 < c \\ 0 & \text{si } s^2 > c. \end{cases}$$

En cambio, si  $\theta_1 > \theta_0$ , se obtiene

$$\varphi = \begin{cases} 1 & \text{si } s^2 > c \\ 0 & \text{si } s^2 < c. \end{cases}$$

El punto crítico  $c$  ha de determinarse, en ambos casos, para que el tamaño del test sea  $\alpha$ . Como  $ns^2/\theta_0^2$  tiene distribución  $\chi_n^2$  (supuesto que  $\theta = \theta_0$ ):

i) Si  $\theta_1 < \theta_0$  ha de ser  $c = n^{-1}\theta_0^2 \chi_{n;1-\alpha}^2$  a fin de que

$$E_{\theta_0}[\varphi] = P_{\theta_0}\{s^2 < n^{-1}\theta_0^2 \chi_{n;1-\alpha}^2\} = \alpha.$$

ii) Si  $\theta_1 > \theta_0$  ha de ser  $c = n^{-1}\theta_0^2 \chi_{n;\alpha}^2$  a fin de que

$$E_{\theta_0}[\varphi] = P_{\theta_0}\{s^2 > n^{-1}\theta_0^2 \chi_{n;\alpha}^2\} = \alpha.$$

b) Si se desea contrastar  $H_0 : \theta \leq \theta_0$  frente a  $H_1 : \theta > \theta_0$ , cabe observar que el contraste

$$\varphi = \begin{cases} 1 & \text{si } s^2 > n^{-1}\theta_0^2 \chi_{n;\alpha}^2 \\ 0 & \text{si } s^2 < n^{-1}\theta_0^2 \chi_{n;\alpha}^2 \end{cases}$$

no depende de  $\theta_1$  y tiene, por tanto, máxima potencia en cualquier  $\theta_1 > \theta_0$ . Además, si  $\theta \leq \theta_0$ ,

$$E_\theta[\varphi] = P_\theta\{s^2 > n^{-1}\theta_0^2 \chi_{n;\alpha}^2\} \leq P_\theta\{s^2 > n^{-1}\theta^2 \chi_{n;\alpha}^2\} = \alpha,$$

luego tiene nivel de significación  $\alpha$  para  $H_0$ . Por consiguiente es el test uniformemente de máxima potencia y nivel de significación  $\alpha$  para las hipótesis en cuestión.

Esta conclusión se deduce directamente del teorema de Karlin-Rubin, si se observa que

$$f_\theta(x_1, \dots, x_n) = c(\theta)e^{q(\theta)s^2}$$

siendo  $q(\theta) = -n/\theta^2$  creciente en  $\theta$ .

La función de potencia del contraste es

$$\beta(\theta) = E_\theta[\varphi] = P_\theta\{s^2 > n^{-1}\theta_0^2 \chi_{n;\alpha}^2\} = P\{\chi_n^2 > (\theta_0/\theta)^2 \chi_{n;\alpha}^2\}$$

que puede obtenerse de la tabla de la  $\chi^2$  y es, desde luego, función creciente de  $\theta$ .

#### Ejercicio 8.4

La función de verosimilitud de  $n$  observaciones es

$$f_\lambda(x_1, \dots, x_n) = \frac{e^{-n\lambda} \lambda^{n\bar{x}}}{x_1! \cdots x_n!} = e^{-n\lambda} \frac{1}{x_1! \cdots x_n!} e^{n\bar{x} \log \lambda},$$

de manera que tiene razón de verosimilitud monótona en  $n\bar{x}$ . Por tanto los tests uniformemente de máxima potencia para contrastar  $\lambda \geq \lambda_0$  frente a  $\lambda < \lambda_0$  tienen función crítica de la forma

$$\varphi = \begin{cases} 1 & \text{si } n\bar{x} < c \\ \gamma & \text{si } n\bar{x} = c \\ 0 & \text{si } n\bar{x} > c. \end{cases}$$

Puesto que  $n\bar{x}$  tiene distribución de Poisson de parámetro  $n\lambda$ , para que el test sea de tamaño  $\alpha$  ha de ser

$$\alpha = P_{\lambda_0}\{n\bar{X} < c\} + \gamma P_{\lambda_0}\{n\bar{X} = c\} = \sum_{k=0}^{c-1} \frac{e^{-n\lambda_0} (n\lambda_0)^k}{k!} + \gamma \frac{e^{-n\lambda_0} (n\lambda_0)^c}{c!}.$$

En el caso  $\lambda_0 = 50$  y  $n = 12$ ,  $n\bar{X}$  se distribuye con distribución de Poisson de parámetro 600, de forma que  $Z = (n\bar{X} - 600)/\sqrt{600}$  es prácticamente  $N(0, 1)$ ; la condición relativa al tamaño es pues, con mucha aproximación

$$\alpha = P\{Z < (c - 600)/\sqrt{600}\};$$

de donde resulta  $c = 600 + z_{1-\alpha}\sqrt{600}$  y concretamente, para  $\alpha = 0'025$ ,  $c = 600 - 1'96\sqrt{600} = 552$ .

La muestra obtenida da un total  $n\bar{x} = 540$  que permite rechazar  $\lambda \geq 50$  con nivel de significación 0'025. Más exactamente, como

$$P\{Z < (540 - 600)/\sqrt{600}\} = P\{Z < -2'45\} = 0'0071,$$

el nivel crítico asociado a la muestra obtenida es 0'0071 y los datos son fuertemente significativos de que la centralita no es rentable.

b) La centralita se cierra o no, según el resultado de un test de función crítica

$$\varphi = \begin{cases} 1 & \text{si } n\bar{x} < c \\ \gamma & \text{si } n\bar{x} = c \\ 0 & \text{si } n\bar{x} > c. \end{cases}$$

Las condiciones impuestas ahora son

$$\begin{aligned} \mathcal{P}_\lambda\{n\bar{X} < c\} &\leq 0'001 && \text{si } \lambda \geq 55, \\ \mathcal{P}_\lambda\{n\bar{X} > c\} &\leq 0'005 && \text{si } \lambda \leq 45, \end{aligned}$$

relativas ambas a la potencia  $\beta(\lambda)$  para diversos valores de  $\lambda$ :  $\beta(\lambda) \leq 0'001$  si  $\lambda \geq 55$  y  $\beta(\lambda) \geq 0'995$  para  $\lambda \leq 45$ . Recuérdese que  $\beta(\lambda)$  es función decreciente de  $\lambda$ .

Puesto que  $n\lambda$  tendrá un valor elevado, sigue siendo válido que el cociente  $Z = (n\bar{X} - n\lambda)/\sqrt{n\lambda}$  tendrá muy aproximadamente distribución  $N(0, 1)$ . Las condiciones anteriores se expresan entonces

$$\begin{aligned} P\{Z < (c - n\lambda)/\sqrt{n\lambda}\} &\leq 0'001 && \text{para } \lambda = 55, \\ P\{Z < (c - n\lambda)/\sqrt{n\lambda}\} &\geq 0'995 && \text{para } \lambda = 45. \end{aligned}$$

Consultando la tabla de la  $N(0, 1)$  se ve que debe ser

$$\frac{c - 55n}{\sqrt{55n}} \leq -3'08 \quad \text{y} \quad \frac{c - 45n}{\sqrt{45n}} \geq 2'58;$$

es decir que  $c$  ha de tomarse en el intervalo  $(45n + 2'58\sqrt{45n}, 55n - 3'08\sqrt{55n})$ . Para que ello sea posible ha de ser

$$45n + 2'58\sqrt{45n} < 55n - 3'08\sqrt{55n}$$

lo cual se produce para  $n \geq 17$ . Hay que observar las llamadas al menos durante 17 días.

### Ejercicio 8.5

a) Las observaciones no constituyen una muestra aleatoria simple de la distribución exponencial puesto que no son independientes, sino cada una superior a la anterior. Concretamente, si  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$  son los instantes aleatorios de fallo de las  $n$  válvulas, ordenados de menor a mayor, la densidad de  $(X_{(1)}, \dots, X_{(r)})$  es

$$\begin{aligned} f_\theta(x_{(1)}, \dots, x_{(r)}) &= \\ &= n\theta e^{-\theta x_{(1)}}(n-1)\theta e^{-\theta x_{(2)}} \cdots (n-r+1)\theta e^{-\theta x_{(r)}}e^{-\theta x_{(r)}(n-r)} \\ &= \frac{n!}{(n-r)!}\theta^r e^{-\theta[x_{(1)}+x_{(2)}+\cdots+(n-r+1)x_{(r)}]} \end{aligned}$$

para  $x_{(1)} < x_{(2)} < \cdots < x_{(r)}$  (que es la probabilidad de que uno de los  $n$  fallos se produzca en el instante  $x_{(1)}$ , uno de los  $n-1$  restantes en el instante posterior  $x_{(2)}, \dots$ , uno de los  $n-r+1$  restantes en el instante  $x_{(r)}$  y los  $n-r$  restantes después de  $x_{(r)}$ ).

Sea  $T$  el estadístico

$$T = \sum_{i=1}^{r-1} X_{(i)} + (n-r+1)X_{(r)}.$$

Los contrastes de máxima potencia de  $\theta = \theta_0$  frente a  $\theta = \theta_1$  tienen región crítica de la forma

$$C = \left\{ \frac{n!}{(n-r)!}\theta_1^r e^{-\theta_1 T} > k \frac{n!}{(n-r)!}\theta_0^r e^{-\theta_0 T} \right\} = \begin{cases} \{T < c\} & \text{si } \theta_1 > \theta_0, \\ \{T > c\} & \text{si } \theta_1 < \theta_0. \end{cases}$$

En ambos casos el valor del punto crítico  $c$  ha de determinarse mediante la condición  $P_{\theta_0}(C) = \alpha$ , lo cual exige conocer la distribución de  $T$ . Para ello, obsérvese que  $Y_1 = X_{(1)}, Y_2 = X_{(2)} - X_{(1)}, \dots, Y_r = X_{(r)} - X_{(r-1)}$ , tienen densidad conjunta

$$\frac{n!}{(n-r)!}\theta^r e^{-\theta y_1}e^{-\theta(y_1+y_2)} \cdots e^{-\theta(y_1+\cdots+y_{r-1})}e^{-\theta(y_1+\cdots+y_r)(n-r+1)}$$

$$= n\theta e^{-n\theta y_1} (n-1)\theta e^{-(n-1)\theta y_2} \cdots (n-r+1)\theta e^{-(n-r+1)\theta y_r}$$

para  $y_1, y_2, \dots, y_r > 0$ ; luego son independientes y con distribuciones exponenciales de parámetros  $n\theta, (n-1)\theta, \dots, (n-r+1)\theta$  respectivamente. Ahora bien, es

$$T = nY_1 + (n-1)Y_2 + \cdots + (n-r+1)Y_r,$$

donde los sumandos son independientes y con distribución exponencial de parámetro  $\theta$ . Por tanto  $T$  tiene distribución  $\gamma(r, \theta)$  o bien  $2\theta T$  distribución  $\chi_{2r}^2$ . Con lo cual resulta

$$c = \begin{cases} \frac{\chi_{2r;1-\alpha}^2}{2\theta_0} & \text{si } \theta_1 > \theta_0, \\ \frac{\chi_{2r;\alpha}^2}{2\theta_0} & \text{si } \theta_1 < \theta_0. \end{cases}$$

b) Como la densidad de la muestra es  $n!/(n-r)! \theta^r e^{-\theta T}$  y  $T$  tiene distribución continua, el teorema de Lehmann afirma que los tests insesgados, uniformemente de máxima potencia, para contrastar  $\theta = \theta_0$  frente a  $\theta \neq \theta_0$  son de la forma:

$$\varphi = \begin{cases} 1 & \text{si } T < c_1 \text{ ó } T > c_2 \\ 0 & \text{si } c_1 < T < c_2, \end{cases}$$

donde, puesto que  $E_\theta[T] = r/\theta$ ,  $c_1$  y  $c_2$  han de determinarse con las condiciones:

$$\begin{cases} E_{\theta_0}[\varphi] = P_{\theta_0}\{T < c_1\} + P_{\theta_0}\{T > c_2\} = \alpha, \\ E_{\theta_0}[\varphi T] = E_{\theta_0}[TI_{\{T < c_1\}}] + E_{\theta_0}[TI_{\{T > c_2\}}] = \alpha r/\theta. \end{cases}$$

En términos de la densidad  $\gamma(r, \theta)$ , estas condiciones se expresan

$$\int_{(0,c_1) \cup (c_2, \infty)} \frac{\theta^r t^{r-1} e^{-\theta t}}{(r-1)!} dt = \alpha = \int_{(0,c_1) \cup (c_2, \infty)} \frac{\theta^{r+1} t^r e^{-\theta t}}{r!} dt$$

y, mediante el cambio  $y = 2\theta t$  (que reduce la  $\gamma(r, \theta)$  a una  $\chi_{2r}^2$ ), se traducen en

$$P\{\chi_{2r}^2 < y_1\} + P\{\chi_{2r}^2 > y_2\} = \alpha = P\{\chi_{2r+2}^2 < y_1\} + P\{\chi_{2r+2}^2 > y_2\},$$

siendo  $y_1 = 2\theta_0 c_1$  e  $y_2 = 2\theta_0 c_2$ .

Con  $r = 9$  y  $\alpha = 0'05$ , la tabla de la  $\chi^2_{18}$ , interpolada linealmente en los intervalos  $(8'231, 9'39)$  y  $(31'53, 34'81)$ , y la tabla de la  $\chi^2_{20}$ , interpolada linealmente en los intervalos  $(8'26, 9'951)$  y  $(31'41, 34'17)$ , permiten aproximar las ecuaciones anteriores por:

$$0'025 + 0'025 \frac{y_1 - 8'231}{9'39 - 8'231} + 0'025 - 0'015 \frac{y_2 - 31'53}{34'81 - 31'53} = 0'05,$$

$$0'01 + 0'015 \frac{y_1 - 8'26}{9'591 - 8'26} + 0'05 - 0'025 \frac{y_2 - 31'41}{34'17 - 31'41} = 0'05;$$

cuya solución es  $y_1 = 8'504$  e  $y_2 = 32'817$ . Como se ve, hay poca diferencia con el test de colas  $\alpha/2$  que, corresponde a tomar  $y_1 = \chi^2_{2r;1-\alpha/2}$  ( $= 8'231$ )  $y_2 = \chi^2_{2r;\alpha/2}$  ( $= 31'53$ ). Finalmente, en cualquier caso, hay que tomar

$$c_1 = y_1/(2\theta_0) \quad \text{y} \quad c_2 = y_2/(2\theta_0).$$

### Ejercicio 8.6

Las densidades de la muestra bajo la hipótesis nula y alternativa son respectivamente

$$f_0(x_1, \dots, x_n) = 1 \quad \text{para } x_1, \dots, x_n \in (0, 1),$$

$$f_1(x_1, \dots, x_n) = 2^n x_1 x_2 \cdots x_n \quad \text{para } x_1, \dots, x_n \in (0, 1).$$

El test de máxima potencia tiene entonces región crítica:

$$C = \left\{ \prod_{i=1}^n x_i > k \right\} = \left\{ -2 \sum_{i=1}^n \log x_i < c \right\}$$

donde  $k$  ó  $c$  han de determinarse mediante la condición  $P_0(C) = 0'05$ .

a) Si  $n = 1$ , ello significa

$$0'05 = P_0\{X_1 > k\} = 1 - k;$$

luego  $k = 0'95$ . La potencia es entonces

$$P_1\{X_1 > 0'95\} = \int_{0'95}^1 2x \, dx = 0'0975.$$

b) Si  $n = 2$ , tiene que ser

$$0'05 = P_0\{X_1 X_2 > k\} = \int_k^1 \int_{k/x_1}^1 dx_2 dx_1 = 1 - k + k \log k;$$

luego  $k \simeq 0'701$ . La potencia es entonces

$$\mathcal{P}_1\{X_1 X_2 > 0'701\} = \int_{0'701}^1 2x_1 \int_{0'701/x_1}^1 2x_2 dx_2 dx_1 = 0'159.$$

c) Si  $n = 3$ , para que

$$\begin{aligned} 0'05 &= \mathcal{P}_0\{X_1 X_2 X_3 > k\} = \int_k^1 \int_{k/x_1}^1 \int_{k/(x_1 x_2)}^1 dx_3 dx_2 dx_1 \\ &= 1 - k + k \log k - k/2 \log^2 k, \end{aligned}$$

debe ser  $k = 0'441$ . La potencia es entonces

$$\mathcal{P}_1\{X_1 X_2 X_3 > 0'441\} = \iiint_{x_1 x_2 x_3 > 0'441} 8x_1 x_2 x_3 dx_1 dx_2 dx_3 = 0'226.$$

Como se ve la potencia aumenta al crecer el tamaño de la muestra.

d) Para  $n$  grande el método anterior no es nada práctico. Ahora bien, bajo la hipótesis nula cada  $X_i$  es uniforme en  $(0, 1)$ ; por tanto,  $-2 \log X_i$  es exponencial de parámetro  $1/2$  y  $-2 \sum_{i=1}^n \log X_i$  tiene distribución  $\chi_{2n}^2$ . Ha de ser pues

$$0'05 = \mathcal{P}_0 \left\{ -2 \sum_{i=1}^n \log X_i < c \right\} = P \{ \chi_{2n}^2 < c \},$$

con lo cual  $c = \chi_{2n;0'95}^2$  (o bien  $k = \exp\{-\chi_{2n;0'95}^2/2\}$ ).

Bajo la hipótesis alternativa,  $-2 \log X_i$  tiene distribución exponencial de parámetro 1; de manera que  $-2 \sum_{i=1}^n \log X_i$  tiene distribución  $\gamma(n, 1)$  o bien  $-4 \sum_{i=1}^n \log X_i$  tiene distribución  $\chi_{2n}^2$ . Con ello la potencia resulta

$$\mathcal{P}_1 \left\{ -2 \sum_{i=1}^n \log X_i < \chi_{2n;0'95}^2 \right\} = P \{ \chi_{2n}^2 < 2\chi_{2n;0'95}^2 \}.$$

Puede observarse en la tabla de la  $\chi^2$  que la potencia sigue aumentando con  $n$ ; por ejemplo:

$$\begin{aligned} \chi_{12;0'95}^2 &= 5'226 & \text{y} & \quad P\{\chi_{12}^2 < 10'452\} \simeq 0'414, \\ \chi_{24;0'95}^2 &= 13'85 & \text{y} & \quad P\{\chi_{24}^2 < 27'7\} \simeq 0'719. \end{aligned}$$

Es curioso que, siendo intuitivamente tan claro que la potencia debe crecer al aumentar el tamaño muestral, no pueda establecerse analíticamente de manera sencilla.

Si  $n$  es suficientemente grande para emplear la aproximación normal, como  $\sqrt{2\chi_{2n}^2} - \sqrt{4n-1} \simeq N(0, 1)$ , será

$$\chi_{2n;0'95}^2 \simeq \frac{1}{2}(\sqrt{4n-1} - 1'64)^2$$

y la potencia se puede expresar aproximadamente en términos de una variable  $Z$  con distribución  $N(0, 1)$ :

$$\begin{aligned} P\left\{\chi_{2n}^2 < (\sqrt{4n-1} - 1'64)^2\right\} &= P\left\{\sqrt{2\chi_{2n}^2} < \sqrt{2}\sqrt{4n-1} - 1'64\right\} \\ &= P\left\{Z < (\sqrt{2}-1)\sqrt{4n-1} - 1'64\right\}, \end{aligned}$$

que crece rápidamente hacia 1 al aumentar  $n$ .

### Ejercicio 8.7

a) La densidad de la muestra es

$$f_\theta(x_1, \dots, x_n) = \frac{1}{\theta^n} I_{\{x_{(n)} < \theta\}}$$

que depende sólo de la mayor observación  $x_{(n)}$ ; como  $\theta$  es desconocido puede obtenerse, en principio, cualquier valor  $x_{(n)} > 0$ . Para contrastar  $\theta = \theta_0$  frente a  $\theta = \theta_1$  ( $\theta_1 < \theta_0$ ), puesto que la distribución de la muestra es continua, debe bastar con un test no aleatorizado que, según el lema de Neyman-Pearson, será de máxima potencia, si su región crítica  $C^*$  está comprendida entre

$$\{f_{\theta_1} > k f_{\theta_0}\} = \begin{cases} \{x_{(n)} < \theta_1\} & \text{si } k < \theta_0^n / \theta_1^n \\ \emptyset & \text{si } k \geq \theta_0^n / \theta_1^n, \end{cases}$$

y

$$\{f_{\theta_1} \geq k f_{\theta_0}\} = \begin{cases} \{x_{(n)} > 0\} & \text{si } k = 0 \\ \{x_{(n)} < \theta_1\} \cup \{x_{(n)} \geq \theta_0\} & \text{si } 0 < k \leq \theta_0^n / \theta_1^n \\ \emptyset & \text{si } k > \theta_0^n / \theta_1^n. \end{cases}$$

Tomar  $k > \theta_0^n / \theta_1^n$ , conduce al test  $C^* = \emptyset$ , que no rechaza nunca y tiene, por tanto, tamaño cero. Tomar  $0 < k < \theta_0^n / \theta_1^n$ , obliga a que sea

$$\{x_{(n)} < \theta_1\} \subset C^* \subset \{x_{(n)} < \theta_1\} \cup \{x_{(n)} > \theta_0\},$$

cuyo tamaño no es  $\alpha$ , sino  $\mathcal{P}_{\theta_0}(C^*) = \theta_1^n / \theta_0^n$ . Hay que tomar pues:

(i) o bien  $k = 0$  y escoger  $C^*$  que verifique

$$\{x_{(n)} < \theta_1\} \subset C^* \subset \{x_{(n)} > 0\},$$

(ii) o bien  $k = \theta_0^n/\theta_1^n$  y escoger  $C^*$  que verifique

$$\emptyset \subset C^* \subset \{x_{(n)} < \theta_1\} \cup \{x_{(n)} \geq \theta_0\}.$$

El lema de Neyman-Pearson asegura que cualquier  $C^*$  que cumpla alguna de las condiciones anteriores tiene máxima potencia (en  $\theta_1$ ), entre los tests con su mismo nivel de significación. La elección entre uno y otro caso depende del tamaño que se quiera alcanzar; los tests del primer tipo tienen siempre tamaño

$$\mathcal{P}_{\theta_0}(C^*) \geq \mathcal{P}_{\theta_0}\{x_{(n)} < \theta_1\} = \theta_1^n/\theta_0^n,$$

mientras que el tamaño de los segundos es siempre inferior a  $\theta_1^n/\theta_0^n$ .

En cualquiera de los dos casos hay múltiples soluciones con tamaño  $\alpha$ . Por ejemplo, si  $\alpha < \theta_1^n/\theta_0^n$ :

$$C_1^* = \left\{ x_{(n)} < \alpha^{1/n} \theta_0 \right\} \quad \text{y} \quad C_2^* = \left\{ (\theta_1^n - \alpha \theta_0^n)^{1/n} < x_{(n)} < \theta_1 \right\}$$

cumplen  $\mathcal{P}_{\theta_0}(C_i^*) = \alpha$  y  $\mathcal{P}_{\theta_1}(C_i^*) = \alpha \theta_0^n/\theta_1^n$ ; lo cual confirma que la potencia es la misma para ambos.

En cambio, si  $\alpha \geq \theta_1^n/\theta_0^n$ ,

$$C_1^* = \left\{ x_{(n)} < \alpha^{1/n} \theta_0 \right\}$$

y

$$C_3^* = \left\{ x_{(n)} < \theta_1 \right\} \cup \left\{ ((1 - \alpha) \theta_0^n + \theta_1^n)^{1/n} < x_{(n)} < \theta_0 \right\}$$

tienen ambos tamaño  $\alpha$  y potencia 1.

$C_1^*$  sirve para cualquier  $\alpha$ , pero no es mejor (ni peor) que las otras soluciones, cuando el valor de  $\alpha$  permite usarlas.

b) Otra ventaja de la solución  $C_1^*$  es que no depende de  $\theta_1$ . Luego es uniformemente de máxima potencia para contrastar  $H_0 : \theta = \theta_0$  frente a  $H_1 : \theta < \theta_0$ . Además, para cualquier  $\theta \geq \theta_0$  es

$$\mathcal{P}_\theta\{x_{(n)} < \alpha^{1/n} \theta_0\} = \alpha \frac{\theta_0^n}{\theta^n} \leq \alpha;$$

luego  $C_1^*$  es la región crítica del test de nivel de significación  $\alpha$  y uniformemente de máxima potencia para contrastar  $\theta \geq \theta_0$  frente a  $\theta < \theta_0$ .

Se puede llegar a la misma conclusión observando que la distribución de la muestra tiene razón de verosimilitud monótona en  $x_{(n)}$ . En efecto, si  $\theta < \theta'$ ,

$$\frac{f_{\theta'}}{f_\theta} = \frac{\theta^n}{\theta'^n} \frac{I_{\{x_{(n)} < \theta'\}}}{I_{\{x_{(n)} < \theta\}}}$$

toma el valor constante  $\theta^n/\theta'^n$  para  $x_{(n)} < \theta$  y vale  $+\infty$  para  $x_{(n)} \in [\theta, \theta']$ <sup>7</sup>; luego es función creciente de  $x_{(n)}$ . El teorema de Karlin-Rubin —o su simétrico, mejor dicho— garantiza entonces que los tests uniformemente de máxima potencia, tienen región crítica de la forma:  $C^* = \{x_{(n)} < c\}$  donde ha de ser  $c = \alpha^{1/n}\theta_0$  para que el contraste tenga tamaño  $\alpha$ .

c) Examinemos ahora el contraste de  $\theta = \theta_0$  frente a  $\theta = \theta_1$ , con  $\theta_1 > \theta_0$ . Como en el apartado (a),

$$\{f_{\theta_1} > k f_{\theta_0}\} = \begin{cases} \{x_{(n)} < \theta_1\} & \text{si } k < \theta_0^n/\theta_1^n \\ \{\theta_0 \leq x_{(n)} < \theta_1\} & \text{si } k \geq \theta_0^n/\theta_1^n, \end{cases}$$

y

$$\{f_{\theta_1} \geq k f_{\theta_0}\} = \begin{cases} \{x_{(n)} > 0\} & \text{si } k \leq \theta_0^n/\theta_1^n \\ \{\theta_0 \leq x_{(n)} < \theta_1\} & \text{si } k > \theta_0^n/\theta_1^n. \end{cases}$$

Si  $k < \theta_0^n/\theta_1^n$  se obtiene un test de tamaño 1. Y si  $k > \theta_0^n/\theta_1^n$ , un test de tamaño cero. Luego hay que tomar  $k = \theta_0^n/\theta_1^n$ , para que las condiciones de optimalidad del lema de Neyman-Pearson sean

$$(iii) \quad \{\theta_0 \leq x_{(n)} < \theta_1\} \subset C^* \subset \{x_{(n)} > 0\}.$$

El test de región crítica

$$C^{**} = \{x_{(n)} < \alpha^{1/n}\theta_0\} \cup \{x_{(n)} \geq \theta_0\}$$

tiene tamaño:  $\mathcal{P}_{\theta_0}(C^{**}) = \alpha$  y es de máxima potencia para cualquier  $\theta_1 \neq \theta_0$ , puesto que cumple las condiciones (i), (ii) y (iii) según que sea  $\theta_1 < \alpha^{1/n}\theta_0$ ,  $\theta_1 \in [\alpha^{1/n}\theta_0, \theta_0)$  ó  $\theta_1 > \theta_0$ . Así pues  $C^{**}$  es un test uniformemente de máxima potencia para contrastar  $\theta = \theta_0$  frente a  $\theta \neq \theta_0$ .

---

<sup>7</sup> Las “muestras” con  $x_{(n)} \geq \theta'$  no forman parte del espacio muestral ni para el valor  $\theta$ , ni para el valor  $\theta'$ .

A pesar de que no pertenezca a la familia exponencial, la distribución uniforme es un caso en el cual existe el test uniformemente de máxima potencia para contrastes bilaterales, sin necesidad de restringirse a los tests insegados.

### Ejercicio 8.8

La situación es muy similar a la del ejercicio anterior y puede de hecho reducirse a ella, mediante el cambio de variable  $Y = e^{-X}$ . Sin embargo, como hemos utilizado el ejercicio anterior para mostrar una utilización delicada del lema de Neyman-Pearson no aleatorizado, procederemos aquí de la misma manera con el lema para tests aleatorizados.

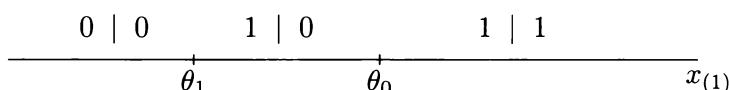
La densidad de la muestra se puede expresar

$$f_\theta(x_1, \dots, x_n) = e^{n(\theta - \bar{x})} I_{\{x_{(1)} \geq \theta\}},$$

de forma que para contrastar  $\theta = \theta_0$  frente a  $\theta = \theta_1$ , son de máxima potencia los tests con función crítica de la forma

$$\begin{aligned} \varphi &= \begin{cases} 1 & \text{si } e^{n(\theta_1 - \bar{x})} I_{\{x_{(1)} \geq \theta_1\}} > k e^{n(\theta_0 - \bar{x})} I_{\{x_{(1)} \geq \theta_0\}} \\ \gamma(x_1, \dots, x_n) & \text{si } e^{n(\theta_1 - \bar{x})} I_{\{x_{(1)} \geq \theta_1\}} = k e^{n(\theta_0 - \bar{x})} I_{\{x_{(1)} \geq \theta_0\}} \\ 0 & \text{si } e^{n(\theta_1 - \bar{x})} I_{\{x_{(1)} \geq \theta_1\}} < k e^{n(\theta_0 - \bar{x})} I_{\{x_{(1)} \geq \theta_0\}} \end{cases} \\ &= \begin{cases} 1 & \text{si } I_{\{x_{(1)} \geq \theta_1\}} > k' I_{\{x_{(1)} \geq \theta_0\}} \\ \gamma(x_1, \dots, x_n) & \text{si } I_{\{x_{(1)} \geq \theta_1\}} = k' I_{\{x_{(1)} \geq \theta_0\}} \\ 0 & \text{si } I_{\{x_{(1)} \geq \theta_1\}} < k' I_{\{x_{(1)} \geq \theta_0\}}. \end{cases} \end{aligned}$$

Si  $\theta_1 < \theta_0$ , a medida que  $x_{(1)}$  crece, la pareja de funciones indicatrices que hay que comparar pasa por los valores



No hay pues más solución que escoger  $k' = 1$  si no se quiere obtener un test de tamaño 0 ó 1. Con  $k' = 1$  resulta

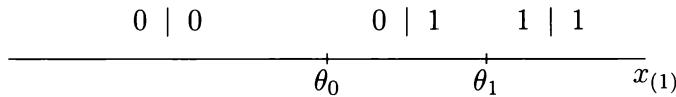
$$\varphi_- = \begin{cases} 1 & \text{si } \theta_1 \leq x_{(1)} < \theta_0 \\ \gamma(x_1, \dots, x_n) & \text{si } x_{(1)} < \theta_1 \text{ ó } x_{(1)} \geq \theta_0. \end{cases}$$

Si el problema consistiese en contrastar  $\theta = \theta_0$  frente a  $\theta = \theta_1$  ( $\theta_1 < \theta_0$ ), entre estos tests óptimos el más simple, de tamaño  $\alpha$ , sería

$$\varphi = \begin{cases} 1 & \text{si } x_{(1)} < c \\ 0 & \text{si } x_{(1)} > c \end{cases}$$

siendo  $c = \theta_0 - n^{-1} \log(1 - \alpha)$  (como se ha obtenido en el ejemplo 8.6).

Análogamente, si  $\theta_1 > \theta_0$ , los términos a comparar valen, según el valor de  $x_{(1)}$ :



Hay ahora dos soluciones óptimas no triviales; la correspondiente a  $k' = 0$ , que es

$$\varphi_+^0 = \begin{cases} 1 & \text{si } x_{(1)} \geq \theta_1 \\ \gamma(x_1, \dots, x_n) & \text{si } x_{(1)} < \theta_1 \end{cases}$$

y la correspondiente a  $k' = 1$ , que es

$$\varphi_+^1 = \begin{cases} \gamma(x_1, \dots, x_n) & \text{si } x_{(1)} < \theta_0 \text{ ó } x_{(1)} \geq \theta_1 \\ 0 & \text{si } \theta_0 \leq x_{(1)} < \theta_1. \end{cases}$$

Si el problema consistiese en contrastar  $\theta = \theta_0$  frente a  $\theta = \theta_1$  ( $\theta_1 > \theta_0$ ) la solución más simple, de tamaño  $\alpha$ , es tomar

$$\varphi = \begin{cases} 1 & \text{si } x_{(1)} > \theta_0 - n^{-1} \log \alpha \\ 0 & \text{si } x_{(1)} < \theta_0 - n^{-1} \log \alpha, \end{cases}$$

que es de la forma  $\varphi_+^0$  en caso de ser  $\theta_0 - n^{-1} \log \alpha < \theta_1$  (es decir para tamaños  $\alpha > e^{n(\theta_0 - \theta_1)}$ ) y de la forma  $\varphi_+^1$  en caso de ser  $\theta_0 - n^{-1} \log \alpha > \theta_1$  (es decir para tamaños  $\alpha < e^{n(\theta_0 - \theta_1)}$ ).

El teorema de Karlin-Rubin ha permitido evitar estas consideraciones en el ejemplo 8.6. Pero ahora, como se trata de contrastar  $\theta = \theta_0$  frente a  $\theta \neq \theta_0$ , no hay más remedio que intentar encontrar un test que cumpla

simultáneamente las restricciones de  $\varphi_-$ ,  $\varphi_+^0$  y  $\varphi_+^1$ , para que sea de máxima potencia para cualquier valor de  $\theta_1 \neq \theta_0$ . No es muy difícil:

$$\varphi^\star = \begin{cases} 1 & \text{si } x_{(1)} < \theta_0 \text{ ó } x_{(1)} \geq c \\ 0 & \text{si } \theta_0 \leq x_{(1)} < c, \end{cases}$$

siendo  $c = \theta_0 - n^{-1} \log \alpha$  para que tenga tamaño  $\alpha$ , cumple:

- ▷ es de la forma  $\varphi_-$  para cualquier  $\theta_1 < \theta_0$ ,
- ▷ es de la forma  $\varphi_+^0$  para cualquier  $\theta_1 \in (\theta_0, \theta_0 - n^{-1} \log \alpha)$ ,
- ▷ es de la forma  $\varphi_+^1$  para cualquier  $\theta_1 > \theta_0 - n^{-1} \log \alpha$ .

Por consiguiente, es uniformemente de máxima potencia.

### Ejercicio 8.9

a) La función de densidad de la muestra es

$$f_\theta(x_1, \dots, x_n) = \frac{2^n}{(x_1 - \theta)^3(x_2 - \theta)^3 \cdots (x_n - \theta)^3} I_{\{x_{(1)} \geq \theta + 1\}}.$$

Sea

$$T = \frac{(x_1 - \theta_0)(x_2 - \theta_0) \cdots (x_n - \theta_0)}{(x_1 - \theta_1)(x_2 - \theta_1) \cdots (x_n - \theta_1)}$$

y obsérvese que  $X_{(1)}$  tiene distribución

$$\mathcal{P}_\theta \{X_{(1)} \leq y\} = 1 - (y - \theta)^{-2n}$$

para valores de  $y \geq 1 + \theta$  (ya que la distribución teórica es  $1 - (x - \theta)^{-2}$  para  $x \geq \theta + 1$ ).

Para diseñar los tests de máxima potencia en el contraste de  $\theta = \theta_0$  frente a  $\theta_1$  (prescindiendo del término constante  $2^n$ ) hay que comparar

$$\frac{I_{\{x_{(1)} \geq \theta_1 + 1\}}}{(x_1 - \theta_1)^3 \cdots (x_n - \theta_1)^3} \quad \text{y} \quad k \frac{I_{\{x_{(1)} \geq \theta_0 + 1\}}}{(x_1 - \theta_0)^3 \cdots (x_n - \theta_0)^3}$$

En el caso  $k = 0$  se obtienen como tests óptimos los de la forma

$$\varphi_0 = \begin{cases} 1 & \text{si } x_{(1)} \geq \theta_1 + 1 \\ \gamma & \text{si } x_{(1)} < \theta_1 + 1, \end{cases}$$

cuyo tamaño:

$$E_{\theta_0}[\varphi_0] = \mathcal{P}_{\theta_0} \{X_{(1)} \geq \theta_1 + 1\} + \gamma \mathcal{P}_{\theta_0} \{X_{(1)} < \theta_1 + 1\}$$

es siempre mayor que  $(\theta_1 + 1 - \theta_0)^{-2n}$ ; mientras que su potencia es

$$E_{\theta_1}[\varphi_0] = P_{\theta_1} \{ X_{(1)} \geq \theta_1 + 1 \} = 1.$$

En el caso  $k > 0$ , la función crítica del test de máxima potencia es de la forma

$$\varphi_1 = \begin{cases} 1 & \text{si } x_{(1)} \geq \theta_1 + 1 \text{ y } T > c \\ \gamma & \text{si } x_{(1)} \geq \theta_1 + 1 \text{ y } T = c \\ 0 & \text{si } \theta_0 + 1 \leq x_{(1)} < \theta_1 + 1 \text{ o bien } x_{(1)} \geq \theta_1 + 1 \text{ y } T < c, \end{cases}$$

cuyo tamaño

$$E_{\theta_0}[\varphi_1] = P_{\theta_0} \{ X_{(1)} \geq \theta_1 + 1, T > c \}$$

es siempre menor que  $P_{\theta_0} \{ X_{(1)} \geq \theta_1 + 1 \} = (\theta_1 + 1 - \theta_0)^{-2n}$ ; aunque su cálculo explícito requeriría averiguar la distribución conjunta de  $X_{(1)}$  y  $T$ , lo cual no es nada sencillo.

El test a emplear depende pues del nivel de significación requerido:

- ▷ si  $\alpha \geq (\theta_1 + 1 - \theta_0)^{-2n}$ , tiene nivel de significación  $\alpha$  un test de la forma  $\varphi_0$ , en el cual se puede tomar  $\gamma = 0$  (haciendo que el tamaño pueda ser inferior a  $\alpha$ ) puesto que ajustar el tamaño no aumenta la potencia, que ya vale 1.
- ▷ si  $\alpha < (\theta_1 + 1 - \theta_0)^{-2n}$ , habría que usar un test de la forma  $\varphi_1$ , para el cual es muy difícil determinar el punto crítico  $c$ .

En otras situaciones de contraste de hipótesis simples se ha indicado que conviene elegir  $\alpha$ , comparándolo con la probabilidad de error de tipo II; este principio aconsejaría en este caso utilizar siempre tests de la forma  $\varphi_1$  (que disminuyen  $\alpha$  a costa de disminuir la potencia). La dificultad de construir efectivamente  $\varphi_1$  hace que ello no sea posible en esta situación.

De hecho, si se puede jugar con el tamaño muestral, nunca se necesitará emplear  $\varphi_1$ ; dado  $\alpha$ , tomando

$$n \geq \frac{-\log \alpha}{2 \log(\theta_1 + 1 - \theta_0)} \quad \text{será} \quad (\theta_1 + 1 - \theta_0)^{-2n} \leq \alpha$$

y el test adecuado es  $\varphi_0$ . Ni siquiera conviene tomar un valor de  $n$  muy superior a esta cota, pues ello no incrementa la potencia.

b) Supongamos que  $\varphi^*$  es la función crítica de un test de nivel de significación  $\alpha$  y uniformemente de máxima potencia para contrastar  $H_0 : \theta = \theta_0$  frente a  $H_1 : \theta > \theta_0$ .

Sea  $\theta^* = \alpha^{-1/(2n)} + \theta_0 - 1$ . Para valores de  $\theta_1 < \theta^*$  es  $\alpha < (\theta_1 + 1 - \theta_0)^{-2n}$  y  $\varphi^*$  tiene que ser de la forma  $\varphi_1$ . Hay que aceptar entonces  $H_0$ , siempre que  $x_{(1)} < \theta_1 + 1$ ; como ello es cierto para cualquier  $\theta_1 < \theta^*$ , resulta

$$\varphi^* = 0 \quad \text{para } x_{(1)} < \theta^* + 1$$

y  $\varphi^*$  no es de la forma  $\varphi_1$  tal y como se requería. No existe pues test uniformemente de máxima potencia, de ningún nivel de significación, para contrastar  $\theta = \theta_0$  frente a  $\theta > \theta_0$ . Menos aún si la hipótesis nula es  $\theta \leq \theta_0$ .

### Ejercicio 8.10

La densidad de la muestra de tamaño  $n$  es

$$f_{\theta, \lambda}(x_1, \dots, x_n) = \frac{1}{\lambda^n} e^{n(\theta - \bar{x})/\lambda} I_{\{x_{(1)} \geq \theta\}}.$$

Como  $\theta_1 < \theta_0$ , si  $\lambda_1 > \lambda_0$ , la desigualdad

$$\frac{1}{\lambda_1^n} e^{n(\theta_1 - \bar{x})/\lambda_1} I_{\{x_{(1)} \geq \theta_1\}} > \frac{1}{\lambda_0^n} e^{n(\theta_0 - \bar{x})/\lambda_0} I_{\{x_{(1)} \geq \theta_0\}}$$

se verifica si

$$\theta_1 \leq x_{(1)} < \theta_0 \quad \text{o bien} \quad x_{(1)} \geq \theta_0 \quad \text{y} \quad n\bar{x} > c,$$

habida cuenta que  $1/\lambda_0 - 1/\lambda_1 > 0$  y siendo  $c$  una constante (que depende de  $k, n, \lambda_0, \lambda_1, \theta_0$  y  $\theta_1$ ). Análogamente, la igualdad entre ambas densidades se produce cuando

$$x_{(1)} < \theta_1 \quad \text{o bien} \quad x_{(1)} \geq \theta_0 \quad \text{y} \quad n\bar{x} = c.$$

Luego los tests de máxima potencia para las hipótesis simples consideradas tienen por función crítica

$$\varphi_a = \begin{cases} 1 & \text{si } \theta_1 \leq x_{(1)} < \theta_0 \quad \text{o bien } x_{(1)} \geq \theta_0 \quad \text{y} \quad n\bar{x} > c_a \\ \gamma & \text{si } x_{(1)} < \theta_1 \quad \text{o bien } x_{(1)} \geq \theta_0 \quad \text{y} \quad n\bar{x} = c_a \\ 0 & \text{si } x_{(1)} \geq \theta_0 \quad \text{y} \quad n\bar{x} < c_a. \end{cases}$$

La opción intermedia tiene probabilidad cero, bajo ambas hipótesis, luego puede incluirse en la primera, dando lugar al test no aleatorizado de región crítica

$$C_a = \{x_{(1)} < \theta_0\} \cup \{x_{(1)} \geq \theta_0, n\bar{x} \geq c_a\}.$$

Si  $\lambda_1 < \lambda_0$ , sólo cambia el sentido de la última desigualdad, de forma que el test de máxima potencia tiene ahora región crítica

$$C_b = \{x_{(1)} < \theta_0\} \cup \{x_{(1)} \geq \theta_0, n\bar{x} \leq c_b\}.$$

Supuesto que  $\theta = \theta_0$  y  $\lambda = \lambda_0$ , cada  $X_i - \theta_0$  tiene distribución exponencial de parámetro  $1/\lambda_0$ ,  $n(\bar{X} - \theta_0)$  tiene distribución  $\gamma(n, 1/\lambda_0)$ , o lo que es lo mismo,  $2n(\bar{X} - \theta_0)/\lambda_0$  tiene distribución  $\chi^2_{2n}$ . Por tanto, para que los tests tengan tamaño  $\alpha$ , debe ser en el primer caso

$$\alpha = \mathcal{P}_{\theta_0, \lambda_0}(C_a) = \mathcal{P}_{\theta_0, \lambda_0}\{n\bar{X} \geq c_a\} = P\left\{\chi^2_{2n} \geq \frac{2}{\lambda_0}(c_a - n\theta_0)\right\},$$

de donde

$$\frac{2}{\lambda_0}(c_a - n\theta_0) = \chi^2_{2n; \alpha} \quad \text{o bien} \quad c_a = n\theta_0 + \frac{\lambda_0}{2} \chi^2_{2n; \alpha}.$$

En cambio, en el segundo caso, ha de ser

$$\alpha = \mathcal{P}_{\theta_0, \lambda_0}(C_b) = \mathcal{P}_{\theta_0, \lambda_0}\{n\bar{X} \leq c_b\} = P\left\{\chi^2_{2n} \leq \frac{2}{\lambda_0}(c_b - n\theta_0)\right\},$$

es decir

$$\frac{2}{\lambda_0}(c_b - n\theta_0) = \chi^2_{2n; 1-\alpha} \quad \text{o bien} \quad c_b = n\theta_0 + \frac{\lambda_0}{2} \chi^2_{2n; 1-\alpha}.$$

Los resultados anteriores permiten concluir:

- i) El test de región crítica  $C_a$  es de máxima potencia para el contraste de  $\theta = \theta_0, \lambda = \lambda_0$  frente a cualquier  $\theta_1 < \theta_0, \lambda_1 > \lambda_0$ . Es pues uniformemente de máxima potencia en el contraste propuesto en (i).
- ii) El test de región crítica  $C_b$  es de máxima potencia para el contraste de  $\theta = \theta_0, \lambda = \lambda_0$  frente a cualquier  $\theta_1 < \theta_0, \lambda_1 < \lambda_0$ . Es pues uniformemente de máxima potencia en el contraste propuesto en (ii).
- iii) Un test uniformemente de máxima potencia para el contraste propuesto en (iii) tiene que ser de la forma  $C_a$  (para tener máxima potencia en  $\theta_1, \lambda_1 > \lambda_0$ ) y de la forma  $C_b$  (para tener máxima potencia en  $\theta_1, \lambda_1 < \lambda_0$ ). Como ambas cosas son incompatibles no existe test uniformemente de máxima potencia.
- iv) Para que un test, sea cual sea su tamaño, tenga máxima potencia en el contraste de  $(\theta, \lambda)$  ( $\theta \geq \theta_0, \lambda \leq \lambda_0$ ) frente a  $(\theta_1, \lambda_1)$ , su región crítica ha de ser de la forma

$$C = \{x_{(1)} < \theta\} \cup \{x_{(1)} \geq \theta, n\bar{x} \geq c\}.$$

Un test con tal región crítica tendría nivel de significación  $\alpha$  si, para cualquier  $\theta \geq \theta_0, \lambda \leq \lambda_0$ , fuese

$$\alpha \geq \mathcal{P}_{\theta, \lambda}(C) = \mathcal{P}_{\theta, \lambda}\{n\bar{X} \geq c\} = P\{Z > c - n\theta\},$$

siendo  $Z$  una variable con distribución  $\gamma(n, 1/\lambda)$ . Está claro que ello no puede cumplirse para cualquier  $\theta > \theta_0$ . No existe, por tanto, ningún test de la forma indicada que tenga nivel de significación  $\alpha$  para el contraste indicado en (iv).

## Capítulo 9

# Métodos de contraste

### 9.1. Introducción

El capítulo anterior se ha centrado, inmediatamente después de las consideraciones generales sobre contraste de hipótesis, en la determinación de tests óptimos —de acuerdo con el criterio de la máxima potencia. Ello sólo es posible en los casos de hipótesis simple frente a simple; en los contrastes unilaterales, bajo ciertas condiciones para la distribución de la muestra; en los contrastes bilaterales para distribuciones exponenciales, restringiéndose a considerar tests insesgados; y poco más. La situación general es que no existe test uniformemente de máxima potencia para contrastar una hipótesis arbitraria frente a otra.

Es importante entonces, para muchos problemas prácticos, disponer de algunos métodos de contraste, basados en principios razonables, que permitan diseñar tests adecuados en cualquier circunstancia. El más importante de ellos lo constituyen los *tests de razón de verosimilitudes*, que coinciden con el resultado del lema de Neyman-Pearson para los contrastes de hipótesis simple frente a simple, pero guardan también una estrecha relación con el principio de máxima verosimilitud.

En el mismo sentido es útil, igualmente, analizar la relación entre las regiones de aceptación de los tests de hipótesis y los intervalos de confianza para el parámetro sobre el cual se contrasta. El paralelismo existente entre ambos conceptos puede servir de ayuda a la hora de emplear cualquiera de los dos métodos de inferencia; pues, por una parte, permite la construcción rápida de tests de hipótesis basados en el correspondiente intervalo de confianza y, por otro lado, provee de nuevos criterios para la selección de intervalos de confianza.

## 9.2. Tests de razón de verosimilitudes

El método de la razón de verosimilitudes es aplicable en cualquier situación en que la distribución de la muestra,  $\mathcal{P}_\theta$ , dependa de un parámetro, que varía en un espacio paramétrico  $\Theta \subset \mathbb{R}^k$ , y se desee contrastar una hipótesis nula  $H_0 : \theta \in \Theta_0$  frente a la alternativa  $H_1 : \theta \in \Theta_1 = \Theta - \Theta_0$ , donde  $\Theta_0$  es cualquier subconjunto de  $\Theta$ .

La idea del método se basa en que, para una muestra fija, su verosimilitud  $f_\theta(x_1, \dots, x_n)$  es una medida de lo bien que “explica” el valor  $\theta$  del parámetro los resultados obtenidos. Por consiguiente,  $\sup_{\theta \in \Theta_0} f_\theta(x_1, \dots, x_n)$  supone un índice acerca de la “mejor explicación” de la muestra que puede obtenerse bajo  $H_0$ ; mientras que  $\sup_{\theta \in \Theta} f_\theta(x_1, \dots, x_n)$  proporciona tal índice entre todos los valores posibles del parámetro. Ambas cantidades son fáciles de calcular pues coinciden respectivamente con  $f_{\hat{\theta}_0}(x_1, \dots, x_n)$  y  $f_{\hat{\theta}}(x_1, \dots, x_n)$ , donde  $\hat{\theta}_0$  y  $\hat{\theta}$  representan los estimadores de máxima verosimilitud de  $\theta$ , según que se considere  $\Theta_0$  ó  $\Theta$  como espacio paramétrico.

Que  $f_{\hat{\theta}_0}(x_1, \dots, x_n)$  sea mucho más bajo que  $f_{\hat{\theta}}(x_1, \dots, x_n)$  significa que la explicación de los resultados, sobre la base de que  $H_0$  es cierta, es mucho peor que la explicación sin tal restricción; parece adecuado entonces rechazar que la distribución de la población verifique  $H_0$ . En caso contrario —es decir si  $f_{\hat{\theta}_0}$  es próximo a  $f_{\hat{\theta}}$  para la muestra obtenida— no hay razones suficientes para descartar que  $H_0$  sea cierta, puesto que ello no mejoraría sensiblemente la verosimilitud de las observaciones.

La comparación de ambas cantidades se realiza formando el cociente entre ellas, para obtener un número entre 0 y 1. Concretamente:

*Se denomina razón de verosimilitudes para contrastar la hipótesis  $H_0 : \theta \in \Theta_0$  frente a  $H_1 : \theta \in \Theta_1$  al cociente*

$$\Lambda(x_1, \dots, x_n) = \frac{\sup_{\theta \in \Theta_0} f_\theta(x_1, \dots, x_n)}{\sup_{\theta \in \Theta_0 \cup \Theta_1} f_\theta(x_1, \dots, x_n)} = \frac{f_{\hat{\theta}_0}(x_1, \dots, x_n)}{f_{\hat{\theta}}(x_1, \dots, x_n)}.$$

*Un test de razón de verosimilitudes para dichas hipótesis es aquél que tiene región crítica de la forma*

$$C = \{\Lambda(x_1, \dots, x_n) < c\},$$

*es decir que rechaza  $H_0$  cuando el cociente  $\Lambda$  toma valores pequeños: inferiores a un cierto valor crítico  $c$ .*

Más en general, el test puede ser aleatorizado con función crítica de la forma

$$\varphi = \begin{cases} 1 & \text{si } \Lambda(x_1, \dots, x_n) < c \\ \gamma & \text{si } \Lambda(x_1, \dots, x_n) = c \\ 0 & \text{si } \Lambda(x_1, \dots, x_n) > c. \end{cases}$$

En ambos casos,  $c \in (0, 1)$  una constante que ha de determinarse de acuerdo con el nivel de significación  $\alpha$  deseado; esto es, para conseguir que sea

$$\sup_{\theta \in \Theta_0} E_\theta[\varphi] = \alpha.$$

En el caso de hipótesis simple ( $\theta = \theta_0$ ) frente a simple ( $\theta = \theta_1$ ), la razón de verosimilitudes se reduce a

$$\Lambda(x_1, \dots, x_n) = \frac{f_{\theta_0}(x_1, \dots, x_n)}{\max[f_{\theta_0}(x_1, \dots, x_n), f_{\theta_1}(x_1, \dots, x_n)]};$$

de forma que la comparación  $\Lambda < c$  (con  $c < 1$ ) es equivalente a la comparación  $f_{\theta_1}/f_{\theta_0} > k$  ( $k = 1/c$ ) necesaria para aplicar el lema de Neyman-Pearson. Por tanto, en este caso, ambos procedimientos son idénticos.

Bastará emplear tests de razón de verosimilitudes no aleatorizados, para cualquier nivel de significación  $\alpha$ , cuando la distribución del estadístico  $\Lambda$  sea continua. En caso contrario, excepto para ciertos valores de  $\alpha$ , habrá que aleatorizar, determinando el valor de  $\gamma$  para que se verifique la restricción sobre el tamaño del test.

Con frecuencia, la desigualdad  $\Lambda < c$  puede expresarse en términos de un estadístico más simple que  $\Lambda$ , cuya distribución es conocida. Si no es así, la principal dificultad en la construcción del test de razón de verosimilitudes consiste en obtener la distribución de  $\Lambda$ , necesaria para determinar el punto crítico  $c$ . A menudo resulta imposible hacerlo con exactitud y hay que emplear la distribución asintótica de  $\Lambda$ , que se expondrá al final de esta sección. Como dicha distribución es continua (una  $\chi^2$  de hecho), la mayor parte de los tests de razón de verosimilitudes que aparecen en las aplicaciones son no aleatorizados.

Conviene antes de nada ver cómo se utilizan los conceptos anteriores.

### Ejemplo 9.1

Una empresa afirma que en su producción hay a lo sumo una proporción  $\theta_0$  de productos defectuosos. Un cliente suspicaz elige  $n$  de ellos al azar para

ver si la afirmación es correcta; es decir, si  $\theta$  representa la proporción real de defectuosos, se trata de contrastar  $H_0 : \theta \leq \theta_0$  frente a  $\theta > \theta_0$ .

Si  $x_i = 1$  ó  $0$ , según que el elemento muestral  $i$ -ésimo sea defectuoso o no, y  $T = \sum_{i=1}^n x_i$ , la función de verosimilitud es

$$f_\theta(x_1, \dots, x_n) = \theta^T (1 - \theta)^{n-T};$$

de forma que el estimador de máxima verosimilitud de  $\theta$  es  $\hat{\theta} = T/n$  y se tiene

$$\sup_{\theta \in [0,1]} f_\theta(x_1, \dots, x_n) = f_{\hat{\theta}}(x_1, \dots, x_n) = \left(\frac{T}{n}\right)^T \left(1 - \frac{T}{n}\right)^{n-T}$$

Por otra parte, como  $f_\theta$  (o su logaritmo) es una función creciente para  $\theta \in (0, T/n)$  y decreciente para  $\theta \in (T/n, 1)$ , el máximo para  $\theta \leq \theta_0$  se alcanza en

$$\hat{\theta}_0 = \begin{cases} \theta_0 & \text{si } T/n > \theta_0 \\ T/n & \text{si } T/n \leq \theta_0 \end{cases}$$

y vale

$$\sup_{\theta \leq \theta_0} f_\theta(x_1, \dots, x_n) = \begin{cases} \theta_0^T (1 - \theta_0)^{n-T} & \text{si } T/n > \theta_0, \\ \left(\frac{T}{n}\right)^T \left(1 - \frac{T}{n}\right)^{n-T} & \text{si } T/n \leq \theta_0. \end{cases}$$

Así pues

$$\Lambda(x_1, \dots, x_n) = \begin{cases} \frac{\theta_0^T (1 - \theta_0)^{n-T}}{(T/n)^T (1 - T/n)^{n-T}} & \text{si } T > n\theta_0, \\ 1 & \text{si } T \leq n\theta_0. \end{cases}$$

Tomando logaritmos se comprueba fácilmente que  $\Lambda$  es una función decreciente de  $T$ ; de forma que la desigualdad  $\Lambda < c$  equivale a  $T > k$ . Por tanto el test de razón de verosimilitudes, de tamaño  $\alpha$ , tiene función crítica

$$\varphi = \begin{cases} 1 & \text{si } T > k \\ \gamma & \text{si } T = k \\ 0 & \text{si } T < k, \end{cases}$$

donde  $k$  y  $\gamma$  han de determinarse con la condición

$$\mathcal{P}_{\theta_0}\{T > k\} + \gamma \mathcal{P}_{\theta_0}\{T = k\} = \alpha,$$

habida cuenta de que  $T$  tiene distribución binomial  $B(n, \theta_0)$  (supuesto que  $\theta = \theta_0$ ).

Nótese que la familia de distribuciones de la muestra tiene razón de verosimilitud monótona en  $T$ ; con lo cual el test de razón de verosimilitudes obtenido es, en este caso, uniformemente de máxima potencia.

Para contrastes sobre poblaciones de tipo normal, los tests de razón de verosimilitudes dan también buenos resultados.

### Ejemplo 9.2

Acerca de una población  $N(\mu, \sigma)$ , se quiere comprobar si la media es  $\mu_0$ . Supuesto que  $\sigma$  es desconocida, los conjuntos del espacio paramétrico que describen las hipótesis nula y alternativa son

$$\Theta_0 = \{(\mu, \sigma) \mid \mu = \mu_0, \sigma > 0\} \quad \text{y} \quad \Theta_1 = \Theta - \Theta_0 = \{(\mu, \sigma) \mid \mu \neq \mu_0, \sigma > 0\}.$$

Para una muestra aleatoria simple de tamaño  $n$ , con la función de verosimilitud  $f_{\mu, \sigma}$  y los estimadores de máxima verosimilitud  $\hat{\mu} = \bar{x}$  y  $\hat{\sigma} = s$ , determinados en el ejemplo 7.4, se obtiene

$$\sup_{(\mu, \sigma) \in \Theta} f_{\mu, \sigma}(x_1, \dots, x_n) = (2\pi s^2)^{-n/2} e^{-\sum_{i=1}^n (x_i - \bar{x})^2 / (2s^2)} = (2\pi s^2)^{-n/2} e^{-n/2}.$$

Mientras que, si  $\mu = \mu_0$ , es  $\hat{\sigma}_0^2 = 1/n \sum_{i=1}^n (x_i - \mu_0)^2$  y

$$\sup_{(\mu, \sigma) \in \Theta_0} f_{\mu, \sigma}(x_1, \dots, x_n) = (2\pi \hat{\sigma}_0^2)^{-n/2} e^{-n/2}.$$

Por consiguiente

$$\Lambda(x_1, \dots, x_n) = \left( \frac{s^2}{\hat{\sigma}_0^2} \right)^{n/2} = \left( \frac{s^2}{s^2 + (\bar{x} - \mu_0)^2} \right)^{n/2} = \left( 1 + \frac{(\bar{x} - \mu_0)^2}{s^2} \right)^{-n/2},$$

puesto que  $\hat{\sigma}_0^2 = 1/n \sum_{i=1}^n (x_i - \bar{x})^2 + (\bar{x} - \mu_0)^2$ .

La desigualdad  $\Lambda < c$  define

$$C = \left\{ \sqrt{n-1} \frac{|\bar{x} - \mu_0|}{s} > k \right\}$$

como región crítica del contraste de razón de verosimilitudes. El test es ahora no aleatorizado puesto que  $\sqrt{n-1} |\bar{x} - \mu_0|/s$  tiene distribución  $t$  de Student con  $n-1$  grados de libertad; lo cual indica además que debe ser  $k = t_{n-1; \alpha/2}$  para que el contraste sea de tamaño  $\alpha$ .

De acuerdo con los resultados citados en el último apartado del capítulo anterior, en esta ocasión, el método de la razón de verosimilitudes ha conducido al test insesgado uniformemente de máxima potencia.

### Ejemplo 9.3

Para la misma población del ejemplo anterior, se desea comprobar ahora si  $\sigma = \sigma_0$ ; es decir hacer el contraste entre las hipótesis definidas por

$$\Theta_0 = \{(\mu, \sigma) \mid \mu \in \mathbb{R}, \sigma = \sigma_0\} \quad \text{y} \quad \Theta_1 = \{(\mu, \sigma) \mid \mu \in \mathbb{R}, \sigma \neq \sigma_0\}.$$

Como en el ejemplo anterior

$$\sup_{(\mu, \sigma) \in \Theta} f_{\mu, \sigma}(x_1, \dots, x_n) = (2\pi s^2)^{-n/2} e^{-n/2},$$

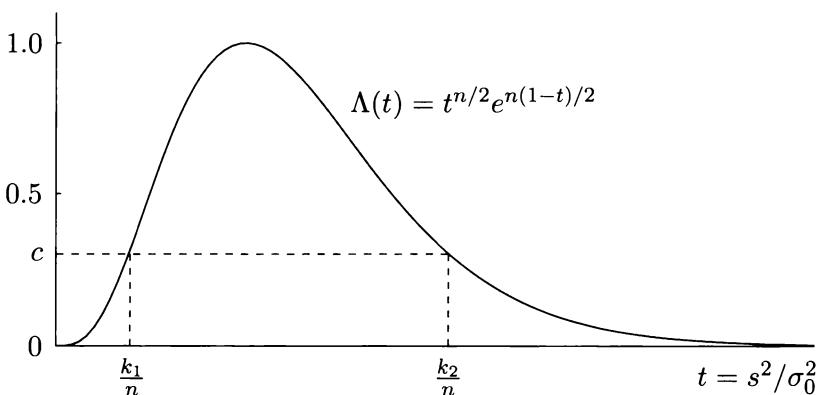
pero ahora

$$\sup_{(\mu, \sigma) \in \Theta_0} f_{\mu, \sigma}(x_1, \dots, x_n) = (2\pi \sigma_0^2)^{-n/2} e^{-ns^2/(2\sigma_0^2)};$$

de manera que

$$\Lambda(x_1, \dots, x_n) = \left( \frac{s^2}{\sigma_0^2} \right)^{n/2} e^{-ns^2/2\sigma_0^2} e^{n/2}.$$

Como función de  $s^2/\sigma_0^2$ ,  $\Lambda$  es una función creciente en  $(0, 1)$  y decreciente en  $(1, \infty)$  (que alcanza el máximo 1 para  $s^2 = \sigma_0^2$ ).



Por consiguiente, la desigualdad  $\Lambda < c$  se verifica en

$$C = \left\{ n \frac{s^2}{\sigma_0^2} < k_1 \right\} \cup \left\{ n \frac{s^2}{\sigma_0^2} > k_2 \right\},$$

donde debe ser

$$(k_1/n)^{n/2} e^{-k_1/2} = (k_2/n)^{n/2} e^{-k_2/2},$$

es decir

$$k_1 - n \log k_1 = k_2 - n \log k_2.$$

Además, como  $ns^2/\sigma_0^2$  tiene distribución  $\chi_{n-1}^2$  cuando la hipótesis nula es cierta, para que el contraste tenga tamaño  $\alpha$ , tiene que cumplirse

$$P\{k_1 < \chi_{n-1}^2 < k_2\} = 1 - \alpha.$$

El contraste de razón de verosimilitudes tiene, por tanto, región crítica de forma similar a la del contraste insesgado uniformemente de máxima potencia citado al final del capítulo anterior; las condiciones de determinación de los puntos críticos  $k_1$  y  $k_2$  son sin embargo distintas. A éste propósito, como no es fácil resolver el sistema de ecuaciones constituido por las dos últimas igualdades, en la práctica se toma también:  $k_1 = \chi_{n-1;1-\alpha/2}^2$ ,  $k_2 = \chi_{n-1;\alpha/2}^2$ .

La misma técnica puede utilizarse para contrastes relativos a dos poblaciones.

#### Ejemplo 9.4

Dos poblaciones independientes tienen distribución  $N(\mu_1, \sigma)$  y  $N(\mu_2, \sigma)$  respectivamente, con la misma desviación típica  $\sigma$ , desconocida. Mediante sendas muestras aleatorias simples,  $(x_1, \dots, x_n)$  e  $(y_1, \dots, y_m)$ , se desea contrastar la igualdad de las medias poblacionales. Es decir la hipótesis nula correspondiente al conjunto

$$\Theta_0 = \{(\mu_1, \mu_2, \sigma) \mid \mu_1 = \mu_2, \sigma > 0\},$$

dentro del espacio paramétrico

$$\Theta = \{(\mu_1, \mu_2, \sigma) \mid \mu_1, \mu_2 \in I\!\!R, \sigma > 0\}.$$

La función de verosimilitud

$$\left( \frac{1}{\sqrt{2\pi}\sigma} \right)^{n+m} e^{-\sum_{i=1}^n (x_i - \mu_1)^2 / 2\sigma^2} e^{-\sum_{i=1}^m (y_i - \mu_2)^2 / 2\sigma^2}$$

(o su logaritmo) alcanza su máximo en

$$\hat{\mu}_1 = \bar{x}, \quad \hat{\mu}_2 = \bar{y}, \quad \hat{\sigma}^2 = \frac{ns_1^2 + ms_2^2}{n+m},$$

donde vale

$$\left(\frac{1}{\sqrt{2\pi}\hat{\sigma}}\right)^{n+m} e^{-ns_1^2/2\hat{\sigma}^2} e^{-ms_2^2/2\hat{\sigma}^2} = \left(\frac{1}{\sqrt{2\pi}\hat{\sigma}}\right)^{n+m} e^{-(n+m)/2}.$$

En cambio en  $\Theta_0$ , el máximo de la verosimilitud se alcanza con

$$\hat{\mu} = \frac{n\bar{x} + m\bar{y}}{n+m}, \quad \hat{\sigma}_0^2 = \frac{1}{n+m} \left( \sum_{i=1}^n (x_i - \hat{\mu})^2 + \sum_{i=1}^m (y_i - \hat{\mu})^2 \right),$$

siendo el valor del máximo

$$\left(\frac{1}{\sqrt{2\pi}\hat{\sigma}_0}\right)^{n+m} e^{-(n+m)/2}.$$

Como  $(n+m)\hat{\sigma}_0^2 = ns_1^2 + (\bar{x} - \hat{\mu})^2 + ms_2^2 + (\bar{y} - \hat{\mu})^2$ , resulta

$$\begin{aligned} \Lambda &= \left(\frac{\hat{\sigma}}{\hat{\sigma}_0}\right)^{n+m} = \left(\frac{ns_1^2 + ms_2^2}{ns_1^2 + (\bar{x} - \hat{\mu})^2 + ms_2^2 + (\bar{y} - \hat{\mu})^2}\right)^{(n+m)/2} \\ &= \left(1 + \frac{m^2 + n^2}{(n+m)^2} \frac{(\bar{x} - \bar{y})^2}{ns_1^2 + ms_2^2}\right)^{-(n+m)/2} \end{aligned}$$

puesto que  $(\bar{x} - \hat{\mu})^2 + (\bar{y} - \hat{\mu})^2 = (n^2 + m^2)(\bar{x} - \bar{y})^2/(n+m)^2$ .

La desigualdad  $\Lambda < c$  define por tanto la región crítica

$$C = \left\{ \frac{|\bar{x} - \bar{y}|}{\sqrt{ns_1^2 + ms_2^2}} > k' \right\} = \left\{ \frac{|\bar{x} - \bar{y}|/\sqrt{1/n + 1/m}}{\sqrt{(ns_1^2 + ms_2^2)/(n+m-2)}} > k \right\}$$

donde el estadístico de la última expresión tiene distribución  $t_{n+m-2}$ ; de forma que ha de ser  $k = t_{n+m-2;\alpha/2}$ , para que el contraste tenga tamaño  $\alpha$ .

Al contrario de lo que sucede en los ejemplos anteriores, hay ocasiones en que el estadístico  $\Lambda$  no puede expresarse en términos de otro estadístico más simple, cuya distribución se conozca.

### Ejemplo 9.5

Se quiere comprobar si en un dado las caras opuestas: 1 y 4, 2 y 5 y 3 y 6, tienen la misma probabilidad. Después de  $n$  lanzamientos, si  $p_i$  representa la probabilidad del resultado  $i$  y  $n_i$  el número de tales resultados obtenidos, la verosimilitud de las observaciones es

$$p_1^{n_1} p_2^{n_2} p_3^{n_3} p_4^{n_4} p_5^{n_5} p_6^{n_6} \left(1 - \sum_{i=1}^5 p_i\right)^{n_6},$$

lo cual da como estimador de máxima verosimilitud de las probabilidades de cada cara  $\hat{p}_i = n_i/n$  (véase el ejemplo 7.6). La hipótesis a contrastar es  $H_0 : p_1 = p_4, p_2 = p_5, p_3 = p_6$ . Supuesto que se verifica, tomando como parámetros  $p_1$  y  $p_2$ , la verosimilitud resulta

$$p_1^{n_1+n_4} p_2^{n_2+n_5} \left( \frac{1 - 2p_1 - 2p_2}{2} \right)^{n_3+n_6},$$

que alcanza su máximo para

$$\tilde{p}_1 = \frac{n_1 + n_4}{2n}, \quad \tilde{p}_2 = \frac{n_2 + n_5}{2n}.$$

La razón de verosimilitudes es entonces

$$\Lambda = \frac{(n_1 + n_4)^{n_1+n_4} (n_2 + n_5)^{n_2+n_5} (n_3 + n_6)^{n_3+n_6}}{2^n n_1^{n_1} n_2^{n_2} n_3^{n_3} n_4^{n_4} n_5^{n_5} n_6^{n_6}}.$$

No hay, en este caso, ninguna simplificación sustancial de la región crítica  $\{\Lambda < c\}$ . A partir de los valores de  $n_i$  hay que obtener el valor de  $\Lambda$  para rechazar cuando sea menor que  $c$ . Por otra parte, para determinar el punto crítico  $c$ , a fin de que el test tenga nivel de significación  $\alpha$ , sería necesario conocer la distribución del estadístico  $\Lambda$ . Pero no hay ninguna manera simple de obtener dicha distribución.

En tales circunstancias, es muy conveniente disponer de un resultado general que proporcione la distribución asintótica de  $\Lambda$  cuando el tamaño muestral crece. Ello permitirá, al menos, una determinación aproximada del punto crítico  $c$ , válida para tamaños muestrales grandes. Este es el sentido del siguiente teorema<sup>1</sup>.

**Teorema 9.1:** Sobre una población, cuya distribución depende de un parámetro  $\theta$  que varía en  $\Theta \subset \mathbb{R}^k$ , se quiere contrastar la hipótesis nula definida por

$$\Theta_0 = \{\theta \in \Theta \mid \theta_i = g_i(\omega_1, \dots, \omega_q) \text{ con } (\omega_1, \dots, \omega_q) \in \Omega\},$$

siendo  $\Omega$  un abierto de  $\mathbb{R}^q$  y  $g_i$  funciones con derivadas parciales de primer orden continuas.

Si se verifican las condiciones (1), (2) y (3) del teorema 7.2 (para cada  $\theta_i$ ), entonces, para cada  $\theta \in \Theta_0$ , se cumple

$$-2 \log \Lambda(X_1, \dots, X_n) \xrightarrow{d_\theta} \chi_{k-q}^2.$$

---

<sup>1</sup> Cf. Rao [1973] página 418.

El resultado está ligado con el comportamiento asintótico de los estimadores de máxima verosimilitud para  $\theta$ :  $\hat{\theta}$  y  $\hat{\theta}_0$ , en  $\Theta$  y  $\Theta_0$  respectivamente; puesto que

$$-2 \log \Lambda(X_1, \dots, X_n) = 2 \log f_{\hat{\theta}}(X_1, \dots, X_n) - 2 \log f_{\hat{\theta}_0}(X_1, \dots, X_n).$$

En la aplicaciones usuales, la conclusión del teorema se da por garantizada, ya que —como ocurría con la normalidad asintótica de los estimadores de máxima verosimilitud— las hipótesis son suficientes, pero no necesarias, para utilizar la distribución  $\chi^2$  en la determinación del punto crítico del test. Por tanto, en situaciones no anómalas, la única precaución importante es observar el número de parámetros  $q$ , bajo la hipótesis nula, y el número de parámetros  $k$  prescindiendo de la hipótesis.

### Ejemplo 9.5 (continuación)

En el ejemplo anterior es  $k = 5$  y  $q = 2$ ; luego, si  $n$  es grande,

$$-2 \log \Lambda \text{ tiene aproximadamente distribución } \chi_3^2.$$

Naturalmente,  $\{\Lambda < c\}$  equivale a  $\{-2 \log \Lambda > c'\}$ , de forma que el test de razón de verosimilitudes, de nivel  $\alpha$ , tiene (aproximadamente) región crítica

$$2n \log 2 + 2 \sum_{i=1}^6 n_i \log n_i - 2 \sum_{i=1}^3 (n_i + n_{i+3}) \log(n_i + n_{i+3}) > \chi_{3;\alpha}^2.$$

### Ejemplo 9.6

Un ejemplo clásico de aplicación del contraste de razón de verosimilitudes es la justificación del test de Bartlett para comprobar la igualdad de varianzas de  $k$  poblaciones independientes y normales,  $N(\mu_i, \sigma_i)$ , cuando todos los parámetros son desconocidos. Es decir se trata de contrastar la hipótesis:

$$H_0 : \sigma_1 = \sigma_2 = \dots = \sigma_k$$

frente a la alternativa de que al menos hay dos diferentes.

Con muestras aleatorias simples de cada una de las poblaciones, de tamaños  $n_1, \dots, n_k$ , tales que  $n = \sum_{i=1}^k n_i$ , la verosimilitud de las observaciones es

$$\frac{1}{(2\pi)^{n/2} \sigma_1^{n_1} \cdots \sigma_k^{n_k}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^k \frac{1}{\sigma_i^2} \sum_{j=1}^{n_i} (x_{ij} - \mu_i)^2 \right\}.$$

Los estimadores de máxima verosimilitud resultan

$$\hat{\mu}_i = \bar{x}_i \quad \text{y} \quad \hat{\sigma}_i^2 = s_i^2$$

y proporcionan a la verosimilitud el valor

$$\frac{1}{(2\pi)^{n/2} s_1^{n_1} \cdots s_k^{n_k}} e^{-n/2}.$$

En cambio, bajo la hipótesis nula la verosimilitud es

$$\frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \mu_i)^2 \right\},$$

que da lugar a los estimadores de máxima verosimilitud

$$\hat{\mu}_i = \bar{x}_i \quad \text{y} \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^k n_i s_i^2,$$

para los cuales la verosimilitud vale

$$\frac{1}{(2\pi)^{n/2} \hat{\sigma}^n} e^{-n/2}.$$

Por consiguiente

$$\Lambda = \frac{s_1^{n_1} \cdots s_k^{n_k}}{\hat{\sigma}^n} \quad \text{y} \quad -2 \log \Lambda = 2n \log \hat{\sigma} - 2 \sum_{i=1}^k n_i \log s_i.$$

Ambos son estadísticos dependientes de las varianzas muestrales (y de su promedio ponderado:  $\hat{\sigma}$ ).

El modelo cuenta con  $2k$  parámetros desconocidos que se reducen a  $k+1$  si la hipótesis nula es correcta; por tanto  $-2 \log \Lambda$  tiene aproximadamente distribución  $\chi_{k-1}^2$ . La región crítica aproximada del test de razón de verosimilitudes, de tamaño  $\alpha$ , es pues

$$C = \{-2 \log \Lambda > \chi_{k-1;\alpha}^2\}.$$

## Contrastes múltiples

El último ejemplo suscita la cuestión de si no sería mejor comparar las varianzas poblacionales dos a dos, en lugar todas a la vez como hace el test de Bartlett. Es decir, la alternativa consiste en contrastar cada hipótesis del tipo  $H_0^{(ij)} : \sigma_i = \sigma_j$  independientemente, mediante las muestras de las poblaciones  $i$  y  $j$ .

Concretamente, la conclusión del ejemplo anterior, con  $k = 2$  indica que el test de razón de verosimilitudes para contrastar  $H_0^{(ij)}$  tiene región crítica

$$\begin{aligned} C_{ij} &= \left\{ s_i^{n_i} s_j^{n_j} \left/ \left( \frac{n_i s_i^2 + n_j s_j^2}{n_i + n_j} \right)^{(n_i+n_j)/2} < c \right. \right\} \\ &= \left\{ \left( 1 + \frac{n_j s_j^2}{n_i s_i^2} \right)^{n_i/2} \left( 1 + \frac{n_i s_i^2}{n_j s_j^2} \right)^{n_j/2} > c' \right\}. \end{aligned}$$

Como la función  $(1+t)^{n_i/2}(1+1/t)^{n_j/2}$  es decreciente para  $t < n_j/n_i$  y creciente para  $t > n_j/n_i$ , la región de aceptación se puede expresar en la forma

$$C_{ij}^c = \left\{ k'_1 < \frac{n_j s_j^2}{n_i s_i^2} < k'_2 \right\} = \left\{ k_1 < \frac{s_j^2/(n_j - 1)}{s_i^2/(n_i - 1)} < k_2 \right\},$$

donde debería ser

$$(1 + k'_1)^{n_i/2} (1 + 1/k'_1)^{n_j/2} = (1 + k'_2)^{n_i/2} (1 + 1/k'_2)^{n_j/2}$$

(lo cual impone una primera condición a  $k_1 = n_i(n_i - 1)k'_1/(n_j(n_j - 1))$  y  $k_2 = n_i(n_i - 1)k'_2/(n_j(n_j - 1))$ ) y, si se quiere que el contraste de  $H_0^{(ij)}$  tenga tamaño  $\alpha'$ , ha de ser

$$P\{k_1 < F_{n_j-1, n_i-1} < k_2\} = 1 - \alpha',$$

habida cuenta que el estadístico  $(n_i - 1)s_j^2/((n_j - 1)s_i^2)$  tiene distribución  $F_{n_j-1, n_i-1}$ , bajo la hipótesis  $H_0^{(ij)}$ .

A costa de hallar la solución  $k_1$  y  $k_2$  de las dos últimas ecuaciones (para cada  $i, j$ ), se consigue así, sin recurrir a resultados aproximados, contrastar con nivel de significación  $\alpha'$ , cada una de las hipótesis  $H_0^{(ij)}$ . Pero muy bien podría ocurrir que las observaciones llevasen a aceptar  $H_0^{(12)}$  y  $H_0^{(23)}$ ,

y sin embargo rechazar  $H_0^{(13)}$  y, por consiguiente,  $H_0$ . Es decir, los datos pueden no contener evidencia suficiente para rechazar  $\sigma_1 = \sigma_2$ , ni  $\sigma_2 = \sigma_3$ , pero resultar significativos de que  $\sigma_1 \neq \sigma_3$ . No queda pues más remedio que efectuar los  $p = {k \choose 2}$  contrastes correspondientes a cada pareja; lo cual significa utilizar como región crítica de  $H_0$  la región:

$$C' = \bigcup_{i < j=1}^k C_{ij}.$$

Lo malo de este procedimiento es la influencia del tamaño de cada test individual en el nivel de significación del contraste global; en general sólo se puede afirmar que

$$\mathcal{P}_{H_0} \left( \bigcup_{i < j=1}^k C_{ij} \right) \leq p\alpha'.$$

De forma que, si se quiere conseguir nivel de significación  $\alpha$  para el contraste de  $H_0$ , habrá que tomar  $\alpha' = \alpha/p$ . El tamaño puede ser entonces bastante menor que  $\alpha$  y la diferencia se traducirá en una notable pérdida de potencia, es decir mayor probabilidad de aceptar  $H_0$  cuando es falsa, que haciendo el contraste global.

En conclusión, ante situaciones como la anterior en que se trata de contrastar una hipótesis múltiple del tipo

$$H_0 : \theta \in \Theta_0 = \bigcap_{l=1}^r \Theta_l,$$

es preferible utilizar la razón de verosimilitudes con  $\Theta_0$ , en lugar de obtener la región crítica  $C_l$  para cada  $H_0^{(l)} : \theta \in \Theta_l$  y formar luego la región crítica

$$C' = \bigcup_{l=1}^r C_l.$$

La segunda manera de proceder únicamente es aceptable, en aquellos casos en que las regiones  $C_l$  sean independientes, de manera que, si cada una tiene tamaño  $\alpha'$ , sea

$$\mathcal{P}_{H_0} \left( \bigcup_{l=1}^r C_l \right) = 1 - (1 - \alpha')^r$$

lo cual permite fijar  $\alpha' = 1 - (1 - \alpha)^{1/r}$ , para que el tamaño del test obtenido sea exactamente igual a  $\alpha$ . Este es el caso, por ejemplo, si se quieren

realizar contrastes simultáneos sobre la media y la varianza de una población normal: el teorema de Fisher garantiza que los estadísticos empleados ( $\sqrt{n-1}(\bar{x} - \mu_0)/s$  y  $ns^2/\sigma_0^2$ ) son independientes; luego se pueden realizar tests individuales, de tamaños  $\alpha_1$  y  $\alpha_2$ , para conseguir  $1 - (1 - \alpha_1)(1 - \alpha_2)$  como tamaño del test conjunto.

En ausencia de independencia, hay que insistir en que el test global es preferible. Sólo en el caso de que este test rechace  $H_0$ , puede resultar de interés realizar contrastes parciales, para analizar con más detalle a cuales de las  $H_0^{(l)}$  es atribuible la falsedad de  $H_0$ . Las técnicas específicas que para ello existen, en ciertos casos, no se expondrán aquí, por ser más propias de un capítulo sobre *Análisis de la varianza*.

### 9.3. Relación entre estimación confidencial y contraste de hipótesis

Es posible apreciar, sobre todo en los ejemplos, una gran analogía entre los resultados de la estimación confidencial y del contraste de hipótesis. Así, en el ejemplo 9.2, se ha concluido que, con nivel de significación  $\alpha$ , debía aceptarse la hipótesis  $\mu = \mu_0$ , acerca de la media de una población  $N(\mu, \sigma)$ , con  $\sigma$  desconocida, en caso de que  $\bar{x}$  pertenezca al intervalo

$$\left[ \mu_0 - t_{n-1;\alpha/2} \frac{s}{\sqrt{n-1}}, \mu_0 + t_{n-1;\alpha/2} \frac{s}{\sqrt{n-1}} \right],$$

lo cual equivale a decir que  $\mu_0$  pertenezca al intervalo

$$\left[ \bar{x} - t_{n-1;\alpha/2} \frac{s}{\sqrt{n-1}}, \bar{x} + t_{n-1;\alpha/2} \frac{s}{\sqrt{n-1}} \right].$$

Pero éste es precisamente el intervalo de nivel de confianza  $1 - \alpha$  construido en las mismas circunstancias para  $\mu$  (véase §4.5 (B)).

La coincidencia no tiene nada de casual, sino que obedece a la similitud entre los planteamientos de la estimación confidencial y el contraste de hipótesis simples.

- ▷ En el primer caso, se trata de construir un conjunto  $S(x_1, \dots, x_n)$  del espacio paramétrico  $\Theta$ , dependiente de la muestra observada, de manera que se cumpla

$$\mathcal{P}_\theta \{S(X_1, \dots, X_n) \ni \theta\} \geq 1 - \alpha,$$

para cada  $\theta \in \Theta$ . La región  $S(x_1, \dots, x_n)$  puede interpretarse entonces como el conjunto de valores de  $\theta$  compatibles con las observaciones, a nivel de confianza  $1 - \alpha$ .

- ▷ Por su parte, el contraste de una hipótesis simple  $H_0 : \theta = \theta_0$  consiste en determinar una región  $C(\theta_0)$  del espacio muestral  $\mathcal{X}$ , de manera que la obtención de una muestra en  $C(\theta_0)$  resulte inconsistente con  $H_0$ , a nivel de significación  $\alpha$ . Lo cual significa que sea

$$\mathcal{P}_{\theta_0}\{(X_1, \dots, X_n) \in C(\theta_0)\} \leq \alpha.$$

Resulta entonces inmediato que, conocidos los diversos  $C(\theta)$ , tomando

$$S(x_1, \dots, x_n) = \{\theta \in \Theta \mid (x_1, \dots, x_n) \notin C(\theta)\}$$

se obtiene una región de confianza para  $\theta$  de nivel de confianza  $1 - \alpha$ . De hecho, para cada  $\theta \in \Theta$ ,

$$\mathcal{P}_\theta\{S(X_1, \dots, X_n) \ni \theta\} = \mathcal{P}_\theta\{(X_1, \dots, X_n) \notin C(\theta)\} \geq 1 - \alpha.$$

O, recíprocamente, dada la región  $S(X_1, \dots, X_n)$ ,

$$C(\theta_0) = \{(x_1, \dots, x_n) \in \mathcal{X} \mid \theta_0 \notin S(x_1, \dots, x_n)\}$$

es la región crítica de un test, de nivel de significación  $\alpha$ , para la hipótesis  $\theta = \theta_0$ . En efecto,

$$\mathcal{P}_{\theta_0}\{(X_1, \dots, X_n) \in C(\theta_0)\} = \mathcal{P}_{\theta_0}\{S(X_1, \dots, X_n) \ni \theta_0\} \leq \alpha.$$

Según lo anterior, una vez construido un intervalo de confianza para un parámetro se puede reinterpretar para obtener un test para contrastar cualquier hipótesis simple sobre él. Por regla general, los tests asociados a intervalos acotados serán adecuados para el contraste frente a hipótesis alternativas bilaterales; mientras que, para hipótesis alternativas unilaterales, son preferibles los tests asociados a cotas de confianza (es decir intervalos de confianza uno de cuyos extremos coincide con el extremo del espacio paramétrico).

Por ejemplo, los sucesivos apartados de la sección 4.5 proporcionan tests con buenas propiedades para contrastar, acerca de los parámetros de una población normal, las hipótesis:

- ▷  $H_0 : \mu = \mu_0$  frente a  $H_1 : \mu \neq \mu_0$ , (A) si  $\sigma$  es conocida, (B) si  $\sigma$  es desconocida.
- ▷  $H_0 : \sigma = \sigma_0$  frente a  $H_1 : \sigma \neq \sigma_0$ , (C) si  $\mu$  es conocida, (D) si  $\mu$  es desconocida.
- ▷  $H_0 : \mu = \mu_0, \sigma = \sigma_0$  frente a  $H_1 : \mu \neq \mu_0, \sigma \neq \sigma_0$  mediante la región obtenida en (E).

O bien, sobre dos poblaciones normales, las hipótesis:

- ▷  $H_0 : \mu_1 = \mu_2$  frente a  $H_1 : \mu_1 \neq \mu_2$ , (F) si  $\sigma_1, \sigma_2$  son conocidas, (G) si son desconocidas, (I) en el caso de datos apareados.
- ▷  $H_0 : \sigma_1 = \sigma_2$  frente a  $H_1 : \sigma_1 \neq \sigma_2$ , mediante el intervalo construido en (H).

El siguiente ejemplo muestra la manera genérica de proceder.

### Ejemplo 9.7

La duración  $X$  de ciertas componentes tiene distribución exponencial de parámetro  $\theta > 0$ . A partir de una muestra aleatoria simple de tamaño  $n$ , la construcción de un intervalo de confianza para  $\theta$  puede hacerse teniendo en cuenta que  $T = \sum_{i=1}^n X_i$  tiene distribución gamma  $\gamma(n, \theta)$ , o bien,  $2\theta T$  distribución  $\chi^2_{2n}$  (cf. ejercicio 3.8). Por consiguiente,

$$\mathcal{P}_\theta \{ \chi^2_{2n;1-\alpha/2} < 2\theta T < \chi^2_{2n;\alpha/2} \} = 1 - \alpha$$

y

$$\left[ \frac{\chi^2_{2n;1-\alpha/2}}{2T}, \frac{\chi^2_{2n;\alpha/2}}{2T} \right]$$

es un intervalo de confianza, de nivel  $1 - \alpha$  para  $\theta$ .

Para contrastar  $H_0 : \theta = \theta_0$  frente a  $H_1 : \theta \neq \theta_0$  se puede entonces adoptar el test que acepta  $H_0$  cuando

$$\theta_0 \in [ \chi^2_{2n;1-\alpha/2}/(2T), \chi^2_{2n;\alpha/2}/(2T) ],$$

es decir, con región crítica

$$C = \left\{ T < \frac{\chi^2_{2n;1-\alpha/2}}{2\theta_0} \right\} \cup \left\{ T > \frac{\chi^2_{2n;\alpha/2}}{(2\theta_0)} \right\}$$

cuyo nivel de significación es  $\alpha$ .

En sentido contrario, puede observarse que la densidad de la muestra

$$f_\theta(x_1, \dots, x_n) = \theta^n e^{-\theta \sum_{i=1}^n x_i}$$

tiene razón de verosimilitud monótona en  $-T$ , de forma que para el contraste de  $H_0 : \theta = \theta_0$  frente a  $H_1 : \theta > \theta_0$  es uniformemente de máxima potencia, a nivel de significación  $\alpha$ , el test de región crítica

$$C(\theta_0) = \{T < \chi^2_{2n;1-\alpha}/(2\theta_0)\}.$$

La región

$$S(x_1, \dots, x_n) = \{\theta > 0 \mid T > \chi^2_{2n;1-\alpha}/(2\theta)\} = \left( \frac{\chi^2_{2n;1-\alpha}}{2T}, \infty \right)$$

proporciona una cota inferior de confianza para el parámetro, de nivel de confianza  $1 - \alpha$ .

La única asimetría que existe entre intervalos de confianza y tests de hipótesis simples, radica en los criterios de calidad con que se escogen unos y otros. Buscar un intervalo de confianza de longitud mínima, aunque es sensato, no es un criterio muy profundo y no conducirá necesariamente a tests óptimos en ningún sentido. En cambio, conseguir un test uniformemente de máxima potencia tiene un significado muy preciso en relación con el contraste de hipótesis, que admite su traducción al problema de la estimación confidencial. Para ello bastará comparar diversas regiones de confianza en base a la probabilidad de que cada una cubra valores “erróneos” del parámetro. Es decir:

*Si  $S(x_1, \dots, x_n)$  es una región de confianza para un parámetro  $\theta$ , se denomina función de precisión de  $S$  a*

$$\mathcal{P}_\theta\{S(X_1, \dots, X_n) \ni \theta'\},$$

*que mide la probabilidad, supuesto que el valor del parámetro es  $\theta$ , de que  $S$  contenga a otro valor diferente  $\theta'$ .*

*Si  $S(x_1, \dots, x_n)$  y  $S'(x_1, \dots, x_n)$  son regiones de confianza para  $\theta$ , con el mismo nivel de confianza  $1 - \alpha$ , la primera se denomina uniformemente más precisa que la segunda si, para cualesquiera  $\theta \neq \theta' \in \Theta$ , se cumple*

$$\mathcal{P}_\theta\{S(X_1, \dots, X_n) \ni \theta'\} \leq \mathcal{P}_\theta\{S'(X_1, \dots, X_n) \ni \theta'\};$$

ya que ello indica que, cualquiera que sea  $\theta$ ,  $S$  tiene menos probabilidad que  $S'$  de contener a los restantes valores  $\theta'$ .

Consecuentemente, una región  $S(x_1, \dots, x_n)$  con nivel de confianza  $1 - \alpha$ , se denomina uniformemente de máxima precisión cuando es uniformemente más precisa que cualquier otra región  $S'$ , de nivel de confianza  $1 - \alpha$ .

Es evidente que la definición anterior está parafraseando los conceptos relativos a tests, en el lenguaje de las regiones de confianza. No es de extrañar por tanto que sea fácil demostrar que:

*Si, para cada  $\theta_0 \in \Theta$ ,  $C(\theta_0)$  es la región crítica de un test uniformemente de máxima potencia para contrastar  $H_0 : \theta = \theta_0$  frente a  $H_1 : \theta \in \Theta - \{\theta_0\}$ , con nivel de significación  $\alpha$ , entonces*

$$S(x_1, \dots, x_n) = \{\theta \in \Theta \mid (x_1, \dots, x_n) \notin C(\theta)\}$$

es una región de confianza, de nivel de confianza  $1 - \alpha$ , uniformemente de máxima precisión. Y recíprocamente.

Según esto, cada resultado acerca de la construcción de tests uniformemente de máxima potencia, admite su contrapartida en términos de regiones de confianza uniformemente de máxima precisión. No es muy útil insistir en ello, puesto que, en la práctica, basta interpretar los problemas de estimación confidencial como contrastes de hipótesis. Así, en el ejemplo 9.7 (supuesto que es  $\Theta = [\theta_0, \infty)$ ), el intervalo  $(\chi^2_{2n;1-\alpha}/(2T), \infty)$  es una región de confianza uniformemente de máxima precisión, de nivel de confianza  $1 - \alpha$ .

### Ejemplo 9.8

En la situación del ejemplo 4.3, la densidad de la muestra es

$$f_\theta(x_1, \dots, x_n) = \theta^n e^{(\theta-1)T},$$

siendo  $T(x_1, \dots, x_n) = \sum_{i=1}^n \log x_i$ . Si queremos obtener un intervalo de confianza para  $\theta$ , pensemos en el contraste de la hipótesis  $H_0 : \theta = \theta_0$  frente a  $H_1 : \theta \neq \theta_0$ . Según el teorema de Lehmann, el test de región de aceptación<sup>2</sup>

$$C^c(\theta_0) = \{c_1 < T < c_2\} = \{k_1 < -T\theta_0 < k_2\}$$

---

<sup>2</sup> Conviene expresarla mediante  $-T\theta_0$  puesto que su distribución es  $\gamma(n, 1)$  (si  $H_0$  es cierta).

es insesgado uniformemente de máxima potencia y de nivel de significación  $\alpha$ , supuesto que las constantes  $c_1, c_2$  (o bien  $k_1, k_2$ ) se determinan mediante las condiciones

$$\left\{ \begin{array}{l} \mathcal{P}_{\theta_0}\{c_1 < T < c_2\} = \int_{k_1}^{k_2} \frac{t^{n-1}e^{-t}}{(n-1)!} dt = 1 - \alpha, \\ \mathbb{E}_{\theta_0}[TI_{\{c_1 < T < c_2\}}] = -\frac{1}{\theta_0} \int_{k_1}^{k_2} \frac{t^n e^{-t}}{(n-1)!} dt = -(1-\alpha)\frac{n}{\theta_0}, \end{array} \right.$$

(habida cuenta que  $\mathbb{E}_{\theta_0}[T] = -n/\theta_0$ ). La región de confianza asociada:

$$S(x_1, \dots, x_n) = \{\theta \mid k_1 < -T\theta < k_2\} = \left( \frac{k_1}{-T}, \frac{k_2}{-T} \right)$$

tiene nivel de confianza  $1 - \alpha$ . Lo único que se ha ganado, respecto a la conclusión del ejemplo 4.3, es que —si nos tomamos la molestia de hallar la solución  $k_1$  y  $k_2$  de las ecuaciones anteriores— el resultado constituye la región de confianza insesgada<sup>3</sup> uniformemente de máxima precisión.

## Ejercicios

### Ejercicio 9.1

Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de una distribución exponencial de media  $\theta$ , con densidad

$$f_\theta(x) = \frac{1}{\theta} e^{-x/\theta} \quad \text{para } x > 0.$$

- a) Determinar el contraste de razón de verosimilitudes para contrastar  $H_0 : \theta \leq \theta_0$  frente a  $H_1 : \theta > \theta_0$ , con nivel de significación  $\alpha$ .
- b) Determinar el contraste de razón de verosimilitudes para contrastar, con nivel de significación  $\alpha$ ,  $H_0 : \theta \notin (3, 5)$  frente a  $H_1 : \theta \in (3, 5)$ . Discutir el tamaño muestral que debe tomarse para hallar una región crítica razonable.

---

<sup>3</sup> Esto es,  $\mathcal{P}_{\theta'}\{S(X_1, \dots, X_n) \ni \theta\} \leq 1 - \alpha$  si  $\theta' \neq \theta$ ; lo cual significa que, en caso de que el valor real del parámetro no fuese  $\theta$ , la probabilidad de que  $S$  incluya el valor  $\theta$  es inferior a la probabilidad con que lo incluye cuando realmente el parámetro vale  $\theta$ .

**Ejercicio 9.2**

Se dispone de una muestra aleatoria simple de una distribución geométrica con función de probabilidad

$$f_{\theta}(x) = (1 - \theta) \theta^x \quad \text{para } x = 0, 1, 2, \dots$$

donde  $\theta$  es una probabilidad desconocida. Determinar el contraste de razón de verosimilitudes para contrastar  $H_0 : \theta \leq \theta_0$  frente a  $H_1 : \theta > \theta_0$ . Comprobar que el test obtenido es uniformemente de máxima potencia.

**Ejercicio 9.3**

Dada una muestra aleatoria simple de tamaño  $n$  de una variable uniformemente distribuida en el intervalo  $[\theta_1, \theta_2]$ , determinar el test de razón de verosimilitudes para contrastar  $H_0 : \theta_1 \leq 0 \leq \theta_2$  frente a  $H_1 : \theta_1 > 0$  ó  $\theta_2 < 0$ .

**Ejercicio 9.4**

Sea  $(X_1, Y_1), \dots, (X_n, Y_n)$  una muestra aleatoria simple de una distribución normal bidimensional, de medias, varianzas y covarianza desconocidas, cuyo coeficiente de correlación se representa por  $\rho$ . Determinar el contraste de razón de verosimilitudes para contrastar  $H_0 : \rho = 0$  frente a  $H_1 : \rho \neq 0$ .

**Ejercicio 9.5**

Sea  $(X_1, Y_1), \dots, (X_n, Y_n)$  una muestra aleatoria simple de una distribución normal bidimensional, de vector de medias  $(\mu_1, \mu_2)$ , cuyas componentes tienen la misma varianza  $\sigma^2$  y con coeficiente de correlación  $\rho$ .

Determinar el test de razón de verosimilitudes para contrastar  $H_0 : \rho = 0$  frente a  $H_1 : \rho \neq 0$ .

**Ejercicio 9.6**

Las fluctuaciones  $X_1, \dots, X_n$  en la cotización de un determinado valor bursátil, respecto a su valor nominal, a lo largo de sesiones consecutivas, se supone que obedece a la ley

$$X_i = \theta X_{i-1} + e_i,$$

siendo  $X_0 = 0$  y  $e_i$  variables independientes y  $N(0, \sigma)$ .

Determinar el contraste de razón de verosimilitudes para contrastar la hipótesis nula de independencia  $H_0 : \theta = 0$  frente a la alternativa  $H_1 : \theta \neq 0$ , distinguiendo los casos en que  $\sigma$  sea conocida (por ser prácticamente común para todos los valores bursátiles) o desconocida.

**Ejercicio 9.7**

En un estudio sociológico se ha observado que la proporción de su renta que una familia gasta en bienes de primera necesidad sigue una distribución con densidad

$$f_{\theta}(x) = \theta x^{\theta-1} \quad \text{para } 0 < x < 1,$$

donde  $\theta$  es un parámetro que mide la pobreza de la sociedad. A fin de comparar dos poblaciones independientes, se ha observado la característica en cuestión en  $n$  familias de la primera y  $m$  familias de la segunda. Determinar el test de razón de verosimilitudes para contrastar  $H_0 : \theta_1 \leq \theta_2$  frente a  $H_1 : \theta_1 > \theta_2$ .

**Ejercicio 9.8**

Con objeto de comparar la virulencia de dos organismos patógenos, se inoculó con el organismo I a  $n$  conejillos de indias, de los que  $x$  manifestaron signos de enfermedad, y a otros  $m$  con el organismo II, entre los cuales  $y$  enfermaron. Los animales se mantuvieron aislados, de forma que puede suponerse que existe independencia entre los resultados en cada uno. Sean  $p_1$  y  $p_2$  las frecuencias con que cada tipo de organismo produce síntomas de enfermedad.

- a) Determinar el test de razón de verosimilitudes para contrastar  $H_0 : p_1 = p_2$  frente a  $H_1 : p_1 \neq p_2$ .
- b) Determinar un intervalo de confianza aproximado para  $p_1 - p_2$  y deducir de ello un test para las mismas hipótesis.

Aplicar los resultados al caso  $n = 60$ ,  $x = 18$  y  $m = 100$ ,  $y = 42$ .

**Ejercicio 9.9**

Con objeto de estudiar la efectividad de un agente diurético, se eligieron al azar 11 pacientes, aplicando a 6 de ellos dicho fármaco y un placebo a los 5 restantes. La variable observada en esta experiencia fue la concentración de sodio en la orina a las 24 horas, la cual dió los resultados siguientes:

Diurético :	20'4	62'5	61'3	44'2	11'1	23'7
Placebo :	1'2	6'9	38'7	20'4	17'2	

Supuesto que las concentraciones de sodio, en ambos casos, tienen distribución normal:  $N(\mu_1, \sigma_1)$  y  $N(\mu_2, \sigma_2)$  respectivamente, contrastar a un nivel de significación del 5 %, si existe diferencia, en el efecto medio, al usar el agente diurético.

**Ejercicio 9.10**

Con objeto de averiguar si la fuerza de gravedad hace disminuir significativamente la estatura de las personas a lo largo del día, se seleccionaron al azar 10 individuos —mujeres de 25 años—, de las que se midió su estatura (en cm.) al levantarse,  $X_i$ , y antes de acostarse,  $Y_i$ , obteniéndose los siguientes datos:

$$X_i : \quad 169'7 \quad 168'5 \quad 165'9 \quad 177'8 \quad 179'6 \quad 168'9 \quad 169'2 \quad 167'9 \quad 181'8 \quad 163'3$$

---

$$Y_i : \quad 168'2 \quad 166'4 \quad 166'7 \quad 177'2 \quad 177'9 \quad 168'0 \quad 169'5 \quad 166'7 \quad 182'5 \quad 161'1$$

¿Puede inferirse, a nivel  $\alpha = 0'05$ , que la estatura disminuye a lo largo de la jornada?

## Solución de los ejercicios

### Ejercicio 9.1

a) La función de verosimilitud

$$f_\theta(x_1, \dots, x_n) = \theta^{-n} e^{-n\bar{x}/\theta}$$

(o su logaritmo) es creciente para  $\theta \in (0, \bar{x})$  y decreciente en  $(\bar{x}, \infty)$ . El estimador de máxima verosimilitud para  $\theta$  es pues  $\bar{x}$  y

$$\sup_{\Theta} f_\theta(x_1, \dots, x_n) = \bar{x}^{-n} e^{-n}.$$

Por otro lado, en  $\Theta_0 = (0, \theta_0]$  el supremo será el mismo si  $\bar{x} \leq \theta_0$ ; si no, se alcanzará en  $\theta_0$ . Por tanto,

$$\Lambda(x_1, \dots, x_n) = \begin{cases} 1 & \text{si } \bar{x} \leq \theta_0, \\ \bar{x}^n \theta_0^{-n} e^{n(1-\bar{x}/\theta_0)} & \text{si } \bar{x} > \theta_0. \end{cases}$$

Como  $\Lambda(x_1, \dots, x_n)$  es función decreciente de  $\bar{x}$  cuando  $\bar{x} > \theta_0$ , la desigualdad  $\Lambda(x_1, \dots, x_n) < c$  equivale a  $\bar{x} > k'$ , o bien  $n\bar{x} > k$ ; luego el test de razón de verosimilitudes tiene por región crítica

$$C = \{n\bar{x} > k\}.$$

Pero  $n\bar{X}$  tiene distribución  $\gamma(n, 1/\theta)$ , es decir que  $2n\bar{X}/\theta$  tiene distribución  $\chi_{2n}^2$ ; para conseguir el test de tamaño  $\alpha$  tendrá que ser entonces

$$\alpha = P_{\theta_0} \{ n\bar{X} > k \} = P \{ \chi_{2n}^2 > 2k/\theta_0 \},$$

de donde  $k = \theta_0 \chi_{2n;\alpha}^2 / 2$ .

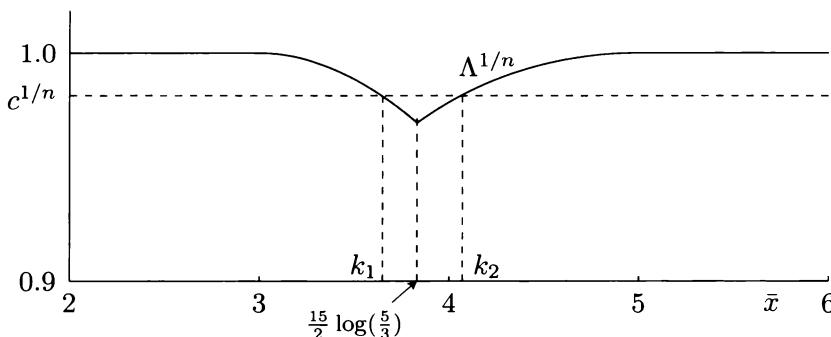
Según el teorema de Karlin-Rubin, los tests obtenidos, de región crítica  $C = \{n\bar{x} > k\}$  son uniformemente de máxima potencia para las hipótesis consideradas en (a). En cambio, para las hipótesis consideradas en (b) seguramente no existe test uniformemente de máxima potencia; sin embargo, el método de la razón de verosimilitudes ofrece la siguiente solución.

(b) De acuerdo con la forma de la función  $f_\theta(x_1, \dots, x_n)$ , si  $\bar{x} \notin (3, 5)$ , el  $\sup_{\theta \notin (3, 5)} f_\theta$  se alcanza en  $\bar{x}$  y vale  $\bar{x}^{-n} e^{-n}$ ; en cambio, si  $\bar{x} \in (3, 5)$ , el máximo será el mayor de los valores  $3^{-n} e^{-n\bar{x}/3}$  y  $5^{-n} e^{-n\bar{x}/5}$ . Pero el primer

valor es mayor si  $3 < \bar{x} < 15/2 \log(5/3) \simeq 3'831$  y menor si  $15/2 \log(5/3) < \bar{x} < 5$ . La razón de verosimilitudes es pues

$$\Lambda(x_1, \dots, x_n) = \begin{cases} 1 & \text{si } \bar{x} \notin (3, 5) \\ 3^{-n} e^{-n\bar{x}/3} \bar{x}^n e^n & \text{si } 3 < \bar{x} \leq \frac{15}{2} \log \frac{5}{3} \\ 5^{-n} e^{-n\bar{x}/5} \bar{x}^n e^n & \text{si } \frac{15}{2} \log \frac{5}{3} \leq \bar{x} < 5 \end{cases}$$

y la representación gráfica de  $\Lambda^{1/n}$  es de la forma



La región crítica  $C = \{\Lambda < c\}$  (con  $0 < c < 1$ ) del test de razón de verosimilitudes corresponde a un intervalo

$$C = \{k_1 < \bar{x} < k_2\},$$

donde debe ser  $3^{-1}k_1 e^{-k_1/3} = 5^{-1}k_2 e^{-k_2/5}$ . Numéricamente no es difícil encontrar la solución  $k_2$  asociada a cada valor de  $k_1$ ; por ejemplo:

$k_1 :$	3'1	3'2	3'3	3'4	3'5	3'6	3'7	3'8
$k_2 :$	4'835	4'681	4'531	4'388	4'250	4'118	3'991	3'868

Si el tamaño muestral está fijo, entre las regiones críticas de este tipo hay que seleccionar la adecuada a la restricción sobre el tamaño. Concretamente, como  $2nX/\theta$  tiene distribución  $\chi_{2n}^2$ ,

$$\begin{aligned} \mathcal{P}_\theta\{k_1 < \bar{X} < k_2\} &= P\{2nk_1/\theta < \chi_{2n}^2 < 2nk_2/\theta\} \\ &= \frac{1}{(n-1)!} \int_{nk_1/\theta}^{nk_2/\theta} y^{n-1} e^{-y} dy \end{aligned}$$

tiene que valer a lo sumo  $\alpha$  para  $\theta \notin (3, 5)$ . La derivada respecto a  $\theta$  de  $P_\theta\{k_1 < \bar{X} < k_2\}$  (prescindiendo de la constante  $1/(n-1)!$ ) es

$$\frac{(nk_1)^n}{\theta^{n+1}} e^{-nk_1/\theta} - \frac{(nk_2)^n}{\theta^{n+1}} e^{-nk_2/\theta},$$

que es positiva para  $\theta < (k_2 - k_1)/(\log k_2 - \log k_1)$  y negativa en caso contrario. Dado que siempre es  $k_1 < (k_2 - k_1)/(\log k_2 - \log k_1) < k_2$ , para  $\theta \notin (3, 5)$ , la máxima probabilidad se alcanza en  $\theta = 3$  ó  $\theta = 5$ . El test de tamaño  $\alpha$  se consigue pues seleccionando  $k_1$  (y el valor asociado de  $k_2$ ) para que sea  $\alpha$  la mayor de las probabilidades

$$P\{2nk_1/3 < \chi^2_{2n} < 2nk_2/3\} \quad \text{y} \quad P\{2nk_1/5 < \chi^2_{2n} < 2nk_2/5\}.$$

Ahora bien, para no obtener una región crítica muy pequeña (y un test, por consiguiente, uniformemente de muy poca potencia), se puede fijar la región crítica,  $\{3'2 < \bar{x} < 4'681\}$  por ejemplo, e incrementar el tamaño muestral hasta que dicha región tenga tamaño  $\alpha$ . Como el tamaño muestral tendrá que ser grande, puede utilizarse que  $Z = \sqrt{n}(\bar{X} - \theta)/\theta$  tiene aproximadamente distribución  $N(0, 1)$ ; de forma que

$$\begin{aligned} \mathcal{P}_3\{3'2 < \bar{X} < 4'681\} &= P\left\{\frac{0'2}{3}\sqrt{n} < Z < \frac{1'681}{3}\sqrt{n}\right\}, \\ \mathcal{P}_5\{3'2 < \bar{X} < 4'681\} &= P\left\{-\frac{1'8}{5}\sqrt{n} < Z < -\frac{0'319}{5}\sqrt{n}\right\} \end{aligned}$$

y debe ser  $\sqrt{n} 0'2/3 \geq z_\alpha$  y  $-\sqrt{n} 0'319/5 \leq -z_\alpha$ ; es decir  $n \geq 245'67 z_\alpha^2$ .

Como era intuitivamente de esperar, el contraste considerado exige un tamaño muestral elevado para que pueda dar un resultado aceptable; afortunadamente no es normal, en la práctica, tener que considerar este tipo de contrastes.

## Ejercicio 9.2

La función de verosimilitud:

$$f_\theta(x_1, \dots, x_n) = (1 - \theta)^n \theta^{x_1 + \dots + x_n} = (1 - \theta)^n \theta^{n\bar{x}}$$

o su logaritmo, son crecientes para  $\theta < \bar{x}/(1 + \bar{x})$  y decrecientes en caso contrario; de forma que  $\hat{\theta} = \bar{x}/(1 + \bar{x})$  es el estimador de máxima verosimilitud de  $\theta$ . Por tanto

$$\sup_{\theta \in (0,1)} f_\theta = \left(\frac{1}{1 + \bar{x}}\right)^n \left(\frac{\bar{x}}{1 + \bar{x}}\right)^{n\bar{x}},$$

mientras que

$$\sup_{\theta \leq \theta_0} f_\theta = \begin{cases} \left(\frac{1}{1+\bar{x}}\right)^n \left(\frac{\bar{x}}{1+\bar{x}}\right)^{n\bar{x}} & \text{si } \hat{\theta} \leq \theta_0, \\ (1-\theta_0)^n \theta_0^{n\bar{x}} & \text{si } \hat{\theta} > \theta_0. \end{cases}$$

La razón de verosimilitudes es entonces

$$\Lambda = \begin{cases} 1 & \text{si } \hat{\theta} \leq \theta_0, \\ \frac{(1-\theta_0)^n \theta_0^{n\bar{x}} (1+\bar{x})^{n(1+\bar{x})}}{\bar{x}^{n\bar{x}}} & \text{si } \hat{\theta} > \theta_0. \end{cases}$$

El logaritmo de la segunda expresión es

$$n \log(1 - \theta_0) + n\bar{x} \log \theta_0 + n(1 + \bar{x}) \log(1 + \bar{x}) - n\bar{x} \log \bar{x},$$

cuya derivada respecto a  $\bar{x}$ :

$$n (\log \theta_0 + \log(1 + \bar{x}) - \log \bar{x})$$

es negativa si  $\hat{\theta} > \theta_0$ , de forma que  $\Lambda$  es decreciente con  $\bar{x}$ . Por consiguiente, la función crítica

$$\varphi = \begin{cases} 1 & \text{si } \Lambda < c \\ \gamma & \text{si } \Lambda = c \\ 0 & \text{si } \Lambda > c \end{cases} \quad \text{es equivalente a} \quad \varphi = \begin{cases} 1 & \text{si } \bar{x} > k \\ \gamma & \text{si } \bar{x} = k \\ 0 & \text{si } \bar{x} < k \end{cases}$$

con  $k > \theta_0$ . Para que el test tenga tamaño  $\alpha$ ,  $k$  y  $\gamma$  deben ser tales que

$$\mathcal{P}_{\theta_0}\{n\bar{x} > nk\} + \gamma \mathcal{P}_{\theta_0}\{n\bar{x} = nk\} = \alpha,$$

donde  $n\bar{x}$ , como suma de variables independientes con distribución geométrica, tiene distribución binomial negativa de parámetros  $n$  y  $\theta_0$ ; es decir

$$\mathcal{P}_{\theta_0}\{n\bar{x} = r\} = \binom{n+r-1}{r} (1-\theta_0)^n \theta_0^r \quad \text{para } r = 0, 1, 2, \dots$$

Por ejemplo, si  $n = 30$ ,  $\theta_0 = 1/2$  y  $\alpha = 0'05$ , se puede calcular

$$\mathcal{P}_{1/2}\{n\bar{x} > 44\} = 0'0402 \quad \text{y} \quad \mathcal{P}_{1/2}\{n\bar{x} > 43\} = 0'0503,$$

luego debe ser  $k = 44/30$  y  $\gamma = 0'9703$ .

La función de verosimilitud  $(1 - \theta)^n e^{n\bar{x} \log \theta}$  tiene razón de verosimilitud monótona en  $\bar{x}$ ; luego, según el teorema de Karlin-Rubin, la función crítica anterior es la de un test uniformemente de máxima potencia para contrastar  $H_0 : \theta \leq \theta_0$  frente a  $H_1 : \theta > \theta_0$ .

### Ejercicio 9.3

La función de verosimilitud de una muestra de tamaño  $n$  es

$$f_{\theta_1, \theta_2}(x_1, \dots, x_n) = \frac{1}{(\theta_2 - \theta_1)^n} I_{\{x_{(1)} \geq \theta_1\}} I_{\{x_{(n)} \leq \theta_2\}},$$

que sólo es no nula para  $\theta_1 \leq x_{(1)}$  y  $\theta_2 \geq x_{(n)}$ , y vale entonces  $1/(\theta_2 - \theta_1)^n$ . El máximo se alcanza, pues, en  $\hat{\theta}_1 = x_{(1)}$ ,  $\hat{\theta}_2 = x_{(n)}$ , donde vale

$$f_{\hat{\theta}_1, \hat{\theta}_2} = (x_{(n)} - x_{(1)})^{-n}.$$

Bajo la restricción  $\theta_1 \leq 0 \leq \theta_2$ , puede ocurrir:

- ▷ que sea  $x_{(1)} \leq 0 \leq x_{(n)}$ , en cuyo caso el máximo de  $f_{\theta_1, \theta_2}$  sigue alcanzándose en los mismos valores y con el mismo resultado;
- ▷ que sea  $0 < x_{(1)} \leq x_{(n)}$ ; el máximo se alcanza entonces para  $\theta_1 = 0$ ,  $\theta_2 = x_{(n)}$  y vale  $x_{(n)}^{-n}$ ;
- ▷ que sea  $x_{(1)} \leq x_{(n)} < 0$ , lo cual da  $(-x_{(1)})^{-n}$  como valor del máximo, para  $\theta_1 = x_{(1)}$  y  $\theta_2 = 0$ .

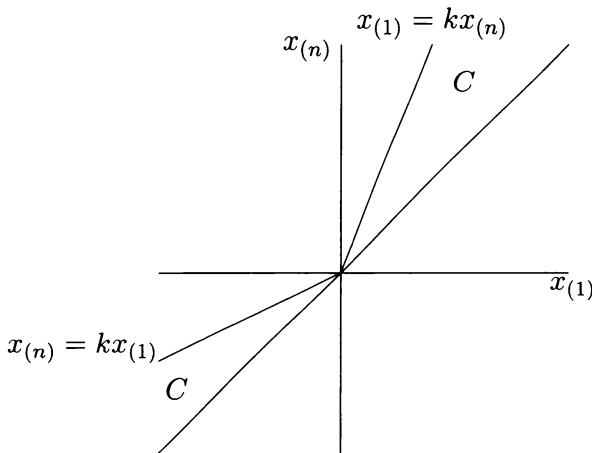
La razón de verosimilitudes es por tanto

$$\Lambda = \begin{cases} 1 & \text{si } x_{(1)} \leq 0 \leq x_{(n)} \\ \left(\frac{x_{(n)} - x_{(1)}}{x_{(n)}}\right)^n = \left(1 - \frac{x_{(1)}}{x_{(n)}}\right)^n & \text{si } x_{(1)} > 0 \\ \left(\frac{x_{(n)} - x_{(1)}}{-x_{(1)}}\right)^n = \left(1 - \frac{x_{(n)}}{x_{(1)}}\right)^n & \text{si } x_{(n)} < 0. \end{cases}$$

La región crítica  $\{\Lambda < c\}$  es entonces

$$\{x_{(1)} > kx_{(n)} > 0\} \cup \{x_{(n)} < kx_{(1)} < 0\}$$

(siendo  $k = 1 - c^{1/n}$ ), tal y como muestra la figura siguiente:



Para que el test tenga tamaño  $\alpha$ , habrá de ser

$$\alpha = \sup_{\theta_1 \leq 0 \leq \theta_2} [\mathcal{P}_{\theta_1, \theta_2} \{X_{(1)} > kX_{(n)} > 0\} + \mathcal{P}_{\theta_1, \theta_2} \{X_{(n)} < kX_{(1)} < 0\}] .$$

Ahora bien, la densidad conjunta de  $X_{(1)}$  y  $X_{(n)}$  es

$$\frac{n(n-1)}{(\theta_2 - \theta_1)^n} (y-x)^{n-2} \quad \text{para } \theta_1 \leq x \leq y \leq \theta_2.$$

Luego la probabilidad a maximizar vale

$$\begin{aligned} & \frac{n(n-1)}{(\theta_2 - \theta_1)^n} \left( \int_0^{\theta_2} \int_{ky}^y (y-x)^{n-2} dx dy + \int_{\theta_1}^0 \int_x^{kx} (y-x)^{n-2} dy dx \right) = \\ &= \frac{n}{(\theta_2 - \theta_1)^n} \left( \int_0^{\theta_2} (y - ky)^{n-1} dy + \int_{\theta_1}^0 (kx - x)^{n-1} dx \right) \\ &= (1-k)^{n-1} \frac{\theta_2^n + (-\theta_1)^n}{(\theta_2 - \theta_1)^n}. \end{aligned}$$

El segundo factor alcanza su valor máximo 1, cuando uno sólo de los parámetros vale 0; luego la condición requerida es que  $(1-k)^{n-1} = \alpha$  o bien  $k = 1 - \alpha^{1/(n-1)}$ .

#### Ejercicio 9.4

La función de verosimilitud de la muestra es

$$\begin{aligned} L &= (2\pi\sigma_1\sigma_2\sqrt{1-\rho^2})^{-n} \exp \left\{ -\frac{1}{2(1-\rho^2)} \cdot \right. \\ &\quad \cdot \left. \left[ \frac{1}{\sigma_1^2} \sum_{i=1}^n (x_i - \mu_1)^2 - \frac{2\rho}{\sigma_1\sigma_2} \sum_{i=1}^n (x_i - \mu_1)(y_i - \mu_2) + \frac{1}{\sigma_2^2} \sum_{i=1}^n (y_i - \mu_2)^2 \right] \right\} \end{aligned}$$

de manera que

$$\log L = -n \log 2\pi - \frac{n}{2} \log \sigma_1^2 - \frac{n}{2} \log \sigma_2^2 - \frac{n}{2} \log(1 - \rho^2) - \frac{1}{2(1 - \rho^2)} \cdot \left[ \frac{1}{\sigma_1^2} \sum_{i=1}^n (x_i - \mu_1)^2 - \frac{2\rho}{\sigma_1 \sigma_2} \sum_{i=1}^n (x_i - \mu_1)(y_i - \mu_2) + \frac{1}{\sigma_2^2} \sum_{i=1}^n (y_i - \mu_2)^2 \right].$$

Para hallar el máximo, hay que imponer en primer lugar

$$\begin{aligned} 0 &= \frac{\partial \log L}{\partial \mu_1} = \frac{1}{\sigma_1^2(1 - \rho^2)} \sum_{i=1}^n (x_i - \mu_1), \\ 0 &= \frac{\partial \log L}{\partial \mu_2} = \frac{1}{\sigma_2^2(1 - \rho^2)} \sum_{i=1}^n (y_i - \mu_2), \end{aligned}$$

de donde se obtiene:  $\hat{\mu}_1 = \bar{x}$  y  $\hat{\mu}_2 = \bar{y}$ . Ello permite expresar  $\log L$  en la forma

$$\begin{aligned} \log L &= -n \log 2\pi - \frac{n}{2} \log \sigma_1^2 - \frac{n}{2} \log \sigma_2^2 - \frac{n}{2} \log(1 - \rho^2) \\ &\quad - \frac{ns_1^2}{2\sigma_1^2(1 - \rho^2)} - \frac{ns_2^2}{2\sigma_2^2(1 - \rho^2)} + \frac{n\rho s_{11}}{\sigma_1 \sigma_2(1 - \rho^2)}. \end{aligned}$$

Ha de cumplirse además

$$\begin{aligned} 0 &= \frac{\partial \log L}{\partial \sigma_1^2} = -\frac{n}{2\sigma_1^2} + \frac{ns_1^2}{2\sigma_1^4(1 - \rho^2)} - \frac{n\rho s_{11}}{2\sigma_1^3 \sigma_2(1 - \rho^2)}, \\ 0 &= \frac{\partial \log L}{\partial \sigma_2^2} = -\frac{n}{2\sigma_2^2} + \frac{ns_2^2}{2\sigma_2^4(1 - \rho^2)} - \frac{n\rho s_{11}}{2\sigma_2^3 \sigma_1(1 - \rho^2)}, \end{aligned}$$

de donde se deduce que debe ser

$$\begin{aligned} (1 - \rho^2)\sigma_1^2 \sigma_2^2 + \rho s_{11} \sigma_1 \sigma_2 - s_1^2 \sigma_2^2 &= 0, \\ (1 - \rho^2)\sigma_1^2 \sigma_2^2 + \rho s_{11} \sigma_1 \sigma_2 - s_2^2 \sigma_1^2 &= 0. \end{aligned}$$

Por tanto,

$$\frac{\hat{\sigma}_1^2}{s_1^2} = \frac{\hat{\sigma}_2^2}{s_2^2} = \lambda,$$

con lo que las dos últimas ecuaciones se reducen a

$$(1 - \rho^2)\lambda s_1^2 s_2^2 + \rho s_{11} s_1 s_2 - s_1^2 s_2^2 = 0. \quad [A]$$

Por último, hay que imponer la condición

$$0 = \frac{\partial \log L}{\partial \rho} = \frac{n\rho}{1-\rho^2} - \frac{ns_1^2\rho}{\sigma_1^2(1-\rho^2)^2} - \frac{ns_2^2\rho}{\sigma_2^2(1-\rho^2)^2} + \frac{ns_{11}(1+\rho^2)}{2\sigma_1\sigma_2(1-\rho^2)^2}$$

de la que se obtiene

$$(1-\rho^2)\rho\sigma_1^2\sigma_2^2 - \rho s_1^2\sigma_2^2 - \rho s_2^2\sigma_1^2 + (1+\rho^2)s_{11}\sigma_1\sigma_2 = 0,$$

o lo que es lo mismo,

$$(1-\rho^2)\lambda s_1^2 s_2^2 - 2s_1^2 s_2^2 + \frac{1+\rho^2}{\rho} s_{11} s_1 s_2 = 0. \quad [B]$$

Restando [B] de [A], resulta:  $\hat{\rho} = \frac{s_{11}}{s_1 s_2}$ . Este resultado, reemplazado en [A], proporciona  $\lambda = 1$ , con lo cual  $\hat{\sigma}_1^2 = s_1^2$ , y  $\hat{\sigma}_2^2 = s_2^2$ .

Sustituidos en la función de verosimilitud,  $L$ , los estimadores de máxima verosimilitud obtenidos, se obtiene

$$\sup L = \left( 2\pi \sqrt{s_1^2 s_2^2 - s_{11}^2} \right)^{-n} e^{-n}.$$

Por otra parte, si  $\rho = 0$  es

$$L = (2\pi\sigma_1\sigma_2)^{-n} \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^n (x_i - \mu_1)^2 - \frac{1}{2\sigma_2^2} \sum_{i=1}^n (y_i - \mu_2)^2 \right\},$$

cuyo logaritmo alcanza su máximo en  $\hat{\mu}_1 = \bar{x}$ ,  $\hat{\mu}_2 = \bar{y}$ ,  $\hat{\sigma}_1^2 = s_1^2$ ,  $\hat{\sigma}_2^2 = s_2^2$ ; de manera que

$$\sup_{\rho=0} L = (2\pi s_1 s_2)^{-n} e^{-n}.$$

La razón de verosimilitudes es entonces

$$\Lambda = \left( \frac{\sqrt{s_1^2 s_2^2 - s_{11}^2}}{s_1 s_2} \right)^n = (1-r^2)^{n/2},$$

siendo  $r$  el coeficiente de correlación muestral. El test de razón de verosimilitudes tiene, por tanto, región crítica

$$C = \{(1-r^2)^{n/2} < c\} = \{|r| > k\} = \{\sqrt{n-2} \frac{|r|}{\sqrt{1-r^2}} > K\}.$$

La última expresión es la más adecuada puesto que, según el teorema 3.2, bajo la hipótesis nula  $\rho = 0$ ,  $\sqrt{n-2}|r|/\sqrt{1-r^2}$  tiene distribución  $t$  de Student con  $n - 2$  grados de libertad. Luego debe ser  $K = t_{n-2;\alpha/2}$  para que el test tenga nivel de significación  $\alpha$ .

### Ejercicio 9.5

Puesto que las varianzas de ambas variables son iguales, la función de verosimilitud será

$$L = \left(2\pi\sigma^2\sqrt{1-\rho^2}\right)^{-n} \exp\left\{-\frac{1}{2\sigma^2(1-\rho^2)} \cdot \left[ \sum_{i=1}^n (x_i - \mu_1)^2 + \sum_{i=1}^n (y_i - \mu_2)^2 - 2\rho \sum_{i=1}^n (x_i - \mu_1)(y_i - \mu_2) \right] \right\}.$$

Para determinar la razón de verosimilitudes hay que obtener  $\sup_{\Theta} L$  y  $\sup_{\Theta_0} L$ , siendo

$$\Theta = \{(\mu_1, \mu_2, \sigma^2, \rho) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^+ \times [-1, 1]\}$$

y  $\Theta_0$  el subconjunto  $\{\rho = 0\}$  de  $\Theta$ . Para determinar el primero, hay que imponer las condiciones:

$$0 = \frac{\partial \log L}{\partial \mu_1} = \frac{1}{\sigma^2(1-\rho^2)} \sum_{i=1}^n (x_i - \mu_1) \quad \text{de donde} \quad \hat{\mu}_1 = \bar{x},$$

$$0 = \frac{\partial \log L}{\partial \mu_2} = \frac{1}{\sigma^2(1-\rho^2)} \sum_{i=1}^n (y_i - \mu_2) \quad \text{de donde} \quad \hat{\mu}_2 = \bar{y}.$$

Y además, una vez reemplazadas las soluciones anteriores,

$$0 = \frac{\partial \log L}{\partial \sigma^2} = -\frac{n}{\sigma^2} + \frac{n(s_1^2 + s_2^2 - 2\rho s_{11})}{2\sigma^4(1-\rho^2)}$$

de donde se obtiene  $2\sigma^2(1-\rho^2) = s_1^2 + s_2^2 - 2\rho s_{11}$ . Por último

$$0 = \frac{\partial \log L}{\partial \rho} = \frac{n\rho}{1-\rho^2} - \frac{n\rho(s_1^2 + s_2^2 - 2\rho s_{11})}{\sigma^2(1-\rho^2)^2} + \frac{ns_{11}}{\sigma^2(1-\rho^2)}$$

que, junto con la ecuación anterior, da  $\sigma^2\rho = s_{11}$ . Por tanto,

$$\hat{\sigma}^2 = \frac{s_1^2 + s_2^2}{2} \quad \text{y} \quad \hat{\rho} = \frac{2s_{11}}{s_1^2 + s_2^2}.$$

Reemplazando en  $L$  las soluciones halladas, resulta

$$\sup_{\Theta} L = \left( \pi \sqrt{(s_1^2 + s_2^2)^2 - 4s_{11}^2} \right)^{-n} e^{-n}.$$

Por otra parte, si  $\rho = 0$ , entonces

$$L = (2\pi\sigma^2)^{-n} \exp \left\{ -\frac{1}{2\sigma^2} \left[ \sum_{i=1}^n (x_i - \mu_1)^2 + \sum_{i=1}^n (y_i - \mu_2)^2 \right] \right\},$$

cuyo logaritmo alcanza su máximo en  $\hat{\mu}_1 = \bar{x}$ ,  $\hat{\mu}_2 = \bar{y}$ , y  $\hat{\sigma}^2 = (s_1^2 + s_2^2)/2$ . Con lo cual

$$\sup_{\Theta_0} L = (\pi(s_1^2 + s_2^2))^{-n} e^{-n}.$$

En definitiva la razón de verosimilitudes es

$$\Lambda = \left( \sqrt{1 - \frac{4s_{11}^2}{(s_1^2 + s_2^2)^2}} \right)^n = \left( \sqrt{1 - \frac{4S_{11}^2}{(S_1^2 + S_2^2)^2}} \right)^n$$

y la región crítica del contraste  $H_0 : \rho = 0$  frente a  $H_1 : \rho \neq 0$  será

$$C = \{\Lambda < c\} = \left\{ \left| \frac{2S_{11}}{S_1^2 + S_2^2} \right| > k \right\},$$

donde  $c$  ha de determinarse mediante la condición  $P_{\rho=0}(C) = \alpha$ , lo cual exige conocer la distribución del estadístico  $R = 2S_{11}/(S_1^2 + S_2^2)$  bajo la hipótesis de que  $\rho = 0$  (y supuesto que las varianzas de ambas variables son iguales).

Incluso, para poder determinar la potencia del test para valores de  $\rho \neq 0$ , interesa conocer la distribución del estadístico, aún cuando no sea  $\rho = 0$ . Para ello sean  $U = (X + Y)/2$  y  $V = (X - Y)/2$ , es decir

$$\begin{pmatrix} U \\ V \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}$$

que son variables con distribución normal de medias  $(\mu_1 + \mu_2)/2$  y  $(\mu_1 - \mu_2)/2$  respectivamente y matriz de covarianzas

$$\frac{1}{4} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \frac{\sigma^2}{2} \begin{pmatrix} 1 + \rho & 0 \\ 0 & 1 - \rho \end{pmatrix};$$

de manera que  $U$  y  $V$  son independientes. Ahora bien, si  $\tilde{S}_1^2$ ,  $\tilde{S}_2^2$  son las cuasivarianzas muestrales y  $\tilde{S}_{11}$  la cuasicovarianza muestral de la muestra transformada  $(u_i, v_i)$ ,

$$\begin{aligned} S_{11} &= \frac{1}{n-1} \sum_{i=1}^n (u_i + v_i - \bar{u} - \bar{v})(u_i - v_i - \bar{u} + \bar{v}) \\ &= \frac{1}{n-1} \sum_{i=1}^n [(u_i - \bar{u})^2 - (v_i - \bar{v})^2] = \tilde{S}_1^2 - \tilde{S}_2^2, \\ S_1^2 &= \frac{1}{n-1} \sum_{i=1}^n (u_i + v_i - \bar{u} - \bar{v})^2 = \tilde{S}_1^2 + \tilde{S}_2^2 + 2\tilde{S}_{11}, \\ S_2^2 &= \tilde{S}_1^2 + \tilde{S}_2^2 - 2\tilde{S}_{11}. \end{aligned}$$

Luego

$$R = \frac{\tilde{S}_1^2 - \tilde{S}_2^2}{\tilde{S}_1^2 + \tilde{S}_2^2} = 1 - \frac{2}{1 + \tilde{S}_1^2/\tilde{S}_2^2} = 1 - \frac{2}{1 + T(1 + \rho)/(1 - \rho)},$$

siendo

$$T = \frac{\tilde{S}_1^2/(1 + \rho)}{\tilde{S}_2^2/(1 - \rho)}$$

que tiene distribución  $F_{n-1, n-1}$ , puesto que el numerador y el denominador (divididos por  $\sigma^2/2$ ) tienen distribución  $\chi^2_{n-1}$ .

Bajo la hipótesis  $\rho = 0$ , la región crítica  $C = \{|R| > k\}$  se expresa

$$C = \left\{ T < \frac{1-k}{1+k} \right\} \cup \left\{ T > \frac{1+k}{1-k} \right\};$$

de tal modo que, para que  $\mathcal{P}_{\rho=0}(C) = \alpha$ , debe ser

$$\frac{1+k}{1-k} = F_{n-1, n-1; \alpha/2};$$

lo cual es equivalente a  $(1-k)/(1+k) = F_{n-1, n-1; 1-\alpha/2}$  y proporciona

$$k = \frac{F_{n-1, n-1; \alpha/2} - 1}{F_{n-1, n-1; \alpha/2} + 1}.$$

La potencia del contraste en  $\rho = \rho_1 \neq 0$  se expresa entonces

$$\mathcal{P}_{\rho=\rho_1}\{|R| > k\} = P\left\{\frac{1+\rho_1}{1-\rho_1} T < \frac{1-k}{1+k}\right\} + P\left\{\frac{1+\rho_1}{1-\rho_1} T > \frac{1+k}{1-k}\right\}$$

$$= P \left\{ T < \frac{1 - \rho_1}{1 + \rho_1} F_{n-1, n-1; 1-\alpha/2} \right\} + P \left\{ T > \frac{1 - \rho_1}{1 + \rho_1} F_{n-1, n-1; \alpha/2} \right\}.$$

### Ejercicio 9.6

Como la distribución condicionada  $(X_i | X_{i-1} = x_{i-1})$  es  $N(\theta x_{i-1}, \sigma)$ , la función de verosimilitud de las observaciones —no independientes—  $x_1, \dots, x_n$  es,

$$f_\theta(x_1, \dots, x_n) = \left( \frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta x_{i-1})^2 \right\}$$

siendo  $x_0 = 0$ .

Si  $\sigma^2$  es conocido, el máximo del logaritmo de la verosimilitud se alcanza cuando

$$0 = \frac{\partial}{\partial \theta} f_\theta = -\frac{1}{2\sigma^2} \sum_{i=1}^n 2(x_i - \theta x_{i-1}) x_{i-1};$$

es decir en

$$\hat{\theta} = \frac{T_1}{T_0} \quad \text{siendo} \quad T_1 = \sum_{i=2}^n x_i x_{i-1} \quad \text{y} \quad T_0 = \sum_{i=2}^n x_{i-1}^2.$$

Lo cual proporciona como valor del máximo

$$\begin{aligned} f_{\hat{\theta}} &= \left( \frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left( x_i - x_{i-1} \frac{T_1}{T_0} \right)^2 \right\} \\ &= \left( \frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left\{ -\frac{1}{2\sigma^2} \left( \sum_{i=1}^n x_i^2 - \frac{T_1^2}{T_0} \right) \right\}. \end{aligned}$$

Mientras que, bajo  $H_0$ , la verosimilitud no depende de ningún parámetro, sino que es

$$f_0(x_1, \dots, x_n) = \left( \frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 \right\}.$$

La razón de verosimilitudes será, entonces

$$\Lambda = \frac{f_0}{f_{\hat{\theta}}} = \exp \left\{ -\frac{1}{2\sigma^2} \frac{T_1^2}{T_0} \right\}$$

y, por tanto, la región crítica de la forma

$$C = \{\Lambda < c\} = \{-2 \log \Lambda > c'\} = \{T_1^2 > kT_0\}.$$

Si  $\sigma^2$  es desconocido, el máximo (de  $f_{\theta, \sigma^2}$ ) se alcanza cuando

$$\begin{cases} -\frac{1}{2\sigma^2} \sum_{i=1}^n 2(x_i - \theta x_{i-1})x_{i-1} = 0, \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \theta x_{i-1})^2 = 0. \end{cases}$$

Es decir, para

$$\hat{\theta} = \frac{T_1}{T_0} \quad \text{y} \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\theta} x_{i-1})^2,$$

que proporcionan como máximo de  $f_{\theta, \sigma^2}$ :

$$\begin{aligned} f_{\hat{\theta}, \hat{\sigma}^2} &= \left( \frac{2\pi}{n} \sum_{i=1}^n \left( x_i - x_{i-1} \frac{T_1}{T_0} \right)^2 \right)^{-n/2} e^{-n/2} \\ &= \left( \frac{2\pi}{n} \right)^{-n/2} \left( \sum_{i=1}^n x_i^2 - \frac{T_1^2}{T_0} \right)^{-n/2} e^{-n/2}. \end{aligned}$$

En cambio, si es  $\theta = 0$ , la verosimilitud se reduce a

$$f_{0, \sigma^2} = \left( \frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 \right\},$$

cuyo logaritmo alcanza su máximo para  $\tilde{\sigma}^2 = \sum_{i=1}^n x_i^2/n$ , siendo para este valor,

$$f_{0, \tilde{\sigma}} = \left( \frac{2\pi}{n} \sum_{i=1}^n x_i^2 \right)^{-n/2} e^{-n/2}.$$

Por tanto, la razón de verosimilitudes resulta

$$\Lambda = \left( \sum_{i=1}^n x_i^2 - \frac{T_1^2}{T_0} \middle/ \sum_{i=1}^n x_i^2 \right)^{n/2} = \left( 1 - \frac{T_1^2}{T_0(T_0 + x_n^2)} \right)^{n/2},$$

con lo que la región crítica del contraste resulta ser

$$C = \{\Lambda < c\} = \{-2 \log \Lambda > c'\} = \{T_1^2 > k T_0 (T_0 + x_n^2)\}.$$

La última expresión de las regiones críticas muestra que, según sea  $\sigma^2$  conocido o desconocido, el contraste se basa en estadísticos diferentes; concretamente:

$$T_1^2/T_0 \quad \text{y} \quad T_1^2/(T_0(T_0 + X_n^2)).$$

Como no es posible calcular la distribución exacta de tales estadísticos, habrá que utilizar en ambos casos la distribución asintótica de  $-2 \log \Lambda$  (dando por supuesto que  $n$  es grande). Será

$$-2 \log \Lambda = \begin{cases} \frac{T_1}{\sigma^2 T_0} & \text{si } \sigma^2 \text{ es conocido,} \\ -n \log \left( 1 - \frac{T_1^2}{T_0(T_0 + X_n^2)} \right) & \text{si } \sigma^2 \text{ es desconocido.} \end{cases}$$

pero su distribución es, en ambos casos, aproximadamente  $\chi_1^2$ . Así que, el valor crítico  $c'$  para  $-2 \log \Lambda$  es  $\chi_{1;\alpha}^2$  para ambas situaciones.

### Ejercicio 9.7

Si  $x_1, \dots, x_n$  designan las proporciones observadas en la primera población e  $y_1, \dots, y_m$  las observaciones de la segunda población, la verosimilitud es

$$f_{\theta_1, \theta_2}(x_1, \dots, x_n, y_1, \dots, y_m) = \theta_1^n \left( \prod_{i=1}^n x_i \right)^{\theta_1-1} \theta_2^m \left( \prod_{i=1}^m y_i \right)^{\theta_2-1}$$

Su logaritmo:

$$n \log \theta_1 + (\theta_1 - 1) \sum_{i=1}^n \log x_i + m \log \theta_2 + (\theta_2 - 1) \sum_{i=1}^m \log y_i$$

se hace máximo para

$$\hat{\theta}_1 = \frac{n}{-\sum_{i=1}^n \log x_i} \quad \text{y} \quad \hat{\theta}_2 = \frac{m}{-\sum_{i=1}^m \log y_i};$$

lo cual proporciona a la verosimilitud el valor máximo

$$\sup f_{\theta_1, \theta_2} = \left( \frac{n}{T_1} \right)^n e^{T_1 - n} \left( \frac{m}{T_2} \right)^m e^{T_2 - m}$$

donde  $T_1 = -\sum_{i=1}^n \log x_i$  y  $T_2 = -\sum_{i=1}^m \log y_i$  son positivos.

Por otra parte, en la región  $\theta_1 \leq \theta_2$ , será

$$\sup_{\theta_1 \leq \theta_2} f_{\theta_1, \theta_2} = \sup f_{\theta_1, \theta_2} \quad \text{si } \hat{\theta}_1 \leq \hat{\theta}_2.$$

Pero, si  $\hat{\theta}_1 > \hat{\theta}_2$ , puesto que no hay ningún punto interior al recinto  $\theta_1 \leq \theta_2$  en el que se anulen las derivadas parciales (de  $\log f_{\theta_1, \theta_2}$ ) respecto a  $\theta_1$  y  $\theta_2$ , el máximo se alcanzará sobre la frontera y, más concretamente, sobre la recta  $\theta_1 = \theta_2 = \theta$ . En ella la verosimilitud se expresa

$$f_\theta(x_1, \dots, x_n, y_1, \dots, y_m) = \theta^{n+m} \left( \prod_{i=1}^n x_i \right)^{\theta-1} \left( \prod_{i=1}^m y_i \right)^{\theta-1},$$

cuyo logaritmo:

$$(n+m) \log \theta + (\theta - 1) \left( \sum_{i=1}^n \log x_i + \sum_{i=1}^m \log y_i \right)$$

se hace máximo en

$$\hat{\theta} = \frac{n+m}{T_1 + T_2},$$

para el cual la verosimilitud vale

$$\left( \frac{n+m}{T_1 + T_2} \right)^{n+m} e^{T_1 + T_2 - n - m}.$$

La razón de verosimilitudes es por tanto

$$\begin{aligned} \Lambda &= \begin{cases} 1 & \text{si } \hat{\theta}_1 \leq \hat{\theta}_2 \\ \left( \frac{n+m}{T_1 + T_2} \right)^{n+m} \left( \frac{T_1}{n} \right)^n \left( \frac{T_2}{m} \right)^m & \text{si } \hat{\theta}_1 > \hat{\theta}_2 \end{cases} \\ &= \begin{cases} 1 & \text{si } T_1/T_2 \geq n/m \\ \frac{(n+m)^{n+m}}{n^n m^m} \left( \frac{T_1}{T_1 + T_2} \right)^n \left( \frac{T_2}{T_1 + T_2} \right)^m & \text{si } T_1/T_2 < n/m. \end{cases} \end{aligned}$$

Con  $U = T_1/T_2$ , se puede expresar

$$\Lambda = \begin{cases} 1 & \text{si } U \geq n/m \\ \frac{(n+m)^{n+m}}{n^n m^m} \left( 1 + \frac{1}{U} \right)^{-n} (1+U)^{-m} & \text{si } U < n/m. \end{cases}$$

Es fácil ver que  $\Lambda$  es función creciente de  $U$ , luego la región crítica del test de razón de verosimilitudes es

$$C = \{\Lambda < c\} = \{U < k\} = \left\{ - \sum_{i=1}^n \log x_i < -k \sum_{i=1}^m \log y_i \right\}.$$

Nótese que,  $-\log X_i$  y  $-\log Y_i$  tienen distribución exponencial de parámetros  $\theta_1$  y  $\theta_2$  respectivamente. Luego la distribución de  $T_j$  es  $\gamma(n, \theta_j)$ ; o, dicho de otro modo,  $2\theta_1 T_1$  y  $2\theta_2 T_2$  tienen distribución  $\chi_{2n}^2$  y  $\chi_{2m}^2$  respectivamente. Por consiguiente

$$\frac{m\theta_1}{n\theta_2} U \quad \text{tiene distribución } F_{2n, 2m}.$$

Para que el test obtenido tenga tamaño  $\alpha$  debe ser

$$\alpha = \sup_{\theta_1 \leq \theta_2} \mathcal{P}\{U < k\} = \sup_{\theta_1 \leq \theta_2} P \left\{ F_{2n, 2m} < \frac{m\theta_1}{n\theta_2} k \right\}.$$

Desde luego, la última probabilidad es máxima (para  $\theta_1 \leq \theta_2$ ) cuando sea  $\theta_1 = \theta_2$ , luego debe ser

$$\frac{m}{n} k = F_{2n, 2m; 1-\alpha} \quad \text{o bien} \quad k = \frac{n}{m} F_{2n, 2m; 1-\alpha}.$$

### Ejercicio 9.8

a) Sea  $x_i = 1, 0$ , según que el  $i$ -ésimo individuo inoculado con el organismo I enferme o no, e  $y_j$  la misma variable para el individuo  $j$ -ésimo inoculado con el organismo II. La función de verosimilitud de las observaciones es

$$f_{p_1, p_2}(x_1, \dots, x_n, y_1, \dots, y_m) = p_1^x (1-p_1)^{n-x} p_2^y (1-p_2)^{m-y}$$

donde  $x = \sum_{i=1}^n x_i$  e  $y = \sum_{j=1}^m y_j$ .

Como los estimadores de máxima verosimilitud de  $p_1$  y  $p_2$  son  $\hat{p}_1 = x/n$  y  $\hat{p}_2 = y/m$ , es

$$\sup_{p_1, p_2 \in [0, 1]} f_{p_1, p_2} = \hat{p}_1^{n\hat{p}_1} (1-\hat{p}_1)^{n(1-\hat{p}_1)} \hat{p}_2^{m\hat{p}_2} (1-\hat{p}_2)^{m(1-\hat{p}_2)}.$$

Por otra parte, supuesto que las frecuencias  $p_1$  y  $p_2$  tienen el mismo valor  $p$ , la verosimilitud se expresa

$$f_p(x_1, \dots, x_n, y_1, \dots, y_m) = p^{x+y} (1-p)^{n+m-x-y}$$

que alcanza su máximo en  $\hat{p} = \frac{x+y}{n+m}$ .

Luego

$$\sup_{p_1=p_2} f_{p_1,p_2} = \hat{p}^{(n+m)\hat{p}}(1-\hat{p})^{(n+m)(1-\hat{p})}.$$

La razón de verosimilitudes es entonces

$$\Lambda(x,y) = \frac{\hat{p}^{(n+m)\hat{p}}(1-\hat{p})^{(n+m)(1-\hat{p})}}{\hat{p}_1^{n\hat{p}_1}(1-\hat{p}_1)^{n(1-\hat{p}_1)}\hat{p}_2^{n\hat{p}_2}(1-\hat{p}_2)^{n(1-\hat{p}_2)}}$$

con lo cual

$$-2 \log \Lambda = (n+m) [\hat{p} \log \hat{p} + (1-\hat{p}) \log (1-\hat{p})]$$

$$-n [\hat{p}_1 \log \hat{p}_1 + (1-\hat{p}_1) \log (1-\hat{p}_1)] - m [\hat{p}_2 \log \hat{p}_2 + (1-\hat{p}_2) \log (1-\hat{p}_2)]$$

tiene aproximadamente distribución  $\chi^2_1$  puesto que  $n$  y  $m$  son grandes y el modelo cuenta con dos parámetros, que se reducen a uno sólo si la hipótesis nula es cierta.

La región crítica del test de nivel de significación  $\alpha$  es por tanto

$$C = \{-2 \log \Lambda > \chi^2_{1,\alpha}\}.$$

Para  $n = 60$ ,  $m = 100$ ,  $x = 18$  e  $y = 42$  resulta  $\hat{p}_1 = 0'3$ ,  $\hat{p}_2 = 0'42$ ,  $\hat{p} = 0'375$  y  $-2 \log \Lambda = 2'338$ . Según la tabla de la  $\chi^2_1$ , el nivel crítico al que tal resultado permite rechazar la hipótesis  $p_1 = p_2$ , es  $0'145$ ; demasiado grande para que los datos se consideren significativos de que existe diferencia entre  $p_1$  y  $p_2$ .

b) Si se trata de construir un intervalo de confianza para  $p_1 - p_2$ , puesto que  $n$  y  $m$  son grandes, puede utilizarse que

$$\hat{p}_i \text{ es aproximadamente } N \left( p_i, \sqrt{p_i(1-p_i)/n} \right).$$

Por consiguiente

$$\hat{p}_1 - \hat{p}_2 \text{ es aproximadamente } N \left( p_1 - p_2, \sqrt{\frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{m}} \right).$$

La varianza puede estimarse mediante  $\bar{\sigma}^2 = \hat{p}_1(1-\hat{p}_1)/n + \hat{p}_2(1-\hat{p}_2)/m$ , de donde resulta para  $p_1 - p_2$ , con nivel de confianza  $1 - \alpha$ , el intervalo de extremos

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2} \bar{\sigma}.$$

Reinterpretado como contraste de la hipótesis  $H_0 : p_1 - p_2 = 0$ , ello lleva a rechazar la hipótesis nula cuando sea

$$|\hat{p}_1 - \hat{p}_2| > z_{\alpha/2} \bar{\sigma},$$

lo cual constituye la región crítica de un test de nivel de significación (aproximadamente) igual a  $\alpha$ .

Por ejemplo, con los datos concretos del enunciado, la estimación de la desviación típica es

$$\bar{\sigma} = \sqrt{\frac{0'3 0'7}{60} + \frac{0'42 0'58}{100}} = 0'077;$$

mientras que  $|\hat{p}_1 - \hat{p}_2| = 0'12$ . Como  $0'12/0'077 = 1'56 = z_{0'0594}$ , el nivel crítico para rechazar la hipótesis nula es  $0'1188$ ; no muy diferente del obtenido mediante la razón de verosimilitudes.

En la práctica se prefiere este método al de la razón de verosimilitudes pero introduciendo en él la siguiente modificación: El test obtenido tendrá región crítica

$$\{ |\hat{p}_1 - \hat{p}_2| > z_{\alpha/2} \tilde{\sigma} \},$$

siendo  $\tilde{\sigma}^2$  la estimación de la varianza de  $\hat{p}_1 - \hat{p}_2$ . Pero, si la hipótesis nula es cierta, la estimación de la varianza  $\tilde{\sigma}^2 = \hat{p}(1 - \hat{p})(1/n + 1/m)$ , será más precisa que  $\bar{\sigma}$ .

En el caso concreto considerado, resulta

$$\tilde{\sigma} = \sqrt{0'375 0'625(1/60 + 1/100)} = 0'079;$$

como  $0'12/0'079 = 1'52 = z_{0'0643}$ , la hipótesis de igualdad puede rechazarse con nivel crítico  $0'1286$ . El test resulta algo más conservador que el anterior, aunque no tanto como el de la razón de verosimilitudes.

Como comentario, cabe señalar que, según estos datos, una diferencia entre el 30 % y el 42 % no es significativa, cuando los tamaños muestrales son 60 y 100 respectivamente. Por ejemplo, si dos centros de enseñanza presentan 60 y 100 alumnos a las pruebas de selectividad, en el primero aprueban el 30 % y en el segundo el 42 %, no puede afirmarse que las diferencias no sean debidas al azar; puesto que, aún siendo así, se observarían diferencias de tal magnitud, con probabilidad en torno a 0'1286, es decir, casi 13 de cada cien veces. A menudo, los medios de comunicación hablan en porcentajes, dando a entender que se refieren al conjunto de la población: "el 42 % de las jóvenes fuma, mientras sólo lo hace el 30 % de los jóvenes". Uno se ve

tentado a pensar que es una gran diferencia. Pero, o se precisan los tamaños muestrales con los que se ha realizado la investigación, o la conclusión no significa prácticamente nada.

### Ejercicio 9.9

La relación existente entre intervalos de confianza y tests de hipótesis bilaterales, permite construir el contraste de la hipótesis  $H_0 : \mu_1 = \mu_2$  frente a  $H_1 : \mu_1 \neq \mu_2$ , mediante la determinación de un intervalo de confianza para la diferencia de medias  $\mu_1 - \mu_2$ , a nivel de confianza uno menos el nivel de significación; bastará después ver si el intervalo contiene o no al punto 0.

En este caso, la elección del intervalo de confianza para la diferencia de medias, requiere saber si las varianzas poblacionales son iguales o no; así que conviene primero efectuar el contraste de  $H_0 : \sigma_1 = \sigma_2$  frente a  $H_1 : \sigma_1 \neq \sigma_2$ . Ello puede llevarse a cabo, también, determinando un intervalo de confianza para el cociente de varianzas y viendo si contiene o no al valor 1.

Concretamente,

$$I = \left[ \frac{S_1^2/S_2^2}{F_{5,4;\alpha/2}}, \quad \frac{S_1^2/S_2^2}{F_{5,4;1-\alpha/2}} \right]$$

es un intervalo de confianza de nivel  $1 - \alpha$  para  $\sigma_1^2/\sigma_2^2$ ; que con los datos del enunciado ( $S_1^2 = 483'12$  y  $S_2^2 = 208'52$ ) y un nivel de confianza del 95 % da

$$I = \left[ \frac{483'12/208'52}{9'3645}, \quad \frac{483'12 \cdot 7'3879}{208'52} \right] = [ 0'247, 17'117 ],$$

puesto que

$$F_{5,4;0'025} = 9'3645 \quad \text{y} \quad F_{5,4;0'975} = 1/F_{4,5;0'025} = 1/7'3879.$$

Como el intervalo obtenido contiene al 1, las varianzas muestrales no indican que las varianzas poblacionales sean distintas.

Por consiguiente, el intervalo de confianza para la diferencia de medias poblacionales tendrá por extremos

$$\bar{x}_1 - \bar{x}_2 \mp t_{n_1+n_2-2;\alpha/2} \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}};$$

es decir, con nivel de confianza 0'95,

$$37'2 - 16'88 \mp 2'262 \sqrt{\frac{5 \cdot 483'12 + 4 \cdot 208'52}{9}} \sqrt{\frac{1}{6} + \frac{1}{5}} = \begin{cases} -5'707, \\ 46'347. \end{cases}$$

Puesto que el valor 0 está incluido entre ambos, las observaciones realizadas no permiten concluir que exista diferencia entre usar o no el diurético.

### Ejercicio 9.10

Evidentemente se trata de un caso de datos apareados, así que, en primer lugar, hay que calcular la variable diferencia  $D_i = X_i - Y_i$

$$D_i : \quad 1'5 \quad 2'1 \quad -0'8 \quad 0'6 \quad 1'7 \quad 0'9 \quad -0'3 \quad 1'2 \quad -0'7 \quad 2'2$$

sobre cuya media poblacional  $\mu$  se desea contrastar  $H_0 : \mu \leq 0$  frente a  $H_1 : \mu > 0$  (la hipótesis a demostrar). Ni que decir tiene que  $X$ ,  $Y$  y, por consiguiente,  $D$  se consideran con distribución normal.

La relación existente entre cotas de confianza y contrastes de hipótesis unilaterales, indica que debe rechazarse  $H_0$ , a nivel  $\alpha$ , si

$$\frac{\bar{d}}{S_d/\sqrt{n}} > t_{n-1;\alpha},$$

siendo  $S_d$  la cuasivarianza de  $D_i$  y  $\bar{d}$  su media muestral.

Como es  $\bar{d} = 0'84$ ,  $S_d^2 = 1'24$  y  $t_{9;0'05} = 1'833$ , resulta

$$\frac{\bar{d}}{S_d/\sqrt{n}} = \frac{0'84}{\sqrt{1'24/10}} = 2'38 > 1'833,$$

que permite rechazar  $H_0$  y concluir que existe una disminución significativa de la estatura a lo largo del día.

De hecho, el nivel crítico:  $P\{t_9 > 2'38\} = 0'022$ , es lo suficientemente pequeño como para obtener, con seguridad razonable, que los diferencias de estatura observadas no son fruto de errores aleatorios de medida.

# Capítulo 10

## Contrastes globales sobre la distribución poblacional

### 10.1. Introducción

Hasta ahora los tests de hipótesis han sido utilizados para contrastar la veracidad de hipótesis acerca de los parámetros de la distribución de una población. Sin embargo, en muchas ocasiones, es necesario emitir un juicio estadístico sobre la distribución poblacional en su conjunto. Los problemas de este tipo que se plantean de manera habitual son los siguientes:

- Decidir, a la vista de una muestra aleatoria de una población, si puede admitirse que la distribución poblacional coincide con una cierta distribución dada ( $N(0, 1)$ , exponencial(2), Poisson(5), etc.) o pertenece a un cierto tipo de distribuciones (normales, exponenciales, geométricas, etc.). Ello se conoce como el problema de *bondad del ajuste* del tipo de distribución especificada a los datos observados.
- Analizar si varias muestras aleatorias provienen de poblaciones con la misma distribución teórica, de forma que pueden ser utilizadas conjuntamente para inferencias posteriores acerca de ésta; o, por el contrario, son muestras de poblaciones con distinta distribución, que no pueden agruparse como información homogénea acerca de una única distribución. Esto se conoce como el problema de *homogeneidad* de varias muestras.
- Estudiar, en el caso de que se observen dos o más características de los elementos de la población (de forma que la distribución teórica no sea

unidimensional), si las características observadas pueden considerarse independientes, y se puede proceder a su análisis por separado, o, por el contrario, existe relación estadística entre ellas.

Cualquiera de estos problemas debe considerarse *no paramétrico*, en el sentido de que no se trata de decidir entre distribuciones  $F_\theta$  que sólo se diferencian en el valor del parámetro  $\theta$ . Por ejemplo, la hipótesis alternativa en un contraste de bondad del ajuste a una distribución uniforme en  $(0, 1)$ , está compuesta por todas las distribuciones diferentes de ella o, al menos, por todas las distribuciones continuas sobre  $(0, 1)$ .

En cuanto a la metodología a utilizar, cabe distinguir dos tipos de procedimientos claramente diferenciados.

Un primer método de trabajo consiste en descomponer el recorrido de la distribución teórica en un número finito de subconjuntos:  $A_1, A_2, \dots, A_k$ . Despues, clasificar las observaciones muestrales, según el subconjunto a que pertenezcan. Y, por último, comparar las frecuencias observadas de cada  $A_i$ , con las probabilidades que les corresponderían con una distribución teórica con las propiedades a contrastar.

Esta manera de proceder da lugar a los que se denominan *tests  $\chi^2$* , debido a que los estadísticos que miden la discrepancia entre frecuencias y probabilidades tienen este tipo de distribución límite.

Los tests  $\chi^2$  discretizan cualquier distribución poblacional y centran su interés en la variable aleatoria  $k$ -dimensional que cuenta el número de observaciones muestrales en cada uno de los subconjuntos  $A_i$ . Formalmente, los tests  $\chi^2$  realizan entonces contrastes de hipótesis sobre los parámetros de la distribución multinomial de esta variable  $k$ -dimensional. Así, el tratamiento de un problema, inicialmente no paramétrico, se reduce a un contraste paramétrico relacionado con la distribución multinomial. Las secciones 10.2, 10.3 y 10.4 se dedicarán al estudio de los tests de este tipo para cada uno de los contrastes anteriormente citados.

Un método alternativo consiste en usar la distribución muestral,  $F_n^*(x)$ , para compararla con la distribución teórica —en el caso de problemas de bondad del ajuste— o para compararlas entre sí —en el caso de problemas de homogeneidad. Midiendo la distancia entre ambas, puede llegar a saberse si es importante o poco significativa, siempre que se tenga un patrón de comparación; es decir, siempre que se conozca la distribución con la que es de esperar que se presenten las diversas distancias.

Un procedimiento para medir las distancias entre distribuciones y los resultados (asintóticos) sobre su distribución, fueron diseñados por Kolmo-

gorov y Smirnov y se considerarán en las secciones 10.5 y 10.6.

La sección 10.7 estudiará procedimientos no paramétricos, para sustituir los tests típicos sobre  $\mu = \mu_0$  (en el caso de una población) o sobre  $\mu_1 = \mu_2$  (en el caso de dos poblaciones), cuando no pueda admitirse que la distribución de la población es normal. El mismo tipo de técnicas se utilizarán en la sección 10.8 para diseñar tests de independencia, alternativos al test  $\chi^2$ .

## 10.2. Contrastes $\chi^2$ de bondad del ajuste

### 10.2.1. Hipótesis simple

Supongamos que se dispone de una muestra aleatoria simple, de tamaño  $n$ , de una población con distribución desconocida  $F$ . Trataremos, en primer lugar, de contrastar si puede aceptarse la hipótesis  $H_0 : F = F_0$ , donde  $F_0$  es una distribución completamente especificada, que no depende de ningún parámetro desconocido. En este sentido, la hipótesis nula es simple, mientras que la alternativa,  $H_1 : F \neq F_0$ , está compuesta por todas las distribuciones diferentes de  $F_0$ .

Si el recorrido de la distribución poblacional se divide en  $k$  conjuntos disjuntos:  $A_1, A_2, \dots, A_k$ , sea  $p_i^0$  la probabilidad de cada  $A_i$  bajo la distribución  $F_0$  y  $p_i$  la probabilidad desconocida que la distribución teórica  $F$  asigna a cada  $A_i$ .

Naturalmente, la variable aleatoria  $(N_1, N_2, \dots, N_k)$ , que cuenta el número de observaciones de la muestra en cada subconjunto  $A_1, A_2, \dots, A_k$ , tiene distribución multinomial con parámetros  $n$  y  $\mathbf{p} = (p_1, p_2, \dots, p_k)$ .

Es decir:

$$P\{N_1 = n_1, N_2 = n_2, \dots, N_k = n_k\} = \frac{n!}{n_1! n_2! \dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}.$$

La idea consiste entonces en sustituir el contraste no paramétrico propuesto inicialmente, por el contraste de la hipótesis  $H_0 : \mathbf{p} = \mathbf{p}^0$ , frente a la alternativa  $H_1 : \mathbf{p} \neq \mathbf{p}^0$ , llevado a cabo sobre la base de los valores  $n_1, \dots, n_k$ , que la muestra ha proporcionado a las variables  $N_1, \dots, N_k$ .

En primer lugar, se puede pensar en utilizar el test de razón de verosimilitudes. Para ello, recuérdese (cf. Ejemplo 7.6) que los estimadores de máxima verosimilitud de los parámetros  $p_i$  son  $\hat{p}_i = N_i/n$ , con lo cual la

razón de verosimilitudes es

$$\Lambda(N_1, N_2, \dots, N_k) = \frac{(p_1^0)^{N_1} (p_2^0)^{N_2} \cdots (p_k^0)^{N_k}}{\hat{p}_1^{N_1} \hat{p}_2^{N_2} \cdots \hat{p}_k^{N_k}} = \prod_{i=1}^k \left( \frac{p_i^0}{\hat{p}_i} \right)^{N_i}$$

El test de razón de verosimilitudes tiene entonces por región crítica

$$\{\Lambda < c\} = \{-2 \log \Lambda > k\}$$

y está basado, por tanto, en el estadístico

$$-2 \log \Lambda = 2 \sum_{i=1}^k N_i (\log \hat{p}_i - \log p_i^0),$$

cuya distribución asintótica es  $\chi_{k-1}^2$  (supuesto que  $\mathbf{p} = \mathbf{p}^0$ ), en virtud del resultado general del capítulo anterior acerca de la distribución de la razón de verosimilitudes (habida cuenta de que la relación  $\sum_{i=1}^k p_i = 1$  establece que hay  $k - 1$  parámetros en el problema).

Nótese que  $-2 \log \Lambda$  constituye una medida ponderada de las diferencias existentes entre las probabilidades teóricas ( $p_i^0$ ) de cada uno de los conjuntos  $A_i$  y las frecuencias observadas ( $\hat{p}_i$ ) en cada uno de ellos. (En el sentido de que se anula cuando coinciden  $\hat{p}_i$  y  $p_i^0$  y crece conforme unas se alejan de las otras).

Esto es lo esencial del método: Primero, disponer de un estadístico que mida la discrepancia entre las frecuencias observadas y las probabilidades indicadas por el modelo a contrastar. Segundo, conocer la distribución en el muestreo de dicho estadístico, para poder juzgar si su valor es suficientemente grande para poner de manifiesto diferencias importantes entre unas y otras.

La tradición estadística en relación con la bondad del ajuste, iniciada por Pearson en 1900, antes del desarrollo de los tests de razón de verosimilitudes, hace que normalmente no se emplee el estadístico  $-2 \log \Lambda$ . Lo usual es sustituirlo por una medida, en cierta forma más natural, de la discrepancia entre las  $\hat{p}_i$  y las  $p_i^0$ , definida por:

$$D = \sum_{i=1}^k \frac{n}{p_i^0} (\hat{p}_i - p_i^0)^2 = \sum_{i=1}^k \frac{(N_i - np_i^0)^2}{np_i^0}$$

que, como se ve en la última expresión, acumula las diferencias al cuadrado entre el número de observaciones en cada conjunto  $A_i$  y el número de

ellas que cabía esperar según  $F_0$ , ponderándolas mediante  $1/np_i^0$ , puesto que no parece lógico dar la misma importancia a una diferencia de 2 —por ejemplo— donde eran de esperar 20 observaciones que donde había que esperar solamente 5.

La intuición de Pearson, al elegir los factores de ponderación  $1/np_i^0$  es sorprendente: no sólo permite determinar la distribución en el muestreo de  $D$  —como veremos enseguida—, sino que proporciona valores muy parecidos a la razón de verosimilitudes. En efecto, admitiendo que  $(p_i^0 - \hat{p}_i)/\hat{p}_i$  sean cantidades pequeñas,

$$\log \frac{p_i^0}{\hat{p}_i} = \log \left( 1 + \frac{p_i^0 - \hat{p}_i}{\hat{p}_i} \right) \simeq \frac{p_i^0 - \hat{p}_i}{\hat{p}_i} - \frac{1}{2} \left( \frac{p_i^0 - \hat{p}_i}{\hat{p}_i} \right)^2,$$

luego

$$\begin{aligned} \log \Lambda &\simeq \sum_{i=1}^k N_i \frac{p_i^0 - \hat{p}_i}{\hat{p}_i} - \frac{1}{2} \sum_{i=1}^k N_i \frac{(p_i^0 - \hat{p}_i)^2}{\hat{p}_i^2} \\ &= n \sum_{i=1}^k (p_i^0 - \hat{p}_i) - \frac{1}{2} \sum_{i=1}^k \frac{n^2}{N_i} (\hat{p}_i - p_i^0)^2 \\ &= -\frac{1}{2} \sum_{i=1}^k \frac{(N_i - np_i^0)^2}{N_i} \end{aligned}$$

y, por consiguiente,

$$-2 \log \Lambda \simeq \sum_{i=1}^k \frac{(N_i - np_i^0)^2}{N_i} \simeq \sum_{i=1}^k \frac{(N_i - np_i^0)^2}{np_i^0}$$

puesto que  $N_i \simeq np_i^0$  para cada  $i$ .

En cualquier caso, la elección del estadístico  $D$  conduce al test de bondad del ajuste de región crítica:  $\{D > c\}$  y es necesario conocer la distribución en el muestreo de  $D$  —bajo la hipótesis nula— para poder determinar  $c$  de acuerdo con el nivel de significación escogido. Este es el objetivo del siguiente teorema.

**Teorema 10.1:** *Cuando  $n \rightarrow \infty$*

$$D = \sum_{i=1}^k \frac{(N_i - np_i)^2}{np_i} \xrightarrow{d} \chi_{k-1}^2.$$

Demostración: Sea

$$\xi_r = (I_{\{X_r \in A_1\}}, I_{\{X_r \in A_2\}}, \dots, I_{\{X_r \in A_{k-1}\}})$$

la variable aleatoria, de dimensión  $k - 1$ , que indica a cual de los conjuntos  $A_1, A_2, \dots, A_{k-1}$  pertenece el elemento muestral  $r$ -ésimo,  $X_r$ . (Es decir  $\xi_r = (0, 0, \dots, 1, \dots, 0)$ , con un 1 en la posición  $i$  tal que  $X_r \in A_i$ ). Será entonces

$$(N_1, N_2, \dots, N_{k-1}) = \sum_{r=1}^n \xi_r,$$

que expresa  $\mathbf{N} = (N_1, N_2, \dots, N_{k-1})$  como suma de las  $n$  variables aleatorias independientes  $\xi_r$ , de vector de medias  $\mathbf{p} = (p_1, p_2, \dots, p_{k-1})$  y matriz de covarianzas  $\Sigma = (\sigma_{ij})$ , donde

$$\sigma_{ij} = E[I_{\{X_r \in A_i\}} I_{\{X_r \in A_j\}}] - p_i p_j = \begin{cases} -p_i p_j & \text{si } i \neq j, \\ p_i(1 - p_i) & \text{si } i = j. \end{cases}$$

Entonces,  $\eta_r = (\xi_r - \mathbf{p})\Sigma^{-1/2}$  es una variable con vector de medias 0 y matriz de covarianzas  $I$  (la matriz identidad de dimensión  $k - 1$ )<sup>1</sup>. Además

$$\frac{\mathbf{N} - \mathbf{p}}{\sqrt{n}} \Sigma^{-1/2} = \frac{1}{\sqrt{n}} \sum_{r=1}^n \eta_r \xrightarrow{d} N_{k-1}(0, I),$$

en virtud del teorema central del límite. Luego

$$\left( \frac{\mathbf{N} - \mathbf{p}}{\sqrt{n}} \right) \Sigma^{-1} \left( \frac{\mathbf{N} - \mathbf{p}}{\sqrt{n}} \right)^t \xrightarrow{d} \chi_{k-1}^2.$$

Una simple comprobación muestra que

$$\Sigma^{-1} = \begin{pmatrix} 1/p_1 + 1/p_k & 1/p_k & 1/p_k \\ 1/p_k & 1/p_2 + 1/p_k & 1/p_k \\ \vdots & \vdots & \vdots \\ 1/p_k & 1/p_k & 1/p_{k-1} + 1/p_k \end{pmatrix},$$

---

<sup>1</sup> Como  $\Sigma$  es definida positiva se puede expresar  $\Sigma = CDC^t$  siendo  $C$  ortogonal y  $D$  diagonal, de términos  $d_{ii}$ . Si  $D^{-1/2}$  es la matriz diagonal, de términos  $1/\sqrt{d_{ii}}$ , es  $\Sigma^{-1/2} = CD^{-1/2}C^t$ . Por tanto

$$E[\eta_r^t \cdot \eta_r] = E[\Sigma^{-1/2}(\xi_r - \mathbf{p})^t(\xi_r - \mathbf{p})\Sigma^{-1/2}] = (\Sigma^{-1/2})^t \Sigma \Sigma^{-1/2} = I.$$

con lo cual

$$\begin{aligned} & \left( \frac{\mathbf{N} - \mathbf{p}}{\sqrt{n}} \right) \Sigma^{-1} \left( \frac{\mathbf{N} - \mathbf{p}}{\sqrt{n}} \right)^t = \\ &= \sum_{i=1}^{k-1} \left( \frac{1}{p_i} + \frac{1}{p_k} \right) \frac{(N_i - np_i)^2}{n} + \sum_{i \neq j} \frac{1}{p_k} \frac{(N_i - np_i)(N_j - np_j)}{n} \\ &= \sum_{i=1}^{k-1} \frac{(N_i - np_i)^2}{np_i} + \frac{1}{np_k} \left( \sum_{i=1}^{k-1} (N_i - np_i) \right)^2 = \sum_{i=1}^k \frac{(N_i - np_i)^2}{np_i}, \end{aligned}$$

puesto que  $N_k - np_k = n - \sum_{i=1}^{k-1} N_i - n \left( 1 - \sum_{i=1}^{k-1} p_i \right) = \sum_{i=1}^{k-1} (np_i - N_i)$ .

En definitiva

$$D = \sum_{i=1}^k \frac{(N_i - np_i)^2}{np_i} \xrightarrow{d} \chi_{k-1}^2.$$

El teorema anterior, usado con  $p_i = p_i^0$ , proporciona, bajo la hipótesis nula, la distribución aproximada de  $D$  cuando  $n$  es grande. Lo grande que debe ser  $n$ , para que la aproximación sea aceptablemente buena, está en relación con los valores de las probabilidades  $p_i$ ; usualmente se considera que debe ser

$$np_i \geq 5 \quad \text{para cada } i = 1, 2, \dots, k.$$

Ello sirve, a menudo, para delimitar los conjuntos  $A_i$ , que han de elegirse de forma que se cumplan las restricciones  $np_i^0 \geq 5$ .

Sin embargo, puesto que el contraste no discrimina entre distribuciones  $F_0$  que asignen la misma probabilidad a los conjuntos  $A_i$ , es aconsejable que el número  $k$  de elementos de la partición no sea inferior a 5 (salvo en ajustes a distribuciones discretas con menor número de valores posibles). Así, tendrá que ser alguna de las  $p_i^0 \leq 1/5$  y, por consiguiente, es necesario que sea  $n \geq 25$ . Aunque, evidentemente, se conseguirá mayor exactitud cuanto más grande sea  $n$ .

Algunos ejemplos pueden clarificar la manera práctica de proceder.

### Ejemplo 10.1

Se ha estimado que el número de accidentes diarios en cada regimiento del ejército sigue una distribución de Poisson de parámetro 2. Un determinado regimiento ha recogido, durante 200 días, los siguientes datos:

nº de accidentes:	0	1	2	3	4	5	6	7
nº de días:	22	53	58	39	20	5	2	1

con los cuales quiere contrastar si se ajusta a la distribución indicada. Debe entonces calcular las probabilidades  $p_i^0 = e^{-2}2^x/x!$  y el número esperado ( $np_i^0 = 200p_i^0$ ) de cada cantidad de accidentes:

$i :$	0	1	2	3	4	$\geq 5$
$p_i^0 :$	0'1353	0'2707	0'2707	0'1804	0'0902	0'0527
$np_i^0 :$	27'06	54'14	54'14	36'08	18'04	10'54

donde se han agrupado los valores de  $i \geq 5$  para tener valores de  $np_i^0 \geq 5$ .

La comparación de las frecuencias observadas con las frecuencias esperadas, se realiza mediante el estadístico

$$D = \sum_{i=0}^5 \frac{(n_i - np_i^0)^2}{np_i^0} = -n + \sum_{i=1}^5 \frac{n_i^2}{np_i^0}.$$

Su valor con los datos disponibles es

$$\frac{22^2}{27'06} + \frac{53^2}{54'14} + \cdots + \frac{8^2}{10'54} - 200 = 2'307,$$

y su distribución, si la hipótesis es correcta, es aproximadamente  $\chi^2$  con 5 grados de libertad (puesto que  $k = 6$ ). Por ejemplo  $P\{\chi_5^2 > 7'29\} = 0'2$ , de manera que únicamente un valor de  $D$  superior a 7'29 permitiría, con nivel de significación 0'2, afirmar que las diferencias entre las frecuencias observadas y esperadas son significativas de que la variable en cuestión no tiene distribución de Poisson de parámetro 2. El valor hallado de  $D$  no permite, por tanto, rechazar la hipótesis con ningún nivel de significación razonable.

La razón de verosimilitudes

$$-2 \log \Lambda = 2 \left[ 22 \log \left( \frac{22}{27'06} \right) + \cdots + 8 \log \left( \frac{8}{10'54} \right) \right] = 2'427,$$

tiene un valor similar al de  $D$  y, con la misma distribución  $\chi_5^2$ , daría lugar a la misma conclusión.

### Ejemplo 10.2

Una máquina, en correcto estado de funcionamiento, produce piezas cuya longitud se distribuye con distribución  $N(10'5, 0'15)$ . En determinado momento se observa la siguiente muestra, de tamaño 40, de la longitud de las

piezas producidas:

10'39	10'66	10'12	10'32	10'25	10'91	10'52	10'83
10'72	10'28	10'35	10'46	10'54	10'72	10'23	10'18
10'62	10'49	10'32	10'61	10'64	10'23	10'29	10'78
10'81	10'39	10'34	10'62	10'75	10'34	10'41	10'81
10'64	10'53	10'31	10'46	10'47	10'43	10'57	10'74

y se desea saber si la muestra avala que la máquina está funcionando correctamente.

Para hacer el contraste  $\chi^2$  de bondad del ajuste, habrá que dividir el recorrido de la variable en intervalos; como  $n = 40$ , se pueden hacer 8 intervalos, con número esperado de observaciones igual a 5, buscando los cuantiles de órdenes 0'125, 0'25, 0'375, ..., 0'875. Ello da lugar a la partición:

$A_i$	$n_i$	$np_i^0$
$\leq 10'33$	10	5
$(10'33, 10'4]$	5	5
$(10'4, 10'45]$	2	5
$(10'45, 10'5]$	4	5
$(10'5, 10'55]$	3	5
$(10'55, 10'6]$	1	5
$(10'6, 10'67]$	6	5
$> 10'67$	9	5
Total	40	40

Al clasificar las observaciones en tales conjuntos se obtiene la segunda columna, mientras que la tercera expresa el número de ellas que eran de esperar en cada conjunto. Entonces es

$$D = \sum_{i=1}^8 \frac{(n_i - np_i^0)^2}{np_i^0} = \frac{5^2 + 0 + 3^2 + 1 + 2^2 + 4^2 + 1 + 4^2}{5} = 14'4.$$

Si la hipótesis fuese correcta, la distribución de  $D$  es  $\chi_7^2$  y la tabla de la  $\chi^2$  indica que

$$P\{\chi_7^2 > 14'4\} = 0'0445.$$

Por tanto, el valor de  $D$  es suficientemente grande para afirmar, con cualquier nivel de significación superior a 0'0445, que las discrepancias ocurridas entre las frecuencias observadas y las esperadas no son fruto del azar, sino que la máquina no está produciendo actualmente piezas con longitud  $N(10'5, 0'15)$ .

### Ejemplo 10.3

Los impactos de 60 bombas volantes sobre la superficie de Londres, considerada cuadrada, fueron clasificados en 9 zonas obtenidas dividiendo cada lado en tres partes iguales, con los siguientes resultados:

8	7	3
5	9	11
6	4	7

Los responsables de la defensa querían averiguar si las bombas perseguían algún objetivo concreto o se distribuían al azar sobre la superficie de la ciudad.

Con distribución uniforme sobre toda la superficie, cada cuadrícula tenía probabilidad  $1/9$  de recibir cada impacto y, por tanto, un número esperado de impactos de  $60/9$ . La discrepancia entre las frecuencias observadas y las esperadas puede medirse mediante

$$D = \sum_{i=1}^9 \frac{(n_i - np_i^0)^2}{np_i^0} = -60 + \frac{9}{60} \sum_{i=1}^9 n_i^2$$

que, con los datos obtenidos, toma el valor  $7'5$ .

Bajo la hipótesis de que la distribución es uniforme,  $D$  tiene distribución  $\chi_8^2$  y, por tanto,

$$P\{\chi_8^2 > 7'5\} = 0'48;$$

valor que no permite rechazar la hipótesis de que los impactos se distribuyen al azar sobre toda la superficie. Incluso con un nivel de significación tan poco exigente como  $0'2$ , debería ser  $D \geq 11'03$ , para poder descartar la aleatoriedad de los impactos.

#### 10.2.2. Hipótesis compuesta

Los resultados y ejemplos anteriores muestran cómo contrastar la bondad del ajuste de una distribución totalmente especificada a una población, mediante una muestra aleatoria simple obtenida de ella. Sin embargo, la situación más usual es aquella en que la hipótesis a contrastar es compuesta y, más concretamente, especifica una familia de distribuciones de forma funcional dada, pero dependiente de algunos parámetros no especificados. Es frecuente, por ejemplo, sospechar que un conjunto de observaciones provienen de una distribución normal, uniforme, etc., pero desconocer cuál de las distribuciones dentro de la familia es la adecuada para describir la población.

A primera vista se puede pensar en realizar la estimación puntual de los parámetros, para contrastar después el ajuste a la distribución poblacional que indiquen los propios datos como la más adecuada. La idea es correcta y no merecería mayor comentario, en el caso de que la estimación de los parámetros se realizase con una muestra distinta (independiente) de aquella que se utiliza para juzgar el ajuste. Sin embargo, si se utiliza la misma muestra para ambos fines, el proceso ha de ser examinado con más atención: al realizar la estimación de  $F_0$ , las probabilidades  $p_i^0$  dependerán de la muestra; no es válido entonces, en el teorema 10.1, sustituir las  $p_i$  por probabilidades  $p_i^0$  que no son constantes (como antes), sino variables aleatorias, función de la misma muestra que proporciona los valores de las variables  $N_i$ . Intuitivamente, parece claro que si la distribución poblacional se selecciona —a través de la estimación de sus parámetros— para que se adapte bien a la muestra obtenida, se puede ser más exigente a la hora de juzgar si las frecuencias observadas responden adecuadamente a las previstas por la distribución escogida. Pero, ¿cómo cuantificar esta sensación? Hay que volver a considerar la cuestión desde el principio.

En la nueva situación, la pretensión será contrastar una hipótesis nula de la forma  $F \in \{F_\theta \mid \theta \in \Theta \subset \mathbb{R}^q\}$ , acerca de la distribución poblacional desconocida  $F$ .

Conservando el modo de operar, se divide el conjunto de posibles valores de las observaciones en  $k$  conjuntos disjuntos,  $A_1, A_2, \dots, A_k$ , y  $N_i$  seguirá representando el número de observaciones muestrales en cada  $A_i$ . La variable  $\mathbf{N} = (N_1, N_2, \dots, N_k)$  sigue teniendo distribución multinomial de parámetros  $n$  y  $\mathbf{p} = (p_1, p_2, \dots, p_k)$ .

La diferencia con el caso anterior es que, ahora, la hipótesis nula sobre la distribución poblacional no determina unas únicas probabilidades  $p_i^0$  de los conjuntos  $A_i$ , sino que cada valor del parámetro  $\theta$  proporciona sus propios valores:  $\mathbf{p} = \pi(\theta)$ , o más explícitamente,

$$\begin{cases} p_i = \pi_i(\theta_1, \theta_2, \dots, \theta_q) \\ i = 1, 2, \dots, k, \end{cases}$$

lo cual define la región

$$\mathcal{R} = \{\pi(\theta) \mid \theta \in \Theta\} \subset \{\mathbf{p} = (p_1, \dots, p_k) \mid p_1, \dots, p_k \in (0, 1) \text{ y } \sum_{i=1}^k p_i = 1\}.$$

La hipótesis  $F \in \{F_\theta \mid \theta \in \Theta\}$  a contrastar, se sustituye entonces por la hipótesis  $H_0 : \mathbf{p} \in \mathcal{R}$ , acerca de los parámetros de la distribución multinomial de  $\mathbf{N}$ .

Diseñar el test de razón de verosimilitudes, en esta nueva situación, exige maximizar

$$p_1^{N_1} p_2^{N_2} \cdots p_k^{N_k}$$

bajo la condición  $\mathbf{p} \in \mathcal{R}$ . Es decir, buscar el valor  $\hat{\theta} \in \Theta$  que hace máximo

$$\sum_{i=1}^k N_i \log p_i = \sum_{i=1}^k N_i \log \pi_i(\theta).$$

Supuesto que las funciones  $\pi_i$  son derivables, ello conduce a resolver el sistema de ecuaciones

$$\begin{cases} \sum_{i=1}^k \frac{N_i}{\pi_i(\theta)} \frac{\partial}{\partial \theta_j} \pi_i(\theta) = 0 \\ j = 1, 2, \dots, q, \end{cases} \quad (10.1)$$

que habitualmente tiene una única solución.

Puesto que el máximo sin restricciones, sobre  $\mathbf{p}$ , se sigue alcanzando en  $\hat{\mathbf{p}}$ , la razón de verosimilitudes será entonces

$$\Lambda(N_1, \dots, N_k) = \prod_{i=1}^k \left( \frac{\pi_i(\hat{\theta})}{\hat{p}_i} \right)^{N_i}$$

Luego el test de razón de verosimilitudes tiene región crítica de la forma:

$$\{\Lambda < c\} = \{-2 \log \Lambda > k\},$$

donde

$$-2 \log \Lambda = 2 \sum_{i=1}^k N_i \left( \log \hat{p}_i - \log \pi_i(\hat{\theta}) \right)$$

tiene distribución asintótica  $\chi_{k-1-q}^2$  (cf. sección 9.2), bajo condiciones de regularidad que habitualmente se verifican. Ello muestra que la estimación de cada parámetro de la distribución poblacional, rebaja en 1 el número de grados de libertad de la distribución asintótica de la razón de verosimilitudes (supuesto que se realiza mediante el sistema de ecuaciones [10.1]).

De la misma manera que en el apartado (A), lo habitual es reemplazar el estadístico  $-2 \log \Lambda$  por

$$D = \sum_{i=1}^k \frac{(N_i - n\pi_i(\hat{\theta}))^2}{n\pi_i(\hat{\theta})}$$

y utilizar el test de región crítica  $\{D > c\}$ . Para ello es necesario conocer la distribución asintótica de  $D$ , que queda precisada en el siguiente resultado<sup>2</sup>:

**Teorema 10.2:** Si  $\Theta$  es un intervalo de  $\mathbb{R}^q$  ( $q < k - 1$ ), en el cual:

- i) Existen y son continuas  $\frac{\partial}{\partial \theta_j} \pi_i(\theta)$  para cada  $i = 1, \dots, k$  y  $j = 1, \dots, q$ ;
- ii) La matriz  $k \times q$  de términos  $\frac{1}{\sqrt{\pi_i(\theta)}} \frac{\partial}{\partial \theta_j} \pi_i(\theta)$  tiene rango  $q$ .

Entonces, si para cada tamaño muestral  $n$ ,  $\hat{\theta}_n$  es una solución del sistema de ecuaciones [10.1], tal que  $\hat{\theta}_n \rightarrow \theta$ , se verifica

$$D = \sum_{i=1}^k \frac{(N_i - n\pi_i(\hat{\theta}_n))}{n\pi_i(\hat{\theta}_n)} \xrightarrow{d} \chi^2_{k-1-q}.$$

El teorema es válido, en realidad, para sucesiones de estimadores  $\hat{\theta}_n$  más generales que las proporcionadas por el sistema de ecuaciones [10.1], que —recuérdese— proviene de aplicar el principio de máxima verosimilitud a los datos agrupados en las clases  $A_i$ . En el caso de poblaciones continuas, si el único objetivo fuese conocer el valor de  $\theta$ , sería preferible hallar la estimación de máxima verosimilitud directamente a partir de la muestra  $X_1, \dots, X_n$ , sin agruparla previamente en los conjuntos  $A_1, \dots, A_k$ , lo cual supone una pérdida evidente de información. Habitualmente el resultado anterior se aplica tomando como  $\hat{\theta}_n$  esta última estimación de máxima verosimilitud, dando por supuesto que su utilización afectará levemente a la distribución asintótica de  $D$ .

También en el caso de poblaciones continuas, es frecuente en la práctica, elegir los conjuntos  $A_1, A_2, \dots, A_k$  en función de la distribución poblacional estimada (con probabilidades aproximadamente iguales, tales que  $np_i \geq 5$ ). Sin embargo, con esta manera de proceder, los conjuntos  $A_i$  dependen de las observaciones muestrales, lo cual supone una clara desviación de las hipótesis teóricas, pues la distribución de  $(N_1, \dots, N_k)$  no será entonces multinomial. Es preferible, por tanto, elegir arbitrariamente un número considerable de conjuntos  $A_i$  y agruparlos después para conseguir que sea  $np_i \geq 5$ .

---

<sup>2</sup> Véase Rao (1973) pág. 391.

### Ejemplo 10.4

Un modelo genético indica que la distribución de la población entre hombres y mujeres, daltónicos o normales, se ajusta a las probabilidades

	Hombres	Mujeres
Normales	$q/2$	$q^2/2 + pq$
Daltónicos	$p/2$	$p^2/2$

siendo  $p = 1 - q$ , la proporción en la población de cromosomas  $X$  portadores del daltonismo. Para comprobar la teoría se examinaron 2000 individuos elegidos al azar, con los siguientes resultados:

	Hombres	Mujeres
Normales	894	1015
Daltónicos	81	10

y se desea saber si las observaciones concuerdan con el modelo.

Puesto que  $q$  no es conocido, habrá que hallar su estimación de máxima verosimilitud. La muestra observada tiene por verosimilitud

$$\frac{2000!}{894! \ 81! \ 1015! \ 10!} \left(\frac{q}{2}\right)^{894} \left(\frac{1-q}{2}\right)^{81} \left[q\left(1-\frac{q}{2}\right)\right]^{1015} \left[\frac{(1-q)^2}{2}\right]^{10},$$

cuyo logaritmo (prescindiendo de términos independientes de  $q$ ) es

$$894 \log q + 81 \log(1-q) + 1015 \log q + 1015 \log(2-q) + 20 \log(1-q)$$

y tiene por derivada respecto a  $q$

$$\frac{1909}{q} - \frac{101}{1-q} - \frac{1015}{2-q}.$$

La estimación de  $q$  es pues  $\hat{q} = 0'91277$  (la única raíz, menor que 1, de la ecuación anterior; que corresponde a un máximo de la verosimilitud). Con dicho valor de  $q$ , los números esperados de observaciones en cada una de los cuatro grupos son

	Hombres	Mujeres
Normales	912'77	992'39
Daltónicos	87'23	7'61

La discrepancia entre las frecuencias observadas y las esperadas, medida por el estadístico  $D$ , toma el valor

$$\frac{(894 - 912'77)^2}{912'77} + \frac{(81 - 87'23)^2}{87'23} + \frac{(1015 - 992'39)^2}{992'39} + \frac{(10 - 7'61)^2}{7'61} = 2'097.$$

Por otra parte, como  $D$  está compuesto de cuatro sumandos y se ha estimado un parámetro ( $q$ ) de la distribución,  $D$  tiene distribución asintótica  $\chi^2_2$ . Ahora bien

$$P\{\chi^2_2 > 2'097\} = 0'35,$$

luego no puede rechazarse la hipótesis de que las observaciones concuerdan con el modelo; haría falta un valor de  $D$  superior a  $4'6$  para poder afirmar, con nivel de significación  $0'1$ , que las diferencias entre las frecuencias observadas y esperadas no son debidas al azar.

Nótese que un valor muy pequeño de  $D$  no es prueba de un muy buen ajuste de las observaciones al modelo teórico; antes bien induce a sospechar que las frecuencias observadas han sido manipuladas para que se ajusten a las esperadas. De hecho, puesto que la distribución de  $D$  es aproximadamente  $\chi^2_2$  (en este caso concreto), lo normal será obtener un valor moderado de  $D$  y la tabla de la  $\chi^2_2$  indica, por ejemplo, que sólo en un 5% de los casos será  $D < 0'103$ .

### Ejemplo 10.5

La muestra del ejemplo 10.2 ha puesto de relieve un funcionamiento incorrecto de la máquina, que no se ajusta a la distribución  $N(10'5, 0'15)$ . Cabe, sin embargo, la posibilidad de que el desajuste haya afectado a la media y la desviación típica, pero que la distribución de la longitud de las piezas producidas siga siendo de tipo normal. Para verificar tal posibilidad hay que estimar (por máxima verosimilitud) la media y la desviación típica poblacional; lo cual da

$$\hat{\mu} = \bar{x} = 10'502 \quad \text{y} \quad \hat{\sigma} = s = 0'2025.$$

En lugar de hacer el contraste de bondad del ajuste a la  $N(10'502, 0'2025)$  como en el ejemplo 10.2 (eligiendo la partición en función de la distribución), es preferible tomar una partición cualquiera, para calcular sus frecuencias observadas y esperadas; por ejemplo:

$A_i$	$n_i$	$np_i$
$\leq 10'2$	2	2'72
$(10'2, 10'3]$	5	3'65
$(10'3, 10'4]$	8	5'92
$(10'4, 10'5]$	6	7'55
$(10'5, 10'6]$	4	7'59
$(10'6, 10'7]$	6	6'00
$(10'7, 10'8]$	5	3'74
$(10'8, 10'9]$	3	1'84
$> 10'9$	1	0'99

donde la segunda columna corresponde al número de observaciones en cada intervalo y la tercera a las probabilidades que la  $N(10'502, 0'2025)$  asigna a cada intervalo, multiplicadas por 40. Los intervalos extremos se agrupan para conseguir un número esperado de observaciones superior a 5:

$A_i$	$n_i$	$np_i$
$\leq 10'3$	7	6'37
$(10'3, 10'4]$	8	5'92
$(10'4, 10'5]$	6	7'55
$(10'5, 10'6]$	4	7'59
$(10'6, 10'7]$	6	6'00
$> 10'7$	9	6'57

Según esta última tabla es

$$D = \sum_{i=1}^6 \frac{(n_i - np_i)^2}{np_i} = 3'708$$

y la distribución de  $D$  es aproximadamente  $\chi_3^2$  (puesto que de los seis sumandos hay que descontar siempre uno y, en este caso, dos más por los dos parámetros estimados a partir de la muestra). Como

$$P\{\chi_3^2 > 3'708\} = 0'295,$$

la muestra no permite rechazar la hipótesis de que la población tiene distribución normal (y más concretamente  $N(10'502, 0'2025)$ ). La conclusión debe ser pues que el desajuste sufrido por la máquina ha aumentado la desviación típica, sin afectar sustancialmente al tipo de distribución, ni a su media.

### 10.3. Contraste $\chi^2$ de homogeneidad

Supongamos que se dispone de  $m$  muestras aleatorias simples de otras tantas poblaciones, cuyos tamaños son respectivamente  $n_1, n_2, \dots, n_m$ . Puede tratarse, por ejemplo, de medidas antropométricas sobre individuos de distintas regiones, de características físicas de piezas producidas en diversas condiciones, etc.

A partir de estos datos, se desea decidir si la distribución poblacional es la misma en todos los casos, y por consiguiente se dispone de una muestra de tamaño  $n = n_1 + n_2 + \dots + n_m$  de una única distribución; o, por el contrario, se trata de poblaciones heterogéneas con diferentes distribuciones.

Nuevamente, el conjunto de posibles valores de las observaciones se dividirá en  $k$  conjuntos disjuntos:  $A_1, A_2, \dots, A_k$ ; clasificando en ellos las observaciones de cada muestra. Si  $n_{ij}$  representa el número de observaciones de la muestra  $i$  que pertenecen al conjunto  $A_j$ , los datos pueden tabularse en lo que se denomina una *tabla de contingencia*  $m \times k$ :

Muestra	$A_1$	$A_2$	...	$A_k$	Total
1	$n_{11}$	$n_{12}$	...	$n_{1k}$	$n_1$
2	$n_{21}$	$n_{22}$		$n_{2k}$	$n_2$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
$m$	$n_{m1}$	$n_{m2}$	...	$n_{mk}$	$n_m$
Total	$n_{\cdot 1}$	$n_{\cdot 2}$	...	$n_{\cdot k}$	$n$

La hipótesis de que las  $m$  poblaciones son homogéneas, se traduce en que cada conjunto  $A_j$  debe tener una probabilidad teórica,  $p_j$ , desconocida, pero que no varía de la población  $i$  a la  $i'$ . En cambio, la heterogeneidad de las poblaciones debe reflejarse (al menos si la partición  $A_j$  es suficientemente fina) en que, para algún  $j$ , las probabilidades  $p_{ij}$ , que la distribución de la población  $i$  asigna al conjunto  $A_j$ , no sean todas iguales.

El contraste de homogeneidad se plantea entonces, en forma simplificada, como el contraste de la hipótesis nula:

$$H_0 : \begin{cases} p_{1j} = p_{2j} = \dots = p_{mj} \quad (= p_j) \\ \text{para cada } j = 1, 2, \dots, k, \end{cases}$$

frente a la alternativa de que alguna de tales igualdades no es cierta.

Naturalmente, si  $\mathbf{N}_i = (N_{i1}, N_{i2}, \dots, N_{ik})$  es la variable aleatoria que cuenta el número de observaciones de la muestra  $i$  dentro de cada conjunto

$A_j$  —y que da lugar a los datos  $(n_{i1}, n_{i2}, \dots, n_{ik})$ — la distribución de  $\mathbf{N}_i$  es una distribución multinomial de parámetros  $n_i$  y  $\mathbf{p}_i = (p_{i1}, p_{i2}, \dots, p_{ik})$ . De forma que el test que se pretende diseñar es el de la igualdad:  $\mathbf{p}_1 = \mathbf{p}_2 = \dots = \mathbf{p}_m$ , de los parámetros de  $m$  distribuciones multinomiales.

Bajo la hipótesis nula, la probabilidad de la clasificación reflejada en la tabla de contingencia es proporcional a

$$p_1^{n_{.1}} p_2^{n_{.2}} \cdots p_k^{n_{.k}}$$

(eliminando factores que no dependen de las  $p_i$ ). Como se sabe dicha probabilidad (sometida a  $\sum_{j=1}^k p_j = 1$ ) se hace máxima cuando  $p_j$  vale  $\hat{p}_{.j} = n_{.j}/n$  para cada  $j$ .

En cambio, sin suponer cierta la hipótesis nula, la probabilidad de obtener los datos de la tabla de contingencia es proporcional a

$$p_{11}^{n_{11}} \cdots p_{1k}^{n_{1k}} p_{21}^{n_{21}} \cdots p_{2k}^{n_{2k}} p_{m1}^{n_{m1}} \cdots p_{mk}^{n_{mk}},$$

cuyo máximo (con las restricciones  $\sum_{j=1}^k p_{ij} = 1$  para cada  $i$ ) se alcanza en  $\hat{p}_{ij} = n_{ij}/n_i$ .

La razón de verosimilitudes es por tanto

$$\Lambda = \frac{\prod_{j=1}^k \left(\frac{n_{.j}}{n}\right)^{n_{.j}}}{\prod_{i=1}^m \prod_{j=1}^k \left(\frac{n_{ij}}{n_i}\right)^{n_{ij}}}$$

y da lugar al test de región crítica

$$\{\Lambda < c\} = \{-2 \log \Lambda > k\},$$

donde

$$\begin{aligned} -2 \log \Lambda &= 2 \sum_{i=1}^m \sum_{j=1}^k n_{ij} \log \frac{n_{ij}}{n_i} - 2 \sum_{j=1}^k n_{.j} \log \frac{n_{.j}}{n} \\ &= 2 \sum_{i=1}^m \sum_{j=1}^k n_{ij} (\log \hat{p}_{ij} - \log \hat{p}_{.j}). \end{aligned}$$

Según los resultados de la sección 9.2, si la hipótesis nula es correcta, la distribución asintótica de  $-2 \log \Lambda$ , es

$$\chi^2_{(mk-m)-(k-1)} = \chi^2_{(m-1)(k-1)}$$

ya que, sin la restricción impuesta por la hipótesis nula hay  $mk - m$  parámetros  $p_{ij}$  en el problema (pues  $\sum_{j=1}^k p_{ij} = 1$  para cada  $i$ ) y, en cambio, bajo la hipótesis nula son  $k - 1$  los parámetros  $p_j$  (pues  $\sum_{j=1}^k p_j = 1$ ).

El resultado tiene una interpretación natural. Si la hipótesis nula es correcta, con tamaños muestrales grandes, no debe haber diferencias importantes entre:

- ▷ las probabilidades estimadas  $\hat{p}_{ij}$  de cada conjunto  $A_j$  con cada una de las muestras y
- ▷ la probabilidad estimada  $\hat{p}_{\cdot j}$  de cada conjunto  $A_j$  con las  $m$  muestras mezcladas;

de manera que  $-2 \log \Lambda$  tendrá un valor relativamente pequeño, debido sólo a la influencia del azar en la elección de cada muestra. Por el contrario, la heterogeneidad de las poblaciones provoca diferencias sistemáticas entre unas estimaciones y otras, razón por la cual el valor de  $-2 \log \Lambda$  tenderá a ser más grande de lo que es razonable esperar por meras causas de azar. La distribución  $\chi^2_{(m-1)(k-1)}$  dictamina con qué probabilidad debe esperarse, en el primer caso, cada valor de  $-2 \log \Lambda$ .

Una vez más, el test clásico de homogeneidad de poblaciones conserva la misma idea, pero difiere del test de razón de verosimilitudes, en la manera de valorar conjuntamente las diferencias entre las  $\hat{p}_{ij}$  y las  $\hat{p}_{\cdot j}$ . Concretamente, en lugar del estadístico  $-2 \log \Lambda$ , utiliza el estadístico:

$$D = \sum_{i=1}^m \sum_{j=1}^k \frac{n_i}{\hat{p}_{\cdot j}} (\hat{p}_{ij} - \hat{p}_{\cdot j})^2 = \sum_{i=1}^m \sum_{j=1}^k \frac{(n_{ij} - n_i n_{\cdot j}/n)^2}{n_i n_{\cdot j}/n}.$$

Se obtiene entonces el test con región crítica de la forma:  $\{D > c\}$  y la distribución asintótica de  $D$ , bajo la hipótesis nula (necesaria para determinar el nivel crítico  $c$ ) vuelve a ser  $\chi^2_{(m-1)(k-1)}$ .

Como argumento acerca de esta conclusión sobre la distribución asintótica de  $D$ , nótese que, si la hipótesis nula es correcta y las probabilidades  $p_j$  de cada conjunto  $A_j$  fuesen conocidas, el teorema 10.1 indicaría que

$$\sum_{j=1}^k \frac{(n_{ij} - n_i p_j)^2}{n_i p_j} \xrightarrow{d} \chi^2_{k-1}.$$

Por tanto, como las  $m$  muestras son independientes entre sí, sería

$$\sum_{i=1}^m \sum_{j=1}^k \frac{(n_{ij} - n_i p_j)^2}{n_i p_j} \xrightarrow{d} \chi^2_{m(k-1)}.$$

Sin embargo, las probabilidades  $p_j$  no son conocidas y han de ser sustituidas por su estimación de máxima verosimilitud,  $\hat{p}_{\cdot j}$ , dando lugar al estadístico  $D$ . Al estimar, por máxima verosimilitud,  $k - 1$  parámetros poblacionales (pues  $\sum_{j=1}^k p_j = 1$ ), el teorema 10.2 asegura que el número de grados de libertad de la distribución asintótica de  $D$  se reduce a  $(m - 1)(k - 1)$ .

En la práctica, para poder emplear la distribución asintótica de  $D$ , o de la razón de verosimilitudes, debe contarse con tamaños muestrales  $n_i$  apreciables (mayores que 20 por lo menos) y, sobre todo, efectuar la partición  $A_j$ , de forma que  $n_i n_{\cdot j}/n$  no sea nunca muy pequeño (superior a 2, suele ser la norma aceptada).

### Ejemplo 10.6

Se quiere estudiar si los distintos grupos sanguíneos se presentan con las mismas frecuencias en tres grupos étnicos diferentes. Para ello se analizaron un cierto número de individuos de cada raza, obteniéndose los resultados siguientes:

Raza	0	$A$	$B$	$AB$	Total
A	32	11	7	2	52
B	47	13	17	9	86
C	23	7	9	6	45
Total	102	31	33	17	183

Aún sin ningún tipo de teoría está claro que las cifras de las distintas filas de la tabla anterior no son comparables entre sí, puesto que se refieren a diferentes tamaños muestrales. En porcentajes, los datos se expresan:

Raza	0	$A$	$B$	$AB$	Total
A	61'54	21'15	13'46	3'85	100
B	54'65	15'12	19'77	10'46	100
C	51'11	15'56	20'00	13'33	100
Total	55'74	16'94	18'03	9'29	100

los cuales, si se refiriesen al conjunto de las poblaciones, revelarían diferencias apreciables, al menos entre la primera raza y las otras dos. Sin embargo

se trata de porcentajes muestrales, y el azar ha podido influir en las diferencias que se aprecian. La cuestión es entonces, si la disparidad es suficiente —en relación con los tamaños muestrales utilizados— para que haya una seguridad razonable de que los porcentajes de cada grupo sanguíneo no son los mismos en las tres razas.

El estadístico  $D$ , simplificado en la forma  $n \left( -1 + \sum_{i=1}^m \sum_{j=1}^k \frac{n_{ij}^2}{n_i n_{\cdot j}} \right)$ , vale

$$D = 183 \left( \frac{32^2}{52 \cdot 102} + \cdots + \frac{6^2}{45 \cdot 17} - 1 \right) = 4'691.$$

Por otra parte, si la hipótesis de homogeneidad es correcta, ha de esperarse un valor de  $D$  regido por la distribución  $\chi_6^2$  (puesto que  $m = 3$  y  $k = 4$ ). De acuerdo con la tabla de dicha distribución el valor obtenido no es nada raro, sino que

$$P\{\chi_6^2 > 4'691\} = 0'584.$$

Haría falta un valor de  $D$  superior a 10'64 para poder rechazar la hipótesis de homogeneidad, al nivel de significación 0'1.

La conclusión es la misma si se utiliza el test de razón de verosimilitudes, puesto que la distribución de  $-2 \log \Lambda$  es también  $\chi_6^2$  y

$$-2 \log \Lambda = 2 \left[ 32 \log \frac{0'6154}{0'5574} + \cdots + 6 \log \frac{0'1333}{0'0929} \right] = 5'079.$$

### Ejemplo 10.7

Para comprobar la eficacia del test  $\chi^2$  de homogeneidad se han simulado dos muestras aleatorias simples, de tamaño 50, de las distribuciones  $N(0, 1)$  y Cauchy (de densidad  $\pi^{-1}(1 + x^2)^{-1}$ ), cuya apariencia gráfica es similar. Las muestras obtenidas han sido:

$N(0, 1)$					Cauchy				
- 0'99	1'54	- 1'02	0'56	- 0'36	- 2'15	1'34	- 2'98	1'22	0'46
0'31	- 0'18	0'41	0'51	- 0'44	- 0'60	0'58	2'18	- 0'63	1'03
- 0'28	0'75	0'26	- 0'89	1'76	- 1'21	7'05	- 5'96	1'23	0'77
0'98	- 0'46	0'07	0'68	1'11	- 16'39	0'03	0'71	- 0'56	- 0'91
0'39	- 0'45	- 0'44	1'27	- 1'13	0'44	- 27'53	0'44	3'77	- 0'69
0'21	1'88	2'57	- 0'80	- 0'16	- 0'52	1'24	- 1'18	- 0'52	0'28
0'89	0'03	0'25	0'58	0'83	- 1'24	0'88	0'66	- 0'96	- 0'29
0'31	0'99	0'15	- 0'13	- 1'56	1'28	1'58	- 1'74	28'33	- 0'58
- 1'24	- 0'64	- 1'34	- 0'99	1'85	0'08	- 0'71	- 4'07	2'45	1'41
- 0'16	0'11	- 1'21	- 0'21	- 0'22	12'89	1'28	1'39	- 3'49	- 1'42

Tales muestras se pueden clasificar en los intervalos:

$A_j$	$n_{1j}$	$n_{2j}$	$n_{\cdot j}$
( $-\infty, -2]$	0	7	7
( $-2, -1'2]$	4	4	8
( $-1'2, -0'9]$	4	3	7
( $-0'9, -0'6]$	3	4	7
( $-0'6, -0'3]$	5	4	9
( $-0'3, 0]$	7	1	8
( $0, 0'3]$	7	3	10
( $0'3, 0'6]$	7	4	11
( $0'6, 0'9]$	4	4	8
( $0'9, 1'2]$	3	1	4
( $1'2, 2]$	5	9	14
( $2, \infty)$	1	6	7
Total	50	50	100

con los números de observaciones en las respectivas muestras consignados en la segunda y tercera columnas de la tabla.

El estadístico  $D$  toma el valor  $20'03$  y tiene distribución  $\chi^2_{11}$ . Puesto que

$$P\{\chi^2_{11} > 20'03\} = 0'045,$$

puede rechazarse la homogeneidad de ambas muestras con nivel crítico  $0'045$ . Puede observarse que, a pesar de la agrupación de los resultados (que no distingue las observaciones  $-27'53$  y  $-3'49$ ), el contraste da el resultado correcto.

## 10.4. Contraste $\chi^2$ de independencia

Supongamos que de  $n$  elementos de una población se han observado dos características  $X$  e  $Y$ , obteniéndose una muestra aleatoria simple bidimensional  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ . Sobre la base de dichas observaciones se desea contrastar si las características poblacionales  $X$  e  $Y$  son independientes o no.

Para ello se dividirá el conjunto de posibles valores de  $X$  en  $k$  conjuntos disjuntos:  $A_1, A_2, \dots, A_k$ ; mientras que el conjunto de posibles valores de  $Y$  será descompuesto en  $r$  conjuntos disjuntos:  $B_1, B_2, \dots, B_r$ . Al clasificar

los elementos de la muestra, aparecerá un cierto número de ellos,  $n_{ij}$ , en cada una de las  $k \times r$  clases así constituidas, dando lugar a una *tabla de contingencia* de la forma:

	$B_1$	$B_2$	$\dots$	$B_r$	Total
$A_1$	$n_{11}$	$n_{12}$	$\dots$	$n_{1r}$	$n_{1\cdot}$
$A_2$	$n_{21}$	$n_{22}$		$n_{2r}$	$n_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
$A_k$	$n_{k1}$	$n_{k2}$	$\dots$	$n_{kr}$	$n_{k\cdot}$
Total	$n_{\cdot 1}$	$n_{\cdot 2}$	$\dots$	$n_{\cdot r}$	$n$

Si  $p_{ij}$  es la probabilidad que la distribución poblacional asigna al suceso  $\{X \in A_i, Y \in B_j\}$ , la probabilidad de la clasificación reflejada en la tabla es

$$n! \prod_{i=1}^k \prod_{j=1}^r \frac{1}{n_{ij}!} p_{ij}^{n_{ij}},$$

que alcanza el máximo (con la restricción  $\sum_{i=1}^k \sum_{j=1}^r p_{ij} = 1$ ) en  $\hat{p}_{ij} = n_{ij}/n$ .

La hipótesis de independencia entre  $X$  e  $Y$  implica las igualdades:

$$p_{ij} = p_{i\cdot} p_{\cdot j} \quad \text{para cada } i, j,$$

donde  $p_{i\cdot}$  es la probabilidad poblacional de  $\{X \in A_i\}$  y  $p_{\cdot j}$  la probabilidad poblacional de  $\{Y \in B_j\}$ . Así que, en realidad, trataremos de contrastar la hipótesis  $H_0$  constituida por las igualdades anteriores.

Bajo  $H_0$ , la probabilidad de las observaciones resumidas en la tabla de contingencia es

$$n! \prod_{i=1}^k \prod_{j=1}^r \frac{1}{n_{ij}!} (p_{i\cdot} p_{\cdot j})^{n_{ij}} = \frac{n!}{\prod_{i=1}^k \prod_{j=1}^r n_{ij}!} \prod_{i=1}^k p_{i\cdot}^{n_{i\cdot}} \prod_{j=1}^r p_{\cdot j}^{n_{\cdot j}},$$

cuyo máximo (sujeto a  $\sum_{i=1}^k p_{i\cdot} = 1$  y  $\sum_{j=1}^r p_{\cdot j} = 1$ ) se alcanza en

$$\hat{p}_{i\cdot} = \frac{n_{i\cdot}}{n}, \quad \hat{p}_{\cdot j} = \frac{n_{\cdot j}}{n}.$$

La razón de verosimilitudes para el contraste de  $H_0$  es pues

$$\Lambda = \frac{\prod_{i=1}^k (n_{i\cdot}/n)^{n_{i\cdot}} \prod_{j=1}^r (n_{\cdot j}/n)^{n_{\cdot j}}}{\prod_{i=1}^k \prod_{j=1}^r (n_{ij}/n)^{n_{ij}}},$$

con lo cual

$$\begin{aligned}-2 \log \Lambda &= 2 \sum_{i=1}^k \sum_{j=1}^r n_{ij} \log \frac{n_{ij}}{n} - 2 \sum_{i=1}^k n_{i \cdot} \log \frac{n_{i \cdot}}{n} - 2 \sum_{j=1}^r n_{ \cdot j} \log \frac{n_{ \cdot j}}{n} \\&= 2 \sum_{i=1}^k \sum_{j=1}^r n_{ij} \left( \log \frac{n_{ij}}{n} - \log \frac{n_{i \cdot}}{n} - \log \frac{n_{ \cdot j}}{n} \right) \\&= 2 \sum_{i=1}^k \sum_{j=1}^r n_{ij} (\log \hat{p}_{ij} - \log \hat{p}_{i \cdot} \hat{p}_{\cdot j}).\end{aligned}$$

La región crítica del contraste de razón de verosimilitudes es pues

$$\{-2 \log \Lambda > k\}$$

y, de acuerdo con los resultados de la sección 9.2, si la hipótesis nula es correcta,  $-2 \log \Lambda$  tiene distribución  $\chi^2$  con  $kr - 1 - (k - 1) - (r - 1) = (k - 1)(r - 1)$  grados de libertad (ya que, sin la restricción impuesta por la hipótesis nula hay  $kr - 1$  parámetros; y, bajo la hipótesis nula, el número de parámetros es  $(k - 1) + (r - 1)$ ).

De nuevo,  $-2 \log \Lambda$  mide la discrepancia total entre las probabilidades  $\hat{p}_{ij}$ , estimadas para cada casilla sin tener en cuenta la hipótesis de independencia, y las probabilidades  $\hat{p}_{i \cdot} \hat{p}_{\cdot j}$ , estimadas aceptando la independencia. Si la hipótesis nula es correcta, el valor de  $-2 \log \Lambda$  será relativamente pequeño; mientras que, si la hipótesis nula es falsa, habrá diferencias mayores entre las  $\hat{p}_{ij}$  y los productos  $\hat{p}_{i \cdot} \hat{p}_{\cdot j}$ . La distribución  $\chi^2_{(k-1)(r-1)}$  indica el valor a partir del cual debe considerarse significativo (para cada nivel de significación dado) el resultado obtenido con  $-2 \log \Lambda$ .

El test clásico de independencia sustituye el estadístico  $-2 \log \Lambda$  por

$$D = \sum_{i=1}^k \sum_{j=1}^r \frac{n}{\hat{p}_{i \cdot} \hat{p}_{\cdot j}} (\hat{p}_{ij} - \hat{p}_{i \cdot} \hat{p}_{\cdot j})^2 = \sum_{i=1}^k \sum_{j=1}^r \frac{(n_{ij} - n_{i \cdot} n_{\cdot j}/n)^2}{n_{i \cdot} n_{\cdot j}/n}$$

Se obtiene así la región crítica:  $\{D > c\}$ , en la cual el nivel crítico  $c$  se determina haciendo uso de la distribución asintótica de  $D$ , bajo la hipótesis nula, que es  $\chi^2_{(k-1)(r-1)}$ .

De hecho, suponiendo  $H_0$  correcta y las probabilidades  $p_{i \cdot}$  y  $p_{\cdot j}$  conocidas, el teorema 10.1 indica que

$$\sum_{i=1}^k \sum_{j=1}^r \frac{(n_{ij} - np_{i \cdot} p_{\cdot j})^2}{np_{i \cdot} p_{\cdot j}} \xrightarrow{d} \chi^2_{kr-1}.$$

Al sustituir las  $k - 1$  probabilidades  $p_i$ , y las  $r - 1$  probabilidades  $p_{.j}$ , por sus estimaciones de máxima verosimilitud, el número de grados de libertad de la distribución asintótica de  $D$  se reduce a  $kr - 1 - (k - 1) - (r - 1) = (k - 1)(r - 1)$ .

Obsérvese que, aún correspondiendo a planteamientos diferenciados, el tratamiento de las tablas de contingencia es el mismo en el caso de independencia entre dos caracteres y en el de homogeneidad de varias muestras.

El contraste de independencia puede, sin embargo, generalizarse fácilmente para el caso de tres (o más) características poblacionales:  $X, Y, Z$ . Se contaría entonces con una tabla de contingencia tridimensional, cuya desviación de la hipótesis de independencia sería medida por

$$D = \sum_{i=1}^k \sum_{j=1}^r \sum_{l=1}^s \frac{(n_{ijl} - n_{i..}n_{.j.}n_{..l}/n)^2}{n_{i..}n_{.j.}n_{..l}/n}$$

con distribución asintótica  $\chi^2_{(k-1)(r-1)(s-1)}$ .

### Ejemplo 10.8

Para estudiar si el grupo sanguíneo de los individuos tiene relación con la predisposición a padecer diabetes, se han seleccionado al azar 400 sujetos de los que se ha determinado el grupo sanguíneo y el nivel de glucosa en sangre en idénticas condiciones experimentales. Clasificada la segunda medida en niveles bajo, medio y alto, los resultados han sido:

	Bajo	Medio	Alto	Total
0	137	86	35	258
A	42	23	11	76
B	19	17	7	43
AB	14	7	2	23
Total	212	133	55	400

La dependencia entre ambas características debe traducirse en un valor relativamente alto del estadístico  $D$ . Con los datos expresados en la tabla se obtiene:

$$D = 400 \left( \frac{137^2}{258 \cdot 212} + \cdots + \frac{2^2}{23 \cdot 55} - 1 \right) = 2'406.$$

Por otra parte,  $D$  tiene distribución  $\chi^2_6$  y la tabla de tal distribución indica que  $P\{\chi^2_6 > 2'204\} = 0'9$ , luego no puede concluirse que exista una relación entre el grupo sanguíneo y la concentración de glucosa en sangre.

## 10.5. Contraste de Kolmogorov-Smirnov de bondad del ajuste

La metodología de los test  $\chi^2$  discretiza las observaciones muestrales en conjuntos de una cierta partición, para comparar después el histograma de frecuencias observadas con el histograma de frecuencias que deberían obtenerse de acuerdo con la hipótesis a contrastar. De esta manera, en el caso de distribuciones poblacionales continuas, dichos tests no hacen el mejor uso posible de la información contenida en la muestra, puesto que ignoran el valor exacto de las observaciones. Con un tamaño muestral relativamente grande, que permita utilizar una partición fina del conjunto de observaciones, el defecto puede no ser muy acusado. Pero con un número pequeño de observaciones y, por tanto, con una partición grosera de ellas, la pérdida de información puede ser importante; además, en tal caso, la distribución exacta del estadístico de contraste puede estar alejada de la distribución  $\chi^2$  límite.

Así pues, en el contraste de bondad del ajuste a una distribución teórica unidimensional de tipo continuo, es a menudo preferible el uso de un procedimiento alternativo que, en lugar de estudiar la diferencia entre los histogramas de frecuencias, se basa en la comparación de la función de distribución muestral con la función de distribución teórica.

Recuérdese que la función de distribución muestral, asociada a una muestra aleatoria  $x_1, x_2, \dots, x_n$  es

$$F_n^*(x) = \frac{\text{número de } x_i \leq x}{n} = \begin{cases} 0 & \text{si } x \leq x_{(1)} \\ i/n & \text{si } x_{(i)} \leq x < x_{(i+1)} \\ 1 & \text{si } x \geq x_{(n)} \end{cases}$$

donde  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$  son los elementos de la muestra ordenada.

Si se trata de una muestra aleatoria simple de una población descrita por la variable aleatoria unidimensional  $X$ , con función de distribución  $F$ , el teorema de Glivenko-Cantelli (cf. Cap. 2) afirma que hay probabilidad uno de obtener una sucesión muestral con la cual  $F_n^*(x)$  converja a  $F(x)$ , uniformemente en  $x$ . Es decir que

$$\sup_{x \in \mathbb{R}} |F_n^*(x) - F(x)| \xrightarrow{\text{c.s.}} 0.$$

En el caso en que  $F$  sea desconocida y se desee contrastar la hipótesis  $H_0 : F = F_0$ , siendo  $F_0$  una distribución dada, parece sensato utilizar

$$\Delta_n = \sup_{x \in \mathbb{R}} |F_n^*(x) - F_0(x)|$$

como medida de la veracidad de la hipótesis, puesto que dicho estadístico tendrá, a la larga, un valor pequeño si  $H_0$  es correcta y un valor no despreciable ( $\sup_{x \in \mathbb{R}} |F(x) - F_0(x)|$ ) en el caso en que  $F \neq F_0$ .

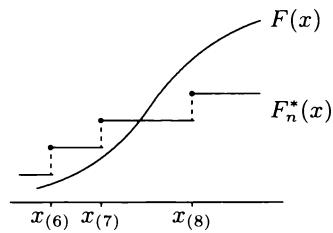
El test de Kolmogorov-Smirnov de bondad del ajuste, se basa entonces en el estadístico  $\Delta_n$  y, más concretamente, tiene como región crítica

$$\{\Delta_n > k\}.$$

No obstante, tal conclusión será de muy poca utilidad mientras no se conozca, bajo la hipótesis nula, la distribución del estadístico de contraste  $\Delta_n$ , que permita determinar el nivel crítico  $k$  a partir del cual sus valores son significativos de la existencia de una desviación de  $F$  respecto a  $F_0$ . A este respecto, una primera observación útil es:

**Lema:** Si  $F = F_0$  y  $F_0$  es continua, la distribución de  $\Delta_n$  no depende de  $F_0$ .

En efecto, gráficamente es claro que el valor de  $\Delta_n$  se alcanza cuando  $x$  coincide con alguno de los elementos de la muestra (o a la izquierda de uno de ellos). Más formalmente, llamando  $X_{(0)} = -\infty$  y  $X_{(n+1)} = +\infty$ , se tiene



$$\begin{aligned} \sup_{x \in \mathbb{R}} [F_n^*(x) - F_0(x)] &= \max_{0 \leq i \leq n} \sup_{X_{(i)} \leq x < X_{(i+1)}} \left[ \frac{i}{n} - F_0(x) \right] \\ &= \max_{1 \leq i \leq n} \left[ \frac{i}{n} - F_0(X_{(i)}) \right]. \end{aligned}$$

Pero  $F_0(X_{(1)}) \leq F_0(X_{(2)}) \leq \dots \leq F_0(X_{(n)})$  es una muestra ordenada de una uniforme<sup>3</sup> en  $(0,1)$ , que puede designarse por  $U_{(1)} \leq U_{(2)} \leq \dots \leq U_{(n)}$ . Luego  $\sup_{x \in \mathbb{R}} [F_n^*(x) - F_0(x)]$  tiene la misma que la distribución que

$$\max_{1 \leq i \leq n} \left[ \frac{i}{n} - U_{(i)} \right],$$

que no depende de  $F_0$ . De manera similar

$$\sup_{x \in \mathbb{R}} [F_0(x) - F_n^*(x)] = \max_{0 \leq i \leq n} \left[ F_0(X_{(i+1)}) - \frac{i}{n} \right] = \max_{1 \leq i \leq n} \left[ F_0(X_{(i)}) - \frac{i-1}{n} \right]$$

<sup>3</sup> Recuérdese que  $F_0(X)$  es uniforme en  $(0,1)$  si  $X$  tiene distribución continua  $F_0$ .

tiene la misma distribución que

$$\max_{1 \leq i \leq n} \left[ U_{(i)} - \frac{i-1}{n} \right]$$

que tampoco depende de  $F_0$ . Como

$$\Delta_n = \max \left( \sup_{x \in \mathbb{R}} [F_n^*(x) - F_0(x)], \sup_{x \in \mathbb{R}} [F_0(x) - F_n^*(x)] \right),$$

la misma conclusión vale para  $\Delta_n$ .

Según ello, para tratar de determinar la distribución de  $\Delta_n$ , puede suponerse que  $F_0$  es la distribución uniforme en  $(0,1)$ . Como la densidad de un muestra ordenada de una uniforme en  $(0,1)$  es

$$n! I_{\{0 < u_1 < u_2 < \dots < u_n < 1\}},$$

es sencillo dar una expresión integral de las distribución buscada:

$$\begin{aligned} P\{\Delta_n \leq z\} &= P \left\{ \max_{1 \leq i \leq n} \left[ \frac{i}{n} - U_{(i)} \right] \leq z, \max_{1 \leq i \leq n} \left[ U_{(i)} - \frac{i-1}{n} \right] \leq z \right\} \\ &= P \left\{ \frac{i}{n} - z \leq U_{(i)} \leq z + \frac{i-1}{n} \quad \text{para } i = 1, \dots, n \right\} \\ &= n! \int_{\frac{1}{n}-z}^z \int_{\frac{2}{n}-z}^{z+\frac{1}{n}} \cdots \int_{\frac{n-1}{n}-z}^{z+\frac{n-2}{n}} \int_{1-z}^{z+\frac{n-1}{n}} I_{\{0 < u_1 < u_2 < \dots < u_n < 1\}} du_1 du_2 \cdots du_{n-1} du_n. \end{aligned}$$

Tal expresión no proporciona un resultado explícito, pero permite tabular la distribución de  $\Delta_n$ , para pequeños valores de  $n$ . Dado el uso al que se destina, lo más útil es que la tabla proporcione, en función de  $n$  y  $\alpha$ , el valor de  $k$  tal que  $P\{\Delta_n > k\} = \alpha$ , como muestra la tabla 7 (debida originalmente a Massey).

Para valores grandes de  $n$  se puede utilizar la distribución asintótica proporcionada por el siguiente resultado, debido a Kolmogorov y Smirnov:

*Si  $F = F_0$  y  $F_0$  es continua, para cada  $z > 0$ , cuando  $n \rightarrow \infty$  se cumple*

$$P\{\sqrt{n}\Delta_n \leq z\} \longrightarrow \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 z^2}.$$

Dicha distribución límite de  $\sqrt{n}\Delta_n$  —conocida como distribución  $K$  de Kolmogorov— fue tabulada por Smirnov y aparece reflejada en la última fila de la tabla 7.

Algunos ejemplos servirán para mostrar como llevar a cabo los cálculos.

### Ejemplo 10.9

Con la muestra del ejemplo 10.2, una vez ordenada, puede construirse la siguiente la tabla cuya segunda columna da el número de observaciones acumuladas hasta cada valor muestral, de él se obtiene (dividiendo por el tamaño de la muestra) el valor de la función de distribución muestral, expresado en la tercera columna:

$x_{(i)}$	$i$	$F^*(x_{(i)})$	$F_0(x_{(i)})$		
10'12	1	0'025	0'0056	0'0194	0'0056
10'18	2	0'050	0'0164	0'0336	-0'0086
10'23	4	0'100	0'0359	0'0641	-0'0141
10'25	5	0'125	0'0478	0'0772	-0'0522
10'28	6	0'150	0'0712	0'0788	-0'0538
10'29	7	0'175	0'0807	0'0943	-0'0693
10'31	8	0'200	0'1026	0'0974	-0'0724
10'32	10	0'250	0'1151	0'1349	-0'0849
10'34	12	0'300	0'1431	0'1569	-0'1069
10'35	13	0'325	0'1587	0'1663	-0'1413
10'39	15	0'375	0'2317	0'1433	-0'0933
10'41	16	0'400	0'2743	0'1257	-0'1007
10'43	17	0'425	0'3204	0'1046	-0'0796
10'46	19	0'475	0'3949	0'0801	-0'0301
10'47	20	0'500	0'4207	0'0793	-0'0543
10'49	21	0'525	0'4734	0'0516	-0'0266
10'52	22	0'550	0'5530	-0'0030	0'0280
10'53	23	0'575	0'5793	-0'0043	0'0293
10'54	24	0'600	0'6051	-0'0051	0'0301
10'57	25	0'625	0'6796	-0'0546	0'0796
10'61	26	0'650	0'7683	-0'1183	0'1433
10'62	28	0'700	0'7881	-0'0881	0'1381
10'64	30	0'750	0'8247	-0'0747	0'1247
10'66	31	0'775	0'8569	-0'0819	0'1069
10'72	33	0'825	0'9288	-0'1038	0'1538
10'74	34	0'850	0'9452	-0'0952	0'1202
10'75	35	0'875	0'9522	-0'0772	0'1022
10'78	36	0'900	0'9690	-0'0690	0'0940
10'81	38	0'950	0'9806	-0'0306	0'0806
10'83	39	0'975	0'9861	-0'0111	0'0361
10'91	40	1	0'9969	0'0031	0'0219

La cuarta columna proporciona los valores de la función de distribución  $N(10'5, 0'15)$ ; las diferencias

$$F_n^*(x_{(i)}) - F_0(x_{(i)}) = \frac{i}{n} - F_0(x_{(i)}) \quad \text{y} \quad F_0(x_{(i)}) - F_n^*(x_{(i)}-) = F_0(x_{(i)}) - \frac{i-1}{n}$$

aparecen en la quinta y sexta columnas respectivamente. El máximo de la quinta columna es 0'1663, mientras que la sexta tiene como máximo 0'1538; así pues

$$\Delta_{40} = 0'1663.$$

La tabla 7 indica (para  $n = 40$ ) que la hipótesis de que la muestra proviene de la distribución  $N(10'5, 0'15)$  puede ser rechazada con nivel de significación 0'2, pero debe ser aceptada si el nivel de significación es ligeramente inferior a 0'2.

Pese a las críticas realizadas sobre el contraste  $\chi^2$ , en este ejemplo, el test de Kolmogorov-Smirnov no detecta falta de ajuste entre la muestra y la distribución  $N(10'5, 0'15)$ , mientras que el test  $\chi^2$ , usado en el ejemplo 10.2, sí lo hacía.

Puesto que todas las conclusiones acerca de la distribución de  $\Delta_n$  exigen que la distribución poblacional sea continua, los contrastes de Kolmogorov-Smirnov sólo pueden utilizarse para juzgar el ajuste de este tipo de distribuciones; en cambio tienen la ventaja, sobre los contrastes  $\chi^2$ , de poderse utilizar para cualquier tamaño muestral.

### Ejemplo 10.10

Un laboratorio farmacéutico afirma que uno de sus productos confiere inmunidad contra la picadura de insectos durante un tiempo exponencial de media 2'5 horas. Probado en 25 sujetos, en un ambiente con gran número de mosquitos, los instantes (en horas) en que recibieron la primera picadura fueron:

0'01	0'01	0'02	0'02	0'02	0'03	0'03	0'23	0'51
0'74	0'96	1'17	1'46	1'62	2'18	2'25	2'79	3'45
3'83	3'92	4'27	5'43	5'79	5'91	6'34		

La muestra, junto con el número de observaciones inferior a cada valor muestral, aparece en las dos primeras columnas de la siguiente tabla, permitiendo el cálculo de la función de distribución muestral.

$x_{(i)}$	$i$	$F^*$	$F_0$		
0'01	2	0'08	0'004	0'076	0'004
0'02	5	0'20	0'008	0'192	-0'072
0'03	7	0'28	0'012	0'268	-0'188
0'23	8	0'32	0'088	0'232	-0'192
0'51	9	0'36	0'185	0'175	-0'135
0'74	10	0'40	0'256	0'144	-0'104
0'96	11	0'44	0'319	0'121	-0'081
1'17	12	0'48	0'374	0'106	-0'066
1'46	13	0'52	0'442	0'078	-0'038
1'62	14	0'56	0'477	0'083	-0'043
2'18	15	0'60	0'582	0'018	0'022
2'25	16	0'64	0'593	0'047	-0'007
2'79	17	0'68	0'672	0'008	0'032
3'45	18	0'72	0'748	-0'028	0'068
3'83	19	0'76	0'784	-0'024	0'064
3'92	20	0'80	0'792	0'008	0'032
4'27	21	0'84	0'819	0'021	0'019
5'43	22	0'88	0'886	-0'006	0'046
5'79	23	0'92	0'901	0'019	0'021
5'91	24	0'96	0'906	0'054	-0'014
6'34	25	1	0'921	0'079	-0'039

La cuarta columna contiene la función de distribución:  $1 - e^{-0'4x}$ , que permite calcular las mismas diferencias del ejemplo anterior. Se obtiene, por tanto,  $\Delta_{25} = 0'268$  y la tabla 7 indica que la hipótesis de que la muestra proviene de la distribución indicada puede ser rechazada con nivel de significación 0'05. El que las mayores diferencias aparezcan al principio de la tabla sugiere que hay una cierta proporción de personas sobre las que el producto no es efectivo mientras que, quizás, el comportamiento de las demás queda bien reflejado por la distribución indicada.

El test  $\chi^2$  puede realizarse dividiendo el recorrido en 5 intervalos, de probabilidad  $1/5$ :

$A_i$	$n_i$	$np_i$
[0, 0'558)	9	5
[0'558, 1'277)	3	5
[1'277, 2'291)	4	5
[2'291, 4'024)	4	5
[4'024, $\infty$ )	5	5

Las frecuencias observadas en cada uno de los cinco intervalos, que aparecen en la segunda fila, confieren al estadístico de contraste el valor  $D = 4'4$  que, comparado con la distribución  $\chi^2_4$ , no permite rechazar la hipótesis de ajuste

ni siquiera con nivel de significación 0'3. Así pues, en este caso, el test  $\chi^2$  no detecta la falta de ajuste detectada por el contraste de Kolmogorov-Smirnov.

Es natural preguntarse si el contraste de Kolmogorov-Smirnov no puede adaptarse para juzgar el ajuste a una familia de distribuciones continuas, dependiente de ciertos parámetros  $\theta$ , de forma similar a como se modifican los tests  $\chi^2$  para contrastar hipótesis nulas compuestas. Por supuesto, habría que dar primero una estimación,  $\hat{\theta}$ , de los parámetros, para calcular después el estadístico:

$$\Delta_n = \sup_{x \in \mathbb{R}} |F_n^*(x) - F_{\hat{\theta}}(x)|.$$

Si se utilizan muestras distintas para la estimación de  $\theta$  y para obtener  $F_n^*(x)$ , los resultados anteriores son aplicables. Pero, si  $\hat{\theta}$  es función de la misma muestra con que se calcula  $F_n^*$ , la distribución de  $\Delta_n$  queda alterada, de manera teóricamente impredecible. Cualitativamente, el adaptar la distribución teórica a los datos —mediante la estimación de sus parámetros— tiende a disminuir considerablemente el valor de  $\Delta_n$ ; de manera que la utilización, en esta situación, de la tabla 7 favorece excesivamente a la hipótesis nula.

En el caso, sumamente importante en la práctica, de tratar de reconocer si una muestra aleatoria simple proviene de una población normal de media y varianza desconocidas, Lilliefors ha realizado estudios empíricos<sup>4</sup> acerca de la distribución de  $\Delta_n$ . La tabla 8 contiene sus conclusiones, proporcionando el nivel crítico para  $\Delta_n$ , según el valor de  $n$  y del nivel de significación.

### Ejemplo 10.11

La muestra utilizada en el ejemplo 10.2, daba lugar, en el ejemplo 10.5, a la estimación  $\mu = 10'502$ ,  $\sigma = 0'2025$ . Puede repetirse entonces el cálculo de  $\Delta_{40}$  para juzgar el ajuste a la distribución normal con tales parámetros, como muestra la tabla siguiente.

Su resultado es  $\Delta_{40} = 0'0986$ , con una reducción sensible respecto al valor hallado en el ejemplo 10.9, como era de esperar al haber estimado la distribución a partir de la muestra. La tabla 8 indica que dicho valor no es suficiente para rechazar la hipótesis de que los datos son normales, ni siquiera con un nivel de significación tan alto como 0'2.

---

<sup>4</sup> Basados en la simulación, para cada tamaño  $n$ , de una colección de muestras  $N(0, 1)$  y en la determinación, para cada una de ellas, de la medida  $\Delta_n$  del ajuste entre la muestra y la  $N(\bar{x}, s)$ .

$x_{(i)}$	$i$	$F^*(x_{(i)})$	$F(x_{(i)})$		
10'12	1	0'025	0'0296	-0'0046	0'0296
10'18	2	0'050	0'0559	-0'0059	0'0309
10'23	4	0'100	0'0896	0'0104	0'0396
10'25	5	0'125	0'1067	0'0183	0'0067
10'28	6	0'150	0'1365	0'0135	0'0115
10'29	7	0'175	0'1476	0'0274	-0'0024
10'31	8	0'200	0'1715	0'0285	-0'0035
10'32	10	0'250	0'1844	0'0656	-0'0156
10'34	12	0'300	0'2119	0'0881	-0'0381
10'35	13	0'325	0'2264	0'0986	-0'0736
10'39	15	0'375	0'2901	0'0849	-0'0349
10'41	16	0'400	0'3248	0'0752	-0'0502
10'43	17	0'425	0'3611	0'0639	-0'0389
10'46	19	0'475	0'4178	0'0572	-0'0072
10'47	20	0'500	0'4372	0'0628	-0'0378
10'49	21	0'525	0'4764	0'0486	-0'0236
10'52	22	0'550	0'5354	0'0146	0'0104
10'53	23	0'575	0'5550	0'0200	0'0050
10'54	24	0'600	0'5744	0'0256	-0'0006
10'57	25	0'625	0'6315	-0'0065	0'0315
10'61	26	0'650	0'7031	-0'0531	0'0781
10'62	28	0'700	0'7200	-0'0200	0'0700
10'64	30	0'750	0'7522	-0'0022	0'0522
10'66	31	0'775	0'7824	-0'0074	0'0324
10'72	33	0'825	0'8592	-0'0342	0'0842
10'74	34	0'850	0'8801	-0'0301	0'0551
10'75	35	0'875	0'8897	-0'0147	0'0397
10'78	36	0'900	0'9151	-0'0151	0'0401
10'81	38	0'950	0'9359	0'0141	0'0359
10'83	39	0'975	0'9474	0'0276	-0'0026
10'91	40	1	0'9780	0'0220	0'0030

## 10.6. Contraste de homogeneidad de Kolmogorov-Smirnov

Supongamos que se dispone de dos muestras aleatorias simples, independientes entre sí:  $X_1, X_2, \dots, X_n$  e  $Y_1, Y_2, \dots, Y_m$ , con las cuales se intenta decidir si las distribuciones poblacionales desconocidas,  $F_X$  y  $F_Y$ , coinciden. Esto es contrastar la hipótesis nula  $H_0 : F_X = F_Y$ .

Si las distribuciones involucradas son de tipo continuo, el contraste  $\chi^2$  adolece de los mismos defectos señalados en relación con el problema de bondad del ajuste: no se tiene en cuenta el valor exacto de las observaciones y sólo puede usarse con tamaños muestrales apreciables. En el caso en que

los tamaños muestrales sean pequeños, se hace imprescindible, por tanto, un método alternativo, que respete más fielmente las observaciones efectuadas.

En esta situación, la misma técnica de los estadísticos de Kolmogorov-Smirnov puede servir para comparar las funciones de distribución muestrales asociadas a ambas muestras. Es decir, considerando las distribuciones muestrales:

$$F_n^*(x) = \frac{1}{n} \sum_{i=1}^n I_{\{X_i \leq x\}} \quad \text{y} \quad G_m^*(x) = \frac{1}{m} \sum_{i=1}^m I_{\{Y_i \leq x\}},$$

se puede formar el estadístico

$$\Delta_{n,m} = \sup_{x \in \mathbb{R}} |F_n^*(x) - G_m^*(x)|,$$

que mide la diferencia global existente entre ellas y se denomina estadístico de Kolmogorov-Smirnov correspondiente a ambas muestras.

Si la hipótesis  $H_0$  es cierta, es probable que  $F_n^*(x)$  y  $G_m^*(x)$  sean próximas (puesto que ambas constituyen estimaciones de la distribución poblacional común) y que  $\Delta_{n,m}$  tenga, por tanto, un valor relativamente pequeño. En cambio, si  $F_X \neq F_Y$ , cada una de las dos distribuciones muestrales tenderá a aproximarse a su distribución poblacional y el valor de  $\Delta_{n,m}$  será más elevado. Esto conduce a contrastar la hipótesis  $H_0$  mediante un test con región crítica de la forma:

$$\{\Delta_{n,m} > k\}$$

y será necesario, por consiguiente, estudiar la distribución, bajo  $H_0$ , de  $\Delta_{n,m}$  para poder determinar el punto crítico  $k$  asociado a cada nivel de significación.

En primer lugar, hay que observar que la distribución, bajo  $H_0$ , de  $\Delta_{n,m}$  no depende del valor común,  $F$ , de  $F_X$  y  $F_Y$ , siempre que  $F$  sea continua. En efecto, cuando  $u$  recorre el intervalo  $(0,1)$ ,  $F^{-1}(u)$ <sup>5</sup> recorre el soporte de la distribución  $F$ , luego

$$\Delta_{n,m} = \sup_{u \in (0,1)} |F_n^*(F^{-1}(u)) - G_m^*(F^{-1}(u))|.$$

Ahora bien

$$F_n^*(F^{-1}(u)) = \frac{\text{número de } X_i \leq F^{-1}(u)}{n} = \frac{\text{número de } F(X_i) \leq u}{n}$$

---

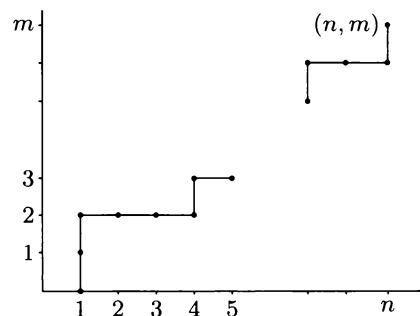
<sup>5</sup> Si  $F$  no es estrictamente creciente, se interpreta  $F^{-1}(u) = \inf\{x \in \mathbb{R} \mid F(x) \leq u\}$ .

es la función de distribución muestral asociada a la muestra  $F(X_i)$ , de una uniforme en  $(0,1)$ . Lo mismo ocurre con  $G_m^*(F^{-1}(u))$  y la distribución de  $\Delta_{n,m}$  es, por tanto, independiente de  $F$ .

La distribución de  $\Delta_{n,m}$  puede entonces obtenerse por métodos puramente combinatorios: Como el valor de  $\Delta_{n,m}$  ha de ser un múltiplo de  $1/nm$  (puesto que es la máxima diferencia entre una distribución con saltos  $1/n$  y otra de saltos  $1/m$ ), basta determinar

$$P\left\{\Delta_{n,m} < \frac{r}{nm}\right\} \quad \text{para } r = 1, 2, \dots, nm.$$

Si se ordenan conjuntamente, de forma creciente, ambas muestras, cada ordenación puede representarse como una poligonal que, empezando en el origen, se mueve una unidad a la derecha por cada observación de la primera muestra y una unidad hacia arriba por cada observación de la segunda; y acaba, por tanto, en el punto  $(n, m)$ .



Hay, desde luego,  $\binom{n+m}{n}$  poligonales posibles, todas ellas equiprobables (puesto que, bajo  $H_0$ , las  $n+m$  observaciones son independientes e idénticamente distribuidas). Por otra parte, las sucesivas coordenadas de los vértices proporcionan los sucesivos valores del par  $(nF_n^*(x), mG_m^*(x))$ ; así que

$$\Delta_{n,m} = \max \left| \frac{a}{n} - \frac{b}{m} \right| = \max \frac{|ma - nb|}{nm},$$

donde el máximo se calcula sobre los vértices  $(a, b)$  de la poligonal. Si se traza la recta  $mx - ny = 0$ , que une el origen con el punto  $(n, m)$ , la longitud del segmento vertical que une el punto  $(a, b)$  a la recta es

$$\left| \frac{ma}{n} - b \right| = m \left| \frac{a}{n} - \frac{b}{m} \right|,$$

de forma que  $m\Delta_{n,m}$  es la mayor distancia vertical de los vértices de la poligonal a la recta. Por consiguiente

$$P\left\{\Delta_{n,m} < \frac{r}{nm}\right\} = \binom{n+m}{n}^{-1} A_r(n, m),$$

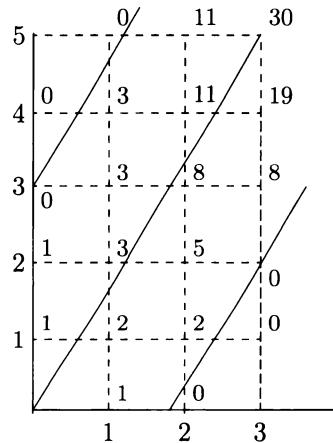
siendo  $A_r(n, m)$  el número de poligonales que alcanzan el punto  $(n, m)$  sin tener ningún vértice a distancia vertical de la recta superior o igual a  $r/n$ .

Fijadas las rectas  $mx - ny = \pm r/n$ , el número  $a_r(p, q)$  de poligonales que alcanzan el punto  $(p, q)$  sin tocar las rectas, se anula salvo en el interior de la banda, donde verifica

$$a_r(p, q) = a_r(p - 1, q) + a_r(p, q - 1),$$

lo cual permite determinarlos sucesivamente, hasta llegar a

$$a_r(n, m) = A_r(n, m).$$



En la figura, con  $n = 3$ ,  $m = 5$  y  $r = 9$ , se indican los sucesivos valores de  $a_9(p, q)$ , hasta llegar a  $A_9(3, 5) = 30$ . De donde

$$P\left\{\Delta_{3,5} < \frac{9}{15}\right\} = \binom{8}{3}^{-1} 30 = \frac{30}{56}.$$

Aunque sencillo, el recuento no es cómodo y, además, no interesa la probabilidad para un  $r$  dado, sino el valor de  $r/nm$  para una probabilidad dada. Los resultados, en este sentido, tabulados por Massey para valores pequeños de  $n$  y  $m$ , aparecen en la tabla 9. Para  $n = m$  los resultados figuran en la tabla 10.

La última fila de ambas tablas corresponde a la distribución asintótica, proporcionada por el siguiente resultado de Smirnov:

*Si  $F$  es continua, para cada  $z > 0$ , cuando  $n, m \rightarrow \infty$*

$$P\left\{\sqrt{\frac{nm}{n+m}}\Delta_{n,m} \leq z\right\} \rightarrow \sum_{i=-\infty}^{\infty} (-1)^i e^{-2i^2 z^2}.$$

### Ejemplo 10.12

Para aplicar el test de Kolmogorov-Smirnov a las muestras consideradas en el ejemplo 10.7 es necesario ordenar conjuntamente ambas muestras, sin perder la información de a qué muestra pertenece cada observación. Ello se lleva a cabo en la tabla siguiente, en la que figuran los valores de ambas distribuciones muestrales y su diferencia.

$x$	$F^*(x)$	$G^*(x)$		$x$	$F^*(x)$	$G^*(x)$		
-27'53	0	0'02	-0'02	0'21	0'56	0'50	0'06	
-16'39	0	0'04	-0'04	0'25	0'58	0'50	0'08	
-5'96	0	0'06	-0'06	0'26	0'60	0'50	0'10	
-4'07	0	0'08	-0'08	0'28	0'60	0'52	0'08	
-3'49	0	0'10	-0'10	0'31 <sub>2</sub>	0'64	0'52	0'12	
-2'98	0	0'12	-0'12	0'39	0'66	0'52	0'14	
-2'15	0	0'14	-0'14	0'41	0'68	0'52	0'16	
-1'74	0	0'16	-0'16	0'44 <sub>2</sub>	0'68	0'56	0'12	
-1'56	0'02	0'16	-0'14	0'46	0'68	0'58	0'10	
-1'42	0'02	0'18	-0'16	0'51	0'70	0'58	0'12	
-1'34	0'04	0'18	-0'14	0'56	0'72	0'58	0'14	
-1'24	-1'24	0'06	0'20	-0'14	0'58	0'74	0'60	0'14
-1'21	-1'21	0'08	0'22	-0'14	0'66	0'74	0'62	0'12
	-1'18	0'08	0'24	-0'16	0'68	0'76	0'62	0'14
-1'13		0'10	0'24	-0'14	0'71	0'76	0'64	0'12
-1'02		0'12	0'24	-0'12	0'75	0'78	0'64	0'14
-0'99 <sub>2</sub>		0'16	0'24	-0'08	0'77	0'78	0'66	0'12
	-0'96	0'16	0'26	-0'10	0'83	0'80	0'66	0'14
	-0'91	0'16	0'28	-0'12	0'88	0'80	0'68	0'12
-0'89		0'18	0'28	-0'10	0'89	0'82	0'68	0'14
-0'8		0'20	0'28	-0'08	0'98	0'84	0'68	0'16
	-0'71	0'20	0'30	-0'10	0'99	0'86	0'68	0'18
-0'64		0'22	0'32	-0'12	1'03	0'86	0'70	0'16
	-0'63	0'22	0'34	-0'12	1'11	0'88	0'70	0'18
	-0'6	0'22	0'36	-0'14	1'22	0'88	0'72	0'16
	-0'58	0'22	0'38	-0'16	1'23	0'88	0'74	0'14
	-0'56	0'22	0'40	-0'18	1'24	0'88	0'76	0'12
	-0'52 <sub>2</sub>	0'22	0'44	-0'22	1'27	0'90	0'76	0'14
-0'46		0'24	0'44	-0'20	1'28 <sub>2</sub>	0'90	0'80	0'10
-0'45		0'26	0'44	-0'18	1'34	0'90	0'82	0'08
-0'44 <sub>2</sub>		0'30	0'44	-0'14	1'39	0'90	0'84	0'06
-0'36		0'32	0'44	-0'12	1'41	0'90	0'86	0'04
	-0'29	0'32	0'46	-0'14	1'54	0'92	0'86	0'06
-0'28		0'34	0'46	-0'12	1'58	0'92	0'88	0'04
-0'22		0'36	0'46	-0'10	1'76	0'94	0'88	0'06
-0'21		0'38	0'46	-0'08	1'85	0'96	0'88	0'08
-0'18		0'40	0'46	-0'06	1'88	0'98	0'88	0'10
-0'16 <sub>2</sub>		0'44	0'46	-0'02	2'18	0'98	0'90	0'08
-0'13		0'46	0'46	0	2'45	0'98	0'92	0'06
0'03	0'03	0'48	0'48	0	2'57	1	0'92	0'08
0'07		0'50	0'48	0'02	3'77	1	0'94	0'06
	0'08	0'50	0'50	0	7'05	1	0'96	0'04
0'11		0'52	0'50	0'02	12'89	1	0'98	0'02
0'15		0'54	0'50	0'04	28'33	1	1	0

La máxima diferencia (en valor absoluto) entre ambas distribuciones muestrales es  $\Delta_{50,50} = 0'22$ , de forma que  $\sqrt{50}\Delta_{50,50} = 1'555$ . La tabla 10 indica que dicho valor es suficiente para rechazar la hipótesis de homogeneidad con nivel de significación  $0'2$ , pero no lo es para niveles de significación ligeramente más bajos. Nuevamente hay que concluir que, en este caso, el test  $\chi^2$  empleado en el ejemplo 10.7 resulta más sensible a la falta de homogeneidad que el contraste de Kolmogorov-Smirnov.

### Ejemplo 10.13

Las estrellas de una galaxia suelen clasificarse en dos poblaciones (I y II) atendiendo básicamente a su posición dentro de ella. Se quiere comprobar si la distribución de las luminosidades es la misma en ambas poblaciones y para ello se ha medido (en la escala logarítmica que habitualmente se emplea) la luminosidad de 8 estrellas de la población I y 12 de la población II. Los resultados obtenidos han sido

Población I:  $2'1, -0'4, 1'2, 1'5, -0'8, -1'6, 1'5, -2'5$ .

Población II:  $0'7, 0'2, -0'6, 1'1, -1'2, -1'4, -0'8, 1'3, 1'8, 0'2, 0'9, -0'8$ .

Ordenando conjuntamente ambas muestras resulta

$x$	$F^*(x)$	$G^*(x)$	
$-2'5$	$1/8$	$0$	$3/24$
$-1'6$	$2/8$	$0$	$6/24$
	$-1'4$	$2/8$	$4/24$
	$-1'2$	$2/8$	$2/24$
$-0'8$	$0'8_2$	$3/8$	$1/24$
	$-0'6$	$3/8$	$-1/24$
$-0'4$	$4/8$	$5/12$	$2/24$
	$0'2_2$	$4/8$	$-2/24$
	$0'7$	$4/8$	$-4/24$
	$0'9$	$4/8$	$-6/24$
	$1'2$	$4/8$	$-8/24$
$1'2$		$5/8$	$-5/24$
		$1'3$	$-7/24$
$1'5_2$		$7/8$	$-1/24$
		$1'8$	$-3/24$
$2'1$		$1$	$0$

Por consiguiente  $\Delta_{8,12} = 8/24$  y la tabla 9 muestra que dicho valor no es suficiente para rechazar que la luminosidad de ambas poblaciones tiene la

misma distribución, ni siquiera con nivel de significación 0'2. Por otra parte, los tamaños muestrales no son suficientemente grandes para poder aplicar el test  $\chi^2$  de homogeneidad.

## 10.7. Contrastes de posición

En ocasiones, acerca de la distribución unidimensional desconocida de una población, sólo interesa su posición sobre la recta porque se da por supuesto que las condiciones en que el fenómeno se observa sólo pueden transladar la distribución sin deformarla. Ejemplos de este tipo de situaciones son las siguientes:

- (A) Una empresa, que tradicionalmente comenzaba sus actividades diarias a las 9<sup>h</sup>, ha cambiado su horario para abrir a las 8<sup>h</sup> y se pregunta si ello habrá afectado los retrasos de sus empleados. Por supuesto, estos son aleatorios, variando de día en día y de un empleado a otro, pero es aceptable pensar que la forma de su distribución no ha variado; el temor es que se haya desplazado hacia la derecha, lo cual supone un incremento del tiempo perdido.
- (B) Cierta comunidad ha modificado la procedencia del agua destinada al consumo doméstico. Tras cierto tiempo, quiere comprobar si ello ha afectado a la concentración de sodio en la sangre de sus habitantes, en el sentido de que la distribución de dicha concentración se haya trasladado hacia uno u otro lado, mientras que la forma de la distribución se supone que no habrá variado apenas.
- (C) Se desea saber si las ventas en dos establecimientos de la misma cadena son análogas. Presumiblemente la forma de la distribución de las ventas diarias será similar para ambas, así que el objetivo es detectar si una está desplazada respecto a la otra.

Supuesto que no pueda admitirse la normalidad de las distribuciones poblacionales<sup>6</sup>, una manera usual de abordar este tipo de problemas de posición se basa en el uso de la mediana muestral para contrastar, en realidad, hipótesis acerca de las medianas poblacionales. Existen varios tests diseñados con tal fin.

<sup>6</sup>Pues, si lo fuesen, lo adecuado es aplicar los tests paramétricos sobre la media de una normal.

### 10.7.1. El test de los signos

Si se dispone de una muestra aleatoria simple, de tamaño  $n$ , de una población con distribución continua desconocida  $F$ , cuya mediana será designada por  $M$ , es posible contrastar la hipótesis nula  $H_0 : M = m_0$  frente a alguna de las alternativas  $M < m_0$ ,  $M > m_0$  ó  $M \neq m_0$ .

Puesto que  $F(M) = 1/2$ , si  $H_0$  es cierta, es de esperar que aproximadamente la mitad de las observaciones sean inferiores a  $m_0$  y la otra mitad superiores a  $m_0$ . Una posibilidad es, entonces, basar el contraste en el estadístico

$$T = \text{número de observaciones muestrales mayores que } m_0,$$

que tiene distribución binomial de parámetros  $n$  y  $1/2$ , si  $H_0$  es correcta. La denominación del test de los signos se debe a que el cálculo de  $T$  sólo requiere saber el número de signos positivos y negativos que resultan al restar  $m_0$  a las observaciones muestrales.

La región crítica del contraste depende naturalmente de la hipótesis alternativa. Por ejemplo, con  $H_1 : M > m_0$ , cuando  $H_0$  no sea cierta la muestra tendrá tendencia a presentar un exceso de valores superiores a  $m_0$ ; de manera que la región crítica adecuada es de la forma:

$$\{T \geq k\},$$

donde el punto crítico  $k$  ha de cumplir

$$\frac{1}{2^n} \sum_{j=k}^n \binom{n}{j} \leq \alpha,$$

para que el contraste tenga nivel de significación  $\alpha$ .

Análogamente, para  $H_1 : M < m_0$ , la región crítica oportuna será

$$\{T \leq k\} \quad \text{con} \quad \frac{1}{2^n} \sum_{j=0}^k \binom{n}{j} \leq \alpha.$$

En cambio, si  $H_1 : M \neq m_0$ , el test debe rechazar para valores anormalmente bajos o altos de  $T$ ; así que la región crítica debe ser

$$\{T \leq k\} \cup \{T \geq n - k\} \quad \text{siendo} \quad \frac{1}{2^n} \sum_{j=1}^k \binom{n}{j} + \frac{1}{2^n} \sum_{j=n-k}^n \binom{n}{j} \leq \alpha.$$

Si el tamaño muestral es apreciable ( $n \geq 20$ ) puede emplearse la aproximación normal a la distribución binomial para determinar los puntos críticos que proporcionen el test del tamaño  $\alpha$  elegido. Para tamaños muestrales pequeños, puesto que la distribución binomial es discreta, será preciso aleatorizar el test si se quiere conseguir un tamaño igual al nivel de significación.

### Ejemplo 10.14

En el supuesto (A) anterior, la empresa sabía que la mediana de los retrasos de sus empleados era de 5 minutos. Con el cambio de horario selecciona 12 empleados y observa, en determinado día, los siguientes retrasos (en minutos):

$$2'5 \quad 1'2 \quad 7 \quad 1'8 \quad 8'3 \quad 6'8 \quad 5'2 \quad 3'4 \quad 4'7 \quad 6'2 \quad 9'1 \quad 5'2.$$

El contraste que desea realizar la empresa es  $H_0 : M = 5$  (los retrasos no han variado respecto a su valor tradicional) frente a  $H_1 : M > 5$  (los retrasos han aumentado). El número de datos superiores a 5 es  $T = 7$  y la distribución binomial  $B(12, 1/2)$  indica que, si  $H_0$  es correcta,

$$P\{T \geq 7\} = 0'3871.$$

Luego no puede rechazar la hipótesis nula con ningún nivel de significación razonable.

Aunque, según la hipótesis de continuidad de la distribución poblacional, la probabilidad de que alguna observación muestral coincida con  $m_0$  es cero, la imprecisión de las medidas hace, en la práctica, que se observen datos iguales a  $m_0$ . Si no se puede incrementar la precisión, es aconsejable excluir tales valores muestrales, disminuyendo consecuentemente el tamaño de la muestra.

Nótese que el test es aplicable, con ligeras modificaciones, a contrastes sobre cuantiles cualesquiera de la distribución poblacional. De hecho, si  $x_p$  es el cuantil de orden  $p$ , para contrastar  $H_0 : x_p = m_0$ , basta emplear como distribución de  $T$  la distribución binomial  $B(n, 1 - p)$ .

El test de los signos puede emplearse también en el caso de datos apareados, aplicándolo a la variable diferencia:

### Ejemplo 10.15

En la misma situación del ejemplo anterior, la empresa seleccionó 16 de sus empleados y midió sus retrasos en dos días, antes y después del cambio de horario. Los resultados fueron:

$$\begin{array}{cccccccccc} 2'1/3'4 & 1'2/5'1 & 4'2/2'6 & 4'6/7'4 & 0'7/2'4 & 3'2/2'7 & 5'6/5'2 & 1'8/2'9 \\ 4'8/6'5 & 2'3/7'3 & 0'4/0'8 & 2'5/2'2 & 3'2/9'8 & 4'7/2'8 & 1'6/2'2 & 6'3/6'5 \end{array}$$

que se traducen en los siguientes aumentos de los retrasos

$$\begin{array}{cccccccc} 1'3 & 3'9 & -1'6 & 2'8 & 1'7 & -0'5 & -0'4 & 1'1 \\ 1'7 & 5'0 & 0'4 & -0'3 & 6'6 & -1'9 & 0'6 & 0'2 \end{array}$$

Si  $M$  es la mediana de la distribución de los incrementos, se puede contrastar, ahora, la hipótesis  $H_0 : M = 0$  (el retraso no se ha incrementado) frente a  $H_1 : M > 0$  (se ha producido un incremento del retraso). Para ello, el número de incrementos positivos es  $T = 11$  y la distribución binomial  $B(16, 1/2)$  proporciona

$$P\{T \geq 11\} = 0'105,$$

se puede entonces rechazar la hipótesis  $M = 0$  con nivel crítico  $0'105$ .

### 10.7.2. El test de Wilcoxon de los rangos signados

Con total desconocimiento acerca de la forma de la distribución teórica, el test de los signos tiene que ignorar las magnitudes de las diferencias entre las observaciones y  $m_0$  y considerar únicamente su signo. En el caso en que la distribución poblacional, además de continua, sea simétrica alrededor de su mediana, Wilcoxon propuso un test, para la misma hipótesis  $H_0 : M = m_0$ , que utiliza mejor los valores de las observaciones.

Si  $D_i = X_i - m_0$  son las diferencias entre las observaciones muestrales y el valor a contrastar para  $M$ , el método consiste en ordenar, en orden creciente, los valores absolutos  $|D_i|$  y anotar el rango (o lugar),  $r(|D_i|)$ , que cada uno ocupa en dicha ordenación. La suma de los rangos de las observaciones mayores que  $m_0$ :

$$T^+ = \sum_{i=1}^n r(|D_i|) I_{\{D_i > 0\}},$$

constituye el estadístico de contraste. Intuitivamente, si se cumple la hipótesis alternativa  $H_1 : M > m_0$ , es de esperar que haya más diferencias positivas que negativas y, a la vez, que aquellas tengan valores absolutos superiores a éstas. Por tanto, habrá que rechazar la hipótesis nula e inclinarse por la hipótesis alternativa, si  $T^+$  tiene un valor alto; es decir utilizar la región crítica  $\{T^+ > k_1\}$ .

Simétricamente, la región crítica  $\{T^+ < k_2\}$  es adecuada cuando la hipótesis alternativa es  $H_1 : M < m_0$ , puesto que entonces habrá tendencia a que el número de observaciones positivas sea menor y sus rangos bajos, lo cual dará un valor pequeño para  $T^+$ . Con  $H_1 : M \neq m_0$ , la región crítica indicada es  $\{T^+ < k'_2\} \cup \{T^+ > k'_1\}$ .

El primer paso para la obtención de la distribución de  $T^+$  —con la cual determinar los puntos críticos— es establecer:

**Lema:** Si  $H_0$  es cierta,  $T^+$  tiene la misma distribución que  $\sum_{r=1}^n rY_r$  siendo  $Y_r$  variables independientes, con  $P\{Y_r = 0\} = P\{Y_r = 1\} = 1/2$ .

En efecto, como la distribución de  $D_i$  es simétrica alrededor del origen, se tiene

$$P\{D_i > 0, |D_i| \leq x\} = P\{0 < D_i \leq x\} = \frac{1}{2}P\{|D_i| \leq x\} = P\{D_i > 0\}P\{|D_i| \leq x\},$$

lo cual indica que  $I_{\{D_i > 0\}}$  es independiente de  $|D_i|$ . Es independiente entonces de  $|D_1|, \dots, |D_n|$  y, por tanto, de  $r(|D_1|), \dots, r(|D_n|)$ . Ello significa que, después de ordenar los  $|D_i|$ , el término que ocupa el lugar  $r$  puede ser positivo o negativo con la misma probabilidad, independientemente de la ordenación que haya resultado.

De acuerdo con ello, si  $a_n(t)$  es el número de asignaciones de signos a los sucesivos rangos, que hacen  $T^+ = t$ , será

$$P\{T^+ = t\} = \frac{a_n(t)}{2^n},$$

lo cual permite calcular fácilmente, para  $n$  pequeño, la distribución de  $T^+$ .

Por ejemplo, con  $n = 4$ :

	1	-	+	-	-	-	+	+	+	-	-	-	-	+	+	+	-	+
	2	-	-	+	-	-	+	-	-	+	+	+	-	+	+	-	+	+
	3	-	-	-	+	-	-	+	-	+	-	+	+	-	+	+	+	+
	4	-	-	-	-	+	-	-	+	-	+	+	-	+	+	+	+	+
$T^+$	0	1	2	3	4	3	4	5	5	6	7	6	7	8	9	10		

de donde

$$P\{T^+ = t\} = \begin{cases} 1/16 & \text{si } t = 0, 1, 2, 8, 9, 10, \\ 2/16 & \text{si } t = 3, 4, 5, 6, 7. \end{cases}$$

La distribución de  $T^+$  siempre es simétrica alrededor de  $n(n + 1)/4$  pues el valor de  $T^+$  obtenido para una asignación de signos y para su opuesta suman  $\sum_{r=1}^n i = n(n + 1)/2$ .

La tabla 11 contiene los resultados de cálculos del tipo anterior, en forma de puntos críticos,  $k_1(\alpha)$ , con diversos niveles de significación y pequeños valores de  $n$ , para el contraste frente a la alternativa  $M > m_0$ . Si la alternativa fuese  $M < m_0$ , por simetría, habría que usar  $k_2(\alpha) = n(n+1)/2 - k_1(\alpha)$ . Mientras que con la alternativa  $M \neq m_0$  es  $k'_1(\alpha) = k_1(\alpha/2)$  y  $k'_2(\alpha) = n(n+1)/2 - k_1(\alpha/2)$ .

Desde luego  $E[T^+] = n(n+1)/4$  y, como  $V(Y_r) = 1/4$ , es

$$V(T^+) = \frac{1}{4} \sum_{i=1}^n r^2 = \frac{n(n+1)(2n+1)}{24};$$

así que, para  $n$  elevado (superior a 20), se puede emplear que  $T^+$  es aproximadamente  $N\left(n(n+1)/4, \sqrt{n(n+1)(2n+1)/24}\right)$ .

### Ejemplo 10.16

En el supuesto (B) anterior, se sabe que la distribución de la cantidad de sodio por unidad de volumen de sangre era simétrica, alrededor de 3'24 g., con el antiguo suministro de agua. Recientemente se han practicado 15 análisis, con los siguientes resultados (en gramos por unidad):

$$\begin{array}{cccccccc} 2'37 & 2'95 & 3'40 & 2'64 & 3'66 & 3'18 & 2'72 & 3'61 \\ 3'87 & 1'97 & 1'66 & 3'72 & 2'10 & 1'83 & 3'03 \end{array}$$

Las diferencias necesarias para aplicar el test de los rangos signados al contraste de  $M = 3'24$  frente a  $M \neq 3'24$ , son:

$$\begin{array}{cccccccc} -0'87_{11} & -0'29_4 & +0'16_2 & -0'69 & +0'42_6 & -0'06_1 & -0'52_8 & +0'37_5 \\ +0'63_{10} & -1'27_{13} & -1'58_{15} & +0'48_7 & -1'14_{12} & -1'41_{14} & -0'21_3 \end{array}$$

cuyos rangos, en la ordenación creciente de sus valores absolutos, se ha indicado junto a cada término. La suma de los rangos de los términos positivos es  $T^+ = 2 + 6 + 5 + 10 + 7 = 30$ . Con nivel de significación  $\alpha = 0'1$ , la tabla 11 indica que la hipótesis  $M = 3'24$  puede ser rechazada si  $T^+ > 89$  ó  $T^+ < 31$ . En cambio, para  $\alpha = 0'05$  la región crítica del test es  $T^+ > 94$  ó  $T^+ < 26$ . Los datos obtenidos permiten, pues, afirmar que la distribución de la cantidad de sodio ha variado, con un riesgo de error próximo al 10 %.

En principio, no deberían presentarse observaciones iguales a  $m_0$  ni empates en la ordenación de los  $|D_i|$ , puesto que la distribución poblacional es continua. Si, por imprecisiones en las medidas, tales fenómenos se presentan, se adopta la solución práctica de ignorar los que coincidan con  $m_0$ .

(reduciendo el tamaño muestral) y asignar el promedio de sus rangos a las rachas de  $|D_i|$  empatados.

La utilización del test de Wilcoxon en el caso de datos apareados se muestra en el ejemplo siguiente:

### Ejemplo 10.17

En 8 personas elegidas al azar se analizó el contenido de sodio por unidad de volumen, cierto tiempo antes y cierto tiempo después del cambio de suministro de agua, con los siguientes resultados:

$$(3'34, 2'58) \quad (2'82, 2'46) \quad (3'06, 3'50) \quad (2'30, 2'16) \\ (4'22, 3'78) \quad (3'55, 3'19) \quad (2'61, 2'94) \quad (2'83, 1'94)$$

Los incrementos han sido

$$-0'76 \quad -0'36 \quad +0'44 \quad -0'14 \quad -0'44 \quad -0'36 \quad +0'33 \quad -0'89 \\ (7) \quad (3'5) \quad (5'5) \quad (1) \quad (5'5) \quad (3'5) \quad (2) \quad (8)$$

con los rangos que se indican en la segunda fila. Ello permite utilizar el test de Wilcoxon para el contraste de  $M = 0$  frente a  $M \neq 0$ , acerca de la mediana  $M$  de la distribución de los incrementos (que será simétrica si lo era la distribución inicial).

El estadístico de contraste toma el valor  $T^+ = 7'5$  y la tabla 11 indica que, con nivel de significación  $0'1$ , la hipótesis  $M = 0$  sólo podría rechazarse si fuese  $T^+ > 30$  ó  $T^+ < 6$ .

Nótese que, si se conoce la mediana poblacional, el test de los rangos signados se convierte en un test sobre la hipótesis subyacente de que la distribución poblacional es simétrica alrededor de la mediana. Por tanto, con tamaños muestrales grandes (que aseguren que la mediana muestral es próxima a la poblacional), puede usarse el test, sustituyendo  $m_0$  por la mediana muestral, para contrastar la simetría de la distribución poblacional.

### 10.7.3. El test de la mediana

Los dos tests anteriores se refieren a la mediana de una única población, y hacen uso de una única muestra (en el caso de datos apareados la población y la muestra que interesan son las diferencias entre las parejas de datos). Sin embargo, con frecuencia se plantean situaciones en las cuales hay que comparar dos poblaciones continuas y tratar de detectar desplazamientos entre ambas distribuciones.

Se contará entonces con dos muestras aleatorias simples:  $x_1, \dots, x_n$  e  $y_1, \dots, y_m$ , correspondientes a cada población e independientes entre sí. Si se ordenan conjuntamente en orden creciente, la mediana,  $z$ , de la muestra combinada es el valor central, en el caso de que  $n + m$  sea impar, y el promedio de los dos valores centrales en el caso de que  $n + m$  sea par. Sea

$$T = \text{número de } x_i \text{ inferiores a } z.$$

Si ambas poblaciones tienen la misma distribución, es de esperar que  $T$  sea próximo a  $n/2$ . En cambio, si  $T$  resulta mucho mayor que  $n/2$ , es razonable suponer que la mediana  $M_x$  de la primera población es inferior a la de la segunda  $M_y$ ; mientras que si  $T$  es mucho menor que  $n/2$ , ello parece indicar que  $M_x$  es superior a  $M_y$ .

El test de la mediana para el contraste de la hipótesis nula  $H_0 : M_x = M_y$  se basa, entonces, en el estadístico  $T$  y utiliza las regiones críticas:

$$\{T \geq k\}, \quad \{T \leq k'\} \quad \text{ó} \quad \{T \leq k_1\} \cup \{T \geq k_2\},$$

según que la hipótesis alternativa sea

$$H_1 : M_x < M_y, \quad H_1 : M_x > M_y \quad \text{ó} \quad H_1 : M_x \neq M_y.$$

Si la distribución de ambas muestras es la misma, la distribución de  $T$  se puede hallar con facilidad. Puesto que las  $n + m$  observaciones son independientes e igualmente distribuidas, las  $\binom{n+m}{n}$  maneras de asignar  $n$  a la primera muestra (y las otras  $m$  a la segunda) son equiprobables. Si  $p$  es la parte entera de  $(n + m)/2$ , hay  $p$  de las  $n + m$  observaciones inferiores a  $z$  y será  $T = t$  en todas las asignaciones en las que resulten  $t$  de la primera muestra entre las  $p$  primeras y  $n - t$  entre las  $n + m - p$  últimas. Luego

$$P\{T = t\} = \frac{\binom{p}{t} \binom{n+m-p}{n-t}}{\binom{m+n}{n}},$$

donde  $t$  puede variar entre  $\max(0, p - m)$  y  $\min(n, p)$ . Se trata pues de una distribución hipergeométrica que puede aproximarse, si  $n$  y  $m$  son grandes, por una

$$N(n/2, \sqrt{nm/4(n+m)}).$$

**Ejemplo 10.18**

Las ventas de los establecimientos  $A$  y  $B$  fueron controladas durante 9 y 12 días respectivamente, con los siguientes resultados (en miles de pesetas):

A:	132'5	167'4	189'8	124'6	136'6	147'5	159'9	117'8	106'3
B:	97'4	108'2	114'1	86'3	101'8	122'6	78'3	136'2	89'5
	118'4	109'2	92'7						

La ordenación conjunta de ambas muestras (sin perder la procedencia de cada dato), figura al margen. La mediana de la muestra conjunta (que ocupa el lugar 11) es  $z = 117'8$  y hay un único término de la primera muestra inferior a  $z$ . Luego  $T = 1$ .

Para contrastar  $M_x = M_y$  frente a  $M_x > M_y$ , con nivel de significación  $\alpha$ , el test de la mediana utiliza la región crítica  $\{T \leq k'\}$  donde ha de ser

$$P\{T \leq k'\} = \sum_{t=0}^{k'} \frac{\binom{10}{t} \binom{11}{9-t}}{\binom{21}{9}} \leq \alpha.$$

Con  $k' = 1$  el nivel de significación resulta 0'0058, de forma que el resultado obtenido permite afirmar que  $M_x > M_y$  con gran seguridad.

El test  $\chi^2$  aplicado a la tabla de contingencia

	< 120	> 120	Tot.
A	2	7	9
B	10	2	12
Tot.	12	9	21

da un valor del estadístico de contraste  $D = 7'84$  que, comparado con la distribución  $\chi_1^2$ , permite descartar la homogeneidad de ambas muestras con nivel de significación inferior a 0'01. Pero, con los tamaños muestrales disponibles y la partición efectuada, el test  $\chi^2$  es seguramente menos fiable que el test de la mediana. Con tamaños muestrales grandes el test  $\chi^2$  es preferible si no hay constancia de que la forma de ambas distribuciones sea la misma, ya que el test de la mediana tiene tendencia a aceptar la homogeneidad si  $M_x = M_y$  aunque la forma de las distribuciones sea diferente.

Por la misma razón es preferible como test de homogeneidad el de Kolmogorov-Smirnov que, en este caso concreto, da  $\Delta_{9,12} = 0'6944$ , lo cual permite rechazar la hipótesis de igual distribución con nivel crítico inferior a  $0'01$ . De hecho, el estadístico  $T$  equivale a calcular la diferencia  $F_n^*(x) - G_m^*(x)$  exclusivamente en el punto  $z$ , mientras que  $\Delta_{n,m}$  proporciona la máxima diferencia y es, por tanto, más sensible a disparidades entre las distribuciones que no sean sólo debidas a diferencias de posición.

#### 10.7.4. El test de Mann-Whitney

Continuemos con el problema de detectar diferencias de posición entre dos poblaciones continuas, a partir de sendas muestras aleatoria simples:  $(x_1, \dots, x_n)$  e  $(y_1, \dots, y_m)$ . Con frecuencia se utiliza entonces el test de Mann-Whitney, basado en el estadístico

$$V = \sum_{i=1}^n \sum_{j=1}^m I_{\{x_i > y_j\}},$$

que suma el número de  $y_j$  que hay por debajo de cada  $x_i$  y toma, por tanto, valores entre 0 y  $nm$ . Un valor elevado de  $V$  indica que las  $x$  están desplazadas a la derecha de las  $y$ ; mientras que un resultado pequeño para  $V$  se produce cuando se da la situación inversa. Nótese, sin embargo, que un valor intermedio para  $V$  no significa forzosamente que las distribuciones  $F_X$  y  $F_Y$  de ambas poblaciones sean parecidas, sino que se produciría, por ejemplo, si ambas distribuciones se diferenciaseen principalmente por tener muy distintas dispersiones. De ahí que el test se use sólo en el supuesto de que no hay diferencias importantes entre la forma de  $F_X$  y  $F_Y$ , para contrastar  $H_0 : F_X = F_Y$  frente a alguna de las alternativas

$$H_1 : F_X < F_Y \quad H_1 : F_X > F_Y \quad \text{ó} \quad H_1 : (F_X < F_Y \text{ ó } F_X > F_Y)$$

En cada uno de los tres casos, lógicamente la región crítica adecuada es

$$\{V > k\} \quad \{V < k'\} \quad \text{ó} \quad \{V < k_1\} \cup \{V > k_2\}$$

Puesto que, bajo  $H_0$ , las  $\binom{n+m}{n}$  ordenaciones conjuntas de ambas muestras son equiprobables, la distribución de  $V$  puede determinarse por simple recuento del número,  $c_{n,m}(v)$ , de ordenaciones en las cuales la suma del número de  $y$  anteriores a cada  $x$  es igual a  $v$ .

El caso  $n = 2, m = 3$  figura al margen y da

	$V$
$xyyy$	0
$xyxyy$	1
$xyyxy$	2
$xyyyx$	3
$yxxyy$	2
$yxyxy$	3
$yxyyx$	4
$yyxx$	4
$yyxyx$	5
$yyyxx$	6

$$P\{V = v\} = \begin{cases} 1/10 & \text{si } v = 0, 1, 5, 6, \\ 2/10 & \text{si } v = 2, 3, 4. \end{cases}$$

También puede procederse recurrentemente, teniendo en cuenta que

$$c_{n,m}(v) = c_{n,m-1}(v) + c_{n-1,m}(v - m)$$

(ya que cada ordenación de  $n$  y  $m$  símbolos que dé lugar al valor  $v$  tendrá una  $y$  después de una ristra de  $n$  y  $m - 1$  símbolos que dé el valor  $v$ , o bien una  $x$  después de una ristra de  $n - 1$  y  $m$  símbolos que dé el valor  $v - m$ ).

Los resultados obtenidos por tales procedimientos figuran en la tabla 12, en forma de puntos críticos  $k$  (con los cuales  $P\{V > k\} \leq \alpha$ ) para el contraste frente a la alternativa  $F_X < F_Y$ . Como la distribución de  $V$  es simétrica alrededor de  $nm/2$  (pues cada ordenación y su opuesta dan valores de  $v$  que suman  $nm$ ), es  $P\{V < nm - k\} = P\{V > k\}$ , lo cual permite obtener los puntos críticos para el contraste frente a  $F_X > F_Y$ . En la tabla sólo figuran los resultados para  $m \geq n$  porque siempre se puede tomar como primera muestra la más pequeña.

Cuando  $n$  y  $m$  son grandes puede usarse el hecho de que, bajo  $H_0$ ,  $V$  es asintóticamente

$$N(nm/2, \sqrt{nm(n+m+1)/12}).$$

### Ejemplo 10.19

Con los datos del ejemplo 18, ordenados conjuntamente, resulta

$$V = 6 + 9 + 11 + 11 + 12 + 12 + 12 + 12 + 12 = 97$$

Pero  $V$  es aproximadamente  $N(54, 14'07)$ , luego

$$P\{V > 96\} \simeq P\left\{\frac{V - 54}{14'07} > 2'98\right\} = 0'0014$$

y el test de Mann-Whitney corrobora, con nivel de significación inferior a 0'005, que las ventas del establecimiento  $A$  son superiores a las de  $B$ .

## 10.8. Contrastes de independencia

Para contrastar la independencia entre dos características poblacionales continuas<sup>7</sup>  $X$  e  $Y$ , existen alternativas al test  $\chi^2$  estudiado en la sección 10.3.

### 10.8.1. El test $\tau$ de Kendall

Si  $(X_1, Y_1), \dots, (X_n, Y_n)$  es una muestra aleatoria simple de una distribución bidimensional,

$$\begin{aligned}\pi_+ &= P\{X_1 < X_2, Y_1 < Y_2\} + P\{X_1 > X_2, Y_1 > Y_2\} \\&= P\{(X_1 - X_2)(Y_1 - Y_2) > 0\}, \\ \pi_- &= P\{X_1 < X_2, Y_1 > Y_2\} + P\{X_1 > X_2, Y_1 < Y_2\} \\&= P\{(X_1 - X_2)(Y_1 - Y_2) < 0\},\end{aligned}$$

son parámetros poblacionales que miden respectivamente la probabilidad de que haya una asociación positiva entre las componentes (en el sentido de que valores superiores de  $X$  se presenten junto con valores mayores de  $Y$ ) y la probabilidad de que haya una asociación negativa (de forma que a valores superiores de  $X$  correspondan valores inferiores de  $Y$ ).

Por la continuidad de la distribución es  $P\{(X_1 - X_2)(Y_1 - Y_2) = 0\} = 0$ , así que  $\pi_- = 1 - \pi_+$ . Entonces

$$\tau = \pi_+ - \pi_- = 2\pi_+ - 1,$$

se denomina *coeficiente de asociación de Kendall* y mide —en cierta forma— la dependencia de las variables. Al menos, si  $X$  e  $Y$  son independientes, es

$$\begin{aligned}\pi_+ &= P\{X_1 < X_2\}P\{Y_1 < Y_2\} + P\{X_1 > X_2\}P\{Y_1 > Y_2\} \\&= P\{X_1 > X_2\}P\{Y_1 < Y_2\} + P\{X_1 < X_2\}P\{Y_1 > Y_2\} = \pi_-\end{aligned}$$

de forma que es  $\tau = 0$ . Aunque el recíproco no es cierto y podría ser  $\tau = 0$  sin que  $X$  e  $Y$  fuesen independientes.

Después de hacer efectivo el muestreo, se puede contrastar la independencia, mediante un estimador de  $\tau$  que nos indique si  $\tau$  es próximo a cero. Para ello, como

$$c_{ij} = \begin{cases} 1 & \text{si } (x_i - x_j)(y_i - y_j) > 0 \\ -1 & \text{si } (x_i - x_j)(y_i - y_j) < 0 \end{cases}$$

---

<sup>7</sup> Cuya distribución conjunta no sea normal, ya que en tal caso son preferibles los contrastes paramétricos sobre el coeficiente de correlación (véase ejercicios 9.4 y 9.5)

tiene media  $\tau$ , se puede formar el promedio de los diversos  $c_{ij}$ . Concretamente, dado que  $c_{ij} = c_{ji}$ , el coeficiente de asociación muestral se define como

$$T = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} c_{ij}.$$

Si  $P$  y  $N$  representan el número de  $c_{ij}$  positivos y negativos entre los  $\binom{n}{2}$  a considerar, es simplemente

$$T = \frac{2}{n(n-1)} (P - N).$$

Y, como debe ser  $P + N = \binom{n}{2}$ , resulta

$$T = \frac{4P}{n(n-1)} - 1.$$

Además,  $P$  se cuenta fácilmente con la muestra ordenada por su primera componente: es el número de parejas  $i < j$  para las cuales  $y_i < y_j$ .

Obviamente  $T$  puede valer entre  $-1$  y  $1$  y un valor de  $T$  no muy próximo a cero, parece indicar que es  $\tau \neq 0$  y que, por tanto, las variables  $X$  e  $Y$  no son independientes. La región crítica  $\{|T| > k\}$  sirve entonces para contrastar la independencia entre  $X$  e  $Y$ . (Si bien debería ser sustituida por  $\{T < k\}$  o  $\{T > k\}$  en el caso en que la hipótesis alternativa precise una correlación positiva o negativa entre las variables).

Para hallar la distribución de  $T$ , se puede observar que  $T$  sólo depende del orden en que queden las componentes  $y_i$ , cuando la muestra se ordena por las primeras componentes  $x_i$ . Es decir que el resultado para  $T$  es el mismo si la muestra, ordenada en orden creciente de su primera componente, se sustituye por

$$(1, r_1), (2, r_2), (3, r_3), \dots, (n, r_n),$$

donde cada  $r_i$  indica el rango (o lugar) que la  $y_i$  ocupa en la ordenación creciente de las segundas componentes. Si  $\tau = 0$ , cada una de las permutaciones  $r_1, r_2, \dots, r_n$  de los  $n$  primeros números tiene la misma probabilidad de aparecer, igual a  $1/n!$ ; así que

$$P\{P = p\} = \frac{d_n(p)}{n!},$$

siendo  $d_n(p)$  el número de permutaciones para las cuales es  $P = p$ .

Con  $n = 4$ , por ejemplo, las 24 permutaciones aparecen al margen; en cada una  $P$  cuenta el número de parejas que mantienen el orden natural y  $T$  es  $2P/6 - 1$ .

La distribución de  $T$  que se obtiene es

$$P\{T = t\} = \begin{cases} 1/24 & \text{si } t = -1, 1 \\ 3/24 & \text{si } t = -4/6, 4/6 \\ 5/24 & \text{si } t = -2/6, 2/6 \\ 6/24 & \text{si } t = 0. \end{cases}$$

Los resultados análogos, para  $n \leq 10$ , aparecen en la tabla 13, tabulados en forma de puntos críticos  $k$  tales que

$$P\{|T| > k\} \leq \alpha.$$

Para  $n > 10$ , puede utilizarse que  $T$  es aproximadamente

$$N\left(0, \sqrt{\frac{2(2n+5)}{9n(n-1)}}\right)$$

(véase Gibbons (1985), pág. 218).

	<i>P</i>	<i>T</i>
1234	6	1
1243	5	4/6
1324	5	4/6
1342	4	2/6
1423	4	2/6
1432	3	0
2134	5	4/6
2143	4	2/6
2314	4	2/6
2341	3	0
2413	3	0
2431	2	-2/6
3124	4	2/6
3142	3	0
3214	3	0
3241	2	-2/6
3412	2	-2/6
3421	1	-4/6
4123	3	0
4132	2	-2/6
4213	2	-2/6
4231	1	-4/6
4312	1	-4/6
4321	0	-1

### Ejemplo 10.20

En 10 empleados de una empresa se ha observado la distancia (en km.) de su domicilio a la sede de la empresa y el retraso (en min.) con el que llegaron al trabajo cierto día. Los resultados fueron

$$\begin{array}{cccccc} (3'2, 5'1) & (2'4, 3'6) & (1'9, 4'2) & (2'8, 6'3) & (1'2, 2'3) \\ (2'7, 3'4) & (4'0, 2'8) & (0'7, 3'2) & (6'1, 5'3) & (3'7, 3'7) \end{array}$$

Ordenada la muestra según la distancia, los retrasos asociados son

$$\begin{array}{cccccccccc} 3'2 & 2'3 & 4'2 & 3'6 & 3'4 & 6'3 & 5'1 & 3'7 & 2'8 & 5'3 \\ (3) & (1) & (7) & (5) & (4) & (10) & (8) & (6) & (2) & (9) \end{array}$$

cuyos rangos (en la ordenación de dichos valores de menor a mayor) se ha indicado debajo de cada uno. El recuento del número de parejas que conservan el orden proporciona

$$P = 7+8+3+4+4+0+1+1+1 = 29, \quad \text{con lo cual} \quad T = 13/45 = 0'288.$$

La tabla 13 muestra que tendría que ser  $T > 0'33$  para poder rechazar la hipótesis de independencia con nivel de significación  $0'1$ . Así que los datos obtenidos no permiten concluir que haya relación entre el retraso y la distancia del domicilio a la empresa.

Cuando la imprecisión de las medidas provoca que aparezcan observaciones con alguna de sus componentes repetida, la solución práctica es tomar el  $c_{ij}$  correspondiente igual a cero; así que sigue siendo  $T = 2(P - N)/n(n - 1)$ , pero ya no es cierto que  $P + N = n(n - 1)/2$ .

### 10.8.2. El test del coeficiente de correlación entre rangos de Spearman

Otra manera de proceder, en la misma situación del apartado anterior, consiste en sustituir la muestra  $(x_1, y_1), \dots, (x_n, y_n)$  por  $(a_1, b_1), \dots, (a_n, b_n)$  donde  $a_i$  es el rango que ocupa el término  $x_i$  una vez ordenadas las  $x$  de menor a mayor y  $b_i$  es el rango que ocupa  $y_i$  en la ordenación creciente de las  $y$ . Puede entonces calcularse el *coeficiente de correlación entre rangos*:

$$R_S = \frac{\sum_{i=1}^n (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_{i=1}^n (a_i - \bar{a})^2 \sum_{i=1}^n (b_i - \bar{b})^2}},$$

que mide —de manera algo más precisa que el estadístico  $T$  de Kendall— si hay dependencia entre el tamaño de la primera componente y el de la segunda<sup>8</sup>. La región crítica  $\{|R_S| > k\}$  sirve entonces para contrastar la independencia entre las variables.

Dado que  $\bar{a} = \bar{b} = 1/n \sum_{i=1}^n i = (n + 1)/2$  y

$$\sum_{i=1}^n (a_i - \bar{a})^2 = \sum_{i=1}^n (b_i - \bar{b})^2 = \sum_{i=1}^n i^2 - n \frac{(n + 1)^2}{4} = \frac{n(n^2 - 1)}{12},$$

se puede expresar

$$R_S = \frac{12}{n(n^2 - 1)} \sum_{i=1}^n (a_i - \bar{a})(b_i - \bar{b}).$$

---

<sup>8</sup> Naturalmente el coeficiente de correlación muestral,  $r$ , entre las  $x_i$  y las  $y_i$  es mejor medida de dependencia que  $R_S$ ; pero el problema es que no es posible determinar la distribución en el muestreo de  $r$  sin hipótesis específicas sobre la expresión funcional de la distribución conjunta de  $X$  e  $Y$ ; en cambio, la distribución de  $R_S$  está determinada, sin más que suponer que dicha distribución es continua.

El cálculo de  $R_S$  todavía se puede simplificar si la muestra se ordena según los valores de la primera componente. En efecto, si  $r_j$  representa el rango de la  $y$  que ocupa, ahora, la posición  $j$ , se tiene

$$\begin{aligned} U &= \sum_{j=1}^n (r_j - j)^2 = \sum_{i=1}^n (b_i - a_i)^2 = \sum_{i=1}^n (b_i - \bar{b} + \bar{a} - a_i)^2 \\ &= \frac{n(n^2 - 1)}{6} - 2 \sum_{i=1}^n (a_i - \bar{a})(b_i - \bar{b}) \\ &= \frac{n(n^2 - 1)}{6} - 2 \frac{n(n^2 - 1)}{12} R_S = \frac{n(n^2 - 1)}{6} (1 - R_S), \end{aligned}$$

de forma que

$$R_S = 1 - \frac{6U}{n(n^2 - 1)}.$$

Bajo la hipótesis de independencia, al igual que en el apartado anterior,  $(r_1, \dots, r_n)$  puede ser indistintamente cualquiera de las  $n!$  permutaciones de  $\{1, \dots, n\}$ . Bastará, pues, obtener el valor de  $U$  y de  $R_S$  con cada una de ellas, para conocer la distribución de  $R_S$ .

Con  $n = 4$ , los cálculos de  $U$  y  $R_S$  aparecen al margen, dando lugar a la distribución de  $R_S$ :

$$P\{R_S = r\} = \begin{cases} 1/24 & \text{si } r = -1, 1 \\ 3/24 & \text{si } r = -0'8, 0'8 \\ 1/24 & \text{si } r = -0'6, 0'6 \\ 4/24 & \text{si } r = -0'4, 0'4 \\ 2/24 & \text{si } r = -0'2, 0, 0'2. \end{cases}$$

Los puntos críticos del contraste aparecen en la tabla 14, para diversos niveles de significación y  $n \leq 10$ . Para  $n > 10$ , puede probarse que  $R_S$  es aproximadamente

$$N\left(0, \frac{1}{\sqrt{n-1}}\right).$$

	$U$	$R_S$
1234	0	1
1243	2	0'8
1324	2	0'8
1342	6	0'4
1423	6	0'4
1432	8	0'2
2134	2	0'8
2143	4	0'6
2314	6	0'4
2341	12	-0'2
2413	10	0
2431	14	-0'4
3124	6	0'4
3142	10	0
3214	8	0'2
3241	14	-0'4
3412	16	-0'6
3421	18	-0'8
4123	12	-0'2
4132	14	-0'4
4213	14	-0'4
4231	18	-4/6
4312	16	-4/6
4321	20	-1

**Ejemplo 10.21**

Con los datos del ejemplo 10.20, ordenados como allí se hizo, resulta

$$U = 2^2 + 1^2 + 4^2 + 1^2 + 1^2 + 4^2 + 1^2 + 2^2 + 7^2 + 1^2 = 94$$

y  $R_S = 1 - 6U/990 = 0'43$ . Según la tabla 14, dicho valor del coeficiente de correlación entre rangos no es suficiente para rechazar la independencia entre las variables ni siquiera con nivel de significación 0'1.

En el caso (teóricamente imposible) de que la muestra contenga empates, entre las  $x_i$  o entre las  $y_i$ , los rangos  $a_i$  y  $b_i$  de cada racha de observaciones empatadas se sustituyen por el promedio de sus rangos. El valor de  $R_S$  ha de ser calculado entonces mediante su expresión inicial, puesto que las simplificaciones efectuadas dejan de ser válidas.

**10.8.3. El test de rachas**

Un problema de independencia distinto de los anteriores, se plantea cuando existe duda de que una muestra sea realmente aleatoria simple; es decir, que las sucesivas observaciones de la población hayan sido efectuadas independientemente. Condiciones de muestreo sin las debidas garantías de aleatoriedad —por ejemplo, un fenómeno observado en días sucesivos, en lugar de en días elegidos al azar— pueden haber afectado a la independencia de las observaciones y dar al traste con la aplicación de todos los métodos basados en el muestreo aleatorio simple.

Si se trata de una variable que sólo puede presentar los valores 0 ó 1 y se presentan las muestras

00000000111111	6	10101010101010
----------------	---	----------------

no cabe duda de que tienen un aspecto poco aleatorio y que cabe dudar de que se hayan observado en condiciones aleatorias. En cambio, la sospecha no surgiría, por ejemplo, ante las muestras:

10011100110001	6	01110010000111
----------------	---	----------------

Ante el desconocimiento de la distribución poblacional —en este caso, de las proporciones de 0 y 1 que deberían presentarse— el único rasgo por el que se puede diferenciar la poca aleatoriedad de las primeras muestras respecto de las segundas es por la colocación de los dígitos y, más exactamente, por el número,  $R$ , de rachas de cada uno que haya presentes. Así, en la primera

es  $R = 2$ , en la segunda  $R = 14$  y, en cambio, es  $R = 7$  y  $R = 6$  en las dos últimas.

Naturalmente son muchas más las secuencias con un número intermedio de rachas que con un número extremo, de forma que una secuencia elegida al azar contendrá, muy probablemente, un número intermedio de rachas. Regiones críticas de la forma

$$\{R < k_1\} \cup \{R > k_2\}$$

sirven entonces como test para el contraste de la hipótesis de que la muestra es aleatoria simple.

Supuesto que hay en la muestra  $n$  ceros y  $m$  unos, la distribución de  $R$ , bajo  $H_0$ , se puede hallar explícitamente:

$$\left. \begin{aligned} P\{R = 2r\} &= 2 \frac{\binom{n-1}{r-1} \binom{m-1}{r-1}}{\binom{n+m}{n}} \\ P\{R = 2r+1\} &= \frac{\binom{n-1}{r-1} \binom{m-1}{r} + \binom{n-1}{r} \binom{m-1}{r-1}}{\binom{n+m}{n}} \end{aligned} \right\} \text{para } r \leq \min(n, m).$$

En efecto,  $n$  ceros y  $m$  unos pueden ordenarse de  $(n+m)!$  maneras, igualmente probables. Para formar secuencias con  $r$  rachas de ceros y  $r$  rachas de unos, los 0 se pueden ordenar de  $n!$  formas, dividiéndolos luego en  $r$  grupos, mediante la elección de  $r-1$  de los  $n-1$  huecos existentes entre ellos. Análogamente, los unos pueden ser ordenados de  $m!$  maneras y divididos en  $r$  grupos de  $\binom{m-1}{r-1}$  formas. Después hay que intercalar los grupos formados, empezando con el primer grupo de ceros, o bien con el primer grupo de unos. En total, la probabilidad de que haya  $r$  rachas de cada tipo es

$$\frac{n! \binom{n-1}{r-1} m! \binom{m-1}{r-1} 2}{(n+m)!} = 2 \frac{\binom{n-1}{r-1} \binom{m-1}{r-1}}{\binom{n+m}{n}}.$$

De manera similar se calcula la probabilidad de que haya  $r$  rachas de ceros y  $r+1$  rachas de unos y la probabilidad de que haya  $r+1$  rachas de ceros y  $r$  rachas de unos, las cuales dan lugar a los dos sumandos de la segunda expresión. (Naturalmente, el número de rachas de un tipo y otro se diferencian, a lo sumo, en 1).

Si  $n$  y  $m$  son grandes (superiores a 10), puede probarse que  $R$  es aproximadamente

$$N \left( \frac{2nm}{(n+m)} + 1, \sqrt{\frac{2nm(2nm-n-m)}{(n+m)^2(n+m-1)}} \right).$$

**Ejemplo 10.22**

Al extraer 19 bolas con reemplazamiento de una bolsa con bolas blancas y negras, se ha obtenido el resultado

$$BBBBNNNBBB BBBB NN$$

que muestra  $R = 4$  rachas. Puesto que hay 12 blancas y 5 negras, el número de rachas podría haber sido entre 2 y 11 y la fórmula anterior permite calcular la probabilidad de cada resultado:

2	3	4	5	6	7	8	9	10	11
0'0003	0'002	0'014	0'046	0'107	0'195	0'213	0'24	0'107	0'075

Incluyendo las probabilidades de menor a mayor, se observa que  $\{R \leq 4\}$  es la región crítica con tamaño  $\alpha = 0'0169$ ; con tamaño  $\alpha = 0'0631$  se podría rechazar para  $R \leq 5$  y para  $\alpha = 0'1377$  se obtendría la región crítica  $\{R \leq 5\} \cup \{R = 11\}$ . Por tanto, con el resultado obtenido, hay una seguridad alta de que la experiencia no se realizó en las condiciones adecuadas; quizás no se agitó convenientemente la bolsa antes de cada extracción y cada una se vio influenciada por el resultado de la anterior.

Existen diversas formas de adaptar el test de rachas al caso de una muestra continua  $(x_1, x_2, \dots, x_k)$ :

- ▷ Se puede por ejemplo elegir un promedio de los valores muestrales (como  $\bar{x}$  o, más habitualmente, la mediana) y reemplazar cada observación por  $-$  ó  $+$ , según que su valor sea inferior o superior a tal promedio. Después basta aplicar el test a la secuencia de signos obtenidos para decidir si el número de rachas es significativo de que las condiciones de experimentación han hecho que exista dependencia entre las observaciones. (Si se elige la mediana, será naturalmente  $n = m$ , después de prescindir de ella en el caso de que haya un número impar de observaciones).
- ▷ También se puede sustituir cada valor muestral (excepto el primero) por  $+$  ó  $-$  según que sea mayor o menor que el anterior. El test se aplica de forma idéntica para detectar si ha habido períodos de observación con tendencias de crecimiento o decrecimiento, no atribuibles al azar.

**Ejemplo 10.23**

Los 12 datos de ventas del establecimiento  $B$ , en el ejemplo 10.18, tienen como mediana 105. Los términos de la muestra original (no de la reordenada,

¡claro está!), comparados con la mediana dan lugar a la secuencia de signos

$$- + + - - + - + - + + -$$

con  $R = 9$  rachas. Con  $n = m = 6$  la distribución de  $R$  es simétrica entre 2 y 12:

2 y 12	3 y 11	4 y 10	5 y 9	6 y 8	7
0'002	0'011	0'054	0'011	0'216	0'216

La región crítica  $\{R \leq 4\} \cup \{R \geq 10\}$  tendría tamaño 0'134, de forma que, con  $R = 9$ , no puede afirmarse que tomar los datos en días consecutivos haya afectado a la independencia de la muestra.

Los crecimientos y decrecimientos de la muestra son:

$$+ + - + + - + - + - -$$

En este caso  $n = 6$  y  $m = 5$  de forma que el número máximo de rachas sería 11 y tampoco las  $R = 8$  rachas observadas permiten descartar la independencia, puesto que la distribución de  $R$  es

2	3	4	5	6	7	8	9	10	11
0'004	0'019	0'086	0'151	0'259	0'216	0'173	0'065	0'022	0'002

El test de rachas podría aplicarse también a los contrastes de posición correspondientes a dos muestras, aunque da un resultado mucho más pobre que el test de Mann-Whitney. Basta ordenar ambas muestras conjuntamente y sustituir los elementos de la primera por 0 y los de la segunda por 1. Así, con la ordenación de las muestras realizada en ejemplo 10.18, es  $R = 8$ .

Si ambas muestras proceden de la misma distribución poblacional, sus valores deberían estar entremezclados, dando lugar a un número elevado de rachas. En cambio, si existe una diferencia de posición entre las distribuciones poblacionales  $F_X$  y  $F_Y$ , los ceros tendrán tendencia a estar en un extremo y los unos en el otro, dando lugar a un número de rachas más reducido. Luego,  $\{R \leq k\}$  puede ser utilizado como región crítica para rechazar la hipótesis  $F_X = F_Y$  frente a la alternativa de que ambas distribuciones se diferencian en su posición.

La distribución de  $R$  sirve, como siempre, para hallar el punto crítico adecuado a cada nivel de significación fijado. Por ejemplo, con  $n = 9$  y  $m = 12$  como en el ejemplo 10.18, es

$$P\{R \leq 8\} = 2 \sum_{r=1}^4 \frac{\binom{8}{r-1} \binom{11}{r-1}}{\binom{21}{9}} + \sum_{r=1}^3 \frac{\binom{8}{r-1} \binom{11}{r} + \binom{8}{r} \binom{11}{r-1}}{\binom{21}{9}} = 0'102,$$

lo cual sólo permite rechazar la igualdad en distribución de las ventas de los dos establecimientos con nivel crítico 0'102.

## Ejercicios

### Ejercicio 10.1

*El número de revoluciones (en millones) de determinadas piezas mecánicas, antes de fallar, tiene en las condiciones habituales una distribución con función de densidad*

$$f(x) = \frac{\beta\alpha^\beta}{x^{\beta+1}} \quad x > \alpha > 0, \quad \beta > 0,$$

donde experiencias previas han determinado como estimaciones:  $\alpha = 2'4$  y  $\beta = 5'5$ .

Después de un cambio en el método de lubricación de las piezas, se observó el comportamiento de 20 de ellas, obteniéndose los siguientes números de revoluciones (en millones) antes del fallo:

2'52	2'54	2'55	2'56	2'57	2'60	2'64	2'65	2'69	2'71
2'76	2'78	2'83	2'87	2'95	3'03	3'07	3'15	3'21	3'60

¿Puede afirmarse que los datos siguen manteniendo la misma distribución teórica? (a) si se emplea el test  $\chi^2$ , (b) si se emplea el test de Kolmogorov-Smirnov. En caso de que tal hipótesis resulte rechazada, ¿puede afirmarse que la distribución es otra de la misma familia paramétrica?

### Ejercicio 10.2

Una afección de la glándula tiroides ha sido investigada en una cierta región durante los años ochenta. El número de casos observados desde Junio de 1986 a Mayo de 1989 vienen dados en la siguiente tabla

	Mes											
	E	F	M	A	M	J	J	A	S	O	N	D
1986						6	9	8	6	8	11	8
1987	5	4	4	2	1	8	8	6	2	2	1	2
1988	7	8	3	1	2	7	7	6	5	5	3	5
1989	1	2	1	1	2							

Se quiere investigar si existe o no alguna periodicidad en dicha enfermedad, contrastando: (a) Si pueden considerarse homogéneas las tres temporadas durante las cuales se recogieron los datos. (b) Si los casos se presentan con variaciones estacionales.

### Ejercicio 10.3

Se quiere averiguar si existe independencia entre las calificaciones en dos asignaturas *A* y *B* de un mismo curso.

Para ello, se anotaron los resultados de 1000 alumnos que cursaron ambas disciplinas, obteniéndose la siguiente tabla de contingencia, en donde aparecen por filas las notas de la asignatura *A*.

	0	1	2	3	4	5	6	7	8	9	10
0	9	9	8	5	4	2	3	2	4	2	0
1	8	16	9	8	5	7	3	2	4	2	3
2	1	10	13	15	12	10	5	6	3	3	0
3	3	19	23	16	19	16	12	9	8	7	7
4	4	7	8	12	11	15	12	12	11	9	9
5	2	8	13	16	18	20	13	10	9	9	9
6	1	13	15	10	9	9	10	9	8	15	6
7	1	2	3	9	7	13	4	9	5	6	1
8	2	4	4	9	15	17	22	16	8	9	11
9	1	4	3	4	4	8	10	7	9	9	11
10	0	2	1	4	8	7	15	9	10	13	10

¿Pueden considerarse independientes ambas variables? ¿Existen diferencias significativas entre las distribuciones de los resultados de ambas asignaturas?

### Ejercicio 10.4

Se quiere averiguar si los enfermos de “hiperparatiroidismo” presentan una concentración de estroncio en el fluido extracecular distinta de lo normal.

Para ello se seleccionaron al azar 7 pacientes con dicha enfermedad y 10 personas aparentemente normales. Sus concentraciones de estroncio (en %) fueron:

Grupo I: 12'5 13'8 16'3 20'2 22'1 13'5 10'5

Grupo II: 8'1 9'3 12'7 11'6 13'7 16'2 15'9 10'3 11'1 13'1

Utilizar el test de Kolmogorov-Smirnov y el test de la mediana para contrastar si existen diferencias entre las concentraciones de ambas poblaciones.

**Ejercicio 10.5**

Se está investigando si dos regiones agrícolas de Puerto Rico, SAN JUAN y ARECIBO, son equivalentes respecto al número de granjas por milla cuadrada. Para ello se seleccionaron al azar varios municipios de cada una de las dos regiones, obteniendo los siguientes resultados:

Región de ARECIBO		Región de SAN JUAN	
Arecibo	4'5	Barranquitas	16'1
Barceloneta/Florida	5'6	Bayamon/Catano	5'2
Camuy	12'1	Carolina	5'5
Ciales	8'5	Ceiba	3'0
Hatillo	8'9	Comercio	14'9
Lares	17'9	Corozal	13'7
Morovis	12'9	Guaynabo	7'5
Quebradillas	10'7	Loiza/Canovanas	3'8
Utuado	18'3	Luquillo	4'2
Vega Baja	4'8	Naranjito	18'6
		Rio Grande	2'8
		San Juan/Rio Piedras	2'7
		Toa Baja	0'7
		Trujillo Alto	9'2
		Vega Alta	6'6

Utilizar el test de Kolmogorov-Smirnov y el de Mann-Whitney para contrastar si existen diferencias entre ambas regiones.

**Ejercicio 10.6**

El peso atómico de 12 muestras de plomo, procedentes de desintegraciones radioactivas, ha sido determinado por dos procedimientos diferentes, obteniéndose como resultados:

I :	206'34	207'42	207'25	206'81	206'53	206'56
	207'38	206'90	207'32	206'72	207'24	207'67
II :	206'37	207'38	207'33	206'69	206'48	206'52
	207'31	206'94	207'21	206'58	207'18	207'65

donde las muestras figuran en el mismo orden en ambos casos.

Utilizar el test de los signos y el test de los rangos signados para contrastar si existe diferencia significativa entre los dos procedimientos de determinación. ¿Bajo qué condiciones es adecuado utilizar cada uno de los tests?

**Ejercicio 10.7**

Un hecho fundamental en la teoría del “Big Bang” es la relación existente entre la velocidad  $V$  de recesión de una galaxia, (medida por el corrimiento

hacia el rojo de la luz procedente de ella) y su distancia  $d$  a La Tierra (deducida de la luminosidad aparente de una cefeida, cuya luminosidad absoluta es función de su periodo de pulsación).

Esta relación,  $V = H_0 \cdot d$ , conocida como ley de Hubble, fue establecida por Edwin Hubble en 1929 observando la velocidad (en km/s) y la magnitud aparente (usualmente medida en escala logarítmica) de un cierto número de galaxias y de cúmulos galácticos; parte de los resultados eran:

Galaxia	Velocidad	Magnitud
Virgo	1259	10'4
Pegaso	3687	12'9
Perseo	5011	13'5
Cabellera de Berenice	7356	14'2
Osa Mayor nº 1	14678	16'1
Leo	18478	16'2
Corona Boreal	21544	16'5
Géminis	19952	16'3
Boyero	39811	17'9
Osa Mayor nº 2	38312	17'7

Emplear el test de Kendall y el test de Spearman para contrastar si existe dependencia significativa entre ambas variables.

### Ejercicio 10.8

Las 100 primeras cifras decimales del número  $\pi$  son

$$\begin{aligned} \pi = & 3'14159265358979323846264338327950288419716939937510 \\ & 58209749445923078164062862089986280348253421170679 \end{aligned}$$

Estudiar si el resultado tiene las propiedades de una secuencia de cifras elegida al azar.

### Ejercicio 10.9

Recientemente ha adquirido gran auge el estudio de sistemas “caóticos”: que se comportan de forma determinística, pero dan lugar a resultados “imprevisibles”. Uno muy simple, que se utiliza como ejemplo de lo que ello quiere decir, es la ecuación recurrente

$$x_{n+1} = rx_n(1 - x_n)x_0 = 0'5.$$

Con un programa de ordenador de una línea se puede observar que, para  $r = 2'6$  la sucesión  $x_n$  se estabiliza muy rápidamente hacia  $0'6153846$ ; para  $r = 3'1$ , la sucesión comienza enseguida a tomar alternativamente los dos valores  $0'5580142$  y  $0'7645665$ ; sin embargo, para  $r = 3'9$  los cien primeros términos (aproximados) son:

0'975	0'095	0'335	0'869	0'442	0'962	0'141	0'475	0'972	0'104
0'363	0'902	0'343	0'879	0'412	0'944	0'203	0'631	0'908	0'325
0'856	0'479	0'973	0'101	0'354	0'892	0'374	0'913	0'309	0'832
0'542	0'967	0'121	0'416	0'947	0'192	0'607	0'930	0'253	0'737
0'754	0'723	0'780	0'667	0'865	0'452	0'966	0'127	0'433	0'957
0'158	0'520	0'973	0'100	0'354	0'891	0'375	0'914	0'303	0'824
0'563	0'959	0'152	0'504	0'974	0'095	0'336	0'870	0'439	0'960
0'147	0'490	0'974	0'096	0'339	0'874	0'428	0'954	0'167	0'545
0'967	0'124	0'424	0'952	0'176	0'566	0'957	0'157	0'517	0'973
0'099	0'349	0'886	0'393	0'930	0'251	0'734	0'760	0'710	0'802

y prosiguen indefinidamente con una pauta de irregularidad semejante. Al aumentar  $r$  por encima de un cierto valor ( $\simeq 3'53$ ) el sistema ha adquirido un comportamiento “caótico”, y no se puede prever lo que valdrá el término  $x_{200}$  sin calcularlo explícitamente.

Contrastar si la secuencia de los resultados anteriores puede considerarse una muestra aleatoria simple de alguna distribución.

### Ejercicio 10.10

Utilizar el test de Lilliefors para contrastar, con nivel de significación  $0'1$ , la normalidad de los últimos 25 datos del ejercicio anterior.

¿Es razonable, en estas circunstancias, el uso realizado del test de normalidad y puede, por consiguiente, concluirse que, una vez perdido el efecto del valor inicial  $x_0$ , el sistema produce valores que tienen aproximadamente distribución normal?

## Solución de los ejercicios

### Ejercicio 10.1

Para realizar el contraste  $\chi^2$  de bondad del ajuste, conviene elegir cuatro clases equiprobables, para que el número esperado de observaciones en cada una no sea menor que 5. Como la función de distribución es

$$F(x) = 1 - \left(\frac{\alpha}{x}\right)^{\beta},$$

con los valores de los parámetros indicados los tres cuartiles son: 2'529, 2'722 y 3'088. Ello da lugar a la clasificación

$A_i$	$n_i$	$p_i$	$np_i$
(2'4, 2'529]	1	1/4	5
(2'529, 2'722]	9	1/4	5
(2'722, 3'088]	7	1/4	5
(3'088, $\infty$ )	3	1/4	5

con un valor para el estadístico de Pearson:  $D = 8$ . La distribución  $\chi_3^2$  indica un nivel crítico entre 0'05 y 0'025, que lleva a rechazar la distribución propuesta.

Para realizar el test de Kolmogorov-Smirnov, la tabla siguiente contiene las observaciones muestrales, la distribución muestral, la distribución teórica:  $1 - (2'4/x)^{5/5}$  y las diferencias entre ambas.

Se obtiene de ella  $\Delta_{20} = 0'2354$  y la Tabla 7 indica que el ajuste de los datos a la distribución indicada no puede rechazarse con nivel de significación 0'1 (o inferior).

Nótese que aplicar ambos contrastes y rechazar la hipótesis nula cuando alguno de los dos detecte falta de ajuste, equivale a utilizar como región crítica:  $C = \{D > c\} \cup \{\Delta > k\}$ . Si  $c$  y  $k$  se determinan para que los contrastes tengan nivel de significación  $\alpha_1$  y  $\alpha_2$  respectivamente, el uso del test definido por  $C$  incrementa la probabilidad de error de tipo I hasta  $P(C) \leq \alpha_1 + \alpha_2$ . En este caso concreto, si se toma  $c = 7'815$  y  $k = 0'294$ , empleando sólo el test  $\chi^2$  se cometerá error de tipo I un 5 % de las veces; empleando sólo el test de Kolmogorov-Smirnov se cometerá error de tipo I en un 5 % de las ocasiones; y, si se emplean ambos (para rechazar cuando alguno de los dos rechace), la frecuencia de errores de tipo I será inferior al 10 %, pero superior al 5 %.

$x_{(i)}$	$F^*(x_{(i)})$	$F_0(x_{(i)})$		
2'52	0'05	0'2354	-0'1854	0'2354
2'54	0'10	0'2679	-0'1679	0'2179
2'55	0'15	0'2835	-0'1335	0'1835
2'56	0'20	0'2988	-0'0988	0'1488
2'57	0'25	0'3137	-0'0637	0'1137
2'60	0'30	0'3561	-0'0561	0'1061
2'64	0'35	0'4080	-0'0580	0'1080
2'65	0'40	0'4202	-0'0202	0'0702
2'69	0'45	0'4660	-0'0160	0'0660
2'71	0'50	0'4873	0'0127	0'0373
2'76	0'55	0'5364	0'0136	0'0364
2'78	0'60	0'5544	0'0456	0'0044
2'83	0'65	0'5960	0'0540	-0'0040
2'87	0'70	0'6261	0'0739	-0'0239
2'95	0'75	0'6785	0'0715	-0'0215
3'03	0'80	0'7225	0'0775	-0'0275
3'07	0'85	0'7418	0'1082	-0'0582
3'15	0'90	0'7759	0'1241	-0'0741
3'21	0'95	0'7980	0'1520	-0'1020
3'60	1	0'8925	0'1075	-0'0575

Si, a la vista de lo anterior, se rechaza que las observaciones se ajustan a la distribución indicada, con los valores paramétricos dados, habrá que proceder a la estimación de máxima verosimilitud de  $\alpha$  y  $\beta$ , antes de contrastar el ajuste a la distribución con tales parámetros. La función de verosimilitud es

$$f(x_1, \dots, x_n) = \frac{\beta^n \alpha^{n\beta}}{(x_1 \cdots x_n)^{\beta+1}} I_{\{x_{(1)} > \alpha\}}$$

y el logaritmo del primer factor:

$$n \log \beta + n\beta \log \alpha - (\beta + 1) \sum_{i=1}^n \log x_i$$

es creciente con  $\alpha$ , así que el mayor valor posible de  $\alpha$ :  $\hat{\alpha} = x_{(1)}$ , constituye su estimación de máxima verosimilitud. Derivando respecto a  $\beta$  se obtiene el estimador de máxima verosimilitud de  $\beta$ :

$$\hat{\beta} = \left( \frac{1}{n} \sum_{i=1}^n \log x_i - \log \alpha \right)^{-1}$$

En particular, con las observaciones realizadas, resulta  $\hat{\alpha} = 2'52$  y  $\hat{\beta} = 9'445$ . Con ello puede procederse a realizar el test  $\chi^2$  mediante la tabla

$A_i$	$n_i$	$p_i$	$np_i$
(2'52, 2'6]	6	0'2556	5'112
(2'6, 2'72]	4	0'2583	5'166
(2'72, 2'93]	4	0'2453	4'906
(2'93, $\infty$ )	6	0'2408	4'816

que da al estadístico de Pearson el valor:  $D = 0'876$ . Como se han estimado 2 parámetros, la distribución de  $D$  es  $\chi_1^2$  y el ajuste de los datos a la distribución tiene que ser aceptado.

El escaso tamaño muestral empleado ha obligado a incumplir ligeramente la norma  $np_i \geq 5$  (sin posibilidad de reducir el número de conjuntos de la partición para no anular los grados de libertad de la distribución  $\chi^2$  correspondiente). Convendría por tanto aumentar el número de observaciones, pero los resultados obtenidos son bastante concluyentes de que el nuevo método de lubricación ha afectado a los parámetros de la distribución y no a su forma funcional.

### Ejercicio 10.2

(a) En primer lugar se trata de detectar si hay una pauta común en los tres ciclos anuales considerados, ya que, en caso contrario, ello significaría que el comportamiento es diferente en cada año. Para ello, conviene considerar los datos agrupados en la forma:

	J	J	A	S	O	N	D	E	F	M	A-M	Total
1986-87	6	9	8	6	8	11	8	5	4	4	3	72
1987-88	8	8	6	2	2	1	2	7	8	3	3	50
1988-89	7	7	6	5	5	3	5	1	2	1	3	45
Total	21	24	20	13	15	15	15	13	14	8	9	167

con los meses de Abril y Mayo sumados para conseguir que sea  $n_i n_{\cdot j} / n \geq 2$ . El estadístico de contraste toma el valor

$$D = 167 \left( -1 + \sum_{i=1}^3 \sum_{j=1}^{11} \frac{n_{ij}}{n_i n_{\cdot j}} \right) = 24'477$$

y  $D$  tiene distribución  $\chi^2_{20}$ , cuya tabla indica que la hipótesis de que las tres temporadas siguen el mismo patrón no puede ser rechazada con nivel de significación 0'1 (el nivel crítico es, de hecho, 0'222).

(b) Admitida la homogeneidad de las tres muestras, los 167 casos, agrupados por meses, se distribuyen como indica la tabla siguiente.

J	J	A	S	O	N	D	E	F	M	A	M
21	24	20	13	15	15	15	13	14	8	4	5

La influencia del mes sobre el número de casos ocurridos tendría que ser descartada si las frecuencias observadas fuesen compatibles con probabilidades de 1/12 para cada uno de ellos; es decir si no puede admitirse que los datos son desviaciones debidas al azar en torno a  $167/12$  casos por mes. El estadístico de Pearson, para dicho contraste vale

$$D = -167 + \frac{12}{167} \sum_{j=1}^{12} n_j^2 = 29'24$$

y tiene distribución  $\chi^2_{11}$ . La hipótesis de uniformidad de la distribución puede rechazarse, por tanto, con nivel de significación 0'005.

Las diferencias entre los tres meses de verano (J, J, A) no son significativas, pues los datos:

J	J	A
21	24	20
65/3	65/3	65/3

dan como valor del estadístico de Pearson

$$D = -65 + 3/65 \sum_{j=1}^3 n_j^2 = 0'4,$$

que, comparado con la distribución  $\chi^2_2$ , no permite rechazar la hipótesis de que los casos se presentan uniformemente distribuidos entre los tres meses.

Lo mismo ocurre con los tres meses de primavera (M, A, M:  $D = 1'53$ , inferior a  $\chi^2_{2,0'1}$ ) y, por supuesto, con los seis meses de otoño-invierno.

En cambio, existen diferencias significativas entre estos tres períodos; por ejemplo, la comparación entre el verano y los seis meses siguientes, realizada mediante la tabla adjunta, da como resultado:

Verano	Otoño-Invierno
65	85
$\frac{1}{3}150$	$\frac{2}{3}150$

$$D = 6'75 > \chi^2_{1,0'01},$$

de manera que no hay un reparto uniforme de los casos entre los tres meses de verano y los seis siguientes.

En definitiva, la incidencia de la enfermedad es más alta en verano y más baja en primavera, respecto del nivel medio durante el resto del año.

En Inferencia estadística —como sin duda ha quedado de manifiesto a lo largo del texto— hay diversos procedimientos de análisis de una misma situación. Los datos de este ejercicio corresponden a una *serie temporal* (un conjunto de observaciones a lo largo del tiempo) y tales situaciones tienen su tratamiento específico que debería ser abordado en un curso independiente. Ello no significa, sin embargo, que el análisis efectuado dé “peores” resultados que las técnicas de series temporales; sólo quiere decir que existen métodos para obtener información complementaria, de cara, por ejemplo, a hacer previsiones sobre la incidencia de la enfermedad en años posteriores.

### Ejercicio 10.3

Completada la tabla de contingencia

	0	1	2	3	4	5	6	7	8	9	10	$n_i$
0	9	9	8	5	4	2	3	2	4	2	0	48
1	8	16	9	8	5	7	3	2	4	2	3	67
2	1	10	13	15	12	10	5	6	3	3	0	78
3	3	19	23	16	19	16	12	9	8	7	7	139
4	4	7	8	12	11	15	12	12	11	9	9	110
5	2	8	13	16	18	20	13	10	9	9	9	127
6	1	13	15	10	9	9	10	9	8	15	6	105
7	1	2	3	9	7	13	4	9	5	6	1	60
8	2	4	4	9	15	17	22	16	8	9	11	117
9	1	4	3	4	4	8	10	7	9	9	11	70
10	0	2	1	4	8	7	15	9	10	13	10	79
$n_j$	32	94	100	108	112	124	109	91	79	84	67	1000

el estadístico de Pearson para el contraste de independencia vale

$$D = 1000 \left( \sum_{i=0}^{10} \sum_{j=0}^{10} \frac{n_{ij}^2}{n_i \cdot n_j} - 1 \right) = 243'64.$$

Por otra parte la distribución de  $D$ , si es cierta la independencia, es  $\chi^2_{100}$ . Como

$$P\{\chi^2_{100} > 243'64\} \simeq P\left\{ \frac{1}{2} \left( Z + \sqrt{199} \right)^2 > 243'64 \right\} = P\{Z > 7'96\} < 0'00001,$$

(donde  $Z$  es una variable con distribución  $N(0, 1)$ ) la independencia de ambas variables puede ser rechazada con gran seguridad.

La distribución de notas de cada asignatura es

	0	1	2	3	4	5	6	7	8	9	10	Tot.
A	48	67	78	139	110	127	105	60	117	70	79	1000
B	32	94	100	108	112	124	109	91	79	84	67	1000
Tot.	80	161	178	247	222	251	214	151	196	154	146	2000

El estadístico para el contraste de homogeneidad de ambas muestras vale

$$D = 2000 \left( \sum_{i=1}^2 \sum_{j=0}^{10} \frac{n_{ij}^2}{1000n_j} - 1 \right) = 30'44$$

y su distribución es  $\chi^2_{10}$ . Como  $P\{\chi^2_{10} > 30'44\} < 0'005$ , la homogeneidad de la distribución de las calificaciones debe ser rechazada.

#### Ejercicio 10.4

Ordenando conjuntamente ambas muestras se obtiene la tabla:

$x$	$F^*(x)$	$G^*(x)$
8'1	0	1/10
9'3	0	2/10
10'3	0	3/10
10'5	1/7	3/10
11'1	1/7	4/10
11'6	1/7	5/10
12'5	2/7	5/10
12'7	2/7	6/10
13'1	2/7	7/10
13'5	3/7	7/10
13'7	3/7	8/10
13'8	4/7	8/10
15'9	4/7	9/10
16'2	4/7	1
16'3	5/7	1
20'2	6/7	1
22'1	1	0

según ello  $\Delta_{7,10} = 3/7$ . La tabla 9 indica, entonces, que el nivel crítico del test de homogeneidad de Kolmogorov-Smirnov es mayor que 0'2, lo que obliga a aceptar la hipótesis nula de igualdad entre ambas poblaciones.

A partir de la tabla anterior, la mediana de la muestra combinada es 13'1 y el número de observaciones de la primera muestra inferiores a 13'1 es  $T = 2$ . La distribución de  $T$ , en el supuesto de que ambas muestras provengan de la misma población, resulta

0	1	2	3	4	5	6	7
0'0019	0'0346	0'1814	0'3628	0'3023	0'1037	0'013	0'0004

de forma que (incluyendo las probabilidades de menor a mayor) la región crítica  $\{T \leq 1\} \cup \{T \geq 5\}$  tiene tamaño 0'1536 y no permite rechazar la igualdad entre ambas poblaciones.

### Ejercicio 10.5

Después de ordenar conjuntamente ambas muestras se obtiene la tabla

$x$	$F^*(x)$	$G^*(x)$	$x$	$F^*(x)$	$G^*(x)$		
0'7	0	1/15	-2/30	8'5	4/10	10/15	-8/30
2'7	0	2/15	-4/30	8'9	5/10	10/15	-5/30
2'8	0	3/15	-6/30	9'2	5/10	11/15	-7/30
3'0	0	4/15	-8/30	10'7	6/10	11/15	-4/30
3'8	0	5/15	-10/30	12'1	7/10	11/15	-1/30
4'2	0	6/15	-12/30	12'9	8/10	11/15	2/30
4'5	1/10	6/15	-9/30	13'7	8/10	12/15	0
4'8	2/10	6/15	-6/30	14'9	8/10	13/15	-2/30
5'2	2/10	7/15	-8/30	16'1	8/10	14/15	-4/30
5'5	2/10	8/15	-10/30	17'9	9/10	14/15	-1/30
5'6	3/10	8/15	-7/30	18'3	1	14/15	2/30
6'6	3/10	9/15	-9/30	18'6	1	1	0
7'5	3/10	10/15	-11/30				

De la que se deduce que es  $\Delta_{10,15} = 12/30 = 2/5$ . La tabla 9 indica que el nivel crítico del test es 0'2. Ello conduce a aceptar la igualdad entre ambas poblaciones.

Por su parte el estadístico de Mann-Whitney toma el valor

$$V = 6 + 6 + 8 + 10 + 10 + 11 + 11 + 11 + 14 + 14 = 101.$$

Utilizando que  $V$  tendrá aproximadamente distribución normal de parámetros  $N(75, 18'03)$ , se obtiene un nivel crítico igual a

$$2P\{V > 100\} = 2P\{Z > 1'39\} = 0'1646,$$

que lleva a aceptar la hipótesis nula, aunque con menor seguridad. Como se ve, el test de Mann-Whitney es menos conservador que el de Kolmogorov-Smirnov.

### Ejercicio 10.6

Las diferencias entre los resultados por ambos procedimientos son

$$-0'03 \quad 0'04 \quad -0'08 \quad 0'12 \quad 0'05 \quad 0'04 \quad 0'07 \quad -0'04 \quad 0'11 \quad 0'14 \quad 0'06 \quad 0'02$$

El test de los signos para el contraste de la hipótesis de que es cero la mediana  $M$  de la distribución de las diferencias se basa en el número de observaciones positivas:  $T = 9$ . Su distribución, si realmente es  $M = 0$ , es binomial de parámetros 12 y  $1/2$ , luego

$$P\{T \leq 3\} + P\{T \geq 9\} = 0'146$$

es el nivel crítico con el cual puede descartarse  $M = 0$  frente a la alternativa  $M \neq 0$ . No es, por tanto, suficientemente pequeño para afirmar que existe diferencia entre ambos procedimientos.

Para aplicar el test de los rangos signados hay que consignar, bajo cada diferencia, el rango que ocupa en la ordenación de sus valores absolutos:

$$\begin{array}{ccccccccccccc} -0'03 & 0'04 & -0'08 & 0'12 & 0'05 & 0'04 & 0'07 & -0'04 & 0'11 & 0'14 & 0'06 & 0'02 \\ (2) & (4) & (9) & (11) & (6) & (4) & (8) & (4) & (10) & (12) & (7) & (1) \end{array}$$

La suma de los rangos de los términos positivos es

$$T^+ = 4 + 11 + 6 + 4 + 8 + 10 + 12 + 7 + 1 = 63.$$

La tabla 11 muestra que para rechazar  $M = 0$  frente a  $M \neq 0$ , con nivel de significación  $0'05$ , habría de ser  $T^+ > 64$  ó  $T^+ < 8$ . En cambio, con nivel de significación  $0'1$ , la hipótesis  $M = 0$  puede ser rechazada puesto que es  $T^+ > 60$ .

Los tests aplicados contrastan únicamente si la mediana de las diferencias es nula; sólo deben utilizarse, por tanto, si se supone que las distribuciones de los resultados obtenidos por ambos procedimientos no se diferencian por su dispersión, asimetría, etc.

Además, el uso de cualquiera de los tests anteriores sólo es adecuado si hay razones para pensar que la distribución de las diferencias no es normal. Verificar esta condición mediante un test de normalidad no es, sin embargo, conveniente puesto que en tales contrastes la hipótesis de normalidad actúa

como hipótesis nula y sólo es rechazada cuando hay una gran evidencia contra ella.

Por último, el test de los rangos signados no debería emplearse si no hay pruebas de que la distribución de las diferencias tiene un centro de simetría.

### Ejercicio 10.7

Ordenada la muestra según la velocidad, las magnitudes esperadas son

$$\begin{array}{cccccccccc} 10'4 & 12'9 & 13'5 & 14'2 & 16'1 & 16'2 & 16'5 & 16'3 & 17'9 & 17'7 \\ (1) & (2) & (3) & (4) & (5) & (6) & (8) & (7) & (10) & (9) \end{array}$$

cuyos rangos —en la ordenación de dichos valores— se han indicado debajo de cada uno. El recuento del número de parejas que conservan el orden proporciona el valor

$$P = 9 + 8 + 7 + 6 + 5 + 4 + 2 + 2 + 0 + 0 = 43,$$

con lo cual  $T = 0'911$ .

La tabla 13 indica entonces un nivel crítico mucho menor que  $0'02$ , lo cual conduce a rechazar la hipótesis de independencia entre ambas variables.

Para determinar el test basado en el coeficiente de correlación entre rangos de Spearman, se calcula primero, a partir de los mismos rangos anteriores,

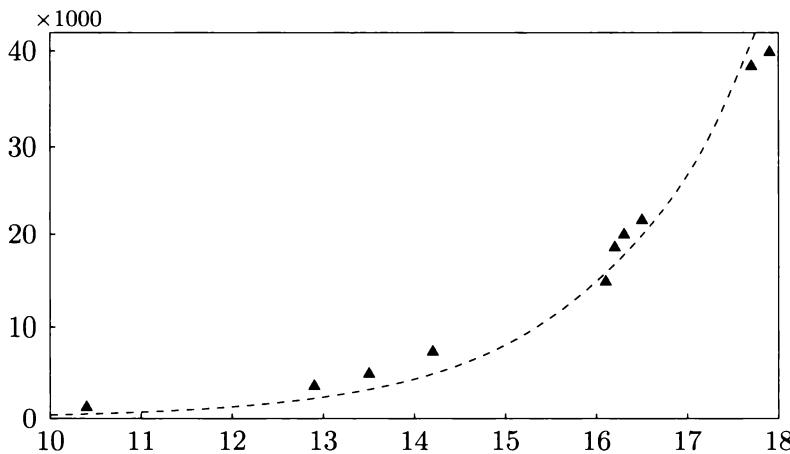
$$U = \sum_{j=1}^{10} (r_j - j)^2 = 4,$$

lo cual proporciona

$$R_S = 1 - \frac{6U}{n(n^2 - 1)} = 0'976.$$

La tabla 14 conduce de nuevo a rechazar la hipótesis nula de independencia entre ambas variables, con un nivel crítico inferior a  $0'001$ .

La representación gráfica de los 10 puntos, que aparece en la figura siguiente, hace bastante inútil la realización de los contrastes anteriores, puesto que a simple vista se observa una dependencia clara entre ambas variables e, incluso, sugiere que la velocidad crece exponencialmente con la magnitud aparente, lo cual se traduce en una dependencia lineal entre la velocidad y la distancia.



### Ejercicio 10.8

Se puede contrastar, en primer lugar, si todas las cifras aparecen con la misma frecuencia,  $1/10$ , como si hubiesen sido elegidas al azar mediante extracciones de una urna con 10 bolas numeradas de 0 a 9.

Para ello han de compararse las frecuencias observadas y esperadas, mediante la tabla

	0	1	2	3	4	5	6	7	8	9
$n_i$	8	8	12	11	10	8	9	8	12	14
$np_i$	10	10	10	10	10	10	10	10	10	10

El valor del estadístico de Pearson resulta

$$D = \frac{1}{10} \sum_{i=0}^9 (n_i - 10)^2 = 4'2,$$

que comparado con la distribución  $\chi^2_9$ , lleva a aceptar la hipótesis de uniformidad con un nivel crítico próximo a  $0'9$ .

Los números reales que presentan en su desarrollo decimal proporciones de cada cifra iguales a  $1/10$ , se denominan *normales*. Se puede probar que son normales casi todos los números reales: todos menos los de un conjunto de longitud cero (que incluiría a todos los números racionales no periódicos, a muchos periódicos y a irracionales como  $0'7077077707777077\dots$ ). Por supuesto, nadie sabe si  $\pi$  es normal o bien adquiere, después de unos cuantos millones de cifras, una estructura “anormal” (del tipo de la del irracional

anterior). Pero la comprobación de la “normalidad” de  $\pi$  ha sido realizada para el primer millón de cifras.

Por otra parte, hay que contrastar si la posición de las cifras parece el resultado de haberlas elegido al azar, sin dependencias entre ellas. Para ello lo adecuado es un test de rachas; por ejemplo, eligiendo 4'5 como promedio de las 10 cifras, se indican con un – o un + aquellos dígitos que sean, respectivamente, menores o mayores o iguales que 4'5; se obtiene así:

1	4	1	5	9	2	6	5	3	5	8	9	7	9	3	2	3	8	4	6
–	–	–	+	+	–	+	+	–	+	+	+	+	+	–	–	–	+	–	+
2	6	4	3	3	8	3	2	7	9	5	0	2	8	8	4	1	9	7	1
–	+	–	–	–	+	–	–	+	+	+	–	–	+	+	–	–	+	+	–
6	9	3	9	9	3	7	5	1	0	5	8	2	0	9	7	4	9	4	4
+	+	–	+	+	–	+	+	–	–	+	+	–	–	+	+	–	+	–	–
5	9	2	3	0	7	8	1	6	4	0	6	2	8	6	2	0	8	9	9
+	+	–	–	–	+	+	–	+	–	–	+	–	+	+	–	–	+	+	+
8	6	2	8	0	3	4	8	2	5	3	4	2	1	1	7	0	6	7	9
+	+	–	+	–	–	–	+	–	+	–	–	–	–	–	+	–	+	+	+

con  $n = 49$  signos – y  $m = 51$  signos + y un total de  $R = 54$  rachas. Como  $n$  y  $m$  son grandes, para que la colocación de las cifras parezca hecha al azar,  $R$  tendría que tener aproximadamente distribución

$$N\left(\frac{2 \cdot 49 \cdot 51}{100} + 1, \sqrt{\frac{2 \cdot 49 \cdot 51(2 \cdot 49 \cdot 51 - 49 - 51)}{990.000}}\right) = N(50'98, 4'97).$$

El nivel crítico resulta

$$2P\{R > 54\} = 2P\{Z > 0'61\} = 0'5418$$

que no permite, en absoluto, afirmar que las cifras no están situadas como si hubiesen sido elegidas independientemente unas de otras.

Otra posibilidad, en la misma dirección, es clasificar las cifras en pares e impares, tratando de detectar alguna regularidad en la localización de unas y otras. Concretamente

1	4	1	5	9	2	6	5	3	5	8	9	7	9	3	2	3	8	4	6
i	p	i	i	i	p	p	i	i	i	p	i	i	i	i	p	i	p	p	p
2	6	4	3	3	8	3	2	7	9	5	0	2	8	8	4	1	9	7	1
p	p	p	i	i	p	i	p	i	i	i	p	p	p	p	i	i	i	i	
6	9	3	9	9	3	7	5	1	0	5	8	2	0	9	7	4	9	4	4
p	i	i	i	i	i	i	i	i	i	p	i	p	p	i	i	p	i	p	p
5	9	2	3	0	7	8	1	6	4	0	6	2	8	6	2	0	8	9	9
i	i	p	i	p	i	p	i	p	p	p	p	p	p	p	p	i	i	i	
8	6	2	8	0	3	4	8	2	5	3	4	2	1	1	7	0	6	7	9
p	p	p	p	p	i	p	p	p	i	i	p	p	i	i	p	p	i	i	

con  $n = 49$  cifras impares,  $m = 51$  cifras pares y  $R = 43$  rachas. La distribución aproximada de  $R$  es la misma normal anterior (si los términos de la secuencia fuesen independientes) y el nivel crítico resulta

$$2P\{R > 43\} = 2P\{Z > -1'6\} = 0'1096,$$

que tampoco permite afirmar que las cifras no están situadas como si hubiesen sido elegidas al azar.

### Ejercicio 10.9

Para aplicar un test de rachas a los 100 resultados obtenidos, puede determinarse la mediana, que vale 0'5435 (los valores que ocupan el lugar 50 y 51, una vez ordenados, son 0'542 y 0'545). Comparando cada término con la mediana se obtienen los signos:

+	-	-	+	-	+	-	-	+	-	-	+	-	+	-	+	+	-	
+	-	+	-	-	+	-	+	-	+	-	-	-	+	-	+	+	-	+
+	+	+	+	+	-	+	-	+	-	-	+	-	-	+	-	+	-	+
+	+	-	-	+	-	-	+	-	+	-	-	+	-	-	+	-	+	-
+	-	-	+	-	+	+	-	+	-	-	+	-	+	-	+	+	+	+

que muestran un total de  $R = 71$  rachas, constituidas por los 50 símbolos de cada tipo. La distribución de  $R$ , si las observaciones fuesen independientes sería  $N(51, 4'97)$ ; luego, con nivel de significación 0'01, se puede rechazar la independencia de las observaciones en el caso de que sea  $R > 51 + 2'58 \cdot 4'97 = 63'8$  ó  $R < 51 - 2'58 \cdot 4'97 = 38'2$ . El test de rachas empleado detecta pues, con gran seguridad, la falta de aleatoriedad de la muestra.

Si cada término se compara con el anterior, aparecen los signos

-	+	+	-	+	-	+	+	-	+	+	-	+	-	+	+	-	
+	-	+	-	+	+	-	+	-	+	-	+	+	-	+	+	-	+
+	-	+	-	+	-	+	+	-	+	+	-	+	-	+	-	+	-
-	+	-	+	+	-	+	+	-	+	+	-	+	-	+	-	+	-
+	-	+	+	-	+	+	-	+	+	-	+	-	+	-	+	-	+

con  $R = 79$  rachas formadas por 39 signos - y 60 signos +. La distribución de  $R$  si la secuencia fuese aleatoria sería ahora  $N(48'27, 4'72)$ ; de manera que, con nivel de significación 0'001, la región crítica es  $\{R < 32'8\} \cup \{R > 63'7\}$ . Este segundo test resulta, pues, todavía más significativo que el primero.

No es nada sorprendente que el sistema determinístico del enunciado no produzca secuencias aleatorias, por mucho que sea caótico.

**Ejercicio 10.10**

Los últimos 25 datos del ejercicio anterior, una vez ordenados aparecen en la primera columna de la siguiente tabla, seguidos de la función de distribución muestral.

$x_{(i)}$	$F^*(x_{(i)})$	$F(x_{(i)})$	$F^*(x_{(i)}) - F(x_{(i)})$	$F(x_{(i)}) - F^*(x_{(i)})$
0'099	0'04	0'0554	-0'0154	0'0554
0'124	0'08	0'0651	0'0149	0'0251
0'157	0'12	0'0799	0'0401	-0'0001
0'167	0'16	0'0849	0'0751	-0'0351
0'176	0'20	0'0895	0'1105	-0'0705
0'251	0'24	0'1359	0'1041	-0'0641
0'349	0'28	0'2180	0'0620	-0'0220
0'393	0'32	0'2625	0'0575	-0'0175
0'424	0'36	0'2965	0'0635	-0'0235
0'428	0'40	0'3011	0'0989	-0'0589
0'517	0'44	0'4087	0'0313	0'0087
0'545	0'48	0'4445	0'0355	0'0045
0'566	0'52	0'4716	0'0484	-0'0084
0'710	0'56	0'6549	-0'0949	0'1349
0'734	0'60	0'6833	-0'0833	0'1233
0'760	0'64	0'7129	-0'0729	0'1129
0'802	0'68	0'7577	-0'0777	0'1177
0'874	0'72	0'8248	-0'1048	0'1448
0'886	0'76	0'8347	-0'0747	0'1147
0'930	0'80	0'8680	-0'0680	0'1080
0'952	0'84	0'8826	-0'0426	0'0826
0'954	0'88	0'8839	-0'0039	0'0439
0'957	0'92	0'8858	0'0342	0'0058
0'967	0'96	0'8920	0'0680	-0'0280
0'973	1	0'8956	0'1044	-0'0644

Como  $\bar{x} = 0'5878$  y  $s = 0'3065$ , la tercera columna contiene los valores de la distribución normal con dichos parámetros. Por último se han incluido las diferencias  $F^*(x_{(i)}) - F(x_{(i)})$  y  $F(x_{(i)}) - F^*(x_{(i)})$ . En definitiva, se obtiene  $\Delta_{25} = 0'1448$ . Por otra parte, la tabla de Lilliefors muestra que para rechazar la normalidad, con nivel de significación 0'1, debe ser  $\Delta_{25} > 0'165$ ; el nivel crítico será de hecho próximo a 0'2.

Habida cuenta de la procedencia de los datos, no hay ninguna razón para suponer que los valores obtenidos se ajusten a una distribución normal; antes al contrario, consta según el ejercicio anterior que no pueden considerarse como una muestra aleatoria de ninguna distribución ni, en particular, de una normal. En tales circunstancias, asignar a la distribución normal el papel de hipótesis nula y aplicar rutinariamente el test de normalidad, con un nivel de significación como 0'1 ó 0'2, no tiene ningún sentido: los resultados no contienen evidencia suficiente para invalidar la hipótesis nula, pero no hay tampoco evidencia anterior alguna que permita situar a la hipótesis de normalidad en tal posición de privilegio.

Sería preciso un nivel crítico muy elevado para que las observaciones aportasen evidencia en favor de la normalidad de su distribución, en vez de limitarse, con un nivel crítico del orden de 0'2, a no contradecirla. Por ejemplo, en el ejercicio 10.8, se ha aceptado que las cifras de  $\pi$  son “como si hubiesen elegidas al azar”, debido a que el contraste ofrecía un nivel crítico del orden de 0'9.

Es frecuente, en determinadas aplicaciones, disponer de un conjunto de datos que ha sido generado por un cierto mecanismo (habitualmente aplicando una transformación a los elementos de una muestra), sin que haya presunción teórica alguna de la normalidad de los resultados. Aplicar rutinariamente en este contexto un test de normalidad, con un nivel crítico como los que figuran en la tabla 8, es muy poco razonable.



# Bibliografía

- Bickel, P.J.-Doksum, K.A. (1977): *Mathematical Statistics*. Holden-Day.
- Casella, G.-Berger, R.L. (1990): *Statistical Inference*. Brooks/Cole.
- Cox, D.R.-Hinkley, D.V. (1974): *Theoretical Statistics*. Chapman and Hall.
- Cramer, H. (1968): *Métodos Matemáticos de Estadística*. Aguilar.
- Dudewicz, E.J.-Mishra, S.N. (1988): *Modern Mathematical Statistics*. Wiley.
- Ferguson, T. (1967): *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press.
- Gibbons, J.D. (1985): *Nonparametric Statistical Inference*. Marcel Decker.
- De Groot, M.H. (1988): *Probabilidad y estadística*. Addison Wesley.
- Lehmann, E.L. (1983): *Theory of Point Estimation*. Wiley.
- Lehmann, E.L. (1986): *Testing Statistical Hypothesis*. Wiley.
- Mood, A.M.-Graybill, F.A. - Boes, D.C. (1974): *Introduction to the Theory of Statistics*. Mc Graw Hill.
- Rao, C.R. (1973): *Linear Statistical Inference and its Applications*. Wiley.
- Rohatgi, V.K. (1976): *An Introduction to Probability Theory and Mathematical Statistics*. Wiley.
- Schmetterer, L. (1974): *Introduction to Mathematical Statistics*, Springer-Verlag.



# Tablas

Las siguientes tablas son útiles en relación con diversos métodos estadísticos, tal y como se ha descrito en el texto. Toda tabla, por supuesto, es imperfecta puesto que sólo da los valores de una determinada función  $f(x)$  en un conjunto finito de puntos  $x_1, \dots, x_k$ . Una aproximación del valor  $f(x)$  para  $x \in (x_i, x_{i+1})$  puede obtenerse por interpolación lineal; es decir

$$f(x) \simeq f(x_i) + \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} (x - x_i)$$

Actualmente las deficiencias de las tablas pueden ser soslayadas debido a la existencia de programas de ordenador que producen el resultado de muchas de ellas con mayor exactitud y comodidad. Si se dispone de uno de tales “calculadores probabilísticos” no hay que dudar en utilizarlo; en caso contrario, el manejo de las tablas puede dar lugar a pequeñas diferencias con los valores que aparecen en algunos ejemplos del texto.

TABLA 1: Distribución Binomial

$$P_p\{X = x\} = \binom{n}{x} p^x (1 - p)^{n-x}$$

$p$	0'05	0'10	0'15	0'20	0'25	0'30	1/3	0'40	0'45	0'50	
$n$	$x$										
2	0	0'9025	0'8100	0'7225	0'6400	0'5625	0'4900	0'4444	0'3600	0'3025	0'2500
	1	0'0950	0'1800	0'2550	0'3200	0'3750	0'4200	0'4444	0'4800	0'4950	0'5000
	2	0'0025	0'0100	0'0225	0'0400	0'0625	0'0900	0'1111	0'1600	0'2025	0'2500
3	0	0'8574	0'7290	0'6141	0'5120	0'4219	0'3430	0'2963	0'2160	0'1664	0'1250
	1	0'1354	0'2430	0'3251	0'3840	0'4219	0'4410	0'4444	0'4320	0'4084	0'3750
	2	0'0071	0'0270	0'0574	0'0960	0'1406	0'1890	0'2222	0'2880	0'3341	0'3750
	3	0'0001	0'0010	0'0034	0'0080	0'0156	0'0270	0'0370	0'0640	0'0911	0'1250
4	0	0'8145	0'6561	0'5220	0'4096	0'3164	0'2401	0'1975	0'1296	0'0915	0'0625
	1	0'1715	0'2916	0'3685	0'4096	0'4219	0'4116	0'3951	0'3456	0'2995	0'2500
	2	0'0135	0'0486	0'0975	0'1536	0'2109	0'2646	0'2963	0'3456	0'3675	0'3750
	3	0'0005	0'0036	0'0115	0'0256	0'0469	0'0756	0'0988	0'1536	0'2005	0'2500
	4	0'0000	0'0001	0'0005	0'0016	0'0039	0'0081	0'0123	0'0256	0'0410	0'0625
5	0	0'7738	0'5905	0'4437	0'3277	0'2373	0'1681	0'1317	0'0778	0'0503	0'0312
	1	0'2036	0'3280	0'3915	0'4096	0'3955	0'3602	0'3292	0'2592	0'2059	0'1562
	2	0'0214	0'0729	0'1382	0'2048	0'2637	0'3087	0'3292	0'3456	0'3369	0'3125
	3	0'0011	0'0081	0'0244	0'0512	0'0879	0'1323	0'1646	0'2304	0'2757	0'3125
	4	0'0000	0'0004	0'0022	0'0064	0'0146	0'0284	0'0412	0'0768	0'1128	0'1562
	5	0'0000	0'0001	0'0003	0'0010	0'0024	0'0041	0'0102	0'0185	0'0312	
6	0	0'7351	0'5314	0'3771	0'2621	0'1780	0'1176	0'0878	0'0467	0'0277	0'0156
	1	0'2321	0'3543	0'3993	0'3932	0'3560	0'3025	0'2634	0'1866	0'1359	0'0938
	2	0'0305	0'0984	0'1762	0'2458	0'2966	0'3241	0'3292	0'3110	0'2780	0'2344
	3	0'0021	0'0146	0'0415	0'0819	0'1318	0'1852	0'2195	0'2765	0'3032	0'3125
	4	0'0001	0'0012	0'0055	0'0154	0'0330	0'0595	0'0823	0'1382	0'1861	0'2344
	5	0'0000	0'0001	0'0004	0'0015	0'0044	0'0102	0'0165	0'0369	0'0609	0'0938
	6	0'0000	0'0000	0'0001	0'0002	0'0007	0'0014	0'0041	0'0083	0'0156	
7	0	0'6983	0'4783	0'3206	0'2097	0'1335	0'0824	0'0585	0'0280	0'0152	0'0078
	1	0'2573	0'3720	0'3960	0'3670	0'3115	0'2471	0'2048	0'1306	0'0872	0'0547
	2	0'0406	0'1240	0'2097	0'2753	0'3115	0'3177	0'3073	0'2613	0'2140	0'1641
	3	0'0036	0'0230	0'0617	0'1147	0'1730	0'2269	0'2561	0'2903	0'2918	0'2734
	4	0'0002	0'0026	0'0109	0'0287	0'0577	0'0972	0'1280	0'1935	0'2388	0'2734
	5	0'0000	0'0002	0'0012	0'0043	0'0115	0'0250	0'0384	0'0774	0'1172	0'1641
	6	0'0000	0'0000	0'0001	0'0004	0'0013	0'0036	0'0064	0'0172	0'0320	0'0547
	7	0'0000	0'0000	0'0000	0'0001	0'0002	0'0005	0'0016	0'0037	0'0078	
8	0	0'6634	0'4305	0'2725	0'1678	0'1001	0'0576	0'0390	0'0168	0'0084	0'0039
	1	0'2793	0'3826	0'3847	0'3355	0'2670	0'1977	0'1561	0'0896	0'0548	0'0312
	2	0'0515	0'1488	0'2376	0'2936	0'3115	0'2965	0'2731	0'2090	0'1569	0'1094
	3	0'0054	0'0331	0'0839	0'1468	0'2076	0'2541	0'2731	0'2787	0'2568	0'2188
	4	0'0004	0'0046	0'0185	0'0459	0'0865	0'1361	0'1707	0'2322	0'2627	0'2734
	5	0'0000	0'0004	0'0026	0'0092	0'0231	0'0467	0'0683	0'1239	0'1719	0'2188
	6	0'0000	0'0000	0'0002	0'0011	0'0038	0'0100	0'0171	0'0413	0'0703	0'1094
	7	0'0000	0'0000	0'0001	0'0004	0'0012	0'0024	0'0079	0'0164	0'0312	
	8	0'0000	0'0000	0'0000	0'0001	0'0002	0'0007	0'0017	0'0039		

**TABLA 1: Distribución Binomial** (continuación)

<i>n</i>	<i>p</i>	0'05	0'10	0'15	0'20	0'25	0'30	1/3	0'40	0'45	0'50
9	0	0'6302	0'3874	0'2316	0'1342	0'0751	0'0404	0'0260	0'0101	0'0046	0'0020
	1	0'2985	0'3874	0'3679	0'3020	0'2253	0'1556	0'1171	0'0605	0'0339	0'0176
	2	0'0629	0'1722	0'2597	0'3020	0'3003	0'2668	0'2341	0'1612	0'1110	0'0703
	3	0'0077	0'0446	0'1069	0'1762	0'2336	0'2668	0'2731	0'2508	0'2119	0'1641
	4	0'0006	0'0074	0'0283	0'0661	0'1168	0'1715	0'2048	0'2508	0'2600	0'2461
	5	0'0000	0'0008	0'0050	0'0165	0'0389	0'0735	0'1024	0'1672	0'2128	0'2461
	6		0'0001	0'0006	0'0028	0'0087	0'0210	0'0341	0'0743	0'1160	0'1641
	7		0'0000	0'0000	0'0003	0'0012	0'0039	0'0073	0'0212	0'0407	0'0703
	8			0'0000	0'0001	0'0004	0'0009	0'0035	0'0083	0'0176	
	9				0'0000	0'0000	0'0001	0'0003	0'0008	0'0020	
10	0	0'5987	0'3487	0'1969	0'1074	0'0563	0'0282	0'0173	0'0060	0'0025	0'0010
	1	0'3151	0'3874	0'3474	0'2684	0'1877	0'1211	0'0867	0'0403	0'0207	0'0098
	2	0'0746	0'1937	0'2759	0'3020	0'2816	0'2335	0'1951	0'1209	0'0763	0'0439
	3	0'0105	0'0574	0'1298	0'2013	0'2503	0'2668	0'2601	0'2150	0'1665	0'1172
	4	0'0010	0'0112	0'0401	0'0881	0'1460	0'2001	0'2276	0'2508	0'2384	0'2051
	5	0'0001	0'0015	0'0085	0'0264	0'0584	0'1029	0'1366	0'2007	0'2340	0'2461
	6	0'0000	0'0001	0'0012	0'0055	0'0162	0'0368	0'0569	0'1115	0'1596	0'2051
	7		0'0000	0'0001	0'0008	0'0031	0'0090	0'0163	0'0425	0'0746	0'1172
	8			0'0000	0'0001	0'0004	0'0014	0'0030	0'0106	0'0229	0'0439
	9				0'0000	0'0000	0'0001	0'0003	0'0016	0'0042	0'0098
	10					0'0000	0'0000	0'0001	0'0003	0'0010	
11	0	0'5688	0'3138	0'1673	0'0859	0'0422	0'0198	0'0116	0'0036	0'0014	0'0005
	1	0'03293	0'3835	0'3248	0'2362	0'1549	0'0932	0'0638	0'0266	0'0125	0'0054
	2	0'0867	0'2131	0'2866	0'2953	0'2581	0'1998	0'1590	0'0887	0'0513	0'0269
	3	0'0137	0'0710	0'1517	0'2215	0'2581	0'2568	0'2384	0'1774	0'1259	0'0806
	4	0'0014	0'0158	0'0536	0'1107	0'1721	0'2201	0'2384	0'2365	0'2060	0'1611
	5	0'0001	0'0025	0'0132	0'0388	0'0803	0'1321	0'1669	0'2207	0'2360	0'2256
	6	0'0000	0'0003	0'0023	0'0097	0'0268	0'0566	0'0835	0'1471	0'1931	0'2256
	7		0'0000	0'0003	0'0017	0'0064	0'0173	0'0298	0'0701	0'1128	0'1611
	8			0'0000	0'0002	0'0011	0'0037	0'0075	0'0234	0'0462	0'0806
	9				0'0000	0'0001	0'0005	0'0012	0'0052	0'0126	0'0269
	10					0'0000	0'0000	0'0001	0'0007	0'0021	0'0054
	11						0'0000	0'0000	0'0002	0'0005	
12	0	0'5404	0'2824	0'1422	0'0687	0'0317	0'0138	0'0077	0'0022	0'0008	0'0002
	1	0'3413	0'3766	0'3012	0'2062	0'1267	0'0712	0'0462	0'0174	0'0075	0'0029
	2	0'0988	0'2301	0'2924	0'2835	0'2323	0'1678	0'1272	0'0639	0'0339	0'0161
	3	0'0173	0'0852	0'1720	0'2362	0'2581	0'2397	0'2120	0'1419	0'0923	0'0537
	4	0'0021	0'0213	0'0683	0'1329	0'1936	0'2311	0'2384	0'2128	0'1700	0'1208
	5	0'0002	0'0038	0'0193	0'0532	0'1032	0'1585	0'1908	0'2270	0'2225	0'1934
	6	0'0000	0'0005	0'0040	0'0155	0'0401	0'0792	0'1113	0'1766	0'2124	0'2256
	7		0'0000	0'0006	0'0033	0'0115	0'0291	0'0477	0'1009	0'1489	0'1934
	8			0'0001	0'0005	0'0024	0'0078	0'0149	0'0420	0'0762	0'1208
	9				0'0000	0'0001	0'0004	0'0015	0'0033	0'0125	0'0277
	10					0'0000	0'0000	0'0002	0'0005	0'0025	0'0068
	11						0'0000	0'0000	0'0003	0'0010	0'0029
	12							0'0000	0'0001	0'0002	

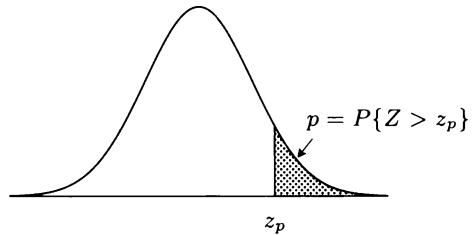
TABLA 2: Distribución de Poisson

$$P_\lambda\{X = x\} = \frac{e^{-\lambda} \lambda^x}{x!}$$

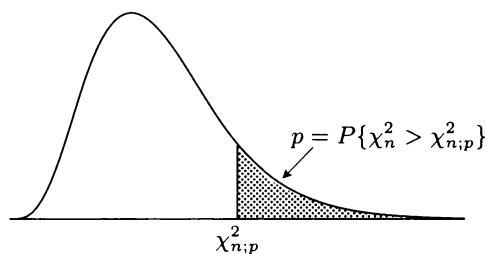
$x$	0	1	2	3	4	5	6	7	8	9	10	11
$\lambda$												
0'1	.9048	.0905	.0045	.0002	.0000							
0'2	.8187	.1637	.0164	.0011	.0001	.0000						
0'3	.7408	.2222	.0333	.0033	.0002	.0000						
0'4	.6703	.2681	.0536	.0072	.0007	.0001	.0000					
0'5	.6065	.3033	.0758	.0126	.0016	.0002	.0000					
0'6	.5488	.3293	.0988	.0198	.0030	.0004	.0000					
0'7	.4966	.3476	.1217	.0284	.0050	.0007	.0001	.0000				
0'8	.4493	.3595	.1438	.0383	.0077	.0012	.0002	.0000				
0'9	.4066	.3659	.1647	.0494	.0111	.0020	.0003	.0000				
1'0	.3679	.3679	.1839	.0613	.0153	.0031	.0005	.0001	.0000			
1'1	.3329	.3662	.2014	.0738	.0203	.0045	.0008	.0001	.0000			
1'2	.3012	.3614	.2169	.0867	.0260	.0062	.0012	.0002	.0000			
1'3	.2725	.3543	.2303	.0998	.0324	.0084	.0018	.0003	.0001	.0000		
1'4	.2466	.3452	.2417	.1128	.0395	.0111	.0026	.0005	.0001	.0000		
1'5	.2231	.3347	.2510	.1255	.0471	.0141	.0035	.0008	.0001	.0000		
1'6	.2019	.3230	.2584	.1378	.0551	.0176	.0047	.0011	.0002	.0000		
1'7	.1827	.3106	.2640	.1496	.0636	.0216	.0061	.0015	.0003	.0001	.0000	
1'8	.1653	.2975	.2678	.1607	.0723	.0260	.0078	.0020	.0005	.0001	.0000	
1'9	.1496	.2842	.2700	.1710	.0812	.0309	.0098	.0027	.0006	.0001	.0000	
2'0	.1353	.2707	.2707	.1804	.0902	.0361	.0120	.0034	.0009	.0002	.0000	
2'2	.1108	.2438	.2681	.1966	.1082	.0476	.0174	.0055	.0015	.0004	.0001	.0000
2'4	.0907	.2177	.2613	.2090	.1254	.0602	.0241	.0083	.0025	.0007	.0002	.0000
2'6	.0743	.1931	.2510	.2176	.1414	.0735	.0319	.0118	.0038	.0011	.0003	.0001
2'8	.0608	.1703	.2384	.2225	.1557	.0872	.0407	.0163	.0057	.0018	.0005	.0001
3'0	.0498	.1494	.2240	.2240	.1680	.1008	.0504	.0216	.0081	.0027	.0008	.0002
3'2	.0408	.1304	.2087	.2226	.1781	.1140	.0608	.0278	.0111	.0040	.0013	.0004
3'4	.0334	.1135	.1929	.2186	.1858	.1264	.0176	.0348	.0148	.0056	.0019	.0006
3'6	.0273	.0984	.1771	.2125	.1912	.1377	.0826	.0425	.0191	.0076	.0028	.0009
3'8	.0224	.0850	.1615	.2046	.1944	.1477	.0936	.0508	.0241	.0102	.0039	.0013
4'0	.0183	.0733	.1465	.1954	.1954	.1563	.1042	.0595	.0298	.0132	.0053	.0019
5'0	.0067	.0337	.0842	.1404	.1755	.1755	.1462	.1044	.0653	.0363	.0181	.0082
6'0	.0025	.0149	.0446	.0892	.1339	.1606	.1606	.1377	.1033	.0688	.0413	.0225
7'0	.0009	.0064	.0223	.0521	.0912	.1277	.1490	.1490	.1304	.1014	.0710	.0452
8'0	.0003	.0027	.0107	.0286	.0573	.0916	.1221	.1396	.1396	.1241	.0993	.0722
9'0	.0001	.0011	.0050	.0157	.0337	.0607	.0911	.1171	.1318	.1381	.1186	.0970
10'0	.0000	.0005	.0023	.0076	.0189	.0378	.0631	.0901	.1126	.1251	.1251	.1137

$x$	12	13	14	15	16	17	18	19	20	21	22	23
$\lambda$												
5'0	.0034	.0013	.0005	.0002	.0000							
6'0	.0113	.0052	.0022	.0009	.0003	.0001	.0000					
7'0	.0263	.0142	.0071	.0033	.0014	.0006	.0002	.0001	.0000			
8'0	.0481	.0296	.0169	.0090	.0045	.0021	.0009	.0004	.0002	.0001	.0000	
9'0	.0728	.0504	.0324	.0194	.0109	.0058	.0029	.0014	.0006	.0003	.0001	.0000
10'0	.0948	.0729	.0521	.0347	.0217	.0128	.0071	.0037	.0019	.0009	.0004	.0002

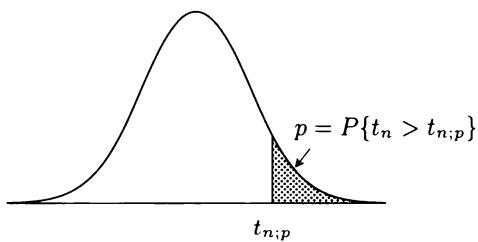
TABLA 3: Distribución Normal (0,1)



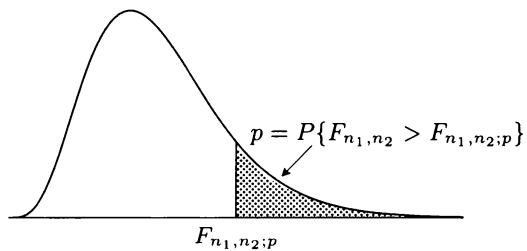
$z_p$	..0	..1	..2	..3	..4	..5	..6	..7	..8	..9
0'0	0'5000	0'4960	0'4920	0'4880	0'4840	0'4801	0'4761	0'4721	0'4681	0'4641
0'1	0'4602	0'4562	0'4522	0'4483	0'4443	0'4404	0'4364	0'4325	0'4286	0'4247
0'2	0'4207	0'4168	0'4129	0'4090	0'4052	0'4013	0'3974	0'3936	0'3897	0'3859
0'3	0'3821	0'3783	0'3745	0'3707	0'3669	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	0'2643	0'2611	0'2578	0'2546	0'2514	0'2483	0'2451
0'7	0'2420	0'2389	0'2358	0'2327	0'2297	0'2266	0'2231	0'2206	0'2177	0'2148
0'8	0'2119	0'2090	0'2061	0'2033	0'2005	0'1977	0'1949	0'1922	0'1984	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	0'1562	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	0'1093	0'1075	0'1056	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	0'0351	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
2'0	0'0228	0'0222	0'0217	0'0212	0'0207	0'0202	0'0197	0'0192	0'0188	0'0183
2'1	0'0179	0'0174	0'0170	0'0166	0'0162	0'0158	0'0154	0'0150	0'0146	0'0143
2'2	0'0139	0'0136	0'0132	0'0129	0'0125	0'0122	0'0119	0'0116	0'0113	0'0110
2'3	0'0107	0'0104	0'0102	0'0099	0'0096	0'0094	0'0091	0'0089	0'0087	0'0084
2'4	0'0082	0'0080	0'0078	0'0075	0'0073	0'0071	0'0069	0'0068	0'0066	0'0064
2'5	0'0062	0'0060	0'0059	0'0057	0'0055	0'0054	0'0052	0'0051	0'0049	0'0048
2'6	0'0047	0'0045	0'0044	0'0043	0'0041	0'0040	0'0039	0'0038	0'0037	0'0036
2'7	0'0035	0'0034	0'0033	0'0032	0'0031	0'0030	0'0029	0'0028	0'0027	0'0026
2'8	0'0026	0'0025	0'0024	0'0023	0'0023	0'0022	0'0021	0'0021	0'0020	0'0019
2'9	0'0019	0'0018	0'0018	0'0017	0'0016	0'0016	0'0015	0'0015	0'0014	0'0014
3'0	0'0013	0'0013	0'0013	0'0012	0'0012	0'0011	0'0011	0'0011	0'0010	0'0010
3'1	0'0010	0'0009	0'0009	0'0009	0'0008	0'0008	0'0008	0'0008	0'0007	0'0007
3'2	0'0007	0'0007	0'0006	0'0006	0'0006	0'0006	0'0006	0'0005	0'0005	0'0005
3'3	0'0005	0'0005	0'0005	0'0004	0'0004	0'0004	0'0004	0'0004	0'0004	0'0003
3'4	0'0003	0'0003	0'0003	0'0003	0'0003	0'0003	0'0003	0'0003	0'0003	0'0002

TABLA 4: Distribución  $\chi_n^2$  de Pearson

$p$	0'995	0'99	0'975	0'95	0'9	0'7	0'3	0'1	0'05	0'025	0'01	0'005
<i>n</i>												
1	0'0 <sup>4</sup> 4	0'0 <sup>3</sup> 2	0'001	0'004	0'016	0'148	1'074	2'706	3'841	5'024	6'635	7'879
2	0'010	0'020	0'051	0'103	0'211	0'713	2'408	4'605	5'991	7'378	9'210	10'60
3	0'072	0'115	0'216	0'352	0'584	1'424	3'665	6'251	7'815	9'348	11'34	12'84
4	0'207	0'297	0'484	0'711	1'064	2'195	4'878	7'779	9'488	11'14	13'28	14'86
5	0'412	0'554	0'831	1'145	1'610	3'000	6'064	9'236	11'07	12'83	15'09	16'75
6	0'676	0'872	1'237	1'635	2'204	3'828	7'231	10'65	12'59	14'45	16'81	18'55
7	0'989	1'239	1'690	2'167	2'833	4'671	8'383	12'02	14'07	16'01	18'48	20'28
8	1'344	1'646	2'180	2'733	3'490	5'527	9'524	13'36	15'51	17'54	20'09	21'96
9	1'735	2'088	2'700	3'325	4'168	6'393	10'66	14'68	16'92	19'02	21'67	23'59
10	2'156	2'558	3'247	3'940	4'865	7'267	11'78	15'99	18'31	20'48	23'21	25'19
11	2'603	3'053	3'816	4'574	5'578	8'148	12'90	17'28	19'68	21'92	24'73	26'76
12	3'074	3'571	4'404	5'226	6'304	9'034	14'01	18'55	21'03	23'34	26'22	28'30
13	3'565	4'107	5'009	5'892	7'042	9'926	15'12	19'81	22'36	24'74	27'69	29'82
14	4'075	4'660	5'629	6'571	7'790	10'82	16'22	21'06	23'69	26'12	29'14	31'32
15	4'601	5'229	6'262	7'261	8'547	11'72	17'32	22'31	25'00	27'49	30'58	32'80
16	5'142	5'812	6'908	7'962	9'312	12'62	18'42	23'54	26'30	28'85	32'00	34'27
17	5'697	6'408	7'564	8'672	10'09	13'53	19'51	24'67	27'59	30'19	33'41	35'72
18	6'265	7'015	8'231	9'390	10'87	14'44	20'60	25'99	28'87	31'53	34'81	37'16
19	6'844	7'633	8'906	10'12	11'65	15'35	21'69	27'20	30'14	32'85	36'19	38'58
20	7'434	8'260	9'591	10'85	12'44	16'27	22'78	28'41	31'41	34'17	37'57	40'00
21	8'034	8'897	10'28	11'59	13'24	17'18	23'86	29'62	32'67	35'48	38'93	41'40
22	8'643	9'542	10'98	12'34	14'04	18'10	24'94	30'81	33'92	36'78	40'29	42'80
23	9'260	10'20	11'69	13'09	14'85	19'02	26'02	32'01	35'17	38'08	41'64	44'18
24	9'886	10'86	12'40	13'85	15'66	19'94	27'10	33'20	36'42	39'36	42'98	45'56
25	10'52	11'52	13'12	14'61	16'47	20'87	28'17	34'38	37'65	40'65	44'31	46'93
26	11'16	12'20	13'84	15'38	17'29	21'79	29'25	35'56	38'89	41'92	45'64	48'29
27	11'81	12'88	14'57	16'15	18'11	22'72	30'32	36'74	40'11	43'20	46'96	49'65
28	12'46	13'56	15'31	16'93	18'94	23'65	31'39	37'92	41'34	44'46	48'28	50'99
29	13'12	14'26	16'05	17'71	19'77	24'58	32'46	39'09	42'56	45'72	49'59	52'34
30	13'79	14'95	16'79	18'49	20'60	25'51	33'53	40'26	43'77	46'98	50'89	53'67

TABLA 5: Distribución  $t$  de Student

$p$	0'4	0'3	0'2	0'1	0'05	0'025	0'01	0'005	0'0025
$n$									
1	0'325	0'727	1'376	3'078	6'314	12'706	31'821	63'657	127'32
2	0'289	0'617	1'061	1'886	2'920	4'303	6'965	9'925	14'089
3	0'277	0'584	0'978	1'638	2'353	3'182	4'541	5'841	7'453
4	0'271	0'569	0'941	1'533	2'132	2'776	3'747	4'604	5'598
5	0'267	0'559	0'920	1'476	2'015	2'571	3'365	4'032	4'773
6	0'265	0'553	0'906	1'440	1'943	2'447	3'143	3'707	4'317
7	0'263	0'549	0'896	1'415	1'895	2'365	2'998	3'499	4'029
8	0'262	0'546	0'889	1'397	1'860	2'306	2'896	3'355	3'833
9	0'261	0'543	0'883	1'383	1'833	2'262	2'821	3'250	3'690
10	0'260	0'542	0'879	1'372	1'812	2'228	2'764	3'169	3'581
11	0'260	0'540	0'876	1'363	1'796	2'201	2'718	3'106	3'497
12	0'259	0'539	0'873	1'356	1'782	2'179	2'681	3'055	3'428
13	0'259	0'538	0'870	1'350	1'771	2'160	2'650	3'012	3'372
14	0'258	0'537	0'868	1'345	1'761	2'145	2'624	2'977	3'326
15	0'258	0'536	0'866	1'341	1'753	2'131	2'602	2'947	3'286
16	0'258	0'535	0'865	1'337	1'746	2'120	2'583	2'921	3'252
17	0'257	0'534	0'863	1'333	1'740	2'110	2'567	2'898	3'222
18	0'257	0'534	0'862	1'330	1'734	2'101	2'552	2'878	3'197
19	0'257	0'533	0'861	1'328	1'729	2'093	2'539	2'861	3'174
20	0'257	0'533	0'860	1'325	1'725	2'086	2'528	2'845	3'153
21	0'257	0'532	0'859	1'323	1'721	2'080	2'518	2'831	3'135
22	0'256	0'532	0'858	1'321	1'717	2'074	2'508	2'819	3'119
23	0'256	0'532	0'858	1'319	1'714	2'069	2'500	2'807	3'104
24	0'256	0'531	0'857	1'318	1'711	2'064	2'492	2'797	3'091
25	0'256	0'531	0'856	1'316	1'708	2'060	2'485	2'787	3'078
26	0'256	0'531	0'856	1'315	1'706	2'056	2'479	2'779	3'067
27	0'256	0'531	0'855	1'314	1'703	2'052	2'473	2'771	3'057
28	0'256	0'530	0'855	1'313	1'701	2'048	2'467	2'763	3'047
29	0'256	0'530	0'854	1'311	1'699	2'045	2'462	2'756	3'038
30	0'256	0'530	0'854	1'310	1'697	2'042	2'457	2'750	3'030
40	0'255	0'529	0'851	1'303	1'684	2'021	2'423	2'704	2'971
60	0'254	0'527	0'848	1'296	1'671	2'000	2'390	2'660	2'915
120	0'254	0'526	0'845	1'289	1'658	1'980	2'358	2'617	2'860
$\infty$	0'253	0'524	0'842	1'282	1'645	1'960	2'326	2'576	2'807

TABLA 6: Distribución  $F$  de Snedecor

	$n_1$	1	2	3	4	5	6	7	8	9	
$n_2$	$p$	0'100	39'864	49'500	53'593	55'833	57'241	58'204	58'906	59'439	59'858
1	0'050	161'45	199'50	215'71	224'58	230'16	233'99	236'77	238'88	240'54	
	0'025	647'79	799'50	864'16	899'58	921'85	937'11	948'22	956'66	963'28	
	0'010	4052'2	4999'5	5403'4	5624'6	5763'6	5859'0	5928'4	5981'1	6022'5	
	0'005	16211	20000	21615	22500	23056	23437	23715	23925	24091	
	0'100	8'5263	9'0000	9'1618	9'2434	9'2926	9'3255	9'3491	9'3668	9'3805	
2	0'050	18'513	19'000	19'164	19'247	19'296	19'330	19'353	19'371	19'385	
	0'025	38'506	39'000	39'165	39'248	39'298	39'331	39'355	39'373	39'387	
	0'010	98'503	99'000	99'166	99'249	99'299	99'332	99'356	99'374	99'388	
	0'005	198'50	199'00	199'17	199'25	199'30	199'33	199'36	199'37	199'39	
	0'100	5'5383	5'4624	5'3908	5'3427	5'3092	5'2847	5'2662	5'2517	5'2400	
3	0'050	10'128	9'5521	9'2766	9'1172	9'0135	8'9406	8'8868	8'8452	8'8123	
	0'025	17'443	16'044	15'439	15'101	14'885	14'735	14'624	14'540	14'473	
	0'010	34'116	30'817	29'457	28'710	28'237	27'911	27'672	27'489	27'345	
	0'005	55'552	49'799	47'467	46'195	45'392	44'838	44'434	44'126	43'882	
	0'100	4'5448	4'3246	4'1908	4'1073	4'0506	4'0098	3'9790	3'9549	3'9357	
4	0'050	7'7086	6'9443	6'5914	6'3883	6'2560	6'1631	6'0942	6'0410	5'9988	
	0'025	12'218	10'649	9'9792	9'6045	9'3645	9'1973	9'0741	8'9796	8'9047	
	0'010	21'198	18'000	16'694	15'977	15'522	15'207	14'976	14'799	14'659	
	0'005	31'333	26'284	24'259	23'155	22'456	21'975	21'622	21'352	21'139	
	0'100	4'0604	3'7797	3'6195	3'5202	3'4530	3'4045	3'3679	3'3393	3'3163	
5	0'050	6'6079	5'7861	5'4095	5'1922	5'0503	4'9503	4'8759	4'8183	4'7725	
	0'025	10'007	8'4336	7'7636	7'3879	7'1464	6'9777	6'8531	6'7572	6'6810	
	0'010	16'258	13'274	12'060	11'392	10'967	10'672	10'456	10'289	10'158	
	0'005	22'785	18'314	16'530	15'556	14'940	14'513	14'200	13'961	13'772	
	0'100	3'7760	3'4633	3'2888	3'1808	3'1075	3'0546	3'0145	2'9830	2'9577	
6	0'050	5'9874	5'1433	4'7571	4'5337	4'3874	4'2839	4'2066	4'1468	4'0990	
	0'025	8'8131	7'2598	6'5988	6'2272	5'9876	5'8197	5'6955	5'5996	5'5234	
	0'010	13'745	10'925	9'7795	9'1483	8'7459	8'4661	8'2600	8'1016	7'9761	
	0'005	18'635	14'544	12'917	12'028	11'464	11'073	10'786	10'566	10'391	
	0'100	3'5894	3'2574	3'0741	2'9605	2'8833	2'8274	2'7849	2'7516	2'7247	
7	0'050	5'5914	4'7374	4'3468	4'1203	3'9715	3'8660	3'7870	3'7257	3'6767	
	0'025	8'0727	6'5415	5'8898	5'5226	5'2852	5'1186	4'9949	4'8994	4'8232	
	0'010	12'246	9'5466	8'4513	7'8467	7'4604	7'1914	6'9928	6'8401	6'7188	
	0'005	16'236	12'404	10'882	10'050	9'5221	9'1554	8'8854	8'6781	8'5138	

TABLA 6: Distribución  $F$  de Snedecor (continuación)

10	12	15	18	20	24	30	40	60	120	$\infty$
60'195	60'705	61'220	61'567	61'740	62'002	62'265	62'529	62'794	63'061	63'328
241'88	243'91	245'95	247'32	248'01	249'05	250'09	251'14	252'20	253'25	254'32
968'63	976'71	984'87	990'36	993'10	997'25	1001'4	1005'6	1009'8	1014'0	1018'3
6055'8	6106'3	6157'3	6191'6	6208'7	6234'6	6260'7	6286'8	6313'0	6339'4	6366'0
24224	24426	24630	24767	24836	24940	25044	25148	25253	25359	25465
9'3916	9'4081	9'4247	9'4358	9'4413	9'4496	9'4579	9'4663	9'4746	9'4829	9'4913
19'396	19'413	19'429	19'440	19'446	19'454	19'462	19'471	19'479	19'487	19'496
39'398	39'415	39'431	39'442	39'448	39'456	39'465	39'473	39'481	39'490	39'498
99'399	99'416	99'432	99'443	99'449	99'458	99'466	99'474	99'483	99'491	99'499
199'40	199'42	199'43	199'44	199'45	199'46	199'47	199'47	199'48	199'49	199'51
5'2304	5'2156	5'2003	5'1898	5'1845	5'1764	5'1681	5'1597	5'1512	5'1425	5'1337
8'7855	8'7446	8'7029	8'6744	8'6602	8'6385	8'6166	8'5944	8'5720	8'5494	8'5265
14'419	14'337	14'253	14'196	14'167	14'124	14'081	14'037	13'992	13'947	13'902
27'229	27'052	26'872	26'751	26'690	26'598	26'505	26'411	26'316	26'221	26'125
43'686	43'387	43'085	42'880	42'778	42'622	42'466	42'308	42'149	41'989	41'829
3'9199	3'8955	3'8689	3'8525	3'8443	3'8310	3'8174	3'8036	3'7896	3'7753	3'7607
5'9644	5'9117	5'8578	5'8209	5'8025	5'7744	5'7459	5'7170	5'6878	5'6581	5'6281
8'8439	8'7512	8'6565	8'5921	8'5599	8'5109	8'4613	8'4111	8'3604	8'3092	8'2573
14'546	14'374	14'198	14'079	14'020	13'929	13'838	13'745	13'652	13'558	13'463
20'967	20'705	20'438	20'257	20'167	20'030	19'892	19'752	19'611	19'468	19'325
3'2974	3'2682	3'2380	3'2171	3'2067	3'1905	3'1741	3'1573	3'1402	3'1228	3'1050
4'7351	4'6777	4'6188	4'5783	4'5581	4'5272	4'4957	4'4638	4'4314	4'3984	4'3650
6'6192	6'5246	6'4277	6'3616	6'3285	6'2780	6'2269	6'1751	6'1225	6'0693	6'0153
10'051	9'8883	9'7222	9'6092	9'5527	9'4665	9'3793	9'2912	9'2020	9'1118	9'0204
13'618	13'384	13'146	12'984	12'903	12'780	12'656	12'530	12'402	12'274	12'144
2'9369	2'9047	2'8712	2'8479	2'8363	2'8183	2'8000	2'7812	2'7620	2'7423	2'7222
4'0600	3'9999	3'9381	3'8955	3'8742	3'8415	3'8082	3'7743	3'7398	3'7047	3'6688
5'4613	5'3662	5'2687	5'2018	5'1684	5'1172	5'0652	5'0125	4'9589	4'9045	4'8491
7'8741	7'7183	7'5590	7'4502	7'3958	7'3127	7'2285	7'1432	7'0568	6'9690	6'8801
10'250	10'034	9'8140	9'6639	9'5888	9'4741	9'3583	9'2408	9'1219	9'0015	8'8793
2'7025	2'6681	2'6322	2'6072	2'5947	2'5753	2'5555	2'5351	2'5142	2'4928	2'4708
3'6365	3'5747	3'5108	3'4666	3'4445	3'4105	3'3758	3'3404	3'3043	3'2674	3'2298
4'7611	4'6658	4'5678	4'5004	4'4667	4'4150	4'3624	4'3089	4'2544	4'1989	4'1423
6'6201	6'4691	6'3143	6'2084	6'1554	6'0743	5'9921	5'9084	5'8236	5'7372	5'6495
8'3803	8'1764	7'9678	7'8253	7'7540	7'6450	7'5345	7'4225	7'3088	7'1933	7'0760

TABLA 6: Distribución  $F$  de Snedecor (continuación)

	$n_1$	1	2	3	4	5	6	7	8	9
	$n_2$	$p$								
8	0'100	3'4579	3'1131	2'9238	2'8064	2'7265	2'6683	2'6241	2'5893	2'5612
	0'050	5'3177	4'4590	4'0662	3'8378	3'6875	3'5806	3'5005	3'4381	3'3881
	0'025	7'5709	6'0595	5'4160	5'0526	4'8173	4'6517	4'5286	4'4332	4'3572
	0'010	11'259	8'6491	7'5910	7'0060	6'6318	6'3707	6'1776	6'0289	5'9106
	0'005	14'688	11'042	9'5965	8'8051	8'3018	7'9520	7'6942	7'4960	7'3386
9	0'100	3'3603	3'0065	2'8129	2'6927	2'6106	2'5509	2'5053	2'4694	2'4403
	0'050	5'1174	4'2563	3'8626	3'6331	3'4817	3'3738	3'2927	3'2296	3'1789
	0'025	7'2093	5'7147	5'0781	4'7181	4'4844	4'3197	4'1971	4'1020	4'0260
	0'010	10'561	8'0215	6'9919	6'4221	6'0569	5'8018	5'6129	5'4671	5'3511
	0'005	13'614	10'107	8'7171	7'9559	7'4711	7'1338	6'8849	6'6933	6'5411
10	0'100	3'2850	2'9245	2'7277	2'6053	2'5216	2'4606	2'4140	2'3772	2'3473
	0'050	4'9646	4'1028	3'7083	3'4780	3'3258	3'2172	3'1355	3'0717	3'0204
	0'025	6'9367	5'4564	4'8256	4'4683	4'2361	4'0721	3'9498	3'8549	3'7790
	0'010	10'044	7'5594	6'5523	5'9943	5'6363	5'3858	5'2201	5'0567	4'9424
	0'005	12'826	9'4270	8'0807	7'3428	6'8723	6'5446	6'3025	6'1159	5'9676
11	0'100	3'2252	2'8595	2'6602	2'5362	2'4512	2'3891	2'3416	2'3040	2'2735
	0'050	4'8443	3'9823	3'5874	3'3567	3'2039	3'0946	3'0123	2'9480	2'8962
	0'025	6'7241	5'2559	4'6300	4'2751	4'0440	3'8807	3'7586	3'6638	3'5879
	0'010	9'6460	7'2057	6'2167	5'6683	5'3160	5'0692	4'8861	4'7445	4'6315
	0'005	12'226	8'9122	7'6004	6'8809	6'4217	6'1015	5'8648	5'6821	5'5368
12	0'100	3'1765	2'8068	2'6055	2'4801	2'3940	2'3310	2'2828	2'2446	2'2135
	0'050	4'7472	3'8853	3'4903	3'2592	3'1059	2'9961	2'9134	2'8486	2'7964
	0'025	6'5538	5'0959	4'4742	4'1212	3'8911	3'7283	3'6065	3'5118	3'4358
	0'010	9'3302	6'9266	5'9526	5'4119	5'0643	4'8206	4'6395	4'4994	4'3875
	0'005	11'754	8'5096	7'2258	6'5211	6'0711	5'7570	5'5245	5'3451	5'2021
13	0'100	3'1362	2'7632	2'5603	2'4337	2'3467	2'2830	2'2341	2'1953	2'1638
	0'050	4'6672	3'8056	3'4105	3'1791	3'0254	2'9153	2'8321	2'7669	2'7144
	0'025	6'4143	4'9653	4'3472	3'9959	3'7667	3'6043	3'4827	3'3880	3'3120
	0'010	9'0738	6'7010	5'7394	5'2053	4'8616	4'6204	4'4410	4'3021	4'1911
	0'005	11'374	8'1865	6'9257	6'2335	5'7910	5'4819	5'2529	5'0761	4'9351
14	0'100	3'1022	2'7265	2'5222	2'3947	2'3069	2'2426	2'1931	2'1539	2'1220
	0'050	4'6001	3'7389	3'3439	3'1122	2'9582	2'8477	2'7642	2'6987	2'6458
	0'025	6'2979	4'8567	4'2417	3'8919	3'6634	3'5014	3'3799	3'2853	3'2093
	0'010	8'8616	6'5149	5'5639	5'0354	4'6950	4'4558	4'2779	4'1399	4'0297
	0'005	11'060	7'9216	6'6803	5'9984	5'5623	5'2574	5'0313	4'8566	4'7173
15	0'100	3'0732	2'6952	2'4898	2'3614	2'2730	2'2081	2'1582	2'1185	2'0862
	0'050	4'5431	3'6823	3'2874	3'0556	2'9013	2'7905	2'7066	2'6408	2'5876
	0'025	6'1995	4'7650	4'1528	3'8043	3'5764	3'4147	3'2934	3'1987	3'1227
	0'010	8'6831	6'3589	5'4170	4'8932	4'5556	4'3183	4'1415	4'0045	3'8948
	0'005	10'798	7'7008	6'4760	5'8029	5'3721	5'0708	4'8473	4'6743	4'5364
16	0'100	3'0481	2'6682	2'4618	2'3327	2'2438	2'1783	2'1280	2'0880	2'0553
	0'050	4'4940	3'6337	3'2389	3'0069	2'8524	2'7413	2'6572	2'5911	2'5377
	0'025	6'1151	4'6867	4'0768	3'7294	3'5021	3'3406	3'2194	3'1248	3'0488
	0'010	8'5310	6'2262	5'2922	4'7726	4'4374	4'2016	4'0259	3'8896	3'7804
	0'005	10'575	7'5138	6'3034	5'6378	5'2117	4'9134	4'6920	4'5207	4'3838

TABLA 6: Distribución  $F$  de Snedecor (continuación)

10	12	15	18	20	24	30	40	60	120	$\infty$
2'5380	2'5020	2'4642	2'4378	2'4246	2'4041	2'3830	2'3614	2'3391	2'3162	2'2926
3'3472	3'2840	3'2184	3'1730	3'1503	3'1152	3'0794	3'0428	3'0053	2'9669	2'9276
4'2951	4'1997	4'1012	4'0334	3'9995	3'9472	3'8940	3'8398	3'7844	3'7279	3'6702
5'8143	5'6668	5'5151	5'4111	5'3591	5'2793	5'1981	5'1156	5'0316	4'9460	4'8588
7'2107	7'0149	6'8143	6'6769	6'6082	6'5029	6'3961	6'2875	6'1772	6'0649	5'9505
2'4163	2'3789	2'3396	2'3121	2'2983	2'2768	2'2547	2'2320	2'2085	2'1843	2'1592
3'1373	3'0729	3'0061	2'9597	2'9365	2'9005	2'8637	2'8259	2'7872	2'7475	2'7067
3'9639	3'8682	3'7694	3'7011	3'6669	3'6142	3'5604	3'5055	3'4493	3'3918	3'3329
5'2565	5'1114	4'9621	4'8594	4'8080	4'7290	4'6486	4'5667	4'4831	4'3978	4'3105
6'4171	6'2274	6'0325	5'8987	5'8318	5'7292	5'6248	5'5186	5'4104	5'3001	5'1875
2'3226	2'2841	2'2435	2'2150	2'2007	2'1784	2'1554	2'1317	2'1072	2'0818	2'0554
2'9782	2'9130	2'8450	2'7977	2'7740	2'7372	2'6996	2'6609	2'6211	2'5801	2'5379
3'7168	3'6209	3'5217	3'4530	3'4186	3'3654	3'3110	3'2554	3'1984	3'1399	3'0798
4'8492	4'7059	4'5582	4'4563	4'4054	4'3269	4'2469	4'1653	4'0819	3'9965	3'9090
5'8467	5'6613	5'4707	5'3396	5'2740	5'1732	5'0705	4'9659	4'8592	4'7501	4'6385
2'2482	2'2087	2'1671	2'1377	2'1230	2'1000	2'0762	2'0516	2'0261	1'9997	1'9721
2'8536	2'7876	2'7186	2'6705	2'6464	2'6090	2'5705	2'5309	2'4901	2'4480	2'4045
3'5257	3'4296	3'3299	3'2607	3'2261	3'1725	3'1176	3'0613	3'0035	2'9441	2'8828
4'5393	4'3974	4'2509	4'1496	4'0990	4'0209	3'9411	3'8596	3'7761	3'6904	3'6025
5'4182	5'2363	5'0489	4'9198	4'8552	4'7557	4'6543	4'5508	4'4450	4'3367	4'2256
2'1878	2'1474	2'1049	2'0748	2'0597	2'0360	2'0115	1'9861	1'9597	1'9323	1'9036
2'7534	2'6866	2'6169	2'5680	2'5436	2'5055	2'4663	2'4259	2'3842	2'3410	2'2962
3'3736	3'2773	3'1772	3'1076	3'0728	3'0187	2'9633	2'9063	2'8478	2'7874	2'7249
4'2961	4'1553	4'0096	3'9088	3'8584	3'7805	3'7008	3'6192	3'5355	3'4494	3'3608
5'0855	4'9063	4'7214	4'5937	4'5299	4'4315	4'3309	4'2282	4'1229	4'0149	3'9039
2'1376	2'0966	2'0532	2'0224	2'0070	1'9827	1'9576	1'9315	1'9043	1'8759	1'8462
2'6710	2'6037	2'5331	2'4836	2'4589	2'4202	2'3803	2'3392	2'2966	2'2524	2'2064
3'2497	3'1532	3'0527	2'9827	2'9477	2'8932	2'8373	2'7797	2'7204	2'6590	2'5955
4'1003	3'9603	3'8154	3'7149	3'6646	3'5868	3'5070	3'4253	3'3413	3'2548	3'1654
4'8199	4'6429	4'4600	4'3335	4'2703	4'1726	4'0727	3'9704	3'8655	3'7577	3'6465
2'0954	2'0537	2'0095	1'9782	1'9625	1'9377	1'9119	1'8852	1'8572	1'8280	1'7973
2'6021	2'5342	2'4630	2'4129	2'3879	2'3487	2'3082	2'2664	2'2230	2'1778	2'1307
3'1469	3'0501	2'9493	2'8789	2'8437	2'7888	2'7324	2'6742	2'6142	2'5519	2'4872
3'9394	3'8001	3'6557	3'5554	3'5052	3'4274	3'3476	3'2656	3'1813	3'0942	3'0040
4'6034	4'4281	4'2468	4'1213	4'0585	3'9614	3'8619	3'7600	3'6553	3'5473	3'4359
2'0593	2'0171	1'9722	1'9403	1'9243	1'8990	1'8728	1'8454	1'8168	1'7867	1'7551
2'5437	2'4753	2'4035	2'3528	2'3275	2'2878	2'2468	2'2043	2'1601	2'1141	2'0658
3'0602	2'9633	2'8621	2'7913	2'7559	2'7006	2'6437	2'5850	2'5242	2'4611	2'3953
3'8049	3'6662	3'5222	3'4220	3'3719	3'2940	3'2141	3'1319	3'0471	2'9595	2'8684
4'4236	4'2498	4'0698	3'9450	3'8826	3'7859	3'6867	3'5850	3'4803	3'3722	3'2602
2'0281	1'9854	1'9399	1'9075	1'8913	1'8656	1'8388	1'8108	1'7816	1'7507	1'7182
2'4935	2'4247	2'3522	2'3011	2'2756	2'2354	2'1938	2'1507	2'1058	2'0589	2'0096
2'9862	2'8890	2'7875	2'7164	2'6808	2'6252	2'5678	2'5085	2'4471	2'3831	2'3163
3'6909	3'5527	3'4089	3'3088	3'2588	3'1808	3'1007	3'0182	2'9330	2'8447	2'7528
4'2719	4'0994	3'9205	3'7963	3'7342	3'6378	3'5388	3'4372	3'3324	3'2240	3'1115

TABLA 6: Distribución  $F$  de Snedecor (continuación)

$n_1$	1	2	3	4	5	6	7	8	9	
$n_2$	$p$									
17	0'100	3'0262	2'6446	2'4374	2'3077	2'2183	2'1524	2'1017	2'0613	2'0284
	0'050	4'4513	3'5915	3'1968	2'9647	2'8100	2'6987	2'6143	2'5480	2'4943
	0'025	6'0420	4'6189	4'0112	3'6648	3'4379	3'2767	3'1556	3'0610	2'9849
	0'010	8'3997	6'1121	5'1850	4'6690	4'3359	4'1015	3'9267	3'7910	3'6822
	0'005	10'384	7'3536	6'1556	5'4967	5'0746	4'7789	4'5594	4'3893	4'2535
18	0'100	3'0070	2'6239	2'4160	2'2858	2'1958	2'1296	2'0785	2'0379	2'0047
	0'050	4'4139	3'5546	3'1599	2'9277	2'7729	2'6613	2'5767	2'5102	2'4563
	0'025	5'9781	4'5597	3'9539	3'6083	3'3820	3'2209	3'0999	3'0053	2'9291
	0'010	8'2854	6'0129	5'0919	4'5790	4'2479	4'0146	3'8406	3'7054	3'5971
	0'005	10'218	7'2148	6'0277	5'3746	4'9560	4'6827	4'4448	4'2759	4'1410
19	0'100	2'9899	2'6056	2'3970	2'2663	2'1760	2'1094	2'0580	2'0171	1'9836
	0'050	4'3808	3'5219	3'1274	2'8951	2'7401	2'6283	2'5435	2'4768	2'4227
	0'025	5'9216	4'5075	3'9034	3'5587	3'3327	3'1718	3'0509	2'9563	2'8800
	0'010	8'1850	5'9259	5'0103	4'5003	4'1708	3'9386	3'7653	3'6305	3'5225
	0'005	10'073	7'0935	5'9161	5'2681	4'8526	4'5614	4'3448	4'1770	4'0428
20	0'100	2'9747	2'5893	2'3801	2'2489	2'1582	2'0913	2'0397	1'9985	1'9649
	0'050	4'3513	3'4928	3'0984	2'8661	2'7109	2'5990	2'5140	2'4471	2'3928
	0'025	5'8715	4'4613	3'8587	3'5147	3'2891	3'1283	3'0074	2'9128	2'8365
	0'010	8'0960	5'8489	4'9382	4'4307	4'1027	3'8714	3'6987	3'5644	3'4567
	0'005	9'9439	6'9865	5'8177	5'1743	4'7616	4'4721	4'2569	4'0900	3'9564
21	0'100	2'9609	2'5746	2'3649	2'2333	2'1423	2'0751	2'0232	1'9819	1'9480
	0'050	4'3248	3'4668	3'0725	2'8401	2'6848	2'5727	2'4876	2'4205	2'3661
	0'025	5'8266	4'4199	3'8188	3'4754	3'2501	3'0895	2'9686	2'8740	2'7977
	0'010	8'0166	5'7804	4'8740	4'3688	4'0421	3'8117	3'6396	3'5056	3'3981
	0'005	9'8295	6'8914	5'7304	5'0911	4'6808	4'3931	4'1789	4'0128	3'8799
22	0'100	2'9486	2'5613	2'3512	2'2193	2'1279	2'0605	2'0084	1'9668	1'9327
	0'050	4'3009	3'4434	3'0491	2'8167	2'6613	2'5491	2'4638	2'3965	2'3419
	0'025	5'7863	4'3828	3'7829	3'4401	3'2151	3'0546	2'9338	2'8392	2'7628
	0'010	7'9454	5'7190	4'8166	4'3134	3'9880	3'7583	3'5867	3'4530	3'3458
	0'005	9'7271	6'8064	5'6524	5'0168	4'6088	4'3225	4'1094	3'9440	3'8116
23	0'100	2'9374	2'5493	2'3387	2'2065	2'1149	2'0472	1'9949	1'9531	1'9189
	0'050	4'2793	3'4221	3'0280	2'7955	2'6400	2'5277	2'4422	2'3748	2'3201
	0'025	5'7498	4'3492	3'7505	3'4083	3'1835	3'0232	2'9024	2'8077	2'7313
	0'010	7'8811	5'6637	4'7649	4'2635	3'9392	3'7102	3'5390	3'4057	3'2986
	0'005	9'6348	6'7300	5'5823	4'9500	4'5441	4'2591	4'0469	3'8822	3'7502
24	0'100	2'9271	2'5383	2'3274	2'1949	2'1030	2'0351	1'9826	1'9407	1'9063
	0'050	4'2597	3'4028	3'0088	2'7763	2'6207	2'5082	2'4226	2'3551	2'3002
	0'025	5'7167	4'3187	3'7211	3'3794	3'1543	2'9946	2'8738	2'7791	2'7027
	0'010	7'8229	5'6136	4'7181	4'2184	3'8951	3'6667	3'4959	3'3629	3'2560
	0'005	9'5513	6'6609	5'5190	4'8898	4'4857	4'2019	3'9905	3'8264	3'6949
25	0'100	2'9177	2'5283	2'3170	2'1843	2'0922	2'0241	1'9714	1'9292	1'8947
	0'050	4'2417	3'3852	2'9912	2'7587	2'6030	2'4904	2'4047	2'3371	2'2821
	0'025	5'6864	4'2909	3'6943	3'3530	3'1287	2'9685	2'8478	2'7531	2'6766
	0'010	7'7698	5'5680	4'6755	4'1774	3'8550	3'6272	3'4568	3'3239	3'2172
	0'005	9'4753	6'5982	5'4615	4'8351	4'4327	4'1500	3'9394	3'7758	3'6447

TABLA 6: Distribución  $F$  de Snedecor (continuación)

10	12	15	18	20	24	30	40	60	120	$\infty$
2'0009	1'9577	1'9117	1'8788	1'8624	1'8362	1'8090	1'7805	1'7506	1'7191	1'6856
2'4499	2'3807	2'3077	2'2562	2'2304	2'1898	2'1477	2'1040	2'0584	2'0107	1'9604
2'9222	2'8249	2'7230	2'6515	2'6158	2'5598	2'5021	2'4422	2'3801	2'3153	2'2474
3'5931	3'4552	3'3117	3'2116	3'1615	3'0835	3'0032	2'9205	2'8348	2'7459	2'6530
4'1423	3'9709	3'7929	3'6692	3'6073	3'5112	3'4124	3'3107	3'2058	3'0971	2'9839
1'9770	1'9333	1'8868	1'8535	1'8368	1'8103	1'7827	1'7537	1'7232	1'6910	1'6567
2'4117	2'3421	2'2686	2'2166	2'1906	2'1497	2'1071	2'0629	2'0166	1'9681	1'9168
2'8664	2'7689	2'6667	2'5949	2'5590	2'5027	2'4445	2'3842	2'3214	2'2558	2'1869
3'5082	3'3706	3'2273	3'1272	3'0771	2'9990	2'9185	2'8354	2'7493	2'6597	2'5660
4'0305	3'8599	3'6827	3'5594	3'4977	3'4017	3'3030	3'2014	3'0962	2'9871	2'8732
1'9557	1'9117	1'8647	1'8310	1'8142	1'7873	1'7592	1'7298	1'6988	1'6659	1'6308
2'3779	2'3080	2'2341	2'1817	2'1555	2'1141	2'0712	2'0264	1'9796	1'9302	1'8780
2'8173	2'7196	2'6171	2'5450	2'5089	2'4523	2'3937	2'3329	2'2695	2'2032	2'1333
3'4338	3'2965	3'1533	3'0532	3'0031	2'9249	2'8442	2'7608	2'6742	2'5839	2'4893
3'9329	3'7631	3'5866	3'4635	3'4020	3'3062	3'2075	3'1058	3'0004	2'8908	2'7762
1'9367	1'8924	1'8449	1'8108	1'7938	1'7667	1'7382	1'7083	1'6768	1'6433	1'6074
2'3479	2'2776	2'2033	2'1506	2'1242	2'0825	2'0391	1'9938	1'9464	1'8963	1'8432
2'7737	2'6758	2'5731	2'5007	2'4645	2'4076	2'3486	2'2873	2'2234	2'1562	2'0853
3'3682	3'2311	3'0880	2'9878	2'9377	2'8594	2'7785	2'6947	2'6077	2'5168	2'4212
3'8470	3'6779	3'5020	3'3792	3'3178	3'2220	3'1234	3'0215	2'9159	2'8058	2'6904
1'9197	1'8750	1'8272	1'7928	1'7756	1'7481	1'7193	1'6890	1'6569	1'6228	1'5862
2'3210	2'2504	2'1757	2'1226	2'0960	2'0540	2'0102	1'9645	1'9165	1'8657	1'8117
2'7348	2'6368	2'5338	2'4611	2'4247	2'3675	2'3082	2'2465	2'1819	2'1141	2'0422
3'3098	3'1729	3'0299	2'9297	2'8796	2'8011	2'7200	2'6359	2'5484	2'4568	2'3603
3'7709	3'6024	3'4270	3'3044	3'2431	3'1474	3'0488	2'9467	2'8408	2'7302	2'6140
1'9043	1'8593	1'8111	1'7764	1'7590	1'7312	1'7021	1'6714	1'6389	1'6042	1'5668
2'2967	2'2258	2'1508	2'0974	2'0707	2'0283	1'9842	1'9380	1'8895	1'8380	1'7831
2'6998	2'6017	2'4984	2'4255	2'3890	2'3315	2'2718	2'2097	2'1446	2'0760	2'0032
3'2576	3'1209	2'9780	2'8776	2'8274	2'7488	2'6675	2'5831	2'4951	2'4029	2'3055
3'7030	3'5350	3'3600	3'2376	3'1764	3'0807	2'9821	2'8799	2'7736	2'6625	2'5455
1'8903	1'8450	1'7964	1'7614	1'7439	1'7159	1'6864	1'6554	1'6224	1'5871	1'5490
2'2747	2'2036	2'1282	2'0745	2'0476	2'0050	1'9605	1'9139	1'8649	1'8128	1'7570
2'6682	2'5699	2'4665	2'3933	2'3567	2'2989	2'2389	2'1763	2'1107	2'0415	1'9677
3'2106	3'0740	2'9311	2'8307	2'7805	2'7017	2'6202	2'5355	2'4471	2'3542	2'2559
3'6420	3'4745	3'2999	3'1776	3'1165	3'0208	2'9221	2'8198	2'7132	2'6016	2'4837
1'8775	1'8319	1'7831	1'7478	1'7302	1'7019	1'6721	1'6407	1'6073	1'5715	1'5327
2'2547	2'1834	2'1077	2'0537	2'0267	1'9838	1'9390	1'8920	1'8424	1'7897	1'7331
2'6396	2'5412	2'4374	2'3640	2'3273	2'2693	2'2090	2'1460	2'0799	2'0099	1'9353
3'1681	3'0316	2'8887	2'7882	2'7380	2'6591	2'5773	2'4923	2'4035	2'3099	2'2107
3'5870	3'4199	3'2456	3'1235	3'0624	2'9667	2'8679	2'7654	2'6585	2'5463	2'4276
1'8658	1'8200	1'7708	1'7353	1'7175	1'6890	1'6589	1'6272	1'5934	1'5570	1'5176
2'2365	2'1649	2'0889	2'0346	2'0075	1'9643	1'9192	1'8718	1'8217	1'7684	1'7110
2'6135	2'5149	2'4110	2'3373	2'3005	2'2422	2'1816	2'1183	2'0517	1'9811	1'9055
3'1294	2'9931	2'8502	2'7496	2'6993	2'6203	2'5383	2'4530	2'3637	2'2695	2'1694
3'5370	3'3704	3'1963	3'0743	3'0133	2'9176	2'8187	2'7160	2'6088	2'4960	2'3765

TABLA 6: Distribución  $F$  de Snedecor (continuación)

	$n_1$	1	2	3	4	5	6	7	8	9
$n_2$	$p$									
26	0'100	2'9091	2'5191	2'3075	2'1745	2'0822	2'0139	1'9610	1'9188	1'8841
	0'050	4'2252	3'3690	2'9751	2'7426	2'5868	2'4741	2'3883	2'3205	2'2655
	0'025	5'6586	4'2655	3'6697	3'3289	3'1048	2'9447	2'8240	2'7293	2'6528
	0'010	7'7213	5'5263	4'6366	4'1400	3'8183	3'5911	3'4210	3'2884	3'1818
	0'005	9'4059	6'5409	5'4091	4'7852	4'3844	4'1027	3'8928	3'7297	3'5989
27	0'100	2'9012	2'5106	2'2987	2'1655	2'0730	2'0045	1'9515	1'9091	1'8743
	0'050	4'2100	3'3541	2'9604	2'7278	2'5719	2'4591	2'3732	2'3053	2'2501
	0'025	5'6331	4'2421	3'6472	3'3067	3'0828	2'9228	2'8021	2'7074	2'6309
	0'010	7'6767	5'4881	4'6009	4'1056	3'7848	3'5580	3'3882	3'2558	3'1494
	0'005	9'3423	6'4885	5'3611	4'7396	4'3402	4'0594	3'8501	3'6875	3'5571
28	0'100	2'8939	2'5028	2'2906	2'1571	2'0645	1'9959	1'9427	1'9001	1'8652
	0'050	4'1960	3'3404	2'9467	2'7141	2'5581	2'4453	2'3593	2'2913	2'2360
	0'025	5'6096	4'2205	3'6264	3'2863	3'0625	2'9027	2'7820	2'6872	2'6106
	0'010	7'6356	5'4529	4'5681	4'0740	3'7539	3'5276	3'3581	3'2259	3'1195
	0'005	9'2838	6'4403	5'3170	4'6977	4'2996	4'0197	3'8110	3'6487	3'5186
29	0'100	2'8871	2'4955	2'2831	2'1494	2'0566	1'9878	1'9345	1'8918	1'8568
	0'050	4'1830	3'3277	2'9340	2'7014	2'5454	2'4324	2'3463	2'2782	2'2229
	0'025	5'5878	4'2006	3'6072	3'2674	3'0438	2'8840	2'7633	2'6686	2'5919
	0'010	7'5976	5'4205	4'5378	4'0449	3'7254	3'4995	3'3302	3'1982	3'0920
	0'005	9'2297	6'3958	5'2764	4'6591	4'2622	3'9830	3'7749	3'6130	3'4832
30	0'100	2'8807	2'4887	2'2761	2'1422	2'0492	1'9803	1'9269	1'8841	1'8490
	0'050	4'1709	3'3158	2'9223	2'6896	2'5336	2'4205	2'3343	2'2662	2'2107
	0'025	5'5675	4'1821	3'5894	3'2499	3'0265	2'8667	2'7460	2'6513	2'5746
	0'010	7'5625	5'3903	4'5097	4'0179	3'6990	3'4735	3'3045	3'1726	3'0665
	0'005	9'1797	6'3547	5'2388	4'6233	4'2276	3'9492	3'7416	3'5801	3'4505
40	0'100	2'8354	2'4404	2'2261	2'0909	1'9968	1'9269	1'8725	1'8289	1'7929
	0'050	4'0848	3'2317	2'8387	2'6060	2'4495	2'3359	2'2490	2'1802	2'1240
	0'025	5'4239	4'0510	3'4633	3'1261	2'9037	2'7444	2'6238	2'5289	2'4519
	0'010	7'3141	5'1785	4'3126	3'8283	3'5138	3'2910	3'1238	2'9930	2'8876
	0'005	8'8278	6'0664	4'9759	4'3738	3'9860	3'7129	3'5088	3'3498	3'2220
60	0'100	2'7914	2'3933	2'1774	2'0410	1'9457	1'8747	1'8194	1'7748	1'7380
	0'050	4'0012	3'1504	2'7581	2'5252	2'3683	2'2540	2'1665	2'0970	2'0401
	0'025	5'2857	3'9253	3'3425	3'0077	2'7863	2'6274	2'5068	2'4117	2'3344
	0'010	7'0771	4'9774	4'1259	3'6491	3'3389	3'1187	2'9530	2'8233	2'7185
	0'005	8'4946	5'7950	4'7290	4'1399	3'7600	3'4918	3'2911	3'1344	3'0083
120	0'100	2'7478	2'3473	2'1300	1'9923	1'8959	1'8238	1'7675	1'7220	1'6843
	0'050	3'9201	3'0718	2'6802	2'4472	2'2900	2'1750	2'0867	2'0164	1'9588
	0'025	5'1524	3'8046	3'2270	2'8943	2'6740	2'5154	2'3948	2'2994	2'2217
	0'010	6'8510	4'7865	3'9491	3'4796	3'1735	2'9559	2'7918	2'6629	2'5586
	0'005	8'1790	5'5393	4'4973	3'9207	3'5482	3'2849	3'0874	2'9330	2'8083
$\infty$	0'100	2'7055	2'3026	2'0838	1'9449	1'8473	1'7741	1'7167	1'6702	1'6315
	0'050	3'8415	2'9957	2'6049	2'3719	2'2141	2'0986	2'0096	1'9384	1'8799
	0'025	5'0239	3'6889	3'1161	2'7858	2'5665	2'4082	2'2875	2'1918	2'1136
	0'010	6'6349	4'6052	3'7816	3'3192	3'0173	2'8020	2'6393	2'5113	2'4073
	0'005	7'8794	5'2983	4'2794	3'7151	3'3499	3'0913	2'8968	2'7444	2'6210

TABLA 6: Distribución  $F$  de Snedecor (continuación)

10	12	15	18	20	24	30	40	60	120	$\infty$
1'8550	1'8090	1'7596	1'7238	1'7059	1'6771	1'6468	1'6147	1'5805	1'5437	1'5036
2'2197	2'1479	2'0716	2'0171	1'9898	1'9464	1'9010	1'8533	1'8027	1'7488	1'6906
2'5895	2'4909	2'3867	2'3128	2'2759	2'2174	2'1565	2'0928	2'0257	1'9545	1'8781
3'0941	2'9579	2'8150	2'7143	2'6640	2'5848	2'5026	2'4170	2'3273	2'2325	2'1315
3'4916	3'3252	3'1515	3'0295	2'9685	2'8728	2'7738	2'6709	2'5633	2'4501	2'3297
1'8451	1'7989	1'7492	1'7131	1'6951	1'6662	1'6356	1'6032	1'5686	1'5313	1'4906
2'2043	2'1323	2'0558	2'0010	1'9736	1'9299	1'8842	1'8361	1'7851	1'7307	1'6717
2'5676	2'4688	2'3644	2'2903	2'2533	2'1946	2'1334	2'0693	2'0018	1'9299	1'8527
3'0618	2'9256	2'7827	2'6820	2'6316	2'5522	2'4699	2'3840	2'2938	2'1984	2'0965
3'4499	3'2839	3'1104	2'9885	2'9275	2'8318	2'7327	2'6296	2'5217	2'4078	2'2867
1'8359	1'7895	1'7395	1'7033	1'6852	1'6560	1'6252	1'5925	1'5575	1'5198	1'4784
2'1900	2'1179	2'0411	1'9861	1'9586	1'9147	1'8687	1'8203	1'7689	1'7138	1'6541
2'5473	2'4484	2'3438	2'2695	2'2324	2'1735	2'1121	2'0477	1'9796	1'9072	1'8291
3'0320	2'8959	2'7530	2'6521	2'6017	2'5223	2'4397	2'3535	2'2629	2'1670	2'0642
3'4117	3'2460	3'0727	2'9508	2'8899	2'7941	2'6949	2'5916	2'4834	2'3689	2'2469
1'8274	1'7808	1'7306	1'6941	1'6759	1'6465	1'6155	1'5825	1'5472	1'5090	1'4670
2'1768	2'1045	2'0275	1'9722	1'9446	1'9005	1'8543	1'8055	1'7537	1'6981	1'6377
2'5286	2'4295	2'3248	2'2503	2'2131	2'1540	2'0923	2'0276	1'9591	1'8861	1'8072
3'0045	2'8685	2'7256	2'6247	2'5742	2'4946	2'4118	2'3253	2'2344	2'1378	2'0342
3'3765	3'2111	3'0379	2'9160	2'8551	2'7594	2'6601	2'5565	2'4479	2'3330	2'2102
1'8195	1'7727	1'7223	1'6856	1'6673	1'6377	1'6065	1'5732	1'5376	1'4989	1'4564
2'1646	2'0921	2'0148	1'9594	1'9317	1'8874	1'8409	1'7918	1'7396	1'6835	1'6223
2'5112	2'4120	2'3072	2'2325	2'1952	2'1359	2'0739	2'0089	1'9400	1'8664	1'7867
2'9791	2'8431	2'7002	2'5992	2'5487	2'4689	2'3860	2'2992	2'2079	2'1107	2'0062
3'3440	3'1787	3'0057	2'8839	2'8230	2'7272	2'6278	2'5241	2'4151	2'2997	2'1760
1'7627	1'7146	1'6624	1'6243	1'6052	1'5741	1'5411	1'5056	1'4672	1'4248	1'3769
2'0772	2'0035	1'9245	1'8674	1'8389	1'7929	1'7444	1'6928	1'6373	1'5766	1'5089
2'3882	2'2882	2'1819	2'1058	2'0677	2'0069	1'9429	1'8752	1'8028	1'7242	1'6371
2'8005	2'6648	2'5216	2'4198	2'3689	2'2880	2'2034	2'1142	2'0194	1'9172	1'8047
3'1167	2'9531	2'7811	2'6593	2'5984	2'5020	2'4015	2'2958	2'1838	2'0635	1'9318
1'7070	1'6574	1'6034	1'5635	1'5435	1'5107	1'4755	1'4373	1'3952	1'3476	1'2915
1'9926	1'9174	1'8364	1'7775	1'7480	1'7001	1'6491	1'5943	1'5343	1'4673	1'3893
2'2702	2'1692	2'0613	1'9834	1'9445	1'8817	1'8152	1'7440	1'6668	1'5810	1'4822
2'6318	2'4961	2'3523	2'2493	2'1978	2'1154	2'0285	1'9360	1'8363	1'7263	1'6006
2'9042	2'7419	2'5705	2'4483	2'3872	2'2898	2'1874	2'0789	1'9622	1'8341	1'6885
1'6524	1'6012	1'5450	1'5031	1'4821	1'4472	1'4094	1'3676	1'3203	1'2646	1'1926
1'9105	1'8337	1'7505	1'6893	1'6587	1'6084	1'5543	1'4952	1'4290	1'3519	1'2539
2'1570	2'0548	1'9450	1'8649	1'8249	1'7597	1'6899	1'6141	1'5299	1'4327	1'3104
2'4721	2'3363	2'1915	2'0869	2'0346	1'9500	1'8600	1'7628	1'6557	1'5330	1'3805
2'7052	2'5439	2'3727	2'2496	2'1881	2'0890	1'9839	1'8709	1'7469	1'6055	1'4311
1'5987	1'5458	1'4871	1'4428	1'4206	1'3832	1'3419	1'2951	1'2400	1'1686	1'000
1'8307	1'7522	1'6664	1'6025	1'5705	1'5173	1'4591	1'3940	1'3180	1'2214	1'000
2'0483	1'9447	1'8326	1'7499	1'7085	1'6402	1'5660	1'4835	1'3883	1'2684	1'000
2'3209	2'1848	2'0385	1'9317	1'8783	1'7908	1'6964	1'5923	1'4730	1'3246	1'000
2'5188	2'3583	2'1868	2'0621	1'9998	1'8983	1'7891	1'6691	1'5325	1'3637	1'000

TABLA 7: Distrib. del estadístico de Kolmogorov-Smirnov  $\Delta_n$ 

$$P\{\Delta_n > x\} = p$$

$p$	0'2	0'1	0'05	0'02	0'01
$n$					
2	0'684	0'776	0'842	0'900	0'929
3	0'565	0'636	0'708	0'785	0'829
4	0'493	0'565	0'624	0'689	0'734
5	0'447	0'509	0'563	0'627	0'669
6	0'410	0'468	0'519	0'577	0'617
7	0'381	0'436	0'483	0'538	0'576
8	0'358	0'410	0'454	0'507	0'542
9	0'339	0'387	0'430	0'480	0'513
10	0'323	0'369	0'409	0'457	0'489
11	0'308	0'352	0'391	0'437	0'468
12	0'296	0'338	0'375	0'419	0'449
13	0'285	0'325	0'361	0'404	0'432
14	0'275	0'314	0'349	0'390	0'418
15	0'266	0'304	0'338	0'377	0'404
16	0'258	0'295	0'327	0'366	0'392
17	0'250	0'286	0'318	0'355	0'381
18	0'244	0'279	0'309	0'346	0'371
19	0'237	0'271	0'301	0'337	0'361
20	0'232	0'265	0'294	0'329	0'352
21	0'226	0'259	0'287	0'321	0'344
22	0'221	0'253	0'281	0'314	0'337
23	0'216	0'247	0'275	0'307	0'330
24	0'212	0'242	0'269	0'301	0'323
25	0'208	0'238	0'264	0'295	0'317
26	0'204	0'233	0'259	0'290	0'311
27	0'200	0'229	0'254	0'284	0'305
28	0'197	0'225	0'250	0'279	0'300
29	0'193	0'221	0'246	0'275	0'295
30	0'190	0'218	0'242	0'270	0'290
31	0'187	0'214	0'238	0'266	0'285
32	0'184	0'211	0'234	0'262	0'281
33	0'182	0'208	0'231	0'258	0'277
34	0'179	0'205	0'227	0'254	0'273
35	0'177	0'202	0'224	0'251	0'269
36	0'174	0'199	0'221	0'247	0'265
37	0'172	0'196	0'218	0'244	0'262
38	0'170	0'194	0'215	0'241	0'258
39	0'168	0'191	0'213	0'238	0'255
40	0'165	0'189	0'210	0'235	0'252
$n > 40$	$1'07/\sqrt{n}$	$1'22/\sqrt{n}$	$1'36/\sqrt{n}$	$1'52/\sqrt{n}$	$1'63/\sqrt{n}$

**TABLA 8: Tabla de Lilliefors. Distribución de  $\Delta_n$  para muestras normales de par metros desconocidos**

$$P\{\Delta_n > x\} = p$$

$p$	0'2	0'1	0'05	0'01
<i>n</i>				
4	0'300	0'352	0'381	0'417
5	0'285	0'315	0'337	0'405
6	0'265	0'294	0'319	0'364
7	0'247	0'276	0'300	0'348
8	0'233	0'261	0'285	0'331
9	0'223	0'249	0'271	0'311
10	0'215	0'239	0'258	0'294
11	0'206	0'230	0'249	0'284
12	0'199	0'223	0'242	0'275
13	0'190	0'214	0'234	0'268
14	0'183	0'207	0'227	0'261
15	0'177	0'201	0'220	0'257
16	0'173	0'195	0'213	0'250
17	0'169	0'189	0'206	0'245
18	0'166	0'184	0'200	0'239
19	0'163	0'179	0'195	0'235
20	0'160	0'174	0'190	0'231
25	0'149	0'165	0'180	0'203
30	0'131	0'144	0'161	0'187
$n > 30$	$0'736/\sqrt{n}$	$0'805/\sqrt{n}$	$0'886/\sqrt{n}$	$1'031/\sqrt{n}$

TABLA 9: Distribución del estadístico de Kolmogorov-Smirnov  $\Delta_{n,m}$ 

$$P\{\Delta_{n,m} > x\} = p$$

	$p$	$0'2$	$0'1$	$0'05$	$0'02$	$0'01$
$n = 1$	$m = 9$	$17/18$				
	$m = 10$	$9/10$				
$n = 2$	$m = 3$	$5/6$				
	$m = 4$	$3/4$				
	$m = 5$	$4/5$	$4/5$			
	$m = 6$	$5/6$	$5/6$			
	$m = 7$	$5/7$	$6/7$			
	$m = 8$	$3/4$	$7/8$	$7/8$		
	$m = 9$	$7/9$	$8/9$	$8/9$		
	$m = 10$	$7/10$	$4/5$	$9/10$		
$n = 3$	$m = 4$	$3/4$	$3/4$			
	$m = 5$	$2/3$	$4/5$	$4/5$		
	$m = 6$	$2/3$	$2/3$	$5/6$		
	$m = 7$	$2/3$	$5/7$	$6/7$	$6/7$	
	$m = 8$	$5/8$	$3/4$	$3/4$	$7/8$	
	$m = 9$	$2/3$	$2/3$	$7/9$	$8/9$	$8/9$
	$m = 10$	$3/5$	$7/10$	$4/5$	$9/10$	$9/10$
	$m = 12$	$7/12$	$2/3$	$3/4$	$5/6$	$11/12$
$n = 4$	$m = 5$	$3/5$	$3/4$	$4/5$	$4/5$	
	$m = 6$	$7/12$	$2/3$	$3/4$	$5/6$	$5/6$
	$m = 7$	$17/28$	$5/7$	$3/4$	$6/7$	$6/7$
	$m = 8$	$5/8$	$5/8$	$3/4$	$7/8$	$7/8$
	$m = 9$	$5/9$	$2/3$	$3/4$	$7/9$	$8/9$
	$m = 10$	$11/20$	$13/20$	$7/10$	$4/5$	$4/5$
	$m = 12$	$7/12$	$2/3$	$2/3$	$3/4$	$5/6$
	$m = 16$	$9/16$	$5/8$	$11/16$	$3/4$	$13/16$
$n = 5$	$m = 6$	$3/5$	$2/3$	$2/3$	$5/6$	$5/6$
	$m = 7$	$4/7$	$23/35$	$5/7$	$29/35$	$6/7$
	$m = 8$	$11/20$	$5/8$	$27/40$	$4/5$	$4/5$
	$m = 9$	$5/9$	$3/5$	$31/45$	$7/9$	$4/5$
	$m = 10$	$1/2$	$3/5$	$7/10$	$7/10$	$4/5$
	$m = 15$	$8/15$	$3/5$	$2/3$	$11/15$	$11/15$
	$m = 20$	$1/2$	$11/20$	$3/5$	$7/10$	$3/4$

**TABLA 9: Distribución del estadístico de Kolmogorov-Smirnov  $\Delta_{n,m}$**   
 (continuación)

	$p$	0'2	0'1	0'05	0'02	0'01
$n = 6$	$m = 7$	23/42	4/7	29/42	5/7	5/6
	$m = 8$	1/2	7/12	2/3	3/4	3/4
	$m = 9$	1/2	5/9	2/3	13/18	7/9
	$m = 10$	1/2	17/30	19/30	7/10	11/15
	$m = 12$	1/2	7/12	7/12	2/3	3/4
	$m = 18$	4/9	5/9	11/18	2/3	13/18
	$m = 24$	11/24	1/2	7/12	5/8	2/3
$n = 7$	$m = 8$	27/56	33/56	5/8	41/56	3/4
	$m = 9$	31/63	5/9	40/63	5/7	47/63
	$m = 10$	33/70	39/70	43/70	7/10	5/7
	$m = 14$	3/7	1/2	4/7	9/14	5/7
	$m = 28$	3/7	13/28	15/28	17/28	9/14
$n = 8$	$m = 9$	4/9	13/24	5/8	2/3	3/4
	$m = 10$	19/40	21/40	23/40	27/40	7/10
	$m = 12$	11/24	1/2	7/12	5/8	2/3
	$m = 16$	7/16	1/2	9/16	5/8	5/8
	$m = 32$	13/32	7/16	1/2	9/16	19/32
$n = 9$	$m = 10$	7/15	1/2	26/45	2/3	31/45
	$m = 12$	4/9	1/2	5/9	11/18	2/3
	$m = 15$	19/45	22/45	8/15	3/5	29/45
	$m = 18$	7/18	4/9	1/2	5/9	11/18
	$m = 36$	13/36	5/12	17/36	19/36	5/9
$n = 10$	$m = 15$	2/5	7/15	1/2	17/30	19/30
	$m = 20$	2/5	9/20	1/2	11/20	3/5
	$m = 40$	7/20	2/5	9/20	1/2	
$n = 12$	$m = 15$	23/60	9/20	1/2	11/20	7/12
	$m = 16$	3/8	7/16	23/48	13/24	7/12
	$m = 18$	13/36	5/12	17/36	19/36	5/9
	$m = 20$	11/30	5/12	7/15	31/60	17/30
$n = 15$	$m = 20$	7/20	2/5	13/30	29/60	31/60
$n = 16$	$m = 20$	27/80	31/80	17/40	19/40	41/80
m y n grandes		$1'07\sqrt{\frac{n+m}{nm}}$	$1'22\sqrt{\frac{n+m}{nm}}$	$1'36\sqrt{\frac{n+m}{nm}}$	$1'52\sqrt{\frac{n+m}{nm}}$	$1'63\sqrt{\frac{n+m}{nm}}$

TABLA 10: Distrib. del estadístico de Kolmogorov-Smirnov  $\Delta_{n,n}$ 

$$P\{\Delta_{n,n} > x\} = p$$

$p$	0'2	0'1	0'05	0'02	0'01
$n$					
3	2/3	2/3			
4	3/4	3/4	3/4		
5	3/5	3/5	4/5	4/5	4/5
6	3/6	4/6	4/6	5/6	5/6
7	4/7	4/7	5/7	5/7	5/7
8	4/8	4/8	5/8	5/8	6/8
9	4/9	5/9	5/9	6/9	6/9
10	4/10	5/10	6/10	6/10	7/10
11	5/11	5/11	6/11	7/11	7/11
12	5/12	5/12	6/12	7/12	7/12
13	5/13	6/13	6/13	7/13	8/13
14	5/14	6/14	7/14	7/14	8/14
15	5/15	6/15	7/15	8/15	8/15
16	6/16	6/16	7/16	8/16	9/16
17	6/17	7/17	7/17	8/17	9/17
18	6/18	7/18	8/18	9/18	9/18
19	6/19	7/19	8/19	9/19	9/19
20	6/20	7/20	8/20	9/20	10/20
21	6/21	7/21	8/21	9/21	10/21
22	7/22	8/22	8/22	10/22	10/22
23	7/23	8/23	9/23	10/23	10/23
24	7/24	8/24	9/24	10/24	11/24
25	7/25	8/25	9/25	10/25	11/25
26	7/26	8/26	9/26	10/26	11/26
27	7/27	8/27	9/27	11/27	11/27
28	8/28	9/28	10/28	11/28	12/28
29	8/29	9/29	10/29	11/29	12/29
30	8/30	9/30	10/30	11/30	12/30
31	8/31	9/31	10/31	11/31	12/31
32	8/32	9/32	10/32	12/32	12/32
34	8/34	10/34	11/34	12/34	13/34
36	9/36	10/36	11/36	12/36	13/36
38	9/38	10/38	11/38	13/38	14/38
40	9/40	10/40	12/40	13/40	14/40
$n > 40$					
$1'52/\sqrt{n}$					
$1'73/\sqrt{n}$					
$1'92/\sqrt{n}$					
$2'15/\sqrt{n}$					
$2'30/\sqrt{n}$					

TABLA 11: Distribución del estadístico de Wilcoxon

$$P\{T^+ > x\} = p$$

<i>p</i>	0'1	0'05	0'025	0'01
<i>n</i>				
3	6	6	6	6
4	9	10	10	10
5	12	14	15	15
6	17	18	20	21
7	22	24	25	27
8	27	30	32	34
9	34	36	39	41
10	40	44	46	49
11	48	52	55	58
12	56	60	64	67
13	64	69	73	78
14	73	79	84	89
15	83	89	94	100
16	93	100	106	112
17	104	111	118	125
18	115	123	130	138
19	127	136	143	152
20	140	149	157	166

TABLA 13: Distribución del estadístico  $\tau$  de Kendall

$$P\{|T| > x\} = p$$

<i>p</i>	0'2	0'1	0'05	0'02
<i>n</i>				
3	1'0000	1'0000	1'0000	1'0000
4	0'6667	0'6667	1'0000	1'0000
5	0'6000	0'6000	0'8000	0'8000
6	0'4667	0'6000	0'7333	0'7333
7	0'4286	0'5238	0'6190	0'7143
8	0'0128	0'5000	0'5714	0'6429
9	0'3333	0'4444	0'5000	0'6111
10	0'3333	0'4222	0'4667	0'5556

TABLA 12: Distribución del estadístico de Mann-Whitney

$$P\{V > x\} = p$$

	<i>m</i>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>
<i>n</i>	<i>p</i>									
<b>2</b>	0'100	4	5	7	8	10	12	13	15	16
	0'050	4	6	8	9	11	13	14	16	18
	0'025	4	6	8	10	12	14	15	17	19
	0'010	4	6	8	10	12	14	16	18	20
<b>3</b>	0'100		7	10	12	14	16	18	21	23
	0'050		8	11	13	15	18	20	22	25
	0'025		9	12	14	16	19	21	24	26
	0'010		9	12	15	18	20	20	25	28
<b>4</b>	0'100			12	15	18	21	24	26	29
	0'050			14	17	20	23	26	29	32
	0'025			15	18	21	24	27	31	34
	0'010			16	19	22	26	29	32	36
<b>5</b>	0'100				19	22	26	29	32	36
	0'050				20	24	28	31	35	38
	0'025				22	26	29	33	37	41
	0'010				23	27	31	35	39	43
<b>6</b>	0'100					26	30	34	38	42
	0'050					28	33	37	41	45
	0'025					30	35	39	43	48
	0'010					32	37	41	46	51
<b>7</b>	0'100						35	39	44	48
	0'050						37	42	47	52
	0'025						40	45	50	55
	0'010						42	48	53	58
<b>8</b>	0'100							44	49	55
	0'050							48	53	59
	0'025							50	56	62
	0'010							54	60	66
<b>9</b>	0'100								55	61
	0'050								59	65
	0'025								63	69
	0'010								66	73
<b>10</b>	0'100									67
	0'050									72
	0'025									76
	0'010									80

TABLA 14: Distribución del estadístico de Spearman

$$P\{R_S > x\} = p$$

<i>p</i>	0'1	0'05	0'025	0'01	0'005	0'001
<i>n</i>						
4	0'8000	0'8000				
5	0'7000	0'8000	0'9000	0'9000		
6	0'6000	0'7714	0'8286	0'8857	0'9429	
7	0'5357	0'6786	0'7450	0'8571	0'8929	0'9643
8	0'5000	0'6190	0'7143	0'8095	0'8571	0'9286
9	0'4667	0'5833	0'6833	0'7667	0'8167	0'9000
10	0'4424	0'5515	0'6364	0'7333	0'7818	0'8667
11	0'4182	0'5273	0'6091	0'7000	0'7545	0'8364
12	0'3986	0'4965	0'5804	0'6713	0'7273	0'8182
13	0'3791	0'4780	0'5549	0'6429	0'6978	0'7912
14	0'3626	0'4593	0'5341	0'6220	0'6747	0'7670
15	0'3500	0'4429	0'5179	0'6000	0'6536	0'7464
16	0'3382	0'4265	0'5000	0'5824	0'6324	0'7265
17	0'3260	0'4118	0'4853	0'5637	0'6152	0'7083
18	0'3148	0'3994	0'4716	0'5480	0'5975	0'6904
19	0'3070	0'3895	0'4579	0'5333	0'5825	0'6737
20	0'2977	0'3789	0'4451	0'5203	0'5684	0'6586
21	0'2909	0'3688	0'4351	0'5078	0'5545	0'6455
22	0'2829	0'3597	0'4241	0'4963	0'5426	0'6318
23	0'2767	0'3518	0'4150	0'4852	0'5306	0'6186
24	0'2704	0'3435	0'4061	0'4748	0'5200	0'6070
25	0'2646	0'3362	0'3977	0'4654	0'5100	0'5962
26	0'2588	0'3299	0'3894	0'4564	0'5002	0'5856
27	0'2540	0'3236	0'3822	0'4481	0'4915	0'5757
28	0'2490	0'3175	0'3749	0'4401	0'4828	0'5660
29	0'2443	0'3113	0'3685	0'4320	0'4744	0'5567
30	0'2400	0'3059	0'3620	0'4251	0'4665	0'5479



# Índice alfabético

- Aproximación de Welch, 73
- Cantidad de información de Fisher, 225
- Coeficiente
  - de asociación de Kendall, 478
  - de correlación, 79
    - entre rangos de Spearman, 481
    - muestral, 79
- Contraste de hipótesis, 330
- Cota
  - de confianza, 114
  - de Fréchet-Cramer-Rao, 225
- Cuantiles muestrales, 35
- Cuasivarianza muestral, 41
- Cuasicovarianza muestral, 79
- Distribución
  - a posteriori, 181
  - a priori, 180
  - de la diferencia de medias, 71
  - de la media muestral, 41, 60, 66, 69
  - de la muestra, 12
  - de la población, 4
  - de la varianza muestral, 66
  - del cociente de varianzas, 75
  - en el muestreo, 17
  - $F$  de Snedecor, 76
  - $\chi^2$  de Pearson, 61
  - muestral, 12
  - teórica, 4
  - $t$  de Student, 68
- Distribución asintótica
  - de la varianza muestral, 48
  - de los cuantiles muestrales, 43
  - de los momentos muestrales, 46, 48
- Eficiencia, 228
  - asintótica, 299
  - de un estimador, 232
- Error
  - cuadrático medio, 147
  - de tipo I, 331
  - de tipo II, 331
- Espacio
  - muestral, 15
  - paramétrico, 4
- Estadístico, 15
  - completo, 216
  - minimal suficiente, 171
  - suficiente, 163
- Estimador, 140
  - admissible, 150
  - asintóticamente normal, 299
  - Bayes, 180
  - centrado, 151
    - uniformemente de mínima varianza, 210
  - consistente, 156
  - de Pitman
    - de un parámetro de escala, 247, 251
    - de un parámetro de posición, 241, 254
  - eficiente, 228
  - inadmisible, 149
  - insesgado, 151
  - invariante, 158
  - máxima verosimilitud 283
  - método de los momentos 279
  - minimax, 176
  - mínimo cuadrático 307
- Familia de tipo exponencial, 174
- Función
  - crítica, 330
  - de potencia, 333
  - de pérdida, 142
  - de precisión, 403
  - de verosimilitud, 283

- indicatriz, 34
- Hipótesis
  - alternativa, 327
  - bilateral, 356
  - compuesta, 328
  - nula, 327
  - simple, 328
  - unilateral, 347
- Intervalo de confianza, 94
- Lema de Neyman-Pearson, 338, 343
- Media muestral, 16, 35
- Mediana muestral, 35
- Momentos muestrales, 35
- Muestra aleatoria, 7
  - simple, 8
- Muestreo, 8
  - bietápico, 8
  - estratificado, 8
  - sin reemplazamiento, 8
  - sistemático, 8
- Nivel
  - crítico, 336
  - de confianza, 94
  - de significación, 332
- Parámetro, 4
  - de escala, 160
  - de posición, 159
- Procedimiento de muestreo, 8
- Punto crítico, 337
- Razón de verosimilitud monótona, 348
- Razón de verosimilitudes, 388
- Región
  - crítica, 330
  - de aceptación, 330
  - de confianza, 96
  - uniformemente más precisa, 403
- Riesgo
  - Bayes, 180
  - de un estimador, 144
  - minimax, 176
- Sesgo de un estimador, 151
- Tabla de contingencia, 445, 450
- Tamaño
  - de la muestra, 7
  - de un test, 332
- Teorema
  - central de la estadística, 33
  - de factorización, 167
  - de Fisher, 66
  - de Glivenko-Cantelli, 37
  - de Karlin-Rubin, 349
  - de Lehman, 357
  - de Lehmann-Scheffé, 218
  - de Rao-Blackwell, 211
  - de Student, 69
- Test
  - de hipótesis, 330
  - de la mediana, 473
  - de los signos, 468
  - de Mann-Whitney, 476
  - de rachas, 483
  - de Spearman, 481
  - de Wilcoxon de los rangos signados, 470
  - insesgado, 357
  - $\tau$  de Kendall, 478
  - uniformemente de máxima potencia, 335
  - uniformemente más potente, 335
- Varianza muestral, 16, 35
- Varianza generalizada, 235