

TEMAS DE MATEMÁTICAS

Una  
introducción  
a la  
estadística  
inferencial

*Luis Rincón*





**Luis Rincón**

**UNA INTRODUCCIÓN  
A LA ESTADÍSTICA INFERENCIAL**

**Facultad de Ciencias, UNAM  
2019**



519.54

Rincón, Luis, autor.

Una introducción a la estadística inferencial / Luis Rincón.  
-- Ciudad de México : Universidad Nacional Autónoma de México,  
Facultad de Ciencias, 2019.

vi, 406 páginas : ilustraciones ; 22 cm. -- (Temas de matemáticas)

Incluye índice.

Bibliografía: páginas 397-400.

ISBN: 978-607-30-2432-7 Fascículo

1. Estadística matemática -- Estudio y enseñanza (Superior). 2. Estimación de parámetros. 3. Prueba de hipótesis estadística. I. Universidad Nacional Autónoma de México. Facultad de Ciencias, editor.

Biblioteca Nacional de México

No. de sistema[000715099] scdd 22

## **Una introducción a la estadística inferencial**

1a. edición, 20 de septiembre de 2019

© DR. 2019. Universidad Nacional Autónoma de México  
Facultad de Ciencias

Ciudad Universitaria, Delegación Coyoacán.

C.P. 04510. Ciudad de México

Coordinación de servicios editoriales: editoriales@ciencias.unam.mx

Plaza Prometeo: tienda.fcencias.unam.mx

**ISBN: 978-607-30-2432-7**

Diseño de portada: Laura Uribe

Prohibida la reproducción total o parcial de la obra, por cualquier medio, sin la autorización por escrito del titular de los derechos.

Impreso y hecho en México.

## Prólogo

La estadística es un área muy amplia y muy diversa de las matemáticas. Sus aplicaciones han abarcado prácticamente todas las disciplinas del quehacer humano. En este trabajo se proporciona una introducción a tres grandes temas clásicos de la estadística inferencial relativos al problema de la estimación de parámetros: la estimación puntual, la estimación por intervalos y las pruebas de hipótesis. En todos los casos el énfasis principal ha sido puesto en la estimación de parámetros de las distribuciones de probabilidad; sin embargo los métodos y las ideas aquí expuestas también son aplicables para tratar otros problemas similares. El enfoque con el que se tratan los temas es principalmente matemático, buscando proveer las demostraciones de casi todos los resultados que se estudian.

Este trabajo contiene material de estudio para un primer curso semestral sobre estadística matemática a nivel universitario. Está dirigido a estudiantes de las carreras de Actuaría, Matemáticas, Matemáticas Aplicadas y otras carreras científicas similares cuyos programas de estudio contemplan cursos formales sobre esta disciplina. Se presupone conocido el material de algún curso semestral de Probabilidad.

Se ha procurado la inclusión de la mayor cantidad de ejercicios con la finalidad de que se resuelvan en las exposiciones de un curso, en tareas o en otros tipos de evaluaciones. La mayoría de estos ejercicios son rutinarios, y tienen la finalidad de practicar o reforzar los métodos y procedimientos de la teoría expuesta. Al final del texto aparecen las soluciones de algunos de ellos.

Luis Rincón  
Noviembre 2019  
Ciudad Universitaria UNAM



# Contenido

<b>Introducción</b>	<b>1</b>
<b>1. Análisis exploratorio de datos</b>	<b>3</b>
1.1. Conceptos elementales . . . . .	3
1.2. Descripciones numéricas . . . . .	18
1.3. Descripciones gráficas . . . . .	60
1.4. Variables aleatorias . . . . .	81
<b>2. Estimación puntual</b>	<b>85</b>
2.1. Introducción . . . . .	85
2.2. Método de momentos . . . . .	94
2.3. Método de máxima verosimilitud . . . . .	108
2.4. Insesgamiento . . . . .	127
2.5. Insesgamiento asintótico . . . . .	138
2.6. Consistencia . . . . .	142
2.7. Sesgo y error cuadrático medio . . . . .	150
2.8. Cota inferior de Cramér-Rao . . . . .	154
2.9. Eficiencia . . . . .	165
2.10. Suficiencia . . . . .	169
2.11. Suficiencia e información . . . . .	181
2.12. Suficiencia conjunta . . . . .	193
2.13. Suficiencia minimal . . . . .	198
2.14. Métodos para probar suficiencia . . . . .	210
2.15. Esperanza condicional . . . . .	211
2.16. Teorema de Rao-Blackwell . . . . .	216
2.17. Completez . . . . .	224

2.18. Teorema de Lehmann-Scheffé . . . . .	229
2.19. Distribuciones tipo exponencial . . . . .	238
<b>3. Estimación por intervalos</b>	<b>245</b>
3.1. Definiciones . . . . .	245
3.2. Distribución Bernoulli . . . . .	248
3.3. Distribución uniforme continua . . . . .	251
3.4. Distribución exponencial . . . . .	255
3.5. Distribución normal . . . . .	256
3.6. Intervalo para la media de una distribución cualquiera . . . . .	263
3.7. Intervalos conjuntos para dos parámetros . . . . .	264
<b>4. Pruebas de hipótesis</b>	<b>267</b>
4.1. Introducción . . . . .	267
4.2. Conceptos elementales . . . . .	273
4.3. Función potencia . . . . .	278
4.4. Ejemplo de una prueba paramétrica . . . . .	282
4.5. Algunas pruebas sobre la distribución normal . . . . .	288
4.6. Lema de Neyman-Person . . . . .	302
<b>Apéndices</b>	
A. Fórmulas varias . . . . .	310
B. Sugerencias a los ejercicios . . . . .	332
<b>Bibliografía</b>	<b>396</b>
<b>Índice analítico</b>	<b>399</b>

# Introducción

La estadística es la ciencia que se encarga de recolectar, organizar, resumir y analizar datos para obtener ciertas afirmaciones a partir de ellos. En su perspectiva clásica, la estadística se clasifica en dos grandes ramas llamadas: estadística descriptiva y estadística inferencial. Para explicar cada una de estas ramas definiremos primero a una población como una colección cualquiera de personas u objetos sobre los cuales nos interesa estudiar algunas de sus características. En muchas situaciones, y por muy diversas razones, no es posible llevar a cabo un estudio exhaustivo sobre el total de la población, de modo que resulta necesario seleccionar un subconjunto, llamado muestra, y sobre este subconjunto se lleva a cabo la investigación.

Las preguntas o mediciones que nos interesa conocer en una población están representadas por variables aleatorias, cuyas características no conocemos completamente. Sólo sabemos los valores que éstas toman para los elementos de la muestra. A estos valores se les llama datos.

En la estadística descriptiva se estudian técnicas que ayudan a describir, mostrar o resumir la información de un conjunto de datos. Los procedimientos de la estadística descriptiva ayudan a comprender la información de una manera breve y resumida, y es particularmente útil cuando la cantidad de datos es grande. Esta descripción de la información se lleva a cabo a través de números, tablas o elementos gráficos. Las conclusiones que se obtienen se refieren únicamente a la muestra observada y no a la población completa.

En el primer capítulo del presente trabajo se proporciona una exposición breve de algunos elementos para la descripción numérica y gráfica de la in-

formación de un conjunto de datos. En los libros [2], [22], [25] y [26], por mencionar algunos, se pueden revisar más detalladamente algunas de las técnicas de la estadística descriptiva.

Por otra parte, en la estadística inferencial se estudian algunas técnicas y procedimientos con el objetivo de que la información de una muestra se generalice o extienda a la población completa. Las afirmaciones que se obtienen poseen necesariamente un cierto grado de imprecisión pues la información a partir de la cual se obtienen es parcial.

El material presentado en este trabajo está enfocado a revisar brevemente algunos aspectos de la estadística descriptiva y a estudiar algunos problemas clásicos de la estadística inferencial. Estudiaremos más particularmente el problema de la estimación de parámetros de los modelos de probabilidad previamente supuestos para las variables aleatorias. Otros textos en donde pueden consultarse con más detalle algunos de los temas expuestos son, por ejemplo, [5], [17], [18], [19] y [27]. A lo largo de la exposición daremos algunas sugerencias de otras fuentes bibliográficas.

# Capítulo 1

## Análisis exploratorio de datos

En este capítulo estudiaremos algunos elementos para la descripción, numérica y gráfica, de la información de un conjunto de datos. Esta exposición está basada principalmente en el libro [22], en donde pueden encontrarse un mayor número de ejemplos y ejercicios a nivel elemental. Supondremos que el lector ha tomado previamente un curso de probabilidad, de modo que será provechoso ligar los conceptos expuestos en este capítulo con algunas nociones elementales de Probabilidad.

### 1.1. Conceptos elementales

En esta sección se explican las nociones de población, muestra y variable. Se mencionan también las escalas de medición que pueden usarse para las variables. Empezaremos explicando la noción de población. Cotidianamente este término se utiliza para referirse a un determinado grupo de personas o seres vivos. Mediante la siguiente definición se amplía su significado y este es el sentido técnico que le daremos en la estadística.

**Definición 1.1** Una **población** es un conjunto de personas, objetos o eventos, de los cuales nos interesa estudiar algunas de sus características.

En un estudio estadístico, la población debe especificarse lo más completamente posible. Esto depende de lo que se desee o se pueda estudiar u observar, y de la forma en la que sea posible medir las características de nuestro interés. Veamos algunos ejemplos.

**Ejemplo 1.1** Los siguientes conjuntos pueden ser considerados como poblaciones para algún estudio estadístico.

- El conjunto de personas afectadas por una cierta enfermedad.
- El conjunto de personas extranjeras que llegan al país en un día.
- El conjunto de artículos defectuosos producidos en una fábrica.
- El conjunto de infracciones de tránsito que se cometen en una ciudad.
- El conjunto de goles que anota un equipo de futbol.
- El conjunto de boletas inválidas en un proceso de electoral.

Como puede verse en el ejemplo anterior, el alcance del concepto de población es muy amplio. Para un estudio estadístico, además de tener definida una población, a veces es conveniente establecer también una unidad de observación.

**Definición 1.2** Una **unidad de observación** es un grupo de elementos de una población del cual se tiene, o es posible obtener, su información de manera conjunta.

La determinación de una unidad de observación depende del problema a estudiar y de la manera en la que la información pueda ser obtenida o que esté disponible. Por ejemplo, en un análisis cuantitativo sobre los resultados de un proceso electoral, la información puede estar disponible por casillas electorales, y en este caso las casillas electorales (grupos de votantes) pueden ser consideradas como unidades de observación. En contraparte, si el

estudio trata acerca de la intención del voto previo a la elecciones, entonces cada persona que vota puede ser considerado como una unidad de observación.

Por simplicidad, consideraremos que cada elemento de una población es una unidad de observación y que nos interesa conocer ciertas características de estos elementos. En particular, al ejercicio cuando se llevan a cabo mediciones en la totalidad de la población se le llama censo. En este caso el análisis estadístico y sus conclusiones se refieren a la población completa. Sin embargo, por muy diversas razones (económicas, técnicas, imposibilidad, etc.) no es posible llevar a cabo mediciones en todos los elementos de la población, de modo que debemos escoger únicamente algunos elementos y de éstos obtener sus características. Por ejemplo, si el proceso de control de calidad de ciertos productos involucra su destrucción parcial o total, entonces no es razonable aplicar ese proceso a todos ellos. Así, a un subconjunto tomado de la población le llamaremos muestra, y a las mediciones que se hagan o que se tengan de una muestra les llamaremos datos.

**Definición 1.3** Una **muestra** es cualquier subconjunto de una población. Al número de elementos de la muestra, que denotaremos por la letra  $n$ , se le llama **tamaño de la muestra**.

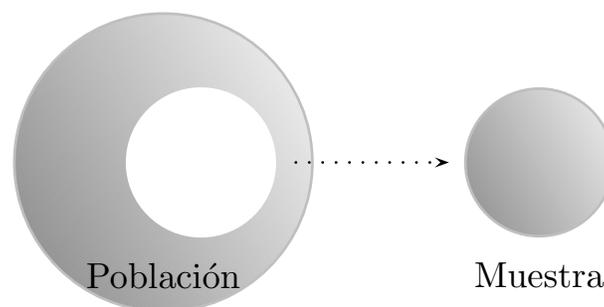


Figura 1.1

En la Figura 1.1 se presenta de manera gráfica y simple la noción de muestra como un subconjunto de una población. Regularmente las muestras se toman mediante un mecanismo azaroso, pero tales procedimientos dependen de lo que se desee estudiar, de la forma en la que se puedan medir las variables de interés y de la estructura o agrupación que posea la población como conjunto. La adecuada selección de una muestra es uno de los temas con mayor relevancia en la estadística.

Debemos mencionar, además, que en ocasiones es necesario definir dos o más poblaciones para llevar a cabo estudios comparativos de ciertas características de interés, o bien obtener dos o más muestras de una misma población. Asimismo puede presentarse la necesidad de incorporar la variable tiempo en el estudio y analizar la evolución temporal de una cierta característica.

## Ejercicios

1. Proponga una posible población en el caso de que se desee llevar a cabo un estudio estadístico sobre los siguientes temas.
  - a) La eficacia de un nuevo medicamento.
  - b) El nivel de consumo de un determinado producto.
  - c) La situación económica de las personas de edad avanzada.
  - d) Las fallas de un programa de cómputo.
  - e) Los defectos en artículos producidos por una maquinaria.

## Variables y datos

A lo que nos interesa medir y registrar en cada elemento de una población le llamaremos variable. Esto es así pues supondremos que una variable es una característica que varía de un elemento a otro de la población.

**Definición 1.4** Una **variable** es una característica de interés que posee cada elemento de una población y que podemos medir.

Una variable también puede considerarse como una pregunta que se le hace a cada elemento de la población, produciendo una respuesta en cada caso. Por ejemplo, en una población humana, podemos considerar la variable o pregunta: “¿Usted fuma?” y obtener como respuesta “si” o “no”. Para una población compuesta por un conjunto de tornillos podemos considerar la variable o pregunta “longitud del tornillo”, y obtener como resultado de la medición un valor numérico.

**Definición 1.5** Mediante el término **datos** se entiende al conjunto de observaciones de una o varias variables de interés para todos los elementos de una muestra.

Generalmente, un conjunto de datos se organiza y almacena en una computadora en la forma de un arreglo como el que se muestra en la Tabla 1.1. En esta tabla cada renglón representa una observación. En este caso tenemos a 5 personas para quienes se han registrado cuatro variables: edad, sexo, peso en kilogramos y estatura en centímetros.

Núm.	Edad	Sexo	Peso (kg.)	Estatura (cm.)
1	25	M	65	170
2	30	F	60	160
3	27	F	55	168
4	23	M	70	173
5	25	F	63	165

Tabla 1.1

De acuerdo al tipo de posibles respuestas que se obtengan, las variables se pueden clasificar en varios tipos. Estudiaremos esto en la siguiente sección.

## Ejercicios

2. Defina una posible variable o pregunta que podría ser de interés para cada una de las poblaciones propuestas en el ejercicio anterior.
3. Determine una posible variable o pregunta que tenga como uno de sus valores el indicado en cada inciso.

*a)* 10 años.

*e)* 23°C.

*b)* 1,265 veces.

*f)* María.

*c)* Pequeño.

*g)* Aprobado.

*d)* Blanco.

*h)* 1,200 hrs.

## Clasificación de variables

Una primera clasificación de variables establece que éstas pueden ser cuantitativas o cualitativas. Como estos nombres lo indican, la primera se refiere a una cantidad, mientras que la segunda se refiere a una cualidad. Veamos las definiciones.

**Definición 1.6** Una variable es **cuantitativa** si sus valores son números y representan una cantidad.

Por ejemplo, el número de hijos en una familia, la longitud de un tornillo, la cantidad de desperfectos de un artículo o el número de años cumplidos de una persona son variables cuantitativas.

**Definición 1.7** Una variable es **cualitativa** si sus valores representan una cualidad, un atributo o una categoría. Se les llama también variables **categorías**.

Por ejemplo, la religión de una persona, su sexo, o su preferencia por algún candidato en un proceso de elección son variables cualitativas pues sus valores son atributos de las personas. El lugar de nacimiento de una persona es otro ejemplo de variable cualitativa o categórica.

Observe que se pueden usar números para etiquetar los valores de una variable cualitativa, pero éstos no representan cantidades, sino que se usan dichos símbolos para denotar alguna cualidad. Por ejemplo, para clasificar la calidad de un producto se pueden usar los símbolos: 2 (bueno), 1 (regular), 0 (malo). En este caso, los símbolos numéricos se usan para clasificar la calidad de un producto, y no se trata realmente de valores numéricos.

Regresemos a las variables cuantitativas, éstas pueden clasificarse, además, en dos categorías de acuerdo al tipo de valores que toman: pueden ser discretas o continuas. Véase la Figura 1.2.

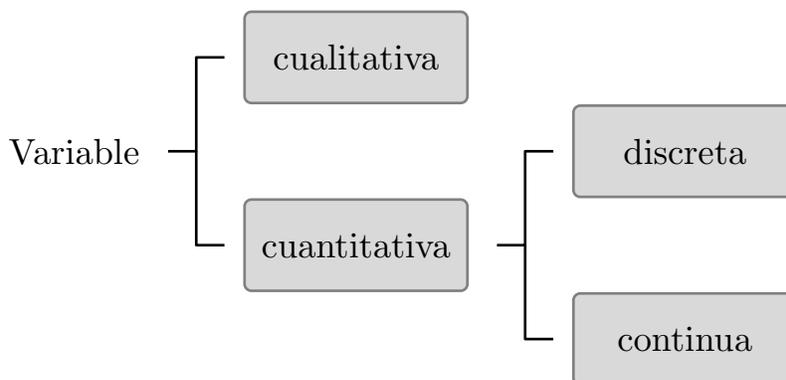


Figura 1.2

**Definición 1.8** Una variable cuantitativa es **discreta** si el conjunto de todos sus posibles valores tiene un número finito de elementos, o bien es infinito, pero se pueden numerar uno por uno de acuerdo al conjunto de números naturales.

Por ejemplo, la colección  $\{0, 1, 2, \dots, 120\}$  puede ser el conjunto de valores de una variable cuantitativa discreta, pues este conjunto tiene un número finito de elementos. Puede corresponder al número de hijos de una persona o el número de años promedio que le quedan por vivir a una persona.

Como otro ejemplo tenemos el conjunto  $\{0, 1, 2, \dots\}$ , que aunque es infinito, es discreto, ya que sus elementos se pueden numerar uno por uno de acuerdo al conjunto de números naturales. Los elementos de este conjunto pueden representar el número aproximado de cigarrillos que una persona fumadora ha consumido en toda su vida hasta el momento del estudio.

**Definición 1.9** Una variable cuantitativa es **continua** si puede tomar todos los valores dentro de un intervalo  $(a, b)$  de números reales y no toma valores aislados.

Por ejemplo, sin considerar la precisión limitada de los aparatos de medición, el tiempo que le toma a una persona llegar a su lugar de trabajo o escuela puede tomar valores continuos en el intervalo  $(0, \infty)$ . Más generalmente, el tiempo que le toma a una persona completar una cierta actividad puede tomar este conjunto de valores.

Pueden existir variables cuantitativas cuyos valores son todos los números dentro de un intervalo  $(a, b)$  y además algunos otros puntos aislados fuera de este intervalo. Estas variables se llaman mixtas; sin embargo, por simplicidad en nuestro tratamiento, no las consideraremos. Supondremos que nuestras variables cuantitativas son únicamente de los dos tipos descritos: discretas o continuas.

Finalmente mencionaremos que, en particular, a una variable que puede tomar únicamente dos valores se le llama variable dicotómica. Este término se aplica tanto para variables cualitativas como cuantitativas. Por ejemplo, el sexo de una persona es una variable cualitativa dicotómica pues puede tomar los valores masculino o femenino.

## Ejercicios

4. Clasifique las siguientes variables en cualitativas o cuantitativas. En caso de ser cuantitativas, diga si son discretas o continuas.
- a) El nivel de producción de una empresa en un año.
  - b) La demanda de taxis en un determinado sitio de una ciudad.
  - c) El nivel de felicidad de una persona.
  - d) El tiempo de vida útil de un foco.
  - e) El número de hijos en las familias.
  - f) La cantidad de agua en una presa.
  - g) La cantidad de dinero en una cuenta bancaria.
  - h) El estado civil de una persona.
  - i) El nivel de contaminación en el aire de una ciudad.
  - j) La precipitación pluvial en una región.
  - k) La preferencia sexual de una persona.
  - l) El tiempo necesario para llevar a cabo un trabajo.
  - m) La temperatura en un cierto lugar en un cierto día.
  - n) El número de mascotas por familia.
  - ñ) La causa de muerte de una persona.
  - o) El nivel de dominio de un idioma extranjero.
  - p) El coeficiente intelectual de una persona.
  - q) El nivel adquisitivo de una persona.
  - r) Los lugares de residencia de una persona previos al actual.
  - s) La actividad u oficio de una persona.
  - t) La capacidad de ahorro de una persona.
  - u) El capital de una persona en el momento de su retiro laboral.
  - v) El resultado de un examen de matemáticas.
  - w) El estado de salud de una persona.
  - x) La estatura promedio de un equipo de básquetbol.

## Escalas de medición

De acuerdo al tipo de valores que pueden tomar las variables, se pueden clasificar éstas de la siguiente manera. Para las variables cualitativas, las escalas de medición pueden ser de dos tipos: nominal u ordinal, mientras que las variables cuantitativas pueden medirse usando dos tipos de escalas: de intervalo o de razón. Explicaremos a continuación cada una de estas escalas. Empezaremos con el caso de las variables cualitativas.

**Definición 1.10** Se dice que una variable cualitativa se mide mediante una **escala nominal**, o que es de tipo nominal, si sus valores son etiquetas o atributos y no existe un orden entre ellos.

Por ejemplo, si nos interesa estudiar la variable cualitativa “sexo” en una población humana, sus dos posibles valores son: masculino y femenino. Estos dos valores son etiquetas, no existe un orden entre ellos y por lo tanto se trata de una variable de tipo nominal. Por otro lado, la variable cualitativa “nacionalidad” también es un ejemplo de una variable de tipo nominal pues sus posibles valores: argentina, española, etc. son atributos y no existe un orden entre ellos. Por simplicidad consideramos en este ejemplo que cada persona tiene una única nacionalidad principal. Como un tercer ejemplo considere la variable cualitativa “religión”, sus posibles valores son: budista, musulmana, católica, etc. y es claro que corresponde a una variable de tipo nominal pues no hay ningún orden natural entre estos valores.

Veamos ahora la definición de la escala ordinal.

**Definición 1.11** Se dice que una variable cualitativa se mide mediante una **escala ordinal**, o que es de tipo ordinal, si sus valores son etiquetas o atributos pero existe un cierto orden entre ellos.

Por ejemplo, podemos considerar que la variable cualitativa “estado en el que se encuentra un artículo” tiene como posibles valores: Malo, Regular y Bueno. Es claro que estos valores son atributos de un artículo y que existe un cierto orden entre estos valores, por lo tanto, se trata de una variable de tipo ordinal.

Como un segundo ejemplo, considere las siguientes calificaciones finales para un alumno en un curso: No Acreditado (NA), Suficiente (S), Bien (B) y Muy Bien (MB). Estos valores son etiquetas, pero es claro que existe un orden ascendente entre estos valores. Por consiguiente, esta variable, medida en el sentido indicado, es un ejemplo de una variable cualitativa de tipo ordinal.

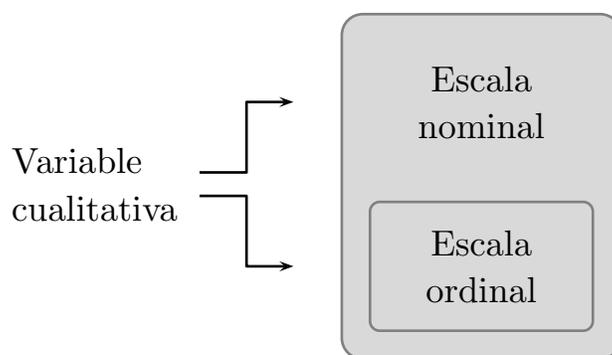


Figura 1.3

En la Figura 1.3 se muestran gráficamente los dos tipos de escala que se usan para variables cualitativas: nominal y ordinal. Observe la contención de conjuntos que se muestra en esta figura. Esta contención significa que toda variable de tipo ordinal puede considerarse como una variable de tipo nominal; ello se logra cuando no se contempla o se ignora el orden entre los valores de la variable. Sin embargo, la consideración contraria no es posible: no se puede crear un orden entre los valores de una variable de tipo nominal sin información o hipótesis adicionales. En la sección de ejercicios se encuentran algunos otros ejemplos de variables cualitativas con escalas de medición nominal y ordinal.

Ahora consideraremos el caso de variables cuantitativas. Recordemos que éstas pueden ser discretas o continuas. Sin embargo, en las siguientes definiciones no hay ninguna distinción a este respecto ya que son las mismas en ambos casos. También recordemos que los valores de una variable cuantitativa son números, por lo tanto ya existe un orden entre estos valores. Agregaremos ahora algunas condiciones adicionales a los valores numéricos de una variable cuantitativa para definir dos nuevos tipos de escalas de medición: la escala de intervalo y la escala de razón. En la Figura 1.4 se muestra la relación general que guardan estos dos tipos de escalas. Veamos primero la definición de escala de intervalo.

**Definición 1.12** Una variable cuantitativa se mide mediante una **escala de intervalo** si existe una noción de distancia entre los valores de la variable, aunque no se pueden realizar operaciones numéricas y no existe necesariamente el valor natural cero.

De esta manera no sólo tenemos la relación de orden entre los valores de una variable cuantitativa, sino que dados cualesquiera dos de sus valores podemos saber la distancia entre ellos. Por ejemplo, la escala Celsius (o Fahrenheit) para medir la temperatura es de tipo intervalo, pues existe una noción de distancia entre dos temperaturas, pero claramente no existe el valor cero natural o absoluto (el cero depende de la escala que se use, la temperatura  $0^{\circ}\text{C}$  no es la misma que  $0^{\circ}\text{F}$ ). Ahora veamos la definición de escala de razón.

**Definición 1.13** Una variable cuantitativa se mide mediante una **escala de razón** si sus valores tienen un sentido físico y existe el cero absoluto.

Por ejemplo, la variable cuantitativa discreta “edad en años cumplidos de una persona” tiene como posibles valores:  $0, 1, \dots, 150$ . Por cuestiones de finitud hemos considerado una edad máxima posible de 150 años. Es claro que puede considerarse que esta variable puede medirse mediante una escala de razón pues la variable puede tomar el valor cero absoluto y existe la

noción física del lapso de 1 año entre un valor y el siguiente en esta escala de medición.

Como un segundo ejemplo, considere la variable cuantitativa (podemos suponer discreta) “peso” de un bebé al nacer. Puesto que siempre existe una precisión finita con la que se efectúan las mediciones, podemos considerar que el conjunto de valores de esta variable cuantitativa tiene un número finito de elementos y que el valor cero está incluido. Esta variable entonces se puede medir mediante una escala de razón.

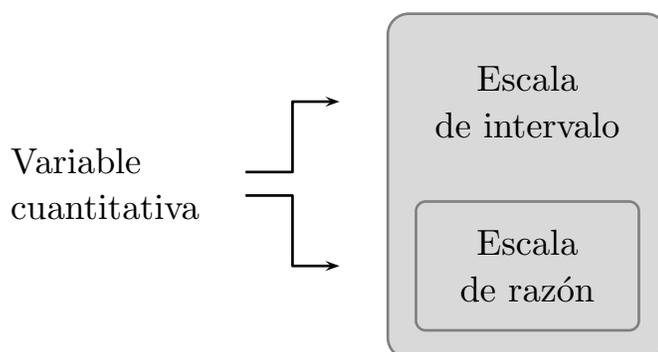


Figura 1.4

En la Figura 1.4 se muestran gráficamente los dos tipos de escala que se usan para variables cuantitativas. Observe que también aquí tenemos una contención de conjuntos. Esta contención significa que toda variable con escala de medición de tipo razón puede considerarse como una variable con escala de medición de tipo intervalo. Esto se consigue cuando no se contempla el sentido físico de la variable o no existe el cero absoluto. La consideración contraria no es posible.

**Advertencia.** Antes de concluir esta sección se debe mencionar que no existe una clasificación única y absoluta para una variable dada. Su tipificación dependerá del tratamiento y uso que de ella se haga. Tal vez la separación más fuerte se encuentre entre variables cualitativas y cuantitativas. De las

segundas, por cuestiones de precisión numérica, una variable continua bien puede considerarse discreta.

## Ejercicios

5. Los siguientes son ejemplos de variables cualitativas. Diga si su escala de medición puede ser nominal u ordinal.
  - a) Día de la semana con mayor tráfico en una ciudad.
  - b) Estado de uso de un artículo.
  - c) Preferencia sexual de una persona.
  - d) Calificación crediticia de un país.
  - e) Nivel de actividad (bajo, medio, alto) en una bolsa de valores durante una jornada.
  - f) Opinión de una persona sobre una decisión del gobierno.
  - g) Color del pelo de una persona.
  - h) La causa de muerte de una persona.
  
6. Los siguientes son ejemplos de variables cuantitativas. Diga si son discretas o continuas y si su escala de medición puede ser de intervalo o de razón.
  - a) Número del día del nacimiento de una persona.
  - b) Número de pruebas no aprobadas en un control de calidad.
  - c) Número de padres vivos de una persona.
  - d) La tasa de interés mensual en una moneda.
  - e) Número de casas en venta en una ciudad.
  - f) Nivel de avance de un estudiante en sus estudios en la universidad.
  - g) La tasa de desempleo en un país.
  - h) Nivel de aprobación de un presidente.

## Agrupamiento de valores

Para una variable cualitativa cualquiera tenemos una cierta cantidad de categorías como sus posibles valores. Algunas de estas categorías pueden agruparse en colecciones de categorías. Por otro lado, los valores de las variables cuantitativas se pueden agrupar en conjuntos  $C_1, C_2, \dots, C_k$ , en donde  $k$  es un cierto número entero. Estos grupos deben ser excluyentes y exhaustivos. Esto significa que cada valor de la variable se clasifica en uno (exhaustividad) y sólo uno de estos grupos (exclusión). En cualquier caso, a los agrupamientos resultantes les llamaremos clases.

**Definición 1.14** Una **clase** es una agrupación de valores de una variable.

Notemos que las clases pueden constar de una o varias categorías en el caso de variables cualitativas, o de uno o varios números en el caso de variables cuantitativas. Observemos, además, que al hacer este tipo de agrupamientos se puede perder información y el tipo de variable puede cambiar. Por ejemplo, la variable “Salario de un trabajador”, que pudo ser considerada originalmente como cuantitativa, se puede transformar en una variable con valores  $C_1 =$  “Salario bajo”,  $C_2 =$  “Salario medio” y  $C_3 =$  “Salario alto”, la cual sería ahora una variable cualitativa de tipo ordinal.

Supongamos entonces que los valores de una variable se agrupan en clases. Al llevar a cabo una observación (u obtener un dato) de la variable, ese valor pertenece a una de las clases definidas y se dice que la clase correspondiente fue observada.

También es posible un proceso contrario. Esto es, hay situaciones en donde los datos disponibles están agrupados en clases y no se tienen las observaciones individuales. Si por alguna razón se desea contar con los datos individuales, estos pueden obtenerse únicamente de manera aproximada eligiendo un dato representante de cada clase. A estos representantes se les llama marcas de clase.

**Definición 1.15** Una **marca de clase** es un dato que representa a una clase.

Por ejemplo, si una determinada clase es un intervalo  $[a, b]$  de  $\mathbb{R}$ , entonces una posible marca de clase puede ser el punto medio del intervalo, en este caso es el valor  $(a + b)/2$ . Cada observación que se tenga de un clase se reemplaza por su representante y de esta manera se genera una colección de valores individuales aproximados de la variable. No existe un método estándar para determinar las marcas de clase. Cada caso debe analizarse por separado y, si es necesario, justificar la elección de cada marca.

## Ejercicios

- Suponga que los valores de la variable “número promedio de tasas de café que una persona toma al día” se agrupan en las categorías: consumo bajo, consumo medio y consumo alto. Determine el tipo de variable creada mediante este agrupamiento y su escala de medición. Justifique la asignación de un representante (marca de clase) para cada categoría.
- Suponga que una cierta variable numérica toma valores en el intervalo  $(-a, a)$ , para algún número real  $a > 0$ . Suponga que se agrupan estos valores en dos categorías: valores positivos y valores negativos. Determine el tipo de variable creada mediante este agrupamiento y su escala de medición. Justifique la asignación de un representante (marca de clase) para cada categoría.

## 1.2. Descripciones numéricas

En esta sección se estudian varias fórmulas que tienen como objetivo resumir la información de un conjunto de datos, principalmente numéricos. Supongamos que tenemos un conjunto de  $n$  mediciones

$$x_1, \dots, x_n,$$

las cuales representan valores observados de cierta variable de interés. Existen varias formas de resumir la información de esta colección de datos. Primero definiremos algunas cantidades que buscan representar un valor central del conjunto de datos. Esta es la razón por la cual a estas cantidades se les llama medidas de tendencia central o medidas de localización.

Estas cantidades son: la media, la moda y la mediana. Definiremos también algunas medidas de dispersión, esto es, cantidades que buscan medir de alguna forma el grado de dispersión o separación entre los datos. Estudiaremos la varianza, la desviación estándar, la desviación media y el rango. Otras cantidades que también se definirán en esta sección y que ayudan a resumir la información de un conjunto de datos son las frecuencias y los cuantiles. En la tabla que aparece en la página 59 se resumen las cantidades que definiremos en esta sección. Empecemos entonces con las medidas de localización.

## Media

La media, o media aritmética, de un conjunto de datos numéricos es la medida de localización más utilizada. Su definición es la siguiente.

**Definición 1.16** La **media** de un conjunto de números  $x_1, \dots, x_n$  se denota por  $\bar{x}$  (se lee  $x$  barra) y se define como el promedio aritmético

$$\bar{x} = \frac{x_1 + \dots + x_n}{n}.$$

La media es un valor promedio que resume y representa a la colección de datos. Se le puede considerar como un representante promedio de los datos, aunque no necesariamente coincide con uno de ellos. La media puede interpretarse como el centro de equilibrio de las observaciones cuando éstas son consideradas como pesos que se colocan sobre un eje horizontal.

En ocasiones las  $n$  observaciones numéricas se encuentran registradas de la siguiente forma: se observan  $k$  valores distintos, los cuales denotaremos también por  $x_1, \dots, x_k$ , pero esta vez se tienen las frecuencias con las que

se han registrado estos valores. A estas frecuencias las denotaremos por  $f_1, \dots, f_k$  y son números enteros mayores o iguales a uno. Tenemos entonces que el dato  $x_i$  fue observado  $f_i$  veces. La suma de todas las frecuencias  $f_i$  es igual al tamaño  $n$  de la muestra, esto es,  $f_1 + \dots + f_k = n$ . La media se calcula como hemos indicado antes pero en este caso se reduce a la siguiente expresión

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

El siguiente resultado no es difícil de comprobar y se deja como ejercicio. Establece el cambio de la media bajo transformaciones lineales de los datos. La multiplicación por una constante corresponde a un cambio de escala, y la suma por alguna otra constante corresponde a una traslación de los datos.

**Proposición 1.1** Sea  $\bar{x}$  la media del conjunto de datos numéricos  $x_1, \dots, x_n$  y sea  $\bar{y}$  la media de los datos transformados  $y_i = ax_i + c$ , para  $i = 1, \dots, n$ , en donde  $a$  y  $c$  son dos constantes arbitrarias. Entonces

$$\bar{y} = a\bar{x} + c.$$

Se debe reiterar que la media es la medida de localización más utilizada, y muchas decisiones importantes son tomadas con base en esta cantidad. En el presente trabajo usaremos la media como punto de referencia para calcular algunas medidas de dispersión.

## Ejercicios

9. Sean  $x_1, \dots, x_n$  observaciones de una cierta variable cuantitativa. Demuestre que

$$a) \sum_{i=1}^n (x_i - \bar{x}) = 0.$$

$$b) \sum_{i=1}^n (x_i - \bar{x})^2 = \left( \sum_{i=1}^n x_i^2 \right) - n\bar{x}^2.$$

10. Considere el conjunto de cinco datos como aparece abajo. Determine el valor del dato faltante  $x_3$  si la media es 2.

$$\begin{aligned}x_1 &= 6, \\x_2 &= 0, \\x_3 &= \blacksquare \\x_4 &= -2, \\x_5 &= 5.\end{aligned}$$

11. Demuestre la Proposición 1.1.
12. Calcule la media de los siguientes conjuntos de datos. La primera columna corresponde al conjunto de datos original. El resto de las columnas se obtiene según la operación indicada en el encabezado.

$x$	$x + 1$	$x - 2$	$2x$	$x/3$
2	3	0	4	$2/3$
5	6	3	10	$5/3$
-1	0	-3	-2	$-1/3$
6	7	4	12	2

13. Calcule la media de los 10 datos que aparecen en la tabla de abajo. Además de los valores observados, se proporciona también la frecuencia con la que cada valor ha sido observado.

Valor	-2	0	2	4	8
Frecuencia	3	2	1	3	1

14. ¿Cuál es la media de un conjunto que consta de

- a) un único dato?  
 b) dos datos idénticos?  
 c) dos datos distintos?  
 d) cien datos idénticos?
15. Diga falso o verdadero:
- a) La media puede ser cero.  
 b) Si la media es cero, alguno de los datos es cero.  
 c) Dos conjuntos de datos distintos pueden tener la misma media.  
 d) Si se añade un cero a un conjunto de datos la media no cambia.  
 e) Si se añade un valor positivo a un conjunto de datos la media aumenta.  
 f) Si se añade un valor negativo a un conjunto de datos la media disminuye.
16. Si a un conjunto de datos se le añade  $k$  veces el valor constante  $a$ , ¿Cuál es la media cuando  $k \rightarrow \infty$ ?
17. ¿Cuál es la media de los primeros  $n$  números naturales?
18. Sea  $\bar{x}_n$  la media del conjunto de datos  $x_1, \dots, x_n$ . Suponga que se añade un dato adicional  $x_{n+1}$  a esta lista y sea  $\bar{x}_{n+1}$  la nueva media. Demuestre que

$$\bar{x}_{n+1} = \frac{1}{n+1} (n\bar{x}_n + x_{n+1}).$$

19. A un conjunto de datos se le añade un dato adicional y resulta que la media no cambia. ¿Cuál es el dato que se añadió?
20. Sea  $\bar{x}_n$  la media del conjunto de datos  $x_1, \dots, x_n$ . Suponga que se omite el dato  $x_i$  de esta lista y sea  $\bar{x}_{n-1}$  la nueva media. Demuestre que

$$\bar{x}_{n-1} = \frac{1}{n-1} (n\bar{x}_n - x_i).$$

21. Sea  $\bar{x}$  la media de  $x_1, \dots, x_n$  y sea  $\bar{y}$  la media de  $y_1, \dots, y_n$ . Demuestre que la media de los datos  $x_1 + y_1, \dots, x_n + y_n$  es

$$\bar{x} + \bar{y}.$$

22. Sea  $\bar{x}$  la media de  $x_1, \dots, x_n$  y sea  $\bar{y}$  la media de  $y_1, \dots, y_m$ . Demuestre que la media de los datos conjuntos  $x_1, \dots, x_n, y_1, \dots, y_m$  es

$$\frac{n}{n+m} \bar{x} + \frac{m}{n+m} \bar{y}.$$

23. Sea  $x_1, \dots, x_n$  un conjunto de datos numéricos con media  $\bar{x} \neq 0$ , y sea  $\bar{y}$  la media de los números  $y_i = x_i/\bar{x}$ , para  $i = 1, \dots, n$ . Demuestre que  $\bar{y} = 1$ .

24. **Media geométrica.** Para un conjunto de datos numéricos  $x_1, \dots, x_n$ , en donde cada uno de ellos es estrictamente positivo, se define la media geométrica como la raíz  $n$ -ésima del producto de todos estos números, es decir,

$$\text{mg}(x) = \sqrt[n]{x_1 \cdots x_n}.$$

Demuestre las siguientes afirmaciones:

a)  $\log(\text{mg}(x)) = \frac{1}{n} \sum_{i=1}^n \log x_i.$

- b) Si  $a > 0$  es una constante y  $ax$  denota el conjunto de números  $ax_1, \dots, ax_n$ , entonces

$$\text{mg}(ax) = a \cdot \text{mg}(x).$$

- c) Si  $y$  denota la colección de números  $y_1, \dots, y_n$ , todos ellos estrictamente positivos, y  $x/y$  denota la colección  $x_1/y_1, \dots, x_n/y_n$ , entonces

$$\text{mg}(x/y) = \frac{\text{mg}(x)}{\text{mg}(y)}.$$

- d) La media geométrica es siempre menor o igual a la media aritmética, es decir,  $\text{mg}(x) \leq \bar{x}$ , o más explícitamente,

$$\sqrt[n]{x_1 \cdots x_n} \leq \frac{x_1 + \cdots + x_n}{n}.$$

25. **Media armónica.** Para un conjunto de datos numéricos  $x_1, \dots, x_n$ , en donde cada uno de ellos es distinto de cero, se define la media armónica como el número

$$\text{ma}(x) = \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{x_i} \right)^{-1}.$$

Suponiendo que  $1/x$  denota la colección de datos  $1/x_1, \dots, 1/x_n$ , demuestre las siguientes fórmulas:

a)  $\text{ma}(1/x) = \frac{1}{\bar{x}}.$

b)  $\text{ma}(x) = \frac{n \cdot (x_1 \cdots x_n)}{\sum_{i=1}^n (x_1 \cdots x_n)/x_i}.$

c) La media armónica es siempre menor o igual a la media geométrica, y por el ejercicio anterior, esta última es menor o igual a la media aritmética, es decir,  $\text{ma}(x) \leq \text{mg}(x) \leq \bar{x}$ . Más explícitamente,

$$\frac{n}{\frac{1}{x_1} + \cdots + \frac{1}{x_n}} \leq \sqrt[n]{x_1 \cdots x_n} \leq \frac{x_1 + \cdots + x_n}{n}.$$

## Moda

A diferencia de la media, la moda se puede calcular tanto para valores numéricos como no numéricos. Su definición es la siguiente.

**Definición 1.17** La **moda** es el valor que aparece con mayor frecuencia en el conjunto de datos, en caso de que lo hubiera.

La moda es una medida de tendencia central de los datos pues indica el valor observado con mayor frecuencia. No existe una notación estándar para la moda. Se puede usar, por ejemplo, la expresión  $\text{Moda}(x)$ , en donde  $x$  representa el conjunto de observaciones  $x_1, \dots, x_n$ .

Sobre el cálculo de la moda tenemos las siguientes observaciones:

- La moda puede no existir, es decir, puede no haber un dato con frecuencia mayor al resto de los datos. En este caso se dice que el conjunto de datos no tiene moda.
- La moda puede existir y ser única. En este caso se dice que el conjunto de datos es unimodal.
- Pueden existir dos o más modas, es decir, puede haber dos o más valores o categorías que aparecen con la misma frecuencia máxima en el conjunto de datos. En este caso se dice que el conjunto de datos es bimodal o multimodal, según sea el caso.
- La moda puede permanecer sin cambio cuando se añaden u omiten datos cuya frecuencia es baja dentro del conjunto de datos.

El siguiente resultado no es difícil de comprobar y establece el cambio de la moda bajo transformaciones lineales, en el caso de datos numéricos.

**Proposición 1.2** Sea  $x_1, \dots, x_n$  un conjunto de datos numéricos con una única moda  $\text{Moda}(x)$ . Defina la colección de datos transformados  $y_i = ax_i + c$ , para  $i = 1, \dots, n$ , en donde  $a \neq 0$  y  $c$  son dos constantes. Entonces el conjunto de datos transformados  $y_1, \dots, y_n$  tiene una única moda dada por

$$\text{Moda}(y) = a \cdot \text{Moda}(x) + c.$$

Se debe reiterar que la moda puede calcularse para cualquier tipo de datos, sean éstos cualitativos o cuantitativos. Además, en el caso de tenerlos agrupados, se puede calcular la moda de estas clases o categorías, y se pueden usar los términos clase modal o intervalo modal, según sea el caso.

## Ejercicios

26. Diga falso o verdadero:

- a) La moda puede ser cero.

- b) La moda puede ser un número negativo.  
 c) La moda, cuando existe, es siempre alguno de los datos observados.  
 d) Si se le añaden ceros a un conjunto de datos, la moda, si existe, no cambia.

27. Demuestre la Proposición 1.2.

28. Calcule la moda de los siguientes conjuntos de datos. El primer renglón corresponde al conjunto de datos original. El resto de los renglones se obtiene según la operación indicada.

$x$	3	3	0	2	2	1	0	3	2	1	2
$x + 2$	5	5	2	4	4	3	2	5	4	3	4
$x/2$	3/2	3/2	0	1	1	1/2	0	3/2	1	1/2	1
$x - 2$	1	1	-2	0	0	-1	-2	1	0	-1	0
$2x$	6	6	0	4	4	2	0	6	4	2	4
$4x$	12	12	0	8	8	4	0	12	8	4	8

## Mediana

Esta es otra medida de tendencia central para datos numéricos. Supongamos nuevamente que tenemos una colección de números  $x_1, \dots, x_n$ . Podemos ordenarlos de menor a mayor, incluyendo repeticiones, y obtener la sucesión ordenada

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)},$$

en donde  $x_{(1)}$  denota el número más pequeño,  $x_{(2)}$  denota el segundo número más pequeño, etcétera, hasta  $x_{(n)}$ , que denota el número más grande. Es claro que algunos de estos números pueden aparecer varias veces en esta ordenación cuando se presentan repeticiones en las observaciones. En este

procedimiento es importante conservar estas repeticiones. La mediana se calcula de la siguiente forma.

**Definición 1.18** La **mediana** de un conjunto de números  $x_1, \dots, x_n$  es el dato ordenado de en medio, esto es,

- Si el número de datos  $n$  es par, entonces existen dos datos ordenados de en medio y la mediana es el promedio de estos dos números, es decir,  $[x_{(n/2)} + x_{(n/2+1)}]/2$ .
- Si el número de datos  $n$  es impar, entonces el dato ordenado de en medio es  $x_{((n+1)/2)}$  y este valor es la mediana.



Figura 1.5

En la Figura 1.5 se ilustra el cálculo de la mediana usando unos pocos datos representados por puntos distintos sobre el eje real en una situación simple. Se muestra el caso cuando el número de datos es par ( $n = 4$ ) y después cuando el número de datos es impar ( $n = 5$ ). De esta manera, la mediana es un valor que separa al conjunto de datos ordenados en dos partes iguales y representa un valor central típico del conjunto de datos, aunque puede no ser ninguno de los valores observados. No existe una notación estándar para la mediana, así es que en este trabajo la denotaremos por  $\tilde{x}$  (se lee  $x$  tilde).

La mediana es entonces un número que separa a los datos en dos partes con igual cantidad de datos a ambos lados: la primera parte son los números que son menores o iguales a la mediana, y la segunda parte corresponde al

conjunto de números que son mayores o iguales a la mediana.

No es difícil comprobar la validez de las siguientes afirmaciones acerca de la mediana.

- La mediana puede ser uno de los datos observados o no serlo. Lo es cuando el número de datos es impar, o cuando el número de datos es par y los dos datos de en medio son iguales.
- La mediana no es uno de los datos observados cuando el número de datos es par y los dos datos de en medio son distintos.
- La mediana es insensible a modificaciones de algunos de los datos, siempre y cuando estos cambios se efectúen dentro de la misma mitad de los datos.

La demostración del siguiente resultado se deja como ejercicio y establece el cambio de la mediana bajo transformaciones lineales de los datos.

**Proposición 1.3** Sea  $\tilde{x}$  la mediana del conjunto de datos numéricos  $x_1, \dots, x_n$  y sea  $\tilde{y}$  la media de los datos transformados  $y_i = ax_i + c$ , para  $i = 1, \dots, n$ , en donde  $a$  y  $c$  son dos constantes arbitrarias. Entonces

$$\tilde{y} = a\tilde{x} + c.$$

## Ejercicios

29. Demuestre la Proposición 1.3.

30. Calcule la mediana del siguiente conjunto de datos.

a) 10, 20, 30, 23, 12.

c) 3, 2, 0, 3.

b) 14, 15, 14, 30, 21.

d) 50, 30, 100, 20.

31. Calcule la mediana de los siguientes conjuntos de datos. La primera columna corresponde al conjunto de datos original. El resto de las columnas se obtiene según la operación indicada en el encabezado.

$x$	$x + 4$	$2x + 3$	$x - 2$	$x/2$	$5x$
4	8	11	2	2	20
6	10	15	4	3	30
6	10	15	4	3	30
2	6	7	0	1	10

32. ¿Cuál es la mediana de un conjunto que consta de
- a) un único dato?
  - b) dos datos idénticos?
  - c) dos datos distintos?
  - d) tres datos idénticos?
  - e) tres datos distintos?
  - f) mil datos idénticos?
33. A un conjunto de datos se le añade un dato adicional y resulta que la mediana no cambia. ¿Cuál es el dato que se añadió?
34. Considere un conjunto de datos cuya mediana es  $\tilde{x}$ . Diga falso o verdadero.
- a) Si se añade un dato a la izquierda de  $\tilde{x}$  y otro dato a la derecha de  $\tilde{x}$ , entonces la mediana no cambia.
  - b) Si se omite un dato a la izquierda de  $\tilde{x}$  y se omite otro dato a la derecha de  $\tilde{x}$ , entonces la mediana no cambia.

35. Calcule la mediana de los dos conjuntos de datos que aparecen en la Figura 1.6. Considere que se tienen tantos datos como puntos aparecen sobre el valor numérico.



Figura 1.6

36. Calcule la mediana de los dos conjuntos de datos que aparecen en la Figura 1.7. Considere que cada valor se observa con la frecuencia indicada arriba de la barra correspondiente.

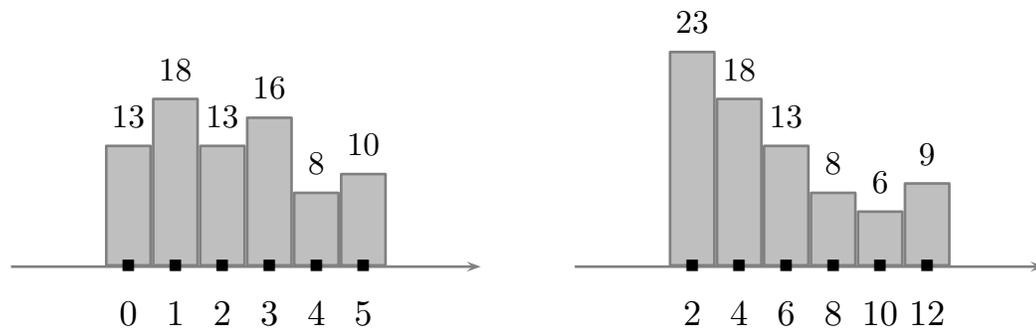


Figura 1.7

37. Diga falso o verdadero: si se le añaden ceros a un conjunto de datos, la mediana no cambia.

## Medidas de dispersión

Estudiaremos ahora algunas cantidades que permiten medir el grado de dispersión de un conjunto de datos numéricos. En casi todas estas medidas es necesario considerar un valor central de los datos como punto de referencia.

Como tal valor central se puede tomar a la media, a la mediana o a la moda. En cada caso se obtendrá una medida de dispersión diferente. Para seguir lo mayormente usado tomaremos como valor central a la media  $\bar{x}$ .

## Varianza

La varianza es un promedio de la distancia al cuadrado de cada uno de los datos  $x_i$  respecto de la media  $\bar{x}$  y es la medida de dispersión más comúnmente usada. Se calcula de la siguiente forma.

**Definición 1.19** La **varianza** de un conjunto de números  $x_1, \dots, x_n$  se denota por  $s^2$  y se define como sigue

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

Para especificar que se trata de la varianza de un conjunto de datos denotado por  $x$ , se escribe  $s_x^2$ ,  $s^2(x)$ , o también  $\text{var}(x)$ . Es claro que para calcularla primero es necesario encontrar la media  $\bar{x}$ . La varianza también puede definirse como se indica en la siguiente fórmula:

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

En esta expresión aparece el denominador  $n-1$  en lugar de  $n$ . Esta fórmula es usada con mucha frecuencia debido a que, cuando se aplica al caso de variables aleatorias, satisface una propiedad estadística importante llamada insesgamiento, la cual estudiaremos más adelante. Así, debe tenerse en cuenta esta diferencia en el cálculo de la varianza, aunque numéricamente la diferencia entre ellas usualmente es pequeña para valores grandes de  $n$ . En el presente capítulo usaremos la fórmula con denominador  $n$ , pues es más natural y consistente con otras cantidades que definiremos más adelante llamadas momentos.

La demostración del siguiente resultado se deja como ejercicio. Muestra el cambio que tiene la varianza bajo transformaciones lineales de los datos. Multiplicar por una constante corresponde a un cambio de escala y sumar una constante corresponde a una traslación.

**Proposición 1.4** Sea  $\text{var}(x)$  la varianza del conjunto de datos numéricos  $x_1, \dots, x_n$  y sea  $\text{var}(y)$  la media de los datos transformados  $y_i = ax_i + c$ , para  $i = 1, \dots, n$ , en donde  $a$  y  $c$  son dos constantes arbitrarias. Entonces

$$\text{var}(y) = a^2 \cdot \text{var}(x).$$

El cálculo de la varianza para datos agrupados puede efectuarse de la siguiente forma: si se tienen  $n$  observaciones de  $k$  valores distintos  $x_1, \dots, x_k$  con frecuencias  $f_1, \dots, f_k$ , la varianza se reduce a la fórmula:

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

## Ejercicios

38. Demuestre la Proposición 1.4.
39. ¿Puede un conjunto de datos numéricos tener varianza cero?
40. Considere el conjunto de datos de dos números:  $x_1$  y  $x_2$ . Encuentre estos números si la media es 2 y la varianza es 9.
41. Sean  $x_1, \dots, x_n$  observaciones numéricas de una cierta variable de interés y sea  $s^2$  su varianza. Demuestre que

$$a) \quad s^2 = \left( \frac{1}{n} \sum_{i=1}^n x_i^2 \right) - \bar{x}^2.$$

$$b) \quad s^2 = \left[ \frac{1}{n} \sum_{i=1}^n (x_i - c)^2 \right] - (\bar{x} - c)^2, \quad c \text{ constante.}$$

42. Sea  $s_x^2$  la varianza del conjunto de datos numéricos  $x_1, \dots, x_n$ . Suponga que estos números se transforman en los datos  $ax_1 + c, \dots, ax_n + c$ , en donde  $a$  y  $c$  son dos constantes. Sean  $y_1, \dots, y_n$  los nuevos datos y sea  $s_y^2$  su varianza. Demuestre que

$$s_y^2 = a^2 \cdot s_x^2.$$

43. Calcule la varianza de los siguientes conjuntos de datos. La primera columna corresponde al conjunto de datos original. El resto de las columnas se obtiene según la operación indicada en el encabezado.

$x$	$x + 2$	$x - 2$	$2x$	$x/2$	$2x + 1$
1	3	-1	2	1/2	3
1	3	-1	2	1/2	3
0	2	-2	0	0	1
4	6	2	8	2	9
2	4	0	4	1	5

44. Sean  $x_1, \dots, x_n$  observaciones numéricas de una cierta variable de interés. Encuentre el valor de  $u$  que minimiza la función

$$g(u) = \sum_{i=1}^n (x_i - u)^2.$$

45. Diga falso o verdadero.

- a) Sea  $x_1, \dots, x_n$  una colección de datos numéricos en donde cada uno de estos registros es igual a un mismo valor. La varianza de esta colección de datos es cero.
- b) Si la varianza de un conjunto de datos es cero, entonces todos los datos son idénticos.

- c) Si se le añaden ceros a un conjunto de datos, la varianza no cambia.
46. Sea  $x_1, \dots, x_n$  un conjunto de datos numéricos con media 2 y varianza 4. Encuentre la media y la varianza de los datos  $y_1, \dots, y_n$ , en donde
- $y_i = 3x_i$ .
  - $y_i = -x_i + 2$ .
  - $y_i = x_i - 3$ .
  - $3x_i + 2y_i = 0$ .
47. Sea  $x_1, \dots, x_n$  un conjunto de datos numéricos con media  $\bar{x} \neq 0$ . Defina  $y_i = x_i/\bar{x}$  para  $i = 1, 2, \dots, n$ . Demuestre que

$$s_y^2 = \frac{1}{\bar{x}^2} \cdot s_x^2.$$

## Desviación estándar

A la raíz cuadrada positiva de la varianza se le llama desviación estándar o desviación típica, y se le denota por la letra  $s$ .

**Definición 1.20** La **desviación estándar** de un conjunto de números  $x_1, \dots, x_n$  se denota por la letra  $s$  y se define como sigue

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

A diferencia de la varianza, la desviación estándar posee la buena cualidad de estar expresada en las mismas unidades de medición que la de los datos observados. Por ejemplo, si los datos son observaciones de una variable de longitud medida en metros, entonces la varianza tiene unidades de medición en metros cuadrado, mientras que la desviación estándar expresa una cantidad en metros.

A continuación se menciona el cambio que tiene la desviación estándar cuando los datos observados se modifican mediante una transformación lineal. Este resultado es una consecuencia inmediata del resultado correspondiente a la varianza.

**Proposición 1.5** Sea  $s(x)$  la desviación estándar del conjunto de datos numéricos  $x_1, \dots, x_n$  y sea  $s(y)$  la desviación estándar de los datos transformados  $y_i = ax_i + c$ , para  $i = 1, \dots, n$ , en donde  $a$  y  $c$  son dos constantes arbitrarias. Entonces

$$s(y) = |a| \cdot s(x).$$

Claramente el cálculo de la desviación estándar para datos agrupados se lleva a cabo de la siguiente forma: si se tienen  $n$  observaciones de  $k$  valores distintos  $x_1, \dots, x_k$  con frecuencias  $f_1, \dots, f_k$ , la desviación estándar es:

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

## Ejercicios

48. Demuestre la Proposición 1.5.
49. Calcule la desviación estándar de los siguientes conjuntos de datos. La primera columna corresponde al conjunto de datos original. El resto de las columnas se obtiene según la operación indicada en el encabezado.

$x$	$x - 1$	$x + 2$	$2x - 2$
0	-1	2	-2
2	1	4	2
2	1	4	2
4	3	6	6

50. Sea  $x_1, \dots, x_n$  un conjunto de datos numéricos con media  $\bar{x}$  y desviación estándar  $s_x > 0$ . Suponga que estos números se transforman en los datos  $(x_1 - \bar{x})/s_x, \dots, (x_n - \bar{x})/s_x$ . Sean  $y_1, \dots, y_n$  los nuevos datos. Demuestre que

- a)  $\bar{y} = 0$ .  
 b)  $s_y^2 = 1$ .

## Desviación media

Al promedio de los valores absolutos de las diferencias entre los datos y la media se le llama desviación media.

**Definición 1.21** La **desviación media** del conjunto de números  $x_1, \dots, x_n$  se denota por  $\text{dm}(x)$  y se calcula de la siguiente forma

$$\text{dm}(x) = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|.$$

La desviación media es otra medida de la dispersión de un conjunto de datos numéricos. Existe también el término desviación media absoluta (*mean absolute deviation*) que se calcula como antes pero tomando a la mediana de los datos como punto central y no la media  $\bar{x}$  como lo hemos hecho aquí. Si se hace uso de una computadora y se emplea alguna función predefinida

para calcular la desviación media, es recomendable verificar el punto central empleado en el cálculo de la función.

La desviación media cambia bajo transformaciones lineales de los datos, como se muestra en el siguiente resultado.

**Proposición 1.6** Sea  $x$  el conjunto de datos numéricos  $x_1, \dots, x_n$  y sea  $ax + c$  el conjunto de datos transformados  $ax_i + c$ , para  $i = 1, \dots, n$ , en donde  $a$  y  $c$  son dos constantes arbitrarias. Entonces

$$\text{dm}(ax + c) = |a| \cdot \text{dm}(x).$$

## Ejercicios

51. Demuestre la Proposición 1.6.
52. Calcule la desviación media de los siguientes conjuntos de datos. La primera columna corresponde al conjunto de datos original. El resto de las columnas se obtiene según la operación indicada en el encabezado.

$x$	$x + 1$	$x - 2$	$2x$	$x/2$	$-5x$
2	3	0	4	1	-10
2	3	0	4	1	-10
0	1	-2	0	0	0
1	2	-1	2	1/2	-5
1	2	-1	2	1/2	-5
0	1	-2	0	0	0

## Rango

Para calcular esta cantidad es necesario identificar el dato más pequeño  $x_{(1)}$  y el dato más grande  $x_{(n)}$  de un conjunto de números  $x_1, \dots, x_n$ . El rango se denota por la letra  $r$  y define como el dato mayor menos el dato menor.

**Definición 1.22** El **rango** de un conjunto de números  $x_1, \dots, x_n$  es

$$r = x_{(n)} - x_{(1)}.$$

Es claro que el rango de un conjunto de números es una medida de dispersión, pues indica la distancia máxima entre cualesquiera dos datos. El rango también puede interpretarse como la longitud del intervalo más pequeño en el que se encuentran todos los datos observados.

Usaremos la expresión  $r_x$ ,  $r(x)$  o  $\text{Rango}(x)$  para denotar al rango de un conjunto de números  $x_1, \dots, x_n$ , aunque ninguna de estas notaciones es estándar. Observe que el rango es una cantidad mayor o igual a cero y que no cambia cuando se añaden u omiten datos, siempre y cuando no se modifique el valor máximo ni el valor mínimo de la colección de datos originales. Otra propiedad interesante del rango se establece en el siguiente recuadro.

**Proposición 1.7** Sea  $x$  el conjunto de datos numéricos  $x_1, \dots, x_n$  y sea  $ax + c$  el conjunto de datos transformados  $ax_i + c$ , para  $i = 1, \dots, n$ , en donde  $a$  y  $c$  son dos constantes arbitrarias. Entonces

$$r(ax + c) = |a| \cdot r(x).$$

Se deja como ejercicio reflexionar sobre la forma de calcular el rango en el caso de que los datos numéricos se encuentren agrupados.

## Ejercicios

53. Demuestre la Proposición 1.7.
54. Encuentre una fórmula para calcular el rango de un conjunto de datos agrupados.
55. Calcule el rango del siguiente conjunto de datos.
- a)  $1, 2, \dots, n$ .
  - b)  $a, 2a, \dots, na$ , con  $a \in \mathbb{R}$ .
  - c)  $a^1, a^1 + a^2, \dots, a^1 + a^2 + \dots + a^n$ , con  $a \geq 1$ .
  - d)  $-1, +1, -1, +1, \dots, (-1)^n$ .
56. Diga falso o verdadero.
- a) El rango de un conjunto de datos puede ser cero.
  - b) El rango de un conjunto de datos puede ser negativo.
  - c) Dos conjuntos de datos distintos pueden tener el mismo rango.
  - d) Si  $x_1 \leq y_1, \dots, x_n \leq y_n$  entonces  $r(x) \leq r(y)$ .

## Coeficiente de variación

Esta es una cantidad con la cual se propone una forma distinta de medir la dispersión de un conjunto de datos numéricos.

**Definición 1.23** Sea  $x_1, \dots, x_n$  una colección de  $n$  observaciones de una variable cuantitativa con media  $\bar{x} \neq 0$  y desviación estándar  $s(x)$ . Al siguiente cociente se le conoce como **coeficiente de variación**.

$$\text{cv}(x) = \frac{s(x)}{\bar{x}}.$$

Recordemos que tanto la desviación estándar  $s(x)$  como la media  $\bar{x}$  poseen las mismas unidades de medición. Por lo tanto, el cociente de estas cantidades no posee unidad de medición y, en consecuencia, este coeficiente puede

servir para comparar la dispersión de dos o más conjuntos de datos numéricos.

A continuación presentamos una propiedad general de este coeficiente, la cual no es difícil demostrar.

**Proposición 1.8** Sea  $x$  el conjunto de datos numéricos  $x_1, \dots, x_n$  y sea  $ax + c$  el conjunto de datos transformados  $ax_i + c$ , para  $i = 1, \dots, n$ , en donde  $a \neq 0$  y  $c$  son dos constantes. Entonces

$$cv(ax + c) = \frac{|a| \cdot s(x)}{a\bar{x} + c}.$$

Para el caso de datos agrupados, las cantidades  $s(x)$  y  $\bar{x}$  se calculan como se ha indicado anteriormente en esta situación y después se aplica directamente la fórmula de la definición de arriba.

## Ejercicios

57. Demuestre la Proposición 1.8.

58. Diga falso o verdadero.

- a) El coeficiente de variación puede ser negativo.
- b) El coeficiente de variación puede ser cero.

59. Sea  $x_1, \dots, x_n$  un conjunto de datos numéricos con media  $\bar{x} \neq 0$ . Defina  $y_i = x_i/\bar{x}$  para  $i = 1, 2, \dots, n$ . Demuestre que

$$cv(y) = \frac{1}{|\bar{x}|} \cdot s_x = \begin{cases} cv(x) & \text{si } \bar{x} > 0, \\ -cv(x) & \text{si } \bar{x} < 0. \end{cases}$$

## Momentos

Las cantidades que hemos definido como media y varianza pueden generalizarse a un concepto más amplio llamado momento.

**Definición 1.24** Sea  $x$  una colección de observaciones  $x_1, \dots, x_n$  de una variable cuantitativa y sea  $k \geq 1$  un número entero. A la cantidad definida a continuación se le llama el  **$k$ -ésimo momento**, o bien, **momento de orden  $k$**  del conjunto de datos.

$$m'_k(x) = \frac{1}{n} \sum_{i=1}^n x_i^k.$$

Se trata simplemente del promedio aritmético de cada uno de los datos elevados a la potencia  $k$ . El valor entero de  $k$  determina el numeral del momento, así por ejemplo, tenemos que el primer momento ( $k = 1$ ) es la media, el segundo momento ( $k = 2$ ) es el promedio de los datos elevados al cuadrado, etcétera. Si  $x$  denota el conjunto de datos  $x_1, \dots, x_n$ , entonces se puede usar el término  $m'_k(x)$  para denotar el  $k$ -ésimo momento de  $x$ .

Cada momento es una medida de cierta característica de los datos. Sin embargo, no se conoce la característica que se está midiendo en cada caso, únicamente se conoce para los primeros pocos momentos. Por ejemplo, el primer momento es la media y esta es una medida de localización o centralidad de los datos; el segundo momento está relacionado con la varianza y esta es una medida de la dispersión de los datos; el tercer momento está relacionado con la asimetría de los datos, el cuarto momento está relacionado con la forma de las colas de la gráfica de frecuencias de los datos, es decir, de la manera en la que decae o se desvanece a cero la gráfica de frecuencias en sus dos extremos: izquierdo y derecho. Y esto es todo, en general no existen interpretaciones bien establecidas para los momentos de orden superior, de la misma forma que no se conocen interpretaciones para todas las derivadas de una función.

Existen además otros tipos de momentos como el siguiente.

**Definición 1.25** El  $k$ -ésimo momento central, o bien, el **momento central de orden  $k$**  del conjunto de datos numéricos  $x_1, \dots, x_n$  es

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

Es decir, tenemos nuevamente un promedio aritmético pero esta vez se trata de los datos centralizados al restarles a cada uno de ellos la media  $\bar{x}$ . No es difícil verificar que  $m_1(x) = 0$  y que  $m_2(x)$  es la varianza de los datos  $x$ .

En la sección de ejercicios se encuentran las expresiones de los momentos bajo transformaciones lineales de los datos.

## Ejercicios

60. Sea  $m_k$  el  $k$ -ésimo momento central de un conjunto de datos. Demuestre que

a)  $m_1 = 0$ .

b)  $m_2 = s^2$ .

c)  $m_2 = m'_2 - (m'_1)^2$ .

61. Sea  $x$  el conjunto de datos numéricos  $x_1, \dots, x_n$ , y sean  $a$  y  $c$  dos constantes. Demuestre que

a)  $m'_k(ax + c) = a^k \cdot \sum_{j=0}^k \binom{k}{j} m'_j(x) (c/a)^{k-j}$ , con  $a \neq 0$ .

b)  $m_k(ax + c) = a^k \cdot m_k(x)$ .

62. Diga cierto o verdadero.

a)  $m'_{2k}(x) \geq 0$ .

b)  $m_{2k}(x) \geq 0$ .

c)  $m_{2k}(x) = 0 \Leftrightarrow$  todos los datos son idénticos.

## Frecuencias

Supongamos que  $C_1, \dots, C_k$  representan  $k$  categorías de una variable cualitativa, o bien agrupamientos excluyentes y exhaustivos de los valores de una variable cuantitativa. A estas categorías o agrupamientos les hemos llamado clases, y la letra  $C$  ayuda a recordar su significado. Al hacer  $n$  observaciones de la variable en estudio se puede contar el número de veces que cada una de estas clases fue observada. Supongamos que la clase  $C_i$  fue observada  $f_i$  veces,  $i = 1, \dots, k$ . A estas cantidades se les llama frecuencias absolutas o simplemente frecuencias.

**Definición 1.26** La **frecuencia** de una clase (categoría o conjunto de valores) es el número de veces que la clase es observada.

Como se tienen  $n$  observaciones de la variable, tenemos que  $f_1 + \dots + f_k = n$ . Esta información puede representarse en forma tabular como se muestra en la Tabla 1.2. En esta tabla están representadas todas las clases consideradas y sus respectivas frecuencias. Esta es una manera muy útil de resumir los datos y se pueden elaborar gráficas como la que aparece en la Figura 1.8. En la gráfica mostrada hemos supuesto que existe algún tipo de orden entre las clases pues hemos colocado primero  $C_1$ , después a  $C_2$ , y así sucesivamente. Estas clases, sin embargo, pueden no tener orden entre ellas y no es relevante el orden en el que se grafican las frecuencias.

Clase	Frecuencia
$C_1$	$f_1$
$\vdots$	$\vdots$
$C_k$	$f_k$

Tabla 1.2

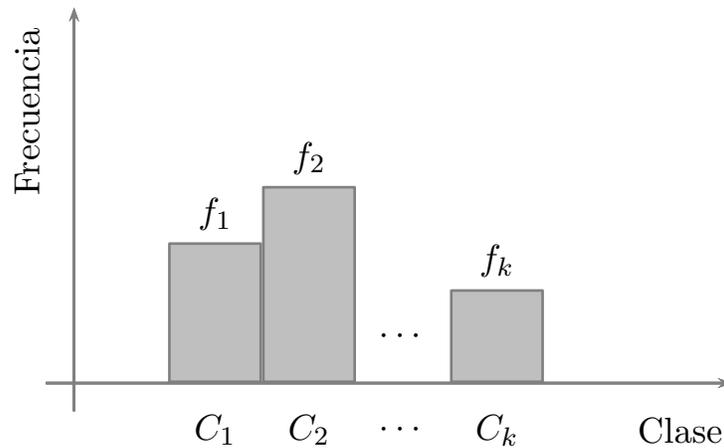


Figura 1.8

Cuando las clases  $C_1, \dots, C_k$  poseen un orden natural y se han definido de menor a mayor como indica el subíndice, es decir,  $C_1 \leq C_2 \leq \dots \leq C_k$ , es útil también considerar las frecuencias acumuladas.

**Definición 1.27** La **frecuencia acumulada** de una clase (categoría o conjunto de valores) es el número total de veces que la clase considerada, junto con las clases anteriores, fueron observadas.

Es decir, si como antes  $f_1, \dots, f_k$  denotan las frecuencias de las clases  $C_1, \dots, C_k$ , entonces la frecuencia acumulada de la clase  $C_j$  es la suma  $f_1 + \dots + f_j$ . Los valores de estas frecuencias acumuladas se muestran en la tercera columna de la Tabla 1.3, y una gráfica general se muestra en la Figura 1.9.

Clase	Frecuencia	Frecuencia acumulada
$C_1$	$f_1$	$f_1$
$C_2$	$f_2$	$f_1 + f_2$
$\vdots$	$\vdots$	$\vdots$
$C_k$	$f_k$	$f_1 + \cdots + f_k$

Tabla 1.3

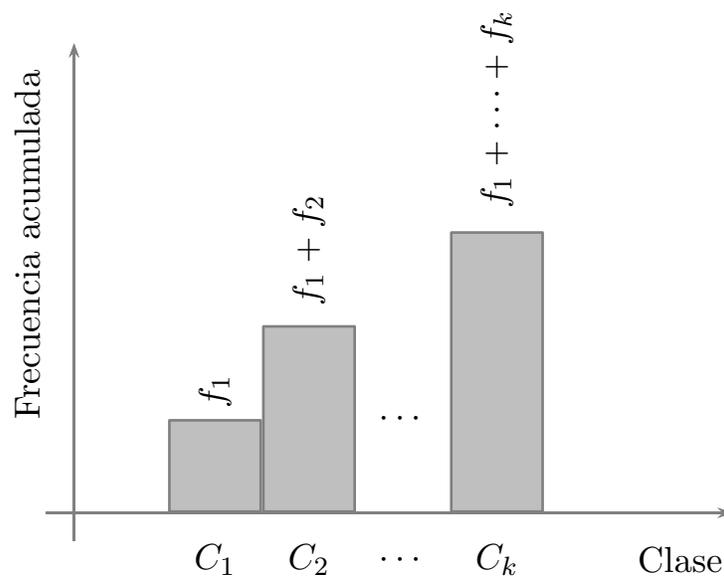


Figura 1.9

Se pueden definir también las frecuencias relativas al dividir cada frecuencia (absoluta) entre el número total de observaciones. A las cantidades así obtenidas se les llama frecuencias relativas. En este caso no es necesario que

haya un orden entre las clases, las frecuencias relativas se pueden calcular también para valores nominales o categóricos.

**Definición 1.28** La **frecuencia relativa** de una clase (categoría o conjunto de valores) es el número de veces que la clase fue observada dividido entre el total de observaciones.

De esta manera, si  $f_1, \dots, f_k$  son las frecuencias absolutas, entonces las cantidades  $f_1/n, \dots, f_k/n$  son las frecuencias relativas, suponiendo que fueron  $n$  observaciones totales. Estas nuevas frecuencias se muestran en la tercera columna de la Tabla 1.4.

Categoría	Frecuencia	Frecuencia relativa	Frecuencia relativa porcentual
$C_1$	$f_1$	$f_1/n$	$100 \cdot f_1/n \%$
$C_2$	$f_2$	$f_2/n$	$100 \cdot f_2/n \%$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$C_k$	$f_k$	$f_k/n$	$100 \cdot f_k/n \%$
Suma	$n$	1	100

Tabla 1.4

Observemos que las frecuencias relativas son números en el intervalo unitario  $[0, 1]$  y que la suma de todas estas cantidades es 1. Cuando estas frecuencias relativas se expresan como porcentajes, es decir, cuando se multiplican por 100, se llaman frecuencias relativas porcentuales. Estas cantidades son equivalentes a las primeras y se muestran en la cuarta columna de la Tabla 1.4.

Considerando nuevamente el caso cuando las categorías  $C_1, \dots, C_k$  poseen un cierto orden natural y se han definido de menor a mayor como indica el subíndice, se pueden definir también las frecuencias relativas acumuladas.

**Definición 1.29** La **frecuencia relativa acumulada** de una clase (categoría o conjunto de valores) es la suma de las frecuencias relativas anteriores e inclusive la clase en cuestión.

Es decir, la frecuencia relativa acumulada de la clase  $C_j$  es la suma

$$f_1/n + \dots + f_j/n.$$

Los valores de estas frecuencias relativas acumuladas se muestran en la tercera columna de la Tabla 1.5. En la cuarta columna aparecen estas mismas cantidades expresadas en porcentaje.

Clase	Frecuencia relativa	Frecuencia relativa acumulada	Frecuencia relativa acumulada porcentual
$C_1$	$f_1/n$	$f_1/n$	$100 \cdot f_1/n \%$
$C_2$	$f_2/n$	$f_1/n + f_2/n$	$100 \cdot (f_1/n + f_2/n) \%$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$C_k$	$f_k/n$	$f_1/n + \dots + f_k/n$	$100 \cdot (f_1/n + \dots + f_k/n) \%$

Tabla 1.5

## Ejercicios

63. Suponga que se tiene una variable cualitativa ordinal con valores ordenados de menor a mayor  $A, B, C, D$ . Suponga además que una serie de observaciones de esta variable produce las frecuencias que aparecen

en la siguiente tabla. Complete esta tabla calculando las frecuencias faltantes. Elabore además una gráfica de la frecuencia y otra de la frecuencia acumulada.

Valor	Frecuencia	Frecuencia acumulada	Frecuencia relativa	Frecuencia relativa acumulada
A	4			
B	2			
C	3			
D	2			

64. Suponga que se tiene una variable cualitativa ordinal con valores ordenados de menor a mayor  $A, B, C, D$ . Suponga además que una serie de 20 observaciones de esta variable produce las frecuencias relativas que aparecen en la siguiente tabla. Complete esta tabla calculando las frecuencias faltantes. Elabore una gráfica de la frecuencia, y otra de la frecuencia acumulada.

Valor	Frecuencia	Frecuencia acumulada	Frecuencia relativa	Frecuencia relativa acumulada
A			0.15	
B				
C			0.20	
D			0.25	

## Cuantiles

Consideremos nuevamente que  $x_1, \dots, x_n$  es un conjunto de  $n$  observaciones de una cierta variable cuantitativa de interés, y que estos valores se ordenan

de menor a mayor, conservando las repeticiones. Un cuantil es un número que separa a los datos en dos partes: un cierto porcentaje de los datos son menores o iguales al cuantil y el porcentaje complementario corresponde a datos que son mayores o iguales al cuantil.

Para dar una definición más precisa de cuantil consideraremos que  $p$  es un número cualquiera conocido tal que  $0 < p \leq 1$ . Este valor determinará los porcentajes de los que hablamos en el párrafo anterior. Por ejemplo, podemos suponer que  $p = 0.5$ . Entonces un cuantil es un número  $c$  tal que la proporción de valores  $x_i$  que son menores o iguales a  $c$  es del 50 %, es decir, la mitad de los datos son menores o iguales al cuantil. Al mismo tiempo debe cumplirse que la proporción de valores  $x_i$  que son mayores o iguales a  $c$  es el porcentaje complementario, esto es, el 50 %. En este caso, al número  $c$  se le llama cuantil de orden  $p = 0.5$  o bien cuantil al 50 % y no es difícil darse cuenta que pueden existir distintos valores  $c$  que cumplan las condiciones mencionadas, en otras palabras, el cuantil puede no ser único. En general, podemos tener cuantiles al 5 %, 10 %, 50 %, o cualquier otro porcentaje dado por la expresión  $100p$  %, con  $0 < p \leq 1$ . Con las ideas introductorias anteriores, podemos ahora dar la definición formal de cuantil para un conjunto de datos numéricos.

**Definición 1.30** Sean  $x_1, \dots, x_n$  observaciones de una variable cuantitativa y sea  $p$  un número tal que  $0 < p \leq 1$ . Un **cuantil- $p$**  del conjunto de datos es un número  $c$  tal que cumple las siguientes dos condiciones al mismo tiempo:

$$\frac{\#\{x_i : x_i \leq c\}}{n} \geq p \quad \text{y} \quad \frac{\#\{x_i : x_i \geq c\}}{n} \geq 1 - p.$$

Recordemos que si  $A$  es un conjunto, entonces la expresión  $\#A$  representa la cardinalidad o número de elementos en el conjunto  $A$ . De este modo la primera desigualdad que aparece en el recuadro anterior establece que la proporción de observaciones menores o iguales al cuantil  $c$  es, por lo menos,  $p$ . La segunda desigualdad establece que la proporción de observaciones que

son mayores o iguales a  $c$  es por los menos  $1 - p$ .

Observemos que se pide que el porcentaje de datos a la izquierda del cuantil sea por lo menos del  $100p\%$  y no necesariamente este porcentaje exacto. Análogamente, se pide que el porcentaje de datos que se encuentran a la derecha del cuantil sea, por lo menos, del  $100(1 - p)\%$  y no necesariamente este porcentaje exacto.

Como hemos mencionado antes, al número  $c$  se le llama cuantil- $p$ , pero también se usa el término cuantil de orden  $p$ , y también cuantil al  $100p\%$ . Para hacer referencia a la probabilidad  $p$ , a un cuantil se le denota por  $c(p)$ , o  $c_p$ . En la literatura pueden encontrarse también los símbolos  $Q(p)$  o  $Q_p$ . La letra  $Q$  proviene del término en inglés *Quantile*.

En ocasiones conviene referirse a los cuantiles que dividen al conjunto de datos en ciertos porcentajes particulares. Tenemos, por ejemplo, los siguientes casos:

- Cuando  $p = 0.25, 0.50$  ó  $0.75$ , a los cuantiles correspondientes se le llama cuartiles, y se usan las expresiones: primer cuartil, segundo cuartil y tercer cuartil, respectivamente.
- Cuando  $p = 0.1, 0.2, \dots, 0.9$ , a los cuantiles correspondientes se les llama deciles. Podemos referirnos al primer decil de un conjunto de datos, al segundo decil, etcétera.
- En otras ocasiones se requiere dividir al conjunto de datos en cien porcentajes iguales, y entonces cuando  $p = 0.01, 0.02, \dots, 0.99$  a los cuantiles correspondientes se les llama percentiles.

En general, no es inmediato el cálculo de los cuantiles, pues debe verificarse con cuidado que se cumplen las dos condiciones que aparecen en la definición. En la sección que trata sobre la función de distribución empírica y que inicia en el página 76, veremos una forma gráfica para calcular los cuantiles de un conjunto de datos numéricos.

## Ejercicios

65. Explique la diferencia entre un cuantil y un cuartil.
66. Calcule el cuantil al 25 %, al 50 % y al 75 % del siguiente conjunto de datos.
- a)  $-2, 0, 1, 4$
- b)  $0, 2, 2, 3, 4$ .
- c)  $10, 50, 0, 30, 30, 20, 10, 0$ .
67. Calcule los deciles del siguiente conjunto de datos.

$0, 0, 0, 3, 5, 5, 6, 8, 8, 8, 9$ .

68. Indique en cada uno de los dos conjuntos de datos que aparecen en la Figure 1.10 los cuantiles al 20 %, 40 %, 60 % y 80 %.



Figura 1.10

## Coeficiente de asimetría (*Skewness*)

La cantidad que llamaremos coeficiente de asimetría (en inglés *skewness*) es una medida de la asimetría (falta de simetría) de un conjunto de datos numéricos  $x_1, \dots, x_n$ . Si  $\bar{x}$  es la media y  $s$  es la desviación estándar, entonces el coeficiente de asimetría se define como el siguiente número.

**Definición 1.31** El **coeficiente de asimetría** (*skewness*) de un conjunto de números  $x_1, \dots, x_n$  es la cantidad

$$\text{sk} = \frac{1}{s^3} \left( \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3 \right).$$

Recordemos que  $s^2$  denota la varianza, en consecuencia, el término  $s^3$  se calcula de la forma siguiente

$$s^3 = (s^2)^{3/2} = \left( \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{3/2}.$$

El coeficiente de asimetría no posee unidad de medición, es un número que puede ser positivo, negativo o cero. Su signo es positivo cuando la gráfica de frecuencias de los datos presenta una cola más alargada hacia la derecha de la media. Este tipo de comportamiento general se muestra en la gráfica derecha de la Figura 1.11 y es un indicativo de que existen datos a la derecha y alejados de la media de tal forma que las cantidades  $(x_i - \bar{x})^3$  son comparativamente grandes y con signo positivo.

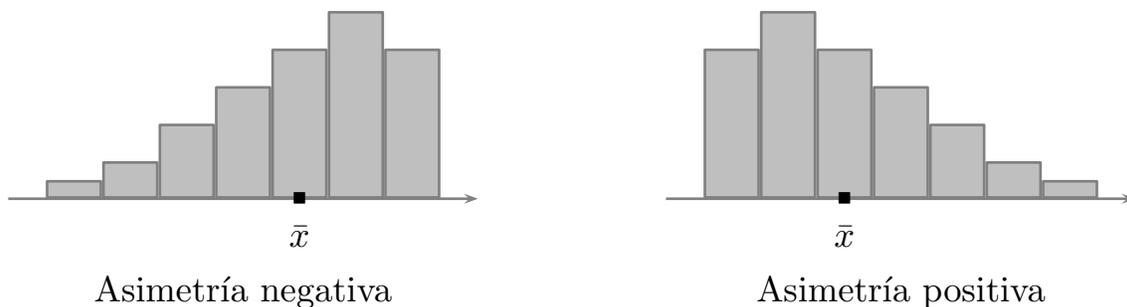


Figura 1.11

En cambio, el signo del coeficiente de asimetría es negativo cuando la gráfica de frecuencias presenta una cola más alargada hacia la izquierda de la

media. Este comportamiento se muestra en la parte izquierda de la Figura 1.11. En este caso existen datos a la izquierda y alejados de la media de tal forma que las cantidades  $(x_i - \bar{x})^3$  son grandes y con signo negativo. Por supuesto, se puede tener una gráfica de frecuencias sin presentar con claridad ninguno de estos dos tipos de comportamientos, pero el coeficiente de asimetría proporciona una cuantificación acerca de la tendencia global de los datos hacia alguno de estos dos posibles escenarios.

Puede comprobarse que, en el caso simétrico, es decir, cuando por cada dato  $x_i$  a la izquierda de  $\bar{x}$  hay otro dato a la derecha y a la misma distancia de este punto central, el coeficiente de asimetría es cero.

Es importante advertir que existen otras formas de definir un coeficiente de asimetría para un conjunto de datos o una distribución. A la definición que hemos visto se le conoce como coeficiente de asimetría de Fisher-Pearson, pero existen otras definiciones alternativas. En términos de los momentos centrales  $m_2$  y  $m_3$ , el coeficiente de asimetría que hemos definido se puede escribir de la siguiente forma

$$\text{sk} = \frac{m_3}{m_2^{3/2}}.$$

El siguiente resultado no es difícil de demostrar y muestra la forma en la que se modifica el coeficiente de asimetría bajo transformaciones lineales.

**Proposición 1.9** Sea  $\text{sk}(x)$  el coeficiente de asimetría del conjunto de datos numéricos  $x_1, \dots, x_n$ . Sean  $a \neq 0$  y  $c$  dos constantes, y considere los datos transformados  $ax_1 + c, \dots, ax_n + c$ . Entonces

$$\text{sk}(ax + c) = \frac{a}{|a|} \cdot \text{sk}(x).$$

## Ejercicios

69. Demuestre la Proposición 1.9.

70. Se dice que el conjunto de datos numéricos  $x_1, \dots, x_n$  es simétrico alrededor de su media  $\bar{x}$  si por cada dato  $x_i$  a la izquierda de  $\bar{x}$  hay otro dato a la derecha y a la misma distancia de este punto central. Demuestre que en esta situación el coeficiente de asimetría es cero.

## Curtosis

La curtosis es un número que denotaremos por la letra  $k$ , y se define de la siguiente manera.

**Definición 1.32** La **curtosis** de un conjunto de números  $x_1, \dots, x_n$  es la cantidad

$$k = \frac{1}{s^4} \left( \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4 \right).$$

Recordemos nuevamente que  $s^2$  denota la varianza, en consecuencia, el término  $s^4$  denota la varianza al cuadrado y se calcula de la siguiente forma

$$s^4 = (s^2)^2 = \left( \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^2.$$

La curtosis es un número positivo que no tiene una unidad de medición. Cuando una observación  $x_i$  dista mucho de la media  $\bar{x}$ , al elevar esta distancia a la cuarta potencia hace que se magnifiquen las distancias grandes. Por lo tanto, una curtosis grande puede indicar un mayor número de datos alejados de la media, hacia uno u otro lado, y por ello a la curtosis se le interpreta como una medida de la forma de las colas de la distribución o del conjunto de datos. Por la expresión “forma de las colas” nos referimos aquí a si éstas son amplias o bien ligeras (o inexistentes). Si son de una forma o de otra, esto afecta la forma de un posible pico que presente la frecuencia de los datos y de allí surgen interpretaciones de la curtosis como una medida del tipo de pico de los datos. Estas interpretaciones están sujetas a debate y por ahora no existe una interpretación aceptada de manera general.

Es claro que en términos de los momentos centrales, la curtosis puede escribirse de la siguiente manera  $k = m_4/m_2^2$ .

El siguiente resultado muestra que la curtosis es invariante bajo transformaciones lineales. Su demostración se deja como ejercicio.

**Proposición 1.10** Sea  $k(x)$  la curtosis del conjunto de datos numéricos  $x_1, \dots, x_n$ . Sean  $a \neq 0$  y  $c$  dos constantes, y considere los datos transformados  $ax_1 + c, \dots, ax_n + c$ . Entonces

$$k(ax + c) = k(x).$$

Se debe advertir que también se denomina con el nombre de curtosis (o *excess kurtosis*) a la cantidad que aparece a continuación. Debido a que la curtosis de la distribución normal estándar es igual a 3, con esta nueva definición, la curtosis de la distribución normal es ahora cero.

$$k_3 = \frac{1}{s^4} \left( \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4 \right) - 3.$$

De esta manera, se toma el tipo de cola de la distribución normal como punto de referencia y se adoptan los siguientes términos:

- **Leptocúrtica** ( $k_3 > 0$ ): Decaimiento rápido, colas ligeras. Este comportamiento se muestra en la Figura 1.12 (a).
- **Mesocúrtica** ( $k_3 = 0$ ): Curva normal. Este comportamiento se muestra en la Figura 1.12 (b).
- **Platicúrtica** ( $k_3 < 0$ ): Decaimiento lento, colas amplias. Este comportamiento se muestra en la Figura 1.12 (c).

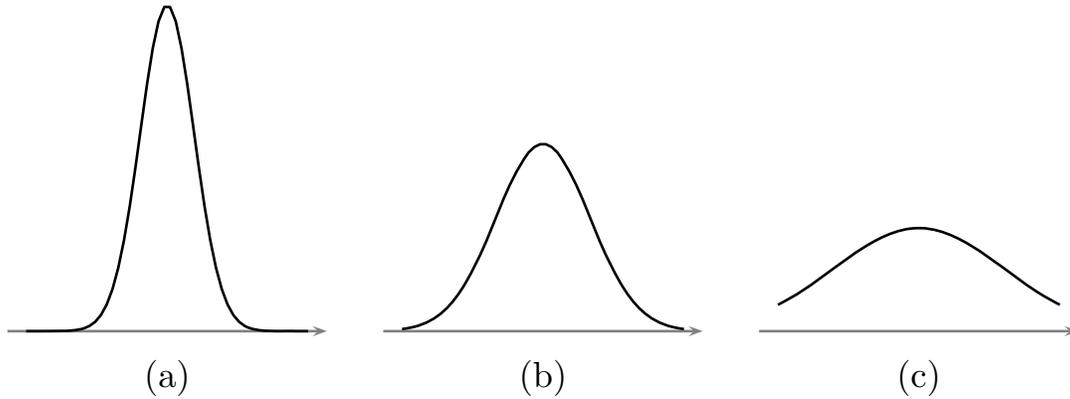


Figura 1.12: (a) Curva leptocúrtica, (b) mesocúrtica (normal) y (c) platicúrtica.

Debe considerarse que los tres tipos de comportamientos indicados son de tipo general y que, dependiendo del signo de la curtosis  $k_3$ , es que esta cantidad puede sugerir una tendencia de los datos hacia uno u otro tipo de comportamiento. Es claro que un conjunto de datos no necesariamente presenta uno de estos tres tipos de forma en su gráfica de frecuencias. El valor de la curtosis es únicamente una insinuación hacia alguno de estos tres tipos de comportamientos.

## Ejercicios

71. Demuestre la Proposición 1.10.
72. Sea  $X$  una variable aleatoria no constante y con cuarto momento finito. Se define la curtosis de  $X$  como el número  $k$  que aparece abajo. Demuestre que la curtosis de la distribución  $N(\mu, \sigma^2)$  es 3.

$$k = \frac{E(X - E(X))^4}{\text{Var}^2(X)}.$$

## Descripciones numéricas para datos agrupados

En ocasiones, la información disponible para una variable cuantitativa se encuentra agrupada en categorías o subconjuntos de valores de la variable. Más

específicamente, supongamos que en lugar de tener las observaciones o registros individuales  $x_1, \dots, x_n$ , tenemos agrupamientos de valores  $C_1, \dots, C_k$  junto con las frecuencias  $f_1, \dots, f_k$  que indican el número de veces que se observó cada agrupamiento. El problema es el siguiente: ¿cómo podemos calcular las descripciones numéricas como la media y la varianza en este caso? Existen por lo menos las siguientes dos soluciones:

- **Primera aproximación.** Se determina una marca de clase para cada categoría y se considera que la marca de clase se observó tantas veces como indica la frecuencia de la categoría. De esta manera, se construyen observaciones individuales aproximadas y se pueden aplicar ahora todas las definiciones y fórmulas antes vistas. En general, la elección de las marcas de clase no es inmediata y alguna argumentación razonable debe proveerse como parte del estudio estadístico.
- **Segunda aproximación.** Se escogen tantos valores numéricos dentro de una categoría como indica la frecuencia. Por ejemplo, pueden escogerse valores equiespaciados si esto es posible. Como antes, se procede a aplicar las fórmulas a la colección de valores numéricos así generados. En este caso, también es conveniente justificar la elección de los valores dentro de una categoría.

Debe enfatizarse que, en cualquiera de las dos perspectivas explicadas, la información producida es únicamente una aproximación, pues se ha perdido información al considerar agrupamientos de valores.

## Ejercicios

73. Calcule la media, la moda y la mediana del conjunto de datos agrupados que aparece en la siguiente tabla. Utilice alguna de las dos perspectivas de selección de la marca de clase.

Clase	Frecuencia
$C_1 = [0, 1)$	$f_1 = 2$
$C_2 = (1, 2]$	$f_2 = 0$
$C_3 = (2, 3]$	$f_3 = 1$
$C_4 = (3, 4]$	$f_4 = 4$

## RESUMEN DE FÓRMULAS

Descripciones numéricas de un conjunto de datos  $x_1, \dots, x_n$

<b>Media</b>	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
<b>Moda</b>	Dato con mayor frecuencia.
<b>Mediana</b>	Dato ordenado de en medio.
<b>Varianza</b>	$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$
<b>Desviación estándar</b>	$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$
<b>Desviación media</b>	$dm = \frac{1}{n} \sum_{i=1}^n  x_i - \bar{x} $
<b>Rango</b>	$r = x_{(n)} - x_{(1)}$
<b>Coefficiente de variación</b>	$cv = \frac{s}{\bar{x}}$
<b>Momentos</b>	$m'_k = \frac{1}{n} \sum_{i=1}^n x_i^k$
<b>Momentos centrales</b>	$m_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k$
<b>Cuantil al 100p%</b>	Al menos el 100p% de los datos son menores al cuantil y al menos 100(1 - p)% de los datos son mayores al cuantil.
<b>Asimetría</b>	$sk = \frac{1}{s^3} \left( \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3 \right)$
<b>Curtosis</b>	$k = \frac{1}{s^4} \left( \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4 \right)$

Tabla 1.6

### 1.3. Descripciones gráficas

Revisaremos ahora algunos elementos gráficos que pueden usarse para representar la información de un conjunto de datos. Estas gráficas puede elaborarse con ayuda de algún paquete de cómputo y tienen el objetivo de transmitir la información de una manera rápida, resumida y de fácil comprensión.

#### Gráfica de barras

Esta es una gráfica simple que consiste de varias barras que representan las categorías (o agrupamiento de valores) de una variable y sus frecuencias. En el eje horizontal se colocan las categorías, se dibuja una barra para cada categoría y la altura de la barra es la frecuencia o número de veces que se observa la categoría. El ancho de la barra no es relevante y puede no ser homogéneo para todas las categorías, en caso de que se decida agrupar algunas de ellas. La Figura 1.13 muestra un ejemplo de una gráfica de barras.

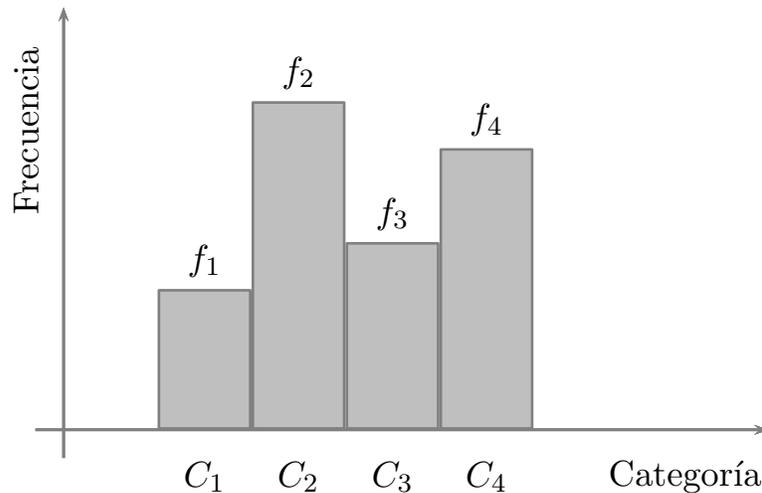


Figura 1.13

En este ejemplo gráfico se han colocado las barras de manera contigua y se han indicado las frecuencias absolutas  $f_i$  en la parte superior de cada

barra. De manera equivalente pueden utilizarse la frecuencia relativa o la frecuencia relativa porcentual para indicar la reiteración de una categoría. Para este tipo de gráficas, la variable puede ser cualitativa o cuantitativa, y las categorías o agrupamiento de valores pueden ser nominales u ordinales. En el caso de variable nominales, las clases o categorías se pueden colocar en cualquier orden. Las gráficas de barras pueden presentarse también en forma horizontal y tener otras variaciones, y es evidente que ayudan a la rápida comprensión de la información numérica. Permiten comparar visualmente las frecuencias de los distintos valores o categorías de una variable.

## Ejercicios

74. Investigue la siguiente información y elabore un histograma con los datos obtenidos.
- a) Las cinco religiones con el mayor número de adeptos en su país.
  - b) El número de veces que un grupo de personas ha visitado al dentista en el último año.
  - c) La composición de las familias: papá-mamá-con hijos, papá-mamá-sin hijos, etcétera.
  - d) Las cinco nacionalidades de mayor número de extranjeros que viven en su país.
  - e) La densidad poblacional de los diez países más poblados.
  - f) La extensión territorial de los diez países más extensos.
75. De un grupo de 50 personas, 18 son fumadoras y 32 no son fumadoras. Elabore un gráfica de barras con esta información.
76. Acerca de la religión que profesa un grupo de 100 personas, se obtuvo la información que se presenta en la siguiente tabla. Elabore un gráfica de barras con esta información.

Religión	
Valor	Frecuencia
Ninguno	28
Catolicismo	21
Cristianismo	19
Islam	15
Budismo	10
Otra	7

77. En la siguiente tabla se muestran los principales destinos de 73 personas que salen del país. Elabore un diagrama de barras con esta información.

Destino	Frecuencia
Canadá	12
Francia	15
Estados Unidos	25
Inglaterra	14
Alemania	7

78. En la tabla que aparece a continuación se muestran los 10 principales países productores de café en el mundo para el año 2013, según datos de la FAO. Elabore una gráfica de barras horizontal con esta información.

Principales países productores de café en 2013	
País	Producción (en toneladas)
Brasil	2,964,538
Vietnam	1,326,688
Indonesia	675,800
Colombia	653,160
Etiopía	392,006
India	318,200
Honduras	280,697
Perú	256,241
Guatemala	248,668
México	231,596

79. Investigue la densidad poblacional por continente y elabore una gráfica de barras con los datos encontrados.
80. Investigue la densidad poblacional de los cinco países más poblados y elabore una gráfica de barras con los datos encontrados.
81. Investigue la extensión territorial de los diez países más grandes en territorio y elabore una gráfica de barras horizontal con los datos encontrados.

## Histograma

Un histograma es una gráfica muy similar a la de barras. Adquiere este nombre cuando existe un orden entre los valores de la variable a graficar. Salvo esta condición, los datos puede ser cualitativos o cuantitativos. Nuevamente, para cada valor, categoría o clase de la variable, se asocia una barra cuya altura es la frecuencia con la que se observa la categoría. Como las categorías tienen un orden, se representan regularmente en el eje horizontal de menor a mayor. Como en las gráficas de barras, y para mayor información, en la parte superior de cada barra se puede colocar la frecuencia absoluta, la frecuencia relativa o la frecuencia porcentual.

A menudo se tiene una gran cantidad de datos numéricos y para elaborar un histograma con esta información se definen agrupaciones de valores, en este caso intervalos, y se calcula el número de datos que quedan en cada intervalo. A partir de estas gráficas se pueden sugerir modelos teóricos de probabilidad para la variable en estudio. Refinando o engrosando los intervalos de valores pueden obtenerse histogramas más claros y sugerentes. Por ejemplo, en la Figura 1.14 se muestra un histograma que claramente se asemeja a la conocida curva en forma de campana y sugiere, por lo tanto, que la variable en estudio puede adoptar el modelo normal o gaussiano.

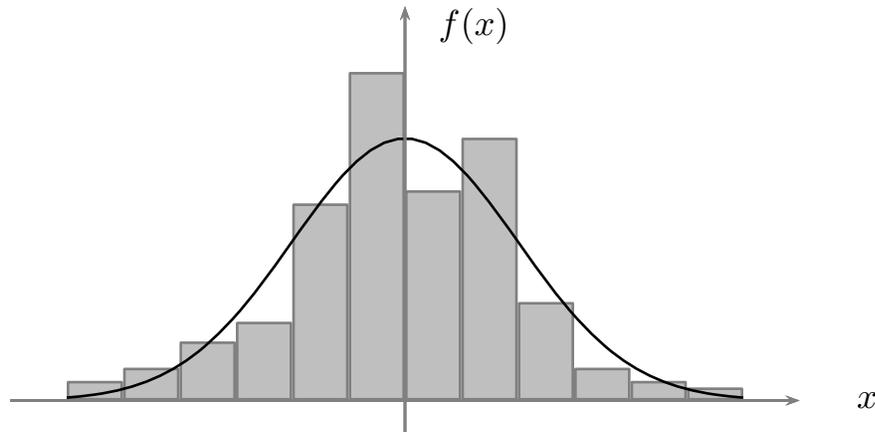


Figura 1.14

Se pueden elaborar también histogramas de frecuencias de ocurrencias de algún suceso a lo largo del tiempo. En este caso en el eje horizontal se colocan los intervalos de tiempo y en el eje vertical las frecuencias observadas.

## Ejercicios

82. ¿Cuál es la diferencia entre un histograma y una gráfica de barras?
83. En la siguiente tabla se muestra el número de desperfectos que tiene cada uno de 105 productos examinados. Elabore un histograma con esta información.

Números de desperfectos por producto	Frecuencia
0	82
1	15
2	5
3	2
4	1

84. De un grupo de 60 familias, se obtuvieron los datos que aparecen en la siguiente tabla acerca del número de automóviles por familia. Elabore un histograma con esta información.

Número de automóviles por familia					
Valor	0	1	2	3	4
Frecuencia	26	20	10	3	1

85. Se consultaron a 73 personas, y se les preguntó por el número de visitas promedio al dentista al año. Se obtuvieron los datos que aparecen en la siguiente tabla. Elabore un histograma con esta información.

Número de visitas al dentista							
Valor	0	1	2	3	4	5	6
Frecuencia	32	1	4	5	10	16	5

86. El número de días consecutivos en que un grupo de trabajadores no pudo asistir a laborar por enfermedad tiene la frecuencia que se muestra en la siguiente tabla. Elabore un histograma con esta información.

Número de días con falta al trabajo					
Valor	1	2	3	4	5
Frecuencia	23	12	5	3	1

## Polígono de frecuencias

Para construir un polígono de frecuencias se marcan los puntos medios en la parte superior de las barras de un histograma y se unen con líneas rectas. A la gráfica resultante se le llama polígono de frecuencias. En la Figura 1.15 se muestra un ejemplo de este tipo de gráficas.

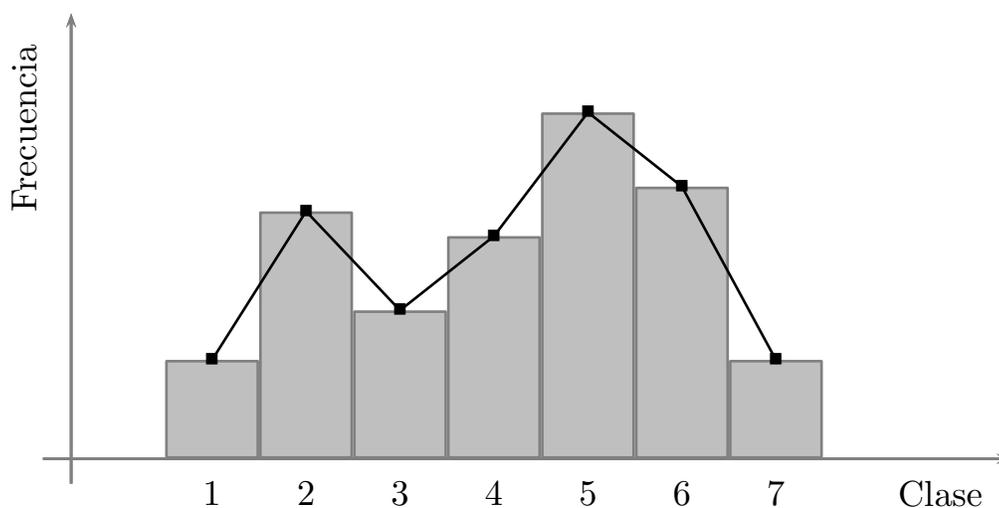


Figura 1.15

La información presentada mediante un polígono de frecuencias es equivalente a la información de un histograma; sin embargo, dado que se trata de líneas rectas, las tendencias de crecimiento y decrecimiento son más evidentes. Esta es una de las utilidades de este tipo de gráficas.

## Polígono de frecuencias acumuladas

Esta es una gráfica equivalente al histograma de frecuencias acumuladas. Para su construcción se marcan los puntos medios en la parte superior de las barras de un histograma de frecuencias acumuladas. Nuevamente, se unen

los puntos con líneas rectas. A la gráfica resultante se le llama polígono de frecuencias acumuladas. En la Figura 1.16 se muestra un ejemplo de este tipo de gráficas.

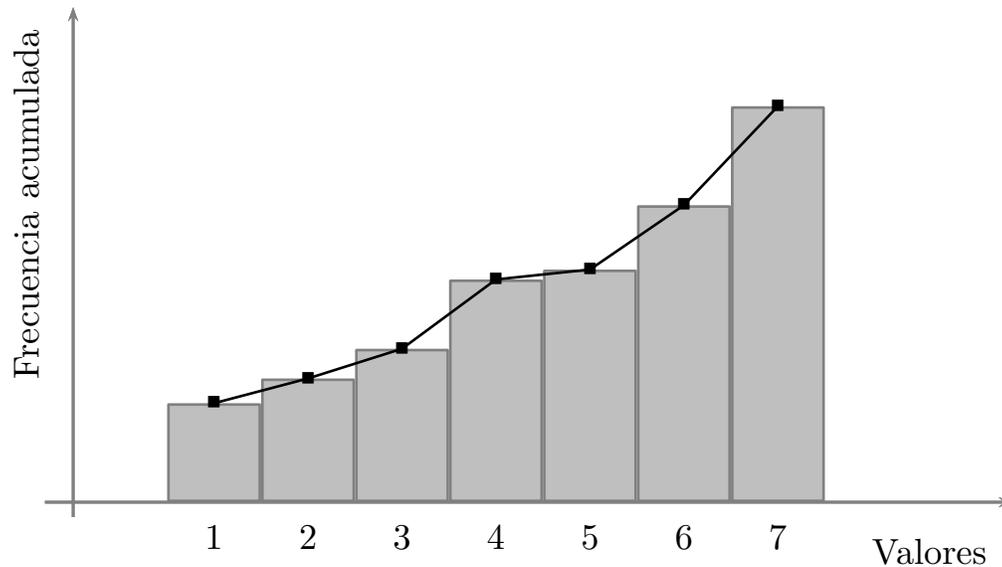


Figura 1.16

Es evidente que el comportamiento creciente de las frecuencias acumuladas es más claramente identificado en este tipo de gráficas. Para mayor información en la gráfica y, si resulta conveniente, se pueden colocar los valores numéricos de las frecuencias acumuladas arriba del punto marcado en cada barra.

## Ojiva

Una ojiva es una curva suave que se traza sobre los puntos de un polígono de frecuencias acumuladas. Se aplica para clases o agrupamientos de valores ordinales, y la curva resultante es más fácil de dibujar cuando el número de clases es grande. Una ojiva es una idealización del comportamiento creciente del polígono de frecuencias acumuladas. En la Figura 1.17 se muestra una de estas gráficas.

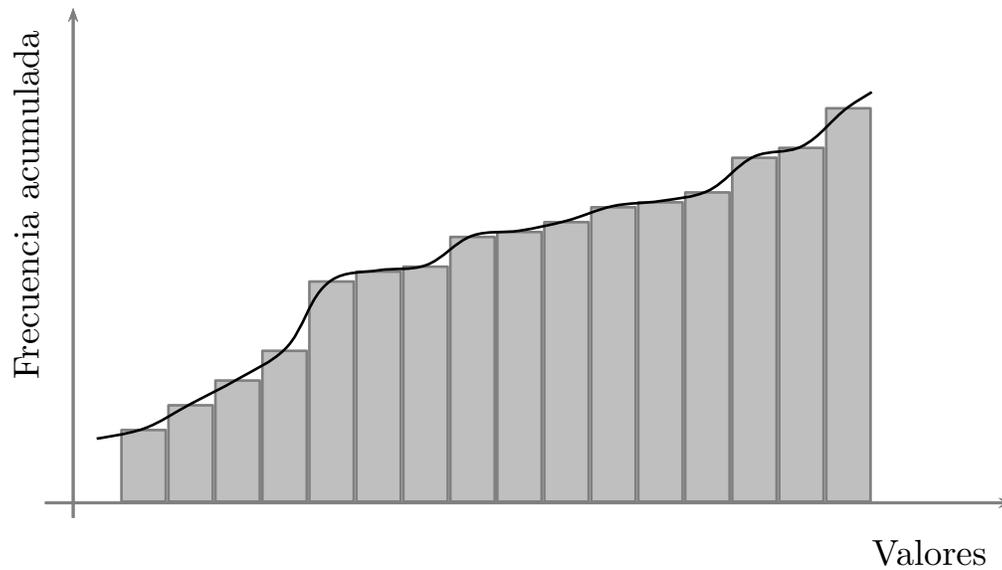


Figura 1.17

## Gráfica de pastel

Para variables cualitativas, o bien, para variables cuantitativas agrupadas, se pueden elaborar gráficas de pastel, también llamadas *pie charts*. Estas gráficas son círculos divididos en sectores que permiten comparar visualmente las frecuencias porcentuales de los valores observados de una variable.

La frecuencia de una categoría o grupo de valores se representa mediante un sector de un círculo, cuyo ángulo se determina de la siguiente forma: una frecuencia relativa, por ejemplo, de 0.2 (véase la Figura 1.18) se asocia con un sector con un ángulo de

$$(0.2) \times (360^\circ) = 72^\circ.$$

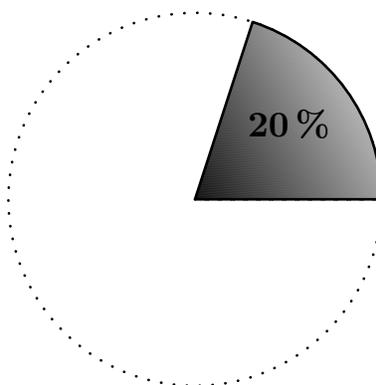


Figura 1.18

De esta manera el círculo completo se llena con los sectores calculados a partir de cada una de las frecuencias. Visualmente las gráficas de pastel son atractivas y logran muy bien el propósito de resumir la información en una gráfica. Se pueden dibujar gráficas de pastel en tercera dimensión y usar colores para hacerlas aún más sugestivas.

## Ejercicios

87. Elabore una gráfica de pastel para los datos que se muestran en la siguiente tabla para la variable número de hijos por familia. En total se consultaron a 120 familias.

Número de hijos por familia				
Valor	0	1	2	3
Frecuencia	24	78	12	6

88. Elabore un gráfica de pastel para los datos que se muestran a continuación relativos a la composición de un país por clase socioeconómica.

Clase socioeconómica			
Valor	Baja	Media	Alta
Porcentaje	50 %	35 %	15 %

89. Elabore un gráfica de pastel para los datos que se muestran en la siguiente tabla para la variable número de padres vivos de una persona, para un conjunto de 60 personas.

Número de padres vivos			
Valor	0	1	2
Frecuencia	5	10	45

90. Elabore una gráfica de pastel para los datos que se muestran a continuación de la variable número de goles anotados por un equipo de fútbol por partido jugado.

Número de goles anotados por partido							
Valor	0	1	2	3	4	5	6
Porcentaje	31 %	40 %	20 %	4 %	3 %	1 %	1 %

91. El número de campeonatos mundiales de fútbol ganados por país, hasta el año 2017, se muestra en la siguiente tabla. Elabore una gráfica de pastel con esta información. Si le es posible actualice la información a la fecha actual.

País	Campeonatos
Brasil	5
Alemania	4
Italia	4
Argentina	2
Uruguay	2
Francia	1
Inglaterra	1
España	1

## Gráficas de tallo y hojas

Esta es otra forma de representar un conjunto de datos numéricos de manera visual. Su aspecto es muy similar al de un histograma dibujado horizontalmente. Daremos varios ejemplos para ilustrar la construcción de este tipo de gráficas. Consideremos el siguiente conjunto de datos

126	102	84	100	67	89
73	124	113	91	92	96
112	70	82	95	121	126
72	84	87	92	107	100

A continuación se separa el dígito menos significativo del resto de los dígitos mediante una línea vertical, por ejemplo, el primer valor 126 se separa en 12|6. Se puede entonces conformar un diagrama como se muestra en la Figura 1.19, en donde se han ordenado los dígitos separados de menor a mayor, incluyendo repeticiones.

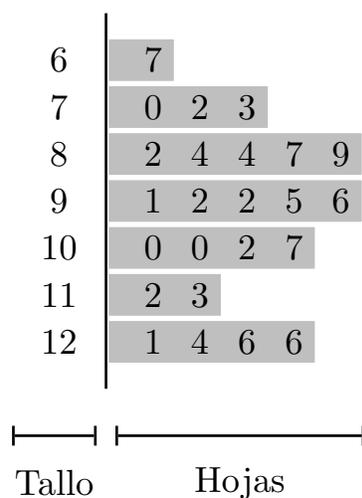


Figura 1.19

Este es un diagrama de tallo y hojas. A los dígitos separados y que aparecen en la parte derecha del diagrama se les llama hojas, y a la parte izquierda se le llama tallo. Si este diagrama se rota 90 grados en el sentido contrario al movimiento de las manecillas del reloj, se obtiene un diagrama similar al de un histograma. En general, los datos deben ser cercanos unos a otros para que los diagramas de tallo y hojas resultantes tengan una forma compacta.

### Variantes

- Si algún tallo tiene demasiadas hojas se pueden separar las hojas en varias partes. Por ejemplo, véase la Figura 1.20, en donde existen muchos datos entre los valores 70 y 80, y se han separado en dos grupos.

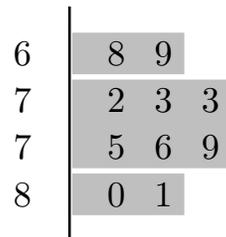


Figura 1.20

- Si resulta conveniente, los datos con muchos dígitos se pueden recortar. Por ejemplo, para el conjunto de números

2104	1757	1562	1756	1730
1992	1683	2133	2013	1684
1710	1881	1961	1672	1855

el primer dato 2104 se puede recortar a 210 y la separación es 21|0. Se elabora entonces el diagrama de la Figura 1.21 y se indica que la unidad de la hoja es 10. En este caso se pierde precisión de los datos originales, pero se gana simplicidad en la presentación.

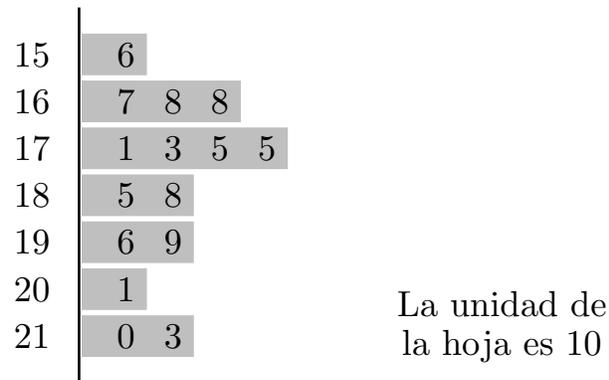


Figura 1.21

La unidad de la hoja indica el número por el que debe multiplicarse el dato graficado para obtener una aproximación del dato original. Por ejemplo, el primer dato graficado 15 | 6 en la Figura 1.21 corresponde a un valor aproximado de 1560. La unidad de las hojas puede ser 100, 10, 1, 0.1, 0.01, etcétera. Por ejemplo, si la unidad de la hoja es 0.1, el dato graficado 15 | 6 corresponde al valor 15.6.

## Ejercicios

92. Elabore un diagrama de tallo y hojas a partir del siguiente conjunto de datos e indique la unidad de la hoja.

a)

49 33 40 37 56 44 46 57 55 32  
 50 52 43 64 40 46 24 30 37 43  
 31 43 50 36 61 27 44 35 31 43  
 52 43 66 50 31 72 26 59 21 47

b)

1266 1354 1402 1107 1296 1389 1425  
 1087 1534 1200 1438 1024 1054 1190  
 1271 1342 1402 1055 1220 1372 1510  
 1124 1050 1199 1203 1355 1510 1426

c)

25.3	28.2	31.4	27.1	30.4	25.0
23.9	24.5	23.1	29.4	28.2	28.1
27.4	26.8	25.2	30.5	29.7	28.4
31.7	29.3	28.5	29.8	30.2	27.6

## Diagramas de caja y brazos

Esta es una forma gráfica de representar algunas características de un conjunto de datos numéricos. Esta representación está compuesta por una caja y por un par de marcas en dos extremos opuestos que asemejan brazos como se muestra en la Figura 1.22. A este tipo de gráficas se les conoce también como diagramas de caja y bigotes, y por los términos en inglés *boxplots* o *whiskers*. Para dibujar estos diagramas se necesita determinar cuatro elementos: el centro de la caja, su altura y los tamaños de los brazos superior e inferior. Explicaremos dos maneras en las que se pueden determinar estos parámetros.

Para el ejemplo mostrado en la Figura 1.22, el centro de la caja es la media  $\bar{x}$ . Se extiende la caja una desviación estándar  $s$  hacia arriba y otra desviación estándar  $s$  hacia abajo. La caja tiene, por lo tanto, una altura de  $2s$  unidades. La marca del brazo superior es igual al máximo valor observado, esto es,  $x_{(n)}$ . La marca del brazo inferior es el mínimo valor observado, es decir,  $x_{(1)}$ . En esta construcción, las longitudes de los brazos pueden ser distintas.

De esta manera, un diagrama de caja y brazos, construido de la forma indicada, es una forma de representar 4 descripciones numéricas de un conjunto de datos en un solo diagrama: el dato menor  $x_{(1)}$ , la media  $\bar{x}$ , la desviación estándar  $s$ , y el dato mayor  $x_{(n)}$ . Se pueden colocar dos o más de estos diagramas, uno junto al otro, a fin de comparar visualmente estas características en distintos conjunto de datos.

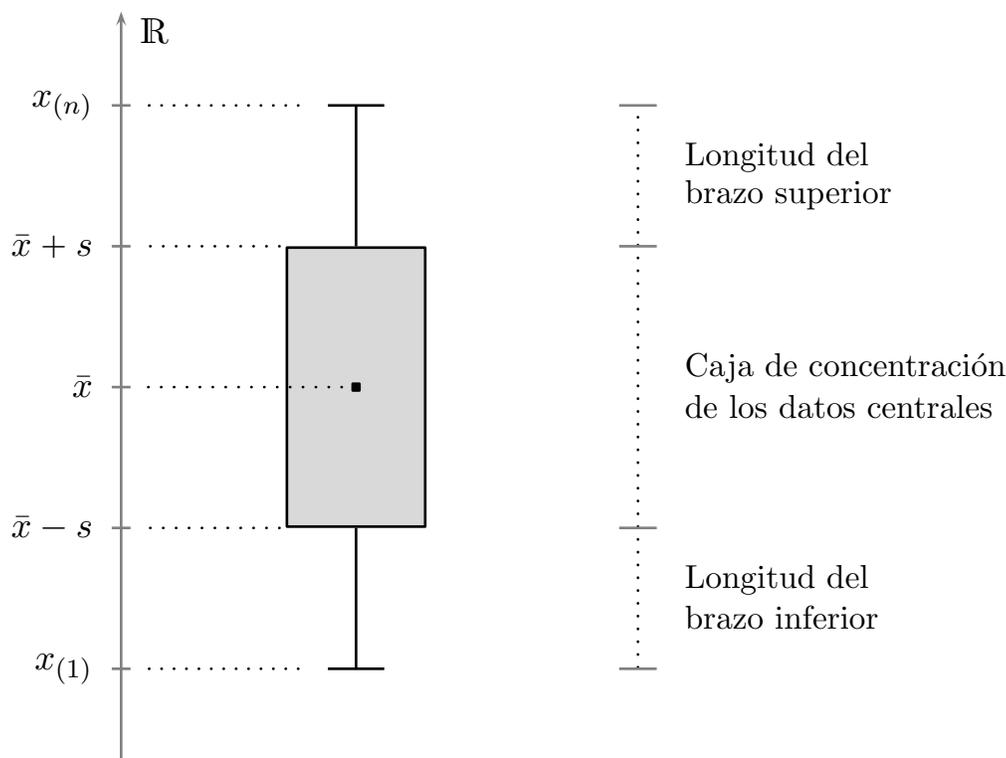


Figura 1.22

Otra manera de construir un diagrama de caja y brazos es a través de los cuantiles. La altura de la caja parte del primer cuartil  $Q_{0.25}$  y se extiende hasta el tercer cuartil  $Q_{0.75}$ . Observe que el segundo cuartil  $Q_{0.5}$ , es decir, la mediana, se encuentra dentro de la caja pero no necesariamente en el centro. La altura de la caja es entonces el así llamado rango intercuartil:  $\text{RIC} = Q_{0.75} - Q_{0.25}$ . El rango intercuartil mide la longitud del intervalo más pequeño que contiene el 50% de los datos centrales alrededor de la mediana. Por su nombre en inglés, el rango intercuartil también se denota por las letras  $\text{IQR}$ , *InterQuartile Range*. Las longitudes de los brazos se puede establecer como 1.5 veces el rango intercuartil  $\text{RIC}$ , y en este caso, los brazos tienen idéntica longitud. A los valores que se encuentren abajo de la marca del brazo inferior o arriba de la marca del brazo superior se les llama valores atípicos. A los valores que se encuentren arriba de  $Q_{0.75} + 3\text{RIC}$  o abajo de  $Q_{0.25} - 3\text{RIC}$  se les llama extremadamente atípicos (*outliers*). Las marcas de los brazos inferior y superior pueden ser los cuantiles al 10% y 90%, respectivamente, o bien los cuantiles al 5% y 95%.

## Ejercicios

93. Usando como parámetros la media, la desviación estándar y los valores máximo y mínimo, construya un diagrama de caja y brazos para el siguiente conjunto de datos.

a) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

b) -20, -1, -1, 0, 0, 1, 1, 15.

c) 2, 20, 4, 30, 5, 0, 10, 20.

94. Usando como parámetros los cuantiles y  $\pm(1.5)$  veces el rango intercuartil  $RIC=Q_{0.75} - Q_{0.25}$ , construya un diagrama de caja y brazos para el siguiente conjunto de datos.

a) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

b) -20, -1, -1, 0, 0, 1, 1, 15.

c) 2, 20, 4, 30, 5, 0, 10, 20.

## Función de distribución empírica

Esta función es otra manera gráfica de representar la información de una colección de observaciones numéricas. Su definición es la siguiente.

**Definición 1.33** La **Función de distribución empírica** de un conjunto de números  $x_1, \dots, x_n$  es la función  $F(x) : \mathbb{R} \rightarrow [0, 1]$  definida como sigue

$$F(x) = \frac{\#\{x_i : x_i \leq x\}}{n}.$$

Es decir, para cada número real  $x$  se debe contar el número de observaciones que son menores o iguales a  $x$  y dividir entre el número total de observaciones  $n$ . Esta es la razón por la que a la función  $F(x)$  se le conoce también como la función de distribución empírica acumulada.

Las gráficas de estas funciones tienen el aspecto de una escalera, presentan un escalón en cada observación  $x_i$  y en donde el tamaño del escalón es la frecuencia relativa del dato  $x_i$ . De esta manera, en la función de distribución empírica está representada toda la información de la colección de datos numéricos. Veamos un ejemplo.

**Ejemplo 1.2** Supongamos que tenemos las siguientes  $n = 5$  observaciones numéricas de una cierta variable de interés:

$$3, 1, 2, 4, 2.$$

Estos pocos valores son suficientes para ilustrar la construcción de una función de distribución empírica. Puede comprobarse que esta función es, para estos datos particulares,

$$F(x) = \begin{cases} 0 & \text{si } x < 1, \\ 1/5 & \text{si } 1 \leq x < 2, \\ 3/5 & \text{si } 2 \leq x < 3, \\ 4/5 & \text{si } 3 \leq x < 4, \\ 1 & \text{si } x \geq 4. \end{cases}$$

La gráfica de esta función se muestra en la Figura 1.23. Observe que, como el dato 2 aparece dos veces, el escalón allí es de magnitud  $2/5$ . Si todos los datos hubieran sido distintos, tendríamos una función de distribución empírica con cinco escalones de magnitud  $1/5$  cada uno.

Así, la función de distribución empírica inicia en el valor cero y se va incrementando mediante saltos hasta llegar al valor uno. En general, mientras mayor sea el número de datos observados, la función de distribución empírica toma un aspecto cada vez más parecido a una curva continua creciente. En una situación real, en donde se tenga una gran cantidad de datos, es necesario el uso de una computadora para graficar esta función.

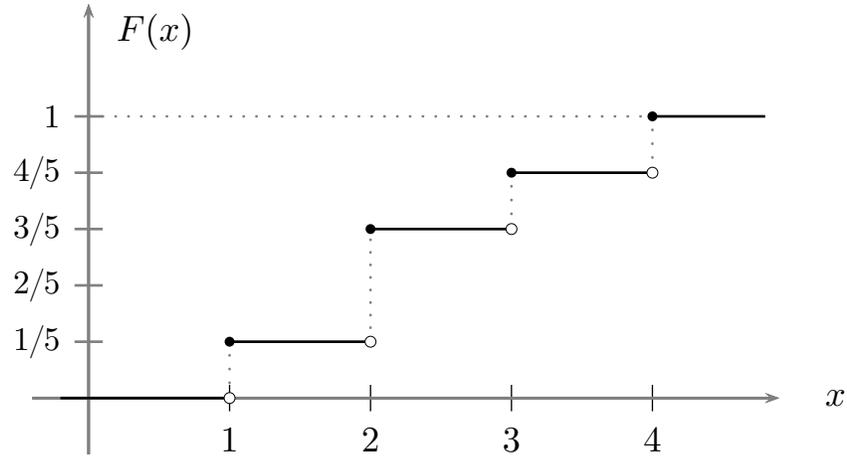


Figura 1.23

Puede verificarse que toda función de distribución empírica  $F(x)$  satisface las siguientes propiedades:

- $F(x) = 0$  para  $x < x_{(1)}$ .
- $F(x) = 1$  para  $x \geq x_{(n)}$ .
- $F(x)$  es creciente, esto es, si  $x \leq y$  entonces  $F(x) \leq F(y)$ .
- $F(x)$  es continua por la derecha.

La función de distribución empírica es importante dentro de la probabilidad y la estadística en general, puesto que, desde el punto de vista teórico, en ella está contenida toda la información obtenida de las observaciones de la variable de interés.

**Ejemplo 1.3** Como una aplicación de esta función, explicaremos una forma gráfica de calcular los cuantiles de una colección de datos a partir de

su función de distribución empírica. Primero, se marca el cuantil en el eje vertical. Después, se traza una línea horizontal hacia la derecha buscando la gráfica de  $F(x)$  (considerando la línea punteada vertical como parte de la gráfica) y al encontrarla se continúa con una línea vertical hacia abajo hasta alcanzar el eje horizontal. El valor  $x$  así encontrado es el cuantil correspondiente. Si el nivel buscado coincide con el piso de un escalón (esto ocurre, por ejemplo, en el caso del cuantil al 20% en el ejemplo anterior), la línea vertical hacia abajo se traza desde el punto central del escalón. ■

## Ejercicios

95. Encuentre la expresión de la función de distribución empírica del siguiente conjunto de datos. Grafique además esta función.

a) 2, 5.

d) 4, 10, 10, 4, 10, 4.

b) -1, 0, 1.

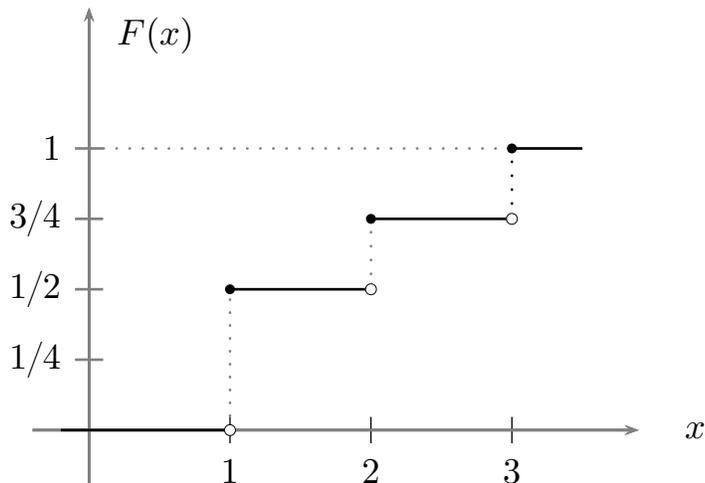
e) 7. (Un solo dato)

c) 2, 0, 0, 1, 5, 3.

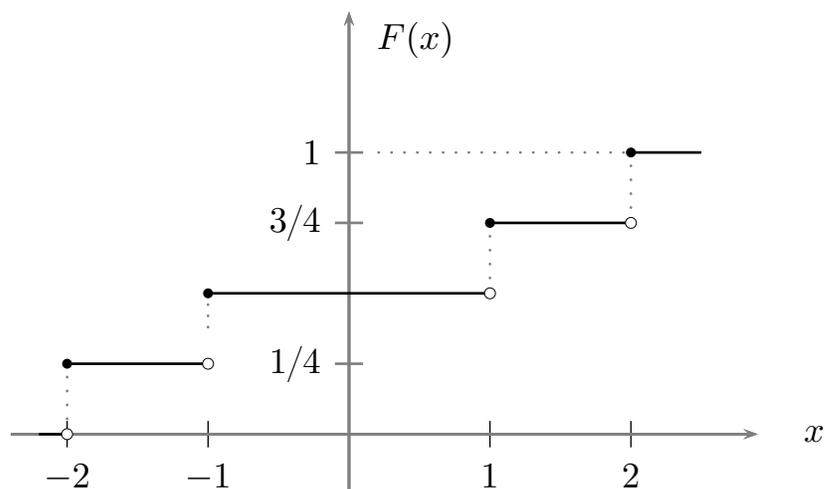
f) 25, 25, 25, 25.

96. Un cierto conjunto de datos  $x_1, \dots, x_n$  produce la función de distribución empírica  $F(x)$  que aparece en cada uno de los siguientes incisos. Encuentre explícitamente a este conjunto de datos y escriba la expresión analítica de la función  $F(x)$ .

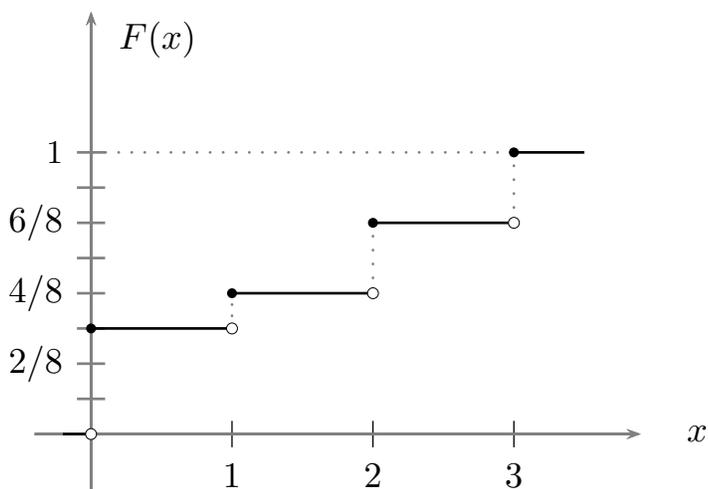
a)



b)



c)



97. Calcule los cuantiles al 20 %, 40 %, 60 % y 80 % del conjunto de datos resumido en la gráfica de la función de distribución empírica que se encuentra en la Figura 1.24.

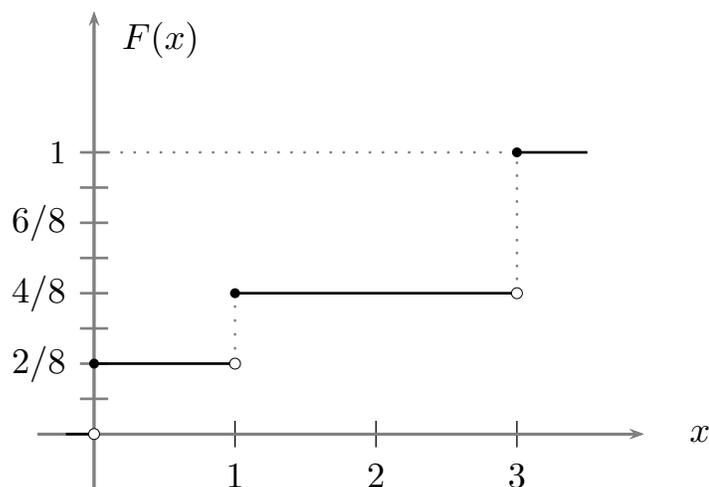


Figura 1.24

## 1.4. Variables aleatorias

En las secciones anteriores hemos considerado variables cualitativas y cuantitativas. En lo que resta de este trabajo consideraremos en su mayor parte variables cuantitativas. Pero no consideraremos valores numéricos observados  $x_1, \dots, x_n$ , sino variables aleatorias. Una variable aleatoria es una función  $X$  definida sobre una población y cuyos valores son números reales. Puede interpretarse esta función como una pregunta o una medición que se hace sobre cada elemento de la población. Uno puede pensar que se toma un elemento al azar de la población (aquí radica la aleatoriedad) y se efectúa la pregunta o medición produciendo una respuesta  $x$ . Debido al carácter aleatorio con el que fue escogido el elemento de la población es que se piensa que el valor  $x$  fue generado al azar y por ello la función  $X$  adquiere el nombre de variable aleatoria, pero vista como una función, no hay ninguna aleatoriedad en ella.

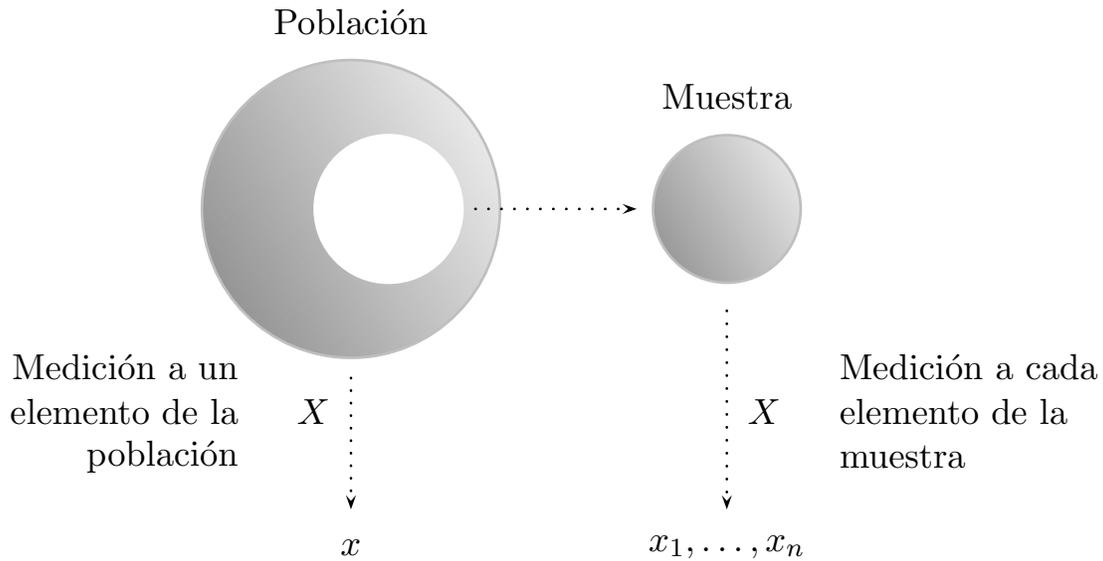


Figura 1.25

Las variables aleatorias representan las características de la población que deseamos conocer. Y cuando no es posible tener la información completa de la población es que se considera la idea de tomar sólo algunos elementos y hacer en ellos las mediciones. En este trabajo hemos supuesto tener una muestra (subconjunto) de la población y hacer la medición sobre estos elementos produciendo los resultados  $x_1, \dots, x_n$ . En la parte derecha de la Figura 1.25 se muestra esta situación.

Las descripciones numéricas para conjuntos de datos numéricos  $x_1, \dots, x_n$  se pueden extender a variables aleatorias  $X$ . En la Tabla 1.7 que aparece más adelante se muestran estas descripciones numéricas en el caso de variables aleatorias continuas. Estas cantidades están ahora expresadas en términos de los conceptos de probabilidad y esperanza. Es interesante comparar estas fórmulas con las que aparecen en la Tabla 1.6 de la página 84.

Cada variable aleatoria tiene asociada una función de distribución. Esta función se define como  $x \mapsto F(x) = P(X \leq x)$ . Se trata de la acumulación de la probabilidad hasta un valor  $x$  cualquiera, y esta expresión es análoga a la que aparece como función de distribución empírica para un conjunto de datos numéricos que hemos mencionado antes. Desde el punto de vista

matemático, la función de distribución es importante, pues contiene toda la información de la variable aleatoria: sus valores y sus probabilidades.

En los siguientes capítulos consideraremos variables aleatorias cuyas distribuciones de probabilidad dependen de un parámetro no especificado que denotaremos por la letra  $\theta$ . Los problemas que estudiaremos serán concernientes a la estimación del valor de este parámetro a la luz de un conjunto de observaciones de la variable aleatoria.

## RESUMEN DE FÓRMULAS

Descripciones numéricas para una variable aleatoria  $X$   
con función de densidad o de probabilidad  $f(x)$

<b>Media</b>	$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx$
<b>Moda</b>	Valor $x$ en donde $f(x)$ es máxima
<b>Mediana</b>	Valor $m$ tal que $P(X \leq m) \geq 1/2$ y $P(X \geq m) \geq 1/2$
<b>Varianza</b>	$\sigma^2 = E(X - \mu)^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$
<b>Desviación estándar</b>	$\sigma = \sqrt{E(X - \mu)^2}$
<b>Desviación media</b>	$E X - \mu  = \int_{-\infty}^{\infty}  x - \mu  f(x) dx$
<b>Rango</b>	Conjunto de valores de la v.a.
<b>Coefficiente de variación</b>	$\sigma/\mu$
<b>Momentos</b>	$\mu'_k = E(X^k) = \int_{-\infty}^{\infty} x^k f(x) dx$
<b>Momentos centrales</b>	$\mu_k = E(X - \mu)^k = \int_{-\infty}^{\infty} (x - \mu)^k f(x) dx$
<b>Cuantil al 100p%</b>	Valor $x$ tal que $P(X \leq x) \geq p$ y $P(X \geq x) \geq 1 - p$
<b>Asimetría</b>	$\mu_3/\sigma^3$
<b>Curtosis</b>	$\mu_4/\sigma^4$

Tabla 1.7

## Capítulo 2

# Estimación puntual

Sea  $X$  una variable aleatoria de interés en un experimento aleatorio, y supongamos que hemos aceptado que  $X$  tiene una función de densidad o de probabilidad conocida  $f(x, \theta)$ , que no está completamente especificada, pues depende de un parámetro desconocido denotado aquí por la letra  $\theta$  (teta). El problema que estudiaremos es el de estimar este parámetro, teniendo como información una serie de observaciones de la variable aleatoria. ¿Cómo se puede llevar a cabo esta estimación? El problema de estimación puntual consiste en encontrar una función de las observaciones, cuyo valor pueda usarse para estimar el parámetro desconocido. En este capítulo veremos algunos métodos para encontrar estimadores puntuales, así como algunas de sus propiedades.

### 2.1. Introducción

Consideremos que  $X$  es una variable aleatoria con función de densidad o de probabilidad conocida  $f(x, \theta)$ , pero dependiente de un parámetro desconocido  $\theta$ . De esta manera, se tiene toda una familia de distribuciones de probabilidad: una distribución para cada valor del parámetro  $\theta$ . Denotaremos por  $\Theta$  al conjunto de valores que puede tomar este parámetro y le llamaremos espacio parametral.

**Definición 2.1** Al conjunto de todos los posibles valores de un parámetro de una distribución de probabilidad se le llama **espacio parametral** y se le denota por la letra  $\Theta$  (teta mayúscula).

En realidad, el parámetro  $\theta$  puede ser una cantidad unidimensional, es decir, un solo parámetro, o bien un vector de dos o más parámetros  $\theta = (\theta_1, \theta_2, \dots)$ . Por otro lado, sabemos bien que existen distribuciones de probabilidad que no dependen de ningún parámetro; sin embargo, aquí estamos considerando la situación en donde por lo menos hay un parámetro involucrado y es desconocido.

Tenemos así la colección o familia parametral  $\{f(x, \theta) : \theta \in \Theta\}$  de funciones de densidad o de probabilidad, en donde la letra  $\theta$  es el nombre genérico que utilizaremos para denotar a un posible parámetro. Veamos algunos ejemplos.

- Para la distribución  $\text{Ber}(\theta)$ , el parámetro  $\theta$  toma valores en el espacio parametral  $\Theta = (0, 1)$ .
- Para la distribución  $\text{bin}(k, p)$ , el parámetro  $\theta$  es el vector de parámetros  $(k, p)$  y el espacio parametral es el producto cartesiano  $\Theta = \{1, 2, \dots\} \times (0, 1)$ .
- Para la distribución  $\text{N}(\mu, \sigma^2)$ , el parámetro  $\theta$  es el vector de parámetros  $(\mu, \sigma^2)$  y el espacio parametral es el conjunto  $\Theta = (-\infty, \infty) \times (0, \infty)$ , correspondiente a la mitad superior del plano cartesiano.

Supongamos ahora que  $x_1, \dots, x_n$  son observaciones independientes que se han obtenido de la variable aleatoria de interés. Es claro que estos valores observados pueden dar algún indicio del valor desconocido del parámetro  $\theta$ . El problema que se plantea es el siguiente: ¿cómo podemos usar estas observaciones para estimar el parámetro  $\theta$  para que de esta manera la función de densidad o de probabilidad  $f(x, \theta)$  quede completamente especificada? Ilustraremos la situación con algunos ejemplos dentro de un contexto práctico.

**Ejemplo 2.1** Se desea conocer la calidad de un lote de 1,000 artículos. Dada la imposibilidad o no conveniencia de someter a prueba a todos ellos, se escogen 20 artículos al azar obteniéndose los siguientes resultados.

Control de calidad de 20 artículos			
$x_1 = 0$	$x_6 = 1$	$x_{11} = 1$	$x_{16} = 1$
$x_2 = 1$	$x_7 = 0$	$x_{12} = 1$	$x_{17} = 1$
$x_3 = 1$	$x_8 = 1$	$x_{13} = 0$	$x_{18} = 1$
$x_4 = 0$	$x_9 = 0$	$x_{14} = 1$	$x_{19} = 1$
$x_5 = 1$	$x_{10} = 1$	$x_{15} = 1$	$x_{20} = 0$

El valor 0 indica que el artículo no pasó el control de calidad y el valor 1 indica que el artículo pasó el control de calidad. Supongamos que  $X$  es la variable que indica si un artículo escogido al azar de la población completa pasa, o no pasa, el control de calidad. Entonces es razonable suponer que  $X$  tiene una distribución  $\text{Ber}(\theta)$ , en donde no conocemos el valor del parámetro  $\theta$ . ¿Cómo podemos estimar el valor de  $\theta$  con base en los datos de la muestra? Al especificar por completo a la distribución Bernoulli en este problema, podemos tener una mejor idea de la cantidad de artículos defectuosos en el lote completo. ■

**Ejemplo 2.2** El tiempo en minutos que un conjunto de 10 personas, escogidas al azar, invierte en trasladarse de la casa al lugar de trabajo, o a la escuela, se muestra en la colección de números que aparece abajo.

Tiempo en minutos	
$x_1 = 100$	$x_6 = 60$
$x_2 = 25$	$x_7 = 75$
$x_3 = 135$	$x_8 = 40$
$x_4 = 120$	$x_9 = 35$
$x_5 = 25$	$x_{10} = 130$

Supongamos que tal variable puede modelarse mediante la distribución  $\exp(\theta)$ , pero no conocemos el valor de  $\theta$ . ¿Cómo podemos estimar el valor de  $\theta$  con base en las observaciones obtenidas? Si se logra especificar completamente a esta distribución exponencial, podemos estimar la cantidad de personas que, para su traslado, ocupan un tiempo dentro de un rango de valores especificado. ■

De esta manera, habiendo supuesto una distribución de probabilidad para una variable aleatoria de interés, en donde la distribución depende de un parámetro no especificado en su valor, el problema consiste en encontrar un mecanismo para estimar el parámetro desconocido tomando como información una serie de observaciones de la variable aleatoria.

En el tratamiento que seguiremos no vamos a considerar observaciones particulares  $x_1, \dots, x_n$ , sino observaciones aleatorias. Escribiremos entonces a éstas como la colección de variables aleatorias  $X_1, \dots, X_n$ , e impondremos dos condiciones fuertes sobre ellas: independencia e idéntica distribución. A esta colección se le llama muestra aleatoria, lo que se abrevia usando las letras iniciales m.a.

**Definición 2.2** Una **muestra aleatoria** es una colección de variables aleatorias  $X_1, \dots, X_n$  que son independientes e idénticamente distribuidas.

Las dos hipótesis mencionadas son características ideales de  $n$  observaciones de la variable aleatoria y que no necesariamente se cumplen en una situación real, pero facilitan considerablemente el análisis probabilístico de los modelos. Sobre la independencia, tenemos que un valor observado para una de las variables no influye o afecta la distribución de probabilidad de cualquier otra variable, siendo esta distribución la misma para obtener cada una de las observaciones. Esto último se refiere a la idéntica distribución. Supondremos, entonces, que todas las variables de una muestra aleatoria tienen la misma función de densidad o de probabilidad  $f(x, \theta)$ .

En particular, la primera observación  $x_1$  puede ser un valor de  $X_1$ , la segunda observación  $x_2$  puede ser un valor de  $X_2$ , etcétera. Así, las variables aleatorias  $X_1, \dots, X_n$  representan  $n$  observaciones al azar e independientes de la variable aleatoria en estudio. Al número entero  $n \geq 1$  se le llama tamaño de la muestra aleatoria y, a menos que se especifique lo contrario, supondremos que este entero es conocido.

Los estimadores que busquemos serán funciones de una muestra aleatoria y a tales funciones les llamaremos estadísticas. Precisamos esta definición a continuación.

**Definición 2.3** Una **estadística** es una función de una muestra aleatoria que no depende de parámetros desconocidos.

Denotaremos por  $T$ , o más explícitamente por  $T(X_1, \dots, X_n)$ , a una de estas funciones de la muestra aleatoria. En nuestro estudio, consideraremos que esta función es una variable aleatoria y que tiene como un posible valor el número  $T(x_1, \dots, x_n)$ . Debe hacerse énfasis en que la expresión mediante la cual se define una estadística no debe depender de parámetros desconocidos, únicamente de las variables de la muestra aleatoria y del tamaño de ésta, pues, justamente, sus valores serán usados como estimaciones para el parámetro desconocido y éstos deben poder determinarse únicamente a través de las variables de la muestra aleatoria. Sin embargo, ocurrirá que la distribución de probabilidad de una estadística dependerá, en general, del parámetro desconocido  $\theta$ .

El concepto de estadística que acabamos de definir es importante. La razón de ello es que nuestros estimadores serán objetos de este tipo. Nos interesará conocer las características y la distribución de probabilidad de estas variables aleatorias, aunque sólo en algunos pocos casos podremos determinar completamente la distribución de una estadística.

Veremos a continuación algunos ejemplos de estadísticas. Algunas de ellas tienen nombre y notación particular por su uso frecuente.

**Ejemplo 2.3** A la estadística denotada por  $\bar{X}$  (se lee  $x$  barra) y que se define a continuación se le llama media muestral. Esta variable aleatoria es simplemente el promedio aritmético de los elementos de la muestra aleatoria, es decir,

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

Si  $x_1, \dots, x_n$  son valores particulares de las variables de la muestra aleatoria, entonces la media muestral es el número  $\bar{x}$  definido antes,

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

Observe el uso de mayúsculas y minúsculas. La estadística  $\bar{X}$  es una variable aleatoria mientras que  $\bar{x}$  es un número real. ■

**Ejemplo 2.4** La siguiente función de una muestra aleatoria es una estadística y se le conoce con el nombre de varianza muestral.

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

Observe que en este promedio aparece el término  $n-1$  en el denominador y no el número de sumandos  $n$ . Más adelante justificaremos esta elección. Si  $x_1, \dots, x_n$  son valores particulares de las variables de la muestra aleatoria, entonces el valor de la varianza muestral es el número

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

■

**Ejemplo 2.5** Sea  $k$  un entero tal que  $1 \leq k \leq n$ . La  $k$ -ésima estadística de orden de una muestra aleatoria de tamaño  $n$  es una variable aleatoria definida de la siguiente forma

$$X_{(k)} = k\text{-ésimo mín } \{X_1, \dots, X_n\}.$$

Esto es,  $X_{(1)}$  es la primera estadística de orden, o bien, puntualmente,  $X_{(1)}(\omega) = \text{mín } \{X_1(\omega), \dots, X_n(\omega)\}$ ,  $X_{(2)}$  es la segunda estadística de orden, etcétera. Se debe observar que estas variables aleatorias no son necesariamente alguna de las variables de la muestra aleatoria, sino que son funciones de todas ellas en la forma indicada arriba. Además, las estadísticas de orden no son independientes pues guardan siempre el orden ascendente  $X_{(1)} \leq \dots \leq X_{(n)}$ . Para denotar a la  $k$ -ésima estadística de orden también se usa el símbolo  $X_{k:n}$ . La ventaja de esta expresión alternativa es que se especifica el tamaño  $n$  de la muestra aleatoria. ■

**Ejemplo 2.6** Sea  $k \geq 1$  un entero. A la estadística que aparece abajo se le conoce con el nombre de  $k$ -ésimo momento muestral. Se trata del promedio aritmético de las variables aleatorias de la muestra elevadas a la potencia  $k$ . Cuando  $k = 1$ , esta estadística se reduce a la media muestral.

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

■

Para mayor claridad, veremos ahora algunos ejemplos de funciones de una muestra aleatoria que no son estadísticas.

**Ejemplo 2.7** Sea  $X_1, \dots, X_n$  una m.a. de la distribución Poisson( $\theta$ ), en donde el parámetro  $\theta > 0$  es desconocido. La variable aleatoria  $T = \theta^{X_1 + \dots + X_n}$  no es una estadística puesto que en su definición aparece el parámetro desconocido  $\theta$ .

■

**Ejemplo 2.8** Sea  $X_1, \dots, X_n$  una m.a. de la distribución  $N(\mu, \sigma^2)$ , en donde los parámetros  $\mu$  y  $\sigma^2$  son desconocidos. La variable aleatoria  $T = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu)/\sigma$  no es una estadística puesto que en su definición aparecen los parámetros desconocidos  $\mu$  y  $\sigma^2$ . Sin embargo, puede demostrarse que la distribución de  $T$  no depende de ningún parámetro desconocido, se trata de la distribución normal estándar.

■

Cuando alguna estadística se proponga o se construya con el objetivo de servir como estimador para un parámetro desconocido  $\theta$  se le denotará, de manera sugerente, por  $\hat{\theta}$ , y se le llamará un estimador. El símbolo  $\hat{\theta}$  se lee “teta circunflejo”. Aquí tenemos la definición.

**Definición 2.4** Un **estimador puntual** para un parámetro desconocido  $\theta$  es una estadística denotada por  $\hat{\theta}$  que se propone para estimar el parámetro.

Observemos que si  $x_1, \dots, x_n$  son valores particulares de las variables de la muestra aleatoria, entonces el número  $\hat{\theta}(x_1, \dots, x_n)$  es una estimación de  $\theta$ , mientras que la variable aleatoria  $\hat{\theta}(X_1, \dots, X_n)$  es un estimador para  $\theta$ . Si se omiten los argumentos, ambos objetos se escriben simplemente como  $\hat{\theta}$ , y puede representar, tal vez con un poco de confusión, tanto una estimación (un número) como un estimador (una variable aleatoria). El contexto y la forma de tratar a  $\hat{\theta}$  determinará si nos referimos a la estimación o al estimador.

Como un ejemplo de estimador tenemos que la media muestral  $\hat{\theta} = \bar{X}$  puede ser usada para estimar el parámetro desconocido  $\theta$  en la distribución  $\text{Ber}(\theta)$ , pues este promedio indica la proporción de valores 1 en el total de la muestra aleatoria. Sin embargo, no es clara la forma de proponer estimadores para el parámetro o parámetros desconocidos de una distribución cualquiera. Surge así el problema de encontrar mecanismos para generar estadísticas que puedan servir como estimadores para los parámetros desconocidos de las distintas distribuciones de probabilidad. ¿Cómo encontrar posibles estimadores para un parámetro desconocido  $\theta$ ? En las siguientes secciones veremos algunos métodos generales para encontrar explícitamente estadísticas que puedan usarse como estimadores para parámetros desconocidos.

## Ejercicios

98. Determine el espacio parametral de las siguientes distribuciones.

- |                          |                                  |
|--------------------------|----------------------------------|
| a) Poisson( $\lambda$ ). | d) gama( $\alpha, \lambda$ ).    |
| b) bin neg( $r, p$ ).    | e) Weibull( $\alpha, \lambda$ ). |
| c) unif( $a, b$ ).       | f) beta( $a, b$ ).               |

99. Conteste las siguientes preguntas.

- ¿Cuál es la diferencia entre un estimador y una estadística?
- ¿Es cierto que toda estadística es un estimador?
- ¿Es cierto que todo estimador es una estadística?

- d) ¿Es cierto que toda estadística es una variable aleatoria?  
 e) ¿Es cierto que toda función de una m.a. es una estadística?  
 f) ¿Es cierto que toda función de una estadística es una estadística?

100. Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución cualquiera. Demuestre las siguientes identidades.

$$a) \sum_{i=1}^n (X_i - \bar{X}) = 0.$$

$$b) \sum_{i=1}^n (X_i - \bar{X})^2 = \left( \sum_{i=1}^n X_i^2 \right) - n\bar{X}^2.$$

$$c) S^2 = \frac{1}{n-1} \left[ \sum_{i=1}^n X_i^2 - \frac{1}{n} \left( \sum_{i=1}^n X_i \right)^2 \right].$$

101. Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución  $f(x, \theta)$ , dependiente de un parámetro desconocido  $\theta$ . Determine si las siguientes variables aleatorias son estadísticas.

$$a) T = X_1.$$

$$h) T = (X_1 + \dots + X_n)^2.$$

$$b) T = (X_1 + X_n)/2.$$

$$i) T = \exp \{X_1 + \dots + X_n\}.$$

$$c) T = X_1 + 2X_2 + \dots + nX_n.$$

$$j) T = \frac{1}{n} \sum_{i=1}^n \frac{X_i - E(X_i)}{\sqrt{\text{Var}(X_i)}}.$$

$$d) T = 1_{(\theta, \infty)}(X_1).$$

$$e) T = (X_1 + \dots + X_n) - \theta.$$

$$k) T = \frac{1}{n} \sum_{i=1}^n \frac{X_i - \bar{X}}{\sqrt{S^2}}.$$

$$f) T = \theta \cdot (X_{(n)} - X_{(1)}).$$

$$g) T = X_1^2 + \dots + X_n^2.$$

$$l) T = (X_1 \dots X_n)^{1/n}.$$

102. Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución uniforme en el conjunto discreto  $\{a_1, \dots, a_m\}$ , en donde los valores  $a_1, \dots, a_m$  y  $m$  son desconocidos. A partir de alguna argumentación intuitiva, proponga un estimador para cada uno de los siguientes parámetros.

a)  $a_1$ .c)  $m$ .b)  $a_m$ .d)  $a_m - a_1$ .

103. Sean  $X_1, \dots, X_n$  una muestra aleatoria de una distribución cualquiera. Demuestre que para cualquier estadística  $T$ ,

$$\sum_{i=1}^n (X_i - \bar{X})^2 \leq \sum_{i=1}^n (X_i - T)^2.$$

104. Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $N(\mu, \sigma^2)$ . Encuentre la distribución de la variable  $Y_i = X_i - \bar{X}$ , para cada  $i = 1, \dots, n$ .

## 2.2. Método de momentos

Este método para encontrar estimadores fue introducido por Karl Pearson<sup>1</sup> a principios del siglo  $XX$ . Consideremos nuevamente que  $f(x, \theta)$  es la función de densidad o de probabilidad de una variable aleatoria  $X$  que depende de un parámetro desconocido  $\theta$ . El método de momentos nos provee de un mecanismo general para estimar  $\theta$ , y para explicarlo necesitamos recordar antes dos conceptos.

**Definición 2.5** Sea  $k \geq 1$  un entero. El  $k$ -ésimo **momento** de una variable aleatoria  $X$ , si existe, es el número  $E(X^k)$ .

A los números  $E(X), E(X^2), E(X^3), \dots$  se les llama también momentos poblacionales. En general, en las expresiones de estas cantidades aparece el parámetro o vector de parámetros  $\theta$ , los cuales son de nuestro interés. Por otro lado, supongamos que  $X_1, \dots, X_n$  es una muestra aleatoria de la distribución en estudio. Tenemos la siguiente definición de otros tipos de momentos.

<sup>1</sup>Karl Pearson (né Carl Pearson, 1857-1936), estadístico inglés.

**Definición 2.6** Sea  $k \geq 1$  un entero. El  $k$ -ésimo **momento** de una muestra aleatoria  $X_1, \dots, X_n$  es la variable aleatoria  $\frac{1}{n} \sum_{i=1}^n X_i^k$ .

A estas variables aleatorias se les llama momentos muestrales. En particular, el primer momento muestral es la media muestral  $\bar{X}$ . Ahora podemos enunciar el método de momentos.

**¿En qué consiste el método de momentos?** Consiste en igualar los momentos muestrales con los correspondientes momentos poblacionales y resolver esta ecuación, o sistema de ecuaciones, para el parámetro o vector de parámetros, cuando ello sea posible.

Se igualan tantos momentos como parámetros haya que estimar, suponiendo que suficientes momentos poblacionales existen para la distribución en cuestión y que son distintos de cero. El método de momentos es muy sencillo de aplicar y lo ilustraremos a continuación con algunos ejemplos.

**Ejemplo 2.9** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{Ber}(\theta)$ , en donde  $\theta$  es desconocido. La estimación del parámetro  $\theta$  por el método de momentos consiste en igualar el primer momento de la distribución, que es  $\theta$ , con el primer momento muestral, que es  $\bar{X}$ . Esta igualación produce directamente la identidad

$$\hat{\theta} = \bar{X}.$$

Observe que cuando se ha hecho la igualación ya no se escribe  $\theta$ , sino  $\hat{\theta}$ , pues resolver la ecuación para este término produce el estimador por el método de momentos. De esta manera, si  $x_1, \dots, x_n$  son los valores de las observaciones, entonces el promedio  $\bar{x} = (x_1 + \dots + x_n)/n$  es la estimación para  $\theta$  por el método de momentos. ■

**Ejemplo 2.10** Sea  $X$  una variable aleatoria continua con función de densidad

$$f(x, \theta) = \begin{cases} \theta x^{\theta-1} & \text{si } 0 < x < 1, \\ 0 & \text{en otro caso,} \end{cases}$$

en donde  $\theta > 0$  es un parámetro desconocido. Supongamos que contamos con una muestra aleatoria  $X_1, \dots, X_n$  de esta distribución. Puede comprobarse, sin mucha dificultad, que  $E(X) = \theta/(1 + \theta)$ . La igualación de esta esperanza con la media muestral produce la ecuación  $\hat{\theta}/(1 + \hat{\theta}) = \bar{X}$ . Observe nuevamente que al escribir esta identidad hemos puesto  $\hat{\theta}$  en lugar de  $\theta$ . Resolviendo para  $\hat{\theta}$  se obtiene el estimador

$$\hat{\theta} = \frac{\bar{X}}{1 - \bar{X}}.$$

Si  $x_1, \dots, x_n$  son los valores numéricos observados, entonces  $\hat{\theta} = \bar{x}/(1 + \bar{x})$  es el valor estimado para  $\theta$  por el método de momentos. ■

En los ejemplos anteriores sólo ha habido un parámetro por estimar. En el siguiente ejemplo consideraremos un caso importante en donde es necesario estimar dos parámetros.

**Ejemplo 2.11** Encontraremos estimadores para los parámetros  $\mu$  y  $\sigma^2$  de una distribución normal mediante el método de momentos. Como se necesitan estimar dos parámetros, se usan los dos primeros momentos. El primer y segundo momentos poblacionales son  $E(X) = \mu$  y  $E(X^2) = \sigma^2 + \mu^2$ . La igualación respectiva de estas cantidades con los dos primeros momentos muestrales produce el sistema de ecuaciones

$$\begin{aligned} \hat{\mu} &= \bar{X}, \\ \hat{\sigma}^2 + \hat{\mu}^2 &= \frac{1}{n} \sum_{i=1}^n X_i^2. \end{aligned}$$

Al hacer la igualación entre los momentos hemos escrito  $\hat{\mu}$  en lugar de  $\mu$  y  $\hat{\sigma}^2$  en lugar de  $\sigma^2$ . Se trata ahora de resolver este sistema de ecuaciones

para  $\hat{\mu}$  y  $\hat{\sigma}^2$ . La primera ecuación es explícita, mientras que la segunda se puede reescribir como sigue

$$\begin{aligned}\hat{\sigma}^2 &= \left( \frac{1}{n} \sum_{i=1}^n X_i^2 \right) - \bar{X}^2 \\ &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \\ &= \frac{n-1}{n} S^2.\end{aligned}$$

La segunda igualdad no es inmediata, pero sólo se requiere llevar a cabo algunas operaciones algebraicas sencillas para obtenerla. De esta manera hemos obtenido estimadores por el método de momentos para los dos parámetros de la distribución normal. Si  $x_1, \dots, x_n$  son las observaciones obtenidas, entonces las estimaciones, por el método de momentos, son

$$\begin{aligned}\hat{\mu} &= \frac{1}{n} \sum_{i=1}^n x_i, \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.\end{aligned}$$

■

En el siguiente ejemplo se muestran algunos problemas técnicos que pueden surgir al aplicar el método de momentos.

**Ejemplo 2.12** Sea  $X$  una variable aleatoria continua con función de densidad  $\text{unif}(-\theta, \theta)$ , en donde  $\theta > 0$  es un parámetro desconocido. Aplicar el método de momentos para encontrar un estimador para  $\theta$  requiere conocer el primer momento de esta distribución. Puede comprobarse que el primer momento es nulo, de modo que la igualación del primer momento poblacional y el primer momento muestral no produce una ecuación útil de la cual puede obtenerse un estimador para  $\theta$ , a saber,  $0 = \bar{X}$ . Se propone entonces igualar los segundos momentos. Como  $E(X^2) = \theta^2/3$ , se obtiene la ecuación

$$\frac{1}{3} \hat{\theta}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2,$$

de donde se obtiene el estimador

$$\hat{\theta} = \sqrt{\frac{3}{n} \sum_{i=1}^n X_i^2}.$$

■

Mostrados ya algunos ejemplos del método de momentos para estimar parámetros, haremos ahora algunas observaciones generales que es bueno tener presente cuando se haga uso de este método.

- **Aplicación.** El método de momentos puede aplicarse sin distinción alguna tanto para distribuciones discretas como continuas.
- **Uso de los momentos.** La idea fundamental del método hace uso del hecho de que, bajo ciertas condiciones, la sucesión de momentos  $E(X), E(X^2), \dots$  determina de manera única a la distribución de probabilidad. En el método sólo se usan los primeros pocos momentos (los necesarios para estimar  $\theta$  y de esta manera determinar completamente a la distribución, pues estamos suponiendo que se conoce su forma). Observemos que, en general, en las expresiones de estos momentos aparece el parámetro  $\theta$ . Por otro lado, la igualdad de estos momentos con los momentos muestrales no es extraña, pues por la ley de los grandes números, cuando el tamaño de muestra  $n$  es grande, el  $k$ -ésimo momento muestral es cercano (en algún sentido) al  $k$ -ésimo momento poblacional. Por ejemplo, para los dos primeros momentos tenemos que

$$\frac{1}{n} \sum_{i=1}^n X_i \approx E(X),$$

$$\frac{1}{n} \sum_{i=1}^n X_i^2 \approx E(X^2).$$

Así, los momentos muestrales son usados para determinar, de manera aproximada, la distribución de probabilidad.

- **Existencia de los momentos.** El método de momentos presupone que existen y que se pueden encontrar expresiones sencillas para los momentos de la distribución en estudio, y que éstas dependen del parámetro o vector de parámetros a estimar. Estas condiciones no necesariamente se cumplen. Por ejemplo, puede comprobarse que la siguiente distribución no posee ningún momento finito: para  $\theta > 0$ ,

$$f(x; \theta) = \begin{cases} \theta x^{-2} & \text{si } x \geq \theta, \\ 0 & \text{en otro caso.} \end{cases}$$

En este caso el método de momentos no puede aplicarse.

- **Solución al sistema de ecuaciones.** El método presupone que la ecuación o sistema de ecuaciones resultante de la igualación de los momentos muestrales y poblacionales tiene una única solución y que ésta es sencilla de encontrar. En general, esto no es así. Cuando se tienen dos o más parámetros, el sistema de ecuaciones puede no ser sencillo de resolver, puesto que las ecuaciones no son necesariamente lineales. Y suponiendo que es posible resolver el sistema de ecuaciones, las expresiones que se encuentran pueden no tener una forma compacta o sencilla. Por ejemplo, considere el caso de la distribución hipergeo( $N, K, n$ ), en donde los tres parámetros son desconocidos. El sistema de ecuaciones resultante no es fácil de resolver.
- **Valores del parámetro.** El método de momentos no garantiza que el estimador encontrado tome valores en el espacio parametral correspondiente. Por ejemplo, si un parámetro toma valores enteros, el método de momentos no necesariamente produce un estimador con valores enteros. Por ejemplo, si consideramos que el parámetro  $p$  es conocido en la distribución  $\text{bin}(k, p)$  y deseamos estimar el parámetro desconocido  $k$  mediante el método de momentos, entonces es inmediato encontrar la solución  $\hat{k} = \bar{X}/p$ , lo que no necesariamente produce un valor entero.

En la Tabla 2.1 se muestran los estimadores por el método de momentos para los parámetros de algunas distribuciones discretas conocidas. Se ha supuesto que  $X_1, \dots, X_n$  es una muestra aleatoria de tamaño  $n$ . En el

Apéndice A al final del texto se puede consultar la expresión y notación de los parámetros para estas distribuciones. Sin embargo, observe que el parámetro  $n$  se reserva para el tamaño de la muestra aleatoria. Para hacer las fórmulas cortas se utiliza la siguiente notación cuando ambos momentos aparecen en la fórmula:

$$m_1 = \frac{1}{n} \sum_{i=1}^n X_i,$$
$$m_2 = \frac{1}{n} \sum_{i=1}^n X_i^2.$$

Es necesario notar que se indica únicamente el resultado producido por el método de momentos, sin garantizar que el estimador tome valores en el espacio parametral correspondiente. Por su complejidad, se ha omitido de esta tabla la distribución hipergeo( $N, K, n$ ).

En la Tabla 2.2 se presentan los estimadores por el método de momentos para los parámetros de algunas distribuciones continuas conocidas. Se incluye el caso de la distribución normal desarrollado antes como ejemplo.

De esta manera, teniendo una distribución de probabilidad dependiente de uno o más parámetros desconocidos, y si existe el número suficiente de sus momentos, uno puede poner en práctica el método de los momentos para intentar obtener estadísticas que pueden proponerse como estimadores de los parámetros desconocidos.

En la siguiente sección veremos un segundo método alternativo general para obtener estimadores para los parámetros desconocidos de una distribución de probabilidad dada.

**Algunos estimadores por el método de momentos**

Distribución	Parámetro(s)	Estimador(es)
$\text{unif}\{1, \dots, k\}$	$k \in \{1, 2, \dots\}$	$\hat{k} = 2\bar{X} - 1$
$\text{Ber}(p)$	$p \in (0, 1)$	$\hat{p} = \bar{X}$
$\text{bin}(k, p)$	$k \in \{1, 2, \dots\}$	$\hat{k} = \frac{m_1^2}{m_1 - (m_2 - m_1^2)}$
	$p \in (0, 1)$	$\hat{p} = 1 - \frac{m_2 - m_1^2}{m_1}$
$\text{geo}(p)$	$p \in (0, 1)$	$\hat{p} = \frac{1}{1 + \bar{X}}$
$\text{bin neg}(r, p)$	$r \in \{1, 2, \dots\}$	$\hat{r} = \frac{m_1^2}{m_2 - m_1^2 - m_1}$
	$p \in (0, 1)$	$\hat{p} = \frac{m_1}{m_2 - m_1^2}$
$\text{Poisson}(\lambda)$	$\lambda \in (0, \infty)$	$\hat{\lambda} = \bar{X}$

Tabla 2.1

**Algunos estimadores por el método de momentos**

Distribución	Parámetro(s)	Estimador(es)
unif( $a, b$ )	$a < b$	$\hat{a} = \frac{4m_1^2 - 3m_2}{2m_1 - 1}$ $\hat{b} = \frac{3m_2 - 2m_1}{2m_1 - 1}$
exp( $\lambda$ )	$\lambda \in (0, \infty)$	$\hat{\lambda} = \frac{1}{\bar{X}}$
gama( $\gamma, \lambda$ )	$\gamma \in (0, \infty)$	$\hat{\gamma} = \frac{m_1^2}{m_2 - m_1^2}$
	$\lambda \in (0, \infty)$	$\hat{\lambda} = \frac{m_1}{m_2 - m_1^2}$
N( $\mu, \sigma^2$ )	$\mu \in (-\infty, \infty)$	$\hat{\mu} = \bar{X}$
	$\sigma^2 \in (0, \infty)$	$\hat{\sigma}^2 = \frac{n-1}{n} S^2$
beta( $a, b$ )	$a \in (0, \infty)$	$\hat{a} = \frac{m_1(m_1 - m_2)}{m_2 - m_1^2}$
	$b \in (0, \infty)$	$\hat{b} = \frac{(1 - m_1)(m_1 - m_2)}{m_2 - m_1^2}$
$\chi^2(k)$	$k \in (0, \infty)$	$\hat{k} = \bar{X}$
$t(k)$	$k \in (0, \infty)$	$\hat{k} = \frac{2m_2}{m_2 - 1}$
F( $a, b$ )	$a \in (0, \infty)$	$\hat{a} = \frac{2m_1^2}{m_1^2 - m_2(2 - m_1)}$
	$b \in (0, \infty)$	$\hat{b} = \frac{2m_1}{m_1 - 1}$

Tabla 2.2

## Ejercicios

105. Suponiendo dada una muestra aleatoria de tamaño  $n$ , encuentre el estimador para  $\theta$  por el método de momentos para cada una de las siguientes distribuciones.

a) Para  $0 < \theta < 4$ ,

$$f(x, \theta) = \begin{cases} \theta/4 & \text{si } x = 1, \\ 1 - \theta/4 & \text{si } x = 2, \\ 0 & \text{en otro caso.} \end{cases}$$

b) Para  $0 < \theta < 6/5$ ,

$$f(x, \theta) = \begin{cases} \theta/2 & \text{si } x = -1, \\ \theta/3 & \text{si } x = 0, \\ 1 - 5\theta/6 & \text{si } x = 1, \\ 0 & \text{en otro caso.} \end{cases}$$

c) Para  $0 < \theta < 3/2$ ,

$$f(x, \theta) = \begin{cases} \theta/3 & \text{si } x = 0, \\ 1 - 2\theta/3 & \text{si } x = 1, \\ \theta/3 & \text{si } x = 2, \\ 0 & \text{en otro caso.} \end{cases}$$

d) Para  $\theta \in \mathbb{N}$ ,

$$f(x, \theta) = \begin{cases} 1/\theta & \text{si } x = 1, 2, \dots, \theta, \\ 0 & \text{en otro caso.} \end{cases}$$

e) Para  $\theta \in \mathbb{N}$ ,

$$f(x, \theta) = \begin{cases} \frac{2x}{\theta(\theta + 1)} & \text{si } x = 1, 2, \dots, \theta, \\ 0 & \text{en otro caso.} \end{cases}$$

f) Para  $0 < \theta < 1$ ,

$$f(x, \theta) = \begin{cases} \theta(1 - \theta)^{x-1} & \text{si } x = 1, 2, \dots \\ 0 & \text{en otro caso.} \end{cases}$$

g) Para  $\theta > 0$ ,

$$f(x, \theta) = \begin{cases} \frac{1}{\theta} & \text{si } 0 < x < \theta, \\ 0 & \text{en otro caso.} \end{cases}$$

h) Para  $\theta > 0$ ,

$$f(x, \theta) = \begin{cases} \frac{2x}{\theta^2} & \text{si } 0 \leq x \leq \theta, \\ 0 & \text{en otro caso.} \end{cases}$$

i) Para cualquier  $\theta$ ,

$$f(x, \theta) = \begin{cases} e^{-(x-\theta)} & \text{si } \theta \leq x < \infty, \\ 0 & \text{en otro caso.} \end{cases}$$

j) Para  $\theta > 0$ ,

$$f(x, \theta) = \begin{cases} \theta x^{\theta-1} & \text{si } 0 < x < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

k) Para  $-1 < \theta < 1$ ,

$$f(x, \theta) = \begin{cases} \frac{1 + \theta x}{2} & \text{si } -1 < x < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

l) Para  $\theta > 0$ ,

$$f(x, \theta) = \begin{cases} \frac{2(\theta - x)}{\theta^2} & \text{si } 0 < x < \theta, \\ 0 & \text{en otro caso.} \end{cases}$$

106. **Distribución doble exponencial.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución doble exponencial que aparece especificada abajo, en donde  $\theta > 0$  es un parámetro desconocido. Encuentre el estimador para  $\theta$  por el método de momentos.

$$f(x, \theta) = \frac{1}{2} \theta e^{-\theta|x|} \quad -\infty < x < \infty.$$

107. **Distribución Rayleigh.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución Rayleigh que aparece especificada abajo, en donde  $\theta > 0$  es un parámetro desconocido. Encuentre el estimador para  $\theta$  por el método de momentos.

$$f(x, \theta) = \begin{cases} 2(x/\theta) e^{-x^2/\theta} & \text{si } x > 0, \\ 0 & \text{en otro caso.} \end{cases}$$

108. Las siguientes distribuciones dependen de dos parámetros: uno desconocido denotado por la letra  $\theta$ , y otro que supondremos conocido y que se denota por una letra distinta. Encuentre el estimador por el método de momentos para el parámetro desconocido  $\theta$ , suponiendo un tamaño de muestra  $n$ .

- |                               |                                  |
|-------------------------------|----------------------------------|
| a) bin( $k, \theta$ ).        | i) N( $\theta, \sigma^2$ ).      |
| b) bin( $\theta, p$ ).        | j) N( $\mu, \theta$ ).           |
| c) bin neg( $r, \theta$ ).    | k) beta( $a, \theta$ ).          |
| d) bin neg( $\theta, p$ ).    | l) beta( $\theta, b$ ).          |
| e) unif( $a, \theta$ ).       | m) Weibull( $\theta, \lambda$ ). |
| f) unif( $\theta, b$ ).       | n) Weibull( $\alpha, \theta$ ).  |
| g) gama( $\theta, \lambda$ ). | ñ) F( $a, \theta$ ).             |
| h) gama( $\gamma, \theta$ ).  | o) F( $\theta, b$ ), $b > 4$ .   |

109. **Algunas distribuciones discretas.** Compruebe que los estimadores por el método de momentos para los parámetros de las distribuciones discretas que aparecen en la Tabla 2.1 son los indicados. Suponga que  $X_1, \dots, X_n$  es una muestra aleatoria de tamaño  $n$  de la distribución

en estudio. En caso necesario, consulte el Apéndice A al final del texto la expresión y notación de los parámetros para estas distribuciones. Sin embargo, observe que el parámetro  $n$  se ha reservado para el tamaño de la muestra aleatoria. Recordemos que  $m_1$  denota el primer momento muestral, y  $m_2$  denota el segundo momento muestral. Se indica únicamente el resultado producido por el método de momentos, sin garantizar que el estimador tome valores en el espacio parametral correspondiente.

110. **Valores al azar.** Los diez números que aparecen en la tabla de abajo son valores al azar generados en  $\mathbb{R}$  de la distribución  $\text{geo}(\theta)$ , mediante el comando `rgeom(10,  $\theta$ )`. Para el parámetro  $\theta$  se usó uno de dos valores:  $\theta = 0.2$  ó  $\theta = 0.4$ . ¿Puede usted determinar el valor de  $\theta$  que se usó? Observe que nunca existirá una confianza absoluta en la respuesta.

Observaciones al azar	
$x_1 = 0$	$x_6 = 3$
$x_2 = 0$	$x_7 = 1$
$x_3 = 0$	$x_8 = 0$
$x_4 = 2$	$x_9 = 3$
$x_5 = 3$	$x_{10} = 6$

111. **Algunas distribuciones continuas.** Compruebe que los estimadores por el método de momentos para los parámetros de las distribuciones continuas que aparecen en la Tabla 2.2 son los indicados. Suponga que  $X_1, \dots, X_n$  es una muestra aleatoria de tamaño  $n$  de la distribución en estudio. En caso necesario, consulte el Apéndice A al final del texto la expresión y notación de los parámetros para estas distribuciones. Se incluye el caso de la distribución normal desarrollado antes como ejemplo.
112. **Valores al azar.** Los diez números que aparecen en la tabla de abajo son valores al azar generados en  $\mathbb{R}$  de la distribución  $\text{exp}(\theta)$ , mediante el comando `rexp(10,  $\theta$ )`. Para el parámetro  $\theta$  se usó uno de dos valores:  $\theta = 2$  ó  $\theta = 5$ . ¿Puede usted determinar el valor de  $\theta$  que se usó? Observe que nunca existirá una confianza absoluta en la respuesta.

Observaciones al azar	
$x_1 = 0.026$	$x_6 = 0.015$
$x_2 = 0.370$	$x_7 = 1.069$
$x_3 = 0.665$	$x_8 = 0.352$
$x_4 = 1.567$	$x_9 = 0.723$
$x_5 = 0.235$	$x_{10} = 0.364$

113. Cuatro focos se ponen a prueba permanente hasta que dejan de funcionar. Los tiempos registrados de vida útil en horas fueron los siguientes.

Tiempo en horas	
$x_1 = 950$	$x_3 = 1020$
$x_2 = 1050$	$x_4 = 985$

Suponga que se acepta la distribución gama( $\gamma, \lambda$ ) como modelo para el tiempo de vida útil de los focos.

$$f(x; \gamma, \lambda) = \begin{cases} \frac{(\lambda x)^{\gamma-1}}{\Gamma(\gamma)} \lambda e^{-\lambda x} & \text{si } x > 0, \\ 0 & \text{en otro caso.} \end{cases}$$

- a) Estime  $\gamma$  y  $\lambda$  por el método de momentos.
- b) Calcule la probabilidad de que un foco nuevo de las mismas características tenga un tiempo de vida mayor a 1000 horas.
114. **Distribución uniforme.** Suponga que las cinco cantidades que aparecen abajo son observaciones de una variable aleatoria con distribución uniforme en el intervalo  $(a, b)$ . Encuentre una estimación por el método de momentos para los parámetros  $a$  y  $b$ .

Observaciones	
$x_1 = 4.0$	$x_4 = 1.5$
$x_2 = 3.2$	$x_5 = 7.2$
$x_3 = 0.3$	

115. **Distribución Bernoulli o binomial.** Al final de cada hora de un día de trabajo en una fábrica se escogieron al azar 10 artículos de una línea de producción para detectar artículos defectuosos y se obtuvieron los resultados que aparecen en la tabla de abajo. Use el método de momentos para estimar la proporción de artículos defectuosos en esta línea de producción.

Hora	1	2	3	4	5	6	7	8
Artículos defectuosos	1	2	1	0	1	2	0	1

### 2.3. Método de máxima verosimilitud

Este importante método para estimar parámetros fue difundido ampliamente por el estadístico inglés Ronald Fisher<sup>2</sup> a través de varios trabajos publicados durante la década de 1920. Sin embargo, la idea fundamental del método había sido usada con anterioridad por varios matemáticos importantes como C. F. Gauss y P. -S. Laplace. La idea que subyace en el método de máxima verosimilitud aparece en la solución de muchos otros problemas de la estadística.

Para explicar este método, primero definiremos una función llamada de verosimilitud. Tomaremos como base una colección de variables aleatorias cuya distribución depende de un parámetro desconocido que se desea estimar.

**Definición 2.7** La **función de verosimilitud** de un vector aleatorio  $(X_1, \dots, X_n)$  cuya distribución depende de un parámetro  $\theta$  se define como la función de densidad o de probabilidad conjunta

$$L(\theta) = f_{X_1, \dots, X_n}(x_1, \dots, x_n, \theta). \quad (2.1)$$

Como la notación lo sugiere, nos interesa estudiar esta función como función del parámetro  $\theta$ . Los valores de este parámetro se encuentran en un

<sup>2</sup>Ronald Aylmer Fisher (1890-1962), estadístico y genetista inglés.

cierto espacio parametral  $\Theta$ , y ese es el dominio de definición de la función de verosimilitud. El parámetro desconocido  $\theta$  puede tomar valores en un conjunto discreto, o bien en todo un continuo de valores, dependiendo de la distribución de probabilidad considerada. Los números  $x_1, \dots, x_n$  son tratados como constantes y son los valores particulares de las variables aleatorias con el correspondiente subíndice.

Observemos que en la definición no se está suponiendo necesariamente que las variables aleatorias  $X_1, \dots, X_n$  constituyen una muestra aleatoria. Sin embargo, cuando sea este el caso, por la hipótesis de independencia, la función de verosimilitud adquiere la forma del siguiente producto

$$L(\theta) = f_{X_1}(x_1, \theta) \cdots f_{X_n}(x_n, \theta). \quad (2.2)$$

Y si ahora se usa la hipótesis de idéntica distribución, entonces se pueden omitir los subíndices de estos factores y escribir

$$L(\theta) = f(x_1, \theta) \cdots f(x_n, \theta).$$

En la mayoría de los casos consideraremos que la información proviene de una muestra aleatoria y, por lo tanto, la última expresión es la que utilizaremos para la función de verosimilitud. La letra  $L$  procede del término en inglés *Likelihood*, que tradicionalmente se ha traducido como verosimilitud.

**¿En qué consiste el método de máxima verosimilitud?** Consiste en encontrar el valor de  $\theta$  que maximiza a la función  $L(\theta)$ . Al valor de  $\theta$  en donde  $L(\theta)$  alcanza su máximo se le llama estimación de máxima verosimilitud o estimación máximo verosímil.

La idea intuitiva es muy natural: se debe encontrar el valor de  $\theta$  de tal forma que los datos observados  $x_1, \dots, x_n$  tengan máxima probabilidad de ser obtenidos. La probabilidad de observar estos valores está directamente relacionada con la función de verosimilitud, y por ello se pide maximizarla. En el caso de una distribución discreta, la función de verosimilitud es exactamente la probabilidad de observar los valores  $x_1, \dots, x_n$ . En ocasiones se usa la expresión  $\hat{\theta}_{mv}$  para denotar el estimador por máxima verosimilitud para  $\theta$ . El significado de las letras que aparecen como subíndices es evidente. Veamos ahora algunos ejemplos.

**Ejemplo 2.13** Encontraremos el estimador máximo verosímil para el parámetro desconocido de una distribución  $\exp(\theta)$ . Suponiendo dada una muestra aleatoria de tamaño  $n$  de esta distribución, la función de verosimilitud es, para  $\theta > 0$ ,

$$\begin{aligned} L(\theta) &= f(x_1, \theta) \cdots f(x_n, \theta) \\ &= \theta e^{-\theta x_1} \cdots \theta e^{-\theta x_n} \\ &= \theta^n e^{-\theta n \bar{x}}. \end{aligned}$$

La gráfica de esta función se muestra en la Figura 2.1.

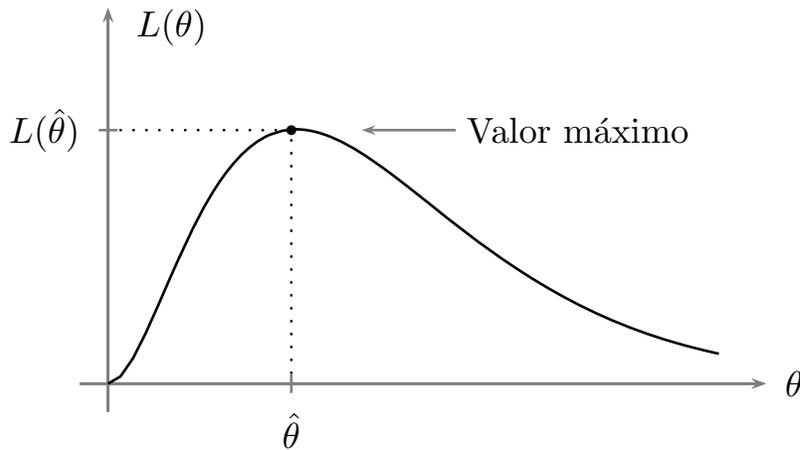


Figura 2.1

Maximizar la función  $L(\theta)$  es equivalente a maximizar  $\ln L(\theta)$ , pues la función logaritmo es continua y monótona creciente en su dominio de definición. Hacemos la operación anterior debido a que la función resultante es más fácil de maximizar como veremos a continuación. Tenemos que

$$\ln L(\theta) = n \ln \theta - \theta n \bar{x}.$$

Derivando respecto a  $\theta$  e igualando a cero, se llega a la ecuación

$$\frac{n}{\theta} - n \bar{x} = 0,$$

de donde se obtiene  $\hat{\theta} = 1/\bar{x}$ . Observe que hemos escrito  $\hat{\theta}$  en lugar de  $\theta$  en esta última expresión. Calculando la segunda derivada se puede comprobar que en este punto la función de verosimilitud tiene, efectivamente, un

máximo. Si  $x_1, \dots, x_n$  son los valores numéricos observados de la muestra aleatoria, entonces el número  $\hat{\theta}(x_1, \dots, x_n) = 1/\bar{x}$  es la estimación máximo verosímil. El estimador máximo verosímil es, entonces, la variable aleatoria

$$\hat{\theta} = \frac{1}{\bar{X}}.$$

■

En el ejemplo anterior fue conveniente maximizar la expresión  $\ln L(\theta)$  en lugar de  $L(\theta)$ . Existe equivalencia entre ambas expresiones en el sentido de que el punto en donde se alcanza el máximo de una de las funciones es el mismo que para la otra función, aunque los valores máximos serán distintos. Observe que no nos interesa calcular el valor máximo de la función de verosímil, sino el punto en el que se alcanza ese valor máximo. Con frecuencia se usan transformaciones de este tipo para encontrar con mayor facilidad el punto buscado.

Por razones de simplicidad hemos escrito la función de densidad de la distribución exponencial como  $f(x, \theta) = \theta e^{-\theta x}$ , sin especificar que  $x > 0$ . En sentido estricto, a la expresión anterior se le debe multiplicar por la función indicadora  $1_{(0, \infty)}(x)$ . Esto no tuvo consecuencias en el cálculo anterior pues en esta función indicadora no aparece el parámetro  $\theta$ . Sin embargo, en aquellas distribuciones en donde el soporte involucra al parámetro a estimar, es crucial incorporar al cálculo la función indicadora correspondiente. Más adelante proporcionaremos un ejemplo de esta situación. Por ahora consideraremos un ejemplo de una distribución discreta.

**Ejemplo 2.14** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución  $\text{geo}(\theta)$ , con parámetro  $\theta$  desconocido. Encontraremos el estimador por máxima verosimilitud para  $\theta$ . La función de verosimilitud es, para  $\theta \in (0, 1)$ ,

$$\begin{aligned} L(\theta) &= f(x_1, \theta) \cdots f(x_n, \theta) \\ &= \theta (1 - \theta)^{x_1} \cdots \theta (1 - \theta)^{x_n} \\ &= \theta^n (1 - \theta)^{n\bar{x}}. \end{aligned}$$

La gráfica de esta función se muestra en la Figura 2.2.

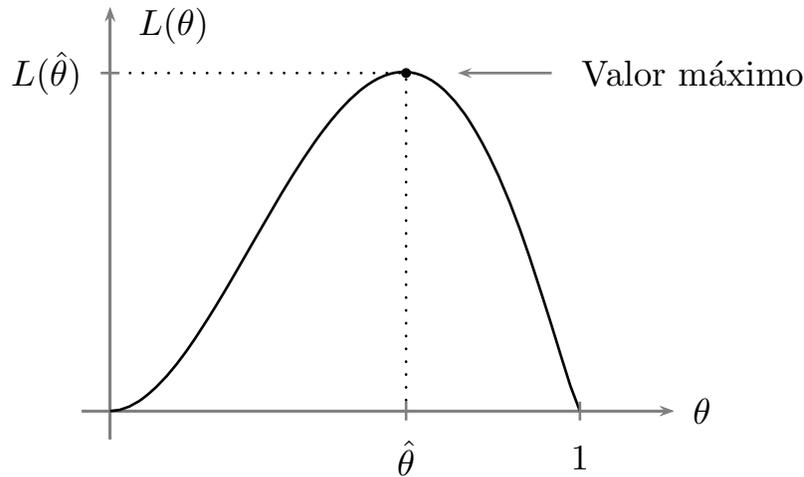


Figura 2.2

Tomando logaritmo se obtiene  $\ln L(\theta) = n \ln \theta + n\bar{x} \ln(1 - \theta)$ . Derivando respecto a  $\theta$  e igualando a cero se llega a la ecuación

$$\frac{n}{\theta} - \frac{n\bar{x}}{1 - \theta} = 0.$$

De donde se obtiene que la estimación es el número  $\hat{\theta} = 1/(1 + \bar{x})$ . Hemos escrito  $\hat{\theta}$  en lugar de  $\theta$ . De esta identidad se sigue que el estimador máximo verosímil es la variable aleatoria

$$\hat{\theta} = \frac{1}{1 + \bar{X}}.$$

Nuevamente, mediante el cálculo de la segunda derivada se puede comprobar que el valor encontrado es un punto crítico en donde la función de verosimilitud tiene efectivamente un máximo global en el espacio parametral  $\Theta = (0, 1)$ . ■

El método de máxima verosimilitud puede aplicarse también en el caso cuando la distribución depende de dos o más parámetros. En el siguiente ejemplo encontraremos los estimadores por máxima verosimilitud para los dos parámetros de la distribución normal.

**Ejemplo 2.15** Dada una muestra aleatoria de tamaño  $n$  de la distribución  $N(\mu, \sigma^2)$ , en donde ambos parámetros son desconocidos, la función de verosimilitud es, para valores  $\mu \in \mathbb{R}$  y  $\sigma^2 > 0$ ,

$$\begin{aligned} L(\mu, \sigma^2) &= f(x_1; \mu, \sigma^2) \cdots f(x_n; \mu, \sigma^2) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_1-\mu)^2/2\sigma^2} \cdots \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_n-\mu)^2/2\sigma^2} \\ &= \left( \frac{1}{2\pi\sigma^2} \right)^{n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right). \end{aligned}$$

La gráfica de esta función se encuentra en la Figura 2.3 para  $n = 2$ ,  $x_1 = 1$  y  $x_2 = 3$ .

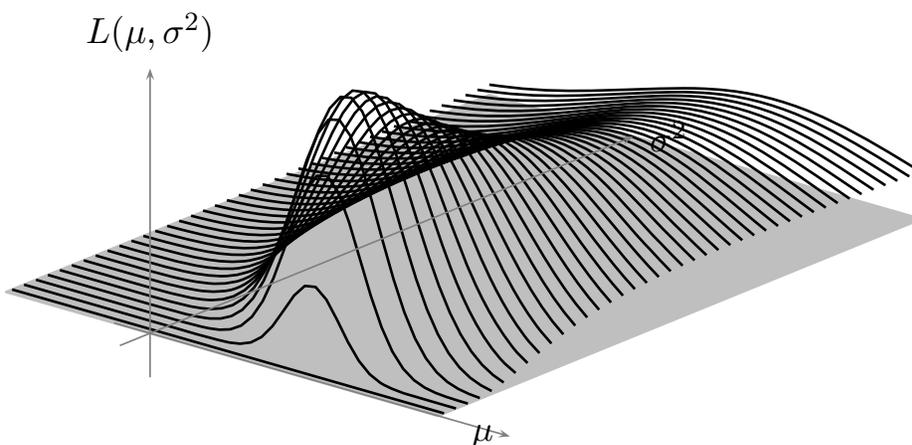


Figura 2.3

Buscamos encontrar el punto en donde esta función de dos variables alcanza su valor máximo. Nuevamente, el logaritmo de esta función adquiere una expresión más sencilla. Tenemos que

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

Por lo tanto,

$$\begin{aligned}\frac{\partial}{\partial \mu} \ln L(\mu, \sigma^2) &= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu), \\ \frac{\partial}{\partial \sigma^2} \ln L(\mu, \sigma^2) &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2.\end{aligned}$$

Igualando a cero ambas derivadas, encontramos un sistema de dos ecuaciones con dos variables,

$$\begin{aligned}\frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) &= 0, \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 &= 0.\end{aligned}$$

De estas ecuaciones se obtiene  $\mu = \frac{1}{n} \sum_{i=1}^n x_i$  y  $\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2$ . Por lo tanto, los estimadores por el método de máxima verosimilitud son

$$\begin{aligned}\hat{\mu} &= \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}, \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n-1}{n} S^2.\end{aligned}$$

Para verificar que la función de verosimilitud tiene, efectivamente, un máximo en el punto encontrado, es necesario calcular la matriz hessiana

$$H(\mu, \sigma^2) = \begin{pmatrix} \frac{\partial^2}{\partial \mu \partial \mu} \ln L(\mu, \sigma^2) & \frac{\partial^2}{\partial \mu \partial \sigma^2} \ln L(\mu, \sigma^2) \\ \frac{\partial^2}{\partial \sigma^2 \partial \mu} \ln L(\mu, \sigma^2) & \frac{\partial^2}{\partial \sigma^2 \partial \sigma^2} \ln L(\mu, \sigma^2) \end{pmatrix}.$$

Se evalúa  $H$  en el punto  $(\hat{\mu}, \hat{\sigma}^2)$ , y se comprueba que la matriz  $H(\hat{\mu}, \hat{\sigma}^2)$  es negativa definida. Véase la página 319 del Apéndice A, en donde se hace una revisión de este procedimiento. Observemos que, para esta distribución, los estimadores por máxima verosimilitud coinciden con los encontrados anteriormente por el método de momentos. Esto no siempre es así. ■

Debe advertirse que la aplicación de las derivadas para encontrar el máximo de una función de verosimilitud no siempre produce expresiones cerradas para el estimador o estimadores, como en los casos mostrados. Por ejemplo, para la distribución  $\text{gama}(\gamma, \lambda)$ , con ambos parámetros desconocidos, se encuentra que  $\hat{\gamma}$  y  $\hat{\lambda}$  satisfacen ciertas ecuaciones que no son fáciles de resolver y algún método numérico debe utilizarse. Como un segundo ejemplo, considere la distribución  $\text{bin}(k, p)$ , con ambos parámetros desconocidos. La dificultad aquí radica en que se debe maximizar la función de verosimilitud para una variable entera  $k \geq 1$  y una variable continua  $p$  en el intervalo  $(0, 1)$ , para cualquier tamaño de muestra  $n$ . En este caso el proceso de maximización no es fácil de llevar a cabo.

El siguiente ejemplo muestra algunas otras dificultades técnicas que pueden surgir al buscar el máximo de una función de verosimilitud.

**Ejemplo 2.16** Consideremos dada una muestra aleatoria tamaño  $n$  de una distribución  $\text{unif}(0, \theta)$ , cuya función de densidad se puede escribir como sigue

$$f(x, \theta) = \frac{1}{\theta} \cdot 1_{(0, \theta)}(x),$$

en donde  $\theta > 0$  es un parámetro desconocido que deseamos estimar. La función de verosimilitud es

$$\begin{aligned} L(\theta) &= \frac{1}{\theta^n} \cdot 1_{(0, \theta)}(x_1) \cdots 1_{(0, \theta)}(x_n) \\ &= \frac{1}{\theta^n} \cdot 1_{(x_{(n)}, \infty)}(\theta) \cdot 1_{(0, \infty)}(x_{(1)}). \end{aligned}$$

Se puede comprobar que la función  $L(\theta)$  es constante cero hasta el valor  $x_{(n)} = \max_i x_i$ , y toma la expresión  $1/\theta^n$  después de ese valor. Véase la Figura 2.4, en donde  $x_{(i)}$  es el  $i$ -ésimo valor ordenado de la muestra. En este ejemplo, consideramos una distribución de probabilidad en donde es decisivo en el análisis incorporar el soporte de la distribución a través de una función indicadora.

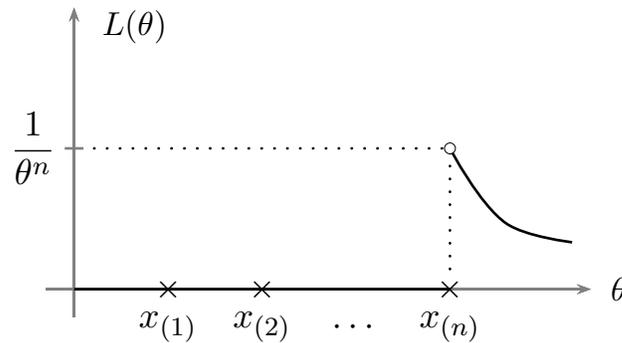


Figura 2.4

Así, la función de verosimilitud nunca alcanza su máximo, y el estimador máximo verosímil no existe. Esta situación puede subsanarse si se considera que la distribución uniforme se tiene sobre el intervalo con extremo derecho cerrado  $(0, \theta]$ . No es difícil darse cuenta que, en este caso, el estimador máximo verosímil existe y es  $\hat{\theta} = X_{(n)}$ . ■

El siguiente es otro ejemplo de una situación inesperada que surge al aplicar el método de máxima verosimilitud.

**Ejemplo 2.17** Consideremos una muestra aleatoria de tamaño  $n$  de una distribución  $\text{unif}(\theta, \theta + 1)$ , en donde  $\theta$  es un parámetro desconocido que puede tomar cualquier valor real y que deseamos estimar. La función de verosimilitud de la muestra aleatoria se puede escribir como sigue

$$\begin{aligned} L(\theta) &= 1_{(\theta, \theta+1)}(x_1) \cdots 1_{(\theta, \theta+1)}(x_n) \\ &= 1_{(-\infty, x_{(1)})}(\theta) \cdot 1_{(x_{(n)}, \infty)}(\theta + 1) \\ &= 1_{(x_{(n)}-1, x_{(1)})}(\theta). \end{aligned}$$

La última igualdad se obtiene de las condiciones  $\theta < x_{(1)}$  y  $\theta + 1 > x_{(n)}$ . Esto significa que la función de verosimilitud es constante 1 para cualquier valor de  $\theta$  en el intervalo  $(x_{(n)} - 1, x_{(1)})$ , y por lo tanto es máxima para cualquier valor del parámetro dentro de este intervalo. Es decir, existe una infinidad no numerable de estimadores máximo verosímiles. ■

Los dos ejemplos anteriores muestran que hay circunstancias en donde el estimador máximo verosímil puede no existir, o bien, no ser único. Sin embargo, en nuestro tratamiento tenderemos a excluir tales casos, y nos referiremos al estimador máximo verosímil como si éste existiera y fuera único, suponiendo implícitamente las condiciones necesarias para que ello ocurra.

Después de haber mostrado algunos ejemplos del método de máxima verosimilitud, haremos ahora algunas observaciones generales sobre este interesante método para estimar parámetros.

- **Aplicación.** El método de máxima verosimilitud puede aplicarse sin distinción alguna, tanto para distribuciones discretas, como continuas. Para el caso de distribuciones discretas, puede convenir usar funciones indicadoras como exponentes para escribir la función de probabilidad como una sola expresión sobre el soporte de la distribución. Esto se muestra a continuación.

$$\begin{aligned}
 f(x) &= \begin{cases} p_1 & \text{si } x = x_1, \\ p_2 & \text{si } x = x_2, \\ \dots & \dots \\ 0 & \text{en otro caso.} \end{cases} \\
 &= \begin{cases} p_1^{1_{\{x_1\}}(x)} \cdot p_2^{1_{\{x_2\}}(x)} \dots & \text{si } x = x_1, x_2, \dots \\ 0 & \text{en otro caso.} \end{cases}
 \end{aligned}$$

- **Momentos vs verosimilitud.** El método de máxima verosimilitud no produce necesariamente los mismos estimadores que el método de momentos. Esto es así porque en cada método se busca el valor de  $\theta$  que cumpla ciertas características, y éstas son diferentes en cada caso.
- **Aplicación general.** En los ejemplos mostrados se aplicó el método de máxima verosimilitud cuando la función de verosimilitud toma la forma del producto en la ecuación (2.2). Esto es consecuencia de la hipótesis de independencia de las variables de la muestra aleatoria. Sin embargo, el método es más general y se puede aplicar también cuando no se tenga esta hipótesis de independencia y la función a maximizar es la que aparece en la ecuación (2.1).

- **Diferenciabilidad.** El procedimiento usual de maximización de la función de verosimilitud a través del cálculo de derivadas puede llevarse a cabo únicamente cuando el parámetro toma un continuo de valores, cuando la función de verosimilitud sea diferenciable y cuando ésta alcance un máximo global en un único punto dentro de su dominio. Sin embargo, el método de máxima verosimilitud no presupone necesariamente el uso de las derivadas para su aplicación. Por ejemplo, si un parámetro toma valores enteros, otra técnica de maximización debe utilizarse.
- **Solubilidad.** Desde el punto de vista práctico, se puede aplicar el método de máxima verosimilitud si no es demasiado difícil encontrar el punto en donde la función de verosimilitud es máxima. Por ejemplo, en el caso de la distribución gama, suponiendo ambos parámetros desconocidos, no existe una fórmula explícita para el punto en donde la función de verosimilitud alcanza su máximo.
- **Valores del parámetro.** Suponiendo la existencia de un estimador máximo verosímil, y a diferencia del método de momentos, el método de máxima verosimilitud garantiza que la estimación toma un valor en el espacio parametral correspondiente. Esto es así por la especificación misma del método: la función de verosimilitud se debe maximizar sobre el espacio parametral.
- **Difeomorfismos.** Como se ha ilustrado en los ejemplos, en algunas ocasiones resulta más conveniente maximizar el logaritmo de la función de verosimilitud que la función de verosimilitud misma. Cualquier otra función monótona y diferenciable puede ser usada convenientemente.
- **Existencia y unicidad.** El estimador máximo verosímil puede no existir como en el caso de la distribución  $\text{unif}(0, \theta)$ , y puede no ser único como en el caso de la distribución  $\text{unif}(\theta, \theta + 1)$ . Ambos ejemplos fueron desarrollados en páginas anteriores.
- **Cambios en el espacio parametral.** Si se reduce el espacio parametral, es decir, si se reduce el dominio en el que está definida la función de verosimilitud, es muy posible que el máximo no sea el mismo, y por lo tanto, el estimador máximo verosímil puede cambiar.

Así pues, considerar cambios en el espacio parametral puede hacer aún más difícil el proceso de encontrar el estimador máximo verosímil para un parámetro.

Existen otros métodos para encontrar estimadores puntuales de parámetros. Dos de ellos son el método de la ji-cuadrada mínima y el método de distancia mínima. Ambos pueden consultarse en [18]. Existe también otra perspectiva distinta para la estadística en general llamada estadística bayesiana. Esta perspectiva provee, en particular, sus propios métodos para la estimación de parámetros. Se puede obtener mayor información, por ejemplo, en [3].

## Funciones parametrales

En ocasiones nos interesará estudiar funciones de un parámetro o conjunto de parámetros de una distribución. Tal concepto se formaliza en la siguiente definición.

**Definición 2.8** Sea  $\theta$  un parámetro o vector de parámetros de una distribución. A cualquier función  $\theta \mapsto \tau(\theta)$  se le llama función parametral.

Veamos algunos ejemplos.

- Si la distribución en estudio es  $\exp(\theta)$ , entonces  $\tau(\theta) = \theta^2 - 1$  es un ejemplo de una función parametral.
- En el caso de la distribución  $\text{bin}(n, p)$ , se puede definir la función parametral correspondiente a la media  $\tau(n, p) = np$ .
- De manera general, los momentos de una distribución (suponiendo su existencia) son funciones de los posibles parámetros.
- Las probabilidades de los distintos eventos son ejemplos de funciones parametrales: si  $X$  es una variable aleatoria con distribución dependiente de uno o varios parámetros, entonces la probabilidad  $P(X \in A)$  es una función parametral para cada conjunto  $A$  de Borel de  $\mathbb{R}$ .
- Los cuantiles de una distribución son ejemplos de funciones parametrales.

Estaremos interesados en encontrar estimadores también para estas funciones parametrales, y estos casos incluyen, por supuesto, a los parámetros individuales.

Supongamos ahora que  $\hat{\theta}$  es el estimador máximo verosímil para  $\theta$ . Si consideramos a una función parametral  $\tau(\theta)$  como un nuevo parámetro que necesita ser estimado por el método de máxima verosimilitud, ¿será cierto que su estimador máximo verosímil es  $\tau(\hat{\theta})$ ? Para responder esta pregunta, observemos que no está claro cuál es la función de verosimilitud asociada a la función parametral  $\tau(\theta)$ . Vamos a definir primero esta función y después daremos respuesta a la pregunta planteada.

**Definición 2.9** La función de verosimilitud asociada a una función parametral  $\tau(\theta)$  se denota por  $L^*$  y se define de la forma siguiente: si  $\eta$  es un posible valor de  $\tau(\theta)$ , entonces

$$L^*(\eta) = \sup \{L(\theta) : \theta \in \tau^{-1}(\eta)\}. \quad (2.3)$$

Al posible valor  $\hat{\eta}$  que maximiza  $L^*(\eta)$  se le llama el estimador máximo verosímil para  $\tau(\theta)$ .

Observemos que el conjunto que aparece en la identidad (2.3) corresponde al conjunto no vacío de todas las evaluaciones  $L(\theta)$  en donde  $\theta$  es una preimagen del valor  $\eta$  y se puede escribir como  $L(\tau^{-1}(\eta))$ , esto corresponde a la aplicación de la función  $L$  en cada elemento del conjunto  $\tau^{-1}(\eta)$ . Al tomar el supremo sobre este conjunto se obtiene la función numérica  $L^*(\eta)$ , la cual estamos definiendo como la función de verosimilitud de la función parametral  $\tau(\theta)$ . Veamos algunos ejemplos.

**Ejemplo 2.18** Sea  $L(\theta)$  la función de verosimilitud de una muestra aleatoria de la distribución  $\text{Ber}(\theta)$ , con  $0 < \theta < 1$ . Daremos dos ejemplos de funciones parametrales y encontraremos las funciones de verosimilitud correspondientes.

- Consideremos la función parametral  $\tau(\theta) = \theta^2$ . En este caso la función parametral también toma valores en el intervalo  $(0, 1)$  como lo hace

$\theta$  y es una función uno a uno. Véase la Figura 2.5(a). La función de verosimilitud para  $\tau(\theta)$  se puede escribir como sigue: para  $0 < \eta < 1$ ,

$$L^*(\eta) = L(\tau^{-1}(\eta)) = L(\sqrt{\eta}).$$

- Consideremos ahora la función parametral  $\tau(\theta) = \theta(1 - \theta)$ . Esta función también toma valores en  $(0, 1)$  pero, para cada uno de sus valores  $\eta$ , hay dos preimágenes  $\theta_1$  y  $\theta_2$  como se muestra en la Figura 2.5(b). Así, la función de verosimilitud para  $\tau(\theta)$  está dada de la siguiente manera: para  $0 < \eta < 1$ ,

$$L^*(\eta) = \max \{L(\theta_1), L(\theta_2)\}.$$

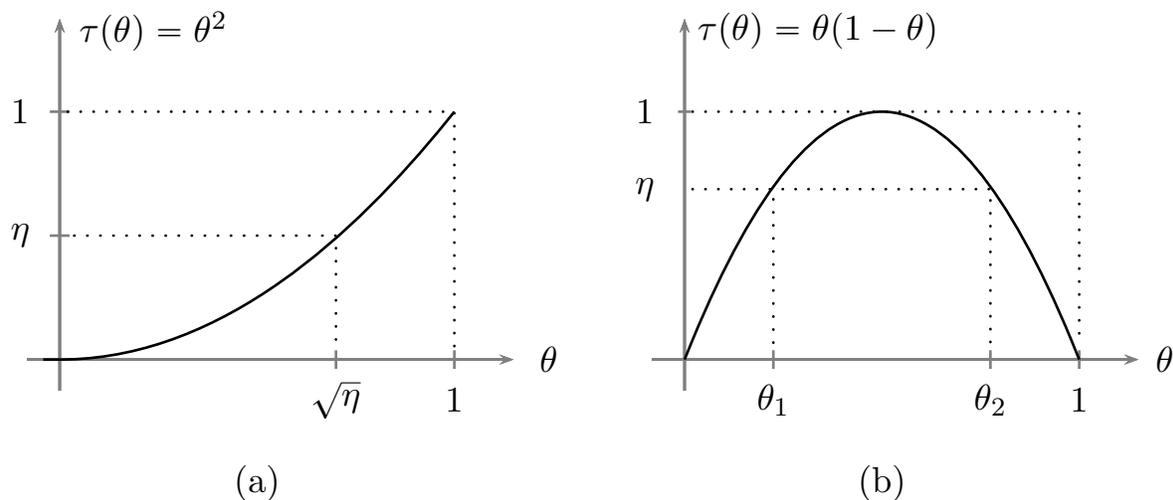


Figura 2.5

El siguiente resultado justifica la forma de definir la función de verosimilitud para una función parametral, pues de esa manera se responde afirmativamente a la pregunta planteada líneas arriba.

**Teorema 2.1 (Principio de invarianza)** Si  $\hat{\theta}$  es el estimador máximo verosímil para un parámetro  $\theta$ , entonces el estimador máximo verosímil para una función parametral  $\tau(\theta)$  es  $\tau(\hat{\theta})$ .

*Demostración.* Consideremos primero el caso cuando la función  $\theta \mapsto \tau(\theta)$  es uno a uno. Entonces la función inversa de  $\tau$  existe y la función de verosimilitud para  $\tau(\theta)$  se puede expresar de la siguiente forma: si  $\eta = \tau(\theta)$ ,

$$L^*(\eta) = L(\tau^{-1}(\eta)) = L(\theta).$$

De esta manera, el máximo de  $L^*(\eta)$  coincide con el máximo de  $L(\theta)$  y este último se alcanza en  $\hat{\theta}$ . Entonces  $L^*(\eta)$  alcanza su máximo en  $\eta = \tau(\hat{\theta})$ .

Veamos ahora el caso cuando  $\theta \mapsto \tau(\theta)$  no necesariamente es una función uno a uno. Por la identidad (2.3), el valor máximo del conjunto de valores  $L^*(\eta)$  coincide con el valor máximo de  $L(\theta)$ . Este último se alcanza en  $\hat{\theta}$ . Por lo tanto, si  $\hat{\eta}$  es el valor  $\tau(\hat{\theta})$ , entonces

$$L^*(\hat{\eta}) = L^*(\tau(\hat{\theta})) = L(\tau^{-1}(\tau(\hat{\theta}))) \ni L(\hat{\theta}).$$

La última afirmación establece que  $L(\hat{\theta})$  es un valor tomado por la función  $L^*(\eta)$ . Como  $L(\hat{\theta})$  es el valor máximo de  $L(\theta)$ , también es el valor máximo de  $L^*(\eta)$  y se alcanza para esta última función en  $\eta = \tau(\hat{\theta})$ . ■

Observemos que el principio de invarianza es también válido cuando el parámetro  $\theta$  es un vector de parámetros. En efecto, en la demostración que hemos presentado no se presupone que  $\theta$  sea un parámetro unidimensional. Veamos algunos ejemplos de este resultado.

**Ejemplo 2.19** El estimador máximo verosímil para el parámetro  $\theta$  en la distribución Bernoulli es  $\bar{X}$ . Entonces el estimador máximo verosímil para la función parametral  $\theta^2$  es  $\bar{X}^2$ . Si ahora consideramos la función parametral  $\theta(1 - \theta)$ , entonces su estimador máximo verosímil es  $\bar{X}(1 - \bar{X})$ . ■

**Ejemplo 2.20** Los estimadores máximo verosímiles para los parámetros de la distribución  $N(\mu, \sigma^2)$  son  $\hat{\mu} = \bar{X}$  y  $\hat{\sigma}^2 = ((n-1)/n)S^2$ . Por el principio de invarianza, el estimador máximo verosímil para la función parametral

- a)  $\mu + 5$  es  $\bar{X} + 5$ .
- b)  $\mu + \sigma$  es  $\bar{X} + \sqrt{(n-1)/n} S$ .
- c)  $\mu/\sigma^2$  es  $(n/(n-1))\bar{X}/S^2$ .

■

## Ejercicios

- 116. Suponiendo una muestra aleatoria de tamaño  $n$ , encuentre el estimador por máxima verosimilitud del parámetro  $\theta$  de cada una de las distribuciones que aparecen en el ejercicio 105 en la página 103.
- 117. **Algunas distribuciones discretas.** Compruebe que los estimadores por el método de máxima verosimilitud para los parámetros de las distribuciones discretas que aparecen en la Tabla 2.3 son los indicados. Suponga que  $X_1, \dots, X_n$  es una muestra aleatoria de tamaño  $n$  de la distribución en estudio. En caso necesario consulte el Apéndice A al final del texto la expresión y notación de los parámetros para estas distribuciones. Como antes, el parámetro  $n$  se reserva para el tamaño de la muestra aleatoria.
- 118. **Distribución binomial.** Suponga que los datos que se muestran en la tabla que aparece abajo corresponden a 50 observaciones de una variable aleatoria  $X$  con distribución  $\text{bin}(k, p)$ , en donde  $k = 5$  y  $p$  es desconocido. Encuentre el estimador máximo verosímil para la probabilidad  $P(X \geq 2)$ .
- 119. **Distribución exponencial.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{exp}(\theta)$ , en donde  $\theta > 0$  es desconocido. Suponga que en lugar de observar esta muestra aleatoria se observan las primeras  $k$  estadísticas de orden  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(k)}$ , en donde  $k \leq n$ . Encuentre el estimador máximo verosímil para  $\theta$  usando  $X_{(1)}, \dots, X_{(k)}$ .

**Algunos estimadores por el  
método de máxima verosimilitud**

Distribución	Parámetro(s)	Estimador(es)
$\text{unif}\{1, \dots, k\}$	$k \in \{1, 2, \dots\}$	$\hat{k} = X_{(n)}$
$\text{Ber}(p)$	$p \in (0, 1)$	$\hat{p} = \bar{X}$
$\text{geo}(p)$	$p \in (0, 1)$	$\hat{p} = \frac{1}{1 + \bar{X}}$
$\text{Poisson}(\lambda)$	$\lambda \in (0, \infty)$	$\hat{\lambda} = \bar{X}$

Tabla 2.3

120. **Distribución Bernoulli.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{Ber}(\theta)$ , con  $0 < \theta < 1$  desconocido. Encuentre el estimador por el método de máxima verosimilitud para la función parametral  $\tau(\theta)$  indicada.
- a)  $\tau(\theta) = \theta^2$ .
- b)  $\tau(\theta) = \theta/(1 - \theta)$ .
- c)  $\tau(\theta) = \theta \cdot (1 - \theta)$ .
121. **Distribución uniforme.** Considere la distribución  $\text{unif}[-\theta, 2\theta]$ , en donde  $\theta > 0$  es desconocido y se desea estimar a través de una muestra aleatoria  $X_1, \dots, X_n$ . Encuentre el estimador para  $\theta$  por el método de máxima verosimilitud.
122. **Distribución doble exponencial.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución doble exponencial que aparece especificada

abajo, en donde  $\theta > 0$  es un parámetro desconocido. Encuentre el estimador para  $\theta$  por el método de máxima verosimilitud.

$$f(x, \theta) = \frac{1}{2} \theta e^{-\theta|x|} \quad -\infty < x < \infty.$$

123. **Distribución Rayleigh.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución Rayleigh, como aparece abajo, en donde  $\theta > 0$  es un parámetro no conocido. Encuentre el estimador para  $\theta$  por máxima verosimilitud.

$$f(x, \theta) = \begin{cases} \frac{2x}{\theta} e^{-x^2/\theta} & \text{si } x > 0, \\ 0 & \text{en otro caso.} \end{cases}$$

124. Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $f(x, \theta)$  que se especifica abajo, en donde  $\theta > 0$  es un parámetro desconocido. Encuentre el estimador por el método de máxima verosimilitud para el parámetro  $\theta$  y para la probabilidad  $P(X > 1)$ .

$$f(x : \theta) = \begin{cases} \frac{1}{\theta} e^{-x/\theta} & \text{si } x > 0, \\ 0 & \text{en otro caso.} \end{cases}$$

125. Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $f(x, \theta)$  que se especifica abajo, en donde  $\theta > -2$  es un parámetro desconocido. Encuentre el estimador por el método de máxima verosimilitud para el parámetro  $\theta$  y para la probabilidad  $P(a < X < b)$ , en donde  $0 < a < b$  son dos constantes conocidas.

$$f(x : \theta) = \begin{cases} (\theta + 2) e^{-(\theta+2)x} & \text{si } x > 0, \\ 0 & \text{en otro caso.} \end{cases}$$

126. Las siguientes distribuciones dependen de dos parámetros: uno desconocido denotado por la letra  $\theta$  y otro que supondremos conocido y que se denota por una letra distinta. Encuentre el estimador por el método de máxima verosimilitud para el parámetro desconocido  $\theta$ , suponiendo un tamaño de muestra  $n$ .

- |                                  |                                       |
|----------------------------------|---------------------------------------|
| a) $\text{bin}(k, \theta)$ .     | e) $\text{gama}(\gamma, \theta)$ .    |
| b) $\text{bin neg}(r, \theta)$ . | f) $N(\theta, \sigma^2)$ .            |
| c) $\text{unif}(a, \theta]$ .    | g) $N(\mu, \theta)$ .                 |
| d) $\text{unif}[\theta, b)$ .    | h) $\text{Weibull}(\alpha, \theta)$ . |

127. **Tres parámetros.** Sean  $X_1, \dots, X_n$  y  $Y_1, \dots, Y_m$  dos muestras aleatorias independientes, la primera de la distribución  $N(\mu_1, \sigma^2)$  y la segunda de la distribución  $N(\mu_2, \sigma^2)$ , en donde todos los parámetros son desconocidos. Observe que la varianza es la misma para ambas distribuciones y que los tamaños de muestra no son necesariamente iguales. Encuentre el estimador por el método de máxima verosimilitud para el vector de parámetros  $(\mu_1, \mu_2, \sigma^2)$ .

128. **Proceso de Poisson.** Un proceso de Poisson de parámetro  $\theta > 0$  es un proceso estocástico a tiempo continuo  $\{X_t : t \geq 0\}$  que satisface las siguientes propiedades.

- $X_0 = 0$  c.s.
- Tiene incrementos independientes.
- $X_t - X_s \sim \text{Poisson}(\theta(t - s))$ , para  $0 \leq s < t$ .

Suponga que el parámetro  $\theta$  es desconocido y que deseamos estimarlo a través de  $n$  observaciones  $X_{t_1}, \dots, X_{t_n}$  de una trayectoria del proceso, en donde  $0 < t_1 < \dots < t_n$  son tiempos fijos. Observe que las variables aleatorias observadas  $X_{t_1}, \dots, X_{t_n}$  no son independientes. Use el método de máxima verosimilitud para estimar  $\theta$ .

129. **Movimiento browniano.** Un movimiento browniano unidimensional de parámetro  $\theta > 0$  es un proceso estocástico a tiempo continuo  $\{B_t : t \geq 0\}$  que satisface las siguientes propiedades.
- $B_0 = 0$  c.s.
  - Las trayectorias son continuas.
  - Tiene incrementos independientes.
  - $B_t - B_s \sim N(0, \theta(t - s))$ , para  $0 \leq s < t$ .

Suponga que el parámetro  $\theta$  es desconocido y que deseamos estimarlo a través de  $n$  observaciones  $B_{t_1}, \dots, B_{t_n}$  de una trayectoria del proceso, en donde  $0 < t_1 < \dots < t_n$  son tiempos fijos. Observe que las variables aleatorias observadas  $B_{t_1}, \dots, B_{t_n}$  no son independientes. Use el método de máxima verosimilitud para estimar  $\theta$ .

## 2.4. Insesgamiento

Teniendo una o posiblemente varias estadísticas que pueden considerarse candidatas para ser usadas como estimadores para los parámetros desconocidos de una distribución de probabilidad, uno puede dedicarse a la tarea de estudiar sus propiedades a fin de escoger el mejor estimador posible. Pero, ¿qué características hacen que un estimador sea bueno? Hay varias respuestas a esta pregunta. En las siguientes secciones veremos que pueden establecerse varias buenas cualidades para un estimador.

Una primera buena propiedad que se le puede pedir a un estimador es que su valor promedio coincida con el parámetro a estimar. Esta idea se formaliza en la siguiente definición.

**Definición 2.10** Un estimador  $\hat{\theta}$  es **insesgado** para el parámetro  $\theta$  si cumple la condición

$$E(\hat{\theta}) = \theta.$$

Esta es una muy buena propiedad para un estimador, pues siendo un estimador una variable aleatoria, y si su objetivo es estimar el valor del parámetro, entonces es alentador saber que su valor promedio es justamente el valor a estimar. En los siguientes ejemplos mostraremos que es posible verificar esta propiedad de insesgamiento, a pesar de no conocer el valor del parámetro.

**Ejemplo 2.21** Comprobaremos que la media muestral es un estimador insesgado para el parámetro de la distribución Poisson( $\theta$ ). Por la propiedad de linealidad de la esperanza tenemos que

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \sum_{i=1}^n \theta = \theta.$$

De esta manera, sin conocer el valor de  $\theta$ , hemos comprobado que la esperanza del estimador  $\bar{X}$  es igual a  $\theta$ . ■

Es interesante observar que el cálculo desarrollado en el ejemplo anterior no depende de la distribución en estudio, de modo que podemos afirmar que la media muestral es siempre un estimador insesgado del posible parámetro o función parametral que pudiera aparecer en la esperanza de la distribución de interés. Por ejemplo, si la distribución en cuestión es  $\text{bin}(k, p)$ , entonces  $\bar{X}$  es un estimador insesgado para la función parametral  $kp$ .

Como uno puede imaginar, los estimadores insesgados no son necesariamente únicos. Pueden proponerse varias estadísticas que resulten ser estimadores insesgados para un mismo parámetro. Esto se muestra en el siguiente ejemplo.

**Ejemplo 2.22** Sea  $X_1, X_2, X_3$  una muestra aleatoria de tamaño  $n = 3$  de la distribución  $\text{Ber}(\theta)$ , con  $\theta > 0$  desconocido. Usando la propiedad de linealidad de la esperanza, se puede comprobar que todos los siguientes estimadores para  $\theta$  son insesgados.

a)  $\hat{\theta}_1 = X_1$ .

b)  $\hat{\theta}_2 = \frac{1}{3}(X_1 + 2X_2)$ .

$$\text{c) } \hat{\theta}_3 = \frac{1}{6} (X_1 + 2X_2 + 3X_3).$$

$$\text{d) } \hat{\theta}_4 = \frac{1}{3} (X_{(1)} + X_{(2)} + X_{(3)}).$$

■

La situación mostrada en el ejemplo anterior plantea ahora el problema de determinar cuándo un estimador insesgado es mejor que otro estimador insesgado. Regresaremos a este problema más adelante. Por ahora seguiremos estudiando cuestiones relativas al insesgamiento. El siguiente es un ejemplo menos evidente e importante de insesgamiento.

**Ejemplo 2.23** Consideremos dada una muestra aleatoria de tamaño  $n$  de la distribución  $N(\mu, \theta)$ , en donde la varianza  $\theta > 0$  es desconocida y es el parámetro que nos interesa estimar. Podemos suponer que el parámetro  $\mu$  es conocido aunque esta hipótesis no es relevante en el siguiente análisis. Recordemos que la varianza muestral es una estadística definida como sigue

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

Comprobaremos que  $S^2$  es un estimador insesgado para  $\theta$ . Esta es la razón por la que aparece el término  $n-1$  como denominador en la definición de varianza muestral, y no  $n$ , como uno inicialmente supondría. Tenemos que

$$\begin{aligned} E(S^2) &= E\left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right) \\ &= \frac{1}{n-1} \sum_{i=1}^n E(X_i^2) - 2E(X_i \bar{X}) + E(\bar{X}^2). \end{aligned} \quad (2.4)$$

Se puede comprobar que

$$E(X_i X_j) = \begin{cases} \mu^2 & \text{si } i \neq j, \\ \theta + \mu^2 & \text{si } i = j. \end{cases}$$

Substituyendo estas expresiones en (2.4) y simplificando se comprueba que  $E(S^2) = \theta$ . Es decir,  $S^2$  es un estimador insesgado para  $\theta$ . ■

Observamos nuevamente que los cálculos anteriores son válidos para cualquier distribución con segundo momento finito, no únicamente para la distribución normal. Hemos usado únicamente la propiedad de la linealidad de la esperanza y las hipótesis de independencia e idéntica distribución de las variables de la muestra aleatoria.

Así, la varianza muestral es siempre un estimador insesgado del posible parámetro o función parametral que pudiera aparecer en la varianza de la distribución de interés. Por ejemplo, si la distribución en cuestión es  $\text{bin}(k, p)$ , entonces  $S^2$  es un estimador insesgado para la función parametral  $kp(1 - p)$ .

En la sección de ejercicios aparecen varios ejemplos en donde se verifica que ni el método de momentos, ni el método de máxima verosimilitud producen necesariamente estimadores que cumplen la propiedad de insesgamiento.

### Insesgamiento para funciones parametrales

Como hemos mostrado antes, el concepto de insesgamiento se aplica no sólo para un parámetro de una distribución de probabilidad, sino también para funciones parametrales. Aquí tenemos entonces una extensión evidente de la definición de insesgamiento dada anteriormente.

**Definición 2.11** Sea  $\theta$  un parámetro o un vector de parámetros y sea  $\tau(\theta)$  una función parametral. Una estadística  $T$  es un **estimador insesgado** para  $\tau(\theta)$  si

$$E(T) = \tau(\theta).$$

Por ejemplo, hemos mostrado que la media muestral es siempre un estimador insesgado para la media de la distribución y que la varianza muestral es insesgado para la varianza de la distribución. Hemos mencionado en los ejemplos anteriores el caso de la distribución binomial. Podemos ahora considerar la distribución  $\text{unif}(a, b)$  y afirmar que el estimador  $\bar{X}$  es insesgado para la función parametral media  $(a+b)/2$ , y que el estimador  $S^2$  es también insesgado para la función parametral varianza  $(b-a)^2/12$ .

## Funciones de estimadores insesgados

Sea  $\hat{\theta}$  un estimador insesgado para un parámetro  $\theta$  y sea  $\varphi$  una función dada, con un dominio de definición adecuado. Nos interesa considerar la estadística  $\varphi(\hat{\theta})$  y el problema es el siguiente: ¿se preserva el insesgamiento bajo transformaciones? Es decir, nos preguntamos si  $\varphi(\hat{\theta})$  es un estimador insesgado para  $\varphi(\theta)$ . La respuesta es, en general, negativa e ilustraremos esto con un ejemplo. Sea  $\varphi(x) = x^2$ . Aplicaremos esta función al estimador insesgado  $\hat{\theta} = \bar{X}$  para el parámetro de la distribución Poisson( $\theta$ ). Tenemos que

$$\begin{aligned}
 E(\hat{\theta}^2) &= E\left(\sum_{i=1}^n X_i\right)^2 \\
 &= \frac{1}{n^2} \sum_{i=1}^n E(X_i^2) + \frac{1}{n^2} \sum_{i \neq j} E(X_i X_j) \\
 &= \frac{n}{n^2}(\theta + \theta^2) + \frac{n(n-1)}{n^2} \theta^2 \\
 &= \frac{\theta}{n} + \theta^2 \\
 &\neq \theta^2.
 \end{aligned}$$

Es decir,  $\hat{\theta}^2$  no es insesgado para  $\theta^2$ . Este hecho es consecuencia de que, en general,  $E(\varphi(\cdot)) \neq \varphi(E(\cdot))$ . Sin embargo, es interesante observar que en este ejemplo en particular se cumple que  $E(\hat{\theta}^2) \rightarrow \theta^2$  cuando  $n \rightarrow \infty$ . A esta propiedad límite de los estimadores le llamaremos insesgamiento asintótico. Ese es el tema que estudiaremos en la siguiente sección y corresponde a una propiedad más débil que el insesgamiento.

Regresando al tema en estudio, dado que la respuesta a la pregunta arriba planteada es negativa, surge de manera inmediata otra pregunta: ¿bajo qué condiciones sobre una transformación se preserva el insesgamiento? Tal transformación debe satisfacer  $E(\varphi(\hat{\theta})) = \varphi(\theta) = \varphi(E(\hat{\theta}))$ . Es decir, la transformación  $\varphi$  debe satisfacer la identidad  $E(\varphi(\cdot)) = \varphi(E(\cdot))$ . Esta ecuación se cumple en muy pocos casos. En efecto, si consideramos que el estimador en cuestión no es constante, entonces se puede comprobar que  $\varphi$

debe ser una función lineal necesariamente. Así, omitiendo los casos triviales de estimadores constantes, únicamente para transformaciones lineales se preserva el insesgamiento de manera general.

## Ejercicios

130. Sea  $\hat{\theta}_n$  un estimador para un parámetro  $\theta$  de una distribución. Demuestre que si  $\hat{\theta}_n \xrightarrow{m} \theta$ , entonces  $\hat{\theta}_n$  es asintóticamente insesgado.
131. **Distribución Bernoulli.** Sea  $X_1, X_2$  una muestra aleatoria de tamaño  $n = 2$  de la distribución Bernoulli con parámetro desconocido  $\theta$ . Demuestre que el siguiente estimador es insesgado para  $\theta$ .

$$\hat{\theta} = \frac{X_{(1)} + X_{(2)}}{2}.$$

132. **Distribución Bernoulli.** Sabemos que  $\hat{\theta} = \bar{X}$  es un estimador insesgado para el parámetro  $\theta$  de la distribución Bernoulli. Demuestre que el estimador  $\hat{\theta}(1 - \hat{\theta})$  no es insesgado para la varianza de esta distribución. Este es otro ejemplo que muestra que el insesgamiento no se preserva bajo transformaciones. Proponga un estimador insesgado para la varianza.
133. **Distribución binomial.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{bin}(k, \theta)$ , en donde el número de ensayos  $k$  es conocido y la probabilidad  $\theta$  es desconocida. Demuestre que los siguientes estimadores son insesgados para el parámetro  $\theta$ .

$$a) \hat{\theta} = \frac{1}{k} X_1$$

$$b) \hat{\theta} = \frac{1}{kn} (X_1 + \dots + X_n).$$

134. **Distribución binomial.** Sean  $X_1, \dots, X_m$  variables aleatorias independientes tal que la  $k$ -ésima variable aleatoria tiene distribución  $\text{bin}(n_k, \theta)$ , para  $k = 1, \dots, m$ . Suponga que los parámetros  $n_1, \dots, n_m$  son conocidos y  $\theta$  es desconocido. Determine si los siguientes estimadores son insesgados para  $\theta$ .

$$a) \hat{\theta} = \frac{X_1 + \cdots + X_m}{n_1 + \cdots + n_m}.$$

$$b) \hat{\theta} = \frac{X_1 + 2X_2 + \cdots + mX_m}{n_1 + 2n_2 + \cdots + mn_m}.$$

135. **Distribución geométrica.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{geo}(\theta)$ , en donde  $\theta$  es desconocido.

- a) Los estimadores para  $\theta$  por el método de momentos y por el método de máxima verosimilitud coinciden y aparece especificado abajo. Demuestre que este estimador no es insesgado.

$$\hat{\theta} = \frac{1}{1 + \bar{X}}.$$

Esto demuestra que el método de momentos y el método de máxima verosimilitud no garantizan la propiedad de insesgamiento.

- b) Demuestre que el siguiente estimador es insesgado para  $\theta$ . Suponga  $n \geq 2$ .

$$\hat{\theta} = \frac{1}{1 + \frac{n}{n-1} \bar{X}}.$$

136. **Distribución binomial negativa.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{bin neg}(r, \theta)$ , en donde la probabilidad  $\theta$  es desconocida y  $r \geq 1$  es un entero conocido.

- a) Los estimadores para  $\theta$  por el método de momentos y por el método de máxima verosimilitud coinciden y aparece especificado abajo. Demuestre que este estimador no es insesgado.

$$\hat{\theta} = \frac{r}{r + \bar{X}}.$$

Este es otro ejemplo en donde se muestra que el método de momentos y el método de máxima verosimilitud no garantizan el insesgamiento.

- b) Demuestre que el siguiente estimador es insesgado para  $\theta$ . Suponga  $nr \geq 2$ .

$$\hat{\theta} = \frac{1}{1 + \frac{n}{nr-1} \bar{X}}.$$

137. **Distribución exponencial.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\exp(\theta)$ , en donde  $\theta$  es desconocido. Aplicando del método de momentos o bien el método de máxima verosimilitud, se obtiene el estimador que aparece abajo. Demuestre que este estimador no es insesgado. Proponga un estimador insesgado.

$$\hat{\theta} = \frac{1}{\bar{X}}.$$

138. **Distribución doble exponencial.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución doble exponencial de parámetro desconocido  $\theta > 0$ .

$$f(x, \theta) = \frac{\theta}{2} e^{-\theta|x|}, \quad -\infty < x < \infty.$$

Demuestre que el estimador por máxima verosimilitud  $\hat{\theta}$ , que aparece abajo, no es insesgado. Proponga un estimador insesgado.

$$\hat{\theta} = \frac{1}{\frac{1}{n} \sum_{i=1}^n |X_i|}.$$

139. **Distribución normal.** Sea  $X_1, \dots, X_4$  una muestra aleatoria de tamaño  $n = 4$  de la distribución  $N(\theta, \sigma^2)$ , en donde la media  $\theta$  es desconocida y la varianza  $\sigma^2$  es conocida. Se proponen los siguientes estimadores para  $\theta$ . Determine cuál de ellos es el mejor en el sentido de ser insesgado y tener varianza menor.

a)  $\hat{\theta}_1 = X_1.$

b)  $\hat{\theta}_2 = X_1 + X_4.$

c)  $\hat{\theta}_3 = (X_1 + X_4)/2.$

d)  $\hat{\theta}_4 = (X_1 + X_4)/3.$

e)  $\hat{\theta}_5 = \bar{X}.$

f)  $\hat{\theta}_6 = X_1 + \hat{\theta}_2 - X_4.$

g)  $\hat{\theta}_7 = \frac{1}{6}(3X_1 + 2X_2 + X_3).$

h)  $\hat{\theta}_8 = \frac{1}{10} \sum_{i=1}^4 iX_i.$

140. **Distribución normal.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $N(\theta, \sigma^2)$  con  $\theta$  y  $\sigma^2$  desconocidos. Demuestre que la estadística que se define a continuación es un estimador insesgado para el parámetro  $\theta$ .

$$\hat{\theta} = \frac{2X_1 + 4X_2 + \dots + 2nX_n}{n(n+1)}.$$

141. **Distribución Rayleigh.** El estimador por máxima verosimilitud para el parámetro  $\theta$  de la distribución Rayleigh, que se especifica abajo, es  $\hat{\theta} = (1/n) \sum_{i=1}^n X_i^2$ . Demuestre que este estimador es insesgado.

$$f(x, \theta) = \begin{cases} \frac{2x}{\theta} e^{-x^2/\theta} & \text{si } x > 0, \\ 0 & \text{en otro caso.} \end{cases}$$

142. Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $f(x, \theta)$  que se especifica abajo, en donde  $\theta$  es un parámetro desconocido. Demuestre que  $\bar{X}$  no es un estimador insesgado para  $\theta$ . Construya uno que lo sea.

$$f(x, \theta) = \begin{cases} e^{-(x-\theta)} & \text{si } x > \theta, \\ 0 & \text{en otro caso.} \end{cases}$$

143. Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución que aparece abajo, en donde  $\theta > 0$  es un parámetro desconocido. Sea  $\hat{\theta}$  el estimador por máxima verosimilitud. Demuestre que  $\hat{\theta}$  no es un estimador insesgado para  $\theta$ . Proponga uno que lo sea.

$$f(x, \theta) = \begin{cases} (\theta + 1)x^\theta & \text{si } 0 < x < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

144. Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución que aparece abajo, en donde  $\theta$  es una constante arbitraria desconocida. Demuestre que  $\bar{X}$  no es un estimador insesgado para  $\theta$ . Construya uno que lo sea.

$$f(x) = \begin{cases} e^{-(x-\theta)} & \text{si } x > \theta, \\ 0 & \text{en otro caso.} \end{cases}$$

145. Considere dada una muestra aleatoria de la distribución  $f(x, \theta)$  como aparece especificada abajo, en donde  $-1 < \theta < 1$  es un parámetro desconocido. Demuestre que el estimador por el método de momentos dado por  $\hat{\theta} = (5/n) \sum_{i=1}^n X_i^3$  es insesgado.

$$f(x, \theta) = \begin{cases} \frac{1 + \theta x}{2} & \text{si } -1 < x < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

146. Considere dada una muestra aleatoria de la distribución  $f(x, \theta)$  como aparece especificada abajo, en donde  $\theta > 0$  es un parámetro desconocido. Demuestre que el estimador por el método de momentos dado por  $\hat{\theta} = 3\bar{X}$  es insesgado.

$$f(x, \theta) = \begin{cases} \frac{2(\theta - x)}{\theta^2} & \text{si } 0 < x < \theta, \\ 0 & \text{en otro caso.} \end{cases}$$

147. Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $f(x, \theta)$  que se especifica abajo, en donde  $\theta > 0$  es desconocido.

$$f(x, \theta) = \begin{cases} \theta x^{\theta-1} & \text{si } 0 < x < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

- a) Demuestre que  $-\ln X_i$  tiene distribución  $\exp(\theta)$ .
  - b) Demuestre que el estimador por el método de máxima verosimilitud  $\hat{\theta} = -n / \sum_{i=1}^n \ln X_i$  no es insesgado.
  - c) Con base en el inciso anterior, encuentre un estimador insesgado para  $\theta$ .
148. Sea  $X_1, \dots, X_n$  una muestra aleatoria de una población con media conocida  $\mu$  y varianza desconocida  $\theta$ . Demuestre que el siguiente estimador es insesgado para  $\theta$ .

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2.$$

149. **Combinación lineal convexa de estimadores insesgados.** Sean  $\hat{\theta}_1$  y  $\hat{\theta}_2$  dos estimadores insesgados para un parámetro  $\theta$ . Demuestre que, para cualquier valor real de  $\alpha$ , el siguiente estimador también es insesgado para  $\theta$ .

$$\hat{\theta} = \alpha \hat{\theta}_1 + (1 - \alpha) \hat{\theta}_2.$$

150. **Distribución normal.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una población con distribución  $N(0, \theta)$ , con  $\theta > 0$  desconocido. Demuestre que el siguiente estimador es insesgado para  $\theta$ .

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n X_i^2.$$

151. Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución dependiente de un parámetro desconocido  $\theta$  y cuya media es este mismo parámetro. Considere la estadística

$$T = \varphi_1(X_1) \cdots \varphi_n(X_n),$$

en donde  $\varphi_1, \dots, \varphi_n$  son funciones lineales de coeficientes conocidos. Demuestre que  $T$  es insesgado para la función parametral

$$\tau(\theta) = \varphi_1(\theta) \cdots \varphi_n(\theta).$$

152. **Función de un estimador insesgado no es necesariamente insesgado.** Sabemos que  $\hat{\theta} = \bar{X}$  es un estimador insesgado para el parámetro  $\theta$  de la distribución Bernoulli. Demuestre directamente que  $\hat{\theta}(1 - \hat{\theta})$  no es insesgado para la varianza de esta distribución pero es, sin embargo, asintóticamente insesgado.

153. Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución con función de densidad o de probabilidad  $f(x, \theta)$ , cuya media es el parámetro  $\theta$ , considerado desconocido. Sea  $\mathcal{E}$  el espacio de todos los estimadores lineales para  $\theta$  como se especifica abajo. Demuestre que  $\bar{X}$  es el único elemento de  $\mathcal{E}$  que es insesgado y tiene varianza mínima.

$$\mathcal{E} = \{a_1 X_1 + \cdots + a_n X_n : a_1, \dots, a_n \in \mathbb{R}\}.$$

154. Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución con función de densidad o de probabilidad  $f(x, \theta)$  dependiente de un parámetro desconocido  $\theta$  y tal que su media es este mismo parámetro. Sean  $a_1, \dots, a_n$  constantes cualesquiera tales que  $a_1 + \dots + a_n \neq 0$ . Demuestre que el siguiente estimador es insesgado para  $\theta$ .

$$\hat{\theta} = \frac{a_1 X_1 + a_2 X_2 + \dots + a_n X_n}{a_1 + \dots + a_n}.$$

155. **Proceso de Poisson.** En el ejercicio 128 se pide encontrar el estimador máximo verosímil para el parámetro  $\theta$  del proceso de Poisson. Demuestre que este estimador, el cual aparece especificado abajo, es insesgado.

$$\hat{\theta} = \frac{X_{t_n}}{t_n}.$$

156. **Movimiento browniano.** En el ejercicio 129 se pide encontrar el estimador máximo verosímil para el parámetro  $\theta$  del movimiento browniano. Demuestre que este estimador, el cual aparece especificado abajo, es insesgado.

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n \frac{(B_{t_i} - B_{t_{i-1}})^2}{t_i - t_{i-1}}.$$

157. **El insesgamiento no se preserva bajo transformaciones.** Sea  $\tau(\theta)$  una función parametral. Si  $\hat{\theta}$  es un estimador insesgado para  $\theta$ , entonces no necesariamente  $\tau(\hat{\theta})$  es insesgado para  $\tau(\theta)$ . Compruebe esta afirmación en el caso del estimador insesgado  $\hat{\theta} = \bar{X}$  para el parámetro de la distribución  $\text{Ber}(\theta)$  y la función parametral  $\tau(\theta) = \theta^2$ .

## 2.5. Insesgamiento asintótico

Si un estimador  $\hat{\theta}$  para un parámetro desconocido  $\theta$  no es insesgado, entonces se dice que es sesgado y a la diferencia  $E(\hat{\theta}) - \theta$  se le llama sesgo. Es posible que este sesgo pueda hacerse cada vez más pequeño conforme el tamaño de la muestra  $n$  crece. Si en el límite cuando  $n \rightarrow \infty$  el sesgo se hace cero, entonces se dice que el estimador es asintóticamente insesgado. Antes de escribir el enunciado formal de esta definición debemos mencionar que escribiremos

$\hat{\theta}_n$  en lugar de  $\hat{\theta}$  cuando deseemos enfatizar la posible dependencia de un estimador del tamaño  $n$  de la muestra aleatoria. Aquí tenemos la definición.

**Definición 2.12** Una estadística  $\hat{\theta}_n$ , basada en una muestra aleatoria de tamaño  $n$ , es un **estimador asintóticamente insesgado** para un parámetro  $\theta$  si

$$\lim_{n \rightarrow \infty} E(\hat{\theta}_n) = \theta. \quad (2.5)$$

Es claro que todo estimador insesgado es asintóticamente insesgado pues la condición (2.5) se cumple sin necesidad de tomar el límite. Por otro lado, más adelante tendremos múltiples oportunidades de mostrar que existen estimadores asintóticamente insesgados que no son insesgados. Estas dos relaciones generales se muestran gráficamente en la Figura 2.6.

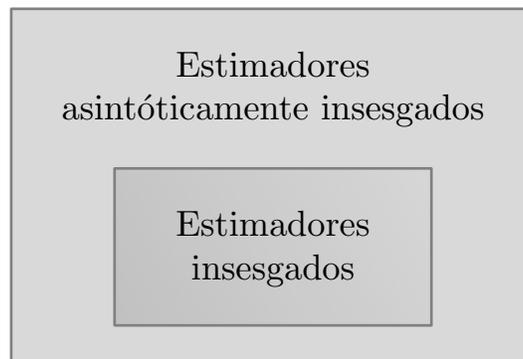


Figura 2.6

A continuación se presenta un ejemplo de insesgamiento asintótico. En la sección de ejercicios se encuentran muchos otros ejemplos.

**Ejemplo 2.24** Consideremos nuevamente el caso de la distribución  $N(\mu, \theta)$ , en donde la varianza  $\theta > 0$  desconocida. Defina el estimador

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Puede comprobarse que  $\hat{\theta}_n$  no es insesgado para  $\theta$  pero es asintóticamente insesgado, pues

$$E(\hat{\theta}_n) = E\left(\frac{n-1}{n} S^2\right) = \frac{n-1}{n} E(S^2) = \frac{n-1}{n} \theta \xrightarrow{n \rightarrow \infty} \theta.$$

De esta manera, aunque  $\hat{\theta}_n$  no cumple la propiedad de ser insesgado, su valor promedio no dista demasiado del valor del parámetro a estimar cuando el tamaño de la muestra es grande. ■

### Funciones de estimadores asintóticamente insesgados

Sea  $\hat{\theta}_n$  un estimador asintóticamente insesgado para un parámetro  $\theta$ , construido a partir de una muestra aleatoria de tamaño  $n$ , y sea  $\varphi$  una función dada, con dominio de definición adecuado. La pregunta que nos planteamos es la siguiente: ¿Se preserva el insesgamiento asintótico bajo transformaciones? Es decir, nos preguntamos si  $\varphi(\hat{\theta}_n)$  también es un estimador asintóticamente insesgado para  $\varphi(\theta)$ . La respuesta es, en general, negativa. Resulta que la propiedad de insesgamiento asintótico no se preserva bajo transformaciones y no es muy difícil dar un ejemplo de esta situación. Considere la función  $\varphi(x) = x^2$  aplicada al estimador insesgado  $\hat{\theta}_n = (X_1 + X_n)/2$  para el parámetro  $\theta$  de la distribución Poisson. Siendo  $\hat{\theta}_n$  insesgado, es asintóticamente insesgado. Sin embargo,  $E(\hat{\theta}_n^2)$  no converge a  $\theta^2$  pues se puede comprobar que  $E(\hat{\theta}_n^2) = \theta^2 + \theta/2$ .

Anteriormente habíamos mencionado que, en general, el insesgamiento se preserva únicamente bajo transformaciones lineales. Como todo estimador insesgado es asintóticamente insesgado, la misma afirmación se cumple para la preservación del insesgamiento asintótico.

### Ejercicios

158. **Distribución Bernoulli.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución  $\text{Ber}(\theta)$ , con  $\theta$  desconocido. Demuestre que el estimador  $\bar{X}(1 - \bar{X})$  es asintóticamente insesgado para la varianza de esta distribución.

159. **Distribución Poisson.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución  $\text{Poisson}(\theta)$ , en donde  $\theta > 0$  es desconocido. Demuestre que  $\bar{X}^2$  es asintóticamente insesgado para  $\theta^2$ .
160. **Distribución uniforme.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución  $\text{unif}(0, \theta)$ , en donde  $\theta > 0$  es desconocido.
- Demuestre que el estimador  $\hat{\theta}_n = X_{(n)}$  no es insesgado para  $\theta$ , sin embargo, es asintóticamente insesgado.
  - Encuentre un estimador insesgado para  $\theta$ .

161. **Distribución exponencial.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución  $\text{exp}(\theta)$ , con  $\theta$  desconocido. Demuestre que el estimador por máxima verosimilitud que aparece abajo es asintóticamente insesgado.

$$\hat{\theta}_n = \frac{1}{\bar{X}}.$$

162. Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución dependiente de un parámetro desconocido  $\theta$ , cuya media es este mismo parámetro y con segundo momento finito. Demuestre que la estadística  $\bar{X}^2$  es un estimador asintóticamente insesgado para  $\theta^2$ .
163. **Máxima verosimilitud no implica insesgamiento.** Sabemos que el estimador máximo verosímil para el parámetro  $\theta$  de la distribución exponencial es  $\hat{\theta} = 1/\bar{X}$ . Demuestre que  $\hat{\theta}$  no es insesgado pero es asintóticamente insesgado.
164. Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución con función de densidad o de probabilidad  $f(x, \theta)$  como aparece abajo, en donde  $\theta$  es un parámetro desconocido y con valores reales. Demuestre que el estimador por máxima verosimilitud  $\hat{\theta} = X_{(1)}$  no es insesgado pero es asintóticamente insesgado para  $\theta$ .

$$f(x, \theta) = \begin{cases} e^{-(x-\theta)} & \text{si } x \geq \theta, \\ 0 & \text{en otro caso.} \end{cases}$$

165. Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $f(x, \theta)$  que se especifica abajo, en donde  $\theta > 0$  es desconocido. Sabemos que el estimador por el método de máxima verosimilitud  $\hat{\theta}_n = -n / \sum_{i=1}^n \ln X_i$

no es insesgado. Véase el ejercicio 147. Demuestre que  $\hat{\theta}_n$  es asintóticamente insesgado.

$$f(x, \theta) = \begin{cases} \theta x^{\theta-1} & \text{si } 0 < x < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

## 2.6. Consistencia

Otra manera de medir la bondad de un estimador es a través de la consistencia. Esta propiedad establece la convergencia en probabilidad del estimador al parámetro a estimar cuando el tamaño de la muestra crece a infinito.

**Definición 2.13** Sea  $\hat{\theta}_n$  un estimador para  $\theta$  basado en una muestra aleatoria de tamaño  $n$ . Se dice que  $\hat{\theta}_n$  es **consistente** para  $\theta$  si  $\hat{\theta}_n \rightarrow \theta$  en probabilidad, cuando  $n \rightarrow \infty$ . Esto es, para cualquier  $\epsilon > 0$ ,

$$\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| > \epsilon) = 0.$$

De esta manera, la cercanía del estimador al parámetro se define en el sentido de la convergencia en probabilidad y se usa la notación  $\hat{\theta}_n \xrightarrow{p} \theta$ . Observe nuevamente que hemos añadido el tamaño de la muestra  $n$  como subíndice en el estimador para enfatizar su dependencia implícita, o explícita, de esta cantidad. Veamos un ejemplo de la propiedad de consistencia.

**Ejemplo 2.25** Sea  $X$  una variable aleatoria con función de densidad o de probabilidad  $f(x, \theta)$ , dependiente de un parámetro desconocido  $\theta$ , el cual se desea estimar a través de una muestra aleatoria de tamaño  $n$ . Supongamos que  $E(X) = \theta$ . Tal situación se presenta, por ejemplo, en la distribución Bernoulli, la distribución Poisson, o la distribución normal. Entonces, por la ley débil de los grandes números, el estimador  $\hat{\theta}_n = \bar{X}$  es consistente para  $\theta$  pues, cuando  $n \rightarrow \infty$ ,

$$\hat{\theta}_n \xrightarrow{p} \theta.$$

■

En general, puede ser una tarea complicada demostrar la convergencia en probabilidad de una sucesión cualquiera de variables aleatorias. Sin embargo, cuando el límite es una constante, en este caso el parámetro a estimar, tenemos el siguiente criterio para demostrar la propiedad de consistencia.

**Proposición 2.1 (Criterio para consistencia)** Sea  $\hat{\theta}_n$  un estimador para  $\theta$ , basado en una muestra aleatoria de tamaño  $n$ . Si

a)  $\lim_{n \rightarrow \infty} E(\hat{\theta}_n) = \theta$  y

b)  $\lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}_n) = 0,$

entonces  $\hat{\theta}_n$  es consistente.

**Demostración.** Se usa la siguiente versión de la desigualdad de Chebyshev: para cualquier  $\epsilon > 0$  y cualquier número real  $a$ ,

$$P(|X - a| > \epsilon) \leq \frac{1}{\epsilon^2} E(X - a)^2.$$

Entonces

$$\begin{aligned} P(|\hat{\theta}_n - \theta| > \epsilon) &\leq \frac{1}{\epsilon^2} E(\hat{\theta}_n - \theta)^2 \\ &= \frac{1}{\epsilon^2} E((\hat{\theta}_n - E(\hat{\theta}_n)) + (E(\hat{\theta}_n) - \theta))^2 \\ &= \frac{1}{\epsilon^2} [\text{Var}(\hat{\theta}_n) + (E(\hat{\theta}_n) - \theta)^2] \\ &\rightarrow 0 \quad \text{cuando } n \rightarrow \infty. \end{aligned}$$

■

Es decir, si un estimador es asintóticamente insesgado y su varianza tiende a cero, entonces es consistente. En particular, cuando se desee probar la propiedad de consistencia para un estimador insesgado, es suficiente verificar que la varianza del estimador converge a cero.

**Ejemplo 2.26** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\exp(\theta)$ . El estimador máximo verosímil para  $\theta$  está dado por  $\hat{\theta}_n = 1/\bar{X}$ . Puede verificarse que este estimador no es insesgado pero es asintóticamente insesgado. Por lo tanto, para verificar la propiedad de consistencia es suficiente demostrar que  $\text{Var}(\hat{\theta}_n)$  tiende a cero cuando  $n$  tiende a infinito. Recordemos que  $X_1 + \dots + X_n$  tiene distribución gama( $n, \theta$ ). Entonces, llevando a cabo las integrales correspondientes puede comprobarse que

$$\begin{aligned} \text{Var}(1/\bar{X}) &= E\left(\frac{n^2}{(X_1 + \dots + X_n)^2}\right) - E^2\left(\frac{n}{X_1 + \dots + X_n}\right) \\ &= \frac{n^2}{(n-1)(n-2)}\theta^2 - \frac{n^2}{(n-1)^2}\theta^2 \\ &\rightarrow 0 \quad \text{cuando } n \rightarrow \infty. \end{aligned}$$

Concluimos que el estimador  $\hat{\theta}_n = 1/\bar{X}$  es, efectivamente, consistente. ■

Por otro lado, es útil recordar que cuando el límite de una sucesión de variables aleatorias es una constante, la convergencia en probabilidad es equivalente a la convergencia en distribución. Puede consultarse este resultado en [12]. Por lo tanto, tenemos que un estimador es consistente si converge en distribución al parámetro a estimar. Esto se escribe  $\hat{\theta}_n \xrightarrow{d} \theta$  y puede resultar un mecanismo alternativo más fácil de verificar para demostrar la propiedad de consistencia.

**Proposición 2.2 (Criterio para consistencia)** El estimador  $\hat{\theta}_n$  es consistente para el parámetro  $\theta$  si, y sólo si, para cualquier  $x \neq \theta$ ,

$$\lim_{n \rightarrow \infty} P(\hat{\theta}_n \leq x) = \begin{cases} 1 & \text{si } x > \theta, \\ 0 & \text{si } x < \theta. \end{cases}$$

**Demostración.** La convergencia en probabilidad y la convergencia en distribución son equivalentes cuando el límite es constante. La función de distribución de la variable aleatoria constante  $\theta$  toma únicamente los valores 1 y 0 como aparece en la expresión de la derecha. ■

Es difícil no preguntarse si existe alguna relación entre el insesgamiento y la consistencia de un estimador. ¿Podría una propiedad implicar la otra? La respuesta es negativa. En el diagrama de la Figura 2.7 se muestra gráficamente que pueden presentarse todas las posibles relaciones entre estos dos conceptos. En este diagrama se pueden identificar 5 regiones disjuntas y en la Tabla 2.4 se dan casos particulares de estimadores que pertenecen a cada una de estas regiones en el caso de estimación del parámetro de la distribución Bernoulli.

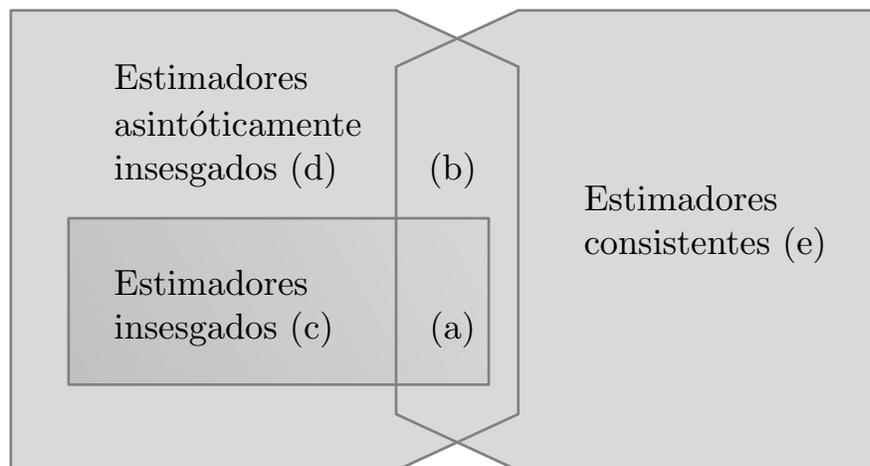


Figura 2.7

No es muy difícil comprobar las afirmaciones que aparecen en la Tabla 2.4. El inciso (e) se desarrolla en el siguiente ejemplo. En la sección de ejercicios se muestran algunos otros ejemplos de las situaciones generales mostradas en la Figura 2.7.

**Estimadores para el parámetro de la distribución  $\text{Ber}(\theta)$**

	Estimador	Insesgado	Asint. insesgado	Consistente
(a)	$\bar{X}$	si	si	si
(b)	$\frac{n}{n-1}\bar{X}$	no	si	si
(c)	$X_1$	si	si	no
(d)	$\frac{n}{n-1}X_1$	no	si	no
(e)	Ejemplo 2.27	no	no	si

Tabla 2.4

**Ejemplo 2.27 (La consistencia no implica el insesgamiento, ni el insesgamiento asintótico.)** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{Ber}(\theta)$ , con  $\theta$  desconocido. Sea  $Z$  otra variable aleatoria con distribución  $\text{Ber}(1/n)$  e independiente de las anteriores. Defina ahora el estimador

$$\hat{\theta}_n = \begin{cases} \bar{X} & \text{si } Z = 0, \\ n & \text{si } Z = 1. \end{cases}$$

Se comprueba que  $\hat{\theta}_n$  no es insesgado, ni asintóticamente insesgado, pues

$$\begin{aligned} E(\hat{\theta}_n) &= E(\hat{\theta}_n | Z = 0) P(Z = 0) + E(\hat{\theta}_n | Z = 1) P(Z = 1) \\ &= \theta \frac{n-1}{n} + 1 \\ &\rightarrow \theta + 1 \quad \text{cuando } n \rightarrow \infty. \end{aligned}$$

Sin embargo,  $\hat{\theta}_n$  es consistente pues para cualquier  $\epsilon > 0$ ,

$$\begin{aligned} P(|\hat{\theta}_n - \theta| > \epsilon) &= P(|\hat{\theta}_n - \theta| > \epsilon | Z = 0) P(Z = 0) \\ &\quad + P(|\hat{\theta}_n - \theta| > \epsilon | Z = 1) P(Z = 1) \\ &= P(|\bar{X} - \theta| > \epsilon) \frac{n-1}{n} + P(|n - \theta| > \epsilon) \frac{1}{n} \\ &\rightarrow 0 \quad \text{cuando } n \rightarrow \infty. \end{aligned}$$

■

Para concluir esta sección y a manera de resumen de las definiciones de insesgamiento, insesgamiento asintótico y consistencia, tenemos la tabla que aparece abajo.

Condición	Propiedad
$E(\hat{\theta}_n) = \theta$	Insesgamiento
$E(\hat{\theta}_n) \rightarrow \theta$	Insesgamiento asintótico
$\hat{\theta}_n \xrightarrow{p} \theta$	Consistencia

Tabla 2.5

Observe que cualquier otro criterio de convergencia o acercamiento de un estimador al parámetro desconocido puede ser considerado como una propiedad deseable y eso genera otras posibles cualidades adicionales para el estimador.

## Ejercicios

166. **Convergencia en distribución.** Demuestre que un estimador  $\hat{\theta}_n$  es consistente para el parámetro  $\theta$  si, y sólo si,  $\hat{\theta}_n$  converge en distribución a la constante  $\theta$ .
167. **Propiedades de la convergencia en probabilidad a una constante.** Sea  $\hat{\theta}_n$  un estimador consistente para  $\theta$ . Demuestre que
- $a\hat{\theta}_n + b$  es consistente para  $a\theta + b$ . Suponga  $a \neq 0$ .
  - $|\hat{\theta}_n|$  es consistente para  $|\theta|$ .
  - $\hat{\theta}_n^2$  es consistente para  $\theta^2$ .
  - $e^{\hat{\theta}_n}$  es consistente para  $e^\theta$ .
  - $1/\hat{\theta}_n$  es consistente para  $1/\theta$ , suponiendo  $\hat{\theta}_n \neq 0, \theta \neq 0$ .

En el siguiente ejercicio se generalizan estos resultados.

168. **Funciones continuas de estimadores consistentes.** Sea  $\hat{\theta}_n$  un estimador consistente para  $\theta$  y sea  $\varphi$  una función continua con dominio adecuado. Demuestre que  $\varphi(\hat{\theta}_n)$  es consistente para la función parametral  $\varphi(\theta)$ .
169. **Desigualdad de Chebyshev.** Sea  $X$  una variable aleatoria con segundo momento finito. Demuestre que para cualquier  $\epsilon > 0$  y cualquier número real  $a$ ,

$$P(|X - a| > \epsilon) \leq \frac{1}{\epsilon^2} E(X - a)^2.$$

Cuando se toma  $a = E(X)$ , se obtiene

$$P(|X - E(X)| > \epsilon) \leq \frac{1}{\epsilon^2} \text{Var}(X).$$

170. **Distribución Bernoulli.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{Ber}(\theta)$ , con  $\theta$  desconocido. Demuestre que  $\bar{X}$  es un estimador consistente para  $\theta$ .
171. **Distribución uniforme.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{unif}(0, \theta)$ , con parámetro  $\theta > 0$  desconocido. Demuestre que  $\max\{X_1, \dots, X_n\}$  es un estimador consistente para  $\theta$ .
172. **Distribución normal.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $N(\mu, \sigma^2)$ , en donde tanto  $\mu$  como  $\sigma^2$  son desconocidos. Demuestre que el estimador  $\hat{\sigma}^2$  que aparece abajo no es insesgado pero es consistente para  $\sigma^2$ .

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

173. Sean  $X_1, \dots, X_n$  variables aleatorias independientes tal que la  $i$ -ésima variable tiene distribución  $\text{bin}(k_i, \theta)$ . Suponga que los parámetros  $k_1, \dots, k_n$  son conocidos, pero  $\theta$  es desconocido. Es inmediato comprobar que los siguientes estimadores son insesgados para  $\theta$ . Demuestre ahora que son consistentes.

$$a) \hat{\theta}_n = \frac{X_1 + \dots + X_n}{k_1 + \dots + k_n}.$$

$$b) \hat{\theta}_n = \frac{X_1 + 2X_2 + \cdots + nX_n}{k_1 + 2k_2 + \cdots + nk_n}.$$

174. **Consistencia  $\not\Rightarrow$  Insesgamiento.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\exp(\theta)$ , con  $\theta > 0$  desconocido. Sabemos que el estimador  $\hat{\theta}$  que aparece abajo no es insesgado para  $\theta$ . Demuestre que  $\hat{\theta}$  es consistente. Este es un ejemplo de un estimador que es consistente pero no es insesgado.

$$\hat{\theta} = \frac{1}{\bar{X}}.$$

175. **Insesgamiento  $\not\Rightarrow$  Consistencia.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{Poisson}(\theta)$ , con  $\theta > 0$  desconocido. Demuestre que el estimador  $\hat{\theta} = (X_1 + X_n)/2$  es insesgado pero no es consistente para  $\theta$ .

176. **Insesgamiento  $\not\Rightarrow$  Consistencia.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\exp(\theta)$ , con  $\theta > 0$  desconocido. Se puede comprobar que el estimador que aparece abajo es insesgado. Demuestre ahora que no es consistente. Este es otro ejemplo de que la propiedad de insesgamiento no implica la consistencia.

$$\hat{\theta} = \frac{1}{1 + \frac{n}{n-1}\bar{X}}.$$

177. Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución que aparece abajo, en donde  $\theta > -1$  es un parámetro desconocido. Demuestre que el estimador por máxima verosimilitud  $\hat{\theta}_n = -1 - n/\sum_{i=1}^n \ln X_i$  es consistente.

$$f(x, \theta) = \begin{cases} (\theta + 1)x^\theta & \text{si } 0 < x < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

178. Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $f(x, \theta)$  como aparece especificada abajo, en donde  $\theta > 0$  es un parámetro desconocido. Demuestre que el estimador por el método de momentos  $\hat{\theta}_n = 3\bar{X}$  es consistente.

$$f(x, \theta) = \begin{cases} \frac{2(\theta - x)}{\theta^2} & \text{si } 0 < x < \theta, \\ 0 & \text{en otro caso.} \end{cases}$$

179. Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $N(\mu, \sigma^2)$ , con  $\mu$  y  $\sigma^2$  desconocidos. Defina la estadística

$$T = \frac{2X_1 + 4X_2 + \dots + 2nX_n}{n(n+1)}.$$

- a) Demuestre que  $T$  insesgado para  $\mu$ .
- b) Demuestre que  $T$  consistente para  $\mu$ .
- c) Determine si  $\max\{0, T\}$  es consistente para  $\mu$ .
180. **Distribución normal.** Demuestre que la varianza muestral  $S^2$  es un estimador consistente para la varianza desconocida  $\sigma^2$  de una distribución normal.
181. Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución con función de densidad o de probabilidad  $f(x, \theta)$  como aparece abajo, en donde  $\theta$  es un parámetro desconocido. Demuestre que  $\hat{\theta}_n = X_{(1)}$  es un estimador consistente para  $\theta$ .

$$f(x, \theta) = \begin{cases} e^{-(x-\theta)} & \text{si } x > \theta, \\ 0 & \text{en otro caso.} \end{cases}$$

182. Considere una distribución con función de densidad  $f(x, \theta)$  como se especifica abajo, en donde  $-1 < \theta < 1$  es un parámetro desconocido. Demuestre que el estimador por el método de momentos dado por  $\hat{\theta}_n = (5/n) \sum_{i=1}^n X_i^3$  es consistente.

$$f(x, \theta) = \begin{cases} \frac{1 + \theta x}{2} & \text{si } -1 < x < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

## 2.7. Sesgo y error cuadrático medio

En el siguiente enunciado formalizamos la definición de sesgo de un estimador que habíamos mencionado en la sección anterior.

**Definición 2.14** El **sesgo** de un estimador  $\hat{\theta}$  para un parámetro  $\theta$  es la diferencia

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta.$$

Aunque la notación que aparece en la definición establece que el sesgo es una operación que se aplica sobre el estimador  $\hat{\theta}$ , el sesgo es una cantidad que depende del valor del parámetro y por lo tanto es una función de éste. Puede ser positivo, negativo o cero. El signo no es relevante pues sólo nos interesa la diferencia entre  $E(\hat{\theta})$  y  $\theta$ . La letra  $B$  proviene del término en inglés *bias*, que se traduce como sesgo o desviación. Es claro que cuando el estimador es insesgado el sesgo es cero. Además, el estimador es asintóticamente insesgado si el sesgo tiende a cero cuando el tamaño de la muestra tiende a infinito.

**Ejemplo 2.28** Para la distribución exponencial con parámetro  $\theta > 0$  desconocido, se puede comprobar que el estimador  $\hat{\theta} = 1/\bar{X}$  no es insesgado pues  $E(\hat{\theta}) = n\theta/(n-1)$ . El sesgo en este caso es la función

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta = \frac{1}{n-1} \theta.$$

■

El sesgo es sólo una de varias maneras en las que se puede medir algún tipo de distancia entre el estimador y el parámetro a estimar. Otra de ellas es el error cuadrático medio que se define a continuación.

**Definición 2.15** Sea  $\hat{\theta}$  un estimador para un parámetro  $\theta$ . El **error cuadrático medio** de  $\hat{\theta}$  es

$$\text{ECM}(\hat{\theta}) = E(\hat{\theta} - \theta)^2.$$

De esta manera, el error cuadrático medio es la distancia cuadrática promedio entre el estimador y el parámetro a estimar, y resulta ser nuevamente una

función del parámetro a estimar. De la fórmula que aparece en la definición anterior, es claro que cuando el estimador es insesgado, el error cuadrático medio es la varianza del estimador, es decir,  $ECM(\hat{\theta}) = \text{Var}(\hat{\theta})$ . Por lo tanto, plantearse el problema de encontrar estimadores insesgados con el error cuadrático medio más pequeño equivale a encontrar estimadores insesgados de varianza mínima. Consideraremos este problema más adelante.

El sesgo y el error cuadrático medio están relacionados mediante las siguientes fórmulas.

**Proposición 2.3** Sea  $\hat{\theta}$  un estimador para un parámetro  $\theta$ . Entonces

1.  $ECM(\hat{\theta}) = \text{Var}(\hat{\theta}) + B^2(\hat{\theta})$ .
2.  $B^2(\hat{\theta}) \leq ECM(\hat{\theta})$ .

**Demostración.** El primer resultado se obtiene a través del análisis que aparece abajo. El segundo resultado es una consecuencia inmediata del primero.

$$\begin{aligned}
 ECM(\hat{\theta}) &= E(\hat{\theta} - \theta)^2 \\
 &= E[(\hat{\theta} - E(\hat{\theta})) + (E(\hat{\theta}) - \theta)]^2 \\
 &= E(\hat{\theta} - E(\hat{\theta}))^2 + 2E(\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta) + (E(\hat{\theta}) - \theta)^2 \\
 &= E(\hat{\theta} - E(\hat{\theta}))^2 + (E(\hat{\theta}) - \theta)^2 \\
 &= \text{Var}(\hat{\theta}) + B^2(\hat{\theta}).
 \end{aligned}$$

■

**Ejemplo 2.29** Considere la distribución  $\exp(\theta)$ , con  $\theta$  desconocido. Se puede comprobar que para el estimador  $\hat{\theta} = 1/\bar{X}$  se cumplen las fórmulas que

aparecen abajo y se verifican las relaciones generales de la proposición anterior. Observe que todas estas cantidades son funciones del parámetro  $\theta$ .

$$\begin{aligned} B(\hat{\theta}) &= \frac{1}{n-1} \theta, \\ \text{Var}(\hat{\theta}) &= \frac{n^2}{(n-1)(n-2)} \theta^2, \\ \text{ECM}(\hat{\theta}) &= \frac{n+2}{(n-1)(n-2)} \theta^2. \end{aligned}$$

■

## Ejercicios

183. Use la desigualdad de Jensen para demostrar, nuevamente, que

$$B^2(\hat{\theta}) \leq \text{ECM}(\hat{\theta}).$$

184. Demuestre las tres afirmaciones del Ejemplo 2.29.

185. **Criterio para la consistencia.** Sea  $\hat{\theta}_n$  un estimador para un parámetro desconocido  $\theta$ , basado en una muestra aleatoria de tamaño  $n$ . Demuestre que si  $\lim_{n \rightarrow \infty} \text{ECM}(\hat{\theta}_n) = 0$ , entonces  $\hat{\theta}_n$  es consistente.

En particular, cuando  $\hat{\theta}_n$  es insesgado,  $\text{ECM}(\hat{\theta}_n) = \text{Var}(\hat{\theta}_n)$  y la hipótesis se expresa como  $\lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}_n) = 0$ .

186. **Inssegamiento no implica ECM menor.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $N(\mu, \theta)$ , en donde la varianza  $\theta > 0$  es desconocida. Suponga  $n \geq 2$ . Se proponen los siguientes dos estimadores para  $\theta$ .

$$\begin{aligned} \hat{\theta}_1 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \\ \hat{\theta}_2 &= \frac{1}{n+1} \sum_{i=1}^n (X_i - \bar{X})^2. \end{aligned}$$

En el Ejemplo 2.23 se demostró que  $\hat{\theta}_1$  es insesgado para  $\theta$  y puede comprobarse que  $\hat{\theta}_2$  es sesgado. Demuestre, sin embargo, que

$$\text{ECM}(\hat{\theta}_2) < \text{ECM}(\hat{\theta}_1).$$

187. **Distribución Bernoulli.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{Ber}(\theta)$ , con  $\theta$  desconocido. Defina el estimador  $\hat{\theta} = \bar{X}$ . Encuentre

- |                                 |                                 |
|---------------------------------|---------------------------------|
| a) $E(\hat{\theta})$ .          | c) $B(\hat{\theta})$ .          |
| b) $\text{Var}(\hat{\theta})$ . | d) $\text{ECM}(\hat{\theta})$ . |

188. **Distribución Poisson.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{Poisson}(\theta)$ , con  $\theta$  desconocido. Defina el estimador  $\hat{\theta} = \bar{X}$ . Encuentre

- |                                 |                                 |
|---------------------------------|---------------------------------|
| a) $E(\hat{\theta})$ .          | c) $B(\hat{\theta})$ .          |
| b) $\text{Var}(\hat{\theta})$ . | d) $\text{ECM}(\hat{\theta})$ . |

## 2.8. Cota inferior de Cramér-Rao

En secciones anteriores hemos estudiado algunos métodos para encontrar posibles estimadores para un parámetro desconocido  $\theta$ . Hemos también establecido el insesgamiento como un primer criterio para determinar la bondad de un estimador y hemos mencionado algunas otras propiedades deseables.

Además del insesgamiento, una segunda buena propiedad de un estimador es que tenga varianza pequeña. Tales estimadores estarán centrados en el valor  $\theta$  y variarán lo menos posible alrededor de esa cantidad. Así, nos interesa buscar estimadores insesgados que tengan la varianza más pequeña posible.

El resultado interesante que estudiaremos a continuación establece que no es posible hacer que la varianza de un estimador insesgado sea arbitrariamente pequeña. En otras palabras, bajo ciertas condiciones generales, existe una

cota inferior para la varianza de cualquier estimador insesgado. Demostraremos este resultado para el problema general de estimar cualquier función parametral  $\tau(\theta)$ .

**Teorema 2.2 (Cota inferior de Cramér-Rao<sup>3</sup>)** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución con función de probabilidad o de distribución  $f(x, \theta)$ , dependiente de un parámetro desconocido  $\theta$ . Sea  $T$  un estimador insesgado para una función parametral  $\tau(\theta)$ . Bajo ciertas condiciones generales que especificaremos más adelante se cumple que

$$\text{Var}(T) \geq \frac{(\tau'(\theta))^2}{n E \left[ \left( \frac{\partial}{\partial \theta} \ln f(X, \theta) \right)^2 \right]}. \quad (2.6)$$

**Demostración.** En los siguientes cálculos llevaremos a cabo algunas operaciones cuya validez supondremos implícitamente. Haremos el análisis suponiendo, además, el caso de variables aleatorias continuas. El caso discreto se analiza de manera semejante.

Sea  $X$  una variable cualquiera de la muestra aleatoria. Como  $\int_{\mathbb{R}} f(x, \theta) dx = 1$ , derivando respecto de  $\theta$  y suponiendo válido el intercambio de la derivada y la integral se tiene que

$$\begin{aligned} 0 &= \frac{d}{d\theta} \int_{\mathbb{R}} f(x, \theta) dx \\ &= \int_{\mathbb{R}} \frac{\partial}{\partial \theta} f(x, \theta) dx \\ &= \int_{\mathbb{R}} \frac{\partial}{\partial \theta} e^{\ln f(x, \theta)} dx \\ &= \int_{\mathbb{R}} f(x, \theta) \frac{\partial}{\partial \theta} \ln f(x, \theta) dx \\ &= E \left[ \frac{\partial}{\partial \theta} \ln f(X, \theta) \right]. \end{aligned} \quad (2.7)$$

<sup>3</sup>Harald Cramér (1893-1985), matemático y estadístico sueco.

<sup>3</sup>Calyampudi Radhakrishna Rao (1920-), matemático y estadístico hindú.

De esta manera hemos comprobado que la variable aleatoria  $(\partial/\partial\theta) \ln f(X, \theta)$  tiene esperanza nula. Suponiendo ahora la diferenciabilidad de la función parametral,

$$\begin{aligned}
\tau'(\theta) &= \frac{d}{d\theta} E(T) \\
&= \frac{d}{d\theta} \int_{\mathbb{R}^n} T(x_1, \dots, x_n) f(x_1, \dots, x_n, \theta) dx_1 \cdots dx_n \\
&= \int_{\mathbb{R}^n} T(x_1, \dots, x_n) \frac{\partial}{\partial\theta} f(x_1, \dots, x_n, \theta) dx_1 \cdots dx_n \quad (2.8) \\
&= \int_{\mathbb{R}^n} T(x_1, \dots, x_n) \frac{\partial}{\partial\theta} e^{\ln f(x_1, \dots, x_n, \theta)} dx_1 \cdots dx_n \\
&= \int_{\mathbb{R}^n} T(x_1, \dots, x_n) \left[ \sum_{i=1}^n \frac{\partial}{\partial\theta} \ln f(x_i, \theta) \right] f(x_1, \dots, x_n, \theta) dx_1 \cdots dx_n \\
&= E\left(T \cdot \sum_{i=1}^n \frac{\partial}{\partial\theta} \ln f(X_i, \theta)\right) \\
&= \text{Cov}\left(T, \sum_{i=1}^n \frac{\partial}{\partial\theta} \ln f(X_i, \theta)\right).
\end{aligned}$$

La última igualdad se obtiene recordando que  $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$  y usando la identidad (2.7). Ahora utilizaremos la desigualdad  $\text{Cov}(X, Y) \leq \sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}$ . Tenemos que

$$\begin{aligned}
(\tau'(\theta))^2 &\leq \text{Var}(T) \cdot \text{Var}\left(\sum_{i=1}^n \frac{\partial}{\partial\theta} \ln f(X_i, \theta)\right) \\
&= \text{Var}(T) \cdot \sum_{i=1}^n \text{Var}\left(\frac{\partial}{\partial\theta} \ln f(X_i, \theta)\right) \\
&= \text{Var}(T) \cdot n \text{Var}\left(\frac{\partial}{\partial\theta} \ln f(X, \theta)\right) \\
&= \text{Var}(T) \cdot n E\left[\left(\frac{\partial}{\partial\theta} \ln f(X, \theta)\right)^2\right].
\end{aligned}$$

■

En el enunciado de la cota inferior de Cramér-Rao y en su demostración hemos usado la letra  $X$  para indicar a cualquier elemento de la muestra

aleatoria  $X_1, \dots, X_n$ . Esto es conveniente notacionalmente pues de esa manera no se hace uso de subíndices, e implícitamente se utiliza la hipótesis de idéntica distribución de las variables de la muestra aleatoria. Por otro lado, es importante observar que el término  $f(X, \theta)$  corresponde a la función de densidad o de probabilidad  $f(x, \theta)$  evaluada en la variable aleatoria  $X$ . Supondremos que tal operación, junto con las que aparecen en la expresión (2.6), produce nuevamente una variable aleatoria y que además su esperanza es finita.

**Definición 2.16** A la expresión del lado derecho de (2.6) se le llama la **cota inferior de Cramér-Rao** (CICR) para la varianza de cualquier estimador insesgado para  $\tau(\theta)$  y se le denota por  $\text{CICR}(\theta)$ .

En general, la CICR es una función del parámetro  $\theta$  y por ello se le escribe como  $\text{CICR}(\theta)$ , aunque en esta notación no se hace referencia a la función parametral  $\tau(\theta)$ . Así es que debemos tener cuidado en que al escribir  $\text{CICR}(\theta)$  no haya duda de la función parametral  $\tau(\theta)$  a la que se hace referencia en dicha cota. En particular, si la función parametral a estimar es el mismo parámetro  $\theta$ , la cota inferior de Cramér-Rao toma la siguiente expresión reducida

$$\text{CICR}(\theta) = \frac{1}{n E \left[ \left( \frac{\partial}{\partial \theta} \ln f(X, \theta) \right)^2 \right]}. \quad (2.9)$$

Cuando no se hace referencia a ninguna función parametral, supondremos implícitamente que la  $\text{CICR}(\theta)$  corresponde a la cota inferior para la varianza de cualquier estimador insesgado para  $\theta$  como aparece en (2.9).

Es interesante observar que el denominador de (2.6) no depende de la función parametral, de modo que conociendo la cota inferior para la varianza de cualquier estimador insesgado para  $\theta$ , es casi inmediato encontrar la cota inferior para la varianza de cualquier estimador insesgado de la función

parametral  $\tau(\theta)$ , simplemente se multiplica la cota anterior por  $(\tau'(\theta))^2$ .

De esta manera, la varianza de cualquier estimador insesgado para una función parametral tiene como valor mínimo la función  $\text{CICR}(\theta)$ . Por lo tanto, en caso de existir un estimador insesgado con varianza igual a  $\text{CICR}(\theta)$  para todo valor del parámetro, sabemos que tal estimador es el mejor en términos de ser insesgado y tener la varianza más pequeña. A tales estimadores les llamaremos estimadores insesgados de varianza mínima uniforme, o por brevedad y por sus siglas en inglés **UMVUE** (*Uniformly Minimum Variance Unbiased Estimator*). El adjetivo uniforme se refiere aquí a que la varianza es la más pequeña para cualquier valor del parámetro dentro del espacio parametral  $\Theta$ .

**Definición 2.17** Se dice que un estimador para un parámetro es un **UMVUE** si es insesgado y tiene varianza mínima dentro del conjunto de todos los estimadores insesgados para el mismo parámetro.

Así, si un estimador insesgado alcanza la **CICR**, es un **UMVUE**. Pero pueden existir estimadores que no alcanzan la **CICR** y ser un **UMVUE** si es que ningún otro estimador insesgado alcanza la varianza más pequeña.

Más adelante retomaremos el problema de determinar la existencia y unicidad de un estimador con estas características. Antes de especificar las condiciones técnicas bajo las cuales se cumple la cota inferior de Cramér-Rao, veamos algunos ejemplos del cálculo de esta cota inferior.

**Ejemplo 2.30** Sea  $X$  una variable aleatoria con distribución  $\text{Ber}(\theta)$ , con  $\theta$  desconocido.

$$f(x, \theta) = \begin{cases} \theta^x(1 - \theta)^{1-x} & \text{si } x = 0, 1, \\ 0 & \text{en otro caso.} \end{cases}$$

Sea  $\hat{\theta}$  cualquier estimador insesgado para el parámetro  $\theta$ , definido a través de una muestra aleatoria de esta distribución. Encontraremos la cota inferior de Cramér-Rao para la varianza de  $\hat{\theta}$ . La función parametral es  $\tau(\theta) = \theta$

y por lo tanto  $\tau'(\theta) = 1$ . Evaluando la función de probabilidad  $f(x, \theta)$  en la variable aleatoria  $X$  y haciendo las operaciones indicadas, es inmediato comprobar que

$$\frac{\partial}{\partial \theta} \ln f(X, \theta) = \frac{X}{\theta} - \frac{1-X}{1-\theta}.$$

Observemos que esta es una variable aleatoria y que tiene esperanza cero. Esto sirve de ejemplo de lo que hemos demostrado antes de manera general. El segundo momento de esta variable aleatoria es

$$\begin{aligned} E\left[\left(\frac{\partial}{\partial \theta} \ln f(X, \theta)\right)^2\right] &= E\left(\frac{X}{\theta} - \frac{1-X}{1-\theta}\right)^2 \\ &= \frac{1}{\theta(1-\theta)}. \end{aligned}$$

Substituyendo esta expresión en la fórmula (2.6) se obtiene que la cota inferior de Cramér-Rao es

$$\text{CICR}(\theta) = \frac{\theta(1-\theta)}{n}, \quad 0 < \theta < 1.$$

En consecuencia, todo estimador insesgado para  $\theta$  y construido a partir de una muestra aleatoria de tamaño  $n$  de la distribución Bernoulli tiene varianza por lo menos esta cantidad. Vista como función de  $\theta$ , la gráfica de la cota inferior de Cramér-Rao es la parábola que se muestra en la Figura 2.8. La varianza de cualquier estimador insesgado debe ser una función de  $\theta$  con valores dentro del área sombreada, es decir, por arriba de la cota inferior indicada mediante la curva continua.

Por ejemplo, consideremos el estimador  $\hat{\theta} = X_1$ . Claramente este estimador es insesgado y su varianza es  $\theta(1-\theta)$ . Se verifica entonces la desigualdad

$$\text{CICR}(\theta) = \frac{\theta(1-\theta)}{n} \leq \theta(1-\theta) = \text{Var}(\hat{\theta}).$$

La gráfica de  $\text{Var}(\hat{\theta}) = \theta(1-\theta)$  es también una parábola pero se encuentra por encima de la parábola dada por  $\text{CICR}(\theta)$ , para  $0 \leq \theta \leq 1$ . Podemos considerar también el estimador insesgado  $\hat{\theta} = \bar{X}$ . Claramente su varianza es  $\theta(1-\theta)/n$  y observamos que coincide con la CICR. En este caso, este es el estimador insesgado para  $\theta$  de varianza mínima, es decir, el UMVUE. ■

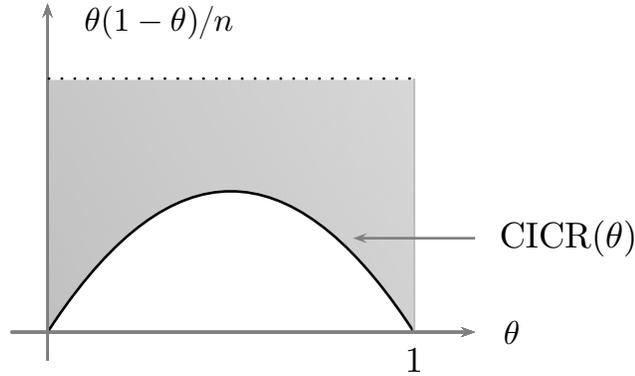


Figura 2.8

**Ejemplo 2.31** Sea  $X$  una variable aleatoria con distribución exponencial de parámetro  $\theta > 0$  desconocido.

$$f(x, \theta) = \begin{cases} \theta e^{-\theta x} & \text{si } x > 0, \\ 0 & \text{en otro caso.} \end{cases}$$

Sea  $\hat{\theta}$  cualquier estimador insesgado para  $\theta$ , definido a través de una muestra aleatoria de esta distribución. Encontraremos la cota inferior de Cramér-Rao para la varianza de  $\hat{\theta}$ . La función parametral es  $\tau(\theta) = \theta$  y por lo tanto  $\tau'(\theta) = 1$ . Evaluando la función de probabilidad  $f(x, \theta)$  en la variable aleatoria  $X$  y llevando a cabo las operaciones indicadas, es inmediato comprobar que

$$\frac{\partial}{\partial \theta} \ln f(X, \theta) = \frac{1}{\theta} - X.$$

Nuevamente esta es una variable aleatoria que tiene esperanza cero. Esto lo hemos demostrado antes de manera general. Por lo tanto,

$$E\left[\left(\frac{\partial}{\partial \theta} \ln f(X, \theta)\right)^2\right] = E(1/\theta - X)^2 = \text{Var}(X) = \frac{1}{\theta^2}.$$

Substituyendo esta expresión en la fórmula (2.6) se obtiene que la cota inferior de Cramér-Rao es

$$\text{CICR}(\theta) = \frac{\theta^2}{n}, \quad \theta > 0.$$

Por lo tanto, todo estimador insesgado para  $\theta$  y construido a partir de una muestra aleatoria de tamaño  $n$  en la distribución exponencial tiene varianza por lo menos esta cantidad. Vista como función de  $\theta > 0$ , la gráfica de la cota inferior de Cramér-Rao se muestra en la Figura 2.9.

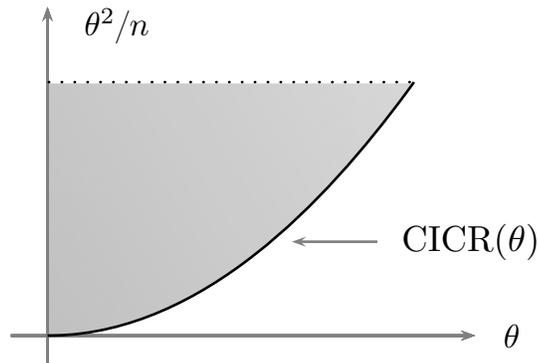


Figura 2.9

Para cada  $\theta > 0$ , la varianza de cualquier estimador insesgado debe ser una función de  $\theta$  con valores dentro del área sombreada. Veamos un ejemplo. Recordemos que el estimador máximo verosímil para el parámetro  $\theta$  de la distribución exponencial es  $1/\bar{X}$  y recordemos que este estimador no es insesgado pues  $E(1/\bar{X}) = (n/(n-1))\theta$ . De aquí puede proponerse el estimador insesgado

$$\hat{\theta} = \frac{n-1}{n} \cdot \frac{1}{\bar{X}},$$

cuya varianza es

$$\begin{aligned} \text{Var}(\hat{\theta}) &= \frac{(n-1)^2}{n^2} \text{Var}(1/\bar{X}) \\ &= \frac{(n-1)^2}{n^2} [E(1/\bar{X})^2 - E^2(1/\bar{X})] \\ &= \frac{(n-1)^2}{n^2} \left[ \frac{n^2}{(n-1)(n-2)} \theta^2 - \frac{n^2}{(n-1)^2} \theta^2 \right] \quad (2.10) \\ &= \frac{1}{n-2} \theta^2. \end{aligned}$$

Las expresiones que aparecen en (2.10) se pueden obtener con facilidad usando la distribución gama. De esta manera, se comprueba que la varianza del estimador insesgado  $\hat{\theta}$  es, efectivamente, mayor o igual que la cota inferior de Cramér-Rao, es decir, para  $n \geq 3$ ,

$$\text{CICR}(\theta) = \frac{1}{n} \theta^2 \leq \frac{1}{n-2} \theta^2 = \text{Var}(\hat{\theta}).$$

■

Teniendo ahora una mayor familiaridad con la cota inferior de Cramér-Rao, vamos a establecer las hipótesis bajo las cuales dicho resultado es válido. A este conjunto de hipótesis se le conoce con el nombre de condiciones de regularidad.

### Condiciones de regularidad

Las siguientes hipótesis son necesarias para la validez de la cota inferior de Cramér-Rao.

- El soporte de  $f(x, \theta)$  dado por el conjunto  $\{x : f(x, \theta) > 0\}$  no depende de  $\theta$ .
- Para todo  $x$  en el soporte de  $f(x, \theta)$ , la siguiente derivada existe

$$\frac{\partial}{\partial \theta} \ln f(x, \theta).$$

- Es válido el siguiente intercambio de derivada e integral.

$$0 = \frac{d}{d\theta} \int_{\mathbb{R}} f(x, \theta) dx = \int_{\mathbb{R}} \frac{\partial}{\partial \theta} f(x, \theta) dx.$$

- $0 < E \left[ \left( \frac{\partial}{\partial \theta} \ln f(X, \theta) \right)^2 \right] < \infty$ .
- Es válido el intercambio de derivada e integral que aparece abajo. Esto se usa en la identidad (2.8) de la demostración. Por brevedad, hacemos uso de la notación  $x^n = (x_1, \dots, x_n)$ .

$$\frac{d}{d\theta} \int_{\mathbb{R}^n} T(x^n) f(x^n, \theta) dx^n = \int_{\mathbb{R}^n} T(x^n) \frac{\partial}{\partial \theta} f(x^n, \theta) dx^n.$$

En la Tabla 2.6 se muestran las expresiones de la cota inferior de Cramér-Rao para algunas distribuciones. El parámetro a estimar se denomina por la letra  $\theta$ , suponiendo que cualquier otro posible parámetro que aparezca en la distribución es conocido. Como siempre, se reserva la letra  $n$  para el tamaño de la muestra. Se deja como ejercicio demostrar estos resultados. Es interesante notar la forma polinomial muy similar de todas estas cotas inferiores.

**Algunos ejemplos de CICR**

Distribución	Parámetro	CICR( $\theta$ )
Ber( $\theta$ )	$0 < \theta < 1$	$\frac{\theta(1-\theta)}{n}$
bin( $k, \theta$ )	$0 < \theta < 1$	$\frac{\theta(1-\theta)}{nk}$
geo( $\theta$ )	$0 < \theta < 1$	$\frac{\theta^2(1-\theta)}{n}$
bin neg( $r, \theta$ )	$0 < \theta < 1$	$\frac{\theta^2(1-\theta)}{nr}$
Poisson( $\theta$ )	$\theta > 0$	$\frac{\theta}{n}$
exp( $\theta$ )	$\theta > 0$	$\frac{\theta^2}{n}$
N( $\mu, \theta$ )	$\theta > 0$	$\frac{2\theta^2}{n}$

Tabla 2.6

En lo que resta del presente capítulo estudiaremos algunas otras propiedades de los estimadores orientadas a la búsqueda de UMVUEs. Para concluir esta sección, planteamos una pregunta que resulta natural después de lo recién estudiado en esta sección: dada una distribución  $f(x, \theta)$ , ¿existe siempre un

estimador insesgado para  $\theta$  que alcance la CICR? Veremos más adelante que la respuesta es negativa.

## Ejercicios

189. **Distribución binomial.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{bin}(k, \theta)$ , con  $0 < \theta < 1$  desconocido. Suponga que  $k \geq 1$  es un entero conocido. Demuestre que

$$\text{CICR}(\theta) = \frac{\theta(1-\theta)}{nk}, \quad 0 < \theta < 1.$$

Demuestre que el estimador  $\hat{\theta} = \bar{X}/k$  es insesgado y que su varianza coincide con la cota inferior de Cramér-Rao, es decir,

$$\text{CICR}(\theta) = \text{Var}(\hat{\theta}), \quad 0 < \theta < 1.$$

Por lo tanto, el estimador  $\bar{X}/k$  es el UMVUE para  $\theta$ , cuando  $k$  es conocido.

190. **Distribución geométrica.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{geo}(\theta)$ , con  $\theta$  desconocido. Demuestre que

$$\text{CICR}(\theta) = \frac{\theta^2(1-\theta)}{n}, \quad 0 < \theta < 1.$$

191. **Distribución binomial negativa.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{bin neg}(r, \theta)$ , con  $\theta$  desconocido. Suponga que  $r \geq 1$  es un entero conocido. Demuestre que

$$\text{CICR}(\theta) = \frac{\theta^2(1-\theta)}{nr}, \quad 0 < \theta < 1.$$

192. **Distribución Poisson.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{Poisson}(\theta)$ , con  $\theta$  desconocido. Demuestre que

$$\text{CICR}(\theta) = \frac{\theta}{n}, \quad \theta > 0.$$

Calcule la varianza de los siguientes estimadores insesgados y compruebe el cumplimiento de la cota inferior de Cramér-Rao.

$$a) \hat{\theta} = X_1.$$

$$b) \hat{\theta} = \bar{X}.$$

193. **Distribución normal: media.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $N(\theta, \sigma^2)$ , con  $\theta$  desconocido y  $\sigma^2$  conocida. Demuestre que

$$\text{CICR}(\theta) = \frac{\sigma^2}{n}. \quad (\text{función constante})$$

194. **Distribución normal: varianza.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $N(0, \theta)$ , con  $\theta > 0$  desconocido. Demuestre que

$$\text{CICR}(\theta) = \frac{2}{n} \theta^2, \quad \theta > 0.$$

Demuestre que el estimador  $\hat{\theta} = (X_1^2 + \dots + X_n^2)/n$  es insesgado y que su varianza coincide con la cota inferior de Cramér-Rao, es decir,

$$\text{CICR}(\theta) = \text{Var}(\hat{\theta}), \quad \theta > 0.$$

Por lo tanto,  $\hat{\theta}$  es un UMVUE para  $\theta$ .

195. **Distribución normal: varianza.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $N(\mu, \sigma^2)$ , con ambos parámetros desconocidos. Suponga  $n \geq 2$ . Recordemos que la varianza muestral  $S^2$  es un estimador insesgado para  $\sigma^2$ .

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

Demuestre que

$$\text{CICR}(\sigma^2) = \frac{2}{n} \sigma^4 < \frac{2}{n-1} \sigma^4 = \text{Var}(S^2).$$

## 2.9. Eficiencia

En esta sección veremos varias definiciones relacionadas con la varianza de un estimador insesgado. Primero veamos una posible manera de comparar dos estimadores insesgados.

**Definición 2.18** Sean  $\hat{\theta}_1$  y  $\hat{\theta}_2$  dos estimadores insesgados para un parámetro  $\theta$ . Se dice que  $\hat{\theta}_1$  es **relativamente más eficiente** que  $\hat{\theta}_2$  si

$$\text{Var}(\hat{\theta}_1) \leq \text{Var}(\hat{\theta}_2). \quad (2.11)$$

De esta manera, de entre dos estimadores insesgados para un mismo parámetro, preferiremos aquel que tenga varianza menor, si es que tal comparación puede llevarse a cabo. Recordemos que la varianza de un estimador es una función del parámetro y la desigualdad (2.11) pudiera no cumplirse para todo valor del parámetro dentro del espacio parametral. En consecuencia, no cualesquiera dos estimadores insesgados pueden compararse uno con el otro de la forma indicada en la definición anterior.

**Ejemplo 2.32** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{Ber}(\theta)$ , con  $\theta$  desconocido. Es claro que los estimadores  $\hat{\theta}_1 = \bar{X}$  y  $\hat{\theta}_2 = X_1$  son insesgados para  $\theta$ . Sin embargo, el estimador  $\hat{\theta}_1$  es relativamente más eficiente que  $\hat{\theta}_2$  pues, para cualquier valor de  $\theta$  en  $(0, 1)$ , se cumple

$$\text{Var}(\hat{\theta}_1) = \frac{\theta(1-\theta)}{n} \leq \theta(1-\theta) = \text{Var}(\hat{\theta}_2).$$

■

Por otro lado, en ocasiones hay estimadores insesgados con la mínima varianza posible dada por la cota inferior de Cramér-Rao. Los llamaremos estimadores eficientes. Estos son casos particulares de los estimadores que hemos denominado como UMVUEs, aquellos que alcanzan la CICR.

**Definición 2.19** Se dice que un estimador insesgado es **eficiente** cuando su varianza alcanza la cota inferior de Cramér-Rao.

Es decir, el estimador insesgado  $\hat{\theta}$  es eficiente si  $\text{Var}(\hat{\theta}) = \text{CICR}(\theta)$  para todo valor de  $\theta$  en el espacio parametral  $\Theta$ . Teniendo como elemento de comparación la cota inferior de Cramér-Rao podemos ahora definir la eficiencia de un estimador insesgado de la siguiente manera.

**Definición 2.20** La **eficiencia** de un estimador insesgado  $\hat{\theta}$  es

$$\text{Efi}(\hat{\theta}) = \frac{\text{CICR}(\theta)}{\text{Var}(\hat{\theta})}. \quad (2.12)$$

Observemos nuevamente que la eficiencia es una función del parámetro  $\theta$  a estimar, es siempre positiva y menor o igual a uno. Más generalmente, la cualidad de ser eficiente para un estimador insesgado puede alcanzarse en el límite cuando el tamaño de la muestra tiende a infinito.

**Definición 2.21** Sea  $\hat{\theta}_n$  un estimador insesgado para  $\theta$ , construido a partir de una muestra aleatoria de tamaño  $n$ . Se dice que  $\hat{\theta}_n$  es **asintóticamente eficiente** si

$$\lim_{n \rightarrow \infty} \text{Efi}(\hat{\theta}_n) = 1.$$

Por supuesto, todo estimador eficiente es asintóticamente eficiente y el recíproco no se cumple. En la Figura 2.10 se muestran gráficamente las relaciones generales entre los conceptos estudiados en esta sección y en el siguiente ejemplo se analizan casos concretos.

Estimadores insesgados

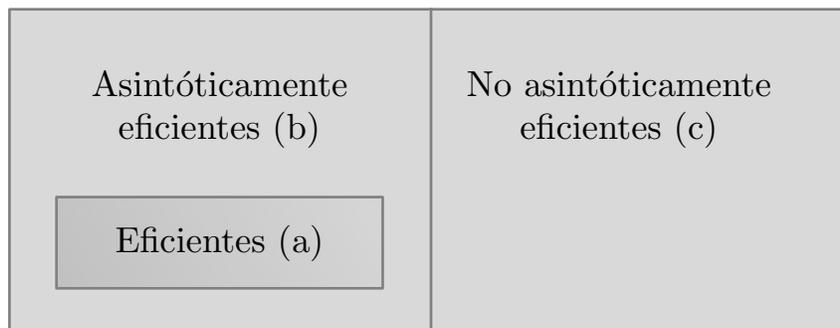


Figura 2.10

**Ejemplo 2.33** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{Ber}(\theta)$ , con  $\theta$  desconocido. Sabemos que la cota inferior de Cramér-Rao para la varianza de cualquier estimador insesgado para  $\theta$  es, para  $0 < \theta < 1$ ,

$$\text{CICR}(\theta) = \frac{\theta(1 - \theta)}{n}.$$

- **Estimador eficiente.** El estimador insesgado  $\hat{\theta} = \bar{X}$  es eficiente pues

$$\text{Var}(\hat{\theta}) = \theta(1 - \theta)/n = \text{CICR}(\theta).$$

Este es un ejemplo de la situación (a) de la Figura 2.10.

- **Estimador no eficiente pero asintóticamente eficiente.** Consideremos el estimador insesgado  $\hat{\theta}_n = (X_1 + \dots + X_{n-1})/(n - 1)$ , es decir, sólo se toma el promedio de las primeras  $n - 1$  variables de la muestra aleatoria. Es claro que  $\hat{\theta}_n$  es insesgado y su varianza es

$$\text{Var}(\hat{\theta}_n) = \frac{\theta(1 - \theta)}{n - 1}.$$

Su eficiencia es

$$\text{Efi}(\hat{\theta}_n) = \frac{n - 1}{n} < 1.$$

Se trata entonces de un estimador que no es eficiente, pero claramente es asintóticamente eficiente. Este es un ejemplo de la situación (b) de la Figura 2.10.

- **Estimador no eficiente ni asintóticamente eficiente.** Consideremos ahora el estimador insesgado

$$\hat{\theta}_n = \frac{2}{n(n + 1)} (X_1 + 2X_2 + \dots + nX_n).$$

Su varianza puede encontrarse como sigue

$$\begin{aligned} \text{Var}(\hat{\theta}_n) &= \frac{4}{n^2(n + 1)^2} \left[ \sum_{k=1}^n k^2 \right] \theta(1 - \theta) \\ &= \frac{2(2n + 1)}{3(n + 1)} \frac{\theta(1 - \theta)}{n}. \end{aligned}$$

Por lo tanto, su eficiencia es

$$\text{Efi}(\hat{\theta}_n) = \frac{3(n+1)}{2(2n+1)} < 1.$$

Entonces, este es un estimador que no es eficiente y tampoco es asintóticamente eficiente pues

$$\lim_{n \rightarrow \infty} \text{Efi}(\hat{\theta}_n) = \frac{3}{4} < 1.$$

Este es un ejemplo de la situación (c) de la Figura 2.10.

■

## Ejercicios

196. Diga falso o verdadero.

- a) Todo estimador eficiente es un UMVUE.
- b) Todo UMVUE es eficiente.

197. Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución normal de media  $\theta$ . Demuestre que  $\bar{X}$  es un estimador insesgado de mínima varianza para  $\theta$ .

## 2.10. Suficiencia

Definiremos a continuación la noción de suficiencia de una estadística para un parámetro de una distribución de probabilidad. Este concepto fue propuesto por Ronald Fisher<sup>4</sup> en 1920 y ha resultado ser de suma importancia dentro de la estadística y sus aplicaciones. En las siguientes secciones tendremos oportunidad de mostrar su utilidad.

---

<sup>4</sup>Ronald Aylmer Fisher (1890-1962), estadístico y genetista inglés.

**Definición 2.22** Una estadística  $T$  es **suficiente** para un parámetro  $\theta$  si la distribución conjunta de la muestra aleatoria  $X_1, \dots, X_n$  condicionada al evento  $(T = t)$  no depende del parámetro  $\theta$ , cualquiera que sea el posible valor  $t$  de la estadística.

Observe que la colección de eventos  $(T = t)$ , un evento para cada valor  $t$  de la estadística, induce una partición en el conjunto de valores  $x^n = (x_1, \dots, x_n)$  de la muestra aleatoria, esto es,  $(T = t) = \{x^n : T(x^n) = t\}$ . Consideremos ahora la función  $x^n \mapsto f(x^n | T = t)$ . Es claro que esta función se anula fuera del subconjunto  $(T = t)$  y se cumple la condición de suficiencia si dentro de este subconjunto la función no depende de  $\theta$ .

Más adelante comprobaremos que la suficiencia se puede interpretar de la siguiente manera: dado un valor  $t$  de la estadística  $T$ , la muestra aleatoria no contiene información adicional sobre el parámetro  $\theta$  que aquella proporcionada por la estadística  $T$ .

Veremos a continuación algunos ejemplos de la forma en la que puede verificarse la propiedad de suficiencia de una estadística mediante la definición anterior. Posteriormente veremos otras maneras equivalentes de verificar esta propiedad.

**Ejemplo 2.34** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución Bernoulli de parámetro desconocido  $\theta$ . Comprobaremos que la estadística  $T = X_1 + \dots + X_n$  es suficiente para  $\theta$ . Notemos que  $T$  tiene distribución  $\text{bin}(n, \theta)$  y que  $T$  no necesariamente es un estimador para  $\theta$ . Observe que si la muestra aleatoria toma el valor  $x^n = (x_1, \dots, x_n)$ , entonces forzosa-mente la estadística  $T$  toma el valor  $T(x^n) = x_1 + \dots + x_n$ . Por brevedad escribiremos a la muestra aleatoria como el vector  $X^n = (X_1, \dots, X_n)$ . Entonces, sea  $t \in \{0, 1, \dots, n\}$  cualquier posible valor de la estadística  $T$  y sea  $x_1, \dots, x_n \in \{0, 1\}$  cualquier valor de la muestra aleatoria. Tenemos que

$$\begin{aligned}
P(X^n = x^n | T = t) &= \frac{P(X^n = x^n, T = t)}{P(T = t)} \\
&= \frac{P(X^n = x^n)}{P(T = t)} \cdot 1_{\{t\}}(x_1 + \cdots + x_n) \\
&= \frac{\theta^{x_1} (1 - \theta)^{1-x_1} \cdots \theta^{x_n} (1 - \theta)^{1-x_n}}{\binom{n}{t} \theta^t (1 - \theta)^{n-t}} \cdot 1_{\{t\}}(n\bar{x}) \\
&= \frac{\theta^{x_1 + \cdots + x_n} (1 - \theta)^{n - (x_1 + \cdots + x_n)}}{\binom{n}{t} \theta^t (1 - \theta)^{n-t}} \cdot 1_{\{t\}}(n\bar{x}) \\
&= \frac{1}{\binom{n}{t}} \cdot 1_{\{t\}}(n\bar{x}).
\end{aligned}$$

Como esta probabilidad no depende de  $\theta$ , concluimos que  $T$  es suficiente para  $\theta$  de acuerdo a la definición anterior. Observe que la condición  $(T = t)$  hace que los posibles valores de la muestra aleatoria se reduzcan a aquellos que cumplen la condición  $x_1 + \cdots + x_n = t$ .

■

**Ejemplo 2.35** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución Poisson con parámetro  $\theta > 0$  desconocido. Comprobaremos que la estadística  $T = X_1 + \cdots + X_n$  es suficiente para  $\theta$ . Observemos que  $T$  tiene distribución Poisson( $n\theta$ ) y que no necesariamente es un estimador para  $\theta$ . Sea  $t \in \{0, 1, \dots\}$  es uno de los posibles valores de  $T$ . Para cualquier valor  $x^n = (x_1, \dots, x_n)$  de la muestra aleatoria, tenemos que

$$\begin{aligned}
P(X^n = x^n | T = t) &= \frac{P(X^n = x^n, T = t)}{P(T = t)} \\
&= \frac{P(X^n = x^n)}{P(T = t)} \cdot 1_{\{t\}}(x_1 + \cdots + x_n) \\
&= \frac{[e^{-\theta} \theta^{x_1} / x_1!] \cdots [e^{-\theta} \theta^{x_n} / x_n!]}{e^{-n\theta} (n\theta)^t / t!} \cdot 1_{\{t\}}(x_1 + \cdots + x_n) \\
&= \frac{e^{-n\theta} \theta^t / (x_1! \cdots x_n!)}{e^{-n\theta} (n\theta)^t / t!} \cdot 1_{\{t\}}(x_1 + \cdots + x_n) \\
&= \frac{t!}{n^t x_1! \cdots x_n!} \cdot 1_{\{t\}}(x_1 + \cdots + x_n).
\end{aligned}$$

Esta probabilidad no depende de  $\theta$  y, por lo tanto,  $T$  es suficiente para  $\theta$ . ■

Ahora veremos un ejemplo en donde no se cumple la propiedad de suficiencia.

**Ejemplo 2.36** Sea  $X_1, X_2, X_3$  una muestra aleatoria de tamaño  $n = 3$  de la distribución  $\text{Ber}(\theta)$ , con  $\theta$  desconocido. Comprobaremos que la estadística  $T = X_1 + 2X_2 + 3X_3$  no es suficiente para  $\theta$ . Para ello es suficiente dar un valor de la muestra aleatoria y un valor de la estadística para los cuales no se cumpla la condición de suficiencia. Tomemos  $(x_1, x_2, x_3) = (1, 1, 0)$  y  $t = 3$ . Entonces

$$\begin{aligned} P(X_1 = 1, X_2 = 1, X_3 = 0 | T = 3) &= \frac{P(X_1 = 1, X_2 = 1, X_3 = 0, T = 3)}{P(T = 3)} \\ &= \frac{P(X_1 = 1, X_2 = 1, X_3 = 0)}{P(T = 3)} \\ &= \frac{\theta^2(1 - \theta)}{\theta^2(1 - \theta) + (1 - \theta)^2\theta} \\ &= \theta. \end{aligned}$$

Claramente esta probabilidad depende del parámetro  $\theta$  y, por lo tanto,  $T$  no es suficiente para  $\theta$ . En conclusión, la estadística  $X_1 + X_2 + X_3$  es suficiente para  $\theta$  pero  $X_1 + 2X_2 + 3X_3$  no lo es. ■

A pesar de lo fácil que resultaron los cálculos en los ejemplos anteriores, en realidad no es sencillo comprobar la suficiencia de una estadística cualquiera usando la definición. Observe que en estos ejemplos fue necesario conocer la distribución de la estadística  $T$  y en los casos mostrados tal distribución fue evidente de encontrar. Esto no siempre es así y los cálculos pueden ser sumamente complicados con casi cualquier otro caso que se considere.

Afortunadamente se cuenta con el siguiente resultado bastante útil, que establece una condición equivalente para la suficiencia. Esta condición es relativamente fácil de verificar y la usaremos con mayor frecuencia que la definición misma de suficiencia. Será un segundo mecanismo para comprobar la suficiencia de una estadística.

Recordemos nuevamente que  $X_1, \dots, X_n$  es una muestra aleatoria de una distribución con función de densidad o de probabilidad  $f(x, \theta)$ , dependiente de un parámetro desconocido  $\theta$ . Y que, por brevedad, escribiremos  $x^n$  en lugar de  $(x_1, \dots, x_n)$  para un valor particular de una muestra aleatoria.

**Teorema 2.3 (Teorema de factorización de J. Neyman)** Una estadística  $T$  es suficiente para  $\theta$  si y sólo si la función de densidad conjunta de la muestra aleatoria se puede factorizar de la siguiente forma

$$f(x^n, \theta) = g(T(x^n), \theta) \cdot h(x^n), \quad (2.13)$$

en donde  $g$  es una función no negativa que depende de los valores de la muestra aleatoria únicamente a través de la estadística  $T$ , y  $h$  es una función no negativa que depende únicamente del valor  $x^n = (x_1, \dots, x_n)$  de la muestra aleatoria.

***Demostración.***

( $\Rightarrow$ ) Supongamos que  $T$  es una estadística suficiente y sea  $x^n = (x_1, \dots, x_n)$  cualquier valor de la muestra aleatoria. Entonces la estadística  $T$  toma el valor  $T(x^n)$ . A la distribución conjunta de la muestra  $f(x^n)$  le añadimos la información redundante  $T = T(x^n)$  y condicionamos de la siguiente forma

$$\begin{aligned} f(x^n) &= f(x^n, T(x^n)) \\ &= f(T(x^n)) \cdot f(x^n | T(x^n)) \end{aligned}$$

El primer factor es una función  $g(T(x^n), \theta)$  que depende del parámetro  $\theta$  y del punto muestral  $x^n$  únicamente a través del valor de la estadística  $T$ . El segundo factor es una función  $h(x^n)$  que depende únicamente del valor de la muestra aleatoria, pues  $T$  es suficiente. De esta forma hemos construido la expresión del lado derecho de la igualdad (2.13).

( $\Leftarrow$ ) Suponga que se cumple la factorización (2.13). Demostraremos que  $T$  es suficiente. Por simplicidad en la escritura consideraremos el caso discreto. Sea  $x^n = (x_1, \dots, x_n)$  cualquier valor de la muestra aleatoria. A partir de este valor definimos el valor de la estadística  $t =$

$T(x_1, \dots, x_n)$ . Ahora consideremos la imagen inversa del valor  $t$  bajo la función  $T$ , es decir,

$$T^{-1}\{t\} = \{y^n : T(y^n) = t\}.$$

Por construcción,  $x^n \in T^{-1}\{t\}$ . Entonces

$$\begin{aligned} P(X_1 = x_1, \dots, X_n = x_n | T = t) &= \frac{P(X^n = x^n, T = t)}{P(T = t)} \\ &= \frac{P(X^n = x^n)}{P(X^n \in T^{-1}\{t\})} \\ &= \frac{P(X^n = x^n)}{\sum_{y^n \in T^{-1}\{t\}} P(X^n = y^n)} \\ &= \frac{g(T(x^n), \theta) h(x^n)}{\sum_{y^n \in T^{-1}\{t\}} g(T(y^n), \theta) h(y^n)} \\ &= \frac{g(t, \theta) h(x^n)}{g(t, \theta) \sum_{y^n \in T^{-1}\{t\}} h(y^n)} \\ &= \frac{h(x^n)}{\sum_{y^n \in T^{-1}\{t\}} h(y^n)}. \end{aligned}$$

Como esta probabilidad no depende de  $\theta$ , concluimos que  $T$  es suficiente. ■

Como una muestra de la forma en la que se aplica el teorema anterior, repetiremos los resultados de los Ejemplos 2.34 y 2.35, pero ahora usando el teorema de factorización.

**Ejemplo 2.37** Sea  $T = X_1 + \dots + X_n$ .

- La estadística  $T$  es suficiente para el parámetro desconocido  $\theta$  en la distribución Bernoulli pues para cualquier valor  $x^n = (x_1, \dots, x_n)$  de la muestra aleatoria,

$$\begin{aligned} P(X^n = x^n) &= \theta^{x_1} (1 - \theta)^{1-x_1} \dots \theta^{x_n} (1 - \theta)^{1-x_n} \\ &= [\theta^{x_1 + \dots + x_n} (1 - \theta)^{n - (x_1 + \dots + x_n)}] \cdot [1] \\ &= g(T(x^n), \theta) \cdot h(x^n). \end{aligned}$$

Por simplicidad en la escritura hemos omitido los factores  $1_{\{0,1\}}(x_i)$ , para  $i = 1, \dots, n$ , los cuales deben incorporarse a la función  $h(x^n)$ .

- La estadística  $T$  también es suficiente para el parámetro desconocido  $\theta$  de la distribución Poisson pues

$$\begin{aligned} P(X^n = x^n) &= e^{-\theta} \frac{\theta^{x_1}}{x_1!} \cdots e^{-\theta} \frac{\theta^{x_n}}{x_n!} \\ &= [e^{-n\theta} \theta^{x_1 + \cdots + x_n}] \cdot \left[ \frac{1}{x_1! \cdots x_n!} \right] \\ &= g(T(x^n), \theta) \cdot h(x^n). \end{aligned}$$

Nuevamente hemos omitido los factores  $1_{\{0,1,\dots\}}(x_i)$ , para  $i = 1, \dots, n$ , los cuales deben incorporarse a la función  $h(x^n)$ . ▪

Algunos otros ejemplos de estadísticas suficientes aparecen en la sección de ejercicios. Observemos que para demostrar que una estadística no es suficiente parece ser más conveniente usar directamente la Definición 2.22 como lo hemos hecho en el Ejemplo 2.36. Para ello se deben encontrar valores particulares  $x_1, \dots, x_n$  de la muestra aleatoria y un valor particular  $t$  de la estadística  $T$ , y verificar que la función  $f(x_1, \dots, x_n | T = t)$  depende del parámetro  $\theta$  a estimar.

En lo que resta de esta sección estudiaremos algunos resultados relativos al concepto de suficiencia. Por ejemplo, uno puede plantearse la siguiente pregunta: ¿Es la transformación de una estadística suficiente también suficiente para el mismo parámetro? Demostraremos a continuación que para que tal propiedad se cumpla, la condición de biyectividad para la transformación es suficiente.

**Proposición 2.4** Funciones biyectivas de estadísticas suficientes son suficientes.

**Demostración.** Usaremos el teorema de factorización. Sea  $T$  una estadística suficiente para un parámetro  $\theta$  y sea  $\varphi$  una función biyectiva definida

sobre el conjunto de valores de  $T$  y con valores reales. Entonces la función inversa de  $\varphi$  existe y podemos escribir  $T = \varphi^{-1} \circ (\varphi \circ T)$ . Como  $T$  es suficiente, por el teorema de factorización tenemos que

$$\begin{aligned} f(x^n, \theta) &= g(T(x^n), \theta) \cdot h(x^n) \\ &= g(\varphi^{-1} \circ (\varphi \circ T)(x^n), \theta) \cdot h(x^n) \\ &= G((\varphi \circ T)(x^n), \theta) \cdot h(x^n), \end{aligned}$$

en donde  $G = g \circ \varphi^{-1}$  es no negativa pues  $g$  es no negativa. Por lo tanto, la composición  $\varphi \circ T$  también es suficiente para  $\theta$ . ■

El resultado anterior también puede demostrarse directamente usando la definición de suficiencia y se deja como ejercicio para el lector.

Por otro lado, observemos que el enunciado y la demostración de la proposición anterior incluye el caso cuando  $T$  es un vector de estadísticas. Para ello, la función biyectiva debe estar bien definida sobre alguna región de  $\mathbb{R}^k$ , por ejemplo, aquella región en donde el vector de estadísticas toma sus valores.

Veamos ahora algunos ejemplos sencillos del uso del resultado recién demostrado.

**Ejemplo 2.38** Sabemos que la estadística  $T = X_1 + \cdots + X_n$  es suficiente para el parámetro  $\theta$  de la distribución Poisson. Tenemos entonces que:

- La estadística  $e^T$  es también suficiente para  $\theta$  pues la función  $\varphi(x) = e^x$  es biyectiva.
- La estadística  $T^2$  es también suficiente para  $\theta$  pues la función  $\varphi(x) = x^2$  es biyectiva sobre el intervalo  $(0, \infty)$ .

El resultado y el ejemplo anteriores sugieren un tercer mecanismo para comprobar la suficiencia de una estadística: verificar que la estadística en cuestión es una función biyectiva de otra estadística que sabemos que es suficiente.

Para concluir esta sección enunciamos un resultado que da respuesta a la siguiente pregunta: si  $T$  una estadística suficiente para  $\theta$ , ¿es  $T$  suficiente para cualquier función parametral  $\tau(\theta)$ ? La respuesta es afirmativa y aquí tenemos el enunciado.

**Proposición 2.5** Toda estadística suficiente para un parámetro  $\theta$  es también suficiente para cualquier función parametral  $\tau(\theta)$ .

*Demostración.* Usaremos la definición. Sea  $T$  una estadística suficiente para  $\theta$ . Entonces la distribución conjunta de la muestra aleatoria condicionada al evento  $(T = t)$  no depende de  $\theta$ , por lo tanto tampoco depende de  $\tau(\theta)$ . ■

Esto nos provee de un cuarto posible método para demostrar la propiedad de suficiencia: en el caso cuando se desee probar suficiencia de una estadística para una función parametral, verificar si la estadística es suficiente para el parámetro en cuestión. Como referencia, véase la sección 2.14, en donde se muestra un resumen de algunos métodos para probar la suficiencia de una estadística.

Más adelante estudiaremos el concepto de suficiencia de un vector de estadísticas para uno o varios parámetros. A tal situación le llamaremos suficiencia conjunta del vector de estadísticas. La definición y los resultados son completamente análogos. En la siguiente sección estudiaremos la información de Fisher. A través de este concepto se le puede dar una interpretación a la suficiencia.

Como un ejemplo general de estadística suficiente, en la sección 2.19 al final del presente capítulo, se presenta una familia amplia de distribuciones de probabilidad llamada familia exponencial. Para cada distribución dentro de esta familia es posible dar la expresión explícita de una estadística suficiente.

**Algunos ejemplos de estadísticas suficientes**

Distribución	Parámetro	Estadística suficiente
$\text{unif}\{1, \dots, \theta\}$	$\theta$	$T = X_{(n)}$
$\text{Ber}(\theta)$	$\theta$	$T = X_1 + \dots + X_n$
$\text{geo}(\theta)$	$\theta$	$T = X_1 + \dots + X_n$
$\text{Poisson}(\theta)$	$\theta$	$T = X_1 + \dots + X_n$
$\text{unif}(0, \theta)$	$\theta$	$T = X_{(n)}$
$\text{N}(\theta, \sigma^2)$	$\theta$	$T = X_1 + \dots + X_n$

Tabla 2.7

## Ejercicios

198. Usando directamente la definición de suficiencia, demuestre que cualquier función biyectiva de una estadística suficiente es suficiente. Considere únicamente el caso discreto.
199. Use el teorema de factorización para demostrar nuevamente que toda estadística suficiente para un parámetro  $\theta$  es también suficiente para cualquier función parametral  $\tau(\theta)$ . Este es el contenido de la Proposición 2.5.
200. Sea  $T$  una estadística suficiente para un parámetro  $\theta$  y sea  $a \neq 0$  una constante conocida. Demuestre directamente de la definición que las siguientes estadísticas también son suficientes.

a)  $T + a$ .

b)  $aT$ .

c)  $e^T$ .

201. Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución especificada abajo, en donde  $\theta$  es un parámetro desconocido. Suponga que cualquier otro parámetro que pudiera aparecer en la distribución es conocido. Demuestre directamente de la definición que la estadística  $T = X_1 + \dots + X_n$  es suficiente.

a)  $\text{bin}(k, \theta)$ .

c)  $N(\theta, \sigma^2)$ .

b)  $\text{geo}(\theta)$ .

d)  $\text{gama}(\gamma, \theta)$ .

202. **Distribución uniforme.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{unif}(0, \theta)$ , con  $\theta > 0$  desconocido. Demuestre que la siguiente estadística es suficiente para  $\theta$ .

$$T = X_{(n)}.$$

203. **Distribución uniforme.** Sea  $X_1$  una muestra aleatoria de tamaño  $n = 1$  de la distribución  $\text{unif}(-\theta, \theta)$ , con  $\theta > 0$  desconocido. Demuestre que la siguiente estadística es suficiente para  $\theta$ .

$$T = |X_1|.$$

204. **Distribución exponencial: no suficiencia.** Sea  $X_1$  una muestra aleatoria de tamaño  $n = 1$  de la distribución  $\text{exp}(\theta)$ , en donde  $\theta > 0$  es desconocido. Demuestre que la siguiente estadística no es suficiente para  $\theta$ .

$$T = 1_{(X_1 > 2)}.$$

205. **Distribución exponencial.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{exp}(\theta)$ , en donde  $\theta > 0$  es desconocido. Demuestre que la siguiente estadística es suficiente para  $\theta$ .

$$T = X_1 + \dots + X_n.$$

206. **Distribución exponencial.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $f(x; \theta)$  especificada abajo, en donde  $\theta$  es un parámetro desconocido. Demuestre que la primera estadística de orden  $X_{(1)}$  es suficiente para  $\theta$ .

$$f(x) = \begin{cases} e^{-(x-\theta)} & \text{si } x > \theta, \\ 0 & \text{en otro caso.} \end{cases}$$

207. **Distribución Rayleigh.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución Rayleigh especificada abajo, en donde  $\theta > 0$  es un parámetro desconocido. Demuestre que la estadística  $U = X_1^2 + \dots + X_n^2$  es suficiente para  $\theta$ .

$$f(x, \theta) = \begin{cases} 2(x/\theta)e^{-x^2/\theta} & \text{si } x > 0, \\ 0 & \text{en otro caso.} \end{cases}$$

208. **Una familia de distribuciones.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución continua con función de densidad

$$f(x, \theta) = \begin{cases} a(\theta)b(x) & \text{si } 0 < x < \theta, \\ 0 & \text{en otro caso,} \end{cases}$$

en donde  $a(\theta)$  y  $b(x)$  son dos funciones no negativas dependientes únicamente de los parámetros indicados con  $\theta > 0$  desconocido. Por ejemplo, cuando  $a(\theta) = 1/\theta$  y  $b(x) = 1$  se obtiene la distribución  $\text{unif}(0, \theta)$ . Demuestre que la siguiente estadística es siempre suficiente para  $\theta$ .

$$T = \text{máx} \{X_1, \dots, X_n\}.$$

209. **Distribución normal.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $N(\mu, \theta)$ , en donde  $\mu$  es conocido y  $\theta > 0$  es desconocido. Sea  $\hat{\theta}$  el estimador para  $\theta$  por el método de máxima verosimilitud.

- a) Encuentre  $\hat{\theta}$ .
- b) Demuestre que  $\hat{\theta}$  es una estadística suficiente.

210. **Distribución normal: no suficiencia.** Sea  $X_1$  una muestra aleatoria de tamaño  $n = 1$  de la distribución  $N(\theta, \sigma^2)$  en donde  $\theta$  es desconocido y  $\sigma^2$  es conocido. Demuestre que la siguiente estadística no es suficiente para  $\theta$ .

$$T = |X_1|.$$

211. **Distribución Bernoulli: no suficiencia.** Sea  $X_1, \dots, X_4$  una muestra aleatoria de tamaño  $n = 4$  de la distribución  $\text{Ber}(\theta)$ , con  $0 < \theta < 1$  desconocido. Demuestre que la siguiente estadística no es suficiente para  $\theta$ .

$$T = X_1(X_2 + X_3) + X_4.$$

212. **Distribución Poisson: no suficiencia.** Sea  $X_1, X_2$  una muestra aleatoria de tamaño  $n = 2$  de la distribución  $\text{Poisson}(\theta)$ , en donde  $\theta$  es desconocido. Demuestre que la siguiente estadística no es suficiente para  $\theta$ .

$$T = X_1 - X_2.$$

213. **Distribución normal: no suficiencia.** Sea  $X_1, X_2$  una muestra aleatoria de tamaño  $n = 2$  de la distribución  $N(\theta, 1)$ , en donde  $\theta$  es desconocido. Demuestre que la siguiente estadística no es suficiente para  $\theta$ .

$$T = X_1 + 2X_2.$$

## 2.11. Suficiencia e información

En esta sección se define el concepto de información de Fisher de una variable aleatoria, o de su distribución  $f(x, \theta)$ , la cual supondremos dependiente de un parámetro desconocido y unidimensional  $\theta$ . Se muestra además la relación entre la información de Fisher y el concepto de suficiencia de una estadística.

**Definición 2.23** Sea  $X$  una variable aleatoria con función de densidad o probabilidad  $f(x, \theta)$ , dependiente de un parámetro  $\theta$ . La **información de Fisher** de  $X$ , o de su distribución, es la función

$$I(\theta) = E \left[ \left( \frac{\partial}{\partial \theta} \ln f(X, \theta) \right)^2 \right]. \quad (2.14)$$

Notemos que la información de Fisher es una función del parámetro  $\theta$  y tiene como dominio de definición el correspondiente espacio parametral. Observemos además con cuidado la expresión  $f(X, \theta)$  que aparece en el enunciado: la función de densidad  $f(x, \theta)$  es evaluada en la variable aleatoria  $X$ , es decir, se trata de una composición de funciones. Supondremos que este término es nuevamente una variable aleatoria y que la función  $\ln f(X, \theta)$  es diferenciable respecto de  $\theta$ . La expresión que define a la información de Fisher es un término que había aparecido antes como parte de la cota inferior de Cramér-Rao, la cual podemos ahora reescribir como sigue: para cualquier estimador insesgado  $T$  para la función parametral  $\tau(\theta)$ , y bajo las hipótesis de regularidad,

$$\text{Var}(T) \geq \frac{(\tau'(\theta))^2}{n \cdot I(\theta)}.$$

Cuando sea necesario especificar la variable aleatoria en cuestión escribiremos  $I_X(\theta)$  y la función de densidad o de probabilidad será  $f_X(x, \theta)$ . Por convención, el logaritmo indicado es el logaritmo natural.

La información de Fisher es una medida de la cantidad de información que una observación de la variable aleatoria contiene acerca del parámetro desconocido  $\theta$ . Veremos a continuación algunos ejemplos del cálculo de esta cantidad.

**Ejemplo 2.39** La información de Fisher de una variable aleatoria  $X$  con distribución  $\text{Ber}(\theta)$  es, para  $0 < \theta < 1$ ,

$$\begin{aligned}
 I(\theta) &= E\left[\left(\frac{\partial}{\partial\theta} \ln \theta^X (1-\theta)^{1-X}\right)^2\right] \\
 &= E\left[\left(\frac{\partial}{\partial\theta} [X \ln \theta + (1-X) \ln(1-\theta)]\right)^2\right] \\
 &= E\left[\left(\frac{X}{\theta} - \frac{1-X}{1-\theta}\right)^2\right] \\
 &= \frac{1}{\theta(1-\theta)}.
 \end{aligned}$$

■

**Ejemplo 2.40** La información de Fisher de una variable aleatoria  $X$  con distribución  $\exp(\theta)$  es, para  $\theta > 0$ ,

$$\begin{aligned}
 I(\theta) &= E\left[\left(\frac{\partial}{\partial\theta} \ln \theta e^{-\theta X}\right)^2\right] \\
 &= E\left[\left(\frac{\partial}{\partial\theta} [\ln \theta - \theta X]\right)^2\right] \\
 &= E\left[(1/\theta - X)^2\right] \\
 &= \text{Var}(X) \\
 &= \frac{1}{\theta^2}.
 \end{aligned}$$

■

Como ejemplos adicionales, en la Tabla 2.8 se muestran las expresiones de la información de Fisher para algunas otras distribuciones de probabilidad. Esta tabla es equivalente a la tabla sobre la cota inferior de Cramér-Rao de la Figura 2.6 en la página 163. El parámetro se denota por la letra  $\theta$ , suponiendo que cualquier otro posible parámetro que aparezca en la distribución es conocido. Como siempre, se reserva la letra  $n$  para el tamaño de la muestra. Se ha dejado comprobar los resultados mostrados en la tabla en la sección de ejercicios.

**Ejemplos de información de Fisher**

Distribución	Parámetro	$I(\theta)$
Ber( $\theta$ )	$0 < \theta < 1$	$\frac{1}{\theta(1-\theta)}$
bin( $k, \theta$ )	$0 < \theta < 1$	$\frac{k}{\theta(1-\theta)}$
geo( $\theta$ )	$0 < \theta < 1$	$\frac{1}{\theta^2(1-\theta)}$
bin neg( $r, \theta$ )	$0 < \theta < 1$	$\frac{r}{\theta^2(1-\theta)}$
Poisson( $\theta$ )	$\theta > 0$	$\frac{1}{\theta}$
exp( $\theta$ )	$\theta > 0$	$\frac{1}{\theta^2}$
gama( $\gamma, \theta$ )	$\theta > 0$	$\frac{\gamma}{\theta^2}$
N( $\theta, \sigma^2$ )	$\theta \in \mathbb{R}$	$\frac{1}{\sigma^2}$
N( $\mu, \theta$ )	$\theta > 0$	$\frac{1}{2\theta^2}$

Tabla 2.8

En la siguiente proposición se presentan dos resultados de utilidad. El primero de ellos establece que la variable aleatoria  $\frac{\partial}{\partial \theta} \ln f(X, \theta)$ , que es parte de la expresión que define a la información de Fisher, siempre tiene esperanza cero. Esto ya había sido demostrado antes cuando estudiamos la cota inferior de Cramér-Rao. El segundo resultado nos provee de una fórmula alternativa para calcular la información de Fisher.

**Proposición 2.6** Sea  $X$  una variable aleatoria con función de densidad o de probabilidad  $f(x, \theta)$  dependiente de un parámetro  $\theta$ . Bajo las hipótesis de regularidad,

1.  $E \left[ \frac{\partial}{\partial \theta} \ln f(X, \theta) \right] = 0.$
2.  $I(\theta) = -E \left[ \frac{\partial^2}{\partial \theta^2} \ln f(X, \theta) \right].$

**Demostración.** Por simplicidad en la escritura supondremos el caso continuo. La prueba es análoga en el caso discreto.

1. Suponiendo válido el intercambio de derivada e integral, tenemos que

$$\begin{aligned}
 E \left[ \frac{\partial}{\partial \theta} \ln f(X, \theta) \right] &= \int_{\mathbb{R}} f(x, \theta) \frac{\partial}{\partial \theta} \ln f(x, \theta) dx \\
 &= \int_{\mathbb{R}} \frac{\partial}{\partial \theta} f(x, \theta) dx \\
 &= \frac{d}{d\theta} \int_{\mathbb{R}} f(x, \theta) dx \\
 &= 0.
 \end{aligned}$$

2. Por el primer resultado, derivando por segunda vez respecto de  $\theta$ , tenemos que

$$\begin{aligned}
 0 &= \frac{\partial}{\partial \theta} E \left[ \frac{\partial}{\partial \theta} \ln f(X, \theta) \right] \\
 &= \frac{\partial}{\partial \theta} \int_{\mathbb{R}} \left[ \frac{\partial}{\partial \theta} \ln f(x, \theta) \right] f(x, \theta) dx \\
 &= \int_{\mathbb{R}} \left[ \left( \frac{\partial^2}{\partial \theta^2} \ln f(x, \theta) \right) f(x, \theta) + \left( \frac{\partial}{\partial \theta} \ln f(x, \theta) \right) \left( \frac{\partial}{\partial \theta} e^{\ln f(x, \theta)} \right) \right] dx \\
 &= \int_{\mathbb{R}} \left[ \left( \frac{\partial^2}{\partial \theta^2} \ln f(x, \theta) \right) f(x, \theta) + \left( \frac{\partial}{\partial \theta} \ln f(x, \theta) \right)^2 f(x, \theta) \right] dx \\
 &= E \left[ \frac{\partial^2}{\partial \theta^2} \ln f(X, \theta) \right] + I_X(\theta).
 \end{aligned}$$

■

Observemos entonces que, como consecuencia de la definición y el primer inciso del resultado anterior, la información de Fisher se puede escribir de la siguiente forma:

$$I(\theta) = \text{Var} \left[ \frac{\partial}{\partial \theta} \ln f(X, \theta) \right].$$

La definición de información de Fisher de una variable aleatoria, o de su distribución, se puede extender fácilmente para el caso de vectores aleatorios, y en particular para muestras aleatorias. Este es el contenido de la siguiente definición y es completamente análoga al caso unidimensional.

**Definición 2.24** Sea  $(X_1, \dots, X_n)$  un vector aleatorio con función de densidad o de probabilidad  $f(x_1, \dots, x_n, \theta)$ , dependiente de un parámetro desconocido  $\theta$ . La información de Fisher del vector  $(X_1, \dots, X_n)$ , o de su distribución, es la función

$$I_{X_1, \dots, X_n}(\theta) = E \left[ \left( \frac{\partial}{\partial \theta} \ln f(X_1, \dots, X_n, \theta) \right)^2 \right]. \quad (2.15)$$

La información de Fisher es una medida de la cantidad de información que una observación del vector aleatorio contiene acerca del parámetro  $\theta$ . Como en el caso unidimensional, observe con cuidado la expresión  $f(X_1, \dots, X_n, \theta)$ , la cual es la función de densidad del vector aleatorio evaluada en el vector mismo. Supondremos que tal expresión es una variable aleatoria y que las operaciones indicadas en (2.15) pueden efectuarse.

Los dos resultados relativos a la información de Fisher presentados en la Proposición 2.6 pueden extenderse al caso de vectores aleatorios. Este es el contenido del siguiente resultado el cual se demuestra de manera análoga al caso unidimensional.

**Proposición 2.7** Sea  $(X_1, \dots, X_n)$  un vector aleatorio con función de densidad o de probabilidad  $f(x, \theta)$ , dependiente de un parámetro  $\theta$ . Bajo las condiciones de regularidad,

1.  $E\left[\frac{\partial}{\partial\theta} \ln f_{X_1, \dots, X_n}(X_1, \dots, X_n, \theta)\right] = 0.$
2.  $I_{X_1, \dots, X_n}(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2} \ln f_{X_1, \dots, X_n}(X_1, \dots, X_n, \theta)\right].$

Bajo la hipótesis de independencia de las variables de un vector y suponiendo que cada una de las distribuciones individuales depende de un mismo parámetro  $\theta$ , en donde estas distribuciones individuales no son necesariamente idénticas, se obtiene el siguiente resultado.

**Proposición 2.8** Sean  $X_1, \dots, X_n$  variables aleatorias independientes, con función de densidad o de probabilidad  $f_1(x; \theta), \dots, f_n(x; \theta)$ , y con información de Fisher  $I_1(\theta), \dots, I_n(\theta)$ , respectivamente. Bajo las condiciones de regularidad para cada una de estas distribuciones,

$$I_{X_1, \dots, X_n}(\theta) = I_1(\theta) + \dots + I_n(\theta).$$

***Demostración.*** El resultado es consecuencia directa de la hipótesis de independencia. Tenemos que

$$\begin{aligned}
I_{X_1, \dots, X_n}(\theta) &= E \left[ \left( \frac{\partial}{\partial \theta} \ln f(X_1, \dots, X_n; \theta) \right)^2 \right] \\
&= E \left[ \left( \frac{\partial}{\partial \theta} \ln f_1(X_1; \theta) \cdots f_n(X_n; \theta) \right)^2 \right] \\
&= E \left[ \left( \sum_{i=1}^n \frac{\partial}{\partial \theta} \ln f_i(X_i; \theta) \right)^2 \right] \\
&= \sum_{i=1}^n E \left[ \left( \frac{\partial}{\partial \theta} \ln f_i(X_i, \theta) \right)^2 \right] \\
&\quad + \sum_{i \neq j} E \left( \frac{\partial}{\partial \theta} \ln f_i(X_i, \theta) \right) E \left( \frac{\partial}{\partial \theta} \ln f_j(X_j, \theta) \right),
\end{aligned}$$

en donde sabemos que la variable aleatoria  $\frac{\partial}{\partial \theta} \ln f_i(X_i, \theta)$  tiene esperanza cero y, en consecuencia, la segunda suma desaparece. La primera suma es igual a  $I_1(\theta) + \cdots + I_n(\theta)$ . ■

En particular, cuando el vector aleatorio constituye una muestra aleatoria, es decir, cuando se tiene la hipótesis de independencia e idéntica distribución dependiente de un parámetro  $\theta$ , se obtiene la siguiente expresión para la información de Fisher de una muestra aleatoria.

**Corolario 2.1** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución dependiente de un parámetro  $\theta$  y la cual satisface las condiciones de regularidad. Entonces

$$I_{X_1, \dots, X_n}(\theta) = n \cdot I_{X_1}(\theta).$$

**Demostración.** Por la independencia y la idéntica distribución,

$$\begin{aligned}
I_{X_1, \dots, X_n}(\theta) &= I_{X_1}(\theta) + \cdots + I_{X_n}(\theta) \\
&= n \cdot I_{X_1}(\theta).
\end{aligned}$$

Para concluir esta sección, demostraremos una relación entre la información de Fisher de una muestra aleatoria y la información de Fisher de cualquier

estadística obtenida de la misma muestra aleatoria. A partir de este resultado, puede obtenerse una interpretación del concepto de suficiencia de una estadística.

**Teorema 2.4** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución dependiente de un parámetro  $\theta$  y sea  $T$  una estadística. Entonces

1.  $I_{X_1, \dots, X_n}(\theta) \geq I_T(\theta)$ .
2. La igualdad se cumple  $\Leftrightarrow T$  es suficiente para  $\theta$ .

*Demostración.* Tenemos que

$$\begin{aligned} I_{X_1, \dots, X_n}(\theta) &= -E\left[\frac{\partial^2}{\partial \theta^2} \ln f(X_1, \dots, X_n)\right] \\ &= -\int_{\mathbb{R}^n} \left[\frac{\partial^2}{\partial \theta^2} \ln f(x^n)\right] f(x^n) dx^n. \end{aligned}$$

Cuando la muestra aleatoria toma el valor  $x^n$ , la estadística  $T$  toma el valor  $T(x^n)$ , de modo que el evento  $T = T(x^n)$  se cumple. A continuación añadimos esta información redundante al cálculo anterior y condicionamos respecto a este evento,

$$\begin{aligned} I_{X_1, \dots, X_n}(\theta) &= -\int_{\mathbb{R}^n} \left[\frac{\partial^2}{\partial \theta^2} \ln f(x^n, T(x^n))\right] f(x^n) dx^n \\ &= -\int_{\mathbb{R}^n} \left[\frac{\partial^2}{\partial \theta^2} \ln f(x^n | T(x^n)) f_T(T(x^n))\right] f(x^n) dx^n \\ &= -\int_{\mathbb{R}^n} \left[\frac{\partial^2}{\partial \theta^2} \ln f(x^n | T(x^n))\right] f(x^n) dx^n \\ &\quad - \int_{\mathbb{R}^n} \left[\frac{\partial^2}{\partial \theta^2} \ln f_T(T(x^n))\right] f(x^n) dx^n \\ &= I_T(\theta) + I_{X_1, \dots, X_n | T}(\theta) \\ &\geq I_T(\theta). \end{aligned}$$

Esto demuestra la primera afirmación. Veamos ahora el segundo resultado. Por lo demostrado antes,

$$\begin{aligned}
 I_{X_1, \dots, X_n}(\theta) = I_T(\theta) &\Leftrightarrow \int_{\mathbb{R}^n} \left( \frac{\partial}{\partial \theta} \ln f(x^n | T(x^n)) \right)^2 f(x^n) dx^n = 0 \\
 &\Leftrightarrow \frac{\partial}{\partial \theta} \ln f(x^n | T(x^n)) = 0 \\
 &\Leftrightarrow \frac{\frac{\partial}{\partial \theta} f(x^n | T(x^n))}{f(x^n | T(x^n))} = 0 \\
 &\Leftrightarrow \frac{\partial}{\partial \theta} f(x^n | T(x^n)) = 0 \\
 &\Leftrightarrow f(x^n | T(x^n)) \text{ no depende de } \theta \\
 &\Leftrightarrow T \text{ es suficiente para } \theta.
 \end{aligned}$$

■

Así, tenemos la siguiente interpretación: una estadística es suficiente si, y sólo si, captura toda la información de Fisher de la muestra aleatoria para estimar el parámetro  $\theta$ .

Por otro lado, este resultado también nos provee de un mecanismo alternativo para demostrar que una estadística es suficiente: su información debe coincidir con la información de la muestra aleatoria. Veamos algunos ejemplos.

**Ejemplo 2.41** Anteriormente comprobamos que la estadística  $T = X_1 + \dots + X_n$  es suficiente para el parámetro  $\theta$  en el caso particular de las distribuciones  $\text{Ber}(\theta)$  y  $\text{Poisson}(\theta)$ . Se demostró lo anterior de dos maneras: una mediante la definición de suficiencia y otra mediante el teorema de factorización. Comprobaremos por tercera ocasión esta afirmación, ahora usando la información de Fisher.

- Puede comprobarse que la información de Fisher de una variable aleatoria  $X$  con distribución  $\text{Ber}(\theta)$  es  $I_X(\theta) = 1/[\theta(1 - \theta)]$ , para  $0 < \theta < 1$ . Por lo tanto, la información de Fisher de una muestra aleatoria  $X_1, \dots, X_n$  de esta distribución es  $I(\theta) = n I_X(\theta) = n/[\theta(1 - \theta)]$ . Esta es exactamente la información de Fisher de la estadística  $T$  cuya distribución es  $\text{bin}(n, \theta)$ . Por lo tanto,  $T$  es suficiente.

- La información de Fisher de una variable aleatoria  $X$  con distribución Poisson( $\theta$ ) es  $I_X(\theta) = 1/\theta$ , para  $\theta > 0$ . Por lo tanto, la información de Fisher de una muestra aleatoria  $X_1, \dots, X_n$  de esta distribución es  $I(\theta) = n I_X(\theta) = n/\theta$ . Por otro lado, puede comprobarse que, como  $T$  tiene distribución Poisson( $n\theta$ ), su información de Fisher respecto del parámetro  $\theta$  es

$$I(\theta) = E \left[ \left( \frac{\partial}{\partial \theta} \ln f(T, n\theta) \right)^2 \right] = \frac{n}{\theta}.$$

Como esto coincide con la información de Fisher de la muestra aleatoria, concluimos que  $T$  es suficiente. ■

Para terminar esta sección señalaremos que la información de Fisher puede también definirse para distribuciones dependientes de dos o más parámetros. Esta extensión y otros resultados pueden consultarse, por ejemplo, en [23].

## Ejercicios

214. **Algunas propiedades de la información de Fisher.** Sea  $X$  una variable aleatoria con distribución dependiente de un parámetro  $\theta$ . Sean  $a$  y  $b$  dos constantes con  $a \neq 0$ . Demuestre las siguientes propiedades.

- $I_X(\theta) \geq 0$ .
- $I_{aX}(\theta) = I_X(\theta)$ .
- $I_{X+b}(\theta) = I_X(\theta)$ .

215. **Reparametrización.** Sea  $X$  una variable aleatoria con función de densidad o de probabilidad  $f(x, \theta)$  dependiente de un parámetro  $\theta$  y con información de Fisher  $I(\theta)$ . Suponga que se tiene una reparametrización  $\theta = \varphi(\eta)$ , en donde  $\varphi$  es una función biyectiva y diferenciable. La función  $f(x, \theta)$  ahora se escribe como  $f(x, \varphi(\eta))$ , en donde  $\eta$  es el nuevo parámetro. Demuestre que la información de Fisher de  $X$  respecto del parámetro  $\eta$  es

$$I(\eta) = (\varphi'(\eta))^2 \cdot I(\theta) \Big|_{\theta=\varphi(\eta)}.$$

216. **Distribución Bernoulli: reparametrización.** Sea  $X$  una variable aleatoria con distribución  $\text{Ber}(\theta)$ . La información de Fisher de  $X$  es, para  $0 < \theta < 1$ ,

$$I(\theta) = \frac{1}{\theta(1-\theta)}.$$

Considere la reparametrización  $\theta = \varphi(\eta) = e^\eta/(1 + e^\eta)$ , en donde  $\eta$  es un nuevo parámetro. Demuestre que, respecto de  $\eta$ , la información de Fisher de  $X$  es, para  $-\infty < \eta < \infty$ ,

$$I(\eta) = \varphi(\eta)(1 - \varphi(\eta)).$$

217. **Proceso de Poisson.** Sea  $\{X_t : t \geq 0\}$  un proceso de Poisson de parámetro  $\theta > 0$  como se definió en la página 126 Sean  $0 < t_1 < \dots < t_n$  tiempos cualesquiera. Encuentre la información de Fisher del vector aleatorio  $(X_{t_1}, \dots, X_{t_n})$ .
218. **Movimiento browniano.** Sea  $\{X_t : t \geq 0\}$  un movimiento browniano de parámetro  $\theta > 0$  como se definió en la página 127. Sean  $0 < t_1 < \dots < t_n$  tiempos cualesquiera. Encuentre la información de Fisher del vector aleatorio  $(X_{t_1}, \dots, X_{t_n})$ .
219. **Varias distribuciones.** Compruebe que la información de Fisher es la indicada para cada una de las distribuciones de probabilidad que aparecen en la tabla de la Figura 2.8.
220. **Distinta distribución, misma información de Fisher.** Demuestre que las siguientes dos distribuciones tienen la misma información de Fisher. Para  $\theta > 0$ ,

$$a) f(x, \theta) = \begin{cases} \theta x^{\theta-1} & \text{si } 0 < x < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

$$b) f(x, \theta) = \frac{\theta}{2} e^{-\theta|x|}, \quad -\infty < x < \infty.$$

221. Calcule la información de Fisher de una variable aleatoria  $X$  con la siguiente distribución: para  $-1 < \theta < 1$ ,

$$f(x, \theta) = \begin{cases} \frac{1 + \theta x}{2} & \text{si } -1 < x < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

222. Calcule la información de Fisher de una variable aleatoria  $X$  con la siguiente distribución: para  $\theta > 0$ ,

$$f(x, \theta) = \begin{cases} \frac{2x}{\theta} e^{-x^2/\theta} & \text{si } x > 0, \\ 0 & \text{en otro caso.} \end{cases}$$

223. Calcule la información de Fisher de una muestra aleatoria  $X_1, \dots, X_n$  de una distribución dependiente de un parámetro  $\theta$  como se indica en cada inciso.

- a)  $N(\theta, \sigma^2)$ .
- b)  $\text{gama}(\gamma, \theta)$ .

224. Usando la información de Fisher, determine si las siguientes estadísticas son suficientes para el parámetro desconocido  $\theta$ .

- a)  $T = X_1^2 + \dots + X_n^2$  para el parámetro  $\theta > 0$  de la distribución Rayleigh especificada abajo y un tamaño de muestra  $n$ .

$$f(x, \theta) = \begin{cases} \frac{2x}{\theta} e^{-x^2/\theta} & \text{si } x > 0, \\ 0 & \text{en otro caso.} \end{cases}$$

- b)  $T = X_1 + 2X_2$  para el parámetro de la distribución  $N(\theta, 1)$ , para un tamaño de muestra  $n = 2$ .
- c)  $T = X_1 - X_2$  para el parámetro de la distribución  $\text{Poisson}(\theta)$ , para un tamaño de muestra  $n = 2$ .

## 2.12. Suficiencia conjunta

En esta breve sección extenderemos el concepto de suficiencia de una estadística para un parámetro al caso de varias estadísticas para varios parámetros. Consideraremos entonces que  $T$  es un vector de estadísticas y  $\theta$  es un

vector de parámetros, no necesariamente de la misma dimensión.

Partiremos nuevamente de una muestra aleatoria  $X_1, \dots, X_n$  de una distribución  $f(x, \theta)$  dependiente de  $\ell$  parámetros  $\theta = (\theta_1, \dots, \theta_\ell)$ . Las definiciones y resultados son análogos al caso unidimensional, aunque pueden surgir ahora nuevas situaciones para las coordenadas del vector de estadísticas y las coordenadas del vector de parámetros.

**Definición 2.25** Se dice que las variables de un vector de estadísticas  $T = (T_1, \dots, T_k)$  son **suficientes conjuntamente** para el vector de parámetros  $\theta = (\theta_1, \dots, \theta_\ell)$  si y sólo si la distribución de la muestra aleatoria  $X_1, \dots, X_n$  condicionada al evento  $T = (t_1, \dots, t_k)$  no depende de  $\theta$ .

Así, por ejemplo, podemos tener las siguientes situaciones:

- $(T_1, T_2)$  es suficiente para  $(\theta_1, \theta_2)$ .
- $(T_1, T_2)$  es suficiente para  $\theta$  (unidimensional).

**Ejemplo 2.42** Cada variable de una muestra aleatoria  $X_1, \dots, X_n$  puede considerarse como una estadística. En efecto, la variable  $X_i$  puede verse como la proyección sobre la  $i$ -ésima coordenada de la muestra aleatoria, esto es,  $T_i(X_1, \dots, X_n) = X_i$ . Así, podemos formar el vector de  $n$  estadísticas  $T = (X_1, \dots, X_n)$ . Es intuitivamente claro, y se puede comprobar sin mucha dificultad, que  $T$  es suficiente para cualquier parámetro o vector de parámetros  $\theta$  del cual dependa la distribución en estudio. También puede demostrarse que el vector de estadísticas de orden  $T = (X_{(1)}, \dots, X_{(n)})$  es siempre suficiente para  $\theta$ . En la sección de ejercicios se pide demostrar estas afirmaciones. ■

**Ejemplo 2.43** Si no tomamos la totalidad de la muestra aleatoria y consideramos que  $T = (X_1, \dots, X_k)$ , en donde  $k < n$ , puede comprobarse que, en

general,  $T$  no es suficiente para  $\theta$ . De hecho, cualquier vector que se pueda formar con un subconjunto propio del conjunto de variables de la muestra aleatoria no será, en general, suficiente para  $\theta$ . De la misma manera, el vector de las primeras  $k$  estadísticas de orden, con  $k < n$ , no es, en general, suficiente para  $\theta$ . Más generalmente, cualquier vector que se pueda formar con cualesquiera  $k$  estadísticas de orden, no será, en general, suficiente para  $\theta$ . ■

El bastante útil teorema de factorización de Jerzy Neyman puede extenderse sin dificultad al caso de vectores de estadísticas. Aquí tenemos el enunciado haciendo uso de la notación  $x^n = (x_1, \dots, x_n)$  y para el caso de suficiencia conjunta de dos parámetros.

**Teorema 2.5 (Teorema de factorización)** Un vector de estadísticas  $(T_1, T_2)$  es suficiente para el parámetro o vector de parámetros  $\theta$  si y sólo si

$$f(x^n, \theta) = g(T_1(x^n), T_2(x^n), \theta) \cdot h(x^n),$$

en donde  $g$  y  $h$  son dos funciones no negativas que dependen únicamente de los argumentos indicados.

Por brevedad en la escritura hemos considerado el caso bidimensional  $(T_1, T_2)$  pero, dejando de lado la longitud de las expresiones, no hay mayor dificultad para enunciar y demostrar el resultado en el caso de vectores de estadísticas  $(T_1, \dots, T_k)$ . La demostración es completamente análoga al caso unidimensional presentada antes. Veamos algunos ejemplos.

**Ejemplo 2.44** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $N(\mu, \sigma^2)$ , en donde  $\mu$  y  $\sigma^2$  son ambos desconocidos. Definamos el vector de estadísticas

$$(T_1, T_2) = \left( \sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2 \right).$$

Demostraremos que  $(T_1, T_2)$  es suficiente para  $(\mu, \sigma^2)$ . Si se quisiera usar la definición de suficiencia conjunta, se tendría que considerar un posible valor

$(t_1, t_2)$  de  $(T_1, T_2)$  y demostrar que la expresión

$$f(x_1, \dots, x_n | t_1, t_2) = \frac{f(x_1, \dots, x_n, t_1, t_2)}{f(t_1, t_2)}$$

no depende de  $\mu$  ni de  $\sigma^2$ . Sin embargo, encontrar la expresión anterior no es sencillo. Utilizaremos entonces el teorema de factorización. Tenemos que

$$\begin{aligned} L(x^n, \mu, \sigma^2) &= \left( \frac{1}{2\pi\sigma^2} \right)^{n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right) \\ &= \left( \frac{1}{2\pi} \right)^{n/2} \cdot \left( \frac{1}{\sigma^2} \right)^{n/2} \exp\left(-\frac{1}{2\sigma^2} \left( \sum_{i=1}^n x_i^2 - 2\mu \sum_{i=1}^n x_i + n\mu^2 \right)\right). \end{aligned}$$

El primer factor es la función constante  $h(x^n)$  y el resto de la expresión corresponde a una función  $g(T_1(x^n), T_2(x^n), \mu, \sigma^2)$ . Por lo tanto,  $(T_1, T_2)$  es suficiente para  $(\mu, \sigma^2)$ . ■

Cuando a un vector de estadísticas suficientes conjuntamente se le aplica una función biyectiva se obtiene otro vector que preserva la propiedad de ser suficiente. Este resultado es análogo al caso unidimensional y se enuncia a continuación. Su demostración es idéntica al caso estudiado antes y se deja como ejercicio.

**Proposición 2.9** Funciones biyectivas de estadísticas suficientes conjuntas son también suficientes.

**Ejemplo 2.45** En el ejemplo anterior se comprobó que el vector de estadísticas  $(T_1, T_2) = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$  es suficiente para el vector de parámetros  $(\mu, \sigma^2)$  en una distribución normal. La transformación

$$(t_1, t_2) \mapsto \left( \frac{t_1}{n}, \frac{nt_2 - t_1^2}{n(n-1)} \right)$$

resulta ser una función biyectiva sobre el espacio parametral  $\Theta = (-\infty, \infty) \times (0, \infty)$ . Después de un cálculo sencillo puede comprobarse que cuando esta función se aplica al vector  $(T_1, T_2)$  se obtiene el vector  $(\bar{X}, S^2)$ . Por lo tanto, este nuevo vector de estadísticas también es suficiente para  $(\mu, \sigma^2)$ . ■

Concluimos esta sección con algunas observaciones generales que no son difíciles de verificar.

- Si  $T_1$  es suficiente para  $\theta$  y si  $T_2$  es otra estadística, entonces  $(T_1, T_2)$  es suficiente para  $\theta$ .
- Si  $T_1$  es suficiente para  $\theta_1$  y si  $T_2$  es suficiente para  $\theta_2$ , entonces  $(T_1, T_2)$  es suficiente para  $(\theta_1, \theta_2)$ .
- Si  $(T_1, T_2)$  es suficiente para  $(\theta_1, \theta_2)$ , entonces no necesariamente  $T_1$  es suficiente para  $\theta_1$  ni  $T_2$  es suficiente para  $\theta_2$ . Por ejemplo, se comprobó que el vector  $(\bar{X}, S^2)$  es suficiente para  $(\mu, \sigma^2)$  en una distribución normal. Es inmediato verificar que el vector en el orden contrario  $(S^2, \bar{X})$  también es suficiente para  $(\mu, \sigma^2)$ . La suficiencia coordinada a coordinada diría que  $S^2$  es suficiente individualmente para  $\mu$  y que  $\bar{X}$  es suficiente para  $\sigma^2$ . Estas afirmaciones son falsas.

### Ejemplos de estadísticas suficientes: dos parámetros

Distribución	Parámetros	Estadística suficiente
$\text{unif}(\theta_1, \theta_2)$	$(\theta_1, \theta_2)$	$T = (X_{(1)}, X_{(n)})$
$N(\theta_1, \theta_2)$	$(\theta_1, \theta_2)$	$T = (\bar{X}, S^2)$
$\text{gama}(\theta_1, \theta_2)$	$(\theta_1, \theta_2)$	$T = \left( \prod_{i=1}^n X_i, \sum_{i=1}^n X_i \right)$

Tabla 2.9

## Ejercicios

225. **Muestra completa.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución dependiente de un parámetro o vector de parámetros des-

conocido  $\theta$ . Demuestre que el vector de estadísticas  $T = (X_1, \dots, X_n)$  es siempre suficiente para  $\theta$ .

226. **Estadísticas de orden.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución dependiente de un parámetro o vector de parámetros desconocido  $\theta$ . Demuestre que el vector de estadísticas de orden  $T = (X_{(1)}, \dots, X_{(n)})$  es siempre suficiente para  $\theta$ .

227. **Distribución Bernoulli.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{Ber}(\theta)$ , con  $\theta$  desconocido. Demuestre que el vector de estadísticas  $(T_1, T_2)$  es suficiente para  $\theta$ , en donde, para  $1 \leq k \leq n-1$ ,

$$\begin{aligned} T_1 &= X_1 + \dots + X_k, \\ T_2 &= X_{k+1} + \dots + X_n. \end{aligned}$$

228. Demuestre que toda función biyectiva de un vector suficiente de estadísticas  $(T_1, \dots, T_k)$  para un vector de parámetros  $(\theta_1, \dots, \theta_\ell)$  es también suficiente. Este es el contenido de la Proposición 2.9.

229. **Información adicional.** Sea  $(T_1, \dots, T_k)$  suficiente para  $(\theta_1, \dots, \theta_\ell)$ . Suponga que  $T_{k+1}$  es una estadística adicional. Demuestre que el vector  $(T_1, \dots, T_{k+1})$  también es suficiente para  $(\theta_1, \dots, \theta_\ell)$ .

### 2.13. Suficiencia minimal

Como hemos visto antes, la cualidad de ser suficiente para una estadística significa que ésta preserva de manera completa la información de la muestra aleatoria para estimar un parámetro desconocido. Sin embargo, pueden existir varias estadísticas suficientes para un mismo parámetro y es posible buscar entre éstas alguna que sea más compacta en un sentido que explicaremos más adelante. A una estadística con esta propiedad se le llama suficiente minimal.

Para precisar el concepto de minimalidad para una estadística suficiente definiremos primero cuándo una estadística es función de otra. Recordemos que  $x^n$  denota el punto muestral  $(x_1, \dots, x_n)$ .

**Definición 2.26** Se dice que una estadística  $T$  es una **función** de otra estadística  $S$  si para cualesquiera dos valores  $x^n$  y  $y^n$  de una muestra aleatoria se cumple la implicación

$$S(x^n) = S(y^n) \Rightarrow T(x^n) = T(y^n).$$

Esta definición de función puede parecer un poco extraña, pero realmente no lo es. El siguiente argumento compatibiliza esta definición con la noción usual de función: recordemos que una relación  $\tau$  de un conjunto  $A$  en un conjunto  $B$  es una función si para cada elemento  $a$  en  $A$  existe un único elemento  $b$  en  $B$  tal que  $\tau(a) = b$ . De manera equivalente,  $\tau$  es una función si se cumple la implicación:  $\tau(a_1) \neq \tau(a_2) \Rightarrow a_1 \neq a_2$ . Por lo tanto,

$$\begin{aligned} T \text{ es una función } \tau \text{ de } S &\Leftrightarrow [\tau(S(x^n)) \neq \tau(S(y^n)) \Rightarrow S(x^n) \neq S(y^n)] \\ &\Leftrightarrow [S(x^n) = S(y^n) \Rightarrow \tau(S(x^n)) = \tau(S(y^n))] \\ &\Leftrightarrow [S(x^n) = S(y^n) \Rightarrow T(x^n) = T(y^n)]. \end{aligned}$$

La última condición es la que aparece en la Definición 2.26 y de esta manera hemos comprobado que es equivalente a la noción usual de función.

Observemos que no hay restricción alguna sobre las dimensiones de las estadísticas  $T$  y  $S$  en la Definición 2.26, de modo que éstas pueden ser vectores de estadísticas de cualquier dimensión. Además, estas dimensiones no necesariamente deben ser coincidentes. Por ejemplo, supongamos que  $S$  es la estadística dada por el vector de la muestra aleatoria, es decir,  $S = (X_1, \dots, X_n)$ . Entonces es claro que toda estadística o vector de estadísticas es función de esta estadística  $S$ . Veamos otros ejemplos.

**Ejemplo 2.46** Sea  $X_1, \dots, X_n$  una muestra aleatoria. Entonces

- La estadística  $T = (X_1 + \dots + X_n)/n$  es función de la estadística  $S = X_1 + \dots + X_n$ , pues  $T = S/n$ .
- El vector de estadísticas  $T = (X_{(1)}, \dots, X_{(n)})$  es función del vector de estadísticas  $S = (X_1, \dots, X_n)$ , pues cada coordenada de  $T$  se puede

escribir en términos de las coordenadas de  $S$ . Por ejemplo, la primera de ellas es  $X_{(1)} = \min \{X_1, \dots, X_n\}$ .

- La estadística  $T = X_1 + \dots + X_n$  es función del vector de estadísticas de orden  $S = (X_{(1)}, \dots, X_{(n)})$ , pues podemos expresar a  $T$  como la suma  $X_{(1)} + \dots + X_{(n)}$ .

▪

Otros ejemplos de situaciones en donde una estadística es, o no es, función de otra estadística se muestran en la sección de ejercicios.

Una observación importante sobre la situación cuando una estadística es función de otra estadística es la siguiente: si  $T$  es función de  $S$ , es decir, si  $T = \tau \circ S$ , entonces la cardinalidad del conjunto de valores de  $T$  es menor o igual a la cardinalidad del conjunto de valores de  $S$ . Esta afirmación es más fácil de comprender si se considera el caso cuando estos rangos de valores son finitos: si  $S$  toma  $k$  valores, entonces  $T$  toma a lo sumo  $k$  valores. En este sentido, consideraremos que  $T$  es más pequeña o más compacta que  $S$ , y esto es una interpretación al concepto de suficiencia minimal que definiremos a continuación.

**Definición 2.27** Se dice que una estadística  $T$  es **suficiente minimal** para un parámetro  $\theta$  si cumple las siguientes dos condiciones:

- a)  $T$  es suficiente para  $\theta$ .
- b)  $T$  es minimal, es decir, es función de cualquier otra estadística suficiente para  $\theta$ .

Por lo tanto, si  $T$  es suficiente minimal, entonces para cada estadística suficiente  $S$  existe una función  $\tau$  tal que  $T$  se puede escribir como la composición  $\tau \circ S$ . Y en consecuencia, la cardinalidad del conjunto de valores de  $T$  es menor o igual a la cardinalidad del conjunto de valores de cualquier estadística suficiente  $S$ . En otras palabras,  $T$  es suficiente minimal si es una estadística

suficiente con el número más pequeño posible de valores. De este hecho proviene el adjetivo minimal en la definición anterior. En general, puede haber varias estadísticas suficientes minimales para un parámetro. Retomaremos este tema más adelante.

Regresando a la definición de suficiencia minimal, observemos que la aplicación directa de la definición puede ser una tarea difícil pues, por la segunda condición, debe comprobarse que la estadística suficiente minimal es función de cualquier otra estadística suficiente. Afortunadamente se cuenta con el siguiente resultado que establece condiciones un tanto más sencillas de comprobar que garantizan la suficiencia minimal de una estadística.

**Teorema 2.6 (Criterio para suficiencia minimal)** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución con función de densidad o de probabilidad  $f(x, \theta)$ , dependiente de un parámetro  $\theta$ . Sea  $T$  una estadística y sean  $x^n$  y  $y^n$  dos valores cualesquiera de la muestra aleatoria. Si se cumplen las dos implicaciones

$$\left[ \frac{f(x^n, \theta)}{f(y^n, \theta)} \text{ no depende de } \theta \right] \Leftrightarrow \left[ T(x^n) = T(y^n) \right], \quad (2.16)$$

entonces  $T$  es suficiente minimal para  $\theta$ .

**Demostración.** Demostraremos primero la suficiencia y para ello utilizaremos el teorema de factorización. Sea  $x^n$  un valor cualquiera de la muestra aleatoria y supongamos que  $t$  es su imagen bajo la estadística  $T$ , es decir,  $T(x^n) = t$ . Sea  $y^n$  otro posible valor de la muestra aleatoria tal que  $T(y^n) = t$ . Este otro valor de la muestra aleatoria no necesariamente es distinto de  $x^n$ , pues puede ser que no haya otro valor con esa propiedad. Es importante observar que, por el orden en que fueron considerados estos objetos,  $y^n$  depende de  $x^n$  únicamente a través del valor  $t$ . Esto se ilustra en la Figura 2.11.

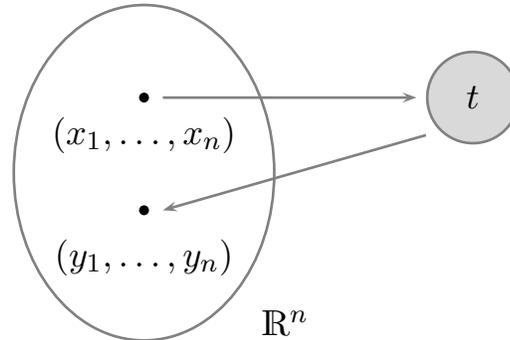


Figura 2.11

Por construcción, se cumple que  $T(x^n) = T(y^n) = t$  y haciendo uso de la implicación de derecha a izquierda de la hipótesis se obtiene que el cociente  $f(x^n, \theta)/f(y^n, \theta)$  no depende de  $\theta$ , es decir,

$$\frac{f(x^n, \theta)}{f(y^n, \theta)} = h_0(x^n, y^n),$$

para alguna función  $h_0$  no negativa, dependiente únicamente de los argumentos indicados. Por lo tanto,

$$\begin{aligned} f(x^n, \theta) &= f(y^n, \theta) \cdot h_0(x^n, y^n) \\ &= g(T(x^n), \theta) \cdot h(x^n), \end{aligned}$$

en donde el factor  $f(y^n, \theta)$  se ha escrito como una función no negativa  $g(T(x^n), \theta)$ , pues  $y^n$  depende de  $x^n$  únicamente a través de  $T(x^n)$ . El segundo factor es una función  $h(x^n)$  dependiente únicamente de  $x^n$  pues, nuevamente, observamos que  $y^n$  depende de  $x^n$ . El teorema de factorización garantiza la suficiencia. Para los puntos muestrales  $x^n$  en donde no existe otro punto muestral tal que bajo  $T$  tome el valor  $t$ , cualquier función de  $x^n$  es función de  $T(x^n) = t$  y las afirmaciones anteriores se cumplen.

Ahora veamos la minimalidad. Sea  $S$  otra estadística suficiente para  $\theta$ . Por el teorema de factorización, para cualquier valor  $x^n$  de la muestra aleatoria,

$$f(x^n, \theta) = g(S(x^n), \theta) \cdot h(x^n),$$

para ciertas funciones no negativas  $g$  y  $h$ . Sean  $x^n$  y  $y^n$  dos valores de la muestra aleatoria tales que  $S(x^n) = S(y^n)$ . Demostraremos que  $T(x^n) = T(y^n)$ . Tenemos que

$$\begin{aligned} \frac{f(x^n, \theta)}{f(y^n, \theta)} &= \frac{g(S(x^n), \theta) \cdot h(x^n)}{g(S(y^n), \theta) \cdot h(y^n)} \\ &= \frac{h(x^n)}{h(y^n)}. \end{aligned}$$

Esto significa que este cociente no depende de  $\theta$ . Usando la implicación de izquierda a derecha de la hipótesis, se obtiene que  $T(x^n) = T(y^n)$ , es decir,  $T$  es función de  $S$ . ■

Observemos que para demostrar la suficiencia se usó únicamente la implicación de derecha a izquierda de la hipótesis, mientras que para demostrar la minimalidad se usó la implicación de izquierda a derecha. Es decir, podemos establecer los resultados parciales en el sentido de suponer una de las implicaciones para obtener las propiedades por separado, aunque, por supuesto, para la minimalidad se requiere primero la suficiencia.

Por otro lado, es crucial observar el significado lógico de las dos implicaciones que aparecen en (2.16). Estas afirmaciones no establecen que si se cumple una de las condiciones para todo par de valores  $x^n$  y  $y^n$  de la muestra aleatoria, entonces se cumple la otra condición. Lo que establecen es que si para algún par de valores  $x^n$  y  $y^n$  se cumple una de las condiciones, entonces para ese par de valores muestrales se cumple la otra condición. Esto incluye la posibilidad de no existan dos puntos muestrales distintos en donde se cumpla alguna de las condiciones.

A continuación veremos algunos ejemplos en donde se muestra la utilidad del teorema anterior.

**Ejemplo 2.47** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{Ber}(\theta)$ . Hemos demostrado antes que la estadística

$$T = X_1 + \dots + X_n$$

es suficiente para  $\theta$ . Demostraremos ahora que  $T$  es además minimal. Sean  $x^n$  y  $y^n$  dos posibles valores de la muestra aleatoria. Después de algunas

simplificaciones se obtiene que

$$\frac{f(x^n, \theta)}{f(y^n, \theta)} = \left( \frac{\theta}{1 - \theta} \right)^{n\bar{x} - n\bar{y}}.$$

Por lo tanto,  $T$  es suficiente minimal pues se verifica que

$$\begin{aligned} \frac{f(x^n, \theta)}{f(y^n, \theta)} \text{ no depende de } \theta &\Leftrightarrow n\bar{x} - n\bar{y} = 0 \\ &\Leftrightarrow T(x^n) = T(y^n). \end{aligned}$$

■

El siguiente ejemplo es particularmente interesante, pues muestra una manera de usar la suficiencia minimal de una estadística para demostrar la no suficiencia de otra estadística.

**Ejemplo 2.48** Sea  $X_1, X_2, X_3$  una muestra aleatoria de tamaño  $n = 3$  de la distribución  $\text{Ber}(\theta)$ . Demostraremos que la estadística

$$S = X_1 \cdot X_2 + X_3$$

no es suficiente para  $\theta$ . Supongamos lo contrario: supongamos que  $S$  es suficiente. Como  $T = X_1 + X_2 + X_3$  es suficiente minimal,  $T$  debe ser función de  $S$ , es decir, debe cumplirse la implicación  $S(x^n) = S(y^n) \Rightarrow T(x^n) = T(y^n)$ . Sin embargo, esto no es así pues  $S(0, 0, 0) = S(0, 1, 0) = 0$  y  $T(0, 0, 0) = 0 \neq T(0, 1, 0) = 1$ . Se concluye que  $T$  no es función de  $S$  y por lo tanto  $S$  no es suficiente. ■

Veamos ahora un ejemplo en el caso vectorial.

**Ejemplo 2.49** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $N(\mu, \sigma^2)$ . Demostraremos que el vector de estadísticas

$$(T_1, T_2) = \left( \sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2 \right)$$

es suficiente minimal para el vector de parámetros  $(\mu, \sigma^2)$ . Sean  $x^n$  y  $y^n$  dos puntos muestrales cualesquiera. Después de algunos cálculos puede comprobarse que

$$\frac{f(x^n, \mu, \sigma^2)}{f(y^n, \mu, \sigma^2)} = \exp \left[ -\frac{1}{2\sigma^2} \left( \sum_{i=1}^n x_i^2 - \sum_{i=1}^n y_i^2 \right) + \frac{\mu}{\sigma^2} \left( \sum_{i=1}^n x_i - \sum_{i=1}^n y_i \right) \right].$$

$$\begin{aligned} \text{Esto no depende de } (\mu, \sigma^2) &\Leftrightarrow \text{el exponente es cero} \\ &\text{para todo valor de } \mu \text{ y } \sigma^2 \\ &\Leftrightarrow \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \quad \text{y} \quad \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i^2 \\ &\Leftrightarrow T_1(x^n) = T_1(y^n) \quad \text{y} \quad T_2(x^n) = T_2(y^n). \end{aligned}$$

■

Demostraremos a continuación que toda función biyectiva de una estadística suficiente minimal es también suficiente minimal. Este resultado es también válido en el caso vectorial.

**Proposición 2.10** Toda función biyectiva de una estadística suficiente minimal es también suficiente minimal.

**Demostración.** Veamos primero la suficiencia. Sabemos que toda función biyectiva de una estadística suficiente es también suficiente por la Proposición 2.4 de la página 175. De modo que esta propiedad ya es conocida. Ahora veamos la minimalidad. Sea  $T$  una estadística suficiente minimal y sea  $\tau$  una función biyectiva. Sea  $S$  otra estadística suficiente. Supongamos que  $x^n$  y  $y^n$  son dos puntos muestrales tales que  $S(x^n) = S(y^n)$ . Como  $T$

es minimal,  $T$  es función de  $S$ , y por lo tanto,  $T(x^n) = T(y^n)$ . Entonces  $(\tau \circ T)(x^n) = (\tau \circ T)(y^n)$ , es decir,  $\tau \circ T$  es función de  $S$ . Por lo tanto, la composición  $\tau \circ T$  es suficiente minimal. ■

Concluimos esta sección mostrando algunos ejemplos del resultado recién demostrado.

**Ejemplo 2.50** Para la distribución  $\text{Ber}(\theta)$ , sabemos que la estadística  $T = X_1 + \cdots + X_n$  es suficiente minimal para  $\theta$ . Definiendo la función biyectiva  $\tau(t) = t/n$  se obtiene que la media muestral  $\tau(T) = \bar{X}$  es también suficiente minimal. ■

**Ejemplo 2.51** Sabemos que el vector de estadísticas  $(T_1, T_2)$  dadas por

$$(T_1, T_2) = \left( \sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2 \right)$$

es suficiente minimal para el vector de parámetros de la distribución  $N(\mu, \sigma^2)$ . Considere la función  $\tau(t_1, t_2)$  especificada abajo. Puede comprobarse que  $\tau$  es biyectiva cuando se le considera definida sobre una región adecuada de  $\mathbb{R}^2$  y que  $\tau(T_1, T_2) = (\bar{X}, S^2)$ . Por lo tanto, la media y varianzas muestrales son estadísticas suficientes minimales para los parámetros de esta distribución.

$$\tau(t_1, t_2) = \left( \frac{t_1}{n}, \frac{nt_2 - t_1^2}{n(n-1)} \right).$$

Como un ejemplo general de estadística suficiente minimal, en la sección 2.19 al final del presente capítulo, se demuestra un resultado que establece que una cierta estadística es suficiente minimal para cada distribución de la familia exponencial.

## Ejercicios

230. Sea  $X_1, \dots, X_n$  una muestra aleatoria. Determine si la estadística  $T$  indicada es función de la estadística  $S$ .

- a)  $S = X_1 - X_2, \quad T = X_1 + X_2.$
- b)  $S = X_1 + X_2, \quad T = X_1.$
- c)  $S = X_1 + X_2, \quad T = X_1 + X_2 + X_3.$
- d)  $S = X_1 + \cdots + X_n, \quad T = (X_1 + \cdots + X_n)^2.$
- e)  $S = X_1 + \cdots + X_n, \quad T = X_{(n)}.$
- f)  $S = (X_1, \dots, X_n), \quad T = X_{(n)}.$
- g)  $S = (X_1, \dots, X_n), \quad T = X_{(1)}.$
- h)  $S = X_1 + \cdots + X_n, \quad T = (X_{(1)}, X_{(n)}).$
- i)  $S = (X_1, \dots, X_n), \quad T = (X_{(1)}, X_{(n)}).$
- j)  $S = X_1 + \cdots + X_k, \quad T = X_1 + \cdots + X_n, \quad 1 \leq k \leq n - 1.$

231. Sea  $X_1, \dots, X_n$  una muestra aleatoria y sean  $S, T$  y  $U$  tres estadísticas. Demuestre las siguientes afirmaciones.

- a) Transitividad: si  $U$  es función de  $T$  y  $T$  es función de  $S$ , entonces  $U$  es función de  $S$ .
- b) Simetría:  $T$  es siempre función de  $T$ .
- c) No reflexividad: si  $T$  es función de  $S$ , entonces no necesariamente  $S$  es función de  $T$ .

232. **El estimador máximo verosímil es función de cualquier estadística suficiente.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución  $f(x, \theta)$ , con  $\theta$  desconocido. Suponga que existe un único estimador  $\hat{\theta}$  para  $\theta$  por el método de máxima verosimilitud. Demuestre que  $\hat{\theta}$  es función de cualquier estadística suficiente para  $\theta$ .

233. Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución especificada abajo, en donde  $\theta$  es un parámetro desconocido. Suponga que cualquier otro parámetro que pudiera aparecer en la distribución es conocido. Demuestre directamente que la estadística  $T = X_1 + \cdots + X_n$  es suficiente minimal.

- a)  $\text{bin}(k, \theta)$ .  
 b)  $\text{Poisson}(\theta)$ .  
 c)  $\text{geo}(\theta)$ .  
 d)  $N(\theta, \sigma^2)$ .  
 e)  $\text{gama}(\gamma, \theta)$ .

234. Sea  $T$  una estadística suficiente minimal para un parámetro  $\theta$  y sea  $a \neq 0$  una constante conocida. Demuestre que las siguientes estadísticas también son suficientes minimales.

- a)  $T + a$ .  
 b)  $aT$ .  
 c)  $e^T$ .

235. **Distribución Bernoulli: no suficiencia.** Sea  $X_1, X_2, X_3$  una muestra aleatoria de tamaño  $n = 3$  de la distribución  $\text{Ber}(\theta)$ . Usando que la estadística  $T = X_1 + X_2 + X_3$  es suficiente minimal para  $\theta$ , demuestre que las siguientes estadísticas no son suficientes.

- a)  $S = X_1 + X_2$ .  
 b)  $S = X_1 + 2X_2 + 3X_3$ .  
 c)  $S = X_1 + 2X_2 + X_3$ .

236. **Distribución Bernoulli: no suficiencia.** Sea  $X_1, \dots, X_4$  una muestra aleatoria de tamaño  $n = 4$  de la distribución  $\text{Ber}(\theta)$ , con  $0 < \theta < 1$  desconocido. Usando el hecho de que  $T = X_1 + X_2 + X_3 + X_4$  es suficiente minimal para  $\theta$ , demuestre que la siguiente estadística no es suficiente.

$$S = X_1(X_2 + X_3) + X_4.$$

237. **Distribución geométrica.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{geo}(\theta)$ , con  $0 < \theta < 1$  desconocido, como se muestra abajo. Demuestre que la estadística  $T = X_1 + \dots + X_n$  es suficiente minimal para  $\theta$ .

$$f(x, \theta) = \begin{cases} \theta(1 - \theta)^x & \text{si } x = 0, 1, \dots \\ 0 & \text{en otro caso.} \end{cases}$$

238. **Distribución Poisson: no suficiencia.** Sea  $X_1, X_2$  una muestra aleatoria de tamaño  $n = 2$  de la distribución  $\text{Poisson}(\theta)$ , en donde  $\theta$  es desconocido. Usando el hecho de que  $T = X_1 + X_2$  es suficiente minimal para  $\theta$ , demuestre que la siguiente estadística no es suficiente.

$$S = X_1 - X_2.$$

239. **Distribución uniforme.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{unif}(0, \theta)$ . Demuestre que la última estadística de orden  $T = X_{(n)}$  es suficiente minimal para  $\theta$ .

240. **Distribución uniforme.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{unif}(\theta - 1, \theta + 1)$ . Demuestre que el vector  $(X_{(1)}, X_{(n)})$  es suficiente minimal para  $\theta$ .

241. **Distribución uniforme.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{unif}(\theta - 1/2, \theta + 1/2)$ , en donde  $\theta$  es desconocido. Determine si

- a)  $X_{(1)}$  es suficiente para  $\theta$ .
- b)  $X_{(n)}$  es suficiente para  $\theta$ .
- c)  $(X_{(1)}, X_{(n)})$  es suficiente para  $\theta$ .

242. **Distribución normal: no suficiencia.** Sea  $X_1, X_2$  una muestra aleatoria de tamaño  $n = 2$  de la distribución  $N(\theta, 1)$ , en donde  $\theta$  es desconocido. Usando el hecho de que  $T = X_1 + X_2$  es suficiente minimal para  $\theta$ , demuestre que la siguiente estadística no es suficiente.

$$S = X_1 + 2X_2.$$

243. **Distribución normal: suficiencia pero no minimalidad.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $N(\theta, \sigma^2)$ . Suponga que  $n$  es par y defina las estadísticas

$$\begin{aligned} T &= X_1 + \dots + X_n, \\ T_1 &= X_1 + X_3 + \dots + X_{n-1}, \\ T_2 &= X_2 + X_4 + \dots + X_n. \end{aligned}$$

Claramente  $T = T_1 + T_2$  y es inmediato comprobar que  $T$  es suficiente para  $\theta$ . Demuestre que

- a)  $(T_1, T_2)$  es suficiente para  $\theta$ .
- b)  $(T_1, T_2)$  no es suficiente minimal para  $\theta$ .

## 2.14. Métodos para probar suficiencia

A manera de resumen, en esta breve sección recolectamos los métodos mencionados antes para demostrar la propiedad de suficiencia para una estadística  $T$ .

- Use directamente la Definición 2.22.
- Aplique el teorema de factorización 2.3.
- Compruebe que la estadística  $T$  es una biyección de otra estadística que se sabe que es suficiente.
- Si se desea probar suficiencia de  $T$  para una función parametral, verifique si la estadística es suficiente para el parámetro en cuestión.
- Demuestre que la información de Fisher de la muestra aleatoria coincide con la información de Fisher de  $T$ .

En el caso que se desee probar que una estadística  $T$  no es suficiente, tenemos las siguientes opciones:

- Use directamente la Definición 2.22 proporcionando un punto muestral  $(x_1, \dots, x_n)$  y un valor  $t$  de la estadística  $T$  tal que la distribución conjunta de la muestra aleatoria evaluada en  $(x_1, \dots, x_n)$  y condicionada al evento  $(T = t)$  dependa del parámetro.
- Suponiendo conocido que otra estadística  $T'$  es suficiente minimal, se puede comprobar que  $T$  no es suficiente suponiendo que lo es y llegando a una contradicción: como  $T'$  es suficiente minimal, es función de  $T$ , es decir, la condición  $T(x^n) = T(y^n)$  implica que  $T'(x^n) = T'(y^n)$ . Así, si se pueden proveer dos puntos muestrales  $x^n$  y  $y^n$  tales que  $T(x^n) = T(y^n)$ , pero que  $T'(x^n) \neq T'(y^n)$ , entonces  $T$  no sería función de  $T'$ , y forzosamente  $T$  no sería suficiente. En el Ejemplo 2.48 que aparece en la página 204 se muestra este procedimiento.

## 2.15. Esperanza condicional

Esta sección contiene una revisión breve sobre el concepto de esperanza condicional de una variable aleatoria respecto de una sigma álgebra. Los resultados que se mencionan aquí se proporcionan sin demostración y pueden estudiarse con mayor detalle, por ejemplo, en el texto de David Williams [28]. El concepto de esperanza condicional nos será de utilidad en la búsqueda de estimadores insesgados de varianza mínima.

Sea  $(\Omega, \mathcal{F}, P)$  un espacio de probabilidad y sea  $X$  una variable aleatoria definida sobre este espacio y con esperanza finita. Sea  $\mathcal{G}$  una subsigma álgebra de  $\mathcal{F}$ , esto significa que  $\mathcal{G}$  es una sigma álgebra de subconjuntos de  $\Omega$  y que  $\mathcal{G} \subseteq \mathcal{F}$ .

**Definición 2.28** La **esperanza condicional** de  $X$  dado  $\mathcal{G}$  es una variable aleatoria que se denota por  $E(X | \mathcal{G})$  y se define mediante las siguientes tres propiedades:

1.  $E(X | \mathcal{G})$  es  $\mathcal{G}$ -medible, esto significa que  $E(X | \mathcal{G})$  es una variable aleatoria respecto de la subsigma álgebra  $\mathcal{G}$ .
2.  $E(X | \mathcal{G})$  tiene esperanza finita.
3. Para cualquier evento  $G$  en  $\mathcal{G}$ ,

$$E(E(X | \mathcal{G}) \cdot 1_G) = E(X \cdot 1_G).$$

Una de las dificultades para entender el concepto de esperanza condicional radica en que ésta no se define de manera explícita, sino a través de las tres propiedades mencionadas. En cursos avanzados de probabilidad se demuestra que la esperanza condicional existe y es la única variable aleatoria, en el sentido casi seguro, que satisface estas propiedades.

En este trabajo vamos a usar la esperanza condicional en el caso cuando la subsigma álgebra  $\mathcal{G}$  es generada por una variable aleatoria  $Y$ , es decir,

cuando

$$\mathcal{G} = \sigma(Y).$$

El término  $\sigma(Y)$  denota la mínima sigma álgebra respecto de la cual  $Y$  es variable aleatoria. En este caso, se escribe  $E(X | Y)$  en lugar de  $E(X | \mathcal{G})$ . Remarcamos esto a continuación.

**Notación.** Cuando  $\mathcal{G} = \sigma(Y)$ , en donde  $Y$  es una variable aleatoria, la esperanza condicional  $E(X | \mathcal{G})$  se escribe  $E(X | Y)$ .

Debido a la propiedad de unicidad casi segura, las igualdades o desigualdades entre una esperanza condicional y otra variable aleatoria son en el sentido casi seguro (c.s.), y a menudo omitiremos tal especificación. En general no es sencillo encontrar expresiones explícitas para la esperanza condicional o para su distribución, ni tampoco la definición implícita que hemos dado líneas arriba permite su manejo directo. La manera de trabajar con la esperanza condicional es a través de sus propiedades. Mencionaremos a continuación algunas de ellas.

- La esperanza condicional es única casi seguramente. Esto significa que si existe una variable aleatoria  $W$  que cumple las tres condiciones de la Definición 2.28, entonces  $W = E(X | \mathcal{G})$  c.s., es decir,

$$P[W = E(X | \mathcal{G})] = 1.$$

- La esperanza condicional es lineal, es decir, si  $X$  y  $Y$  son variables aleatorias con esperanza finita y  $a$  es una constante, entonces

$$E(aX + Y | \mathcal{G}) = aE(X | \mathcal{G}) + E(Y | \mathcal{G}).$$

- La esperanza condicional es monótona creciente, es decir, si  $X \leq Y$  son variables aleatorias con esperanzas finitas, entonces

$$E(X | \mathcal{G}) \leq E(Y | \mathcal{G}).$$

- La esperanza de la variable aleatoria  $E(X | \mathcal{G})$  es idéntica a la esperanza de  $X$ , es decir,

$$E(E(X | \mathcal{G})) = E(X).$$

- Si  $X$  es  $\mathcal{G}$ -medible, entonces es inmediato comprobar que  $X$  mismo cumple las tres condiciones de la Definición 2.28 y por la propiedad de unicidad tenemos que

$$E(X | \mathcal{G}) = X.$$

- Si  $X$  es independiente de  $\mathcal{G}$ , entonces

$$E(X | \mathcal{G}) = E(X).$$

- Si  $Y$  es  $\mathcal{G}$ -medible y acotada, entonces

$$E(X \cdot Y | \mathcal{G}) = Y \cdot E(X | \mathcal{G}).$$

- Si  $\mathcal{G}_1 \subseteq \mathcal{G}_2$  son dos subsigmas álgebras, entonces

$$E(E(X | \mathcal{G}_1) | \mathcal{G}_2) = E(E(X | \mathcal{G}_2) | \mathcal{G}_1) = E(X | \mathcal{G}_1).$$

Las siguientes propiedades son de particular interés en el cálculo explícito de la esperanza condicional.

- Sea  $Y$  una variable aleatoria y sea  $\omega$  un punto en el espacio muestral. Suponga que  $Y(\omega) = y$ . Entonces el número  $E(X | Y = y)$  es el valor de la esperanza condicional  $E(X | Y)$  evaluada en  $\omega$ , es decir,

$$E(X | Y)(\omega) = E(X | Y = y).$$

- Si  $Y$  es discreta con valores  $0, 1, \dots$  entonces  $E(X | Y)$  también es discreta y toma los siguiente valores

$$\begin{aligned} E(X | Y)(\omega) &= \sum_{y=0}^{\infty} E(X | Y = y) \cdot 1_{(Y=y)}(\omega) \\ &= \begin{cases} E(X | Y = 0) & \text{si } Y = 0, \\ E(X | Y = 1) & \text{si } Y = 1, \\ \vdots & \vdots \end{cases} \end{aligned}$$

Además, la esperanza condicional toma el valor  $E(X | Y = y)$  con probabilidad dada por la suma de las probabilidades de los eventos  $(Y = y)$  que produzcan tal valor.

El siguiente ejemplo es un caso particular de la última propiedad y ayuda a entender mejor el concepto de esperanza condicional.

**Ejemplo 2.52** Sea  $X$  una variable aleatoria con esperanza finita y sea  $Y$  otra variable aleatoria con distribución  $\text{Ber}(\theta)$ . Entonces

$$E(X | Y) = E(X | Y = 0) \cdot 1_{(Y=0)} + E(X | Y = 1) \cdot 1_{(Y=1)},$$

en donde las esperanzas condicionales que aparecen en el lado derecho son las usuales de la probabilidad elemental. Más explícitamente,

$$E(X | Y)(\omega) = \begin{cases} E(X | Y = 0) & \text{si } Y(\omega) = 0, \\ E(X | Y = 1) & \text{si } Y(\omega) = 1. \end{cases}$$

De esta manera, la variable aleatoria  $W := E(X | Y)$  toma los dos valores que aparecen en la expresión anterior y su distribución es

$$\begin{aligned} P(W = E(X | Y = 0)) &= 1 - \theta, \\ P(W = E(X | Y = 1)) &= \theta. \end{aligned}$$

■

El concepto de esperanza condicional será de utilidad para entender el teorema de Rao-Blackwell y el teorema de Lehmann-Scheffé que estudiaremos en las siguientes secciones.

## Ejercicios

244. A partir de la Definición 2.28 que aparece en la página 211, demuestre las siguientes propiedades de la esperanza condicional.

a)  $E(E(X | \mathcal{G})) = E(X)$ .

b) Si  $X$  es  $\mathcal{G}$ -medible entonces  $E(X | \mathcal{G}) = X$ .

245. Sea  $X$  una variable aleatoria con esperanza finita y sea  $c$  una constante. Encuentre las siguientes esperanzas condicionales.

- a)  $E(c | X)$ .  
 b)  $E(X | c)$ .  
 c)  $E(cX | X)$ .  
 d)  $E(X | cX)$ . ( $c \neq 0$ )  
 e)  $E(X + c | X)$ .  
 f)  $E(X | X + c)$ .

246. Sea  $X$  una variable aleatoria con esperanza finita y sea  $Y$  una variable aleatoria discreta. Demuestre directamente que

$$E(E(X | Y)) = E(X).$$

247. Sea  $X$  una variable aleatoria con esperanza finita e independiente de la variable aleatoria discreta  $Y$ . Demuestre directamente que

$$E(X | Y) = E(X).$$

248. Sea  $(X, Y)$  un vector aleatorio discreto con función de probabilidad como aparece abajo. Encuentre la distribución de las variables aleatorias  $E(X | Y)$  y  $E(Y | X)$ .

$x \setminus y$	0	1
a)	0	1/8    1/4
	1	1/4    3/8

$x \setminus y$	-1	1
b)	1	1/8    1/4
	2	1/8    1/8
	3	1/4    1/8

$x \setminus y$	-1	0	1
c)	1	1/9    1/9    1/9	
	2	1/9    1/9    1/9	
	3	1/9    1/9    1/9	

249. Sean  $X_1, \dots, X_n$  variables aleatorias con idéntica distribución y con esperanza finita. Demuestre que

$$E(X_1 | X_1 + \dots + X_n) = \frac{1}{n} (X_1 + \dots + X_n).$$

## 2.16. Teorema de Rao-Blackwell

El siguiente resultado establece un procedimiento importante para mejorar un estimador insesgado a través de una estadística suficiente. La mejoría consiste en proponer un nuevo estimador insesgado con varianza menor o igual a la varianza del estimador insesgado original. Para ello se necesita el cálculo de una esperanza condicional.

**Teorema 2.7 (Rao-Blackwell<sup>5</sup>)** Sea  $T$  un estimador insesgado para una función parametral unidimensional  $\tau(\theta)$  y sea  $U$  una estadística suficiente para  $\theta$ . Entonces la variable aleatoria  $E(T | U)$  es una estadística que es función de  $U$  y cumple lo siguiente:

1.  $E(T | U)$  es insesgado para  $\tau(\theta)$ .
2.  $\text{Var}(E(T | U)) \leq \text{Var}(T)$ ,  
con igualdad  $\Leftrightarrow T = E(T | U)$  c.s.

**Demostración.** Veamos primero que la esperanza condicional  $E(T | U)$  es una estadística. Para cada valor  $u$  de  $U$ , tenemos que

$$\begin{aligned} E(T | U = u) &= E(T(X_1, \dots, X_n) | U = u) \\ &= \int_{\mathbb{R}^n} T(x^n) f(x^n | U = u) dx^n. \end{aligned}$$

El primer factor del integrando no depende de  $\theta$  pues  $T$  es una estadística. El segundo factor tampoco depende de  $\theta$  pues  $U$  es suficiente. Concluimos que la variable aleatoria  $E(T | U)$  no depende de  $\theta$  y por lo tanto es una estadística. Este es el único punto en la demostración en donde se hace uso de la hipótesis de que  $U$  es suficiente.

Veamos ahora que la estadística  $E(T | U)$  es función de la estadística  $U$ . Para enfatizar que  $U$  es una función de una muestra aleatoria, a esta esperanza condicional la escribiremos como  $E(T | U(X_1, \dots, X_n))$ . Sean  $x^n$  y

<sup>5</sup>Calyampudi Radhakrishna Rao (1920–), matemático y estadístico hindú.

<sup>5</sup>David Harold Blackwell (1919–2010), estadístico estadounidense.

$y^n$  dos valores de la muestra aleatoria tales que  $U(x^n) = U(y^n)$ . Entonces claramente  $E(T | U = U(x^n)) = E(T | U = U(y^n))$ .

Ahora podemos demostrar las últimas dos afirmaciones de este teorema.

1. La propiedad de insesgamiento de la esperanza condicional es una consecuencia inmediata de la misma propiedad para  $T$ , pues

$$E(E(T | U)) = E(T) = \tau(\theta).$$

2. Finalmente veamos que la varianza de la esperanza condicional es menor o igual a la varianza del estimador insesgado original. Tenemos que

$$\begin{aligned} \text{Var}(T) &= E(T - \tau(\theta))^2 \\ &= E[(T - E(T | U)) + (E(T | U) - \tau(\theta))]^2 \\ &= E[T - E(T | U)]^2 + E[E(T | U) - \tau(\theta)]^2 \\ &\quad + 2 \cdot E[T - E(T | U)] \cdot E[E(T | U) - \tau(\theta)] \\ &= E[T - E(T | U)]^2 + E[E(T | U) - \tau(\theta)]^2 \\ &\quad + 2 \cdot [E(T) - E(T)] \cdot [E(T) - \tau(\theta)] \\ &= E[T - E(T | U)]^2 + E[E(T | U) - \tau(\theta)]^2 \\ &\geq E[E(T | U) - \tau(\theta)]^2 \\ &= \text{Var}(E(T | U)). \end{aligned}$$

Además, esta desigualdad es una igualdad si y sólo si  $E[T - E(T | U)]^2 = 0$ . Pero la esperanza de esta variable aleatoria no negativa es cero si y sólo si la variable misma es cero casi seguramente, esto es,  $T = E(T | U)$  c.s.

■

De esta manera, un estimador insesgado  $T$  puede mejorarse en el sentido de producir a través de él otro estimador insesgado de varianza menor o igual a la varianza de  $T$ . Este mejoramiento se logra calculando su esperanza condicional respecto de alguna estadística suficiente.

En lo que resta de esta sección daremos varios ejemplos de aplicación de este procedimiento, en donde el cálculo de la esperanza condicional puede efectuarse sin demasiada dificultad. Este no siempre es el caso.

**Ejemplo 2.53** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{Ber}(\theta)$ , con  $\theta$  desconocido. Es inmediato comprobar que la estadística  $T := X_1$  es un estimador insesgado para  $\theta$ . Por otro lado, sabemos que  $U := X_1 + \dots + X_n$  es suficiente para  $\theta$ . El procedimiento de Rao-Blackwell sugiere calcular  $E(T | U)$  encontrando así otro estimador insesgado para  $\theta$  con posiblemente una varianza más pequeña.

Sea  $u \in \{0, 1, \dots, n\}$  un posible valor de  $U$ . Entonces, como  $T$  tiene distribución Bernoulli,

$$\begin{aligned}
 E(T | U = u) &= 1 \cdot P(T = 1 | U = u) + 0 \cdot P(T = 0 | U = u) \\
 &= P(X_1 = 1 | X_1 + \dots + X_n = u) \\
 &= \frac{P(X_1 = 1) \cdot P(X_2 + \dots + X_n = u - 1)}{P(X_1 + \dots + X_n = u)} \\
 &= \frac{\theta \cdot \binom{n-1}{u-1} \theta^{u-1} (1-\theta)^{(n-1)-(u-1)}}{\binom{n}{u} \theta^u (1-\theta)^{n-u}} \\
 &= \frac{\binom{n-1}{u-1}}{\binom{n}{u}} \\
 &= \frac{1}{n} u.
 \end{aligned}$$

Como lo anterior se cumple para cada valor  $u \in \{0, 1, \dots, n\}$ , se concluye que

$$E(T | U) = \frac{1}{n} U.$$

Es decir,  $E(T|U) = \bar{X}$ . Se verifica entonces lo siguiente

$$\begin{aligned} \text{Var}(E(T|U)) &= \frac{1}{n}\theta(1-\theta) \\ &\leq \theta(1-\theta) \\ &= \text{Var}(T). \end{aligned}$$

Las gráficas de estas varianzas, como funciones de  $\theta$ , se muestran en la Figura 2.12.

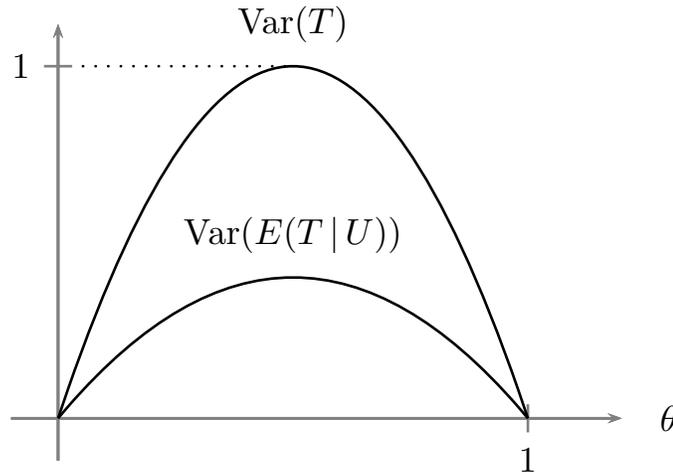


Figura 2.12

■

El siguiente ejemplo es un caso en donde la distribución en estudio es continua.

**Ejemplo 2.54** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $N(\theta, \sigma^2)$ , en donde  $\theta$  es desconocido y  $\sigma^2$  es conocida. El estimador  $T := X_1$  es insesgado para  $\theta$ . Por otro lado, la estadística  $U := \bar{X}$  es suficiente para  $\theta$ . Deseamos encontrar el estimador insesgado mejorado  $E(T|U)$ . Para

cualquier valor  $t$  de  $T$  y cualquier valor  $u$  de  $U$ ,

$$\begin{aligned}
 f_{T|U}(t|u) &= \frac{f_{T,U}(t,u)}{f_U(u)} \\
 &= \frac{f_{X_1,(X_1+\dots+X_n)/n}(t,u)}{f_U(u)} \\
 &= \frac{f_{X_1,X_1+\dots+X_n}(t,nu)}{f_U(u)} \\
 &= \frac{f_{X_1,X_2+\dots+X_n}(t,nu-t)}{f_U(u)} \\
 &= \frac{f_{X_1}(t) f_{X_2+\dots+X_n}(nu-t)}{f_U(u)}.
 \end{aligned}$$

Las tres funciones de densidad que aparecen en la última expresión son normales con ciertos parámetros. Substituyendo estas funciones y simplificando se encuentra que esta función de densidad es  $N(u, (1 - 1/n)\sigma^2)$ . Por lo tanto, la esperanza de esta función de densidad condicional es  $E(T | U = u) = u = \bar{x}$ . Como esta identidad se cumple para cualquier valor de  $u$ , se concluye que

$$E(T | U) = \bar{X}.$$

La varianza de esta variable aleatoria es  $\text{Var}(E(T | U)) = \sigma^2/n$ , constante respecto de  $\theta$ . Se verifica entonces la desigualdad

$$\begin{aligned}
 \text{Var}(E(T | U)) &= \sigma^2/n \\
 &\leq \sigma^2 \\
 &= \text{Var}(T).
 \end{aligned}$$

Las gráficas de estas varianzas, funciones constantes respecto de  $\theta$ , se muestran en la Figura 2.13.

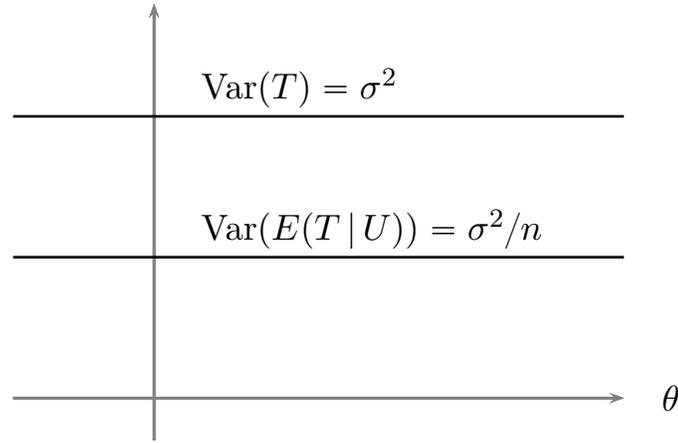


Figura 2.13

■

Ahora veremos una situación general que incluye los dos ejemplos anteriores.

**Ejemplo 2.55** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución dependiente de un parámetro desconocido  $\theta$  y tal que su media es  $\theta$  mismo. Es claro que la estadística  $T := X_1$  es un estimador insesgado para  $\theta$ , y supongamos, por otro lado, que la estadística  $U := X_1 + \dots + X_n$  es suficiente para  $\theta$ . Las distribuciones  $\text{Ber}(\theta)$ ,  $\text{Poisson}(\theta)$  y  $\text{N}(\theta, \sigma^2)$  son ejemplos en donde se cumplen estas dos hipótesis.

Encontraremos  $E(T|U)$ . Para cualquier posible valor  $u$  de  $U$ , por la hipótesis de idéntica distribución tenemos que

$$\begin{aligned} E(T|U = u) &= E(X_1 | X_1 + \dots + X_n = u) \\ &= \frac{1}{n} E(X_1 + \dots + X_n | X_1 + \dots + X_n = u) \\ &= \frac{1}{n} u. \end{aligned}$$

Esto demuestra que  $E(T|U) = U/n = \bar{X}$ . Este es el estimador insesgado mejorado por el procedimiento de Rao-Blackwell y su varianza es

$\text{Var}(E(T | U)) = (1/n) \text{Var}(T)$ . Por lo tanto, se verifica la desigualdad

$$\begin{aligned} \text{Var}(E(T | U)) &= \frac{1}{n} \text{Var}(T) \\ &\leq \text{Var}(T). \end{aligned}$$

■

Ahora veremos un ejemplo en donde el parámetro a estimar es una función parametral. Los cálculos se vuelven un poco más elaborados.

**Ejemplo 2.56** Supongamos nuevamente que  $X_1, \dots, X_n$  es una muestra aleatoria de la distribución  $\text{Ber}(\theta)$ , con  $\theta$  desconocido. Sea  $\tau(\theta) := \theta(1 - \theta)$ . La estadística  $T := X_1(1 - X_2)$  es un estimador insesgado para la función parametral  $\tau(\theta)$  pues, por la hipótesis de independencia,

$$\begin{aligned} E(T) &= E(X_1(1 - X_2)) \\ &= E(X_1) E(1 - X_2) \\ &= \theta(1 - \theta). \end{aligned}$$

Sea  $U := X_1 + \dots + X_n$ . Sabemos que  $U$  es suficiente para  $\theta$  y por lo tanto también lo es para  $\tau(\theta)$ . Encontraremos el estimador insesgado mejorado por el procedimiento de Rao-Blackwell para  $\tau(\theta)$  usando el estimador insesgado inicial  $T$  y la estadística suficiente  $U$ . Sea  $u \in \{0, 1, \dots, n\}$  un posible valor de  $U$ . Entonces

$$\begin{aligned}
E(T | U = u) &= E(X_1(1 - X_2) | U = u) \\
&= E(X_1 | U = u) - E(X_1 X_2 | U = u) \\
&= \frac{u}{n} - 1 \cdot P(X_1 = 1, X_2 = 1 | U = u) \\
&= \frac{u}{n} - \frac{P(X_1 = 1, X_2 = 1, X_3 + \cdots + X_n = u - 2)}{P(X_1 + \cdots + X_n = u)} \\
&= \frac{u}{n} - \frac{\theta^2 \binom{n-2}{u-2} \theta^{u-2} (1-\theta)^{n-u}}{\binom{n}{u} \theta^u (1-\theta)^{n-u}} \\
&= \frac{u}{n} - \frac{\binom{n-2}{u-2}}{\binom{n}{u}} \\
&= \frac{u}{n} - \frac{u(u-1)}{n(n-1)} \\
&= \frac{n}{n-1} \frac{u}{n} \left(1 - \frac{u}{n}\right).
\end{aligned}$$

Como esta identidad se cumple para cualquier valor  $u \in \{0, 1, \dots, n\}$ , se concluye que

$$E(T | U) = \frac{n}{n-1} \frac{U}{n} \left(1 - \frac{U}{n}\right).$$

Este es el estimador insesgado mejorado para  $\tau(\theta) = \theta(1 - \theta)$ . Puede comprobarse que

$$\text{Var}(T) = \theta(1 - \theta)(1 - \theta + \theta^2).$$

Usando la fórmula recursiva para los momentos de la distribución binomial, se puede demostrar también que

$$\text{Var}(E(T | U)) = \frac{2\theta(1 - \theta)(1 - 3\theta + 3\theta^2)}{n(n-1)}.$$

Aunque no es inmediato, se puede comprobar que  $\text{Var}(E(T | U)) \leq \text{Var}(T)$ . Por ejemplo, para  $n = 3$ , las gráficas de estas funciones de  $\theta$  se muestran en la Figura 2.14.

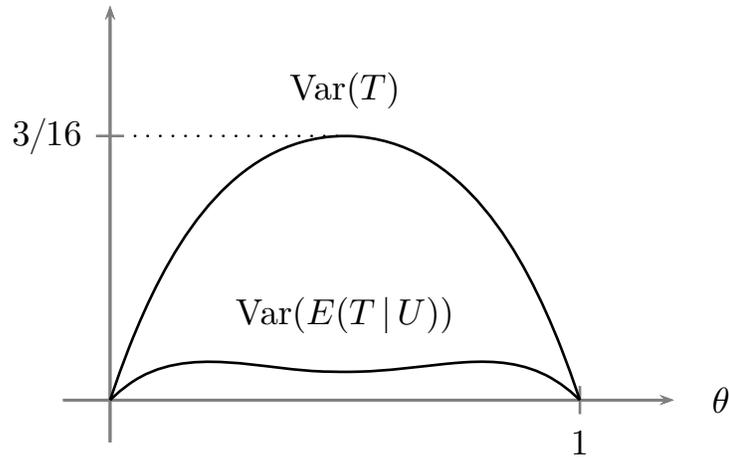


Figura 2.14

■

## Ejercicios

250. **Distribución Bernoulli:**  $\tau(\theta) = \theta^2$ . Sea  $X_1, \dots, X_n$  una muestra aleatoria de tamaño  $n \geq 2$  de la distribución Bernoulli de parámetro  $\theta$ . Suponga conocido que la estadística  $U = X_1 + \dots + X_n$  es suficiente para  $\theta$ . Defina la función parametral  $\tau(\theta) = \theta^2$  y la estadística  $T = X_1 \cdot X_2$ .

- Demuestre que  $T$  es insesgado para  $\tau(\theta)$ .
- Encuentre  $\text{Var}(T)$ .
- Encuentre  $E(T | U)$ .
- Encuentre  $\text{Var}(E(T | U))$ .
- Demuestre que  $\text{Var}(E(T | U)) \leq \text{Var}(T)$ .

## 2.17. Completez

En esta sección estudiaremos una propiedad adicional que pueden cumplir algunas estadísticas y que será fundamental en el resultado que veremos

más adelante acerca de estimadores insesgados de varianza mínima.

Consideremos nuevamente que  $X_1, \dots, X_n$  una muestra aleatoria de una distribución con función de densidad o de probabilidad  $f(x, \theta)$ , dependiente de un parámetro  $\theta$ . Supongamos que  $\theta$  toma valores en un cierto espacio parametral  $\Theta$ . Sea  $T$  una estadística y sea  $f_T(t, \theta)$  su función de densidad o de probabilidad, que también depende, en general, de  $\theta$ . A continuación definiremos la noción de completez para la familia de funciones de densidad o de probabilidad  $\{f_T(t, \theta) : \theta \in \Theta\}$ .

**Definición 2.29** Se dice que una estadística  $T$ , o su familia de funciones de densidad o de probabilidad  $\{f_T(t, \theta) : \theta \in \Theta\}$ , es **completa** si para cualquier función  $h$  se cumple la implicación

$$E(h(T)) = 0 \quad \Rightarrow \quad h(T) = 0 \quad \text{c.s.} \quad (2.17)$$

Observe que, por simplicidad, no hemos especificado el dominio de la función  $h$ , pero éste debe contener al conjunto de valores de la estadística  $T$ , de tal forma que la composición  $h(T)$  tiene sentido. Supondremos además que esta composición es también una variable aleatoria y que tiene esperanza finita. Otra observación importante es que, en general, la esperanza  $E(h(T))$  depende del parámetro desconocido  $\theta$ , así es que la condición  $E(h(T)) = 0$  debe cumplirse para todo valor posible del parámetro  $\theta$ .

En la siguiente sección veremos la utilidad de la propiedad de completez de una estadística cuando se conjunte con la propiedad de suficiencia. Estas propiedades para una estadística aparecen como hipótesis en el teorema de Lehmann-Scheffé. Regresando a la definición de completez, en general no es fácil comprobar su cumplimiento. El siguiente ejemplo, sin embargo, es particularmente sencillo.

**Ejemplo 2.57** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{Ber}(\theta)$ . Demostraremos que la estadística  $T = X_1 + \dots + X_n$  es completa. Sea  $h$  una función cualquiera tal que  $E(h(T)) = 0$ . Como  $T$  tiene distribución

$\text{bin}(n, \theta)$ , tenemos que

$$\begin{aligned} E(h(T)) &= \sum_{t=0}^n h(t) \binom{n}{t} \theta^t (1-\theta)^{n-t} \\ &= (1-\theta)^n \sum_{t=0}^n h(t) \binom{n}{t} \left(\frac{\theta}{1-\theta}\right)^t. \end{aligned}$$

La última suma corresponde a un polinomio en la variable  $\alpha = \theta/(1-\theta)$ . Para que este polinomio en  $\alpha$  sea cero para cualquier posible valor de  $\alpha$ , sus coeficientes deben ser todos forzosamente cero, esto es, para cada  $t = 0, 1, \dots, n$ , se tiene que  $h(t) \binom{n}{t} = 0$ . Esto implica que  $h(t) = 0$  para cada  $t = 0, 1, \dots, n$ , es decir,  $h(T) = 0$ . De esta manera hemos comprobado que la estadística  $T$ , o la familia de distribuciones binomial  $\{f_T(t, \theta) : 0 < \theta < 1\}$ , es completa. ■

Veamos otro ejemplo, esta vez cuando la distribución de probabilidad involucrada es continua.

**Ejemplo 2.58** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{unif}(0, \theta)$ , con  $\theta > 0$ . Demostraremos que la estadística  $T = \max\{X_1, \dots, X_n\}$  es completa. Observemos primero que  $T$  tiene como posibles valores el intervalo  $(0, \theta)$  y recordemos que su función de distribución es la siguiente: para  $0 < t < \theta$ ,

$$\begin{aligned} F_T(t) &= P(\max\{X_1, \dots, X_n\} \leq t) \\ &= (P(X_1 \leq t))^n \\ &= \left(\frac{t}{\theta}\right)^n. \end{aligned}$$

Por lo tanto,

$$f_T(t) = \begin{cases} \frac{n}{\theta} \left(\frac{t}{\theta}\right)^{n-1} & \text{si } 0 < t < \theta, \\ 0 & \text{en otro caso.} \end{cases}$$

Sea entonces  $h$  una función cualquiera tal que  $E(h(T)) = 0$ . Para cualquier valor  $\theta > 0$ ,

$$0 = \int_0^\theta h(t) \frac{n}{\theta} \left(\frac{t}{\theta}\right)^{n-1} dt = \frac{n}{\theta^n} \int_0^\theta h(t) t^{n-1} dt.$$

Esto implica que la integral se anula para cualquier  $\theta > 0$ . Derivando esta integral respecto de  $\theta$  y suponiendo continuidad para la función  $h$ , se obtiene que  $h(\theta)\theta^{n-1} = 0$  para cualquier  $\theta > 0$ . Esto se cumple cuando  $h(\theta) = 0$  para cualquier  $\theta > 0$ , es decir,  $h(T) = 0$ . Esto demuestra que la estadística  $T$ , o la familia de funciones de densidad  $\{f_T(t, \theta) : \theta > 0\}$ , es completa. ■

Observemos que para demostrar la no completez de una estadística  $T$  es suficiente dar una función  $h$  que no sea idénticamente cero en el conjunto de valores de  $T$  y tal que  $E[h(T)] = 0$ . Veremos a continuación un ejemplo de esta situación.

**Ejemplo 2.59** Supongamos que una estadística  $T$  tiene función de densidad  $f(t, \theta)$  dada por la distribución  $N(0, \theta)$ , es decir,

$$f(t, \theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-t^2/2\theta}, \quad -\infty < t < \infty,$$

en donde el parámetro  $\theta > 0$  es la varianza de la distribución y la media es cero. Entonces es fácil comprobar que  $T$ , o la familia de densidades  $\{f(t, \theta) : \theta > 0\}$ , no es completa pues para la función  $h(t) = t$  se cumple la condición

$$\int_{-\infty}^{\infty} h(t) f(t, \theta) dt = 0.$$

y sin embargo,  $h(t)$  no es idénticamente cero. ■

Es interesante observar que la propiedad de completez de una estadística depende fuertemente del espacio parametral  $\Theta$  que se considere como conjunto de posibles valores para  $\theta$ . En efecto, la implicación (2.17) que aparece en la Definición 2.29 debe cumplirse para todo valor de  $\theta$  en  $\Theta$ . Si este conjunto se reduce la completez puede perderse.

La definición de completez para una estadística  $T$  es también válida para un vector de estadísticas  $T = (T_1, \dots, T_k)$ . En este caso la función real  $h$  que se utiliza debe tener como dominio un subconjunto de  $\mathbb{R}^k$ .

Como en el caso unidimensional, se pueden dar ejemplos de vectores de estadísticas que no satisfacen la propiedad de completez.

Demostraremos a continuación que la propiedad de completez permanece invariante bajo transformaciones biyectivas.

**Teorema 2.8** Toda función biyectiva de una estadística completa también es completa.

**Demostración.** Sea  $T$  una estadística completa y sea  $\varphi$  una función biyectiva tal que  $\varphi(T)$  es una variable aleatoria con esperanza finita. Sea  $h$  una función tal que  $E(h(\varphi(T))) = 0$ , es decir,  $E((h \circ \varphi)(T)) = 0$ . La completez de  $T$  implica que  $(h \circ \varphi)(T) = 0$  c.s., es decir,  $h(\varphi(T)) = 0$  c.s. Esto demuestra la completez de  $\varphi(T)$ . El mismo argumento se aplica cuando  $T$  es un vector de estadísticas. ■

Para concluir esta sección mencionaremos que en la sección 2.19 se enuncia un ejemplo general de la propiedad de completez para una cierta estadística para distribuciones dentro de la familia exponencial.

## Ejercicios

251. Sea  $a$  una constante. Demuestre que una estadística  $T$  es completa si y sólo si
- $T + a$  es completa.
  - $aT$  es completa,  $a \neq 0$ .
252. Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución especificada abajo, dependiente del parámetro  $\theta$ . Suponga que cualquier otro parámetro que pudiera aparecer en la distribución tiene un valor fijo conocido. Demuestre directamente que la estadística  $T = X_1 + \dots + X_n$  es completa.
- $\text{bin}(k, \theta)$ .
  - $\text{Poisson}(\theta)$ .
  - $\text{geo}(\theta)$ .
  - $\text{N}(\theta, \sigma^2)$ .
  - $\text{gama}(\gamma, \theta)$ .

253. **Distribución uniforme.** Sea  $T = |X_1|$ , en donde  $X_1$  es una muestra aleatoria de tamaño  $n = 1$  de la distribución  $\text{unif}(-\theta, \theta)$ , con  $\theta > 0$ . Determine si  $T$  es una estadística completa.
254. **Distribución Poisson.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{Poisson}(\theta)$  con  $\theta > 0$ . Demuestre que la estadística
- $T = X_1 + \dots + X_k$  es completa,  $1 \leq k \leq n$ .
  - $T = (X_1, \dots, X_k)$  no es completa,  $2 \leq k \leq n$ .
255. **No completez.** Sea  $f(x, \theta)$  la función de densidad de la distribución  $\text{unif}(-\theta, \theta)$ , con  $\theta > 0$ . Demuestre que la familia de densidades  $\{f(x, \theta) : 0 < \theta < \infty\}$  no es completa.
256. **No completez.** Sea  $f(x, \theta)$  la función de densidad de la distribución  $N(0, \theta)$ , con  $\theta > 0$ . Demuestre que la familia de densidades  $\{f(x, \theta) : \theta > 0\}$  no es completa.

## 2.18. Teorema de Lehmann-Scheffé

En esta sección se presenta un resultado importante que permite construir estimadores insesgados de varianza mínima cuando se cumplen ciertas condiciones.

Sabemos que bajo las condiciones de regularidad, cualquier estimador insesgado para una función parametral tiene varianza por lo menos la cota inferior de Cramér-Rao. Recordemos que, tanto la varianza del estimador como la cota inferior de Cramér-Rao son funciones del parámetro desconocido  $\theta$ . A aquel estimador insesgado cuya varianza sea la más pequeña para cada valor de  $\theta$  le hemos llamado estimador insesgado de varianza mínima uniforme (en inglés *uniformly minimum variance unbiased estimator*), y por brevedad se le llama UMVUE. El calificativo de uniforme se refiere a que la minimalidad de su varianza se cumple para todo valor de  $\theta$  en el espacio parametral  $\Theta$ . El siguiente resultado establece la forma de encontrar este estimador.

**Teorema 2.9 (Lehmann-Scheffé<sup>6</sup>)** Sea  $T$  un estimador insesgado para una función parametral unidimensional  $\tau(\theta)$ . Sea  $U$  una estadística suficiente y completa para  $\theta$ . Entonces  $E(T|U)$  es

1. El único estimador que satisface ser función de  $U$  y ser insesgado para  $\tau(\theta)$ .
2. El UMVUE para  $\tau(\theta)$ , es decir, tiene varianza mínima de entre todos los estimadores insesgados para  $\tau(\theta)$ .

***Demostración.***

1. Veamos primero la unicidad. Hemos demostrado antes que la esperanza condicional  $E(T|U)$  es una estadística que es función de  $U$  y es insesgado para  $\tau(\theta)$ . Supongamos que  $W$  es otro estimador para  $\tau(\theta)$  con estas dos características. Defina la función  $h(U) = W - E(T|U)$ . Entonces

$$E(h(U)) = E(W) - E(E(T|U)) = \tau(\theta) - \tau(\theta) = 0.$$

Como  $U$  es completa,  $h(U) = 0$  c.s. Es decir,  $W = E(T|U)$  c.s. De esta manera, la hipótesis de completez para  $U$  lleva a concluir que  $E(T|U)$  es el único estimador insesgado que es función de  $U$ .

2. Sea  $W$  cualquier estimador insesgado para  $\tau(\theta)$  sin solicitar necesariamente que sea función de  $U$ . Consideremos el estimador  $E(W|U)$ , el cual es insesgado y es función de  $U$ . Por la propiedad de unicidad, tenemos que este estimador es idéntico a  $E(T|U)$ . Por el teorema de Rao-Blackwell,

$$\text{Var}(W) \geq \text{Var}(E(W|U)) = \text{Var}(E(T|U)).$$

■

<sup>6</sup>Erich Leo Lehmann (1917-2009), estadístico estadounidense.

<sup>6</sup>Henry Scheffé (1907-1977), matemático y estadístico estadounidense.

Del resultado general anterior se desprenden los siguientes casos particulares que permiten encontrar el UMVUE para una función parametral  $\tau(\theta)$ .

**Corolario 2.2** Sea  $U$  una estadística suficiente y completa para  $\theta$ . Si la función  $g(U)$  es un estimador insesgado para  $\tau(\theta)$ , entonces  $g(U)$  es el UMVUE para  $\tau(\theta)$ .

*Demostración.* Como  $E(g(U) | U) = g(U)$  c.s., se sigue del teorema de Lehmann-Scheffé que  $g(U)$  es el UMVUE para  $\theta$ . ■

**Corolario 2.3** Si  $T$  es un estimador insesgado para  $\tau(\theta)$ , y suficiente y completo para  $\theta$ , entonces  $T$  es el UMVUE para  $\tau(\theta)$ .

*Demostración.* Como  $T$  es una estadística suficiente y completa también para  $\tau(\theta)$ , la identidad  $E(T | T) = T$  c.s. y el teorema de Lehmann-Scheffé aseguran que  $T$  es el UMVUE para  $\tau(\theta)$ . ■

Por ejemplo, hemos comprobado antes que la estadística  $\bar{X}$  es un estimador insesgado para el parámetro de la distribución  $\text{Ber}(\theta)$ . Puede comprobarse que  $\bar{X}$  es también suficiente y completa para  $\theta$ . Por lo tanto,  $\bar{X}$  es el UMVUE para  $\theta$ .

## Ejercicios

257. **Distribución  $\text{Ber}(\theta)$ : UMVUE para  $\tau(\theta) = \theta + (1 - \theta)e^2$ .**

Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{Ber}(\theta)$ , con  $\theta$  desconocido. Defina la función parametral  $\tau(\theta) = \theta + (1 - \theta)e^2$ .

- Encuentre un estimador insesgado  $T$  para  $\tau(\theta)$  y compruebe que lo es.
- Considere la estadística suficiente y completa  $U = X_1 + \dots + X_n$ . Para cada valor  $u$  de  $U$ , calcule  $E(T | U = u)$ .

- c) Use el teorema de Lehmann-Scheffé para encontrar el UMVUE para  $\tau(\theta)$ .

258. **Distribución  $\text{geo}(\theta)$ : UMVUE para  $\theta$ .**

Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{geo}(\theta)$ , con  $\theta$  desconocido. Nos interesa estimar el parámetro  $\theta$ . Sabemos que la estadística  $U = X_1 + \dots + X_n$  es suficiente y completa para  $\theta$ . Considere el estimador

$$T = \frac{1}{1 + \frac{n}{n-1} \bar{X}}.$$

Demuestre los siguientes resultados que llevan a encontrar el UMVUE para  $\theta$ . Se verifica que la varianza del UMVUE alcanza la cota inferior de Cramér-Rao.

- Demuestre que  $T$  es insesgado para  $\theta$ .
- Calcule  $\text{Var}(T)$ .
- Calcule  $E(T | U)$ . Este es el UMVUE para  $\theta$ .
- Calcule  $\text{Var}(E(T | U))$ .
- Calcule  $\text{CICR}(\theta)$ .
- Demuestre que  $\text{CICR}(\theta) = \text{Var}(E(T | U)) \leq \text{Var}(T)$ .

259. **Distribución  $\text{Poisson}(\theta)$ : UMVUE para  $\theta$ .**

Sea  $X_1, \dots, X_n$  una muestra aleatoria de tamaño  $n \geq 2$  de la distribución  $\text{Poisson}(\theta)$ , con  $\theta > 0$  desconocido. Sabemos que la estadística  $U = X_1 + \dots + X_n$  es suficiente y completa para  $\theta$ . Defina ahora el estimador

$$T = \frac{1}{2} (X_1 + X_2).$$

Demuestre los siguientes resultados que llevan a encontrar el UMVUE para  $\theta$ . Se verifica que la varianza del UMVUE alcanza la cota inferior de Cramér-Rao.

- $T$  es insesgado para  $\theta$ .
- $\text{Var}(T) = \theta/2$ .
- $E(T | U) = \bar{X}$ . Este es el UMVUE para  $\theta$ .
- $\text{Var}(E(T | U)) = \theta/n$ .

$$e) \text{CICR}(\theta) = \theta/n.$$

$$f) \text{CICR}(\theta) = \text{Var}(E(T|U)) \leq \text{Var}(T).$$

260. **Distribución Poisson( $\theta$ ): UMVUE para  $\tau(\theta) = e^{-\theta}$ .**

Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución Poisson( $\theta$ ), con  $\theta > 0$  desconocido.

a) Demuestre que la estadística  $T = 1_{\{0\}}(X_1)$  es un estimador insesgado para la función parametral  $\tau(\theta) = e^{-\theta}$ .

b) Demuestre que la estadística  $U = X_1 + \dots + X_n$  es suficiente para  $\theta$ .

c) Demuestre que  $U$  es una estadística suficiente minimal para  $\theta$ .

d) El procedimiento de Rao-Blackwell sugiere encontrar  $E(T|U)$ . Demuestre que

$$E(T|U) = \left(\frac{n-1}{n}\right)^U.$$

e) Demuestre que  $\text{Var}(T) = e^{-\theta}(1 - e^{-\theta})$ .

f) Recuerde que si  $X$  es una variable aleatoria con distribución Poisson( $\theta$ ), entonces su f.g.p. está dada por

$$G(t) = E(t^X) = e^{\theta(t-1)}.$$

Use la expresión anterior para demostrar que

$$\text{Var}(E(T|U)) = e^{-2\theta}(e^{\theta/n} - 1).$$

g) Demuestre que para la función parametral  $\tau(\theta) = e^{-\theta}$ ,

$$\text{CICR}(\theta) = \frac{\theta}{n} e^{-2\theta}.$$

h) Demuestre que

$$\text{CICR}(\theta) < \text{Var}(E(T|U)) \leq \text{Var}(T).$$

i) Con únicamente la información anterior, ¿qué puede decir de  $E(T|U)$ ?

- j) Demuestre que  $U$  es una estadística completa.  
 k) ¿Qué puede decir ahora de  $E(T | U)$ ?

**261. Distribución Poisson( $\theta$ ): UMVUE para  $\theta e^{-\theta}$ .**

Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución Poisson( $\theta$ ), con  $\theta > 0$  desconocido. Nos interesa estimar la función parametral  $\tau(\theta) = \theta e^{-\theta}$ . Sabemos que la estadística  $U = X_1 + \dots + X_n$  es suficiente y completa para  $\theta$ . Defina el estimador

$$T = 1_{\{1\}}(X_1).$$

Demuestre los siguientes resultados que llevan a encontrar el UMVUE para  $\tau(\theta)$ . Se verifica que la varianza del UMVUE alcanza la cota inferior de Cramér-Rao.

- a)  $T$  es insesgado para  $\tau(\theta)$ .  
 b)  $\text{Var}(T) = \theta e^{-\theta}(1 - \theta e^{-\theta})$ .  
 c)  $E(T | U) = \left(\frac{n-1}{n}\right)^{n\bar{X}-1} \bar{X}$ . Este es el UMVUE para  $\tau(\theta)$ .  
 d)  $\text{Var}(E(T | U)) = e^{-2\theta + \theta/n} \frac{\theta}{n} (1 + (n-1)^2 \frac{\theta}{n}) - e^{-2\theta} \theta^2$ .  
 e)  $\text{CICR}(\theta) = e^{-2\theta} \frac{\theta(1-\theta)^2}{n}$ , para  $\tau(\theta) = \theta e^{-\theta}$ .  
 f)  $\text{CICR}(\theta) \leq \text{Var}(E(T | U)) \leq \text{Var}(T)$ .

**262. Distribución exp( $\theta$ ): UMVUE para  $\theta$ .**

Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución exp( $\theta$ ), con  $\theta > 0$  desconocido. Sabemos que la estadística  $T = X_1 + \dots + X_n$  es suficiente y completa para  $\theta$ .

- a) Demuestre que la estadística  $(n-1)/T$  es un estimador insesgado para  $\theta$ .  
 b) Concluya que  $(n-1)/T$  es el UMVUE para  $\theta$ .  
 c) Calcule la varianza del UMVUE encontrado en el inciso anterior y compare con la cota inferior de Cramér-Rao.

**263. Distribución N( $\theta, \sigma^2$ ): UMVUE para  $\theta$ .**

Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución N( $\theta, \sigma^2$ ), con

$\theta$  desconocido y  $\sigma^2$  conocido. Nos interesa estimar el parámetro  $\theta$ . Sabemos que la estadística  $U = X_1 + \cdots + X_n$  es suficiente y completa para  $\theta$ . Defina el estimador

$$T = X_1.$$

Demuestre las siguientes afirmaciones:

- a)  $T$  es insesgado para  $\theta$ .
- b)  $\text{Var}(T) = \sigma^2$ .
- c)  $E(T | U) = \bar{X}$ .
- d)  $\text{Var}(E(T | U)) = \sigma^2/n$ .
- e)  $\text{CICR}(\theta) = \sigma^2/n$ . (constante)
- f)  $\text{CICR}(\theta) = \text{Var}(E(T | U)) \leq \text{Var}(T)$ .

Por el Teorema de Lehmann-Scheffé, se concluye que  $E(T | U)$  es el UMVUE para  $\theta$ . Además su varianza alcanza la cota inferior de Cramér-Rao.

**264. Distribución  $N(\theta, \sigma^2)$ : UMVUE para  $\theta^2$ .**

Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $N(\theta, \sigma^2)$ , con  $\theta$  desconocido y  $\sigma^2$  conocido. Nos interesa estimar la función parametral  $\tau(\theta) = \theta^2$ . Sabemos que la estadística  $U = X_1 + \cdots + X_n$  es suficiente y completa para  $\theta$ . Defina el estimador

$$T = X_1^2 - \sigma^2.$$

- a) Demuestre que  $T$  es insesgado para  $\theta^2$ .
- b) Calcule  $\text{Var}(T)$ .
- c) Calcule  $E(T | U)$ .
- d) Calcule  $\text{Var}(E(T | U))$ .
- e) Calcule  $\text{CICR}(\theta)$  para  $\tau(\theta) = \theta^2$ .
- f) Compruebe que  $\text{CICR}(\theta) \leq \text{Var}(E(T | U)) \leq \text{Var}(T)$ .

Se concluye que  $E(T | U)$  es el UMVUE para  $\theta^2$ , pues es insesgado y su varianza alcanza la cota inferior de Cramér-Rao.

265. **Distribución  $N(\theta, \sigma^2)$ : UMVUE para  $P(X_1 > a)$ .**

Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $N(\theta, \sigma^2)$ , con  $\theta$  desconocido y  $\sigma^2$  conocido. Sea  $a$  una constante cualquiera. Nos interesa estimar la función parametral  $\tau(\theta) = P(X_1 > a)$ . Sabemos que la estadística  $U = X_1 + \dots + X_n$  es suficiente y completa para  $\theta$ . Defina el estimador

$$T = 1_{(a, \infty)}(X_1).$$

Demuestre los siguientes resultados que llevan a encontrar el UMVUE para  $P(X_1 > a)$ . Se verifica que la varianza del UMVUE alcanza la cota inferior de Cramér-Rao.

- a) Demuestre que  $T$  es insesgado para  $\tau(\theta) = P(X_1 > a)$ .
- b) Calcule  $\text{Var}(T)$ .
- c) Calcule  $E(T|U)$ . Este es el UMVUE para  $\tau(\theta) = P(X_1 > a)$ .
- d) Calcule  $\text{Var}(E(T|U))$ .
- e) Calcule  $\text{CICR}(\theta)$ .
- f) Compruebe que  $\text{CICR}(\theta) \leq \text{Var}(E(T|U)) \leq \text{Var}(T)$ .

266. **Distribución  $N(\theta, \sigma^2)$ : UMVUE para  $P(|X_1| \leq a)$ .**

Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $N(\theta, \sigma^2)$ , con  $\theta$  desconocido y  $\sigma^2$  conocido. Sea  $a > 0$  una constante. Nos interesa estimar la función parametral  $\tau(\theta) = P(|X_1| \leq a)$ . Sabemos que la estadística  $U = X_1 + \dots + X_n$  es suficiente y completa para  $\theta$ . Defina el estimador

$$T = 1_{(0, a)}(|X_1|).$$

Demuestre los siguientes resultados que llevan a encontrar el UMVUE para  $P(|X_1| \leq a)$ . Se verifica que la varianza del UMVUE alcanza la cota inferior de Cramér-Rao.

- a) Demuestre que  $T$  es insesgado para  $\tau(\theta) = P(|X_1| \leq a)$ .
- b) Calcule  $\text{Var}(T)$ .
- c) Calcule  $E(T|U)$ . Este es el UMVUE para  $\tau(\theta) = P(|X_1| \leq a)$ .
- d) Calcule  $\text{Var}(E(T|U))$ .
- e) Calcule  $\text{CICR}(\theta)$ .

f) Compruebe que  $\text{CICR}(\theta) \leq \text{Var}(E(T|U)) \leq \text{Var}(T)$ .

267. Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $f(x, \theta)$  que se especifica abajo, en donde  $\theta > 0$  es desconocido.

$$f(x, \theta) = \begin{cases} e^{-(x-\theta)} & \text{si } x > \theta, \\ 0 & \text{en otro caso.} \end{cases}$$

a) Demuestre que la estadística  $X_{(1)} - 1/n$  es suficiente, completa e insesgada para  $\theta$ .

b) Encuentre el UMVUE para  $\theta$ .

268. Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $f(x, \theta)$  que se especifica abajo, en donde  $\theta > 0$  es un parámetro desconocido. Por el ejercicio 147, sabemos que la estadística  $T = -(n-1)/\sum_{i=1}^n \ln X_i$  es un estimador insesgado para  $\theta$ .

$$f(x, \theta) = \begin{cases} \theta x^{\theta-1} & \text{si } 0 < x < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

a) Demuestre que la media geométrica  $U = (X_1 \cdots X_n)^{1/n}$  es una estadística suficiente y completa para  $\theta$ .

b) Encuentre el UMVUE para  $\theta$ .

269. Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $f(x, \theta)$  que se especifica abajo, en donde  $\theta > 0$  es un parámetro desconocido.

$$f(x, \theta) = \begin{cases} \theta^2 x e^{-\theta x} & \text{si } x > 0, \\ 0 & \text{en otro caso.} \end{cases}$$

a) Demuestre que la estadística  $T = X_1 + \cdots + X_n$  es suficiente y completa para  $\theta$ .

b) Calcule  $E(1/T)$ .

c) Encuentre una función de  $T$  que sea insesgada para  $\theta$ . Use el teorema de Lehmann-Scheffé para concluir que esta función es el UMVUE para  $\theta$ .

## 2.19. Distribuciones tipo exponencial

En esta sección se define una colección amplia de distribuciones de probabilidad llamada familia exponencial. Esta familia agrupa a varias de las distribuciones de probabilidad discretas y continuas más conocidas, todas ellas compartiendo una misma forma para la función de densidad o de probabilidad. Se considera primero el caso cuando sólo hay un parámetro involucrado, y después cuando la distribución depende de varios parámetros. Aquí tenemos la definición en el primer caso.

**Definición 2.30** Una distribución dependiente de un parámetro  $\theta$  es de **tipo exponencial** si su función de probabilidad o de densidad, es de la forma

$$f(x, \theta) = a(\theta) b(x) e^{c(\theta) d(x)}, \quad -\infty < x < \infty, \quad (2.18)$$

en donde  $a(\theta) \geq 0$ ,  $b(x) \geq 0$ ,  $c(\theta)$  y  $d(x)$  son funciones reales que dependen únicamente de los argumentos indicados.

Como hemos señalado, la familia de distribuciones tipo exponencial incluye distribuciones tipo discreto y continuo. La expresión de la fórmula (2.18) justifica el término exponencial en el nombre de esta familia de distribuciones. En la siguiente tabla se muestran algunos ejemplos de distribuciones particulares que pertenecen a la familia exponencial.

Por brevedad en las expresiones que aparecen en la tabla, no hemos escrito la forma completa de la función  $b(x)$  en cada caso. Por ejemplo, para la distribución Bernoulli se debe escribir  $b(x) = 1 \cdot 1_{\{0,1\}}(x)$ , mientras que para la distribución binomial  $b(x) = \binom{n}{x} \cdot 1_{\{0,1,\dots,k\}}(x)$ , indicando así el soporte de la distribución.

**Algunas distribuciones tipo exponencial de un parámetro**

Distribución	$a(\theta)$	$b(x)$	$c(\theta)$	$d(x)$
Ber( $\theta$ )	$1 - \theta$	1	$\ln \frac{\theta}{1-\theta}$	$x$
bin( $k, \theta$ )	$(1 - \theta)^k$	$\binom{k}{x}$	$\ln \frac{\theta}{1-\theta}$	$x$
Poisson( $\theta$ )	$e^{-\theta}$	$\frac{1}{x!}$	$\ln \theta$	$x$
geo( $\theta$ )	$\theta$	1	$\ln(1 - \theta)$	$x$
bin neg( $k, \theta$ )	$(\frac{\theta}{1-\theta})^k$	$\binom{k+x-1}{k-1}$	$\ln(1 - \theta)$	$k + x$
N( $\theta, \sigma^2$ )	$\frac{1}{\sqrt{2\pi}} e^{-\theta^2/2\sigma^2}$	$\frac{1}{\sigma} e^{-x^2/2\sigma^2}$	$\frac{\theta}{\sigma^2}$	$x$
N( $\mu, \theta$ )	$\frac{1}{\theta} e^{-\mu^2/2\theta^2}$	$\frac{1}{\sqrt{2\pi}}$	$\frac{1}{\theta^2}$	$-\frac{1}{2}(x - \mu)^2$
gama( $\theta, \lambda$ )	$\frac{\lambda^\theta}{\Gamma(\theta)}$	$\frac{1}{x} e^{-\lambda x}$	$\theta$	$\ln x$
gama( $\gamma, \theta$ )	$\frac{\theta^\gamma}{\Gamma(\gamma)}$	$x^{\gamma-1}$	$-\theta$	$x$

Tabla 2.10

Observemos que en la tabla aparecen distribuciones que dependen de dos parámetros. En estos casos se considera que la distribución depende del parámetro  $\theta$ , entendiéndose que el segundo parámetro, indicado con otra letra, es constante y conocido. Substituyendo las expresiones mostradas en la tabla para las funciones  $a(\theta)$ ,  $b(x)$ ,  $c(\theta)$  y  $d(x)$  puede comprobarse en cada caso que se obtiene la función de probabilidad o de densidad correspondiente.

**Ejemplo 2.60** Es interesante observar que la representación (2.18) no es única para todas las distribuciones dentro de esta familia. Por ejemplo, para la distribución geométrica, para cada valor  $k = 0, 1, \dots$  las expresiones que

aparecen abajo producen la función de probabilidad  $\text{geo}(\theta)$ .

$$\begin{aligned} a(\theta) &= \theta/(1 - \theta)^k, \\ b(x) &= 1, \\ c(\theta) &= \ln(1 - \theta), \\ d(x) &= k + x. \end{aligned}$$

■

Ahora consideraremos el caso cuando la distribución tipo exponencial depende de varios parámetros.

**Definición 2.31** Una distribución dependiente de  $\ell$  parámetros  $\theta = (\theta_1, \dots, \theta_\ell)$  es de **tipo exponencial** si su función de probabilidad o de densidad, es de la forma

$$f(x, \theta) = a(\theta) b(x) e^{c(\theta) \cdot d(x)}, \quad -\infty < x < \infty,$$

en donde  $a(\theta) \geq 0$  y  $b(x) \geq 0$  son funciones reales que dependen únicamente de los argumentos indicados, y  $c(\theta) = (c_1(\theta), \dots, c_k(\theta))$  y  $d(x) = (d_1(x), \dots, d_k(x))$  son funciones vectoriales, y la expresión  $c(\theta) \cdot d(x)$  indica el producto punto.

Escribiendo explícitamente el producto punto  $c(\theta) \cdot d(x)$ , tenemos que  $f(x, \theta)$  se escribe como sigue

$$f(x, \theta) = a(\theta) b(x) \exp \left[ \sum_{j=1}^k c_j(\theta) d_j(x) \right].$$

En la siguiente tabla se muestran algunos ejemplos de distribuciones tipo exponencial dependientes de dos parámetros.

**Algunas distribuciones tipo exponencial de dos parámetros**

Distribución	$a(\theta_1, \theta_2)$	$b(x)$	$c_1(\theta_1, \theta_2)$	$c_2(\theta_1, \theta_2)$	$d_1(x)$	$d_2(x)$
gama( $\theta_1, \theta_2$ )	$\frac{\theta_2^{\theta_1}}{\Gamma(\theta_1)}$	$\frac{1}{x}$	$\theta_1$	$\ln x$	$-\theta_2$	$x$
N( $\theta_1, \theta_2$ )	$\frac{1}{\sqrt{2\pi\theta_2^2}} e^{-\theta_1^2/2\theta_2^2}$	1	$\frac{\theta_1}{\theta_2^2}$	$-\frac{1}{2\theta_2^2}$	$x$	$x^2$

Tabla 2.11

Es inmediato comprobar que substituyendo las expresiones de las funciones  $a$ ,  $b$ ,  $c_1$ ,  $c_2$ ,  $d_1$  y  $d_2$  indicadas en la tabla se obtiene la correspondiente función de densidad. En particular, las distribuciones gama y normal pertenecen a la familia exponencial considerando un parámetro a la vez, o ambos parámetros al mismo tiempo. Nuevamente, por brevedad, hemos omitido la expresión completa para  $b(x)$ . Tal función debe especificar el soporte de la distribución.

**Ejemplo 2.61** La distribución bin neg( $k, p$ ) no pertenece a la familia exponencial biparamétrica. ■

En el siguiente resultado se muestra explícitamente la existencia de una estadística suficiente, minimal y completa para el vector de parámetros de toda distribución dentro de la familia exponencial. La propiedad de completez es más complicada de demostrar y únicamente indicaremos algunas ideas generales de la prueba.

**Proposición 2.11** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución tipo exponencial dependiente del parámetro  $\theta = (\theta_1, \dots, \theta_\ell)$ . El vector de estadísticas  $T$  especificado abajo es una estadística suficiente minimal y completa para  $\theta$ .

$$T = \left( \sum_{i=1}^n d_1(X_i), \dots, \sum_{i=1}^n d_k(X_i) \right)$$

**Demostración.**

*Suficiencia minimal.* Sean  $(x_1, \dots, x_n)$  y  $(y_1, \dots, y_n)$  dos posibles valores de la muestra aleatoria, no necesariamente distintos. Entonces

$$\begin{aligned} \frac{f(x_1, \dots, x_n, \theta)}{f(y_1, \dots, y_n, \theta)} &= \left( \prod_{i=1}^n \frac{b(x_i)}{b(y_i)} \right) \exp \left[ c(\theta) \left( \sum_{i=1}^n d(x_i) - \sum_{i=1}^n d(y_i) \right) \right] \\ &= \left( \prod_{i=1}^n \frac{b(x_i)}{b(y_i)} \right) \exp \left[ c(\theta) (T(x_1, \dots, x_n) - T(y_1, \dots, y_n)) \right]. \end{aligned}$$

Esta cantidad no depende  $\theta$  si y sólo si el exponente es nulo para cualquier valor de  $\theta$ . Esto lleva a la condición  $T(x_1, \dots, x_n) = T(y_1, \dots, y_n)$ . Por el Teorema 2.6 concluimos que  $T$  es suficiente minimal para  $\theta$ .

*Completez.* Sea  $h$  una función tal que  $E[h(T)] = 0$ . Entonces

$$\begin{aligned} 0 &= E(h(T)) \\ &= \int_{\mathbb{R}^n} h\left(\sum_{i=1}^n d(x_i)\right) f(x_1, \dots, x_n) dx_1 \cdots dx_n \\ &= \int_{\mathbb{R}^n} h\left(\sum_{i=1}^n d(x_i)\right) (a(\theta))^n \left(\prod_{i=1}^n b(x_i)\right) \exp \left[ c(\theta) \sum_{i=1}^n d(x_i) \right] dx_1 \cdots dx_n. \end{aligned}$$

El factor  $(a(\theta))^n$  puede colocarse fuera de la integral y, dado que la integral es nula, este factor puede omitirse. Ahora consideraremos un caso particular de esta integral. Tomando el caso particular  $d(x) = x$ , la integral resultante corresponde a la transformada de Laplace bilateral respecto de la medida de Lebesgue, de dimensión  $n$ , y evaluada en el punto  $(c(\theta), \dots, c(\theta))$  de  $\mathbb{R}^n$ .

Por la propiedad de unicidad de la transformada de Laplace, esta integral es cero si, y sólo si, la función que es factor de la exponencial es cero, es decir,  $h(\sum_{i=1}^n d(x_i)) = 0$  c.s. En el caso cuando la función  $d(x)$  no es la identidad se hace un cambio de variable  $u = d(x)$  y se obtiene nuevamente una transformada de Laplace pero esta vez respecto de una medida que no es necesariamente la de Lebesgue. Nuevamente, la propiedad de unicidad de la transformada de Laplace produce nuevamente el resultado buscado. ■

La teoría matemática de la transformada de Laplace multidimensional puede consultarse en [4]. Por otro lado, la propiedad de suficiencia, sin la minimalidad, se puede demostrar directamente de la definición, o bien mediante el teorema de factorización. Se deja verificar esta afirmación como un ejercicio. Veamos algunos ejemplos.

### Ejemplo 2.62

- a) En el caso  $\text{Ber}(\theta)$ , tenemos que  $d(x) = x$ . Por lo tanto, la estadística  $T = X_1 + \cdots + X_n$  es suficiente minimal y completa para  $\theta$ .
- b) En el caso  $N(\theta_1, \theta_2)$ , tenemos que  $d_1(x) = x$  y  $d_2(x) = x^2$ . Por lo tanto, la estadística  $T = (X_1 + \cdots + X_n, X_1^2 + \cdots + X_n^2)$  es suficiente minimal y completa para  $(\theta_1, \theta_2)$ .

■

## Ejercicios

270. Demuestre la propiedad de suficiencia de la estadística que aparece en la Proposición 2.11 usando
- a) La definición.
  - b) El teorema de factorización de Neyman.



## Capítulo 3

# Estimación por intervalos

En algunos casos es preferible no dar un número como estimación de un parámetro desconocido, sino un intervalo de posibles valores. En este tipo de estimación se busca un intervalo de extremos aleatorios de tal forma que se pueda afirmar, con cierto grado de confiabilidad, que dicho intervalo contiene el verdadero valor del parámetro desconocido. A este tipo de intervalos se les llama intervalos de confianza y fueron introducidos por Jerzy Neyman<sup>1</sup> en 1937.

En este capítulo estudiaremos brevemente los conceptos básicos sobre la estimación por intervalos y proporcionaremos algunos ejemplos particulares de la forma en la que pueden encontrarse este tipo de intervalos.

### 3.1. Definiciones

Como antes, consideremos que tenemos una cierta variable aleatoria de nuestro interés y que ésta tiene función de densidad o de probabilidad conocida  $f(x, \theta)$ , pero dependiente de un parámetro desconocido  $\theta$ , el cual deseamos estimar con base en una muestra aleatoria de esta distribución. Aquí tenemos la definición de intervalo de confianza.

---

<sup>1</sup>Jerzy Neyman (1894-1981), matemático y estadístico polaco.

**Definición 3.1** Sea  $\alpha \in (0, 1)$  un número fijo dado. Un **intervalo de confianza** para un parámetro desconocido  $\theta$  de una distribución de probabilidad es un intervalo aleatorio de la forma  $(\hat{\theta}_1, \hat{\theta}_2)$ , en donde  $\hat{\theta}_1$  y  $\hat{\theta}_2$  son dos estadísticas que satisfacen

$$P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = 1 - \alpha. \quad (3.1)$$

A las estadísticas  $\hat{\theta}_1$  y  $\hat{\theta}_2$  se les conoce como límites inferior y superior, respectivamente, del intervalo de confianza. Al número  $1 - \alpha$  se le conoce como grado o coeficiente de confianza. En general, se toma el valor de  $\alpha$  cercano a cero de tal forma que el grado de confianza,  $1 - \alpha$ , sea cercano a uno. En la práctica es común tomar  $\alpha = 0.05$ , de modo que el grado de confianza es  $1 - \alpha = 0.95$ . Decimos entonces que el grado de confianza es del 95 %.

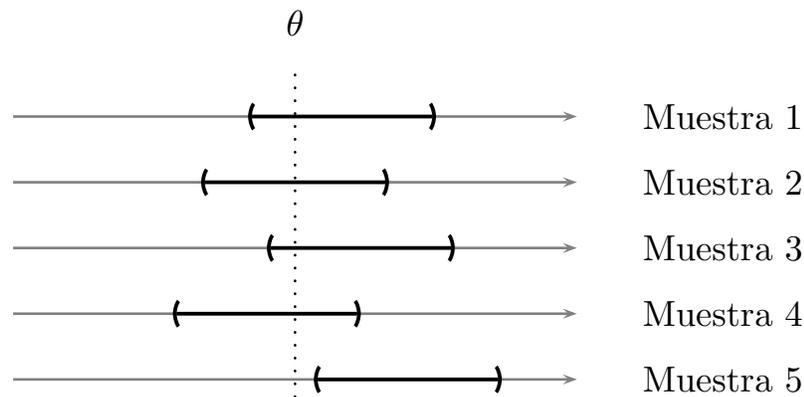


Figura 3.1

Siendo que los límites inferior y superior de un intervalo de confianza son funciones de una muestra aleatoria  $X_1, \dots, X_n$ , al tomar estas variables aleatorias distintos valores se generan distintas realizaciones del intervalo aleatorio. Esta situación se ilustra en la Figura 3.1, en donde se ha indicado gráficamente el valor desconocido de  $\theta$ . Algunas realizaciones del intervalo

de confianza contendrán el valor del parámetro y algunas realizaciones no lo contendrán. Usando la interpretación frecuentista de la probabilidad, podemos afirmar que en un gran número de realizaciones del intervalo aleatorio, el  $(1 - \alpha)100\%$  de la veces el intervalo contendrá el valor del parámetro a estimar.

Observe además que no es correcto decir “la probabilidad de que  $\theta$  pertenezca al intervalo  $(\hat{\theta}_1, \hat{\theta}_2)$  es  $1 - \alpha$ ”, pues, en nuestra perspectiva clásica, el parámetro  $\theta$  no es un elemento aleatorio. En cambio, se dice “la probabilidad de que el intervalo  $(\hat{\theta}_1, \hat{\theta}_2)$  contenga el valor de  $\theta$  es  $1 - \alpha$ ”. De esta forma se entiende que  $\theta$  es constante, aunque desconocido, y el intervalo es el que cambia dependiendo de la muestra aleatoria. Naturalmente el problema fundamental es el siguiente:

¿Cómo encontrar  $\hat{\theta}_1$  y  $\hat{\theta}_2$  de tal forma que la igualdad (3.1) se cumpla?

Este problema no es fácil de resolver. En muchas ocasiones sólo se pueden encontrar intervalos de confianza aproximados, es decir, las estadísticas  $\hat{\theta}_1$  y  $\hat{\theta}_2$  que se encuentran son tales que la igualdad (3.1) se cumple sólo de manera aproximada. En los ejemplos que estudiaremos se observará además que los extremos del intervalo de confianza no se encuentran por separado sino de manera paralela.

El así llamado método pivotal es una manera general de resolver el problema planteado, aunque presupone poder encontrar una variable aleatoria con ciertas características. Explicaremos a continuación este método.

### Método pivotal

Este método supone poder encontrar una función de la muestra y del parámetro desconocido, denotémosla por  $q(X_1, \dots, X_n, \theta)$ , con distribución de probabilidad completamente conocida (no dependiente de  $\theta$ ), de tal manera que puedan determinarse dos números  $a < b$  tales que

$$P(a < q(X_1, \dots, X_n, \theta) < b) = 1 - \alpha.$$

Después, a partir de esta expresión, se debe buscar desprender el término  $\theta$  del evento determinado por las desigualdades y encontrar una expresión de la forma (3.1). A la función  $q(X_1, \dots, X_n, \theta)$  se le llama cantidad pivotal pues de ella se busca obtener el término  $\theta$ .

En las siguientes secciones estudiaremos formas de encontrar intervalos de confianza para los parámetros de algunas distribuciones de probabilidad conocidas. Usaremos principalmente el método pivotal. En general, el problema de encontrar intervalos de confianza para algún parámetro o función parametral de una distribución dada no es sencillo.

### 3.2. Distribución Bernoulli

Supongamos que una cierta variable aleatoria de interés tiene distribución  $\text{Ber}(\theta)$ , en donde el parámetro  $\theta$  es desconocido. Deseamos estimar este parámetro mediante un intervalo de confianza. Sea  $X_1, \dots, X_n$  una muestra aleatoria de esta distribución. Haremos uso del hecho de que un estimador puntual para  $\theta$  es  $\bar{X}$ , en donde  $E(\bar{X}) = \theta$  y  $\text{Var}(\bar{X}) = \theta(1 - \theta)/n$ . Por el teorema central del límite, de manera aproximada,

$$\frac{\bar{X} - \theta}{\sqrt{\theta(1 - \theta)/n}} \sim N(0, 1).$$

En tablas de probabilidades de la distribución normal se pueden encontrar dos valores  $a$  y  $b$  tales que la probabilidad de que esta variable aleatoria tome un valor entre  $a$  y  $b$  sea igual a  $1 - \alpha$ . Como es deseable que la longitud del intervalo  $(a, b)$  sea la más pequeña posible y como la distribución normal estándar es simétrica alrededor del origen, resulta que el intervalo  $(a, b)$  de longitud mínima debe ser también simétrico alrededor del origen. Así, puede encontrarse un valor positivo, que denotaremos por  $z_{\alpha/2}$ , tal que se cumple lo siguiente:

$$P(-z_{\alpha/2} < \frac{\bar{X} - \theta}{\sqrt{\theta(1 - \theta)/n}} < z_{\alpha/2}) \approx 1 - \alpha. \quad (3.2)$$

Véase la Figura 3.2 y observe que el intervalo indicado es el de longitud más pequeña que cumple (3.2).

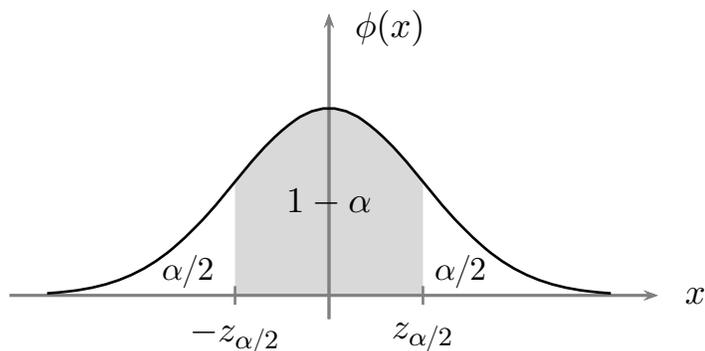


Figura 3.2

El problema aquí es encontrar  $\theta$  a partir de las dos desigualdades que aparecen en (3.2). Presentamos a continuación tres formas en que tal tarea puede llevarse a cabo de manera aproximada.

### Primera solución

Una simplificación al problema planteado consiste en substituir el denominador  $\theta(1 - \theta)/n$  por la estimación puntual  $\bar{X}(1 - \bar{X})/n$ . Es necesario admitir que esta substitución es un tanto burda, pero como resultado se obtendrá una cantidad pivotal a partir de la cual se producirá con facilidad una aproximación al intervalo buscado. Tenemos entonces la expresión

$$P(-z_{\alpha/2} < \frac{\bar{X} - \theta}{\sqrt{\bar{X}(1 - \bar{X})/n}} < z_{\alpha/2}) \approx 1 - \alpha.$$

Resolviendo las dos desigualdades para  $\theta$  se obtiene el siguiente resultado.

**Proposición 3.1** Un intervalo de confianza aproximado para el parámetro de la distribución  $\text{Ber}(\theta)$  está dado por

$$P\left(\bar{X} - \frac{z_{\alpha/2}}{\sqrt{n}} \sqrt{\bar{X}(1 - \bar{X})} < \theta < \bar{X} + \frac{z_{\alpha/2}}{\sqrt{n}} \sqrt{\bar{X}(1 - \bar{X})}\right) \approx 1 - \alpha.$$

Observemos que este intervalo aleatorio tiene como centro la media muestral y se extiende a la derecha y a la izquierda la misma cantidad aleatoria. Por

lo tanto, es un intervalo simétrico y su longitud total es la variable aleatoria  $L = 2z_{\alpha/2} \sqrt{\bar{X}(1-\bar{X})}/\sqrt{n}$ .

### Segunda solución

Otra alternativa para desprender de manera aproximada el parámetro  $\theta$  en la ecuación (3.2) es usar la desigualdad  $\theta(1-\theta) \leq 1/4$  para el denominador que aparece en esa ecuación. Esta desigualdad produce la siguiente cota superior

$$\sqrt{\theta(1-\theta)/n} \leq \frac{1}{2\sqrt{n}}.$$

Utilizando esto en las dos desigualdades de (3.2) se obtiene

$$-\frac{z_{\alpha/2}}{2\sqrt{n}} < -z_{\alpha/2} \sqrt{\theta(1-\theta)/n} < \bar{X} - \theta < z_{\alpha/2} \sqrt{\theta(1-\theta)/n} < \frac{z_{\alpha/2}}{2\sqrt{n}}.$$

En consecuencia, tenemos el siguiente intervalo aproximado.

**Proposición 3.2** Un intervalo de confianza aproximado para el parámetro de la distribución  $\text{Ber}(\theta)$  está dado por

$$P\left(\bar{X} - \frac{z_{\alpha/2}}{2\sqrt{n}} < \theta < \bar{X} + \frac{z_{\alpha/2}}{2\sqrt{n}}\right) \approx 1 - \alpha.$$

Observemos que la longitud de este intervalo es no aleatoria  $L = z_{\alpha/2}/\sqrt{n}$ , y que esta cantidad crece conforme la confianza  $1 - \alpha$  se acerca a 1, y decrece conforme el tamaño de la muestra crece.

### Tercera solución

Como una tercera alternativa para producir una cantidad pivotal de la ecuación (3.2), observemos que el evento en cuestión puede escribirse como  $(|\bar{X} - \theta| < z_{\alpha/2} \sqrt{\theta(1-\theta)/n})$ . Elevando al cuadrado y desarrollando se llega a la desigualdad

$$\theta^2(1 + z_{\alpha/2}^2/n) + \theta(-2\bar{X} - z_{\alpha/2}^2/n) + \bar{X}^2 < 0.$$

Considerando la igualdad, las raíces de esta ecuación cuadrática en  $\theta$  son

$$\begin{aligned}\theta_1 &= \bar{X}/(1 + z_{\alpha/2}^2/n), \\ \theta_2 &= (\bar{X} + z_{\alpha/2}^2/n)/(1 + z_{\alpha/2}^2/n).\end{aligned}$$

Por lo tanto, la ecuación cuadrática es negativa cuando  $\theta_1 < \theta < \theta_2$ , es decir, se tiene entonces el siguiente resultado.

**Proposición 3.3** Un intervalo de confianza aproximado para el parámetro de la distribución  $\text{Ber}(\theta)$  está dado por

$$P\left(\frac{\bar{X}}{1 + z_{\alpha/2}^2/n} < \theta < \frac{\bar{X} + z_{\alpha/2}^2/n}{1 + z_{\alpha/2}^2/n}\right) \approx 1 - \alpha.$$

El intervalo encontrado sigue siendo una aproximación pues tiene como punto de partida la expresión (3.2). Es un intervalo no simétrico de longitud no aleatoria  $L = z_{\alpha/2}^2/(n + z_{\alpha/2}^2)$ .

### 3.3. Distribución uniforme continua

Encontraremos un intervalo de confianza para cada parámetro de la distribución  $\text{unif}(a, b)$ , considerando siempre un parámetro conocido y el otro desconocido. Empezaremos con un caso particular.

#### Primer caso

Consideremos una distribución  $\text{unif}(0, \theta)$ , en donde el parámetro  $\theta > 0$  es desconocido. Encontraremos un intervalo de confianza para este parámetro a partir de una muestra aleatoria  $X_1, \dots, X_n$ . Puede comprobarse que la máxima estadística de orden  $X_{(n)}$  es una estadística suficiente para  $\theta$  y que la variable  $(1/\theta)X_{(n)}$  tiene función de densidad

$$f(x) = \begin{cases} nx^{n-1} & \text{si } 0 < x < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

La gráfica de esta función se muestra en la Figura 3.3 y la correspondiente función de distribución es la siguiente,

$$F(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ x^n & \text{si } 0 < x < 1, \\ 1 & \text{si } x \geq 1. \end{cases}$$

Entonces, dado un valor de  $\alpha \in (0, 1)$ , se pueden encontrar dos valores  $a$  y  $b$  tales que  $0 < a < b < 1$  con

$$P\left(\frac{1}{\theta} X_{(n)} < a\right) = \alpha/2,$$

$$P\left(\frac{1}{\theta} X_{(n)} > b\right) = \alpha/2.$$

Véase la Figura 3.3. A partir de la expresión de la función de distribución puede comprobarse con facilidad que los valores  $a = (\alpha/2)^{1/n}$  y  $b = (1 - \alpha/2)^{1/n}$  satisfacen las condiciones arriba indicadas.

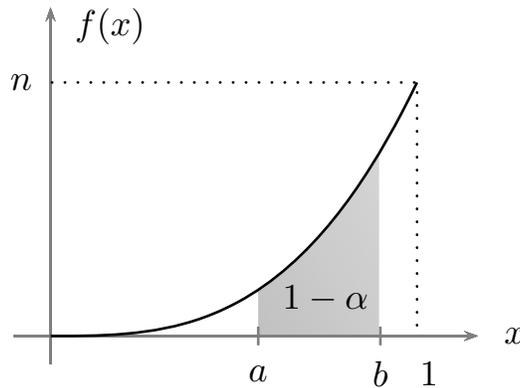


Figura 3.3

Estos dos valores de  $a$  y  $b$  no necesariamente producen un intervalo de longitud mínima, pero son tales que

$$P\left((\alpha/2)^{1/n} < \frac{1}{\theta} X_{(n)} < (1 - \alpha/2)^{1/n}\right) = 1 - \alpha,$$

de donde se obtiene el siguiente resultado.

**Proposición 3.4** Un intervalo de confianza para el parámetro de la distribución  $\text{unif}(0, \theta)$  está dado por

$$P\left(\frac{2^{1/n} X_{(n)}}{(2 - \alpha)^{1/n}} < \theta < \frac{2^{1/n} X_{(n)}}{\alpha^{1/n}}\right) = 1 - \alpha.$$

### Segundo caso

Consideremos ahora la distribución  $\text{unif}(c, \theta)$  con  $c$  conocido y  $\theta$  desconocido. Encontraremos un intervalo de confianza para  $\theta$ . Sea  $X_1, \dots, X_n$  una muestra aleatoria de esta distribución. Entonces  $X_1 - c, \dots, X_n - c$  es una muestra aleatoria de la distribución  $\text{unif}(0, \theta - c)$  y estamos nuevamente en la situación del caso estudiado antes. Puede comprobarse que la estadística  $X_{(n)} - c$  es suficiente para  $\theta - c$  y el cociente  $(X_{(n)} - c)/(\theta - c)$  tiene función de densidad

$$f(x) = \begin{cases} nx^{n-1} & \text{si } 0 < x < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

Entonces, dado un valor de  $\alpha \in (0, 1)$ , se pueden encontrar dos cantidades  $a$  y  $b$  tales que  $0 < a < b < 1$  con

$$P\left(\frac{X_{(n)} - c}{\theta - c} < a\right) = \alpha/2,$$

$$P\left(\frac{X_{(n)} - c}{\theta - c} > b\right) = \alpha/2.$$

Esta situación corresponde nuevamente a la que se muestra en la Figura 3.3, en donde  $a = (\alpha/2)^{1/n}$  y  $b = (1 - \alpha/2)^{1/n}$ . El intervalo encontrado no tiene longitud mínima, sin embargo, tenemos que

$$P\left((\alpha/2)^{1/n} < \frac{X_{(n)} - c}{\theta - c} < (1 - \alpha/2)^{1/n}\right) = 1 - \alpha,$$

de donde se obtiene el siguiente resultado.

**Proposición 3.5** Un intervalo de confianza para el parámetro  $\theta$  de una distribución  $\text{unif}(c, \theta)$ , en donde  $c$  es conocido, está dado por

$$P\left(c + \frac{2^{1/n}(X_{(n)} - c)}{(2 - \alpha)^{1/n}} < \theta < c + \frac{2^{1/n}(X_{(n)} - c)}{\alpha^{1/n}}\right) = 1 - \alpha.$$

Este intervalo se reduce al encontrado antes cuando  $c = 0$ .

### Tercer caso

Finalmente consideremos la distribución  $\text{unif}(\theta, c)$ , con  $c$  conocido y  $\theta$  desconocido. Encontraremos un intervalo de confianza para  $\theta$ . Sea  $X_1, \dots, X_n$  una muestra aleatoria de esta distribución. Entonces  $X_1 - c, \dots, X_n - c$  es una muestra aleatoria de la distribución  $\text{unif}(\theta - c, 0)$ . Multiplicando por  $-1$  tenemos que  $c - X_1, \dots, c - X_n$  es una muestra aleatoria de la distribución  $\text{unif}(0, c - \theta)$ . Procedemos como antes. Puede comprobarse que la estadística  $c - X_{(1)}$  es suficiente para  $c - \theta$  y el cociente  $(c - X_{(1)})/(c - \theta)$  tiene función de densidad

$$f(x) = \begin{cases} nx^{n-1} & \text{si } 0 < x < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

Entonces, dado un valor de  $\alpha \in (0, 1)$ , se pueden encontrar constantes  $a$  y  $b$  tales que  $0 < a < b < 1$  con

$$P\left(\frac{c - X_{(1)}}{c - \theta} < c_1\right) = \alpha/2,$$

$$P\left(\frac{c - X_{(1)}}{c - \theta} > c_2\right) = \alpha/2.$$

Véase nuevamente la Figura 3.3, en donde  $a = (\alpha/2)^{1/n}$  y  $b = (1 - \alpha/2)^{1/n}$ . Estos valores de  $a$  y  $b$  no satisfacen que  $b - a$  sea mínimo pero son tales que

$$P\left((\alpha/2)^{1/n} < \frac{c - X_{(1)}}{c - \theta} < (1 - \alpha/2)^{1/n}\right) = 1 - \alpha.$$

Y de aquí se obtiene el siguiente resultado.

**Proposición 3.6** Un intervalo de confianza para el parámetro  $\theta$  de la distribución  $\text{unif}(\theta, c)$ , en donde  $c$  es conocido, está dado por

$$P\left(c - \frac{2^{1/n}(c - X_{(1)})}{\alpha^{1/n}} < \theta < c - \frac{2^{1/n}(c - X_{(1)})}{(2 - \alpha)^{1/n}}\right) = 1 - \alpha.$$

### 3.4. Distribución exponencial

Encontraremos un intervalo de confianza exacto para el parámetro de la distribución exponencial, a partir de una cantidad pivotal que construiremos a continuación. Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{exp}(\theta)$ . Sabemos que la variable aleatoria  $X_1 + \dots + X_n$  tiene distribución  $\text{gama}(n, \theta)$ . Por otro lado, para cualquier constante  $c > 0$  y para cualquier variable aleatoria continua  $X$  con función de distribución  $F(x)$  y función de densidad  $f(x)$ , se cumple que

$$\begin{aligned} F_{cX}(x) &= F_X(x/c), \\ f_{cX}(x) &= \frac{1}{c} f_X(x/c). \end{aligned}$$

Se pueden usar estos resultados para comprobar que, para el caso en estudio,

$$c(X_1 + \dots + X_n) \sim \text{gama}(n, \theta/c).$$

Tomando  $c = \theta$  se encuentra que  $\theta(X_1 + \dots + X_n) \sim \text{gama}(n, 1)$ . Esta variable aleatoria involucra al parámetro  $\theta$  y su distribución está ahora completamente especificada. Esta es la cantidad pivotal buscada. Así, para cualquier valor  $\alpha \in (0, 1)$ , se pueden encontrar dos valores positivos  $a < b$  tales que

$$P(a < \theta(X_1 + \dots + X_n) < b) = 1 - \alpha.$$

Una manera de determinar los valores de  $a$  y  $b$  es a través de las siguientes dos condiciones:

$$\begin{aligned} P(\theta(X_1 + \dots + X_n) < a) &= \alpha/2, \\ P(\theta(X_1 + \dots + X_n) > b) &= \alpha/2. \end{aligned}$$

Véase la Figura 3.4 en donde se muestra la función de densidad de la distribución  $\text{gama}(n, 1)$ , el valor  $a$  se denota por  $\gamma_{1-\alpha/2}$  y el valor  $b$  por  $\gamma_{\alpha/2}$ . Dado un valor de  $\alpha$ , los valores para  $\gamma_{1-\alpha/2}$  y  $\gamma_{\alpha/2}$  pueden obtenerse de manera aproximada usando algún paquete computacional.

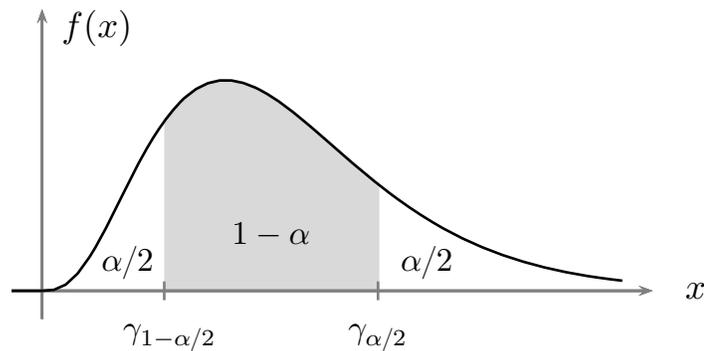


Figura 3.4

Observe que el intervalo considerado no necesariamente es el de longitud más pequeña, sin embargo, permite obtener el siguiente intervalo de confianza.

**Proposición 3.7** Un intervalo de confianza para el parámetro de la distribución  $\exp(\theta)$  está dado por

$$P\left(\frac{\gamma_{1-\alpha/2}}{n\bar{X}} < \theta < \frac{\gamma_{\alpha/2}}{n\bar{X}}\right) = 1 - \alpha.$$

### 3.5. Distribución normal

Encontraremos intervalos de confianza para los parámetros de una distribución normal en varias situaciones.

### Intervalo para la media cuando la varianza es conocida

Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución normal con media desconocida  $\theta$  y varianza conocida  $\sigma^2$ . Encontraremos un intervalo de confianza para  $\theta$ . Como cada una de las variables de la muestra tiene distribución  $N(\theta, \sigma^2)$ , la media muestral  $\bar{X}$  tiene distribución  $N(\theta, \sigma^2/n)$ . De modo que, estandarizando,

$$\frac{\bar{X} - \theta}{\sigma/\sqrt{n}} \sim N(0, 1).$$

En esta situación, esta es la cantidad pivotal que nos ayudará a encontrar el intervalo de confianza buscado. Para cualquier valor de  $\alpha \in (0, 1)$  podemos encontrar un valor  $z_{\alpha/2}$  en tablas de probabilidad normal estándar, véase la Figura 3.5, tal que

$$P(-z_{\alpha/2} < \frac{\bar{X} - \theta}{\sigma/\sqrt{n}} < z_{\alpha/2}) = 1 - \alpha.$$

Como la función de densidad normal estándar es simétrica alrededor del origen, el intervalo de longitud más pequeña y sobre el cual esta función de densidad cubre un área igual a  $1 - \alpha$ , es necesariamente un intervalo simétrico alrededor del origen. Así, el intervalo propuesto es de longitud mínima.

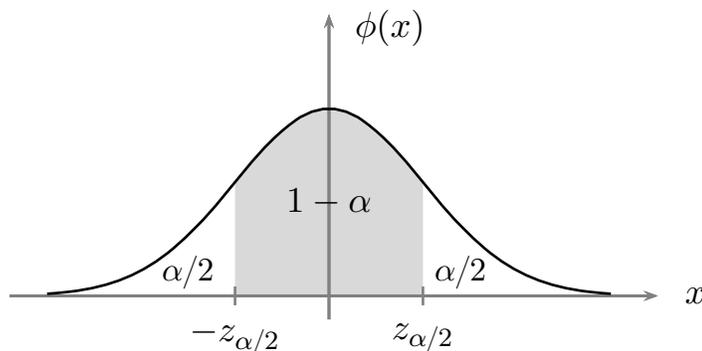


Figura 3.5

Despejando la constante desconocida  $\theta$  se obtiene el siguiente resultado.

**Proposición 3.8** Un intervalo de confianza para la media  $\theta$  de una distribución normal con varianza conocida  $\sigma^2$  está dado por

$$P\left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \theta < \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha. \quad (3.3)$$

Observe que la longitud del intervalo de confianza encontrado es no aleatorio  $L = 2z_{\alpha/2} \cdot \sigma/\sqrt{n}$ . De aquí pueden obtenerse varias observaciones:

- La longitud del intervalo decrece conforme el tamaño de la muestra crece, es decir, mientras mayor información se tenga más preciso es el intervalo. En el límite cuando  $n \rightarrow \infty$ , el intervalo se colapsa en el estimador puntual  $\bar{X}$ .
- Si la confianza requerida crece, es decir, si  $1 - \alpha$  aumenta, entonces  $z_{\alpha/2}$  crece, véase la Figura 3.5, y por lo tanto la longitud del intervalo también crece.
- Si la dispersión de los datos es alta, es decir, si la desviación estándar  $\sigma$  es grande, entonces la longitud del intervalo tiende a ser grande.

### Intervalo para la media cuando la varianza es desconocida

Consideremos nuevamente una distribución normal con media desconocida  $\theta$  pero ahora con varianza desconocida. El resultado teórico que utilizaremos es el siguiente:

$$\frac{\bar{X} - \theta}{S/\sqrt{n}} \sim t(n - 1).$$

Observe que esta es la distribución exacta de esta variable aleatoria, sin importar el tamaño  $n \geq 2$  de la muestra y sobre todo, sin suponer que la varianza es conocida. A partir de lo anterior podemos construir un intervalo de confianza para el parámetro  $\theta$  de forma análoga al caso normal mencionado antes. Para cualquier valor de  $\alpha \in (0, 1)$  podemos encontrar un valor

$t_{\alpha/2} > 0$  en tablas de probabilidad de la distribución  $t$  de  $n - 1$  grados de libertad (véase la Figura 3.6) tal que

$$P(-t_{\alpha/2} < \frac{\bar{X} - \theta}{S/\sqrt{n}} < t_{\alpha/2}) = 1 - \alpha.$$

Debido a la simetría alrededor del origen de la función de densidad de la distribución  $t(n - 1)$ , el intervalo indicado es el de longitud mínima.

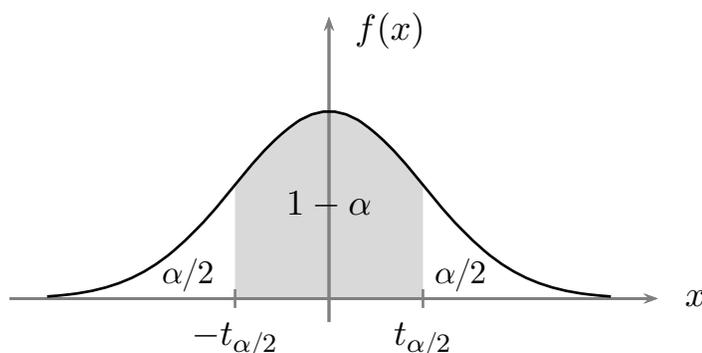


Figura 3.6

Despejando la constante desconocida  $\theta$  de las dos desigualdades anteriores se obtiene el siguiente resultado.

**Proposición 3.9** Un intervalo de confianza para la media  $\theta$  de una distribución normal está dado por la siguiente expresión

$$P\left(\bar{X} - t_{\alpha/2} \frac{S}{\sqrt{n}} < \theta < \bar{X} + t_{\alpha/2} \frac{S}{\sqrt{n}}\right) = 1 - \alpha. \quad (3.4)$$

De este modo, el intervalo aleatorio  $(\bar{X} - t_{\alpha/2} \frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2} \frac{S}{\sqrt{n}})$  es un intervalo de confianza para la media de una población normal sin suponer la varianza conocida. No lo hemos escrito de manera explícita en la fórmula anterior pero el valor  $t_{\alpha/2}$  corresponde a la distribución  $t$  con  $n - 1$  grados de libertad. Para mayor precisión se escribe también  $t_{\alpha/2, n-1}$ .

### Intervalo para la varianza

Encontraremos un intervalo de confianza para la varianza  $\theta^2 > 0$  de una distribución normal. En este caso el resultado teórico de utilidad es el siguiente: dada una muestra aleatoria de tamaño  $n$  de esta distribución,

$$(n - 1) \frac{S^2}{\theta^2} \sim \chi^2(n - 1).$$

Esta es la cantidad pivotal que nos ayudará a encontrar el intervalo buscado. Dado un valor de  $\alpha$ , usando algún paquete computacional o mediante una tabla de probabilidades de la distribución  $\chi^2(n - 1)$ , se pueden encontrar dos valores  $0 < \chi_{1-\alpha/2}^2 < \chi_{\alpha/2}^2$  tales que

$$P\left((n - 1) \frac{S^2}{\theta^2} < \chi_{1-\alpha/2}^2\right) = \alpha/2,$$

$$P\left((n - 1) \frac{S^2}{\theta^2} > \chi_{\alpha/2}^2\right) = \alpha/2.$$

Véase la Figura 3.7. El intervalo  $(\chi_{1-\alpha/2}^2, \chi_{\alpha/2}^2)$  no es necesariamente el de longitud mínima, pero es tal que

$$P(\chi_{1-\alpha/2}^2 < (n - 1) \frac{S^2}{\theta^2} < \chi_{\alpha/2}^2) = 1 - \alpha.$$

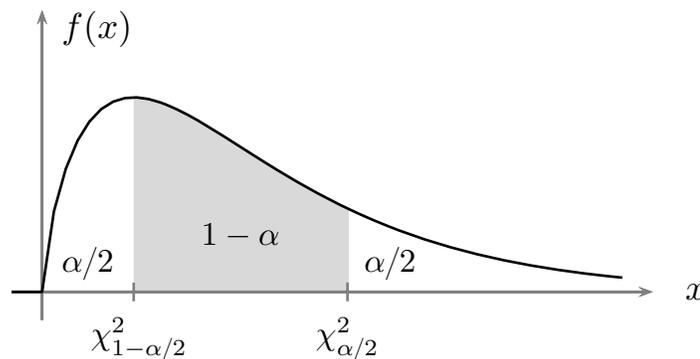


Figura 3.7

Despejando la constante desconocida  $\theta^2$  de las dos desigualdades anteriores se obtiene el siguiente intervalo de confianza.

**Proposición 3.10** Un intervalo de confianza para la varianza desconocida  $\theta^2$  de una distribución normal está dado por

$$P\left(\frac{(n-1)S^2}{\chi_{\alpha/2}^2} < \theta^2 < \frac{(n-1)S^2}{\chi_{1-\alpha/2}^2}\right) = 1 - \alpha. \quad (3.5)$$

De este resultado puede derivarse un intervalo de confianza para la desviación estándar  $\theta$ . Por simplicidad hemos escrito  $\chi_{\alpha/2}^2$ , la expresión completa, incluyendo los grados de libertad, debe ser  $\chi_{\alpha/2, n-1}^2$ . Análogamente para  $\chi_{1-\alpha/2, n-1}^2$ .

### Intervalo para la diferencia de dos medias cuando las varianzas son conocidas

Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución  $N(\theta_1, \sigma_1^2)$  y sea  $Y_1, \dots, Y_m$  otra muestra aleatoria, independiente de la primera, de una distribución  $N(\theta_2, \sigma_2^2)$ . Consideraremos que las medias  $\theta_1$  y  $\theta_2$  son desconocidas y deseamos encontrar un intervalo de confianza para la diferencia  $\theta_1 - \theta_2$ . En esta sección consideraremos el caso cuando las varianzas  $\sigma_1^2$  y  $\sigma_2^2$  son conocidas pero pueden ser diferentes. Como  $\bar{X} \sim N(\theta_1, \sigma_1^2/n)$  y  $\bar{Y} \sim N(\theta_2, \sigma_2^2/m)$ , tenemos que

$$\frac{(\bar{X} - \bar{Y}) - (\theta_1 - \theta_2)}{\sqrt{\sigma_1^2/n + \sigma_2^2/m}} \sim N(0, 1).$$

Puede entonces encontrarse un valor  $z_{\alpha/2}$  de la distribución normal estándar tal que

$$P(-z_{\alpha/2} < \frac{(\bar{X} - \bar{Y}) - (\theta_1 - \theta_2)}{\sqrt{\sigma_1^2/n + \sigma_2^2/m}} < z_{\alpha/2}) = 1 - \alpha.$$

En este caso, el intervalo simétrico indicado es el de longitud mínima que satisface la condición anterior. De aquí se puede obtener el intervalo de confianza buscado.

**Proposición 3.11** Un intervalo de confianza al  $(1 - \alpha)100\%$  para la diferencia de medias  $\theta_1 - \theta_2$  de dos distribuciones normales  $N(\theta_1, \sigma_1^2)$  y  $N(\theta_2, \sigma_2^2)$ , cuando las varianzas son conocidas está dado por

$$(\bar{X} - \bar{Y}) \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}.$$

### Intervalo para la diferencia de dos medias cuando las varianzas son desconocidas pero iguales

Considere nuevamente que  $X_1, \dots, X_n$  es una muestra aleatoria de una distribución  $N(\theta_1, \sigma^2)$  y sea  $Y_1, \dots, Y_m$  otra muestra aleatoria, independiente de la primera, de una distribución  $N(\theta_2, \sigma^2)$ . Observe que estamos en la situación cuando la varianza  $\sigma^2$  es común a ambas distribuciones. Consideraremos que los tres parámetros  $\theta_1, \theta_2$  y  $\sigma^2$  son desconocidos. Deseamos encontrar un intervalo de confianza para la diferencia  $\theta_1 - \theta_2$ . Definamos las siguientes varianzas muestrales.

$$\begin{aligned} S_X^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \\ S_Y^2 &= \frac{1}{m-1} \sum_{i=1}^m (Y_i - \bar{Y})^2, \\ S^2 &= \frac{1}{n+m-2} [(n-1)S_X^2 + (m-1)S_Y^2]. \end{aligned}$$

El último término es una varianza muestral combinada de las dos muestras. Recordemos los siguientes resultados:

- $(n-1) S_X^2 / \sigma^2 \sim \chi^2(n-1)$ .
- $(m-1) S_Y^2 / \sigma^2 \sim \chi^2(m-1)$ .
- Las variables  $(n-1) S_X^2 / \sigma^2$  y  $(m-1) S_Y^2 / \sigma^2$  son independientes.
- $(n-1) S_X^2 / \sigma^2 + (m-1) S_Y^2 / \sigma^2 \sim \chi^2(n+m-2)$ .

Tenemos además estas otras afirmaciones:

- $\frac{\bar{X} - \theta_1}{S_X/\sqrt{n}} \sim t(n - 1).$
- $\frac{\bar{Y} - \theta_2}{S_Y/\sqrt{m}} \sim t(m - 1).$
- $\frac{(\bar{X} - \bar{Y}) - (\theta_1 - \theta_2)}{S\sqrt{1/n + 1/m}} \sim t(n + m - 2).$

El último de estos resultados es el que tomaremos como cantidad pivotal. Observe que en el denominador de esta última variable aleatoria aparece la varianza muestral combinada  $S$  definida antes. Se puede encontrar un valor  $t_{\alpha/2} > 0$  de la distribución  $t(n + m - 2)$  tal que

$$P(-t_{\alpha/2} < \frac{(\bar{X} - \bar{Y}) - (\theta_1 - \theta_2)}{S\sqrt{1/n + 1/m}} < t_{\alpha/2}) = 1 - \alpha,$$

de donde se obtiene el intervalo de confianza buscado.

**Proposición 3.12** Un intervalo de confianza al  $(1 - \alpha)100\%$  para la diferencia de medias  $\theta_1 - \theta_2$  de dos distribuciones normales  $N(\theta_1, \sigma^2)$  y  $N(\theta_2, \sigma^2)$ , cuando las varianzas son iguales y desconocidas, está dado por

$$(\bar{X} - \bar{Y}) \pm t_{\alpha/2} S \sqrt{\frac{1}{n} + \frac{1}{m}}.$$

### 3.6. Intervalo para la media de una distribución cualquiera

Consideremos una distribución cualquiera cuya media es un parámetro desconocido  $\theta$  y una muestra aleatoria de tamaño  $n$  de esta distribución. Si  $n$

es suficientemente grande, por ejemplo puede ser  $n \geq 30$ , con cierta confianza puede aplicarse el teorema central del límite, y entonces de manera aproximada tenemos que

$$\frac{\bar{X} - \theta}{S/\sqrt{n}} \sim N(0, 1).$$

Ahora, para cualquier valor de  $\alpha \in (0, 1)$  podemos encontrar un valor  $z_{\alpha/2}$  en tablas de probabilidad normal estándar tal que

$$P(-z_{\alpha/2} < \frac{\bar{X} - \theta}{S/\sqrt{n}} < z_{\alpha/2}) \approx 1 - \alpha.$$

Resolviendo para  $\theta$  en las dos desigualdades anteriores se obtiene el siguiente intervalo de confianza.

**Proposición 3.13** Un intervalo de confianza aproximado para la media  $\theta$  de una distribución cualquiera está dado por

$$P(\bar{X} - z_{\alpha/2} \frac{S}{\sqrt{n}} < \theta < \bar{X} + z_{\alpha/2} \frac{S}{\sqrt{n}}) \approx 1 - \alpha.$$

### 3.7. Intervalos conjuntos para dos parámetros

Sea  $f(x; \theta_1, \theta_2)$  una distribución de probabilidad dependiente de dos parámetros desconocidos. Supongamos que  $I_1$  e  $I_2$  son dos intervalos de confianza para cada uno de estos parámetros, suponiendo en cada caso que no se conoce el otro parámetro. Suponga que la confianza del primer intervalo es  $1 - \alpha_1$  y la del segundo intervalo es  $1 - \alpha_2$ . El objetivo es encontrar la confianza conjunta de estos dos intervalos.

Recordemos que para cualesquiera dos eventos  $A$  y  $B$  se cumple la desigualdad  $P(A \cap B) \geq 1 - P(A^c) - P(B^c)$ . Por lo tanto, tenemos que

$$\begin{aligned} P(\theta_1 \in I_1, \theta_2 \in I_2) &\geq 1 - P(\theta_1 \notin I_1) + P(\theta_2 \notin I_2) \\ &= 1 - (\alpha_1 + \alpha_2). \end{aligned}$$

Así, la confianza conjunta es por lo menos  $1 - (\alpha_1 + \alpha_2)$ . Si se desea que este valor sea  $1 - \alpha$ , entonces puede solicitarse inicialmente que  $\alpha_1 = \alpha/2$  y  $\alpha_2 = \alpha/2$ . Esto significa que se necesita una confianza mayor para cada intervalo de manera individual para garantizar una confianza igual a  $1 - \alpha$  para el intervalo conjunto.

**Ejemplo 3.1** Consideremos una distribución  $N(\mu, \sigma^2)$ , con ambos parámetros desconocidos. En esta situación, anteriormente hemos encontrado los siguientes intervalos de confianza individuales con confianza  $1 - \alpha_1$  y  $1 - \alpha_2$ , respectivamente, para estos parámetros:

$$I_1 = \left( \bar{X} - t_{\alpha_1/2} \frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha_1/2} \frac{S}{\sqrt{n}} \right),$$

$$I_2 = \left( \frac{(n-1)S^2}{\chi_{\alpha_2/2}^2}, \frac{(n-1)S^2}{\chi_{1-\alpha_2/2}^2} \right).$$

Tomando  $\alpha_1 = \alpha_2 = \alpha/2$ , tenemos que  $P(\mu \in I_1, \sigma^2 \in I_2) \geq 1 - \alpha$ . ■

La confianza conjunta para  $n$  intervalos de confianza puede acotarse por abajo usando la siguiente fórmula general para  $n$  eventos, la cual generaliza al caso  $n = 2$  mencionado antes.

$$P(A_1 \cap \cdots \cap A_n) \geq 1 - \sum_{i=1}^n P(A_i^c).$$

## Ejercicios

271. Sea  $X_1$  una muestra aleatoria de tamaño  $n = 1$  de la distribución  $f(x, \theta)$  especificada abajo, en donde  $\theta > 0$  es desconocido. Considerando la cantidad pivotal  $X_1/\theta$ , encuentre un intervalo de confianza exacto para  $\theta$ .

$$f(x, \theta) = \begin{cases} \frac{2(\theta - x)}{\theta^2} & \text{si } 0 < x < \theta, \\ 0 & \text{en otro caso.} \end{cases}$$

272. **Distribución uniforme.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{unif}(0, \theta)$ , con  $\theta > 0$  desconocido. A través de la cantidad

pivotal  $X_{(n)}/\theta$ , encuentre el intervalo de confianza exacto para  $\theta$  de longitud mínima y con nivel de confianza  $1 - \alpha$ .

273. **Distribución uniforme.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{unif}(-\theta, \theta)$ , en donde  $\theta > 0$  es desconocido. Considerando la cantidad pivotal  $\max_{1 \leq i \leq n} |X_i|/\theta$ , encuentre un intervalo de confianza exacto para  $\theta$ .

274. **Distribución exponencial.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $f(x, \theta)$  especificada abajo, en donde  $a$  es una constante conocida y  $\theta > 0$  es desconocido. Encuentre un intervalo de confianza exacto para  $\theta$ .

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-(x-a)/\theta} & \text{si } x > a, \\ 0 & \text{en otro caso.} \end{cases}$$

275. **Distribución normal.** Se quiere estimar la estatura promedio de un grupo de personas suponiendo una distribución  $N(\theta, \sigma^2)$ , en donde  $\theta$  es desconocido y  $\sigma^2 > 0$  es conocido. Se requiere un intervalo al 95% de confianza pero con una longitud de 2 cm. ¿De qué tamaño debe ser la muestra para hacer esto posible?

# Capítulo 4

## Pruebas de hipótesis

En este capítulo se presenta una breve introducción al tema de pruebas de hipótesis. Estudiaremos este tema con énfasis principalmente en la estimación de parámetros de las distribuciones de probabilidad, aunque las pruebas de hipótesis pueden aplicarse también en otras situaciones.

### 4.1. Introducción

Ilustraremos las ideas básicas de una prueba de hipótesis mediante un ejemplo sencillo e interesante. Más adelante formalizaremos los conceptos para el caso de pruebas concernientes a los parámetros de las distribuciones de probabilidad.



Figura 4.1

Consideremos una situación en la que se efectúa sólo uno de los siguientes dos experimentos aleatorios: se lanza un dado equilibrado y se registra el número obtenido, o bien se lanza una moneda cinco veces y se registra el número de cruces totales que se obtienen. Supondremos que los lados de cada moneda se denominan cara y cruz. Véase la Figura 4.1.

El problema radica en que únicamente conocemos el resultado reportado  $x$  y no conocemos el experimento aleatorio efectuado. Deseamos determinar cuál de los dos experimentos se realizó con base en el número  $x$  observado. Tenemos entonces una situación de dos hipótesis:

$$H_0 : \text{“Se lanzó el dado”} \quad vs \quad H_1 : \text{“Se lanzó la moneda”}.$$

Como única información sobre este experimento tenemos un número  $x$  dentro del conjunto  $\{0, 1, 2, 3, 4, 5, 6\}$ , y con base en él debemos decidir si se llevó a cabo un experimento o el otro. La pregunta que nos planteamos es ¿qué decisión tomar para cada valor de  $x$ ? Observemos que si el número reportado es 0, entonces con seguridad se realizó el experimento de la moneda. En cambio, si se reporta el número 6, entonces con seguridad el dado fue lanzado. ¿Qué decisión tomar para cualquier otro valor de  $x$ ? Una forma de responder esta pregunta es usando máxima verosimilitud. En la Tabla 4.1 se muestran las probabilidades de obtener los posibles valores de  $x$  bajo cada uno de los dos experimentos.

	Número $x$						
	0	1	2	3	4	5	6
(D) Dado	0	1/6	1/6	1/6	1/6	1/6	1/6
(M) Moneda	1/32	5/32	10/32	10/32	5/32	1/32	0

Tabla 4.1

Es claro que cuando se efectúa el experimento de lanzar el dado, cada uno de los resultados 1, 2, 3, 4, 5, 6 se obtiene con probabilidad 1/6. Estas probabilidades aparecen en el correspondiente renglón de la tabla. Por otro

lado, cuando se efectúa el experimento de lanzar la moneda equilibrada, la probabilidad de obtener cualquiera de los números  $x = 0, 1, 2, 3, 4, 5$  es  $\binom{5}{x}(1/2)^x(1/2)^{5-x}$ , y estas probabilidades aparecen en el último renglón de la tabla. Es intuitivamente claro que una buena estrategia consiste en decidir por el experimento que tenga mayor probabilidad de producir el valor  $x$  observado. Estas probabilidades máximas se encuentran sombreadas en la tabla. Siguiendo esta idea se llega a la siguiente regla de decisión:

**Regla de decisión**

Si  $x \in \mathcal{C} = \{0, 2, 3\}$ , se rechaza  $H_0$ ,  
en caso contrario, no se rechaza  $H_0$ .

Por razones evidentes, al conjunto  $\mathcal{C}$  se le llama región de rechazo de la hipótesis  $H_0$ . La regla de decisión anterior es razonable, sin embargo, no está libre de errores. Por ejemplo, si  $x = 2$ , se decide por el experimento de la moneda, pero el resultado bien pudo provenir del dado. Igualmente, si  $x = 1$ , se decide por el dado pero es factible que el resultado haya sido obtenido por la moneda. Para este ejemplo y para las situaciones que estudiaremos más adelante, cualquier regla de decisión no estará exenta de errores.

Los dos tipos de errores que se pueden presentar se denominan error tipo I y error tipo II, y se muestran en la Tabla 4.2.

	$H_0$ cierta	$H_0$ falsa
Rechazar $H_0$	Error tipo I	✓
No rechazar $H_0$	✓	Error tipo II

Tabla 4.2

Se usan las letras  $\alpha$  y  $\beta$  para denotar a las probabilidades de cometer los errores tipo I y II, respectivamente. Cada uno de estos errores se definen y calculan como las siguientes probabilidades condicionales:

$$\begin{aligned}\alpha &= P(\text{“Rechazar } H_0 \text{”} \mid \text{“}H_0 \text{ es verdadera”}), \\ \beta &= P(\text{“No rechazar } H_0 \text{”} \mid \text{“}H_0 \text{ es falsa”}).\end{aligned}$$

El tratamiento que se les da a estas probabilidades condicionales no es el usual de la probabilidad elemental, pues no tenemos la información para calcular la probabilidad de los eventos condicionantes. Supondremos que  $H_0$  es cierta (en el caso para calcular  $\alpha$ ) o falsa (en el caso para  $\beta$ ), y en cada situación veremos si la información supuesta es suficiente para calcular estas probabilidades.

Para el ejemplo que estamos analizando las probabilidades de estos errores se calculan de la siguiente manera: suponiendo que el número reportado es una variable aleatoria y por lo tanto la denotaremos por la letra  $X$ , que  $D$  denota el evento de lanzar el dado y  $M$  el evento de lanzar la moneda, entonces

$$\begin{aligned}\alpha &= P(\text{“Error tipo I”}) \\ &= P(\text{“Rechazar } H_0 \text{”} \mid \text{“}H_0 \text{ es verdadera”}) \\ &= P(X \in \{0, 2, 3\} \mid D) \\ &= 2/6.\end{aligned}$$

Por otro lado,

$$\begin{aligned}\beta &= P(\text{“Error tipo II”}) \\ &= P(\text{“No rechazar } H_0 \text{”} \mid \text{“}H_0 \text{ es falsa”}) \\ &= P(X \in \{1, 4, 5, 6\} \mid M) \\ &= 11/32.\end{aligned}$$

Observe que estas probabilidades no suman 1 pues los eventos condicionantes son distintos. Por otro lado, es claro que deseamos que estas probabilidades de error sean lo más pequeñas posible, sin embargo disminuir una de estas probabilidades puede aumentar la otra. Veamos algunos ejemplos para ilustrar estos posibles comportamientos. Tomaremos como referencia la región de rechazo  $\mathcal{C} = \{0, 2, 3\}$  en donde hemos obtenido que  $\alpha = 2/6$  y  $\beta = 11/32$ .

- Si se toma  $\mathcal{C} = \{0, 1, 2, 3\}$ , entonces

$$\begin{aligned}\alpha &= P(\text{“Error tipo I”}) \\ &= P(X \in \{0, 1, 2, 3\} \mid D) \\ &= 3/6.\end{aligned}$$

$$\begin{aligned}\beta &= P(\text{“Error tipo II”}) \\ &= P(X \in \{4, 5, 6\} \mid M) \\ &= 6/32.\end{aligned}$$

Observamos que  $\alpha$  aumenta y  $\beta$  disminuye. Comparativamente no podemos decir que una región de rechazo sea mejor que la otra, a menos que fijemos prioridades en los dos tipos de error.

- Si se toma  $\mathcal{C} = \{2, 3\}$ , entonces

$$\begin{aligned}\alpha &= P(\text{“Error tipo I”}) \\ &= P(X \in \{2, 3\} \mid D) \\ &= 2/6.\end{aligned}$$

$$\begin{aligned}\beta &= P(\text{“Error tipo II”}) \\ &= P(X \in \{0, 1, 4, 5, 6\} \mid M) \\ &= 12/32.\end{aligned}$$

En este caso  $\alpha$  permanece sin cambio pero  $\beta$  aumenta. Comparativamente, preferimos la primera región de rechazo.

En general, dos regiones de rechazo pueden no ser comparables desde el punto de vista de las probabilidades de error, pues un tipo de error puede ser menor para una región de rechazo y el otro tipo de error puede ser mayor. En términos generales, seguiremos el siguiente criterio para la comparación de dos regiones de rechazo, cuando sea posible:

Se fija un valor de  $\alpha$  y se busca dentro de todas las regiones de rechazo cuya probabilidad de error tipo I sea  $\alpha$ , aquella que tenga probabilidad de error tipo II más pequeña.

Estaremos entonces interesados en encontrar la mejor región de rechazo a un nivel  $\alpha$ , en el sentido especificado en el recuadro.

Por ejemplo, en la Tabla 4.3 se muestran distintas regiones de rechazo  $\mathcal{C}$  con el mismo valor  $\alpha = 2/6$  y para las cuales se ha calculado la probabilidad  $\beta$ . El renglón sombreado, y que corresponde al obtenido usando máxima verosimilitud, es la mejor región de rechazo para  $\alpha = 2/6$  pues la probabilidad  $\beta$  es la menor posible.

Región de rechazo	$\alpha$	$\beta$
$\mathcal{C} = \{0, 1, 2\}$	2/6	16/32
$\mathcal{C} = \{0, 1, 3\}$	2/6	16/32
$\mathcal{C} = \{0, 1, 4\}$	2/6	21/32
$\mathcal{C} = \{0, 1, 5\}$	2/6	25/32
$\mathcal{C} = \{0, 1, 6\}$	2/6	26/32
$\mathcal{C} = \{0, 2, 3\}$	2/6	11/32
$\mathcal{C} = \{0, 2, 4\}$	2/6	16/32
$\mathcal{C} = \{0, 2, 5\}$	2/6	20/32
$\mathcal{C} = \{0, 2, 6\}$	2/6	21/32
$\mathcal{C} = \{0, 3, 4\}$	2/6	16/32
$\mathcal{C} = \{0, 3, 5\}$	2/6	20/32
$\mathcal{C} = \{0, 3, 6\}$	2/6	21/32
$\mathcal{C} = \{0, 4, 5\}$	2/6	25/32
$\mathcal{C} = \{0, 4, 6\}$	2/6	26/32
$\mathcal{C} = \{0, 5, 6\}$	2/6	30/32

Tabla 4.3

El valor  $x = 0$  puede omitirse en cada una de las regiones de rechazo de la tabla anterior, construyendo así otras regiones de rechazo con el mismo valor  $\alpha = 2/6$ . Sin embargo, este valor omitido se traslada a la región de no rechazo y ello incrementa el valor de  $\beta$ . De esta manera, estas regiones de rechazo adicionales no son mejores y por lo tanto se han suprimido en la búsqueda de la mejor región de rechazo de tamaño  $\alpha = 2/6$ .

## Ejercicios

276. Considere nuevamente el experimento del dado y la moneda estudiado en la presente sección. Determine una región de rechazo tal que
- a)  $\alpha = 0$  y que el valor de  $\beta$  sea mínimo.
  - b)  $\beta = 0$  y que el valor de  $\alpha$  sea mínimo.
277. Considere el ejemplo del experimento en donde se lanza un dado o una moneda estudiado en esta sección. Suponga que se lanza el dado o la moneda con probabilidad  $1/2$  cada uno. Encuentre la distribución del número reportado  $X$ .
278. Considere el ejemplo del experimento en donde se lanza un dado o una moneda estudiado en esta sección. Encuentre la mejor región de rechazo con  $\alpha = 1/6$ .

## 4.2. Conceptos elementales

Formalizaremos ahora algunos de los conceptos mencionados en la sección introductoria. Estudiaremos pruebas de hipótesis principalmente en el contexto de la estimación de parámetros en las distribuciones de probabilidad. A tales pruebas se les llama pruebas paramétricas. Aquí tenemos la definición de hipótesis estadística.

**Definición 4.1** Una **hipótesis estadística**, o simplemente **hipótesis**, es una afirmación o conjetura acerca de la distribución de una o más variables aleatorias.

Particularmente las hipótesis a las que haremos mayor referencia serán afirmaciones o conjeturas acerca del valor de los parámetros de las distribuciones de probabilidad. Por ejemplo, si  $X$  es una variable aleatoria con distribución  $\text{bin}(10, p)$ , entonces la afirmación “ $p = 0.2$ ” es una hipótesis. En este caso hemos aceptado la distribución binomial para esta variable aleatoria y conjeturamos acerca del valor de uno de sus parámetros. Del mismo modo, si  $X$  es una variable aleatoria con distribución  $N(\mu, \sigma^2)$ , entonces la afirmación “ $\mu > 0$ ” es otro ejemplo de una hipótesis estadística. Muy diversas hipótesis pueden formularse acerca de una distribución de probabilidad y más adelante tendremos oportunidad de mencionar algunas de ellas.

Establecer con precisión las hipótesis a contrastar depende fuertemente del estudio que se esté llevando a cabo, de la pregunta que se desee contestar, y de la información adicional que se tenga acerca del problema particular en estudio. Nuestra perspectiva será que las hipótesis a contrastar nos son dadas, o que son evidentes de proponer de acuerdo al enunciado del problema.

La siguiente definición establece una clasificación de dos tipos generales de hipótesis que pueden considerarse relativas a la especificación de los parámetros de una distribución.

**Definición 4.2** Una hipótesis es **simple** si especifica por completo la distribución de probabilidad en cuestión, en caso contrario, la hipótesis se llama **compuesta**.

Por ejemplo, si  $X$  es una variable aleatoria con distribución  $\exp(\lambda)$ , entonces la afirmación “ $\lambda = 5$ ” es una hipótesis simple. Si  $X$  tiene distribución  $N(\mu, 1)$ , entonces la afirmación “ $\mu = 0$ ” es otro ejemplo de hipótesis simple. En cambio, si  $X$  tiene distribución  $\text{Poisson}(\lambda)$ , entonces “ $\lambda > 20$ ” es una hipótesis compuesta, pues no se especifica completamente la distribución de la variable aleatoria. Si  $X$  tiene distribución  $\chi^2(n)$ , entonces “ $n \neq 5$ ” es otro ejemplo de una hipótesis compuesta.

En general, contrastaremos dos hipótesis de acuerdo al siguiente esquema y notación.

$$H_0 : (\text{hipótesis nula}) \quad vs \quad H_1 : (\text{hipótesis alternativa}).$$

Esto es, a la hipótesis que aparezca del lado izquierdo le llamaremos hipótesis nula y la denotaremos por  $H_0$ . A la hipótesis que aparezca en el lado derecho le llamaremos hipótesis alternativa y la denotaremos por  $H_1$ . Tanto la hipótesis nula  $H_0$  como la hipótesis alternativa  $H_1$  pueden ser simples o compuestas. De este modo tenemos cuatro diferentes contrastes de tipos de hipótesis: simple vs simple, simple vs compuesta, compuesta vs simple, y compuesta vs compuesta. Las tres últimas son más difíciles de analizar.

Llevar a cabo una prueba de hipótesis significa aplicar una regla para decidir si se acepta la hipótesis nula o se rechaza en favor de la hipótesis alternativa. La información para obtener una regla de decisión que nos lleve a rechazar o no rechazar una hipótesis estadística provendrá de una muestra aleatoria de la distribución en estudio. Por otro lado, al aceptar una hipótesis no se afirma que ésta sea absolutamente cierta, sino simplemente que es consistente con los datos de la muestra aleatoria y la regla de decisión adoptada. Si la información de la muestra o la regla de decisión cambia, muy posiblemente también cambie la decisión de rechazar o no rechazar.

La regla para decidir si se acepta la hipótesis nula o se rechaza en favor de la hipótesis alternativa se expresa en términos de un conjunto llamado región de rechazo. Este conjunto consta de aquellos valores de la muestra aleatoria para los cuales se ha acordado rechazar la hipótesis nula. Es claro que existen tantas regiones de rechazo como subconjuntos de valores de la muestra aleatoria.

**Definición 4.3** Una **región de rechazo** es un subconjunto de valores de una muestra aleatoria para los cuales se rechaza la hipótesis nula. A una región de rechazo se le llama también **región crítica**.

Desde el punto de vista matemático, uno de los problemas principales en las pruebas de hipótesis es el de construir de manera justificada una región de rechazo. Con base en la región de rechazo encontrada se puede entonces llevar a cabo el proceso de decisión de rechazar o no rechazar la hipótesis nula.

Como hemos mencionado antes, al tomar una decisión en una prueba de hipótesis, existe siempre el riesgo de cometer errores. Los dos tipos de errores que pueden surgir se formalizan en las siguientes dos definiciones.

**Definición 4.4** El **error tipo I** se comete cuando se rechaza la hipótesis nula  $H_0$  cuando ésta es verdadera. A la probabilidad de cometer el error tipo I se le denota por la letra  $\alpha$ , y se calcula mediante la siguiente probabilidad condicional:

$$\begin{aligned}\alpha &= P(\text{“Error tipo I”}) \\ &= P(\text{“Rechazar } H_0 \text{”} \mid \text{“} H_0 \text{ es verdadera”}).\end{aligned}$$

A este valor  $\alpha$  se le conoce también como el **tamaño de la región crítica**, el **tamaño de la región de rechazo**, o bien como el **nivel de significancia de la prueba**.

**Definición 4.5** El **error tipo II** se comete cuando no se rechaza la hipótesis nula  $H_0$  cuando ésta es falsa. A la probabilidad de cometer el error tipo II se le denota por la letra  $\beta$ , y se calcula mediante la siguiente probabilidad condicional:

$$\begin{aligned}\beta &= P(\text{“Error tipo II”}) \\ &= P(\text{“No rechazar } H_0 \text{”} \mid \text{“} H_0 \text{ es falsa”}).\end{aligned}$$

Las probabilidades  $\alpha$  y  $\beta$  arriba definidas no son complementarias, es decir, no necesariamente suman 1, pues los eventos condicionantes que aparecen en las probabilidades condicionales anteriores son distintos. Es claro que

deseamos que estas probabilidades tomen valores pequeños. Sin embargo, al solicitar que una de estas probabilidades sea pequeña la otra puede aumentar, así es que puede no ser posible hacer ambas probabilidades tan pequeñas como uno desee. Cuando sea posible, procederemos de la siguiente forma: fijaremos un valor para  $\alpha$  y buscaremos aquella posible región de rechazo de tamaño  $\alpha$  que tenga probabilidad  $\beta$  más pequeña. De esta manera se le da mayor importancia al error tipo I pues se controla su probabilidad de ocurrencia.

Observemos que si  $H_0$  es una hipótesis simple, entonces la distribución de probabilidad en estudio queda completamente especificada y la probabilidad  $\alpha$  podrá ser calculada de manera exacta, aunque en ocasiones puede usarse una aproximación con el fin de dar una expresión corta para esta cantidad. En cambio, si  $H_0$  es una hipótesis compuesta, entonces no podrá calcularse  $\alpha$  pues en tales situaciones se desconoce el valor exacto del parámetro o parámetros en estudio. La misma situación ocurre para  $\beta$  cuando  $H_1$  es simple o compuesta, únicamente en el caso cuando  $H_1$  es simple se puede calcular el valor de  $\beta$  de manera exacta.

Suponiendo el caso del contraste de dos hipótesis simples, un problema consiste en considerar todas las posibles regiones de rechazo de tamaño  $\alpha$  y encontrar aquella que tenga probabilidad  $\beta$  más pequeña. Es claro que estamos interesados en encontrar este tipo de regiones de rechazo óptimas y la solución a este problema es el contenido del así llamado lema de Neyman-Pearson que estudiaremos más adelante.

## Ejercicios

279. Suponga que se tiene una moneda en donde la probabilidad de obtener una de las caras es un parámetro desconocido  $\theta$ , aunque se conoce que sólo puede haber dos casos:  $\theta = 1/2$  ó  $\theta = 7/12$ . Con base en los resultados  $x_1, \dots, x_n$  de  $n$  lanzamientos de la moneda se desea llevar a cabo la prueba de hipótesis

$$H_0 : \theta = 1/2 \quad vs \quad H_1 : \theta = 7/12,$$

en donde se ha convenido en definir la región de rechazo como

$$\mathcal{C} = \{(x_1, \dots, x_n) : \bar{x} \leq 13/14\}.$$

Use el teorema central del límite para aproximar las probabilidades de los errores tipo I y II.

280. Sea  $X_1, X_2, X_3$  una muestra aleatoria de tamaño  $n = 3$  de la distribución  $N(\theta, 4)$ , en donde  $\theta$  es desconocido. Encuentre las probabilidades de cometer los errores tipo I y II para la prueba

$$H_0 : \theta = 2 \quad vs \quad H_1 : \theta = 5,$$

considerando como región de rechazo

- a)  $\mathcal{C} = \{(x_1, x_2, x_3) : x_1 > 4.7\}$ .
- b)  $\mathcal{C} = \{(x_1, x_2, x_3) : (x_1 + 2x_2)/3 > 4.5\}$ .
- c)  $\mathcal{C} = \{(x_1, x_2, x_3) : (x_1 + x_3)/2 > 4.2\}$ .
- d)  $\mathcal{C} = \{(x_1, x_2, x_3) : \bar{x} > 4.1\}$ .

### 4.3. Función potencia

Establecida una región de rechazo para una prueba de hipótesis, la función potencia se define como la probabilidad de rechazar la hipótesis nula  $H_0$  para cada posible valor del parámetro  $\theta$ . Esto es lo que significa la notación un tanto ambigua de la probabilidad condicional que aparece en la siguiente definición.

**Definición 4.6** Suponiendo dada una región de rechazo, la **función potencia** de una prueba de hipótesis sobre un parámetro desconocido  $\theta$  es la función

$$\theta \mapsto \pi(\theta) = P(\text{“Rechazar } H_0\text{”} \mid \theta).$$

Por lo tanto, la función potencia está definida en cada punto del espacio parametral correspondiente. Como veremos más adelante, esta función puede

ser útil para comparar dos regiones de rechazo. Cuando se contrastan dos hipótesis simples  $H_0 : \theta = \theta_0$  vs  $H_1 : \theta = \theta_1$ , las dos probabilidades de error se pueden expresar en términos de la función potencia como sigue

$$\begin{aligned}\alpha &= \pi(\theta_0), \\ \beta &= 1 - \pi(\theta_1).\end{aligned}$$

Veamos un ejemplo del cálculo de la función potencia.

**Ejemplo 4.1** Consideremos el contraste de hipótesis simples

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta = \theta_1,$$

en donde  $\theta_0 < \theta_1$  son dos valores fijos del parámetro desconocido  $\theta$  de una distribución Bernoulli. Debido a que la media de una muestra aleatoria de esta distribución se acerca al valor del parámetro cuando el tamaño de la muestra crece a infinito, convengamos en adoptar como región de rechazo el conjunto

$$\mathcal{C} = \{(x_1, \dots, x_n) : \bar{x} \geq (\theta_0 + \theta_1)/2\},$$

en donde  $(\theta_0 + \theta_1)/2$  es el punto medio del intervalo con extremo izquierdo  $\theta_0$  y extremo derecho  $\theta_1$ . De esta manera, si la media muestral se separa de  $\theta_0$  hacia la derecha a partir del punto  $(\theta_0 + \theta_1)/2$  en adelante, se rechaza  $H_0$  y se prefiere  $H_1$ . Véase la Figura 4.2, en donde hemos llamado también región de rechazo al conjunto de valores de  $\bar{x}$  tales que  $\bar{x} \geq (\theta_0 + \theta_1)/2$ .

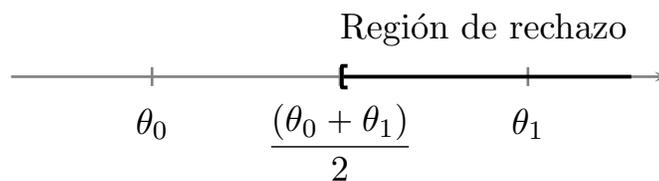


Figura 4.2

Teniendo definida esta región de rechazo, podemos ahora calcular de manera aproximada las probabilidades de los errores tipo I y II, y más generalmente la función potencia de la siguiente forma: por el teorema central del límite,

$$\begin{aligned}\alpha &= P(\bar{X} \geq (\theta_0 + \theta_1)/2 \mid \theta = \theta_0) \\ &= P\left(\frac{\bar{X} - \theta_0}{\sqrt{\theta_0(1 - \theta_0)/n}} \geq \frac{(\theta_1 - \theta_0)/2}{\sqrt{\theta_0(1 - \theta_0)/n}} \mid \theta = \theta_0\right) \\ &\approx 1 - \Phi\left(\frac{(\theta_1 - \theta_0)/2}{\sqrt{\theta_0(1 - \theta_0)/n}}\right).\end{aligned}$$

$$\begin{aligned}\beta &= P(\bar{X} < (\theta_0 + \theta_1)/2 \mid \theta = \theta_1) \\ &= P\left(\frac{\bar{X} - \theta_1}{\sqrt{\theta_1(1 - \theta_1)/n}} < \frac{(\theta_0 - \theta_1)/2}{\sqrt{\theta_1(1 - \theta_1)/n}} \mid \theta = \theta_1\right) \\ &\approx \Phi\left(\frac{(\theta_0 - \theta_1)/2}{\sqrt{\theta_1(1 - \theta_1)/n}}\right).\end{aligned}$$

Recordemos que los valores de  $\theta_0$ ,  $\theta_1$  y  $n$  son conocidos y, por lo tanto, las cantidades anteriores pueden calcularse explícitamente. Usando nuevamente el teorema central del límite, la función potencia se puede aproximar de la siguiente manera: para  $\theta \in (0, 1)$ ,

$$\begin{aligned}\pi(\theta) &= P(\bar{X} \geq (\theta_0 + \theta_1)/2 \mid \theta) \\ &\approx 1 - \Phi\left(\frac{(\theta_0 + \theta_1)/2 - \theta}{\sqrt{\theta(1 - \theta)/n}}\right).\end{aligned}$$

Haciendo un análisis cualitativo del comportamiento de esta función para valores de  $\theta$  cercanos a cero y a uno, se puede comprobar que la gráfica de esta función es la curva creciente que se muestra en la Figura 4.3.

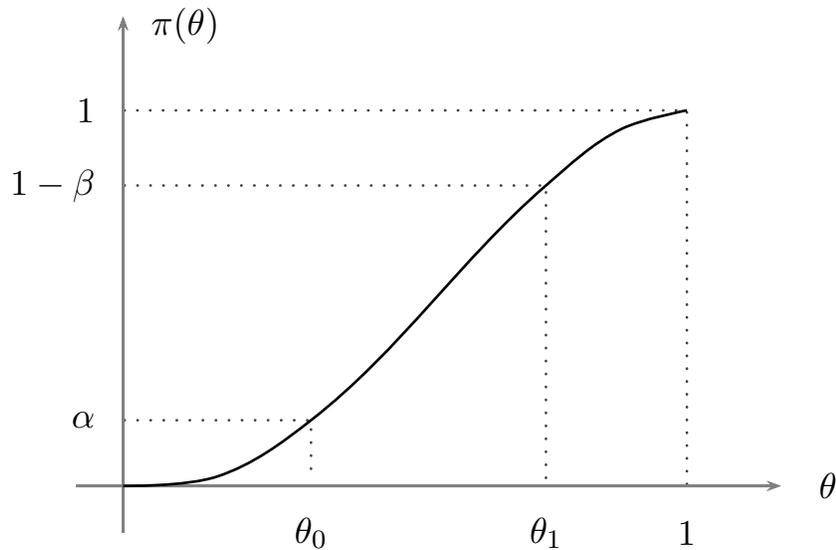


Figura 4.3

Esta es la función potencia asociada a la región de rechazo especificada. Esta función toma el valor  $\alpha$  cuando  $\theta = \theta_0$ , y toma el valor  $1 - \beta$  en  $\theta = \theta_1$ . En general, para valores de  $\theta$  alrededor de  $\theta_0$  la probabilidad de rechazar la hipótesis nula es baja. En cambio, para valores de  $\theta$  cercanos a  $\theta_1$  la probabilidad de rechazar es alta dando preferencia a este valor del parámetro. ■

## Ejercicios

281. **Distribución exponencial.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\exp(\theta)$ , en donde el parámetro  $\theta$  es desconocido. Sea  $\theta_0 > 0$  un valor particular fijo y conocido de  $\theta$  y considere la prueba

$$H_0 : \theta \leq \theta_0 \quad \text{vs} \quad H_1 : \theta > \theta_0.$$

Defina la región de rechazo

$$\mathcal{C} = \{(x_1, \dots, x_n) : \bar{x} > 1/\theta_0\}.$$

a) Calcule y grafique la función potencia de esta prueba.

b) Calcule  $\sup_{\theta \in (0, \theta_0]} \pi(\theta)$ .

c) Calcule  $\sup_{\theta \in (\theta_0, \infty)} (1 - \pi(\theta))$ .

282. **Distribución normal.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $N(\theta, \sigma^2)$ , en donde la media  $\theta$  es desconocida y la varianza  $\sigma^2$  es conocida. Sea  $\theta_0$  un valor particular fijo y conocido de  $\theta$  y considere la prueba

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta \neq \theta_0.$$

Defina la región de rechazo

$$\mathcal{C} = \{(x_1, \dots, x_n) : |\bar{x} - \theta_0| \geq c\},$$

en donde  $c$  es una constante. Encuentre el valor de la constante  $c$  de tal manera que esta región de rechazo sea de un tamaño  $\alpha$  prefijado. Calcule y grafique la función potencia de esta prueba.

#### 4.4. Ejemplo de una prueba paramétrica

En esta sección se desarrolla otro ejemplo en donde se ilustran los conceptos generales y el procedimiento para llevar a cabo una prueba de hipótesis. Esta vez nos referiremos a hipótesis relativas al valor de un parámetro de una distribución.

Suponga que tenemos una moneda y que necesitamos tomar una decisión respecto de si la moneda está o no está equilibrada. Para llegar a alguna conclusión lanzamos la moneda  $n$  veces y con base en los resultados obtenidos decidimos si los lados de la moneda, que llamaremos cara y cruz, tienen la misma probabilidad de ocurrir, o no tienen.



Figura 4.4

Por ejemplo, si de cien lanzamientos se obtienen cincuenta cruces, entonces uno tiende a pensar que esto puede ser una evidencia para creer que la moneda está equilibrada, aunque debe tenerse presente que tal resultado puede también obtenerse con una moneda no equilibrada. Pero, ¿qué decisión tomar si únicamente se obtienen 45 cruces? ¿y si se obtienen 60 cruces? Es claro que en estos últimos casos la decisión no es tan evidente. Vamos a plantear y resolver este problema de decisión a través del contraste de dos hipótesis.

Denotemos por  $X_1, \dots, X_n$  los resultados de  $n$  lanzamientos de la moneda. Convengamos en definir

$$X_i = \begin{cases} 1 & \text{si la moneda cae cruz,} \\ 0 & \text{si la moneda cae cara.} \end{cases}$$

Es decir, cada una de estas variables tiene distribución Bernoulli de parámetro  $\theta$ , en donde este parámetro es la probabilidad desconocida de obtener cruz en cada lanzamiento. Supondremos la hipótesis de independencia para esta colección de variables aleatorias. El problema planteado se formaliza y se traduce en llevar a cabo la prueba de hipótesis

$$H_0 : \theta = 1/2 \quad \text{vs} \quad H_1 : \theta \neq 1/2.$$

Estamos ante una situación de una hipótesis simple contra una hipótesis compuesta. Construiremos a continuación una región de rechazo para esta prueba. Por la ley de los grandes números, la media muestral  $\bar{X}$  se acerca al verdadero valor de  $\theta$  cuando el número de lanzamientos es cada vez más grande, y por lo tanto, es una aproximación para el valor desconocido de  $\theta$ . Cuando  $\bar{X}$  diste mucho de  $1/2$  es razonable pensar que la moneda no

está equilibrada. Es por ello que se propone rechazar la hipótesis  $H_0$  cuando  $|\bar{X} - 1/2| \geq c$ , para algún número  $c$  que encontraremos más adelante, y esto lo haremos estableciendo un valor particular para la probabilidad del error tipo I. Es decir, se propone como región de rechazo al conjunto

$$\mathcal{C} = \{(x_1, \dots, x_n) : |\bar{x} - 1/2| \geq c\}. \quad (4.1)$$

En el caso cuando  $H_0$  es cierta, pero se toma la decisión de rechazar  $H_0$ , se está en la situación de cometer el error tipo I, y establecemos que la probabilidad de que ello ocurra es  $\alpha$ , es decir,

$$P(|\bar{X} - 1/2| \geq c | \theta = 1/2) = \alpha.$$

A partir de esta ecuación se puede encontrar el valor de  $c$  de la siguiente manera: cuando  $H_0$  es verdadera, es decir, cuando  $\theta = 1/2$ , la media muestral tiene distribución aproximada normal de media  $1/2$  y varianza  $(1/2)(1 - 1/2)/n = 1/(4n)$ , y por lo tanto, de manera aproximada,

$$\frac{\bar{X} - 1/2}{1/(2\sqrt{n})} \sim N(0, 1).$$

En consecuencia,

$$\begin{aligned} \alpha &= P(|\bar{X} - 1/2| \geq c | \theta = 1/2) \\ &= 1 - P(|\bar{X} - 1/2| < c | \theta = 1/2) \\ &= 1 - P(-c < \bar{X} - 1/2 < c | \theta = 1/2) \\ &= 1 - P\left(\frac{-c}{1/(2\sqrt{n})} < \frac{\bar{X} - 1/2}{1/(2\sqrt{n})} < \frac{c}{1/(2\sqrt{n})} \mid \theta = 1/2\right) \\ &\approx 2\left(1 - \Phi\left(\frac{c}{1/(2\sqrt{n})}\right)\right). \end{aligned}$$

De donde se obtiene que

$$c \approx \frac{1}{2\sqrt{n}} \Phi^{-1}(1 - \alpha/2).$$

Así, dado un valor para el tamaño de muestra  $n$  y un valor convenido para  $\alpha$ , este es el valor de la constante  $c$  que hace que la región de rechazo (4.1)

sea de tamaño  $\alpha$ . Por ejemplo, para  $n = 100$  y  $\alpha = 0.01$ , puede comprobarse que  $c = 0.128$ . Observemos por otro lado que, cuando  $n$  crece, la constante  $c$  disminuye y ello hace que el área de la región de rechazo disminuya, y por lo tanto la probabilidad  $\alpha$  disminuye. Como función de la probabilidad  $\alpha$ , la constante  $c$  se comporta como se muestra en la Figura 4.5.

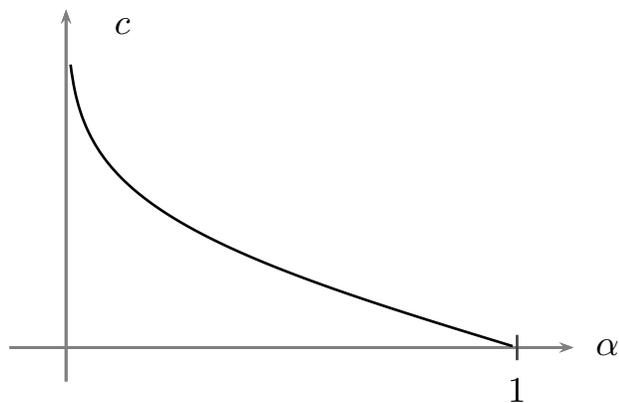


Figura 4.5

De esta forma la estadística de la prueba es la variable aleatoria  $\bar{X}$  y su valor determina el resultado de la prueba: si  $|\bar{X} - 1/2| \geq 0.128$ , se rechaza la hipótesis  $H_0$ . Esta región de rechazo, referida a  $\bar{X}$ , se puede escribir como la unión de intervalos

$$[0, 1/2 - c] \cup [1/2 + c, 1],$$

lo que se muestra gráficamente en la Figura 4.6.

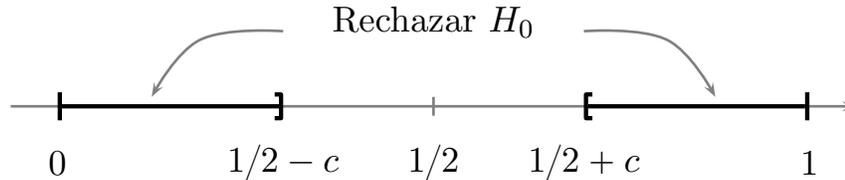


Figura 4.6

De esta manera, si  $\bar{X}$  es menor o igual a  $1/2 - c$ , o mayor o igual a  $1/2 + c$ , decidimos que la diferencia entre  $\bar{X}$  y  $1/2$  no es debido a fluctuaciones azarosas, sino que se debe a que la moneda no está equilibrada y por lo tanto rechazamos  $H_0$ . La probabilidad de un error al tomar tal decisión es  $\alpha$ , de modo que se está tomando un riesgo del  $100\alpha\%$  de clasificar una moneda equilibrada como no equilibrada.

De manera equivalente podemos expresar la región de rechazo en términos del número de cruces obtenidos de la siguiente manera. Considerando nuevamente  $n = 100$  y  $\alpha = 0.01$ , al multiplicar por este valor de  $n$  la condición de rechazo  $|\bar{X} - 1/2| \geq 0.128$  se obtiene la condición equivalente  $|\sum_{i=1}^{100} X_i - 50| \geq 12.8$ , en donde la suma indicada corresponde al número de cruces obtenidos. Es decir, cuando el número de cruces difiera de 50 en más de 12.8, rechazamos la hipótesis nula. Los valores mayores a 50 que satisfacen esta condición son: 63, 64, ..., 100, mientras que los valores menores a 50 son: 37, 36, ..., 0.

Por otro lado, observemos que no podemos calcular la probabilidad del error tipo II pues la hipótesis alternativa es compuesta. Específicamente esta probabilidad es

$$\beta = P(|\bar{X} - \theta| < c \mid \theta \neq 1/2).$$

La imposibilidad de calcular esta probabilidad radica en que el valor de  $\theta$  no está determinado y, en consecuencia, la distribución de  $\bar{X}$  no está plenamente especificada. Una manera parcial de calcular la probabilidad

del error tipo II es considerarla como una función del parámetro, esto es, se fija un valor  $\theta_1$  distinto de  $1/2$  y se calcula la siguiente cantidad

$$\begin{aligned} \beta(\theta_1) &= P(|\bar{X} - \theta| < c \mid \theta = \theta_1) \\ &= P(-c < \bar{X} - \theta < c \mid \theta = \theta_1) \\ &= P\left(\frac{-c}{\sqrt{\theta_1(1-\theta_1)/n}} < \frac{\bar{X} - \theta_1}{\sqrt{\theta_1(1-\theta_1)/n}} < \frac{c}{\sqrt{\theta_1(1-\theta_1)/n}} \mid \theta = \theta_1\right) \\ &\approx \Phi\left(\frac{c}{\sqrt{\theta_1(1-\theta_1)/n}}\right) - \Phi\left(\frac{-c}{\sqrt{\theta_1(1-\theta_1)/n}}\right). \end{aligned}$$

La gráfica de esta función se muestra en la Figura 4.7. Conforme el valor de  $\theta_1$  se aleja de  $1/2$ , la probabilidad de equivocarse del tipo II disminuye.

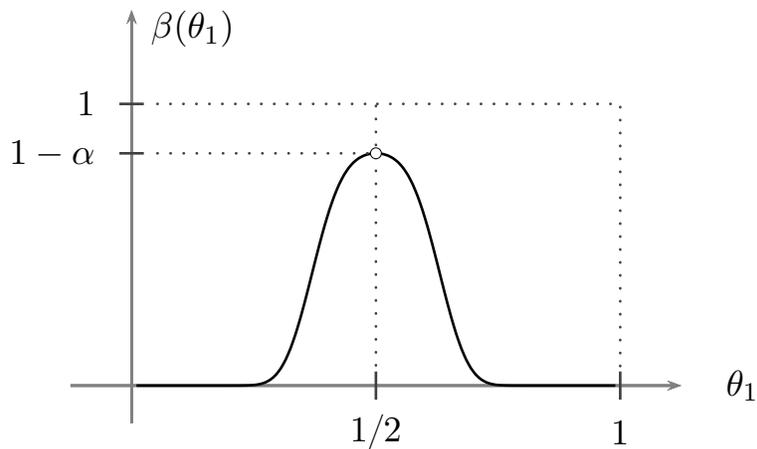


Figura 4.7

Observe que hemos aplicado nuevamente el teorema central del límite para obtener una aproximación de esta probabilidad. De esta forma el error tipo II queda expresado como una función del valor de  $\theta_1$  distinto de  $1/2$ . La función potencia de esta prueba se muestra en la Figura 4.8.

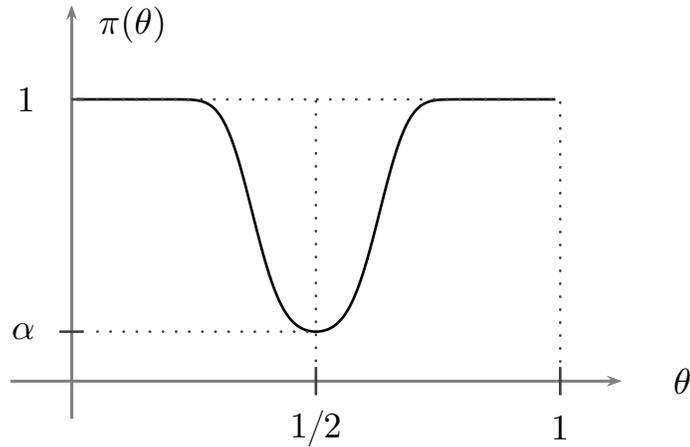


Figura 4.8

Más adelante consideraremos el caso sencillo cuando tenemos dos hipótesis simples para un parámetro  $H_0 : \theta = \theta_0$  contra  $H_1 : \theta = \theta_1$ , y el problema que estudiaremos será el de encontrar una región de rechazo que sea de tamaño predeterminado  $\alpha$  y cuya probabilidad de error tipo II sea mínima.

## 4.5. Algunas pruebas sobre la distribución normal

En esta sección estudiaremos varias pruebas de hipótesis para los parámetros de la distribución normal. Entre otras cantidades, los resultados quedarán expresados en términos de algún valor  $z_\alpha$ , el cual se define como aquel número real tal que  $P(Z > z_\alpha) = \alpha$ , en donde  $\alpha \in (0, 1)$  y  $Z$  denota una variable aleatoria con distribución  $N(0, 1)$ . En palabras,  $z_\alpha$  denota aquel número que acumula a la derecha una probabilidad igual a  $\alpha$  en la distribución normal estándar. También será útil recordar que a la función de distribución normal estándar la denotaremos por  $\Phi(x)$ . Esto es,  $\Phi(x) = P(Z \leq x)$ , definida para cualquier número real  $x$ . Así, por ejemplo, tenemos que  $\Phi(z_\alpha) = 1 - \alpha$ .

### Pruebas para la media con varianza conocida

Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $N(\theta, \sigma^2)$ , en donde la media  $\theta$  es desconocida y consideremos que  $\sigma^2$  es conocida. Entonces la media muestral  $\bar{X}$  tiene distribución  $N(\theta, \sigma^2/n)$ . Por lo tanto,

$$\frac{\bar{X} - \theta}{\sigma/\sqrt{n}} \sim N(0, 1).$$

Las pruebas que consideraremos hacen referencia a un valor particular  $\theta_0$  del parámetro desconocido  $\theta$ .

- **Prueba de dos colas.** Deseamos contrastar las hipótesis

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta \neq \theta_0.$$

El problema es encontrar una regla para decidir cuándo rechazar  $H_0$  en favor de  $H_1$ , con base en los datos de la muestra aleatoria. Cuando  $H_0$  es cierta, esto es, cuando  $\theta$  es efectivamente  $\theta_0$ , tenemos que  $\bar{X} \sim N(\theta_0, \sigma^2/n)$ , y por lo tanto,

$$Z_0 = \frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} \sim N(0, 1).$$

La estadística  $Z_0$  es claramente una medida de la distancia entre  $\bar{X}$  (un estimador de  $\theta$ ), y su valor esperado  $\theta_0$  cuando  $H_0$  es cierta. Es entonces razonable rechazar  $H_0$  cuando la variable  $Z_0$  sea grande. Esta es la razón por la que tomamos como criterio de decisión rechazar  $H_0$  cuando  $|Z_0| \geq c$ , para una cierta constante  $c$ . ¿Cómo encontramos el número  $c$ ? En una tabla de la distribución normal podemos encontrar un valor  $z_{\alpha/2}$  tal que  $P(|Z| \geq z_{\alpha/2}) = \alpha$ , para un valor de  $\alpha$  preestablecido. Véase la Figura 4.9. Este valor  $z_{\alpha/2}$  es precisamente la constante  $c$  buscada pues con ello se logra que la región de rechazo sea de tamaño  $\alpha$ .

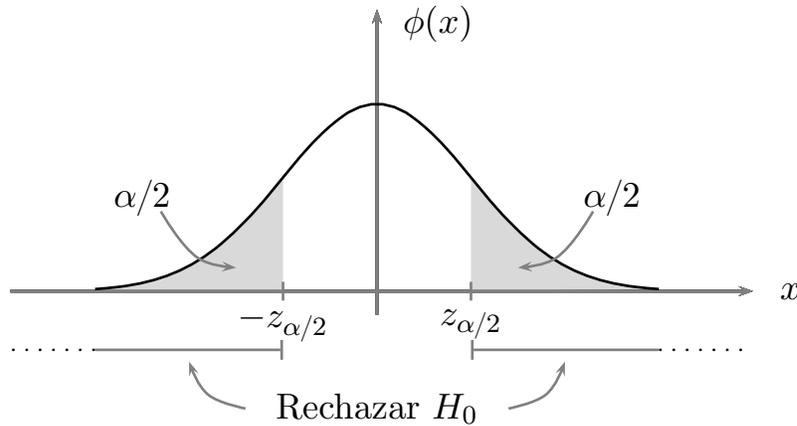


Figura 4.9

A la variable aleatoria  $Z_0$  se le llama la estadística de la prueba, y la prueba se denomina prueba de dos colas pues la región de rechazo consta de las dos colas de la distribución normal estándar que se muestran en la Figura 4.9. Llevar a cabo esta prueba de hipótesis consiste simplemente en usar los datos de la muestra para encontrar el valor de  $Z_0$ . Si resulta que  $|Z_0| \geq z_{\alpha/2}$ , entonces se rechaza  $H_0$ , en caso contrario no se rechaza  $H_0$ . Es decir, la región de rechazo de tamaño  $\alpha$  es

$$\mathcal{C} = \{(x_1, \dots, x_n) : |\frac{\bar{x} - \theta_0}{\sigma/\sqrt{n}}| \geq z_{\alpha/2}\}.$$

Puede comprobarse que la probabilidad de no rechazar la hipótesis nula cuando  $\theta = \theta_1$  (un valor distinto de  $\theta_0$ ) es

$$\begin{aligned} \beta(\theta_1) &= P(|Z_0| < z_{\alpha/2} \mid \theta = \theta_1) \\ &= \Phi(z_{\alpha/2} + \frac{\theta_0 - \theta_1}{\sigma/\sqrt{n}}) - \Phi(-z_{\alpha/2} + \frac{\theta_0 - \theta_1}{\sigma/\sqrt{n}}). \end{aligned}$$

- **Prueba de cola izquierda.** Ahora consideremos la prueba

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta < \theta_0.$$

A esta prueba se le llama prueba de cola inferior, pues la región de rechazo consta de la cola izquierda de la distribución normal estándar

como se muestra en la Figura 4.10. El procedimiento de construcción de la región de rechazo es similar a la prueba de dos colas. Ahora se rechaza la hipótesis  $H_0$  sólo cuando  $\bar{X}$  toma un valor muy a la izquierda de  $\theta_0$ . Así, tenemos nuevamente que, cuando  $H_0$  es cierta,

$$Z_0 = \frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} \sim N(0, 1),$$

y se rechaza la hipótesis  $H_0$  si  $Z_0 \leq c$ , para alguna constante negativa  $c$ . Como se desea que esta región de rechazo tenga tamaño  $\alpha$ , la constante  $c$  debe ser igual a  $-z_\alpha$ . Esto se ilustra en la Figura 4.10.

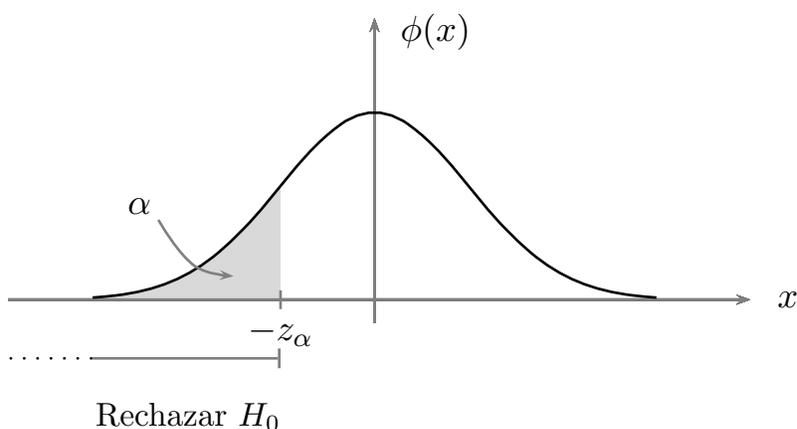


Figura 4.10

Así, la región de rechazo de tamaño  $\alpha$  es

$$\mathcal{C} = \{(x_1, \dots, x_n) : \frac{\bar{x} - \theta_0}{\sigma/\sqrt{n}} \leq -z_\alpha\}.$$

Puede comprobarse que la probabilidad de no rechazar la hipótesis nula cuando  $\theta = \theta_1$  (con  $\theta_1 < \theta_0$ ) es

$$\begin{aligned} \beta(\theta_1) &= P(Z_0 > -z_\alpha \mid \theta = \theta_1) \\ &= 1 - \Phi\left(-z_\alpha + \frac{\theta_0 - \theta_1}{\sigma/\sqrt{n}}\right). \end{aligned}$$

- **Prueba de cola derecha.** Consideremos ahora la prueba

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta > \theta_0,$$

llamada prueba de cola superior. Se rechaza la hipótesis  $H_0$  cuando  $\bar{X}$  toma un valor muy a la derecha de  $\theta_0$ . Tenemos nuevamente que, cuando  $H_0$  es cierta,

$$Z_0 = \frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} \sim N(0, 1),$$

y se rechaza la hipótesis  $H_0$  si  $Z_0 \geq c$ , para alguna constante positiva  $c$ . Como se desea que esta región de rechazo tenga tamaño  $\alpha$ , la constante  $c$  debe ser igual al cuantil  $z_\alpha$ . Es decir, la región de rechazo de tamaño  $\alpha$  es

$$\mathcal{C} = \{(x_1, \dots, x_n) : \frac{\bar{x} - \theta_0}{\sigma/\sqrt{n}} \geq z_\alpha\}.$$

Esto se ilustra en la Figura 4.11.

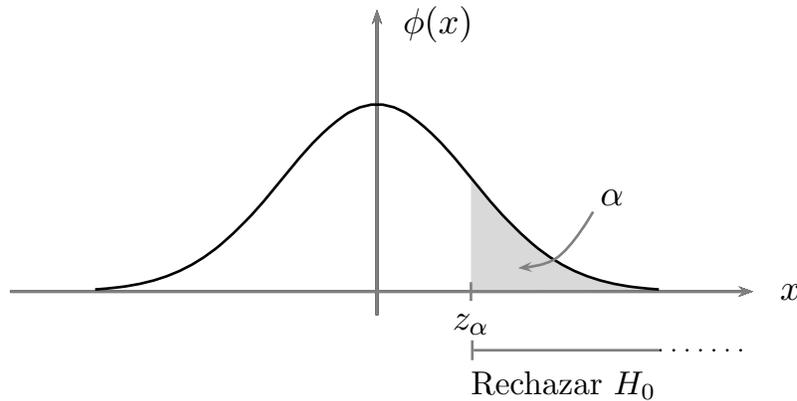


Figura 4.11

Nuevamente, es inmediato comprobar que la probabilidad de no rechazar la hipótesis nula cuando  $\theta = \theta_1$  (con  $\theta_1 > \theta_0$ ) es

$$\begin{aligned} \beta(\theta_1) &= P(Z < z_\alpha \mid \theta = \theta_1) \\ &= \Phi\left(z_\alpha + \frac{\theta_0 - \theta_1}{\sigma/\sqrt{n}}\right). \end{aligned}$$

### Pruebas para la media con varianza desconocida

Consideremos nuevamente una muestra aleatoria  $X_1, \dots, X_n$  de la distribución normal con media desconocida  $\theta$ , pero ahora con varianza desconocida  $\sigma^2$ . Nos interesa encontrar una regla de decisión para llevar a cabo ciertas pruebas de hipótesis sobre el valor desconocido del parámetro  $\theta$ . El procedimiento es muy similar al caso cuando  $\sigma^2$  es conocida y el resultado teórico que es de utilidad aquí es que

$$\frac{\bar{X} - \theta}{S/\sqrt{n}} \sim t(n-1),$$

en donde  $S$  es la varianza muestral. Como en el caso anterior, las pruebas que consideraremos hacen referencia a un valor particular  $\theta_0$  del parámetro desconocido  $\theta$ .

- **Prueba de dos colas.** Consideremos la prueba

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta \neq \theta_0.$$

Es razonable rechazar  $H_0$  cuando la diferencia entre la media muestral  $\bar{X}$  y el valor  $\theta_0$  es grande, es decir, cuando  $|\bar{X} - \theta_0| \geq c$ , para alguna constante  $c$ . Así, es de utilidad saber que

$$T_0 = \frac{\bar{X} - \theta_0}{S/\sqrt{n}} \sim t(n-1),$$

pues esta variable aleatoria es una medida de la distancia entre  $\bar{X}$  y  $\theta_0$ . Además, si  $t_{\alpha/2, n-1}$  denota el número real tal que el área bajo la función de densidad de la distribución  $t(n-1)$  a la derecha de ese valor es  $\alpha/2$ , entonces una región de rechazo de tamaño  $\alpha$  para esta prueba es

$$\mathcal{C} = \{(x_1, \dots, x_n) : \left| \frac{\bar{x} - \theta_0}{s/\sqrt{n}} \right| \geq t_{\alpha/2, n-1}\}.$$

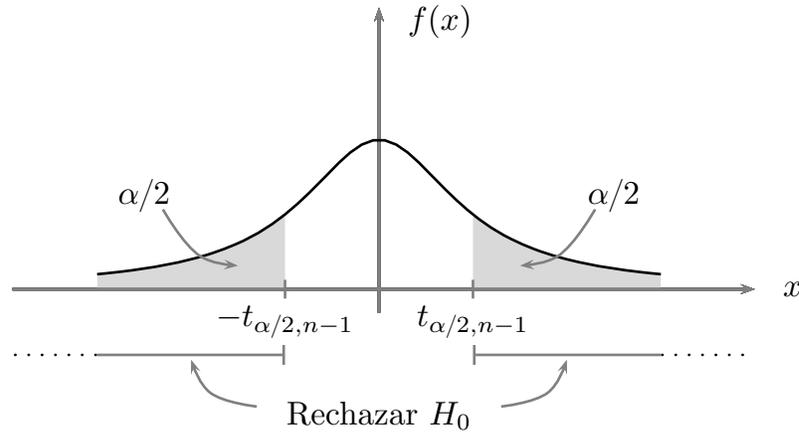


Figura 4.12

Véase la Figura 4.12. Respecto del error tipo II para esta región de rechazo, puede comprobarse que la probabilidad de no rechazar  $H_0$  cuando  $\theta = \theta_1$  (con  $\theta_1$  distinto de  $\theta_0$ ) es

$$\begin{aligned} \beta(\theta_1) &= P(|T_0| < t_{\alpha/2, n-1} \mid \theta = \theta_1) \\ &\approx F\left(t_{\alpha/2, n-1} + \frac{\theta_0 - \theta_1}{s/\sqrt{n}}\right) - F\left(-t_{\alpha/2, n-1} + \frac{\theta_0 - \theta_1}{s/\sqrt{n}}\right), \end{aligned}$$

en donde  $F$  es la función de distribución  $t$  con  $n - 1$  grados de libertad. Observe que la última expresión es una aproximación pues se ha substituido la desviación estándar muestral  $S$  por su valor observado  $s$ .

- **Prueba de cola izquierda.** La prueba  $H_0 : \theta = \theta_0$  vs  $H_1 : \theta < \theta_0$  se llama nuevamente prueba de cola inferior y una región de rechazo de tamaño prefijado  $\alpha$  está dada por

$$\mathcal{C} = \left\{ (x_1, \dots, x_n) : \frac{\bar{x} - \theta_0}{s/\sqrt{n}} \leq t_{\alpha, n-1} \right\},$$

en donde  $t_{\alpha, n-1}$  es el cuantil de la distribución  $t(n - 1)$  al nivel  $\alpha$ . Respecto del error tipo II para esta región de rechazo, la probabilidad

de no rechazar  $H_0$  cuando  $\theta = \theta_1$ , en donde  $\theta_1$  es un valor menor a  $\theta_0$ , se puede comprobar que

$$\begin{aligned}\beta(\theta_1) &= P\left(\frac{\bar{X} - \theta_0}{S/\sqrt{n}} > t_{\alpha, n-1} \mid \theta = \theta_1\right) \\ &\approx 1 - F\left(t_{\alpha, n-1} + \frac{\theta_0 - \theta_1}{s/\sqrt{n}}\right),\end{aligned}$$

en donde  $F$  denota la función de distribución  $t(n-1)$ . La última expresión es únicamente una aproximación a la probabilidad buscada pues se ha reemplazado la desviación estándar muestral  $S$  por su valor observado  $s$ .

- **Prueba de cola derecha.** Finalmente para la prueba de cola superior  $H_0 : \theta = \theta_0$  vs  $H_1 : \theta > \theta_0$  se conocen los siguientes resultados. Una región de rechazo de tamaño prefijado  $\alpha$  está dada por

$$\mathcal{C} = \{(x_1, \dots, x_n) : \frac{\bar{x} - \theta_0}{s/\sqrt{n}} \geq t_{\alpha, n-1}\},$$

en donde  $t_{\alpha, n-1}$  es el cuantil de la distribución  $t(n-1)$  al nivel  $\alpha$ . Respecto del error tipo II para esta región de rechazo, la probabilidad de no rechazar  $H_0$  cuando  $\theta = \theta_1$ , en donde  $\theta_1$  es un valor mayor a  $\theta_0$ , se puede comprobar que

$$\begin{aligned}\beta(\theta_1) &= P\left(\frac{\bar{X} - \theta_0}{S/\sqrt{n}} < t_{\alpha, n-1} \mid \theta = \theta_1\right) \\ &\approx F\left(t_{\alpha, n-1} + \frac{\theta_0 - \theta_1}{s/\sqrt{n}}\right),\end{aligned}$$

en donde  $F$  denota la función de distribución  $t(n-1)$ . Nuevamente, se ha escrito sólo una aproximación a la verdadera probabilidad del error tipo II pues se ha substituido la desviación estándar muestral  $S$  por su valor observado  $s$ .

**Ejemplo 4.2** Se desea determinar si la aplicación de un cierto medicamento afecta la presión arterial sistólica en el ser humano. Para ello se escogen al

azar diez personas, se les mide la presión arterial, después se les aplica el medicamento y una vez que éste ha tenido efecto se mide nuevamente la presión. Se calcula entonces la diferencia entre la primera medición de la presión y la segunda. Los números obtenidos fueron los siguientes:

$$2, -1, 0, -5, 3, 2, 5, -3, 0, 4.$$

Supondremos que la diferencia calculada puede modelarse mediante una variable aleatoria con distribución normal con media  $\theta$  y varianza  $\sigma^2$  desconocidas. Deseamos llevar a cabo la prueba de hipótesis

$$H_0 : \theta = 0 \quad vs \quad H_1 : \theta \neq 0.$$

La primera hipótesis establece que el medicamento no influye significativamente en la presión arterial de las personas. La segunda hipótesis indica que el medicamento sí afecta la presión arterial. Con los datos obtenidos podemos calcular la media y la varianza muestral

$$\begin{aligned}\bar{x} &= 0.7, \\ s^2 &= 9.7888,\end{aligned}$$

y entonces el valor de la estadística de la prueba es

$$t = \frac{\bar{x} - \theta_0}{s/\sqrt{n}} = 0.6712.$$

Para llevar a cabo la prueba tenemos que comparar este valor con  $t_{\alpha/2, n-1}$ . Tomaremos  $\alpha = 0.1$ , y de la tabla de la distribución  $t$  encontramos que  $t_{\alpha/2, n-1} = 1.833$ . La regla de decisión es rechazar  $H_0$  cuando  $|t| > t_{\alpha/2, n-1}$ , pero ello no sucede, por lo tanto concluimos que, con base en la muestra obtenida y la prueba estadística aplicada, no existen evidencias para afirmar que el medicamento afecte la presión arterial de las personas. ■

### Pruebas para la diferencia entre dos medias con varianzas conocidas

Sean  $X_1, \dots, X_n$  y  $Y_1, \dots, Y_m$  dos muestras aleatorias independientes de dos poblaciones, con distribución  $N(\theta_X, \sigma_X^2)$  y  $N(\theta_Y, \sigma_Y^2)$ , respectivamente.

Supondremos que las medias  $\theta_X$  y  $\theta_Y$  son desconocidas y que las varianzas  $\sigma_X^2$  y  $\sigma_Y^2$  son conocidas y pueden ser diferentes. Observe que el tamaño de las muestras puede ser distinto. En esta sección encontraremos un criterio para probar la hipótesis simple

$$H_0 : \theta_X - \theta_Y = \delta,$$

contra alguna hipótesis alternativa, en donde  $\delta$  es una constante. Mediante estas pruebas se puede decidir si las medias de las dos poblaciones normales difieren en la constante  $\delta$  o en una cantidad diferente. El procedimiento es muy similar a las pruebas presentadas antes sobre la media desconocida de una distribución normal con varianza conocida.

Sean  $\bar{X}$  y  $\bar{Y}$  las correspondientes medias muestrales. Sabemos que  $\bar{X}$  tiene distribución  $N(\theta_X, \sigma_X^2/n)$  y  $\bar{Y}$  tiene distribución  $N(\theta_Y, \sigma_Y^2/m)$ . Entonces

$$\bar{X} - \bar{Y} \sim N(\theta_X - \theta_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}).$$

Este es el resultado que nos llevará a encontrar una regla para decidir cuándo rechazar  $H_0$  en favor de alguna hipótesis alternativa, con base en los datos de las muestras aleatorias.

- **Prueba de dos colas.** Consideraremos primero el caso cuando la hipótesis alternativa es

$$H_1 : \theta_X - \theta_Y \neq \delta.$$

Cuando  $H_0$  es cierta, esto es, cuando  $\theta_X - \theta_Y = \delta$ , tenemos que  $\bar{X} - \bar{Y}$  tiene distribución  $N(\delta, \sigma_X^2/n + \sigma_Y^2/m)$ , y por lo tanto,

$$Z_0 := \frac{\bar{X} - \bar{Y} - \delta}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \sim N(0, 1).$$

La estadística  $Z_0$  es nuevamente una medida de la distancia entre la diferencia  $\bar{X} - \bar{Y}$  y  $\delta$ . Es entonces razonable rechazar  $H_0$  cuando la variable  $Z_0$  sea grande en valor absoluto. Es por ello que tomamos como criterio de decisión rechazar  $H_0$  cuando  $|Z_0| \geq c$ , para cierta

constante  $c$ . En una tabla de la distribución normal estándar podemos encontrar un valor  $z_{\alpha/2}$  tal que  $P(|Z| \geq z_{\alpha/2}) = \alpha$ , en donde  $\alpha$  es valor preestablecido para la probabilidad del error tipo I. Este valor  $z_{\alpha/2}$  es la constante  $c$  buscada y con ello se logra que la región de rechazo sea de tamaño  $\alpha$ .

$$\mathcal{C} = \{(x_1, \dots, x_n) : |Z_0| \geq z_{\alpha/2}\}.$$

Respecto del error tipo II, para un valor  $\delta_1$  distinto de  $\delta$ , es inmediato comprobar que la probabilidad de no rechazar la hipótesis nula dado que  $\theta_X - \theta_Y = \delta_1$  es

$$\begin{aligned} \beta(\delta_1) &= P(|Z_0| < z_{\alpha/2} \mid \theta_X - \theta_Y = \delta_1) \\ &= \Phi\left(z_{\alpha/2} + \frac{\delta - \delta_1}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}\right) - \Phi\left(-z_{\alpha/2} + \frac{\delta - \delta_1}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}\right). \end{aligned}$$

- **Prueba de cola izquierda.** Consideremos ahora como hipótesis alternativa

$$H_1 : \theta_X - \theta_Y < \delta,$$

Siguiendo el mismo razonamiento que antes, ahora rechazamos  $H_0$  cuando la estadística  $Z_0$  toma un valor menor a una cierta constante. El valor de esta constante que hace que la región de rechazo sea de tamaño  $\alpha$  es  $-z_\alpha$ . Así, la región de rechazo propuesta es

$$\mathcal{C} = \{(x_1, \dots, x_n) : Z_0 \leq -z_\alpha\}.$$

Respecto del error tipo II, para un valor  $\delta_1$  menor a  $\delta$ , la probabilidad de no rechazar  $H_0$  dado que  $\theta_X - \theta_Y = \delta_1$  es

$$\begin{aligned} \beta(\delta_1) &= P(Z_0 > -z_\alpha \mid \theta_X - \theta_Y = \delta_1) \\ &= 1 - \Phi\left(-z_\alpha + \frac{\delta - \delta_1}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}\right). \end{aligned}$$

- **Prueba de cola derecha.** Finalmente consideremos como hipótesis alternativa

$$H_1 : \theta_X - \theta_Y > \delta.$$

Ahora rechazamos  $H_0$  cuando la estadística  $Z_0$  toma un valor mayor a una cierta constante. El valor de esta constante que hace que la región

de rechazo sea de tamaño  $\alpha$  es  $z_\alpha$ . Así, la región de rechazo propuesta es

$$\mathcal{C} = \{(x_1, \dots, x_n) : Z_0 \geq z_\alpha\}.$$

Respecto del error tipo II, para un valor  $\delta_1$  mayor a  $\delta$ , la probabilidad de no rechazar  $H_0$  dado que  $\theta_X - \theta_Y = \delta_1$  es

$$\begin{aligned} \beta(\delta_1) &= P(Z_0 < z_\alpha \mid \theta_X - \theta_Y = \delta_1) \\ &= \Phi\left(z_\alpha + \frac{\delta - \delta_1}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}\right). \end{aligned}$$

### Pruebas para la varianza

Consideremos nuevamente una muestra aleatoria  $X_1, \dots, X_n$  proveniente de  $n$  observaciones de una variable aleatoria con distribución  $N(\mu, \sigma^2)$ , con ambos parámetros desconocidos. Nos interesa ahora encontrar un mecanismo para probar la hipótesis nula  $H_0 : \sigma^2 = \sigma_0^2$  contra alguna hipótesis alternativa. Una manera de encontrar una regla de decisión para estas pruebas hace uso del resultado teórico que establece que

$$\chi_0^2 := \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1),$$

cuando la varianza desconocida  $\sigma^2$  es, efectivamente,  $\sigma_0^2$ . Como antes, el término  $S^2$  denota la varianza muestral. Por otro lado, recordemos que la esperanza de una variable aleatoria con distribución  $\chi^2(n-1)$  es el parámetro  $n-1$ , y por lo tanto,  $E(\chi_0^2) = n-1$ . De esta manera se propone rechazar la hipótesis  $H_0$  cuando la variable aleatoria  $\chi_0^2$  tome un valor lejano de su valor central  $n-1$ .

- **Prueba de dos colas.** Para la prueba  $H_0 : \sigma^2 = \sigma_0^2$  vs  $H_1 : \sigma^2 \neq \sigma_0^2$ , se propone rechazar  $H_0$  cuando la variable  $\chi_0^2$  está alejada de su valor central tomando un valor en una de las dos colas de su distribución. Estas dos colas se establecen en la siguiente región de rechazo, la cual tiene tamaño  $\alpha$ :

$$\mathcal{C} = \{(x_1, \dots, x_n) : \chi_0^2 < \chi_{1-\alpha/2}^2 \quad \text{ó} \quad \chi_0^2 > \chi_{\alpha/2}^2\},$$

en donde  $\chi_{\alpha/2}^2$  es el número real tal que la distribución  $\chi^2(n-1)$  acumula a la derecha probabilidad  $\alpha/2$ . Análogamente, la probabilidad

a la derecha del número  $\chi_{1-\alpha/2}^2$  es  $1 - \alpha/2$ . Véase la Figura 4.13 para una representación gráfica de estas cantidades, así como de la región de rechazo de esta prueba. Por simplicidad hemos omitido especificar los grados de libertad en la notación para los valores  $\chi_{\alpha/2}^2$  y  $\chi_{1-\alpha/2}^2$ . En la página 327 aparece una tabla que muestra las cantidades  $\chi_{\alpha,n}^2$  para algunos valores de los parámetros  $\alpha$  y  $n$ .

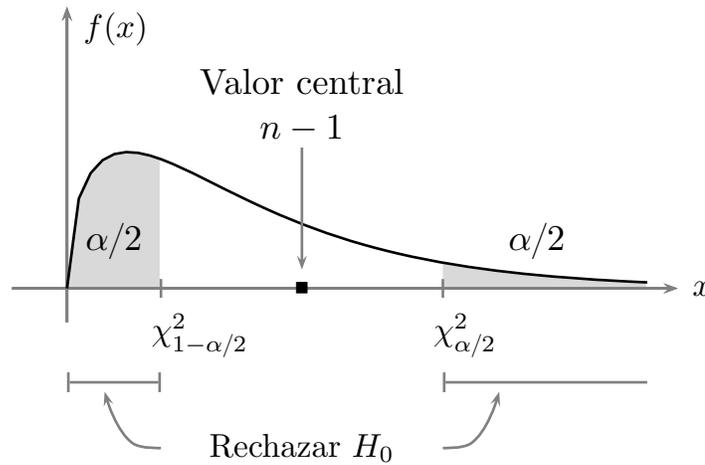


Figura 4.13

Sea  $\sigma_1^2$  cualquier número positivo distinto de  $\sigma_0^2$ . La probabilidad de no rechazar  $H_0$  cuando el valor de la varianza es  $\sigma_1^2$  es

$$\begin{aligned}
 \beta(\sigma_1^2) &= P(\text{"No rechazar } H_0" \mid \sigma^2 = \sigma_1^2) \\
 &= P(\chi_{1-\alpha/2}^2 < \frac{(n-1)S^2}{\sigma_0^2} < \chi_{\alpha/2}^2 \mid \sigma^2 = \sigma_1^2) \\
 &= P(\chi_{1-\alpha/2}^2 \cdot \frac{\sigma_0^2}{\sigma_1^2} < \frac{(n-1)S^2}{\sigma_1^2} < \chi_{\alpha/2}^2 \cdot \frac{\sigma_0^2}{\sigma_1^2} \mid \sigma^2 = \sigma_1^2) \\
 &= F(\chi_{\alpha/2}^2 \cdot \sigma_0^2/\sigma_1^2) - F(\chi_{1-\alpha/2}^2 \cdot \sigma_0^2/\sigma_1^2),
 \end{aligned}$$

en donde  $F$  es la función de distribución  $\chi^2(n-1)$ .

- **Prueba de cola izquierda.** Para la prueba que tiene como hipótesis alternativa  $H_1 : \sigma^2 < \sigma_0^2$  se propone como región de rechazo

$$\mathcal{C} = \{(x_1, \dots, x_n) : \chi_0^2 < \chi_{1-\alpha}^2\},$$

en donde  $\chi_{1-\alpha}^2$  es el número real tal que la distribución  $\chi^2(n-1)$  acumula a la derecha probabilidad  $1-\alpha$ . Así, la región de rechazo se puede identificar como la cola de la izquierda de área  $\alpha$  de la distribución  $\chi^2(n-1)$ . Esta es, por lo tanto, una región de rechazo de tamaño  $\alpha$ . Para cualquier valor positivo  $\sigma_1^2$  menor a  $\sigma_0^2$ , la probabilidad de no rechazar  $H_0$  cuando el valor de la varianza es  $\sigma_1^2$  es

$$\begin{aligned}\beta(\sigma_1^2) &= P(\text{“No rechazar } H_0\text{”} \mid \sigma^2 = \sigma_1^2) \\ &= P\left(\frac{(n-1)S^2}{\sigma_0^2} > \chi_{1-\alpha}^2 \mid \sigma^2 = \sigma_1^2\right) \\ &= P\left(\frac{(n-1)S^2}{\sigma_1^2} > \chi_{1-\alpha}^2 \cdot \frac{\sigma_0^2}{\sigma_1^2} \mid \sigma^2 = \sigma_1^2\right) \\ &= 1 - F(\chi_{1-\alpha}^2 \cdot \sigma_0^2/\sigma_1^2),\end{aligned}$$

en donde  $F$  es la función de distribución  $\chi^2(n-1)$ .

- **Prueba de cola derecha.** Y finalmente para la prueba con hipótesis alternativa  $H_1 : \sigma^2 > \sigma_0^2$  se propone como región de rechazo

$$\mathcal{C} = \{(x_1, \dots, x_n) : \chi_0^2 > \chi_\alpha^2\},$$

en donde  $\chi_\alpha^2$  es el número real tal que la distribución  $\chi^2(n-1)$  acumula a la derecha probabilidad  $\alpha$ . Así, la región de rechazo se puede identificar como la cola de la derecha de área  $\alpha$  de la distribución  $\chi^2(n-1)$ . Esta es, por lo tanto, una región de rechazo de tamaño  $\alpha$ . Para cualquier valor  $\sigma_1^2$  mayor a  $\sigma_0^2$ , la probabilidad de no rechazar  $H_0$  cuando el valor de la varianza es  $\sigma_1^2$  es

$$\begin{aligned}\beta(\sigma_1^2) &= P(\text{“No rechazar } H_0\text{”} \mid \sigma^2 = \sigma_1^2) \\ &= P\left(\frac{(n-1)S^2}{\sigma_0^2} < \chi_\alpha^2 \mid \sigma^2 = \sigma_1^2\right) \\ &= P\left(\frac{(n-1)S^2}{\sigma_1^2} < \chi_\alpha^2 \cdot \frac{\sigma_0^2}{\sigma_1^2} \mid \sigma^2 = \sigma_1^2\right) \\ &= F(\chi_\alpha^2 \cdot \sigma_0^2/\sigma_1^2),\end{aligned}$$

en donde  $F$  es la función de distribución  $\chi^2(n-1)$ .

## Ejercicios

283. En ciertas zonas de la ciudad y durante varios años se ha calculado el pago por el consumo de agua suponiendo un consumo promedio de 20,000 litros mensuales en cada casa. Para determinar si tal cantidad ha cambiado, se han medido los consumos mensuales de 15 casas escogidas al azar, obteniéndose los siguientes resultados: 23456, 18325, 21982, 22371, 13292, 25073, 22601, 20930, 18788, 19162, 21442, 23935, 20320, 19095, 17421. ¿Debe cambiar el consumo promedio mensual estimado para el cálculo de los pagos o permanecer igual? Suponga  $\sigma = 2000$ .
284. En una muestra aleatoria, el tiempo promedio en el que 50 mujeres terminaron una prueba escrita fue de 30 minutos, mientras que 45 hombres terminaron la prueba en un promedio de 35 minutos. Para fines ilustrativos supondremos una varianza de 9 unidades en ambas poblaciones. ¿Hay alguna diferencia entre hombres y mujeres en el tiempo promedio real para concluir la prueba?

## 4.6. Lema de Neyman-Pearson

Consideremos una distribución de probabilidad dependiente de un parámetro desconocido  $\theta$ . Nos interesa llevar a cabo el contraste de dos hipótesis simples

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta = \theta_1,$$

en donde  $\theta_0$  y  $\theta_1$  son dos posibles valores distintos del parámetro  $\theta$ , los cuales supondremos fijos y conocidos. En esta situación, a la probabilidad complementaria del error tipo II, esto es, al número  $1 - \beta$  le hemos llamado potencia de la prueba. Considerando todas las posibles regiones de rechazo de tamaño  $\alpha$ , a aquella que tenga potencia mayor se le llama prueba más potente.

El siguiente resultado, llamado lema de Neyman-Pearson, resuelve el problema de encontrar la región de rechazo más potente para la prueba indicada, es decir, proporciona la región de rechazo con probabilidad de error tipo II más pequeña.

**Proposición 4.1 (Lema de Neyman-Pearson<sup>1</sup>)** La región de rechazo de tamaño  $\alpha$  más potente para el contraste de dos hipótesis simples

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta = \theta_1,$$

está dada por

$$\mathcal{C} = \{(x_1, \dots, x_n) : \frac{L(x_1, \dots, x_n, \theta_1)}{L(x_1, \dots, x_n, \theta_0)} \geq c\}, \quad (4.2)$$

en donde  $L(x_1, \dots, x_n; \theta)$  es la función de verosimilitud de una muestra aleatoria y  $c$  es una constante que hace que esta región de rechazo sea de tamaño  $\alpha$ .

**Demostración.** Por brevedad en la escritura, consideraremos únicamente el caso continuo y escribiremos  $x^n$  en lugar del vector  $(x_1, \dots, x_n)$ . Considerando la región de rechazo  $\mathcal{C}$  definida en el enunciado y observando que la función de verosimilitud  $L(x^n, \theta)$  es la función de densidad del vector de la muestra aleatoria evaluada en el punto  $x^n$ , las probabilidades de cometer los errores tipo I y II son:

$$\begin{aligned} \alpha &= P((X_1, \dots, X_n) \in \mathcal{C} \mid \theta = \theta_0) = \int_{\mathcal{C}} L(x^n, \theta_0) dx^n, \\ \beta &= P((X_1, \dots, X_n) \in \mathcal{C}^c \mid \theta = \theta_1) = \int_{\mathcal{C}^c} L(x^n, \theta_1) dx^n. \end{aligned}$$

Sea  $\mathcal{C}'$  cualquier otra región de rechazo de tamaño  $\alpha$  y sea  $\beta'$  la correspondiente probabilidad de cometer el error tipo II. En la Figura 4.14 se ilustra gráficamente la situación general de estos dos conjuntos. Demostraremos que  $\beta' \geq \beta$ .

<sup>1</sup>Jerzy Neyman (1894-1981), matemático y estadístico polaco.

<sup>1</sup>Egon Sharpe Pearson (1895-1980), estadístico inglés. Hijo de Karl Pearson.

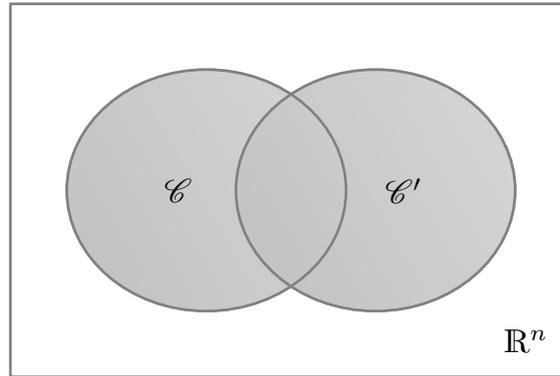


Figura 4.14

Por definición tenemos que

$$\begin{aligned}
 \beta' &= P((X_1, \dots, X_n) \notin \mathcal{C}' \mid \theta = \theta_1) \\
 &= 1 - \int_{\mathcal{C}'} L(x^n, \theta_1) dx^n \\
 &= \left[ \int_{\mathcal{C}} L(x^n, \theta_1) dx^n + \int_{\mathcal{C}^c} L(x^n, \theta_1) dx^n \right] - \int_{\mathcal{C}'} L(x^n, \theta_1) dx^n \\
 &= \left[ \int_{\mathcal{C}} L(x^n, \theta_1) dx^n - \int_{\mathcal{C}'} L(x^n, \theta_1) dx^n \right] + \int_{\mathcal{C}^c} L(x^n, \theta_1) dx^n \\
 &= \left[ \int_{\mathcal{C} - \mathcal{C}'} L(x^n, \theta_1) dx^n - \int_{\mathcal{C}' - \mathcal{C}} L(x^n, \theta_1) dx^n \right] + \int_{\mathcal{C}^c} L(x^n, \theta_1) dx^n.
 \end{aligned}$$

Como la primera integral se calcula para valores  $x^n$  dentro de la región de rechazo  $\mathcal{C}$  y la segunda se calcula fuera de esta región de rechazo, tenemos que

$$\begin{aligned}
 \beta' &\geq c \left[ \int_{\mathcal{C} - \mathcal{C}'} L(x^n, \theta_0) dx^n - \int_{\mathcal{C}' - \mathcal{C}} L(x^n, \theta_0) dx^n \right] + \int_{\mathcal{C}^c} L(x^n, \theta_1) dx^n \\
 &= c \left[ \int_{\mathcal{C}} L(x^n, \theta_0) dx^n - \int_{\mathcal{C}'} L(x^n, \theta_0) dx^n \right] + \int_{\mathcal{C}^c} L(x^n, \theta_1) dx^n \\
 &= \int_{\mathcal{C}^c} L(x^n, \theta_1) dx^n \\
 &= \beta.
 \end{aligned}$$

Las dos integrales que aparecen dentro del último paréntesis son iguales a  $\alpha$ , pues, por hipótesis, tanto  $\mathcal{C}$  como  $\mathcal{C}'$  son regiones de rechazo de tamaño  $\alpha$ . ■

Observaciones:

- El lema de Neyman-Pearson es válido tanto para distribuciones discretas como continuas. Sin embargo, en el caso discreto podría presentarse la situación de no existencia de regiones de rechazo de tamaño exactamente un valor particular de  $\alpha$ . En tales casos se buscan posibles regiones de rechazo de tamaño  $\alpha'$  cercano a  $\alpha$  con  $\alpha' \leq \alpha$ .
- El parámetro  $\theta$  en el enunciado del lema de Neyman-Pearson puede ser un vector de parámetros. En este caso, las regiones de rechazo pueden ser más difíciles de identificar y las probabilidades de error pueden presentar mayor dificultad en su cálculo.
- Durante la prueba del lema de Neyman-Pearson no se hace distinción entre los casos:  $\theta_0 < \theta_1$  ó  $\theta_1 < \theta_0$ , de modo que la región de rechazo más potente (4.2) es la misma en ambas situaciones.

Veamos ahora algunos ejemplos del uso del resultado de Neyman y Pearson.

**Ejemplo 4.3** Sea  $X$  una variable aleatoria con distribución  $N(\theta, \sigma^2)$ , en donde  $\theta$  es desconocido pero  $\sigma^2$  es conocida. Supongamos que deseamos tomar una decisión respecto del parámetro desconocido  $\theta$  de acuerdo al siguiente contraste de hipótesis simples

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta = \theta_1.$$

Supondremos que los valores  $\theta_0$  y  $\theta_1$  son fijos, conocidos y, sin pérdida de generalidad, consideraremos que guardan la relación  $\theta_0 < \theta_1$ . Con base en una muestra aleatoria  $X_1, \dots, X_n$  de esta distribución y usando el lema de Neyman-Pearson, encontraremos la región de rechazo óptima de tamaño  $\alpha$ . Tenemos que la función de verosimilitud es

$$L(x^n, \theta) = \left( \frac{1}{2\pi\sigma^2} \right)^{n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2\right).$$

Por lo tanto, el cociente de verosimilitudes (4.2) es

$$\begin{aligned} \frac{L(x^n, \theta_1)}{L(x^n, \theta_0)} &= \exp \left( -\frac{1}{2\sigma^2} \sum_{i=1}^n [(x_i - \theta_1)^2 - (x_i - \theta_0)^2] \right) \\ &= \exp \left( -\frac{1}{2\sigma^2} [n(\theta_1^2 - \theta_0^2) - 2n\bar{x}(\theta_1 - \theta_0)] \right). \end{aligned}$$

Después de algunos cálculos puede comprobarse que la condición de que la expresión anterior sea mayor o igual a una constante  $c$  es equivalente a la condición  $\bar{x} \geq c$ , para alguna otra constante  $c$  a la que denotaremos por la misma letra. Aquí se usa el supuesto de que  $\theta_0 < \theta_1$ . La región de rechazo más potente es entonces

$$\mathcal{C} = \{(x_1, \dots, x_n) : \bar{x} \geq c\}.$$

Para especificar de manera completa a este conjunto, resta encontrar el valor de la constante  $c$  que hace que esta región de rechazo sea de tamaño  $\alpha$ , es decir,  $c$  debe ser tal que

$$\begin{aligned} \alpha &= P(\bar{X} \geq c \mid \theta = \theta_0) \\ &= P\left(\frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} \geq \frac{c - \theta_0}{\sigma/\sqrt{n}} \mid \theta = \theta_0\right) \\ &= 1 - \Phi\left(\frac{c - \theta_0}{\sigma/\sqrt{n}}\right). \end{aligned}$$

De donde se obtiene que

$$c = \theta_0 + \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1 - \alpha).$$

Este es el valor de la constante  $c$  que hace que la probabilidad del error tipo I sea igual a  $\alpha$ . Por otro lado, la probabilidad del error tipo II es

$$\begin{aligned} \beta &= P(\bar{X} < c \mid \theta = \theta_1) \\ &= \Phi\left(\frac{c - \theta_1}{\sigma/\sqrt{n}}\right) \\ &= \Phi\left(\frac{\theta_0 - \theta_1}{\sigma/\sqrt{n}} + \Phi^{-1}(1 - \alpha)\right). \end{aligned}$$

■

**Ejemplo 4.4** Sea  $X$  una variable aleatoria con distribución  $\text{Poisson}(\theta)$ , en donde el parámetro  $\theta > 0$  es desconocido. Nos interesa estimar el valor de  $\theta$  mediante el contraste de hipótesis simples

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta = \theta_1,$$

en donde  $0 < \theta_0 < \theta_1$  son dos valores fijos y conocidos. Usaremos el lema de Neyman-Pearson para encontrar la región de rechazo de tamaño  $\alpha$  más potente. Tenemos que el cociente de verosimilitudes (4.2) es

$$\begin{aligned} \frac{L(x^n, \theta_1)}{L(x^n, \theta_0)} &= \frac{e^{-\theta_1} \theta_1^{x_1}/x_1! \cdots e^{-\theta_1} \theta_1^{x_n}/x_n!}{e^{-\theta_0} \theta_0^{x_1}/x_1! \cdots e^{-\theta_0} \theta_0^{x_n}/x_n!} \\ &= e^{-n(\theta_1 - \theta_0)} (\theta_1/\theta_0)^{n\bar{x}}. \end{aligned}$$

Después de algunos cálculos puede comprobarse que la condición de que la expresión anterior sea mayor o igual a una constante  $c$  es equivalente a la condición  $x_1 + \cdots + x_n \geq c$ , para alguna otra constante  $c$  que hemos escrito bajo el mismo nombre. Aquí se ha usado la hipótesis de que  $\theta_0 < \theta_1$ . La región de rechazo óptima es entonces

$$\mathcal{C} = \{(x_1, \dots, x_n) : x_1 + \cdots + x_n \geq c\},$$

en donde la constante  $c$  es tal que la probabilidad de cometer el error tipo I es  $\alpha$ , es decir,  $c$  es tal que

$$\alpha = P(X_1 + \cdots + X_n \geq c \mid \theta = \theta_0),$$

en donde, bajo la hipótesis  $\theta = \theta_0$ , la variable  $X_1 + \cdots + X_n$  tiene distribución  $\text{Poisson}(n\theta_0)$ . Observe que, como  $X_1 + \cdots + X_n$  es una variable aleatoria discreta, es posible que la identidad anterior no se cumpla de manera exacta, de modo que se toma el valor entero  $c$  más pequeño tal que

$$P(X_1 + \cdots + X_n \geq c) \leq \alpha.$$

La probabilidad de cometer el error tipo II es

$$\beta = P(X_1 + \cdots + X_n < c \mid \theta = \theta_1),$$

en donde, bajo la hipótesis  $\theta = \theta_1$ , la variable  $X_1 + \cdots + X_n$  tiene ahora distribución  $\text{Poisson}(n\theta_1)$ . ■

Con esto concluimos nuestra breve exposición sobre pruebas de hipótesis. Existen muchas otras pruebas para rechazar o no rechazar muy diversos tipos de hipótesis estadísticas que el lector interesado puede localizar en textos como [7] y [16], o en la literatura especializada en el área de aplicación.

## Ejercicios

285. **Distribución Bernoulli.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{Ber}(\theta)$ , en donde  $\theta \in (0, 1)$  es desconocido. Sea  $\alpha \in (0, 1)$ . Encuentre la región de rechazo óptima de tamaño  $\alpha$  para el contraste de hipótesis simples

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta = \theta_1,$$

en donde  $\theta_0$  y  $\theta_1$  son dos valores parametrales fijos, conocidos y tales que  $0 < \theta_0 < \theta_1 < 1$ . Calcule además la probabilidad de cometer el error tipo II.

286. **Distribución binomial.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{bin}(k, \theta)$ , en donde  $\theta$  es desconocido y el entero  $k \geq 1$  es conocido. Sea  $\alpha \in (0, 1)$ . Encuentre la región de rechazo de tamaño  $\alpha$  más potente para el contraste de hipótesis simples

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta = \theta_1,$$

en donde  $\theta_0$  y  $\theta_1$  son dos valores parametrales fijos, conocidos y tales que  $0 < \theta_0 < \theta_1 < 1$ . Calcule además la probabilidad de cometer el error tipo II.

287. **Distribución geométrica.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\text{geo}(\theta)$ , en donde  $\theta$  es desconocido. Sea  $\alpha \in (0, 1)$ . Encuentre la región de rechazo de tamaño  $\alpha$  más potente para el contraste de hipótesis simples

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta = \theta_1,$$

en donde  $\theta_0$  y  $\theta_1$  son dos valores parametrales fijos, conocidos y tales que  $0 < \theta_0 < \theta_1 < 1$ . Calcule además la probabilidad de cometer el error tipo II.

288. **Distribución exponencial.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $\exp(\theta)$ , en donde  $\theta > 0$  es desconocido. Sea  $\alpha \in (0, 1)$ . Encuentre la región de rechazo óptima de tamaño  $\alpha$  para el contraste de hipótesis simples

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta = \theta_1,$$

en donde  $\theta_0$  y  $\theta_1$  son dos valores parametrales fijos, conocidos y tales que  $0 < \theta_0 < \theta_1$ . Calcule además la probabilidad de cometer el error tipo II.

289. **Distribución normal.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $N(\theta, \sigma^2)$ , en donde ambos parámetros son desconocidos. Sea  $\alpha \in (0, 1)$ . Encuentre la región de rechazo óptima de tamaño  $\alpha$  para el contraste de hipótesis simples

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta = \theta_1,$$

en donde  $\theta_0$  y  $\theta_1$  son dos valores fijos, conocidos y tales que  $\theta_0 < \theta_1$ . Calcule además la probabilidad de cometer el error tipo II.

Sugerencia: use el hecho de que

$$\frac{\bar{X} - \theta}{S/\sqrt{n}} \sim t(n-1).$$

290. **Distribución normal.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución  $N(\mu, \sigma^2)$ , en donde  $\mu$  es conocido y  $\sigma^2$  es desconocida. Sea  $\alpha \in (0, 1)$ . Encuentre la región de rechazo más potente de tamaño  $\alpha$  para la prueba de hipótesis simples

$$H_0 : \sigma^2 = \sigma_0^2 \quad vs \quad H_1 : \sigma^2 = \sigma_1^2,$$

en donde  $\sigma_0^2$  y  $\sigma_1^2$  son dos valores fijos, conocidos y tales que  $0 < \sigma_0^2 < \sigma_1^2$ . Calcule además la probabilidad de cometer el error tipo II.

# Apéndice A

## Fórmulas varias

### Notación

$\mathbb{N}$	Conjunto de números naturales $1, 2, 3, \dots$
$\mathbb{Z}$	Conjunto de números enteros $0, \pm 1, \pm 2, \pm 3, \dots$
$\mathbb{Q}$	Conjunto de números racionales $a/b$ en donde $a, b \in \mathbb{Z}$ con $b \neq 0$ .
$\mathbb{R}$	Conjunto de números reales.
$x^+$	$\text{máx} \{x, 0\}$ .
$x^-$	$\text{mín} \{x, 0\}$ .
$f(x+)$	Límite por la derecha de la función $f$ en el punto $x$ .
$f(x-)$	Límite por la izquierda de la función $f$ en el punto $x$ .
$x \mapsto f(x)$	Función $f(x)$ .
$:=$	Se define como.

## El alfabeto griego

---

A $\alpha$	alfa	I $\iota$	iota	P $\rho, \varrho$	ro
B $\beta$	beta	K $\kappa$	kapa	$\Sigma \sigma, \varsigma$	sigma
$\Gamma \gamma$	gama	$\Lambda \lambda$	lambda	T $\tau$	tau
$\Delta \delta$	delta	M $\mu$	mu	$\Upsilon \upsilon$	upsilon
E $\epsilon, \varepsilon$	epsilon	N $\nu$	nu	$\Phi \phi, \varphi$	fi
Z $\zeta$	zeta	$\Xi \xi$	xi	X $\chi$	ji
H $\eta$	eta	O $\omicron$	omicron	$\Psi \psi$	psi
$\Theta \theta, \vartheta$	teta	$\Pi \pi$	pi	$\Omega \omega$	omega

---

## Exponentes

- $x^1 = x$ .
- $x^0 = 1, \quad x \neq 0$ .
- $x^{-1} = \frac{1}{x}, \quad x \neq 0$ .
- $x^n \cdot x^m = x^{n+m}$ .
- $\frac{x^n}{x^m} = x^{n-m}$ .
- $(x^n)^m = x^{nm}$ .
- $(xy)^n = x^n y^n$ .
- $\left(\frac{x}{y}\right)^n = \frac{x^n}{y^n}$ .
- $x^{-n} = \frac{1}{x^n}, \quad x \neq 0$ .
- $x^{m/n} = \sqrt[n]{x^m}$ .

## Logaritmos

- $\log ab = \log a + \log b.$
- $\log \frac{a}{b} = \log a - \log b.$
- $\log a^n = n \log a.$
- $\log \sqrt[n]{a} = \frac{1}{n} \log a.$
- $\log 1 = 0.$
- $\log_a a = 1.$

## Identidades trigonométricas

- $\operatorname{sen}^2 x + \operatorname{cos}^2 x = 1.$
- $\operatorname{sen}(x \pm y) = \operatorname{sen} x \operatorname{cos} y \pm \operatorname{cos} x \operatorname{sen} y.$
- $\operatorname{cos}(x \pm y) = \operatorname{cos} x \operatorname{cos} y \mp \operatorname{sen} x \operatorname{sen} y.$
- $\operatorname{cos}(\operatorname{arc} \operatorname{sen} x) = \operatorname{sen}(\operatorname{arc} \operatorname{cos} x) = \sqrt{1 - x^2}$  si  $-1 \leq x \leq 1.$

## Fórmulas para sumas

- $\sum_{k=m}^n x_k = x_m + x_{m+1} + \cdots + x_n, \quad m \leq n.$
- $\sum_{k=1}^n c = nc, \quad c \text{ constante.}$
- $\sum_{k=1}^n k = \frac{n(n+1)}{2}.$
- $\sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{6}.$

- $\sum_{k=1}^n k^3 = \left[ \frac{n(n+1)(2n+1)}{2} \right]^2.$
- $\sum_{k=m}^n a^k = \frac{a^m - a^{n+1}}{1-a}, \quad a \neq 1, \quad m \leq n.$
- $\sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x, \quad x \in \mathbb{R}.$
- $\sum_{k=0}^n \binom{n}{k} a^k b^{n-k} = (a+b)^n, \quad a, b \in \mathbb{R}, n \in \mathbb{N}.$  (Teorema del binomio).
- $\sum_{k=1}^{\infty} \frac{1}{k}$  es divergente.
- $\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} = \ln 2.$
- $\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$  (Fórmula de Euler).
- $\sum_{x=0}^{\infty} \binom{a}{x} t^x = (1+t)^a, \quad |t| < 1, a \in \mathbb{R},$  en donde  $\binom{a}{x} := \frac{a(a-1)\cdots(a-x+1)}{x!}.$

## Fórmulas de derivación

- $\frac{d}{dx} c = 0, \quad c \text{ constante.}$
- $\frac{d}{dx} x = 1.$
- $\frac{d}{dx} x^n = n x^{n-1}.$
- $\frac{d}{dx} e^x = e^x.$

- $\frac{d}{dx} \ln x = \frac{1}{x}.$
- $\frac{d}{dx} \operatorname{sen} x = \cos x.$
- $\frac{d}{dx} \cos x = -\operatorname{sen} x.$
- $\frac{d}{dx} \tan x = \sec^2 x.$
- $\frac{d}{dx} \operatorname{arc} \operatorname{sen} x = \frac{1}{\sqrt{1-x^2}}.$
- $\frac{d}{dx} \operatorname{arc} \cos x = -\frac{1}{\sqrt{1-x^2}}.$
- $\frac{d}{dx} [f(x) \pm g(x)] = f'(x) \pm g'(x).$
- $\frac{d}{dx} [f(x)g(x)] = f(x)g'(x) + f'(x)g(x).$
- $\frac{d}{dx} \frac{f(x)}{g(x)} = \frac{g(x)f'(x) - f(x)g'(x)}{g^2(x)}.$
- $\frac{d}{dx} f(g(x)) = f'(g(x))g'(x)$  (Regla de la cadena).

## Fórmulas de integración

- $\int df(x) = \int f'(x) dx = f(x) + c.$
- $\int c dx = c \int dx, \quad c \text{ constante.}$
- $\int x^n dx = \frac{x^{n+1}}{n+1} + c, \quad n \neq -1.$
- $\int \frac{dx}{x} = \ln x + c.$

- $\int e^{ax} dx = \frac{1}{a} e^{ax} + c.$
- $\int \ln x dx = x \ln x - x + c.$
- $\int \operatorname{sen} x dx = -\cos x + c.$
- $\int \cos x dx = \operatorname{sen} x + c.$
- $\int u dv = uv - \int v du$  (Integración por partes).

## Fórmula de Stirling

Para  $n$  grande,

$$n! \approx \sqrt{2\pi} n^{n+1/2} e^{-n}.$$

$n$	$n!$	Stirling
1	1	0.92
2	2	1.91
3	6	5.83
4	24	23.50
5	120	118.01
6	720	710.07
7	5040	4980.39
8	40320	39902.39
...	...	...

## Fórmula de Leibnitz

El siguiente resultado es un caso particular de la fórmula de Leibnitz que permite intercambiar una derivada con una integral. Sea  $f(x, y)$  una función

continua tal que la derivada  $\frac{\partial}{\partial x} f(x, y)$  también es continua. Entonces

$$\frac{d}{dx} \int_a^b f(x, y) dy = \int_a^b \frac{d}{dx} f(x, y) dy.$$

Este tipo de operaciones aparecen en el presente trabajo en las condiciones de regularidad que se solicitan para las funciones de densidad  $f(x, \theta)$  y que se requieren para demostrar, por ejemplo, la cota inferior de Cramér-Rao. Para mayor información sobre la fórmula anterior y algunas generalizaciones véase [8], [9], o [14].

## Función gama

Para valores reales  $\gamma$  fuera del conjunto  $\{\dots, -2, -1, 0\}$ , la siguiente integral es convergente y se le llama la función gama.

$$\Gamma(\gamma) = \int_0^{\infty} t^{\gamma-1} e^{-t} dt,$$

En su dominio de definición, esta función satisface las siguientes propiedades:

- $\Gamma(\gamma + 1) = \gamma \Gamma(\gamma)$ .
- $\Gamma(2) = \Gamma(1) = 1$ .
- $\Gamma(\gamma + 1) = \gamma!$  para  $\gamma = 0, 1, \dots$
- $\Gamma(1/2) = \sqrt{\pi}$ .
- $\Gamma(\gamma + 1/2) = \frac{1 \cdot 3 \cdot 5 \cdots (2\gamma - 1)}{2^\gamma} \sqrt{\pi}$  para  $\gamma = 1, 2, \dots$

## Función beta

Al término  $B(a, b)$  se le conoce como la función beta, y se define para  $a > 0$  y  $b > 0$  como sigue

$$B(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx.$$

Esta función satisface las siguientes propiedades:

- $B(a, b) = B(b, a)$ .
- $B(a, 1) = 1/a$ .
- $B(1, b) = 1/b$ .
- $B(a + 1, b) = \frac{a}{b} B(a, b + 1)$ .
- $B(a + 1, b) = \frac{a}{a + b} B(a, b)$ .
- $B(a, b + 1) = \frac{b}{a + b} B(a, b)$ .
- $B(1/2, 1/2) = \pi$ .
- $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)}$ .

## Convergencia de variables aleatorias

Sea  $X_1, X_2, \dots$  una sucesión infinita de variables aleatorias definidas en un mismo espacio de probabilidad. A continuación revisamos las definiciones de algunos tipos de convergencia para este tipo de sucesiones.

- **Convergencia puntual.** Si  $\lim_{n \rightarrow \infty} X_n(\omega)$  existe para cada  $\omega$  y se le denota por  $X(\omega)$ , entonces se dice que la sucesión es convergente puntualmente y la función límite  $X$  es una variable aleatoria.
- **Convergencia casi segura.** Sea  $X$  una variable aleatoria. Se dice que la sucesión converge casi seguramente a  $X$ , y se escribe  $X_n \xrightarrow{c.s.} X$ , si

$$P\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\} = 1.$$

- **Convergencia en probabilidad.** Sea  $X$  una variable aleatoria. Se dice que la sucesión converge en probabilidad a  $X$ , y se escribe  $X_n \xrightarrow{p} X$ , si para cualquier  $\epsilon > 0$ ,

$$\lim_{n \rightarrow \infty} P\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \epsilon\} = 0.$$

- **Convergencia en media.** Sea  $X$  una variable aleatoria. Se dice que la sucesión converge en media a  $X$ , y se escribe  $X_n \xrightarrow{m} X$ , si

$$\lim_{n \rightarrow \infty} E|X_n - X| = 0.$$

- **Convergencia en media cuadrática.** Sea  $X$  una variable aleatoria. Se dice que la sucesión converge en media cuadrática a  $X$ , y se escribe  $X_n \xrightarrow{m.c.} X$ , si

$$\lim_{n \rightarrow \infty} E|X_n - X|^2 = 0.$$

- **Convergencia en distribución.** Sea  $X$  una variable aleatoria. Se dice que la sucesión converge en distribución a  $X$ , y se escribe  $X_n \xrightarrow{d} X$ , si para cualquier punto de continuidad  $x_0$  de  $F_X(x)$ ,

$$\lim_{n \rightarrow \infty} F_{X_n}(x_0) = F_X(x_0).$$

A este tipo de convergencia se le llama también **convergencia débil**.

En [11], [12] o [21] se puede encontrar mayor información sobre estos tipos de convergencia de variables aleatorias.

## Dos teoremas de convergencia

Sea  $X_1, X_2, \dots$  una sucesión de variables aleatorias que es convergente en el sentido casi seguro a la variable aleatoria  $X$ . El problema consiste en determinar si la sucesión numérica  $E(X_n)$  es convergente a  $E(X)$ , es decir, nos preguntamos si se cumple la igualdad

$$\lim_{n \rightarrow \infty} E(X_n) = E(\lim_{n \rightarrow \infty} X_n).$$

Esta identidad equivale a poder intercambiar las operaciones de límite y esperanza. Se pueden dar ejemplos en donde este intercambio de operaciones no es válido. ¿Bajo qué condiciones se cumple esta igualdad? Aquí tenemos dos resultados importantes que establecen condiciones suficientes para que un límite y la esperanza se puedan intercambiar.

- **Teorema de convergencia monótona.** Sea  $0 \leq X_1 \leq X_2 \leq \dots$  una sucesión de variables aleatorias convergente casi seguramente a una variable  $X$ . Entonces

$$\lim_{n \rightarrow \infty} E(X_n) = E(X).$$

- **Teorema de convergencia dominada.** Sea  $X_1, X_2, \dots$  una sucesión de variables aleatorias para la cual existe otra variable  $Y$  con esperanza finita tal que  $|X_n| \leq Y$ , para  $n \geq 1$ . Si  $\lim_{n \rightarrow \infty} X_n = X$  c.s., entonces  $X$  y  $X_n$  tienen esperanza finita y

$$\lim_{n \rightarrow \infty} E(X_n) = E(X).$$

Se puede encontrar mayor información sobre estos resultados en [6], [12], [28].

## Puntos críticos para funciones de varias variables

Sea  $f(x, y)$  una función real definida sobre un rectángulo  $(a, b) \times (c, d)$  de  $\mathbb{R}^2$  y cuyas derivadas de segundo orden son continuas en este rectángulo. Se dice que  $f(x, y)$  tiene un punto crítico en  $(x_0, y_0)$  si

$$\begin{aligned} \frac{\partial f}{\partial x}(x_0, y_0) &= 0, \\ \text{y } \frac{\partial f}{\partial y}(x_0, y_0) &= 0. \end{aligned}$$

Estamos interesados en recordar algunos criterios para determinar si un punto crítico es un máximo o un mínimo. Esto se utiliza en el presente trabajo en la aplicación del método de máxima verosimilitud cuando se tienen dos o más parámetros. Antes de explicar la manera en la que se puede determinar si un punto crítico es un máximo o un mínimo, vamos a definir los menores principales de una matriz cuadrada.

Sea  $A = (a_{ij})$  una matriz de  $n \times n$  y sea  $k$  un entero tal que  $1 \leq k \leq n$ . El menor principal de orden  $k$  se define como el determinante de la submatriz cuadrada  $(a_{ij})$ ,  $i, j = 1, \dots, k$ , esto es,

$$\begin{aligned} a_{11} &= \text{Primer menor principal} \quad (k = 1), \\ \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} &= \text{Segundo menor principal} \quad (k = 2), \\ &\vdots \\ |A| &= n\text{-ésimo menor principal} \quad (k = n). \end{aligned}$$

Por otro lado, para la función  $f(x, y)$  considerada antes, se define la matriz hessiana como la matriz simétrica

$$H(x, y) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(x, y) & \frac{\partial^2 f}{\partial y \partial x}(x, y) \\ \frac{\partial^2 f}{\partial x \partial y}(x, y) & \frac{\partial^2 f}{\partial y^2}(x, y) \end{pmatrix}. \quad (3)$$

- **Condiciones para un máximo.** La función  $f(x, y)$  tiene un máximo en  $(x_0, y_0)$  si la matriz  $H(x_0, y_0)$  es tal que
  - a) todos sus menores principales de orden impar son negativos y
  - b) todos sus menores principales de orden par son positivos.

Para la matriz (3) esto se reduce a las desigualdades

$$\frac{\partial^2 f}{\partial x^2}(x_0, y_0) < 0 \quad \text{y} \quad |H(x_0, y_0)| > 0.$$

Estas condiciones son equivalentes a solicitar que la matriz  $H(x_0, y_0)$  sea negativa definida. Ello significa que se deben cumplir las siguientes dos condiciones:

- a)  $(x, y)H(x_0, y_0)\begin{pmatrix} x \\ y \end{pmatrix} \leq 0$  para todo  $(x, y) \in \mathbb{R}^2$ .
  - b)  $(x, y)H(x_0, y_0)\begin{pmatrix} x \\ y \end{pmatrix} = 0 \iff (x, y) = (0, 0)$ .
- **Condiciones para un mínimo.** La función  $f(x, y)$  tiene un mínimo en  $(x_0, y_0)$  si la matriz  $H(x_0, y_0)$  es tal que

a) todos sus menores principales son positivos.

En el caso de la matriz (3) esto se reduce a las desigualdades

$$\frac{\partial^2 f}{\partial x^2}(x_0, y_0) > 0 \quad \text{y} \quad |H(x_0, y_0)| > 0.$$

Estas condiciones son equivalentes a solicitar que la matriz  $H(x_0, y_0)$  sea positiva definida. Esto significa que se debe cumplir lo siguiente:

a)  $(x, y)H(x_0, y_0)\begin{pmatrix} x \\ y \end{pmatrix} \geq 0$  para todo  $(x, y) \in \mathbb{R}^2$ .

b)  $(x, y)H(x_0, y_0)\begin{pmatrix} x \\ y \end{pmatrix} = 0 \iff (x, y) = (0, 0)$ .

Por simplicidad en la exposición, hemos considerado funciones reales de únicamente dos variables, sin embargo, los criterios anteriores pueden extenderse al caso de funciones dependientes de cualquier número finito de variables. Para mayor información sobre estos resultados, véase [9], [14], o [15].

## Fórmula recursiva para los momentos en la familia exponencial

En esta sección se presenta una fórmula recursiva para el  $n$ -ésimo momento de ciertas distribuciones dentro de la familia exponencial, la cual fue definida en la sección 2.19. Esta subfamilia de distribuciones corresponde al caso particular cuando  $d(x) = x$  en la expresión general (2.18). Adicionalmente, y sin pérdida de generalidad, se escribe  $\ln c(\theta)$  en lugar de la función  $c(\theta)$ . Esto da lugar a la siguiente expresión para la función de densidad

$$f(x, \theta) = a(\theta) b(x) (c(\theta))^x, \quad -\infty < x < \infty. \quad (4)$$

Aquí tenemos una fórmula recursiva general para los momentos de este tipo de distribuciones.

**Proposición .2** Sea  $X$  una variable aleatoria con función de densidad tipo exponencial de la forma (4). El  $n$ -ésimo momento de  $X$  existe y satisface la ecuación

$$E(X^n) = \frac{c(\theta)}{c'(\theta)} \left( -\frac{a'(\theta)}{a(\theta)} + \frac{d}{d\theta} \right) E(X^{n-1}). \quad (5)$$

**Demostración.** Se considera únicamente el caso continuo. En el caso discreto, la demostración sigue el mismo procedimiento.

$$\begin{aligned}
 E(X^n) &= \int a(\theta) b(x) x^{n-1} [x (c(\theta))^{x-1}] c(\theta) dx \\
 &= \int b(x) x^{n-1} \left[ \frac{d}{d\theta} (c(\theta))^x \right] a(\theta) \frac{c(\theta)}{c'(\theta)} dx \\
 &= \int b(x) x^{n-1} \left[ \frac{d}{d\theta} \left( a(\theta) \frac{(c(\theta))^{x+1}}{c'(\theta)} \right) - (c(\theta))^x \frac{d}{d\theta} \left( a(\theta) \frac{c(\theta)}{c'(\theta)} \right) \right] dx \\
 &= \frac{d}{d\theta} \left( \frac{c(\theta)}{c'(\theta)} E(X^{n-1}) \right) - \left( \frac{1}{a(\theta)} \frac{d}{d\theta} \left( a(\theta) \frac{c(\theta)}{c'(\theta)} \right) \right) E(X^{n-1}) \\
 &= \frac{c(\theta)}{c'(\theta)} \left( -\frac{a'(\theta)}{a(\theta)} + \frac{d}{d\theta} \right) E(X^{n-1}).
 \end{aligned}$$

■

La fórmula anterior permite escribir el  $n$ -ésimo momento en términos de un operador diferencial aplicado al momento  $n - 1$ . Esta fórmula puede considerarse una ecuación diferencial y en diferencias. Y puede resolverse empezando con el valor  $E(X^0) = 1$  y procediendo de manera iterativa. En particular, usando (5) para  $n = 1$  y  $n = 2$  se pueden demostrar las siguientes fórmulas generales para la esperanza y la varianza.

$$\begin{aligned}
 E(X) &= -\frac{c(\theta)}{c'(\theta)} \cdot \frac{a'(\theta)}{a(\theta)}, \\
 \text{Var}(X) &= \frac{c(\theta)}{c'(\theta)} \cdot \frac{d}{d\theta} E(X).
 \end{aligned}$$

Las funciones de densidad de la forma (4) pueden corresponder a distribuciones discretas o continuas. A continuación se proporcionan algunos ejemplos en donde se especifican las funciones  $a(\theta)$ ,  $b(x)$ ,  $c(\theta)$ , y se encuentra la forma particular de la fórmula (5). Por simplicidad, se omite la especificación del soporte de la distribución.

- La función de probabilidad  $\text{bin}(k, \theta)$  puede escribirse en la forma (4) tomando  $a(\theta) = (1 - \theta)^n$ ,  $b(x) = \binom{n}{x}$  y  $c(\theta) = \theta/(1 - \theta)$ . Por lo tanto,

el  $n$ -ésimo momento satisface

$$E(X^n) = \left( k\theta + \theta(1 - \theta) \frac{d}{d\theta} \right) E(X^{n-1}).$$

Empezando con  $E(X^0) = 1$ , se obtiene  $E(X) = k\theta$ ,  $E(X^2) = (k\theta)^2 + k\theta(1 - \theta)$ ,  $E(X^3) = (k\theta)^3 + 3(k\theta)^2(1 - \theta) + k\theta(1 - \theta)(1 - 2\theta)$ , etcétera.

- La función de probabilidad bin neg( $r, \theta$ ) puede escribirse en la forma (4) tomando  $a(\theta) = \theta^r$ ,  $b(x) = \binom{r+x-1}{x}$  y  $c(\theta) = 1 - \theta$ . Por lo tanto, el  $n$ -ésimo momento satisface

$$E(X^n) = \left( \frac{r(1 - \theta)}{\theta} - (1 - \theta) \frac{d}{d\theta} \right) E(X^{n-1}).$$

Empezando con  $E(X^0) = 1$ , se obtiene  $E(X) = r(1 - \theta)/\theta$ ,  $E(X^2) = r^2(1 - \theta)^2/\theta^2 + r(1 - \theta)/\theta^2$ ,  $E(X^3) = r^3(1 - \theta)^3/\theta^3 + 3r^2(1 - \theta)^2/\theta^3 + 2r(1 - \theta)/\theta^3 - r(1 - \theta)/\theta^2$ , etcétera. El caso particular  $r = 1$  corresponde a la distribución  $\text{geo}(\theta)$ .

- La función de probabilidad Poisson( $\theta$ ) puede escribirse en la forma (4) tomando  $a(\theta) = e^{-\theta}$ ,  $b(x) = 1/x!$  y  $c(\theta) = \theta$ . Por lo tanto, el  $n$ -ésimo momento de esta distribución satisface

$$E(X^n) = \left( \theta + \theta \frac{d}{d\theta} \right) E(X^{n-1}).$$

Empezando con  $E(X^0) = 1$ , se obtiene  $E(X) = \theta$ ,  $E(X^2) = \theta^2 + \theta$ ,  $E(X^3) = \theta^3 + 3\theta^2 + \theta$ ,  $E(X^4) = \theta^4 + 6\theta^3 + 7\theta^2 + \theta$ , etcétera. Usando el método de inducción puede comprobarse la fórmula

$$E(X^n) = \theta \sum_{k=0}^{n-1} \binom{n-1}{k} E(X^k).$$

- La función de densidad gama( $\alpha, \theta$ ) puede escribirse en la forma (4) tomando  $a(\theta) = \theta^\alpha$ ,  $b(x) = x^{\alpha-1}/\Gamma(\alpha)$  and  $c(\theta) = e^{-\theta}$ . Esto incluye a

la distribución  $\exp(\theta)$  cuando  $\alpha = 1$ . Por lo tanto, el  $n$ -ésimo momento de esta distribución satisface

$$E(X^n) = \left( \frac{\alpha}{\theta} - \frac{d}{d\theta} \right) E(X^{n-1}).$$

Empezando con  $E(X^0) = 1$ , se obtiene  $E(X) = \alpha/\theta$ ,  $E(X^2) = \alpha^2/\theta^2 + \alpha/\theta^2$ ,  $E(X^3) = \alpha^3/\theta^3 + 3\alpha^2/\theta^3 + 2\alpha/\theta^3$ , etcétera. Factorizando términos se puede recuperar la fórmula

$$E(X^n) = \frac{\alpha(\alpha + 1) \cdots (\alpha + n - 1)}{\theta^n}.$$

El caso  $\alpha = 1$  produce los momentos de la distribución  $\exp(\theta)$ , los cuales son

$$E(X^n) = \frac{n!}{\theta^n}.$$

- Finalmente consideramos el caso normal. La función de densidad  $N(\theta, \sigma^2)$  puede escribirse en la forma (4) tomando  $a(\theta) = e^{-\theta^2/2\sigma^2}$ ,  $b(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}$  y  $c(\theta) = e^{\theta/\sigma^2}$ . Por lo tanto, el  $n$ -ésimo momento de esta distribución satisface

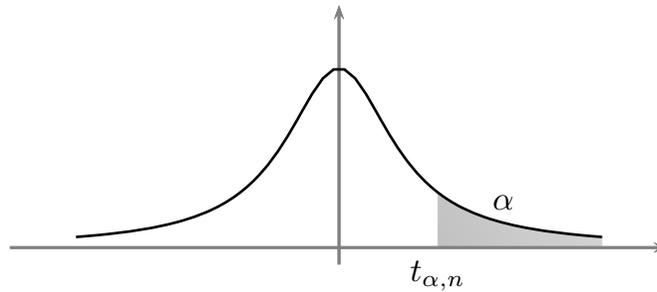
$$E(X^n) = \left( \theta + \sigma^2 \frac{d}{d\theta} \right) E(X^{n-1}).$$

Empezando con  $E(X^0) = 1$ , se obtiene  $E(X) = \theta$ ,  $E(X^2) = \theta^2 + \sigma^2$ ,  $E(X^3) = \theta^3 + 3\theta\sigma^2$ ,  $E(X^4) = \theta^4 + 6\theta^2\sigma^2 + 3\sigma^4$ , etcétera.

Algunos de estos ejemplos muestran que las distribuciones de la forma (4) pueden tener más de un parámetro. Los resultados siguen siendo válidos siempre y cuando se considere a uno de ellos el parámetro principal para la representación, aquí denotado por la letra  $\theta$ , y el resto de los parámetros se consideren como constantes.



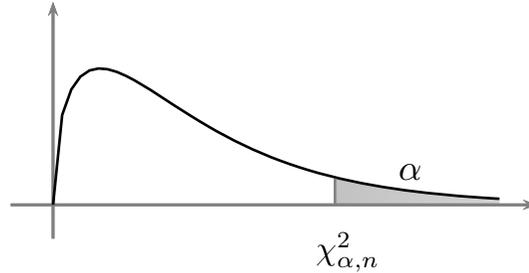
## Tabla de la distribución $t(n)$



$$P(X \geq t_{\alpha, n}) = \alpha$$

$n \setminus \alpha$	0.005	0.01	0.025	0.05	0.1
1	63.657	31.821	12.706	6.314	3.078
2	9.925	6.965	4.303	2.920	1.886
3	5.841	4.541	3.182	2.353	1.638
4	4.604	3.474	2.776	2.132	1.533
5	4.032	3.365	2.571	2.015	1.476
6	3.707	3.143	2.447	1.943	1.440
7	3.499	2.998	2.365	1.895	1.415
8	3.355	2.896	2.306	1.860	1.397
9	3.250	2.821	2.262	1.833	1.383
10	3.169	2.764	2.228	1.812	1.372
11	3.106	2.718	2.201	1.796	1.363
12	3.055	2.681	2.179	1.782	1.356
13	3.012	2.650	2.160	1.771	1.350
14	2.977	2.624	2.145	1.761	1.345
15	2.947	2.602	2.131	1.753	1.341
16	2.921	2.583	2.120	1.746	1.337
17	2.898	2.567	2.110	1.740	1.333
18	2.878	2.552	2.101	1.734	1.330
19	2.861	2.539	2.093	1.729	1.328
20	2.845	2.528	2.086	1.725	1.325
21	2.831	2.518	2.080	1.721	1.323
22	2.819	2.508	2.074	1.717	1.321
23	2.807	2.500	2.069	1.714	1.319
24	2.797	2.492	2.064	1.711	1.318
25	2.787	2.485	2.060	1.708	1.316
26	2.779	2.479	2.056	1.706	1.315
27	2.771	2.473	2.052	1.703	1.314
28	2.763	2.467	2.048	1.701	1.313
29	2.756	2.462	2.045	1.699	1.311
$\infty$	2.576	2.326	1.960	1.645	1.282

# Tabla de la distribución $\chi^2(n)$



$$P(X \geq \chi_{\alpha, n}^2) = \alpha$$

$n \setminus \alpha$	0.995	0.990	0.975	0.950	0.900	0.100	0.050	0.025	0.010	0.005
1	0.0	0.0	0.0	0.0	0.02	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	10.65	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.87	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	34.28	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96	49.65
28	12.46	13.57	15.31	16.93	18.94	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	85.53	90.53	95.02	100.42	104.22
80	51.17	53.54	57.15	60.39	64.28	96.58	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	73.29	107.57	113.14	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	82.36	118.50	124.34	129.56	135.81	140.17

## Tabla de distribuciones discretas

Distribución	Función de probabilidad	Parámetros	Esperanza
Uniforme discreta	$f(x) = 1/n$ para $x = x_1, \dots, x_n$	$x_1, \dots, x_n \in \mathbb{R}$ $n = 1, 2, \dots$	$\frac{1}{n} \sum_{i=1}^n x_i$
Bernoulli	$f(x) = p^x (1-p)^{1-x}$ para $x = 0, 1$	$0 < p < 1$	$p$
Binomial	$f(x) = \binom{k}{x} p^x (1-p)^{k-x}$ para $x = 0, 1, \dots, k$	$n = 1, 2, \dots$ $0 < p < 1$	$kp$
Geométrica	$f(x) = p(1-p)^x$ para $x = 0, 1, \dots$	$0 < p < 1$	$\frac{1-p}{p}$
Binomial negativa	$f(x) = \binom{r+x-1}{x} p^r (1-p)^x$ para $x = 0, 1, \dots$	$r = 1, 2, \dots$ $0 < p < 1$	$\frac{r(1-p)}{p}$
Hipergeométrica	$f(x) = \frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}}$ para $x = 0, 1, \dots, n$	$K = 1, 2, \dots$ $N - K = 1, 2, \dots$ $n \leq \min\{K, N - K\}$	$\frac{nK}{N}$
Poisson	$f(x) = e^{-\lambda} \frac{\lambda^x}{x!}$ para $x = 0, 1, \dots$	$\lambda > 0$	$\lambda$

Varianza	Momentos $E(X^k)$	Función generadora de probabilidad $E(t^X)$	Función generadora de momentos $E(e^{tX})$
$\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$	$\frac{1}{n} \sum_{i=1}^n x_i^k$	$\frac{1}{n} \sum_{i=1}^n t^{x_i}$	$\frac{1}{n} \sum_{i=1}^n e^{tx_i}$
$p(1-p)$	$p$	$1-p+pt$	$1-p+pe^t$
$kp(1-p)$	[1]	$(1-p+pt)^k$	$(1-p+pe^t)^k$
$\frac{1-p}{p^2}$	[1]	$\frac{p}{1-(1-p)t}$ si $ t  < 1/(1-p)$	$\frac{p}{1-(1-p)e^t}$ si $ t  < -\ln(1-p)$
$\frac{r(1-p)}{p^2}$	[1]	$(\frac{p}{1-(1-p)t})^r$ si $ t  < 1/(1-p)$	$(\frac{p}{1-(1-p)e^t})^r$ si $ t  < -\ln(1-p)$
$n \frac{K}{N} \frac{N-K}{N} \frac{N-n}{N-1}$	[2]	[3]	[3]
$\lambda$	[1]	$e^{\lambda(t-1)}$	$e^{\lambda(e^t-1)}$

[1] Vea la fórmula recursiva general (5) en la página 321.

[2] No existe una fórmula compacta.

[3] La definición de esta función no produce una fórmula reducida.

## Tabla de distribuciones continuas

Distribución	Función de densidad	Parámetros	Esperanza
Uniforme continua	$f(x) = \frac{1}{b-a}$ para $x \in (a, b)$	$a < b$	$\frac{a+b}{2}$
Exponencial	$f(x) = \lambda e^{-\lambda x}$ para $x > 0$	$\lambda > 0$	$\frac{1}{\lambda}$
Gama	$f(x) = \frac{(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} \lambda e^{-\lambda x}$ para $x > 0$	$\alpha > 0$ $\lambda > 0$	$\frac{\alpha}{\lambda}$
Beta	$f(x) = \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1}$ para $0 < x < 1$	$a > 0$ $b > 0$	$\frac{a}{a+b}$
Weibull	$f(x) = \lambda \alpha (\lambda x)^{\alpha-1} e^{-(\lambda x)^\alpha}$ para $x > 0$	$\alpha > 0$ $\lambda > 0$	$\frac{\Gamma(1+1/\alpha)}{\lambda}$
Normal	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$ para $-\infty < x < \infty$	$-\infty < \mu < \infty$ $\sigma^2 > 0$	$\mu$
Ji-cuadrada	$f(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}$ para $x > 0$	$n > 0$	$n$
t	$f(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} (1 + \frac{x^2}{n})^{-(n+1)/2}$ para $-\infty < x < \infty$	$n > 0$	0 si $n > 1$
F	$f(x) = \frac{\Gamma((a+b)/2)}{\Gamma(a/2)\Gamma(b/2)} (a/b)^{a/2} x^{a/2-1} (1 + \frac{a}{b}x)^{-(a+b)/2}$ para $x > 0$	$a > 0$ $b > 0$	$\frac{b}{b-2}$ si $b > 2$

Varianza	Momentos $E(X^k)$	Función generadora de momentos $E(e^{tX})$
$\frac{(b-a)^2}{12}$	$\frac{b^{k+1}-a^{k+1}}{(k+1)(b-a)}$	$\frac{e^{bt}-e^{at}}{t(b-a)}$ si $t \neq 0$
$\frac{1}{\lambda^2}$	$\frac{k!}{\lambda^k}$	$\frac{\lambda}{\lambda-t}$ si $t < \lambda$
$\frac{\alpha}{\lambda^2}$	$\frac{\alpha(\alpha+1)\cdots(\alpha+k-1)}{\lambda^k}$	$(\frac{\lambda}{\lambda-t})^\alpha$ si $t < \lambda$
$\frac{ab}{(a+b+1)(a+b)^2}$	$\frac{B(a+k,b)}{B(a,b)}$	No existe fórmula reducida
$\frac{\Gamma(1+2/\alpha)-\Gamma^2(1+1/\alpha)}{\lambda^2}$	$\frac{\Gamma(1+k/\alpha)}{\lambda^k}$	No existe fórmula reducida
$\sigma^2$	[1]	$\exp(\mu t + \sigma^2 t^2/2)$
$2n$	$\frac{2^k \Gamma(n/2+k)}{\Gamma(n/2)}$	$(\frac{1}{1-2t})^{n/2}$ si $t < 1/2$
$\frac{n}{n-2}$ si $n > 2$	No existe fórmula reducida	No existe
$\frac{2b^2(a+b-2)}{a(b-2)^2(b-4)}$ si $b > 4$	$(\frac{b}{a})^k \frac{\Gamma(a/2+k)}{\Gamma(a/2)} \frac{\Gamma(b/2-k)}{\Gamma(b/2)}$ si $2k < b$	No existe

[1] Vea la fórmula recursiva general (5) en la página 321.

# Apéndice B

## Sugerencias a los ejercicios

Esta sección contiene algunas sugerencias para la solución de los ejercicios planteados. Para algunos ejercicios es necesario ser más explícito al dar una solución o al justificar una respuesta. Considere, por tanto, que este material contiene simplemente ideas para generar una solución completa, correcta y bien escrita. La mayoría de las gráficas han sido omitidas. Recuerde además que los métodos empleados o sugeridos para llegar a una solución no son necesariamente únicos.

1.
  - a) El conjunto de personas a las que se les suministra el medicamento.
  - b) El conjunto de personas con poder de adquisición del producto.
  - c) El conjunto de personas de edad avanzada.
  - d) El conjunto de computadoras que pueden hacer uso de ese programa.
  - e) El conjunto de todos los artículos producidos por la maquinaria.
2.
  - a) El grado de eficacia del medicamento.
  - b) La cantidad de producto consumido en un determinado tiempo.
  - c) El nivel de ingreso de las personas de edad avanzada.
  - d) El número de fallas detectadas en un periodo determinado.
  - e) El número de defectos en cada artículo producido.
3.
  - a) La edad de un niño.
  - b) El número de veces que una página web es consultada en un día.
  - c) Tamaño del vaso de café que una persona llega a comprar a una cafetería.

- d)* Color de un producto.
  - e)* Temperatura promedio de un día.
  - f)* El nombre de una persona.
  - g)* El resultado de la prueba de control de calidad de un producto.
  - h)* Tiempo de vida restante de un foco.
4. Dependiendo de la escala de medición utilizada, la clasificación de las variables puede cambiar.
- a)* Cuantitativa discreta.
  - b)* Cuantitativa discreta.
  - c)* Cualitativa.
  - d)* Cuantitativa continua.
  - e)* Cuantitativa discreta.
  - f)* Cuantitativa continua.
  - g)* Cuantitativa discreta.
  - h)* Cualitativa.
  - i)* Cuantitativa continua.
  - j)* Cuantitativa continua.
  - k)* Cualitativa.
  - l)* Cuantitativa continua.
  - m)* Cuantitativa continua.
  - n)* Cuantitativa discreta.
  - ñ)* Cualitativa.
  - o)* Cualitativa.
  - p)* Cuantitativa discreta.
  - q)* Cuantitativa discreta.
  - r)* Cualitativa.
  - s)* Cualitativa.
  - t)* Cuantitativa discreta.
  - u)* Cuantitativa discreta.
  - v)* Cuantitativa discreta.
  - w)* Cualitativa.

- x*) Cuantitativa continua.
5.
    - a*) Ordinal.
    - b*) Nominal.
    - c*) Nominal.
    - d*) Ordinal.
    - e*) Ordinal.
    - f*) Nominal.
    - g*) Nominal.
    - h*) Nominal.
  6.
    - a*) Discreta, escala de intervalo.
    - b*) Discreta, escala de razón.
    - c*) Discreta, escala de razón.
    - d*) Continua, escala de razón.
    - e*) Discreta, escala de razón.
    - f*) Continua, escala de razón.
    - g*) Continua, escala de razón.
    - h*) Continua, escala de razón.
  7. La variable creada es de tipo cualitativo con escala de medición ordinal. Si el consumo bajo se define como tomar 1 tasa de café o menos, entonces una marca de clase para esta categoría puede ser 0.5 tasa de café. Si el consumo medio se define como tomar entre 1 y 2 tasas de café, entonces una marca de clase para esta categoría puede ser 1.5 tasas de café. Y si el consumo alto se define como tomar por lo menos 2 tasas de café, entonces una marca de clase para esta categoría puede ser, por ejemplo, 2.5 tasas de café. Observe que el conjunto de las marcas de clase corresponde a los valores de una variable cuantitativa discreta.
  8. La variable creada es de tipo cualitativo con escala de medición ordinal. El valor medio de cada subintervalo puede servir como marca de clase, esto es  $a/2$  para la categoría números positivos y el valor  $-a/2$  para la categoría números negativos. El valor 0 puede ser incluido en cualquiera de las dos categorías.
  9. Se usan las propiedades de la sumatoria.

$$a) \sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - n\bar{x} = n\bar{x} - n\bar{x} = 0.$$

$$\begin{aligned}
 b) \sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) = \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + n\bar{x}^2 \\
 &= \sum_{i=1}^n x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2.
 \end{aligned}$$

10.  $x_3 = 1$ .

11.  $\bar{y} = \frac{1}{n} \sum_{i=1}^n (ax_i + c) = a \frac{1}{n} \sum_{i=1}^n x_i + \frac{1}{n} \sum_{i=1}^n c = a\bar{x} + c$ .

12.  $\bar{x} = 3$ ,  $\overline{x+1} = 4$ ,  $\overline{x-2} = 1$ ,  $\overline{2x} = 6$ ,  $\overline{x/3} = 1$ .

13.  $\bar{x} = 1.6$ .

14. a) El mismo dato.

b) El mismo dato.

c) El promedio de ellos.

d) El mismo dato.

15. a) Verdadero.

b) Falso.

c) Verdadero.

d) Falso, a menos que la media original haya sido cero.

e) Puede ser falso o verdadero dependiendo del dato añadido. Si el dato adicional es menor a la media original, la nueva media disminuye. Si el dato adicional es mayor a la media original, la nueva media aumenta. La nueva media no cambia si el dato adicional es la media original.

f) Misma situación que en el inciso anterior.

16. La constante  $a$ .

17.  $(n+1)/2$ .

18.

$$\begin{aligned}
 \bar{x}_{n+1} &= \frac{1}{n+1} (x_1 + \cdots + x_{n+1}) \\
 &= \frac{1}{n+1} (x_1 + \cdots + x_n) + \frac{1}{n+1} x_{n+1} \\
 &= \frac{n}{n+1} \frac{1}{n} (x_1 + \cdots + x_n) + \frac{1}{n+1} x_{n+1} \\
 &= \frac{1}{n+1} n\bar{x} + \frac{1}{n+1} x_{n+1} \\
 &= \frac{1}{n+1} (n\bar{x} + x_{n+1}).
 \end{aligned}$$

19. La misma media. Use la fórmula del ejercicio anterior para demostrar esta afirmación.

20.

$$\begin{aligned}
 \bar{x}_{n-1} &= \frac{1}{n-1} (x_1 + \cdots + x_{i-1} + x_{i+1} + \cdots + x_n) \\
 &= \frac{1}{n-1} (x_1 + \cdots + x_n - x_i) \\
 &= \frac{1}{n-1} (x_1 + \cdots + x_n) - \frac{1}{n-1} x_i \\
 &= \frac{n}{n-1} \frac{1}{n} (x_1 + \cdots + x_n) - \frac{1}{n-1} x_i \\
 &= \frac{1}{n-1} n \bar{x} - \frac{1}{n-1} x_i \\
 &= \frac{1}{n-1} (n \bar{x} - x_i).
 \end{aligned}$$

$$21. \overline{x+y} = \frac{1}{n} \sum_{i=1}^n (x_i + y_i) = \frac{1}{n} \sum_{i=1}^n x_i + \frac{1}{n} \sum_{i=1}^n y_i = \bar{x} + \bar{y}.$$

22.

$$\begin{aligned}
 \frac{1}{n+m} ((x_1 + \cdots + x_n) + (y_1 + \cdots + y_m)) &= \frac{n}{n+m} \frac{1}{n} (x_1 + \cdots + x_n) \\
 &\quad + \frac{m}{n+m} \frac{1}{m} (y_1 + \cdots + y_m) \\
 &= \frac{n}{n+m} \bar{x} + \frac{m}{n+m} \bar{y}.
 \end{aligned}$$

$$23. \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{n} \sum_{i=1}^n \frac{x_i}{\bar{x}} = \frac{1}{\bar{x}} \cdot \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{\bar{x}} \cdot \bar{x} = 1.$$

24. a) Esto es consecuencia de las leyes de los logaritmos.

$$\begin{aligned}
 b) \operatorname{mg}(ax) &= \sqrt[n]{ax_1 \cdots ax_n} = \sqrt[n]{a^n(x_1 \cdots x_n)} = (a^n)^{1/n} \sqrt[n]{x_1 \cdots x_n} \\
 &= a \cdot \operatorname{mg}(x).
 \end{aligned}$$

$$c) \operatorname{mg}(x/y) = \sqrt[n]{\frac{x_1 \cdots x_n}{y_1 \cdots y_n}} = \frac{\sqrt[n]{x_1 \cdots x_n}}{\sqrt[n]{y_1 \cdots y_n}} = \frac{\operatorname{mg}(x)}{\operatorname{mg}(y)}.$$

- d) La desigualdad de Jensen establece que si  $X$  es una v.a. con esperanza finita y  $\varphi$  es una función cóncava, entonces  $\varphi(E(X)) \geq E(\varphi(X))$ . Tomaremos  $X$  con distribución uniforme sobre los valores  $x_1, \dots, x_n$  y la función cóncava  $\ln x$ . Tenemos que

$$\begin{aligned} \ln \left( \frac{x_1 + \dots + x_n}{n} \right) &\geq \frac{1}{n} \sum_{i=1}^n \ln x_i \\ &= \sum_{i=1}^n \ln x_i^{1/n} \\ &= \ln (x_1^{1/n} \dots x_n^{1/n}). \end{aligned}$$

25. a) Evidente.  
 b) Simplemente multiplique y divida  $\ln(x)$  por  $(x_1 \dots x_n)$ .  
 c) Se usa nuevamente la desigualdad de Jensen como en el último inciso del ejercicio anterior. Tomaremos  $X$  nuevamente con distribución uniforme pero esta vez sobre los valores  $1/x_1, \dots, 1/x_n$  y la función cóncava  $\ln x$ . Tenemos que

$$\begin{aligned} \ln \left( \frac{1/x_1 + \dots + 1/x_n}{n} \right) &\geq \frac{1}{n} \sum_{i=1}^n \ln 1/x_i \\ &= \sum_{i=1}^n \ln (1/x_i)^{1/n} \\ &= \ln ((1/x_1)^{1/n} \dots (1/x_n)^{1/n}). \end{aligned}$$

Omitiendo el logaritmo y tomando el recíproco se obtiene el resultado.

26. a) Verdadero.  
 b) Verdadero.  
 c) Verdadero.  
 d) Falso, puede cambiar.
27. Si  $a = 0$  entonces la colección de datos transformados consta del valor  $c$  repetidas veces. Este valor  $c$  es la moda y la fórmula se cumple. Si  $a \neq 0$  entonces la transformación  $x_i \mapsto ax_i + c$  establece una relación biunívoca entre los datos observados y los datos transformados. En consecuencia, la frecuencia de cada dato observado es la misma frecuencia que la del dato transformado. Por lo tanto, si  $x^*$  es el dato original con mayor frecuencia, entonces  $ax^* + c$  es el dato transformado que tiene mayor frecuencia, es decir, es la moda de los datos transformados.

28. Se usa el símbolo  $\text{Moda}(x)$  para denotar a la posible colección de modas del conjunto de datos  $x$ .

- a)  $\text{Moda}(x) = \{2\}$ .
- b)  $\text{Moda}(x + 2) = \text{Moda}(x) + 2 = \{4\}$ .
- c)  $\text{Moda}(x/2) = (1/2) \cdot \text{Moda}(x) = \{1\}$ .
- d)  $\text{Moda}(x - 2) = \text{Moda}(x) - 2 = \{0\}$ .
- e)  $\text{Moda}(2x) = 2 \cdot \text{Moda}(x) = \{4\}$ .
- f)  $\text{Moda}(4x) = 4 \cdot \text{Moda}(x) = \{8\}$ .

29. Supongamos  $a \geq 0$ . Los datos originales ordenados de menor a mayor  $x_{(1)} \leq \dots \leq x_{(n)}$  se transforman en los datos ordenados

$$ax_{(1)} + c \leq \dots \leq ax_{(n)} + c.$$

Si  $n$  es impar, entonces el dato de en medio de esta nueva colección es

$$ax_{((n+1)/2)} + c = a\tilde{x} + c.$$

Si  $n$  es par, entonces el dato de en medio es

$$\begin{aligned} \frac{(ax_{(n/2)} + c) + (ax_{(n/2+1)} + c)}{2} &= a \frac{x_{(n/2)} + x_{(n/2+1)}}{2} + c \\ &= a\tilde{x} + c. \end{aligned}$$

El mismo resultado se obtiene cuando  $a < 0$ .

30.
  - a)  $\tilde{x} = 20$ .
  - b)  $\tilde{x} = 15$ .
  - c)  $\tilde{x} = 2.5$ .
  - d)  $\tilde{x} = 40$ .
31. Denotemos por  $\text{med}(x)$  a la mediana del conjunto de datos  $x$ . Entonces
- $\text{med}(x) = 5$ ,
  - $\text{med}(x + 4) = 9$ ,
  - $\text{med}(2x + 3) = 13$ ,
  - $\text{med}(x - 2) = 3$ ,
  - $\text{med}(x/2) = 2.5$ ,
  - $\text{med}(5x) = 25$ .

32. a) El mismo dato.  
 b) El mismo dato.  
 c) El promedio de los dos datos distintos.  
 d) El mismo dato.  
 e) El dato de en medio.  
 f) El mismo dato.
33. El dato añadido debe ser la misma mediana.
34. a) Falso. Considere el conjunto de dos datos: 2, 4. La mediana es 3. Si se añaden los datos 2.2 y 3.2, la nueva mediana es 2.7.  
 b) Falso. Considere el conjunto de cuatro datos: 0, 2, 4, 8. La mediana es 3. Si se omiten los datos 0 y 4, los datos ahora son: 2, 8, y la nueva mediana es 5.
35. Para ambos conjuntos de datos la mediana es  $\tilde{x} = 2$ .
36. Para el primer conjunto el total de datos es  $n = 78$ . Por lo tanto, la mediana es  $\tilde{x} = (x_{(39)} + x_{(40)})/2 = (2 + 2)/2 = 2$ . Para el segundo conjunto el total de datos es  $n = 77$ . Por lo tanto, la mediana es el dato de en medio, es decir,  $\tilde{x} = x_{(39)} = 4$ .
37. Falso. Considere el conjunto de dos datos: 0, 2. La mediana es 1. Si se añade un 0 la nueva mediana es 0.
38. 
$$\text{var}(y) = \frac{1}{n} \sum_{i=1}^n ((ax_i + c) - (a\bar{x} + c))^2 = a^2 \cdot \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = a^2 \cdot \text{var}(x).$$
39. Si, cuando todos los datos son idénticos.
40.  $x_1 = -1, x_2 = 5$ .
41. a) 
$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) = \left(\frac{1}{n} \sum_{i=1}^n x_i^2\right) - 2\bar{x}^2 + \bar{x}^2$$

$$= \left(\frac{1}{n} \sum_{i=1}^n x_i^2\right) - \bar{x}^2.$$
- b) 
$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n ((x_i - c) + (c - \bar{x}))^2$$

$$= \left[\frac{1}{n} \sum_{i=1}^n (x_i - c)^2\right] + \frac{2}{n} \sum_{i=1}^n (x_i - c)(c - \bar{x}) + \frac{1}{n} \sum_{i=1}^n (c - \bar{x})^2$$

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

$$\begin{aligned}
42. \quad s_y^2 &= \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n} \sum_{i=1}^n ((ax_i + c) - (a\bar{x} + c))^2 = \frac{1}{n} \sum_{i=1}^n (ax_i - a\bar{x})^2 \\
&= \frac{1}{n} \sum_{i=1}^n a^2 \cdot (x_i - \bar{x})^2 = a^2 \cdot s_x^2.
\end{aligned}$$

43. Puede comprobarse que  $\bar{x} = 1.6$  y  $\text{var}(x) = 1.84$ . Por lo tanto,

$$\begin{aligned}
\overline{x+2} &= 3.6, & \text{var}(x+2) &= 1.84, \\
\overline{x-2} &= -0.4, & \text{var}(x-2) &= 1.84, \\
\overline{2x} &= 3.2, & \text{var}(2x) &= 7.36, \\
\overline{x/2} &= 0.8, & \text{var}(x/2) &= 0.46, \\
\overline{2x+1} &= 4.2, & \text{var}(2x+1) &= 7.36.
\end{aligned}$$

44. Derivando e igualando a cero se encuentra que  $u = \bar{x}$ .

45. a) Verdadero.

b) Verdadero.

c) Falso.

46. a)  $\bar{y} = 3\bar{x} = 6$ ,  $\text{var}(y) = 9\text{var}(x) = 36$ .

b)  $\bar{y} = -\bar{x} + 2 = 0$ ,  $\text{var}(y) = \text{var}(x) = 4$ .

c)  $\bar{y} = \bar{x} - 3 = -1$ ,  $\text{var}(y) = \text{var}(x) = 4$ .

d)  $\bar{y} = -(3/2)\bar{x} = -3$ ,  $\text{var}(y) = (9/4)\text{var}(x) = 9$ .

47. Es inmediato comprobar que  $\bar{y} = 1$ . Por lo tanto,

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

48. Tome raíz cuadrada del resultado para la varianza.

49. Puede comprobarse que  $\bar{x} = 2$  y  $\text{var}(x) = 2$ . Por lo tanto,  $s(x) = \sqrt{2}$ .

Entonces

$$s(x-1) = s(x) = \sqrt{2},$$

$$s(x+2) = s(x) = \sqrt{2},$$

$$s(2x-2) = 2s(x) = 2\sqrt{2}.$$

$$50. \quad a) \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})/s_x = (\bar{x} - \bar{x})/s_x = 0.$$

$$b) \quad s_y^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2/s_x^2 = s_x^2/s_x^2 = 1.$$

51.

$$\begin{aligned} \text{dm}(ax + c) &= \frac{1}{n} \sum_{i=1}^n |(ax_i + c) - (a\bar{x} + c)| \\ &= \frac{1}{n} \sum_{i=1}^n |a| \cdot |x_i - \bar{x}| \\ &= |a| \cdot \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}| \\ &= |a| \cdot \text{dm}(x). \end{aligned}$$

52. Puede comprobarse que  $\bar{x} = 1$  y  $\text{dm}(x) = 2/3$ . Por lo tanto,

$$\text{dm}(x + 1) = \text{dm}(x) = 2/3,$$

$$\text{dm}(x - 2) = \text{dm}(x) = 2/3,$$

$$\text{dm}(2x) = 2 \cdot \text{dm}(x) = 4/3,$$

$$\text{dm}(x/2) = (1/2) \cdot \text{dm}(x) = 1/3,$$

$$\text{dm}(-5x) = 5 \cdot \text{dm}(x) = 10/3.$$

53. Suponga  $a \geq 0$ . Entonces el valor máximo de los datos transformados es  $a x_{(n)} + c$  y el valor mínimo es  $a x_{(1)} + c$ . Esto implica que

$$r(ax + c) = (a x_{(n)} + c) - (a x_{(1)} + c) = a(x_{(n)} - x_{(1)}) = a \cdot r(x).$$

Si  $a < 0$ , el máximo es  $a x_{(1)} + c$  y el mínimo es  $a x_{(n)} + c$ . Por lo tanto,

$$r(ax + c) = (a x_{(1)} + c) - (a x_{(n)} + c) = a(x_{(1)} - x_{(n)}) = -a(x_{(n)} - x_{(1)}) = |a| \cdot r(x).$$

54. El hecho de que los datos estén agrupados no modifica el cálculo del valor máximo ni el valor mínimo. El cálculo del rango es, por lo tanto, el mismo,  $r(x) = x_{(n)} - x_{(1)}$ .

55. a)  $r = n - 1$ .

b)  $r = (n - 1) \cdot |a|$ .

c)  $r = 0$  si  $a = 1$ , y  $r = a^2(1 - a^{n-1})/(1 - a)$  si  $a > 1$ .

d)  $r = 0$  si  $n = 1$ , y  $r = 2$  si  $n \geq 2$ .

56. a) Verdadero. Esto ocurre si, y sólo si, todos los datos son idénticos.  
 b) Falso. Esto nunca ocurre pues  $x_{(1)} \leq x_{(n)}$ .  
 c) Verdadero.  
 d) Falso. Puede tomarse al conjunto de datos  $x$  como  $1, 2, \dots, n$  y al conjunto de datos  $y$  como el valor  $n$  observado  $n$  veces. Entonces  $r(x) = n - 1$ , mientras que  $r(y) = 0$ .

57. Recordemos que  $s(ax + c) = |a| \cdot s(x)$  y que  $\overline{ax + c} = a\bar{x} + c$ . Por lo tanto,

$$cv(ax + c) = \frac{|a| \cdot s(x)}{a\bar{x} + c}.$$

58. a) Verdadero. Esto ocurre si, y sólo si,  $\bar{x} < 0$ .  
 b) Verdadero. Esto ocurre si, y sólo si, todos los datos son idénticos a un valor distinto de cero.
59. Es inmediato comprobar que  $\bar{y} = 1$ . Por lo tanto,

$$\begin{aligned} cv(y) &= \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{x_i}{\bar{x}} - 1\right)^2} \\ &= \sqrt{\frac{1}{n} \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{\bar{x}^2}} \\ &= \frac{1}{|\bar{x}|} \cdot s_x \\ &= \begin{cases} cv(x) & \text{si } \bar{x} > 0, \\ -cv(x) & \text{si } \bar{x} < 0. \end{cases} \end{aligned}$$

60. Sea  $x_1, \dots, x_n$  el conjunto de datos  $x$ . Entonces

$$a) m_1(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^1 = \left(\frac{1}{n} \sum_{i=1}^n x_i\right) - \bar{x} = 0.$$

$$b) m_2(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = s^2(x).$$

$$\begin{aligned} c) m_2(x) &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) = \left(\frac{1}{n} \sum_{i=1}^n x_i^2\right) - \bar{x}^2 \\ &= m'_2 - (m'_1)^2. \end{aligned}$$

61. a) Se usa el teorema del binomio,

$$\begin{aligned}
 m'_k(ax + c) &= \frac{1}{n} \sum_{i=1}^n (ax_i + c)^k \\
 &= a^k \cdot \frac{1}{n} \sum_{i=1}^n (x_i + (c/a))^k \\
 &= a^k \cdot \frac{1}{n} \sum_{i=1}^n \sum_{j=0}^k \binom{k}{j} x_i^j (c/a)^{k-j} \\
 &= a^k \cdot \sum_{j=0}^k \binom{k}{j} \left( \frac{1}{n} \sum_{i=1}^n x_i^j \right) (c/a)^{k-j} \\
 &= a^k \cdot \sum_{j=0}^k \binom{k}{j} m'_j(x) (c/a)^{k-j}.
 \end{aligned}$$

$$\begin{aligned}
 b) \quad m_k(ax + c) &= \frac{1}{n} \sum_{i=1}^n ((ax_i + c) - (a\bar{x} + c))^k = a^k \cdot \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k = \\
 &= a^k \cdot m_k(x).
 \end{aligned}$$

62. Para cada inciso, cualquiera de las dos opciones es correcta.

63. Se omiten las gráficas. El total de observaciones es  $n = 12$ . Algunas diferencias son provocadas por el redondeo en los cálculos.

Valor	Frecuencia	Frecuencia acumulada	Frecuencia relativa	Frecuencia relativa acumulada
A	4	4	0.33	0.33
B	2	6	0.16	0.5
C	3	9	0.25	0.75
D	3	12	0.25	1

64. Se omiten las gráficas. Es importante tomar en cuenta que el total de observaciones es  $n = 20$ .

Valor	Frecuencia	Frecuencia acumulada	Frecuencia relativa	Frecuencia relativa acumulada
A	3	8	0.15	0.15
B	8	11	0.40	0.55
C	4	15	0.20	0.75
D	5	20	0.25	1

65. Un cuantil es el término genérico para aquella cantidad que separa un conjunto de datos numéricos en dos partes, dejando a la izquierda una cierta proporción de ellos y a la derecha la proporción restante. Un cuartil es el caso particular cuando las proporciones son 25% – 75% (primer cuartil), 50% – 50% (segundo cuartil) y 75% – 25% (tercer cuartil).
66. a)  $c(0.25) = -1$ ,  $c(0.50) = 0.5$ ,  $c(0.75) = 2.5$ .  
 b)  $c(0.25) = 2$ ,  $c(0.50) = 2$ ,  $c(0.75) = 3$ .  
 c)  $c(0.25) = 5$ ,  $c(0.50) = 15$ ,  $c(0.75) = 30$ .
67.  $c(0.1) = 0$ ,  $c(0.2) = 0$ ,  $c(0.3) = 3$ ,  $c(0.4) = 5$ ,  $c(0.5) = 5$ ,  
 $c(0.6) = 6$ ,  $c(0.7) = 8$ ,  $c(0.8) = 8$ ,  $c(0.9) = 8$ .
68. a)  $c(0.2) = -1$ ,  $c(0.40) = 0$ ,  $c(0.60) = 1$ ,  $c(0.80) = 2$ .  
 b)  $c(0.2) = 2$ ,  $c(0.40) = 3$ ,  $c(0.60) = 4$ ,  $c(0.80) = 5$ .
- 69.

$$\begin{aligned}
 \text{sk}(ax + c) &= \frac{\left(\frac{1}{n} \sum_{i=1}^n ((ax_i + c) - (a\bar{x} + c))^3\right)}{\left(\frac{1}{n} \sum_{i=1}^n ((ax_i + c) - (a\bar{x} + c))^2\right)^{3/2}} \\
 &= \frac{\left(\frac{1}{n} \sum_{i=1}^n (a(x_i - \bar{x}))^3\right)}{\left(\frac{1}{n} \sum_{i=1}^n (a(x_i - \bar{x}))^2\right)^{3/2}} \\
 &= \frac{a^3}{|a|^3} \cdot \frac{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3\right)}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right)^{3/2}} \\
 &= \frac{a}{|a|} \cdot \text{sk}(x).
 \end{aligned}$$

70. Por cada dato  $x_i$  tal que  $x_i < \bar{x}$  existe otro dato  $x_j$  tal que  $x_j > \bar{x}$  y  $(x_j - \bar{x}) = -(x_i - \bar{x})$ . Por lo tanto,

$$\text{sk} = \frac{1}{s^3} \left( \frac{1}{n} \sum_{x_i < \bar{x}} (x_i - \bar{x})^3 \right) + \frac{1}{s^3} \left( \frac{1}{n} \sum_{x_j > \bar{x}} (x_j - \bar{x})^3 \right) = 0.$$

71.

$$\begin{aligned} k(ax + c) &= \frac{\left( \frac{1}{n} \sum_{i=1}^n ((ax_i + c) - (a\bar{x} + c))^4 \right)}{\left( \frac{1}{n} \sum_{i=1}^n ((ax_i + c) - (a\bar{x} + c))^2 \right)^2} \\ &= \frac{\left( \frac{1}{n} \sum_{i=1}^n (a(x_i - \bar{x}))^4 \right)}{\left( \frac{1}{n} \sum_{i=1}^n (a(x_i - \bar{x}))^2 \right)^2} \\ &= \frac{a^4 \cdot \left( \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4 \right)}{a^4 \cdot \left( \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^2} \\ &= \text{sk}(x). \end{aligned}$$

72. Los momentos centrales impares de la distribución normal son cero. En cambio, los momentos centrales pares tienen la expresión que aparece abajo.

$$E(X - \mu)^{2n} = \frac{(2n)!}{2^n n!} (\sigma^2)^n, \quad n = 1, 2, \dots$$

Tomando  $n = 1$  se obtiene  $\text{Var}(X) = E(X - \mu)^2 = \sigma^2$ , mientras que para  $n = 2$ ,  $E(X - \mu)^4 = 3\sigma^4$ . Entonces  $k = E(X - \mu)^4 / \text{Var}^2(X) = 3$ .

73. Supongamos que tomamos como las marcas de clase los puntos medios de los intervalos. Estos valores son: 0.5, 1.5, 2.5, 3.5, los cuales tienen frecuencias 2, 0, 1, 4, respectivamente. Con esta información se obtiene que

- a)  $\bar{x} = 2.5$ .
- b)  $\text{Moda}(x) = 3.5$ .
- c)  $\tilde{x} = 3.5$ .

74. Gráfica omitida.
75. Gráfica omitida.
76. Gráfica omitida.
77. Gráfica omitida.
78. Gráfica omitida.

79. Gráfica omitida.
80. Gráfica omitida.
81. Gráfica omitida.
82. En un histograma existe un orden entre los valores de la variable a graficar.
83. Gráfica omitida.
84. Gráfica omitida.
85. Gráfica omitida.
86. Gráfica omitida.
87. Gráfica omitida.
88. Gráfica omitida.
89. Gráfica omitida.
90. Gráfica omitida.
91. Gráfica omitida.
92. Gráfica omitida.
93. Gráfica omitida.
94. Gráfica omitida.
95. Gráficas omitidas.
96. Solución omitida.
97. Solución omitida.
- 98.
- |                                                             |                                                        |
|-------------------------------------------------------------|--------------------------------------------------------|
| <i>a</i> ) $\Theta = (0, \infty)$ .                         | <i>d</i> ) $\Theta = (0, \infty) \times (0, \infty)$ . |
| <i>b</i> ) $\Theta = \{1, 2, \dots\} \times (0, 1)$ .       | <i>e</i> ) $\Theta = (0, \infty) \times (0, \infty)$ . |
| <i>c</i> ) $\Theta = \{(a, b) \in \mathbb{R}^2 : a < b\}$ . | <i>f</i> ) $\Theta = (0, \infty) \times (0, \infty)$ . |
99. *a*) Como indica la definición, una estadística es una función de una m.a. que no depende de parámetros desconocidos. Por otro lado, un estimador es una estadística que se usa para estimar un parámetro desconocido. Por lo tanto, la diferencia radica en el uso que se le da a la estadística.
- b*) Falso.



global en  $t = \bar{x}$ . Por lo tanto,  $g(\bar{x}) \leq g(t)$ . Como esto se cumple para cualquier valor de la muestra aleatoria y cualquier valor de  $t$ , la desigualdad se extiende al caso de variables aleatorias.

104. Cada variable  $Y_i$  tiene distribución  $N(0, \frac{n-1}{n^2} \sigma^2)$ . Es suficiente tomar el caso  $i = 1$ .

105. Algunas de estas expresiones pueden tomar valores fuera del espacio parametral.

a)  $\hat{\theta} = 4(2 - \bar{X})$ .

b)  $\hat{\theta} = (3/4)(1 - \bar{X})$ .

c)  $\hat{\theta} = (3/2)(\frac{1}{n} \sum_{i=1}^n X_i^2 - 1)$ .

d)  $\hat{\theta} = 2\bar{X} - 1$ .

e)  $\hat{\theta} = (3\bar{X} - 1)/2$ .

f)  $\hat{\theta} = 1/\bar{X}$ .

g)  $\hat{\theta} = 2\bar{X}$ .

h)  $\hat{\theta} = 3\bar{X}/2$ .

i)  $\hat{\theta} = \bar{X} - 1$ .

j)  $\hat{\theta} = \bar{X}/(1 - \bar{X})$ .

k) El primer momento poblacional se anula y el segundo momento no depende del parámetro. La igualación de los terceros momentos produce el estimador  $\hat{\theta} = (5/n) \sum_{i=1}^n X_i^3$ .

l)  $\hat{\theta} = 3\bar{X}$ . Este es un ejemplo en donde el estimador por momentos puede no tomar valores dentro del espacio parametral.

106. El primer momento poblacional es cero. Usando el segundo momento se obtiene  $\hat{\theta} = \sqrt{2n / \sum_{i=1}^n X_i^2}$ .

107. Puede comprobarse que el primer momento poblacional es  $E(X) = \sqrt{\pi\theta}/2$ . De aquí se obtiene el estimador  $\hat{\theta} = (2\bar{X})^2/\pi$ .

108. Sea  $m_2 = (1/n) \sum_{i=1}^n X_i^2$ . Entonces

a)  $\hat{\theta} = \bar{X}/k$ .

e)  $\hat{\theta} = -a + 2\bar{X}$ .

b)  $\hat{\theta} = \bar{X}/p$ .

f)  $\hat{\theta} = -b + 2\bar{X}$ .

c)  $\hat{\theta} = r/(r + \bar{X})$ .

g)  $\hat{\theta} = \lambda\bar{X}$ .

d)  $\hat{\theta} = p\bar{X}/(1 - p)$ .

h)  $\hat{\theta} = \gamma/\bar{X}$ .

$$\begin{array}{ll}
 i) \hat{\theta} = \bar{X}. & n) \hat{\theta} = \Gamma(1 + 1/\alpha)/\bar{X}. \\
 j) \hat{\theta} = m_2 - \mu^2. & \\
 k) \hat{\theta} = a(1 - \bar{X})/\bar{X}. & \tilde{n}) \hat{\theta} = 2\bar{X}/(\bar{X} - 1). \\
 l) \hat{\theta} = b\bar{X}/(1 - \bar{X}). & \\
 m) \Gamma(1 + 1/\hat{\theta}) = \lambda\bar{X}. & o) \hat{\theta} = \frac{2b^2}{(b-2)(b-4)m_2 - b^2}.
 \end{array}$$

109. Aplique el método de momentos.

110. La estimación por el método de momentos para  $\theta$  es  $\hat{\theta} = 1/(1 + \bar{x}) = 0.3571429$ . Siendo este valor más cercano a 0.4, determinamos que probablemente éste es el valor para  $\theta$  que se usó. Observe que no hay completa seguridad en nuestra afirmación.

111. Aplique el método de momentos.

112. La estimación por el método de momentos para  $\theta$  es  $\hat{\theta} = 1/\bar{x} = 1.86$ . Siendo este valor más cercano a 2, determinamos que probablemente éste es el valor para  $\theta$  que se usó. No hay completa seguridad en nuestra afirmación.

113. Los siguientes cálculos se han obtenido usando R.

a) Usando las fórmulas de los estimadores para los parámetros de esta distribución, tenemos que  $m_1 = (1/n) \sum_{i=1}^n x_i = 1001.25$  y  $m_2 = (1/n) \sum_{i=1}^n x_i^2 = 1003906$ . Por lo tanto,

$$\begin{aligned}
 \hat{\gamma} &= \frac{m_1^2}{m_2 - m_1^2} = 713.683, \\
 \hat{\lambda} &= \frac{m_1}{m_2 - m_1^2} = 0.712792.
 \end{aligned}$$

b) Sea  $X$  una variable aleatoria con distribución gama( $\hat{\gamma}, \hat{\lambda}$ ). La probabilidad buscada es  $P(X > 1000) = 0.5083321$ .

114. Los siguientes cálculos se han obtenido usando R.

Tenemos que  $m_1 = (1/n) \sum_{i=1}^n x_i = 4.05$  y  $m_2 = (1/n) \sum_{i=1}^n x_i^2 = 16.084$ . Por lo tanto,

$$\begin{aligned}
 \hat{a} &= \frac{4m_1^2 - 3m_2}{2m_1 - 1} = -1.142628, \\
 \hat{b} &= \frac{3m_2 - 2m_1}{2m_1 - 1} = 7.622628.
 \end{aligned}$$

115. Este problema puede modelarse de las dos formas siguientes:

- a) Considerando la información completa de cada hora, la distribución de interés es  $\text{bin}(k, \theta)$  con  $k = 10$  y  $\theta$  desconocido. En este caso, el estimador por el método de momentos para  $\theta$  es  $\hat{\theta} = \bar{X}/10$ . Como  $\bar{x} = 8/8 = 1$ , tenemos que  $\hat{\theta} = 1/10$ .
- b) Considerando cada artículo seleccionado, la distribución de interés es  $\text{Ber}(\theta)$ , con  $\theta$  desconocido. El número de observaciones aquí es 80, de las cuales 8 fueron éxitos. Por lo tanto,  $\hat{\theta} = 8/80 = 1/10$ .

116.

- |                                                                     |                                                                        |
|---------------------------------------------------------------------|------------------------------------------------------------------------|
| a) $\hat{\theta} = 4(2 - \bar{X})$ .                                | h) $\hat{\theta} = X_{(n)}$ .                                          |
| b) $\hat{\theta} = \frac{6}{5n}(n - \sum_{i=1}^n 1_{\{0\}}(X_i))$ . | i) $\hat{\theta} = X_{(1)}$ .                                          |
| c) $\hat{\theta} = \frac{3}{2n}(n - \sum_{i=1}^n 1_{\{1\}}(X_i))$ . | j) $\hat{\theta} = -n / \sum_{i=1}^n \ln X_i$ .                        |
| d) $\hat{\theta} = X_{(n)}$ .                                       | k) $\sum_{i=1}^n 1/(1 + \hat{\theta}X_i) = n$<br>(Solución implícita). |
| e) $\hat{\theta} = X_{(n)}$ .                                       | l) No existe.                                                          |
| f) $\hat{\theta} = 1/\bar{X}$ .                                     |                                                                        |
| g) $\hat{\theta} = X_{(n)}$ .                                       |                                                                        |

117. Aplique el método de máxima verosimilitud.

118. La probabilidad a estimar es la función parametral  $\tau(p) = P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - (1 - p)^5 - 5p(1 - p)^4$ . Aplicando el método de máxima verosimilitud se encuentra que la estimación para  $p$  es  $\hat{p} = \bar{x}/5 = 2.12/5 = 0.424$ . Por lo tanto, la estimación para  $\tau(p)$  es  $\tau(\hat{p}) = 1 - (1 - 0.424)^5 - 5(0.424)(1 - 0.424)^4 = 0.703237$ .

119. Se debe recordar que la función de densidad conjunta de las primeras  $k$  estadísticas de orden  $X_{(1)}, \dots, X_{(k)}$ , para  $1 \leq k \leq n$  es, para  $x_1 < \dots < x_k$ ,

$$f_{X_{(1)}, \dots, X_{(k)}}(x_1, \dots, x_k) = \binom{n}{k} k! f(x_1) \cdots f(x_k) [1 - F(x_k)]^{n-k}.$$

Substituyendo las expresiones para  $f(x)$  y  $F(x)$  en el caso  $\exp(\theta)$ , se encuentra la función de verosimilitud  $L(\theta)$ . Maximizando esta función se llega a que el estimador para  $\theta$  es

$$\hat{\theta} = \frac{k}{k\bar{X}_{(k)} + (n - k)X_{(k)}},$$

en donde  $\bar{X}_{(k)} = (X_{(1)} + \dots + X_{(k)})/k$ . Observe que el estimador encontrado se reduce a  $1/\bar{X}$  cuando  $k = n$ .

120. El estimador máximo verosímil para  $\theta$  es  $\hat{\theta} = \bar{X}$ . Por el principio de invarianza tenemos que

$$a) \widehat{\tau(\theta)} = \bar{X}^2.$$

$$b) \widehat{\tau(\theta)} = \bar{X}/(1 - \bar{X}).$$

$$c) \widehat{\tau(\theta)} = \bar{X}(1 - \bar{X}).$$

121.  $\hat{\theta} = \max\{-X_1, \dots, -X_n, \frac{1}{2}X_1, \dots, \frac{1}{2}X_n\}$ .

$$122. \hat{\theta} = \frac{n}{\sum_{i=1}^n |X_i|}.$$

$$123. \hat{\theta} = \frac{1}{n} \sum_{i=1}^n X_i^2.$$

124.  $\hat{\theta} = \bar{X}$ . El estimador para  $P(X > 1)$  es  $e^{-1/\bar{X}}$ .

125.  $\hat{\theta} = (1/\bar{X}) - 2$ . El estimador para  $P(a < X < b)$  es  $e^{-a/\bar{X}} - e^{-b/\bar{X}}$ .

$$126. a) \hat{\theta} = \bar{X}/k.$$

$$f) \hat{\theta} = \bar{X}.$$

$$b) \hat{\theta} = r/(r + \bar{X}).$$

$$g) \hat{\theta} = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2.$$

$$c) \hat{\theta} = X_{(n)}.$$

$$d) \hat{\theta} = X_{(1)}.$$

$$h) \hat{\theta} = \left(\frac{1}{n} \sum_{i=1}^n x_i^\alpha\right)^{-1/\alpha}.$$

$$e) \hat{\theta} = \gamma/\bar{X}.$$

127. La función de verosimilitud es

$$L(\mu_1, \mu_2, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n+m}{2}} \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{i=1}^n (x_i - \mu_1)^2 + \sum_{i=1}^m (y_i - \mu_2)^2\right)\right].$$

Derivando respecto de  $\mu_1$ ,  $\mu_2$  y  $\sigma^2$ , igualando estas derivadas a cero y resolviendo se encuentra que los estimadores son

$$\hat{\mu}_1 = \bar{X},$$

$$\hat{\mu}_2 = \bar{Y},$$

$$\hat{\sigma}^2 = \frac{1}{n+m} \left(\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^m (Y_i - \bar{Y})^2\right).$$

128. La función de verosimilitud es, para  $0 \leq x_1 \leq \dots \leq x_n$  enteros,

$$\begin{aligned} L(\lambda) &= f_{X_{t_1}, \dots, X_{t_n}}(x_1, \dots, x_n) \\ &= f_{X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}}(x_1, x_2 - x_1, \dots, x_n - x_{n-1}) \\ &= f_{X_{t_1}}(x_1) f_{X_{t_2} - X_{t_1}}(x_2 - x_1) \cdots f_{X_{t_n} - X_{t_{n-1}}}(x_n - x_{n-1}) \\ &= e^{-\lambda t_n} \cdot \lambda^{x_n} \cdot \prod_{i=1}^n \frac{(t_i - t_{i-1})^{x_i - x_{i-1}}}{(x_i - x_{i-1})!}. \end{aligned}$$

Se define  $t_0 = 0$  y  $x_0 = 0$ . Tomando logaritmo, derivando respecto de  $\lambda$  e igualando a cero se obtiene  $\lambda = x_n/t_n$ . Calculando la segunda derivada del logaritmo de la función de verosimilitud y evaluando en este valor de  $\lambda$ , se verifica que este punto es un máximo. El estimador máximo verosímil para  $\lambda$  es entonces

$$\hat{\lambda} = \frac{X_{t_n}}{t_n}.$$

129. La función de verosimilitud es

$$\begin{aligned} L(\sigma^2) &= f_{B_{t_1}, \dots, B_{t_n}}(x_1, \dots, x_n) \\ &= f_{B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}}}(x_1, x_2 - x_1, \dots, x_n - x_{n-1}) \\ &= f_{B_{t_1}}(x_1) f_{B_{t_2} - B_{t_1}}(x_2 - x_1) \cdots f_{B_{t_n} - B_{t_{n-1}}}(x_n - x_{n-1}) \\ &= (2\pi\sigma^2)^{n/2} \cdot \prod_{i=1}^n (t_i - t_{i-1})^{-1/2} \cdot \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n \frac{(x_i - x_{i-1})^2}{t_i - t_{i-1}}\right\} \end{aligned}$$

Se define  $t_0 = 0$  y  $x_0 = 0$ . Tomando logaritmo, derivando respecto de  $\sigma^2$  e igualando a cero se obtiene

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n \frac{(x_i - x_{i-1})^2}{t_i - t_{i-1}}.$$

Calculando la segunda derivada del logaritmo de la función de verosimilitud y evaluando en el valor de  $\sigma^2$  anterior, se verifica que este punto es un máximo. Así, el estimador máximo verosímil para  $\sigma^2$  es

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \frac{(B_{t_i} - B_{t_{i-1}})^2}{t_i - t_{i-1}}.$$

130. Tomando esperanza en la desigualdad  $\hat{\theta}_n - \theta \leq |\hat{\theta}_n - \theta|$  y después tomando el límite, se obtiene  $\lim_{n \rightarrow \infty} E(\hat{\theta}_n) \leq \theta$ . Ahora se parte de  $\theta - \hat{\theta}_n \leq |\hat{\theta}_n - \theta|$  y siguiendo el mismo razonamiento se llega a la desigualdad límite contraria. Esto demuestra la igualdad.

131. El resultado es evidente a partir de la observación de que  $X_{(1)} + X_{(2)} = X_1 + X_2$  y que la media muestral es insesgado. Alternativamente, no es difícil comprobar que estas estadísticas de orden tienen distribución Bernoulli. El promedio de sus esperanzas produce el parámetro  $\theta$ .

132.

$$\begin{aligned}
 E(\bar{X}(1 - \bar{X})) &= E(\bar{X} - \bar{X}^2) \\
 &= \theta - E\left[\left(\frac{1}{n} \sum_{i=1}^n X_i\right)\left(\frac{1}{n} \sum_{j=1}^n X_j\right)\right] \\
 &= \theta - \frac{1}{n^2} (n E(X_1^2) + n(n-1) E(X_1 X_2)) \\
 &= \theta - \frac{1}{n} \theta - \frac{n(n-1)}{n^2} \theta^2 \\
 &= \frac{n-1}{n} \theta(1 - \theta).
 \end{aligned}$$

Se propone como estimador insesgado a  $\frac{n}{n-1} \bar{X}(1 - \bar{X})$ .

133. a)  $E(X_1/k) = k\theta/k = \theta$ .

b)  $E((X_1 + \dots + X_n)/kn) = n(k\theta)/kn = \theta$ .

134. a) Es insesgado.

b) Es insesgado.

135. a) Considere el caso  $n = 1$ . Puede usarse la siguiente estimación.

$$\begin{aligned}
 E(\hat{\theta}) &= E\left(\frac{1}{1+X}\right) \\
 &= \sum_{x=0}^{\infty} \frac{1}{1+x} \theta(1-\theta)^x \\
 &= \theta + \sum_{x=1}^{\infty} \frac{1}{1+x} \theta(1-\theta)^x \\
 &> \theta.
 \end{aligned}$$

Considere ahora cualquier valor natural de  $n$ . Puede usarse el segundo inciso de este problema y la siguiente estimación.

$$E\left(\frac{1}{1+\bar{X}}\right) > E\left(\frac{1}{1+\frac{n}{n-1}\bar{X}}\right) = \theta.$$

b) Se usa el hecho de que  $X_1 + \cdots + X_n$  tiene distribución bin neg( $n, \theta$ ).

$$\begin{aligned}
 E(\hat{\theta}) &= E\left(\frac{n-1}{(n-1) + (X_1 + \cdots + X_n)}\right) \\
 &= \sum_{x=0}^{\infty} \frac{n-1}{n-1+x} \binom{n+x-1}{x} \theta^n (1-\theta)^x \\
 &= \theta \sum_{x=0}^{\infty} \binom{(n-1)+x-1}{x} \theta^{n-1} (1-\theta)^x \\
 &= \theta.
 \end{aligned}$$

136. El procedimiento es similar al caso de la distribución geo( $\theta$ ).

a) Si  $r = 1$ , esto se reduce al caso de la distribución geo( $\theta$ ) y el estimador no es insesgado. Suponga  $r \geq 2$ . Usaremos el segundo inciso de este problema y la siguiente estimación.

$$\begin{aligned}
 E\left(\frac{r}{r + \bar{X}}\right) &= 1 - E\left(\frac{\bar{X}}{r + \bar{X}}\right) \\
 &= 1 - E\left(\frac{n\bar{X}}{nr + n\bar{X}}\right) \\
 &< 1 - E\left(\frac{n\bar{X}}{(nr-1) + n\bar{X}}\right) \\
 &= E\left(\frac{nr-1}{(nr-1) + n\bar{X}}\right) \\
 &= E\left(\frac{1}{1 + \frac{n}{nr-1}\bar{X}}\right) \\
 &= \theta.
 \end{aligned}$$

b) Se usa el hecho de que  $X_1 + \cdots + X_n$  tiene distribución bin neg( $nr, \theta$ ).

$$\begin{aligned}
 E(\hat{\theta}) &= E\left(\frac{nr-1}{(nr-1) + (X_1 + \cdots + X_n)}\right) \\
 &= \sum_{x=0}^{\infty} \frac{nr-1}{nr-1+x} \binom{nr+x-1}{x} \theta^{nr} (1-\theta)^x \\
 &= \theta \sum_{x=0}^{\infty} \binom{(nr-1)+x-1}{x} \theta^{nr-1} (1-\theta)^x \\
 &= \theta.
 \end{aligned}$$

137. Observe que  $X_1 + \cdots + X_n$  tiene distribución  $\text{gamma}(n, \theta)$ . Por lo tanto,

$$\begin{aligned} E(\hat{\theta}) &= \int_0^\infty \frac{n}{x} \frac{(\theta x)^{n-1}}{(n-1)!} \theta e^{-\theta x} dx \\ &= \frac{n}{n-1} \theta \int_0^\infty \frac{(\theta x)^{n-2}}{(n-2)!} \theta e^{-\theta x} dx \\ &= \frac{n}{n-1} \theta. \end{aligned}$$

Se propone como estimador insesgado  $(n-1)\hat{\theta}/n = (n-1)/\sum_{i=1}^n X_i$ .

138. Puede comprobarse que la variable aleatoria  $|X_i|$  tiene distribución  $\exp(\theta)$ , y por lo tanto,  $|X_1| + \cdots + |X_n|$  tiene distribución  $\text{gamma}(n, \theta)$ . El cálculo es el mismo que el del ejercicio anterior. Se propone como estimador insesgado  $(n-1)\hat{\theta}/n = (n-1)/\sum_{i=1}^n |X_i|$ .

139. La media muestral resulta ser el estimador insesgado de menor varianza.

- a) Insesgado con varianza  $\sigma^2$ .
- b) No insesgado.
- c) Insesgado con varianza  $\sigma^2/2$ .
- d) No es insesgado.
- e) Insesgado con varianza  $\sigma^2/4$ .
- f) No es insesgado.
- g) Insesgado con varianza  $(7/18)\sigma^2$ .
- h) Insesgado con varianza  $(3/10)\sigma^2$ .

140. Use la propiedad de linealidad de la esperanza.

141. Use integración por partes.

$$E(\hat{\theta}) = E(X^2) = \int_0^\infty x^2 \cdot \frac{2x}{\theta} e^{-x^2/\theta} dx = \cdots = \theta.$$

142. a)  $E(\bar{X}) = \theta + 1$ .

b)  $\bar{X} - 1$  es insesgado.

143. El estimador por máxima verosimilitud es  $\hat{\theta} = -1 - n/\sum_{i=1}^n \ln X_i$ . Se comprueba que  $-\ln X_i$  tiene distribución  $\exp(\theta + 1)$ , y por lo tanto,  $-\sum_{i=1}^n \ln X_i$  tiene distribución  $\text{gamma}(n, \theta + 1)$ . Usando estos resultados se demuestra que

$$E(\hat{\theta}) = \frac{1 + n\theta}{n-1} \neq \theta.$$

De la igualdad anterior se encuentra que un estimador insesgado es  $\hat{\theta}_0 = ((n-1)\hat{\theta} - 1)/n$ .

144. a)  $E(\bar{X}) = E(X_1) = 1 + \theta$ .

b) Por el inciso anterior, se sugiere  $\hat{\theta} = \bar{X} - 1$ .

145. Sea  $X$  una variable aleatoria con la distribución especificada. Puede comprobarse que  $E(X^3) = \theta/5$ . Por lo tanto,  $E(\hat{\theta}) = 5E(X^3) = \theta$ .

146.  $E(\hat{\theta}) = 3E(X) = \theta$ .

147. a) Para cualquier  $x > 0$ ,

$$P(-\ln X_i > x) = P(X_i < e^{-x}) = \int_0^{e^{-x}} \theta u^{\theta-1} du = u^\theta \Big|_0^{e^{-x}} = e^{-\theta x}.$$

b) Por el inciso anterior,  $-\sum_{i=1}^n \ln X_i$  tiene distribución  $\text{gama}(n, \theta)$ . Entonces

$$\begin{aligned} E\left(\frac{n}{-\sum_{i=1}^n \ln X_i}\right) &= n \int_0^\infty \frac{1}{x} \frac{(\theta x)^{n-1}}{(n-1)!} \theta e^{-\theta x} dx \\ &= \frac{n}{n-1} \theta. \end{aligned}$$

c) Por el inciso anterior, el estimador  $-(n-1)/\sum_{i=1}^n \ln X_i$  es insesgado.

148. Por la propiedad de linealidad de la esperanza,

$$\begin{aligned} E(\hat{\theta}) &= \frac{1}{n} \sum_{i=1}^n E(X_i - \mu)^2 \\ &= \frac{1}{n} \sum_{i=1}^n (E(X_i^2) - 2\mu E(X_i) + \mu^2) \\ &= \frac{1}{n} (nE(X_1^2) - 2n\mu E(X_1) + n\mu^2) \\ &= \frac{1}{n} (n(\theta + \mu^2) - 2n\mu^2 + n\mu^2) \\ &= \theta. \end{aligned}$$

149. Por la propiedad de la linealidad de la esperanza,

$$E(\hat{\theta}) = \alpha E(\hat{\theta}_1) + (1 - \alpha) E(\hat{\theta}_2) = \alpha \theta + (1 - \alpha) \theta = \theta.$$

150.  $E(\hat{\theta}) = E(X_1^2) = \theta$ .

151.  $E(T) = E(\varphi_1(X_1) \cdots \varphi_n(X_n)) = E(\varphi_1(X_1)) \cdots E(\varphi_n(X_n))$   
 $= \varphi_1(E(X_1)) \cdots \varphi_n(E(X_n)) = \varphi_1(\theta) \cdots \varphi_n(\theta) = \tau(\theta)$ .

152. Como  $\sum_{i=1}^n X_i$  tiene distribución  $\text{bin}(n, p)$ , se tiene que

$$\begin{aligned} E(\bar{X}(1 - \bar{X})) &= E(\bar{X}) - E[(\bar{X})^2] \\ &= p - (1/n^2) E[(\sum_{i=1}^n X_i)^2] \\ &= p - (1/n^2)(np(1-p) + n^2p^2) \\ &= p - (1/n)p(1-p) - p^2, \end{aligned}$$

lo cual es distinto de  $p(1-p)$ . Sin embargo,  $p - (1/n)p(1-p) - p^2 \rightarrow p(1-p)$  cuando  $n \rightarrow \infty$  y por lo tanto  $\bar{X}(1 - \bar{X})$  es asintóticamente insesgado.

153. Tomando  $a_n = 1/n$  se comprueba que  $\bar{X}$  pertenece a la colección  $\mathcal{E}$ . Cualquier estimador que pertenece a  $\mathcal{E}$  y que es insesgado es tal que  $a_1 + \dots + a_n = 1$ . Por otro lado, la varianza de un estimador en esta colección es  $(a_1^2 + \dots + a_n^2) \text{Var}(X_1)$ . Por lo tanto, encontrar el estimador dentro de esta colección que es insesgado y tiene varianza mínima equivale a minimizar la función  $g(a_1, \dots, a_n) = a_1^2 + \dots + a_n^2$ , sujeto a  $a_1 + \dots + a_n = 1$ . Este problema es equivalente a minimizar la función

$$h(a_1, \dots, a_{n-1}) = a_1^2 + \dots + a_{n-1}^2 + (1 - a_1 - \dots - a_{n-1})^2.$$

Sea  $a_n = 1 - a_1 - \dots - a_{n-1}$ . Derivando respecto de  $a_i$  para  $i = 1, \dots, n-1$  e igualando a cero se encuentra que  $a_i = a_n$ , es decir,  $a_i$  es constante igual a  $1/n$ . Sin mucha dificultad puede demostrarse que se trata de un mínimo comprobando que la matriz hessiana es positiva definida. Véase el apéndice donde se revisa este criterio.

154. Aplique la propiedad de linealidad de la esperanza.

155.  $E(\hat{\lambda}) = \frac{1}{t_n} E(X_{t_n}) = \lambda.$

156.

$$E(\hat{\sigma}^2) = \frac{1}{n} \sum_{i=1}^n \frac{E(B_{t_i} - B_{t_{i-1}})^2}{t_i - t_{i-1}} = \frac{1}{n} \sum_{i=1}^n \frac{\sigma^2(t_i - t_{i-1})}{t_i - t_{i-1}} = \sigma^2.$$

157.  $E(\hat{\theta}^2) = E(\bar{X}^2) = \frac{1}{n^2} (\sum_{i=1}^n E(X_i^2) + \sum_{i \neq j} E(X_i)E(X_j)) = \frac{\theta}{n} + \frac{n-1}{n}\theta^2 \neq \theta^2.$

158. Se puede comprobar que el estimador  $\hat{\theta}_n = \bar{X}(1 - \bar{X})$  no es insesgado pues se cumple la identidad de abajo. A partir de ella es inmediato verificar que este estimador es asintóticamente insesgado.

$$E(\hat{\theta}_n) = \frac{n-1}{n} \theta(1 - \theta).$$

159. Se usan las hipótesis de independencia e idéntica distribución de las variables de la muestra aleatoria.

$$\begin{aligned}
 E(\bar{X}^2) &= \frac{1}{n^2} \sum_{i,j=1}^n E(X_i X_j) \\
 &= \frac{1}{n^2} (nE(X_1^2) + n(n-1)E(X_1)E(X_2)) \\
 &= \frac{1}{n^2} (n(\theta + \theta^2) + n(n-1)\theta^2) \\
 &= \frac{1}{n}\theta + \theta^2 \\
 &\rightarrow \theta^2.
 \end{aligned}$$

160. a)  $E(X_{(n)}) = \int_0^\theta x \cdot \frac{n}{\theta^n} x^{n-1} dx = \frac{n}{n-1} \theta$ . Claramente  $\lim_{n \rightarrow \infty} E(X_{(n)}) = \theta$ .

b) Por el inciso anterior, se sugiere  $\hat{\theta} = \frac{n-1}{n} X_{(n)}$ .

161. Se puede comprobar que el estimador  $\hat{\theta}_n$  no es insesgado pues se cumple la identidad de abajo. A partir de ella es inmediato verificar que este estimador es asintóticamente insesgado.

$$E(\hat{\theta}_n) = \frac{n}{n-1} \theta.$$

162. Desarrollando el cuadrado, tenemos que

$$\begin{aligned}
 E(\bar{X}^2) &= \frac{1}{n^2} \left( \sum_{i=1}^n E(X_i^2) + \sum_{i \neq j} E(X_i X_j) \right) \\
 &= \frac{1}{n^2} (nE(X_1^2) + n(n-1)E(X_1)E(X_2)) \\
 &= \frac{1}{n^2} (nE(X_1^2) + n(n-1)\theta^2) \\
 &\rightarrow \theta^2 \quad \text{cuando } n \rightarrow \infty.
 \end{aligned}$$

163. Como  $\sum_{i=1}^n X_i$  tiene distribución gama( $n, \lambda$ ) y suponiendo  $n \geq 2$ , se tiene que

$$\begin{aligned} E(\hat{\lambda}) &= E(1/\bar{X}) \\ &= E(n/\sum_{i=1}^n X_i) \\ &= \int_0^{\infty} \frac{n}{x} \frac{(\lambda x)^{n-1}}{(n-1)!} \lambda e^{-\lambda x} dx \\ &= \frac{n}{n-1} \lambda \int_0^{\infty} \frac{(\lambda x)^{n-2}}{(n-2)!} \lambda e^{-\lambda x} dx \\ &= \frac{n}{n-1} \lambda. \end{aligned}$$

Esta última cantidad es distinta de  $\lambda$  y por lo tanto  $\hat{\lambda}$  no es insesgado. Sin embargo, el límite de dicha cantidad cuando  $n$  tiende a infinito es  $\lambda$  y de esta manera se cumple la propiedad de insesgamiento asintótico.

164. Puede comprobarse que la función de densidad del estimador  $X_{(1)}$  es

$$f_{X_{(1)}}(x; \theta) = \begin{cases} n e^{-n(x-\theta)} & \text{si } x \geq \theta, \\ 0 & \text{en otro caso.} \end{cases}$$

La esperanza de esta variable aleatoria es  $E(X_{(1)}) = \theta + 1/n$ .

165. En la solución del Ejercicio 147 se demuestra la igualdad que aparece abajo. El insesgamiento asintótico se sigue inmediatamente.

$$E(\hat{\theta}_n) = \frac{n}{n-1} \theta.$$

- 166.

( $\Rightarrow$ ) Suponga  $\hat{\theta}_n \xrightarrow{p} \theta$ . Se necesita demostrar que, para  $x \neq \theta$ ,

$$\lim_{n \rightarrow \infty} F_{\hat{\theta}_n}(x) = \begin{cases} 1 & \text{si } x > \theta, \\ 0 & \text{si } x < \theta. \end{cases}$$

Suponga  $x > \theta$ . Entonces

$$\begin{aligned} F_{\hat{\theta}_n}(x) &= P(\hat{\theta}_n - \theta \leq x - \theta) \\ &= P(|\hat{\theta}_n - \theta| \leq x - \theta) + P(\hat{\theta}_n - \theta < \theta - x) \\ &\rightarrow 1. \end{aligned}$$

La última afirmación se sigue del hecho de que el primer término converge a uno y el segundo término converge a cero pues

$$\begin{aligned} P(\hat{\theta}_n - \theta < \theta - x) &\leq P(\hat{\theta}_n - \theta < \theta - x) + P(\hat{\theta}_n - \theta > x - \theta) \\ &= P(|\hat{\theta}_n - \theta| > x - \theta) \\ &\rightarrow 0. \end{aligned}$$

Suponga ahora  $x < \theta$ . Entonces

$$\begin{aligned} F_{\hat{\theta}_n}(x) &= P(\hat{\theta}_n - \theta \leq x - \theta) \\ &\leq P(\hat{\theta}_n - \theta \leq x - \theta) + P(\hat{\theta}_n - \theta \geq \theta - x) \\ &= P(|\hat{\theta}_n - \theta| \geq \theta - x) \\ &\rightarrow 0. \end{aligned}$$

( $\Leftarrow$ ) Suponga  $\hat{\theta}_n \xrightarrow{d} \theta$ . Sea  $\epsilon > 0$  arbitraria. Entonces

$$\begin{aligned} P(|\hat{\theta}_n - \theta| \leq \epsilon) &= P(-\epsilon \leq \hat{\theta}_n - \theta \leq \epsilon) \\ &= F_{\hat{\theta}_n}(\theta + \epsilon) - P(\hat{\theta}_n - \theta < -\epsilon) \\ &\rightarrow 1. \end{aligned}$$

Claramente el primer término converge a uno por hipótesis y el segundo término converge a cero pues

$$\begin{aligned} P(\hat{\theta}_n - \theta < -\epsilon) &\leq P(\hat{\theta}_n - \theta < -\epsilon) + P(\hat{\theta}_n - \theta > \epsilon) \\ &= P(|\hat{\theta}_n - \theta| > \epsilon) \\ &\rightarrow 0. \end{aligned}$$

167. Sea  $\epsilon > 0$  arbitrario.

$$a) P(|(a\hat{\theta}_n + b) - (a\theta + b)| > \epsilon) = P(|\hat{\theta}_n - \theta| > \epsilon/|a|) \rightarrow 0.$$

b) Se usa la desigualdad  $||a| - |b|| \leq |a - b|$ . Para cualquier  $\epsilon > 0$ ,

$$\begin{aligned} 1 &= \lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| \leq \epsilon) \\ &\leq \lim_{n \rightarrow \infty} P(||\hat{\theta}_n| - |\theta|| \leq \epsilon). \end{aligned}$$

c) Sea  $\epsilon > 0$  arbitraria. Defina los eventos

$$\begin{aligned} A_n &= \{\omega : |\theta_n - \theta| > \epsilon\}, \\ B_n &= \{\omega : |\theta_n^2 - \theta^2| > \epsilon\}. \end{aligned}$$

Entonces

$$\begin{aligned} P(B_n) &= P(B_n \cap A_n) + P(B_n \cap A_n^c) \\ &\leq P(A_n) + P(|\theta_n + \theta| \cdot |\theta_n - \theta| > \epsilon, |\theta_n - \theta| \leq \epsilon). \end{aligned}$$

Ahora analizamos el evento  $(|\theta_n - \theta| \leq \epsilon)$ . Tenemos que

$$\begin{aligned} (|\theta_n - \theta| \leq \epsilon) &= (\theta - \epsilon \leq \theta_n \leq \theta + \epsilon) \\ &= (2\theta - \epsilon \leq \theta_n + \theta \leq 2\theta + \epsilon) \\ &\subseteq (-2|\theta| - \epsilon \leq \theta_n + \theta \leq 2|\theta| + \epsilon) \\ &= (|\theta_n + \theta| \leq 2|\theta| + \epsilon). \end{aligned}$$

Retomando el cálculo de las probabilidades anteriores, tenemos que

$$\begin{aligned} P(B_n) &\leq P(A_n) + P(|\theta_n - \theta| > \epsilon/|\theta_n + \theta|, |\theta_n - \theta| \leq \epsilon) \\ &\leq P(A_n) + P(|\theta_n - \theta| > \epsilon/(2|\theta| + \epsilon), |\theta_n - \theta| \leq \epsilon) \\ &\leq P(A_n) + P(|\theta_n - \theta| > \epsilon/(2|\theta| + \epsilon)) \\ &\rightarrow 0. \end{aligned}$$

d) Sea  $\epsilon > 0$  arbitrario. Defina las cantidades

$$\begin{aligned} \epsilon_1 &= \ln(1 - \epsilon e^{-\theta}) < 0, \\ \epsilon_2 &= \ln(1 + \epsilon e^{-\theta}) > 0. \end{aligned}$$

Estos números se pueden hacer tan pequeños como se desee haciendo  $\epsilon$  suficientemente pequeño, y se definen de la forma anterior pues son tales que

$$\begin{aligned} 0 &= \lim_{n \rightarrow \infty} P(\epsilon_1 \leq \hat{\theta}_n - \theta \leq \epsilon_2) \\ &\quad \vdots \\ &= \lim_{n \rightarrow \infty} P(|e^{\hat{\theta}_n} - e^\theta| \leq \epsilon). \end{aligned}$$

e) Sea  $\epsilon > 0$  arbitraria. Defina los eventos

$$\begin{aligned} A_n &= (|\theta_n - \theta| > \epsilon), \\ B_n &= (|1/\theta_n - 1/\theta| > \epsilon). \end{aligned}$$

Entonces

$$\begin{aligned} P(B_n) &= P(B_n \cap A_n) + P(B_n \cap A_n^c) \\ &\leq P(A_n) + P(|1/\theta_n - 1/\theta| > \epsilon, |\theta_n - \theta| \leq \epsilon). \end{aligned}$$

Ahora analizamos el evento  $(|\theta_n - \theta| \leq \epsilon)$ . Tenemos que

$$\begin{aligned} (|\theta_n - \theta| \leq \epsilon) &= (\theta - \epsilon \leq \theta_n \leq \theta + \epsilon) \\ &= (|\theta|(\theta - \epsilon) \leq \theta_n \cdot |\theta| \leq |\theta|(\theta + \epsilon)) \\ &\subseteq (-|\theta|(|\theta| + \epsilon) \leq \theta_n \cdot |\theta| \leq |\theta|(|\theta| + \epsilon)) \\ &= (|\theta_n \cdot \theta| \leq |\theta|(|\theta| + \epsilon)). \end{aligned}$$

Retomando el cálculo de las probabilidades anteriores, tenemos que

$$\begin{aligned} P(B_n) &\leq P(A_n) + P(|\theta_n - \theta|/|\theta_n \cdot \theta| > \epsilon, |\theta_n - \theta| \leq \epsilon) \\ &= P(A_n) + P(|\theta_n - \theta| > \epsilon \cdot |\theta_n \cdot \theta|, |\theta_n - \theta| \leq \epsilon) \\ &\leq P(A_n) + P(|\theta_n - \theta| > \epsilon/(|\theta|(|\theta| + \epsilon))) \\ &\rightarrow 0. \end{aligned}$$

168. Sea  $\epsilon > 0$  arbitraria. Como  $\varphi$  es continua, existe  $\delta > 0$  tal que si ocurre el evento  $(|\theta_n - \theta| \leq \delta)$ , entonces ocurre el evento  $(|\varphi(\theta_n) - \varphi(\theta)| \leq \epsilon)$ . Defina los eventos  $A_n = (|\theta_n - \theta| > \delta)$  y  $B_n = (|\varphi(\theta_n) - \varphi(\theta)| > \epsilon)$ . Entonces

$$\begin{aligned} P(B_n) &= P(B_n \cap A_n) + P(B_n \cap A_n^c) \\ &\leq P(A_n) + P(|\varphi(\theta_n) - \varphi(\theta)| > \epsilon, |\theta_n - \theta| \leq \delta) \\ &= P(A_n) \\ &\rightarrow 0. \end{aligned}$$

169. Tenemos que

$$\begin{aligned} E(X - a)^2 &= E[(X - a)^2 \cdot 1_{(|X-a|>\epsilon)}] + E[(X - a)^2 \cdot 1_{(|X-a|\leq\epsilon)}] \\ &\geq E[(X - a)^2 \cdot 1_{(|X-a|>\epsilon)}] \\ &\geq \epsilon^2 \cdot E(1_{(|X-a|>\epsilon)}) \\ &= \epsilon^2 \cdot P(|X - a| > \epsilon). \end{aligned}$$

170. Como  $X_1 + \dots + X_n$  tiene distribución  $\text{bin}(n, \theta)$ , se tiene que, cuando  $n \rightarrow \infty$ ,

$$\text{Var}(\bar{X}) = n\theta(1 - \theta)/n^2 \rightarrow 0.$$

Siendo  $\bar{X}$  insesgado, se concluye que este estimador es consistente.

171. Puede comprobarse que el estimador  $\hat{\theta}_n = \text{máx}\{X_1, \dots, X_n\}$  es asintóticamente insesgado pues  $E(\hat{\theta}_n) = n\theta/(n + 1)$ . Por lo tanto, es suficiente comprobar que su varianza tiene a cero cuando  $n \rightarrow \infty$ .

$$\text{Var}(\hat{\theta}_n) = E(\hat{\theta}_n^2) - E^2(\hat{\theta}_n) = \frac{n}{n+2}\theta - \frac{n^2}{(n+1)^2}\theta = \frac{n}{(n+2)(n+1)^2}\theta \rightarrow 0.$$

172. Sabemos que el estimador  $\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  es insesgado para  $\sigma^2$ .

173. Dado que los estimadores son insesgados, por la desigualdad de Chebyshev,

$$a) P(|\hat{\theta}_n - \theta| > \epsilon) \leq \frac{1}{\epsilon^2} \text{Var}(\hat{\theta}_n) = \frac{1}{\epsilon^2} \frac{k_1 + \dots + k_n}{(k_1 + \dots + k_n)^2} \theta(1 - \theta) \rightarrow 0.$$

$$b) P(|\hat{\theta}_n - \theta| > \epsilon) \leq \frac{1}{\epsilon^2} \text{Var}(\hat{\theta}_n) = \frac{1}{\epsilon^2} \frac{k_1 + 4k_2 + \dots + n^2k_n}{(k_1 + 2k_2 + \dots + nk_n)^2} \theta(1 - \theta) \rightarrow 0.$$

174. Por la ley débil de los grandes números,  $\bar{X} \xrightarrow{p} 1/\theta$ . De aquí se obtiene que  $1/\bar{X} \xrightarrow{p} \theta$ .

175. Claramente el estimador es insesgado. Además el estimador no es constante y tiene la misma distribución de probabilidad para cualquier  $n$ , por lo tanto, no puede converger en probabilidad al parámetro.

176. Por la ley débil de los grandes números,  $\bar{X} \xrightarrow{p} 1/\theta$ . Puede comprobarse que  $\hat{\theta} \xrightarrow{p} \theta/(1 + \theta)$ .

177. El estimador por máxima verosimilitud es  $\hat{\theta} = -1 - n/\sum_{i=1}^n \ln X_i$ . Defina la variable aleatoria  $Y_i = -\ln X_i$ . Se puede comprobar que  $Y_i$  tiene distribución  $\exp(\theta+1)$ , y su media, por lo tanto, es  $1/(\theta+1)$ . Por la ley débil de los grandes números,

$$-\frac{1}{n} \sum_{i=1}^n \ln X_i \xrightarrow{p} \frac{1}{\theta + 1}.$$

Tomando inverso multiplicativo y restando uno de cada lado, se demuestra que  $\hat{\theta} \xrightarrow{p} \theta$ .

178. Dado que el estimador es insesgado, por la desigualdad de Chebyshev,

$$P(|\hat{\theta}_n - \theta| > \epsilon) \leq \frac{1}{\epsilon^2} \text{Var}(\hat{\theta}_n) = \frac{1}{\epsilon^2} \frac{9}{n} \text{Var}(X) \rightarrow 0.$$

179. a)

$$\begin{aligned}
 E(T) &= \frac{2 + 4 + \cdots + 2n}{n(n+1)} \mu \\
 &= \frac{2(1 + 2 + \cdots + n)}{n(n+1)} \mu \\
 &= \mu.
 \end{aligned}$$

b) Puede comprobarse que  $T$  tiene distribución normal con media  $\mu$  y varianza

$$\begin{aligned}
 \text{Var}(T) &= \frac{2^2 + 4^2 + \cdots + (2n)^2}{n^2(n+1)^2} \sigma^2 \\
 &= \frac{2^2(1^2 + 2^2 + \cdots + n^2)}{n^2(n+1)^2} \sigma^2 \\
 &= \frac{2^2 \left( \frac{n(n+1)(2n+1)}{n^2(n+1)^2} \right)}{n^2(n+1)^2} \sigma^2 \\
 &= \frac{2(2n+1)}{3n(n+1)} \sigma^2.
 \end{aligned}$$

Conforme  $n$  tiende a infinito, la varianza se aproxima a cero. Esto implica que  $T$  tiende en probabilidad a la media  $\mu$ .

c) Si  $\mu$  es negativa, no puede haber convergencia en probabilidad de  $T_1 = \text{máx}\{0, T\}$  a  $\mu$  pues  $\mu < 0 \leq T_1$ . Supongamos ahora que  $\mu \geq 0$ . Recordemos que la convergencia en probabilidad a una constante es equivalente a la convergencia en distribución a la misma constante. Entonces

$$\begin{aligned}
 P(T_1 \leq x) &= P(\text{máx}\{0, T\} \leq x) \\
 &= P(0 \leq x, T \leq x) \\
 &= P(T \leq x) \cdot 1_{[0, \infty)}(x) \\
 &= \begin{cases} 1 & \text{si } x \geq \mu, \\ 0 & \text{si } x < \mu. \end{cases}
 \end{aligned}$$

Por lo tanto, cuando  $\mu \geq 0$ ,  $T_1$  converge en probabilidad a  $\mu$ , es decir, es un estimador consistente.

180. Recordemos que, en este caso, la varianza muestral  $S^2$  es tal que  $(n-1)S^2/\sigma^2$  tiene distribución  $\chi^2(n-1)$ . Por lo tanto,  $\text{Var}((n-1)S^2/\sigma^2) = 2(n-1)$ .

De donde se obtiene que  $\text{Var}(S^2) = 2\sigma^4/(n-1)$ . Por la desigualdad de Chebyshev,

$$P(|S^2 - \sigma^2| > \epsilon) \leq \frac{\text{Var}(S^2)}{\epsilon^2} = \frac{2\sigma^4}{\epsilon^2(n-1)} \rightarrow 0.$$

181. Sea  $\hat{\theta}_n = X_{(1)}$ . Para cualquier  $\epsilon > 0$ ,

$$\begin{aligned} P(|\hat{\theta}_n - \theta| > \epsilon) &= P(\hat{\theta}_n > \theta + \epsilon) \\ &= (P(X_1 > \theta + \epsilon))^n \\ &= e^{-n\epsilon} \\ &\rightarrow 0, \quad \text{cuando } n \rightarrow \infty. \end{aligned}$$

182. Como el estimador es insesgado, por la desigualdad de Chebyshev, es suficiente demostrar que  $\text{Var}(\hat{\theta}) \rightarrow 0$ . En efecto,

$$P(|\hat{\theta}_n - \theta| > \epsilon) \leq \frac{1}{\epsilon^2} \text{Var}(\hat{\theta}_n) = \frac{1}{\epsilon^2} \frac{25}{n} (E(X^6) - E(X^3)) \rightarrow 0.$$

183. Tome la función convexa  $\varphi(x) = x^2$ .

184. a) Claramente  $E(X_1) = \mu$  pero  $X_1$  es una v.a. que no depende del tamaño de muestra  $n$ , y por lo tanto no converge en probabilidad a  $\mu$ .
- b) Claramente  $E(\bar{X}) = \mu$  y  $\bar{X}$  converge en probabilidad a la constante  $\mu$ . Este último resultado es la ley débil de los grandes números.
- c) Puede comprobarse que  $E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2$ , y por lo tanto  $\hat{\sigma}^2$  no es insesgado para  $\sigma^2$ . Para demostrar la consistencia se usa la siguiente desigualdad de Chebyshev: para cualquier  $\epsilon > 0$  y cualquier número real  $a$ ,

$$P(|X - a| > \epsilon) \leq \frac{1}{\epsilon^2} E(X - a)^2.$$

Por lo tanto,

$$\begin{aligned} P(|\hat{\sigma}^2 - \sigma^2| > \epsilon) &\leq \frac{1}{\epsilon^2} E(\hat{\sigma}^2 - \sigma^2)^2 \\ &= \frac{1}{\epsilon^2} E(\hat{\sigma}^2 - E(\hat{\sigma}^2) - \frac{\sigma^2}{n})^2 \\ &= \frac{1}{\epsilon^2} [\text{Var}(\hat{\sigma}^2) - \frac{\sigma^4}{n^2}] \\ &\rightarrow 0, \end{aligned}$$

cuando  $n \rightarrow \infty$ , pues puede verificarse que

$$\text{Var}(\hat{\sigma}^2) = \frac{2(n-1)}{n^2} \sigma^4.$$

185. Se aplica la desigualdad de Chebyshev: para cualquier  $\epsilon > 0$  y cualquier número real  $a$ ,

$$P(|X - a| > \epsilon) \leq \frac{1}{\epsilon^2} E(X - a)^2.$$

Por lo tanto,

$$P(|\hat{\theta}_n - \theta| > \epsilon) = \frac{1}{\epsilon^2} E(\hat{\theta}_n - \theta)^2 = \frac{1}{\epsilon^2} \text{ECM}(\hat{\theta}_n) \rightarrow 0.$$

186. Recordemos que, en este caso, la varianza muestral  $S^2$  es tal que  $(n-1)S^2/\theta$  tiene distribución  $\chi^2(n-1)$ . Por lo tanto,  $\text{Var}((n-1)S^2/\theta) = 2(n-1)$ . De donde se obtiene que  $\text{Var}(S^2) = 2\theta^2/(n-1)$ . Esta es la expresión para  $\text{ECM}(\hat{\theta}_1)$ , pues  $\hat{\theta}_1$  es insesgado. Por otro lado, como  $\hat{\theta}_2 = ((n-1)/(n+1))\hat{\theta}_1$ ,

$$\begin{aligned} \text{ECM}(\hat{\theta}_2) &= \text{Var}(\hat{\theta}_2) + \left(\frac{n-1}{n+1}\theta - \theta\right)^2 \\ &= \left(\frac{n-1}{n+1}\right)^2 \text{Var}(\hat{\theta}_1) + \frac{4}{(n+1)^2} \theta^2 \\ &= \frac{2(n-1)}{(n+1)^2} \theta^2 + \frac{4}{(n+1)^2} \theta^2 \\ &= \frac{2}{n+1} \theta^2 \\ &< \frac{2}{n-1} \theta^2 \\ &= \text{ECM}(\hat{\theta}_1). \end{aligned}$$

187.  $E(\hat{\theta}) = \theta$ ,  $\text{Var}(\hat{\theta}) = \theta(1-\theta)/n$ ,  $B(\hat{\theta}) = 0$  y  $\text{ECM}(\hat{\theta}) = \theta(1-\theta)/n$ .

188.  $E(\hat{\theta}) = \theta$ ,  $\text{Var}(\hat{\theta}) = \theta/n$ ,  $B(\hat{\theta}) = 0$  y  $\text{ECM}(\hat{\theta}) = \theta/n$ .

189. Claramente  $E(\hat{\theta}) = (1/k)(k\theta) = \theta$ . Además,

$$\text{Var}(\hat{\theta}) = \frac{1}{k^2} \frac{1}{n} k\theta(1-\theta) = \frac{\theta(1-\theta)}{nk}.$$

Se aplica la definición para comprobar que  $\text{CICR}(\theta)$  tiene esta misma expresión.

190. Llevando a cabo las operaciones indicadas, puede comprobarse el cálculo de la esperanza que aparece abajo. De allí se obtiene el resultado buscado.

$$E\left[\left(\frac{\partial}{\partial\theta} \ln f(X, \theta)\right)^2\right] = \frac{1}{\theta^2(1-\theta)}.$$

191. Llevando a cabo las operaciones indicadas, puede comprobarse el cálculo de la esperanza que aparece abajo. De allí se obtiene el resultado buscado.

$$E\left[\left(\frac{\partial}{\partial\theta} \ln f(X, \theta)\right)^2\right] = \frac{r}{\theta^2(1-\theta)}.$$

192. Como  $\tau(\theta) = \theta$ , se tiene que  $\tau'(\theta) = 1$ . Es inmediato comprobar que

$$\frac{\partial}{\partial\theta} \ln f(X; \theta) = \frac{X}{\theta} - 1.$$

Por lo tanto,

$$E\left[\left(\frac{\partial}{\partial\theta} \ln f(X; \theta)\right)^2\right] = \frac{1}{\theta}.$$

La cota inferior de Cramér-Rao es entonces

$$\text{CICR}(\hat{\theta}) = \frac{\theta}{n}, \quad \theta > 0.$$

- a) Claramente  $E(\hat{\theta}) = \theta$  y  $\text{Var}(\hat{\theta}) = \theta$ . Se verifica entonces que

$$\text{CICR}(\hat{\theta}) = \frac{\theta}{n} \leq \theta = \text{Var}(\hat{\theta}).$$

- b)  $\hat{\theta}$  es insesgado para  $\theta$  y su varianza es  $\theta/n$ . Este es un ejemplo en donde la varianza del estimador insesgado alcanza (coincide con) la cota inferior de Cramér-Rao.

193. Como  $\tau(\theta) = \theta$ , se tiene que  $\tau'(\theta) = 1$ . Es inmediato comprobar que

$$\frac{\partial}{\partial\theta} \ln f(X; \theta) = \frac{X - \theta}{\sigma^2}.$$

Por lo tanto,

$$E\left[\left(\frac{\partial}{\partial\theta} \ln f(X; \theta)\right)^2\right] = \frac{1}{\sigma^2}.$$

La cota inferior de Cramér-Rao es entonces

$$\text{CICR}(\hat{\theta}) = \frac{\sigma^2}{n}.$$

194. Claramente  $E(\hat{\theta}) = E(X_1^2) = \theta$ . Además,

$$\text{Var}(\hat{\theta}) = \frac{1}{n} \text{Var}(X_1^2) = \frac{1}{n} (E(X^4) - E^2(X^2)) = \frac{1}{n} (3\theta^2 - \theta^2) = \frac{2}{n} \theta^2.$$

Se aplica la definición para comprobar que  $\text{CICR}(\hat{\theta})$  tiene esta misma expresión. Se usa nuevamente que  $E(X^4) = 3\theta^2$ .

195. La varianza muestral  $S^2$  es tal que  $(n-1)S^2/\sigma^2$  tiene distribución  $\chi^2(n-1)$ . Por lo tanto,  $\text{Var}((n-1)S^2/\theta) = 2(n-1)$ . De donde se obtiene que  $\text{Var}(S^2) = 2\sigma^2/(n-1)$ . Por otro lado, al aplicar la definición de CICR se obtiene que  $\text{CICR}(S^2) = 2\sigma^2/n$ .
196. a) Verdadero.  
b) Falso.
197. Es inmediato comprobar que la media muestral es un estimador insesgado para el parámetro  $\theta$ . Y puede comprobarse que este estimador es eficiente pues su varianza alcanza la cota inferior de Cramér-Rao, es decir, si la varianza de la distribución se denota por  $\sigma^2$ , entonces  $\text{Var}(\bar{X}) = \text{CICR}(\theta) = \sigma^2/n$ . En el Ejercicio 193 se pide demostrar esta expresión para la cota inferior de Cramér-Lundberg. Observe que esta cantidad es constante respecto de  $\theta$ .
198. Sea  $T$  suficiente para  $\theta$  y sea  $S : \text{Rango}(T) \rightarrow \mathbb{R}$  biyectiva. Para cada valor  $t$  de  $T$  existe un único valor  $s$  de  $S \circ T$ . Por lo tanto,

$$\begin{aligned} P(X_1 = x_1, \dots, X_n = x_n \mid S \circ T = s) &= \frac{P(X_1 = x_1, \dots, X_n = x_n, S \circ T = s)}{P(S \circ T = s)} \\ &= \frac{P(X_1 = x_1, \dots, X_n = x_n, T = t)}{P(T = t)} \\ &= P(X_1 = x_1, \dots, X_n = x_n \mid T = t). \end{aligned}$$

Como esta probabilidad no depende de  $\theta$  pues  $T$  es suficiente, se concluye que  $S \circ T$  también es suficiente.

199. Sea  $\tau(\theta)$  una función parametral. Sea  $\theta$  un valor fijo del parámetro y sea  $\tau(\theta)$  su imagen bajo  $\tau$ . La imagen inversa del punto  $\tau(\theta)$  bajo la función  $\tau$  es el conjunto  $\tau^{-1}(\tau(\theta)) = \{\eta : \tau(\eta) = \tau(\theta)\}$ . Este conjunto contiene por lo menos al valor  $\theta$ , pero puede contener muchos otros puntos. Si se elige un punto de cada uno de estos conjuntos se puede tener una función inversa de  $\tau$ , la cual denotaremos también por el símbolo  $\tau^{-1}$ . De esta manera, si  $T$  es una estadística suficiente, entonces por el teorema de factorización se cumple la factorización  $L(x_1, \dots, x_n; \theta) = g(T(x_1, \dots, x_n); \theta) \cdot h(x_1, \dots, x_n)$ , la cual se puede escribir como  $g(T(x_1, \dots, x_n); \tau^{-1}(\tau(\theta))) \cdot h(x_1, \dots, x_n)$ , o bien como  $G(T(x_1, \dots, x_n); \tau(\theta)) \cdot h(x_1, \dots, x_n)$ , en donde  $G$  es una nueva función que depende de los términos indicados. Esto demuestra que  $T$  también es suficiente para  $\tau(\theta)$ .
200. a) Sea  $t$  cualquier posible valor de  $T+a$ . Entonces la distribución conjunta de la muestra aleatoria dado  $T+a = t$  se reduce a la misma distribución

conjunta pero condicionada al evento equivalente  $T = t - a$ . Por ser  $T$  suficiente, esta distribución condicional tampoco depende de  $\theta$ .

- b) Aplique el mismo argumento que en el inciso anterior: si  $t$  es cualquier valor de la estadística  $aT$ , entonces  $(aT = t) = (T = t/a)$ .
- c) Aplique el mismo argumento que en el primer inciso: si  $t$  es cualquier valor de la estadística  $e^T$ , entonces  $(e^T = t) = (T = \ln t)$ .
201. a) Observe que  $T$  tiene distribución  $\text{bin}(nk, \theta)$ . Además, si  $x_1, \dots, x_n$  son los valores de la muestra aleatoria, entonces la estadística  $T$  toma el valor  $t = x_1 + \dots + x_n$ . Así, tenemos que

$$\begin{aligned} & P(X_1 = x_1, \dots, X_n = x_n | T = t) \\ &= \frac{P(X_1 = x_1, \dots, X_n = x_n, T = t)}{P(T = t)} \cdot 1_{\{t\}}(x_1 + \dots + x_n) \\ &= \frac{P(X_1 = x_1, \dots, X_n = x_n)}{P(T = t)} \cdot 1_{\{t\}}(x_1 + \dots + x_n) \\ &= \frac{\binom{k}{x_1} \theta^{x_1} (1 - \theta)^{k-x_1} \dots \binom{k}{x_n} \theta^{x_n} (1 - \theta)^{k-x_n}}{\binom{nk}{t} \theta^t (1 - \theta)^{nk-t}} \cdot 1_{\{t\}}(x_1 + \dots + x_n) \end{aligned}$$

Después de algunas simplificaciones, se comprueba que esta probabilidad condicional no depende de  $\theta$ .

- b) Observe que  $T$  tiene distribución  $\text{bin neg}(n, \theta)$ . Además, si  $x_1, \dots, x_n$  son los valores de la muestra aleatoria, entonces la estadística  $T$  toma el valor  $t = x_1 + \dots + x_n$ . Así, tenemos que

$$\begin{aligned} & P(X_1 = x_1, \dots, X_n = x_n | T = t) \\ &= \frac{P(X_1 = x_1, \dots, X_n = x_n, T = t)}{P(T = t)} \cdot 1_{\{t\}}(x_1 + \dots + x_n) \\ &= \frac{P(X_1 = x_1, \dots, X_n = x_n)}{P(T = t)} \cdot 1_{\{t\}}(x_1 + \dots + x_n) \\ &= \frac{\theta(1 - \theta)^{x_1} \dots \theta(1 - \theta)^{x_n}}{\binom{n+t-1}{t} \theta^n (1 - \theta)^t} \cdot 1_{\{t\}}(x_1 + \dots + x_n) \end{aligned}$$

Después de algunas simplificaciones, se comprueba que esta probabilidad condicional no depende de  $\theta$ .

- c) Observe que  $T$  tiene distribución  $N(n\theta, n\sigma^2)$ . Además, si  $x_1, \dots, x_n$  son los valores de la muestra aleatoria, entonces la estadística  $T$  toma el valor  $t = x_1 + \dots + x_n$ . Así, tenemos que

$$\begin{aligned} & f(x_1, \dots, x_n | T = t) \\ &= \frac{f(x_1, \dots, x_n, T = t)}{f_T(t)} \cdot 1_{\{t\}}(x_1 + \dots + x_n) \\ &= \frac{f(x_1, \dots, x_n)}{f_T(t)} \cdot 1_{\{t\}}(x_1 + \dots + x_n) \\ &= \frac{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_1-\theta)^2/2\sigma^2} \dots \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_n-\theta)^2/2\sigma^2}}{\frac{1}{\sqrt{2\pi n\sigma^2}} e^{-(t-n\theta)^2/2n\sigma^2}} \cdot 1_{\{t\}}(x_1 + \dots + x_n) \end{aligned}$$

Después de algunas simplificaciones, se comprueba que esta probabilidad condicional no depende de  $\theta$ .

- d) Observe que  $T$  tiene distribución gama( $n\gamma, \theta$ ). Además, si  $x_1, \dots, x_n$  son los valores de la muestra aleatoria, entonces la estadística  $T$  toma el valor  $t = x_1 + \dots + x_n$ . Así, tenemos que

$$\begin{aligned} & f(x_1, \dots, x_n | T = t) \\ &= \frac{f(x_1, \dots, x_n, T = t)}{f_T(t)} \cdot 1_{\{t\}}(x_1 + \dots + x_n) \\ &= \frac{f(x_1, \dots, x_n)}{f_T(t)} \cdot 1_{\{t\}}(x_1 + \dots + x_n) \\ &= \frac{\frac{(\theta x_1)^{\gamma-1}}{(\gamma-1)!} \theta e^{-\theta x_1} \dots \frac{(\theta x_n)^{\gamma-1}}{(\gamma-1)!} \theta e^{-\theta x_n}}{\frac{(\theta t)^{n\gamma-1}}{(n\gamma-1)!} \theta e^{-\theta t}} \cdot 1_{\{t\}}(x_1 + \dots + x_n) \end{aligned}$$

Después de algunas simplificaciones, se comprueba que esta probabilidad condicional no depende de  $\theta$ .

202. Se puede usar el teorema de factorización de Neyman. Tenemos que

$$\begin{aligned} L(x_1, \dots, x_n; \theta) &= \frac{1}{\theta^n} 1_{(0,\theta)}(x_1) \dots 1_{(0,\theta)}(x_n) \\ &= \frac{1}{\theta^n} 1_{(0,\theta)}(x_{(n)}) \\ &= g(x_{(n)}; \theta) \cdot 1. \end{aligned}$$

203. Puede comprobarse que  $T$  tiene distribución  $\text{unif}(0, \theta)$ . Por lo tanto,

$$\begin{aligned} f(x_1 | T = t) &= \frac{f(x_1, T = t)}{f_T(t)} \cdot 1_{\{t\}}(x_1) \\ &= \frac{f(x_1)}{f_T(t)} \cdot 1_{\{t\}}(x_1) \\ &= \frac{1/2\theta}{1/\theta} \cdot 1_{\{t\}}(x_1) \\ &= \frac{1}{2} \cdot 1_{\{t\}}(x_1). \end{aligned}$$

204. Claramente  $T$  tiene distribución Bernoulli con probabilidad de éxito igual a  $P(X_1 > 2) = e^{-2\theta}$ . Tomemos  $x_1 = 3$ . Entonces  $T$  toma el valor  $t = 1$ . Por lo tanto,

$$\begin{aligned} f(x_1 | T = t) &= \frac{f(x_1, T = t)}{f_T(t)} \\ &= \frac{f(x_1)}{f_T(t)} \\ &= \frac{\theta e^{-\theta x_1}}{e^{-2t\theta}(1 - e^{-2\theta})^{1-t}} \\ &= \frac{\theta e^{-3\theta}}{e^{-2\theta}} \\ &= \theta e^{-\theta}. \end{aligned}$$

Esta cantidad depende de  $\theta$  y, por lo tanto,  $T$  no es suficiente.

205. Puede usarse el teorema de factorización de Neyman. Tenemos que

$$\begin{aligned} f(x_1, \dots, x_n; \theta) &= \prod_{i=1}^n \theta e^{\theta x_i} \cdot 1_{(0, \infty)}(x_i) \\ &= \theta^n e^{\theta T(x_1, \dots, x_n)} \cdot \prod_{i=1}^n 1_{(0, \infty)}(x_i) \\ &= g(T(x_1, \dots, x_n); \theta) \cdot h(x_1, \dots, x_n). \end{aligned}$$

206. Puede usarse el teorema de factorización de Neyman. Tenemos que

$$\begin{aligned} f(x_1, \dots, x_n; \theta) &= \prod_{i=1}^n e^{-(x_i - \theta)} \cdot 1_{(\theta, \infty)}(x_i) \\ &= e^{n\bar{x}} \cdot e^{n\theta} \cdot 1_{(\theta, \infty)}(x_{(1)}) \\ &= h(x_1, \dots, x_n) \cdot g(x_{(1)}; \theta). \end{aligned}$$

207. Por el teorema de factorización de Neyman,

$$\begin{aligned}
 f_{X_1, \dots, X_n}(x_1, \dots, x_n) &= f_{X_1}(x_1; \theta) \cdots f_{X_n}(x_n; \theta) \\
 &= 2(x_1/\theta)e^{-x_1^2/\theta} \cdots 2(x_n/\theta)e^{-x_n^2/\theta} \\
 &= (2/\theta)^n \left[ \prod_{i=1}^n x_i \right] e^{-(1/\theta) \sum_{i=1}^n x_i^2} \\
 &= (2/\theta)^n e^{-(1/\theta) \sum_{i=1}^n x_i^2} \cdot \left[ \prod_{i=1}^n x_i \right] \\
 &= g(T(x_1, \dots, x_n); \theta) \cdot h(x_1, \dots, x_n).
 \end{aligned}$$

208. Este resultado se puede obtener del teorema de factorización de Neyman,

$$\begin{aligned}
 f(x_1, \dots, x_n; \theta) &= a^n(\theta) b(x_1) \cdots b(x_n) 1_{(0, \theta)}(x_1) \cdots 1_{(0, \theta)}(x_n) \\
 &= b(x_1) \cdots b(x_n) \cdot a^n(\theta) 1_{(0, \theta)}(x_{(n)}) \\
 &= h(x_1, \dots, x_n) \cdot g(x_{(n)}; \theta).
 \end{aligned}$$

209. a)  $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2.$

b) Puede usarse el teorema de factorización de Neyman.

$$\begin{aligned}
 f(x_1, \dots, x_n; \theta) &= (2\pi\theta)^{-n/2} \exp\left\{-\sum_{i=1}^n (x_i - \mu)^2/2\theta\right\} \\
 &= (2\pi\theta)^{-n/2} \exp\{-n\hat{\theta}(x_1, \dots, x_n)/2\theta\} \\
 &= g(\hat{\theta}(x_1, \dots, x_n); \theta) \cdot 1.
 \end{aligned}$$

210. Observe que la función de densidad de  $T$  es

$$f_T(t) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}} [e^{-(t-\theta)^2/2\sigma^2} + e^{-(-t-\theta)^2/2\sigma^2}] & \text{si } t \geq 0, \\ 0 & \text{en otro caso.} \end{cases}$$

Entonces, para  $t \geq 0$ ,

$$\begin{aligned}
 f(x_1 | T = t) &= \frac{f(x_1, T = t)}{f_T(t)} \\
 &= \frac{f(x_1)}{f_T(t)} \cdot 1_{\{t\}}(x_1) \\
 &= \frac{f_{X_1}(t)}{f_T(t)} \\
 &= \frac{e^{-(t-\theta)^2/2\sigma^2}}{e^{-(t-\theta)^2/2\sigma^2} + e^{-(-t-\theta)^2/2\sigma^2}}.
 \end{aligned}$$

Esta expresión depende de  $\theta$  cuando  $t > 0$ . Por lo tanto,  $T$  no es suficiente.

211. Tome, por ejemplo,  $(x_1, x_2, x_3, x_4) = (0, 1, 1, 0)$ . Entonces  $T(x_1, x_2, x_3, x_4) = 0$ . Además,  $P(T = 0) = P(X_1 = 0, X_4 = 0) + P(X_2 = 0, X_3 = 0, X_4 = 0) = (1 - \theta)^2 + (1 - \theta)^3$ . Por lo tanto,

$$\begin{aligned} & P(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0 | T = 0) \\ &= \frac{P(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0, T = 0)}{P(T = 0)} \\ &= \frac{P(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0)}{P(T = 0)} \\ &= \frac{\theta^2(1 - \theta)^2}{(1 - \theta)^2 + (1 - \theta)^3} \\ &= \frac{\theta^2}{2 - \theta}. \end{aligned}$$

Claramente esta probabilidad depende del parámetro  $\theta$  y, por lo tanto,  $T$  no es suficiente para  $\theta$ .

212. Tome, por ejemplo,  $(x_1, x_2) = (0, 0)$ . Entonces  $T(x_1, x_2) = 0$ . Además,  $P(T = 0) = \sum_{n=0}^{\infty} P(X_1 = n, X_2 = n) = \sum_{n=0}^{\infty} e^{-2\theta} \theta^{2n} / (n!)^2$ . Por lo tanto,

$$\begin{aligned} P(X_1 = 0, X_2 = 0 | T = 0) &= \frac{P(X_1 = 0, X_2 = 0, T = 0)}{P(T = 0)} \\ &= \frac{P(X_1 = 0, X_2 = 0)}{P(T = 0)} \\ &= \frac{e^{-2\theta}}{e^{-2\theta} \sum_{n=0}^{\infty} \theta^{2n} / (n!)^2} \\ &= \frac{1}{\sum_{n=0}^{\infty} \theta^{2n} / (n!)^2}. \end{aligned}$$

Esta expresión depende del parámetro  $\theta$  y, por lo tanto,  $T$  no es suficiente para  $\theta$ .

213. Observe que  $T$  tiene distribución  $N(3\theta, 5)$ . Tome, por ejemplo,  $(x_1, x_2) = (0, 0)$ . Entonces  $T(x_1, x_2) = t = 0$ . Por lo tanto, para estos valores de la

muestra aleatoria,

$$\begin{aligned}
 f(x_1, x_2 | T = t) &= \frac{f(x_1, x_2, T = t)}{f_T(t)} \\
 &= \frac{f(x_1, x_2)}{f_T(t)} \\
 &= \frac{(2\pi)^{-1} e^{-\theta^2}}{(10\pi)^{-1/2} e^{-9\theta^2/10}} \\
 &= \sqrt{\frac{5}{2}} e^{-\theta^2/10}.
 \end{aligned}$$

Esta expresión depende del parámetro  $\theta$  y, por lo tanto,  $T$  no es suficiente para  $\theta$ .

214. a) La afirmación es evidente pues se trata de la esperanza de una variable aleatoria positiva.
- b) Si  $a \neq 0$ , entonces puede comprobarse que  $f_{aX}(x) = (1/a) f_X(x/a)$ . Por lo tanto,  $f_{aX}(aX) = (1/a) f_X(X)$ . Entonces

$$\begin{aligned}
 I_{aX}(\theta) &= E\left[\left(\frac{\partial}{\partial\theta} \log \frac{1}{a} + \frac{\partial}{\partial\theta} \log f_X(X; \theta)\right)^2\right] \\
 &= E\left[\left(\frac{\partial}{\partial\theta} \log f_X(X; \theta)\right)^2\right] \\
 &= I_X(\theta).
 \end{aligned}$$

- c) Puede comprobarse que  $f_{X+b}(x) = f_X(x - b)$ . Por lo tanto,  $f_{X+b}(X + b) = f_X(X)$ . De esta identidad se sigue el resultado pues

$$\begin{aligned}
 I_{X+b}(\theta) &= E\left[\left(\frac{\partial}{\partial\theta} \log f_{X+b}(X + b; \theta)\right)^2\right] \\
 &= E\left[\left(\frac{\partial}{\partial\theta} \log f_X(X; \theta)\right)^2\right] \\
 &= I_X(\theta).
 \end{aligned}$$

215.

$$\begin{aligned}
 I(\eta) &= E\left[\left(\frac{\partial}{\partial\eta} \log f(X, \varphi(\eta))\right)^2\right] \\
 &= E\left[\left(\frac{\partial\varphi(\eta)}{\partial\eta} \frac{\partial}{\partial\varphi(\eta)} \log f(X, \varphi(\eta))\right)^2\right] \\
 &= (\varphi'(\eta))^2 \cdot E\left[\left(\frac{\partial}{\partial\varphi(\eta)} \log f(X, \varphi(\eta))\right)^2\right] \\
 &= (\varphi'(\eta))^2 \cdot I(\theta) \Big|_{\theta=\varphi(\eta)}.
 \end{aligned}$$

216. Observe que  $(\varphi'(\eta))^2 = \varphi^2(\eta)(1 - \varphi(\eta))^2$ . Ahora aplique la fórmula del ejercicio anterior recordando que  $I(\theta) = 1/(\theta(1 - \theta))$ . Alternativamente, obtenga el resultado de la expresión

$$I(\eta) = E \left[ \left( \frac{\partial}{\partial \eta} \log f(X, \varphi(\eta)) \right)^2 \right].$$

217. Por la propiedad de incrementos independientes, tenemos que, para valores enteros  $0 \leq x_1 \leq x_2 \leq \dots \leq x_n$ , la función de probabilidad del vector  $(X_{t_1}, \dots, X_{t_n})$  evaluada en  $(x_1, \dots, x_n)$  es

$$\begin{aligned} f(x_1, \dots, x_n) &= P(X_{t_1} = x_1, \dots, X_{t_n} = x_n) \\ &= P(X_{t_1} = x_1, X_{t_2} - X_{t_1} = x_2 - x_1, \dots, X_{t_n} - X_{t_{n-1}} = x_n - x_{n-1}) \\ &= P(X_{t_1} = x_1)P(X_{t_2} - X_{t_1} = x_2 - x_1) \cdots P(X_{t_n} - X_{t_{n-1}} = x_n - x_{n-1}) \\ &= e^{-\theta t_1} \frac{(\theta t_1)^{x_1}}{x_1!} \cdot e^{-\theta(t_2 - t_1)} \frac{(\theta(t_2 - t_1))^{x_2 - x_1}}{(x_2 - x_1)!} \cdots \\ &\quad \cdots e^{-\theta(t_n - t_{n-1})} \frac{(\theta(t_n - t_{n-1}))^{x_n - x_{n-1}}}{(x_n - x_{n-1})!} \\ &= e^{-\theta t_n} \theta^{x_n} \frac{t_1^{x_1} (t_2 - t_1)^{x_2 - x_1} \cdots (t_n - t_{n-1})^{x_n - x_{n-1}}}{x_1! (x_2 - x_1)! \cdots (x_n - x_{n-1})!}. \end{aligned}$$

Por lo tanto,

$$\frac{\partial}{\partial \theta} \log f(X_1, \dots, X_n; \theta) = -t_n + \frac{1}{\theta} X_{t_n}.$$

Al hacer el resto de las operaciones se obtiene que  $I(\theta) = t_n/\theta$ . Alternativamente, observe que las variables aleatorias incremento  $X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$  son independientes, con distribución reparametrizada Poisson( $\theta(t_i - t_{i-1})$ ), respectivamente. Defina  $t_0 = 0$ . Entonces, usando la fórmula para la información de Fisher cuando se tiene una reparametrización, se tiene que

$$\begin{aligned} I(\theta) &= t_1^2 \frac{1}{\theta t_1} + (t_2 - t_1)^2 \frac{1}{\theta(t_2 - t_1)} + \cdots + (t_n - t_{n-1})^2 \frac{1}{\theta(t_n - t_{n-1})} \\ &= \frac{t_n}{\theta}. \end{aligned}$$

218. Por la propiedad de incrementos independientes, tenemos que

$$\begin{aligned}
 & f(x_1, \dots, x_n) \\
 &= f_{X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}}(x_1, x_2 - x_1, \dots, x_n - x_{n-1}) \\
 &= f_{X_{t_1}}(x_1) f_{X_{t_2} - X_{t_1}}(x_2 - x_1) \cdots f_{X_{t_n} - X_{t_{n-1}}}(x_n - x_{n-1}) \\
 &= \frac{1}{\sqrt{2\pi\theta t_1}} e^{-x_1^2/(2\theta t_1)} \cdots \frac{1}{\sqrt{2\pi\theta(t_n - t_{n-1})}} e^{-(x_n - x_{n-1})^2/(2\theta(t_n - t_{n-1}))} \\
 &= \left(\frac{1}{2\pi}\right)^{n/2} \theta^{-n/2} \frac{1}{\sqrt{t_1(t_2 - t_1) \cdots (t_n - t_{n-1})}} \\
 &\quad \cdot \exp\left\{-\frac{1}{2\theta} \left(\frac{x_1^2}{t_1} + \cdots + \frac{(x_n - x_{n-1})^2}{t_n - t_{n-1}}\right)\right\}
 \end{aligned}$$

Por lo tanto,

$$\frac{\partial}{\partial \theta} \log f(X_1, \dots, X_n; \theta) = -\frac{n}{2\theta} + \frac{1}{2\theta} \left[ \left(\frac{X_1}{\sqrt{\theta t_1}}\right)^2 + \cdots + \left(\frac{X_n - X_{n-1}}{\sqrt{\theta(t_n - t_{n-1})}}\right)^2 \right].$$

La suma de los cuadrados que aparece en esta expresión es una variable aleatoria con distribución  $\chi^2(n)$ , y por lo tanto su media es  $n$  y su segundo momento es  $n(n+2)$ . Al hacer el resto de las operaciones se obtiene que  $I(\theta) = n/(2\theta^2)$ . Alternativamente, observe que las variables aleatorias incremento  $X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$  son independientes, con distribución reparametrizada  $N(0, \theta(t_i - t_{i-1}))$ , respectivamente. Defina  $t_0 = 0$ . Entonces, usando la fórmula para la información de Fisher cuando se tiene una reparametrización, se tiene que

$$\begin{aligned}
 I(\theta) &= t_1^2 \frac{1}{2\theta^2 t_1^2} + (t_2 - t_1)^2 \frac{1}{2\theta^2 (t_2 - t_1)^2} \\
 &\quad + \cdots + (t_n - t_{n-1})^2 \frac{1}{2\theta^2 (t_n - t_{n-1})^2} \\
 &= \frac{n}{2\theta^2}.
 \end{aligned}$$

219. En cada caso aplique la definición de información de Fisher.

220. Para  $\theta > 0$ ,

$$a) I(\theta) = -E\left[\frac{\partial^2}{\partial \theta^2} (\ln \theta + (\theta - 1) \ln X)\right] = 1/\theta^2.$$

$$b) I(\theta) = -E\left[\frac{\partial^2}{\partial \theta^2} (\ln \theta - \theta |X|)\right] = 1/\theta^2.$$

221. Para  $-1 < \theta < 1$ ,

$$\begin{aligned} I(\theta) &= -E\left[\frac{\partial^2}{\partial\theta^2} \ln(1 + \theta X)\right] \\ &= E\left[\frac{X^2}{(1 + \theta X)^2}\right] \\ &= \frac{1}{2} \int_{-1}^1 \frac{x^2}{1 + \theta x} dx \\ &= \frac{1}{2\theta^3} \left(\ln\left(\frac{1 + \theta}{1 - \theta}\right) - 2\theta\right). \end{aligned}$$

222. Puede comprobarse que  $E(X^2) = 1$ . Por lo tanto, para  $\theta > 0$ ,

$$\begin{aligned} I(\theta) &= -E\left[\frac{\partial^2}{\partial\theta^2} \left(-\ln\theta - \frac{X^2}{\theta}\right)\right] \\ &= -E\left[\frac{1}{\theta^2} - \frac{2}{\theta^3} X^2\right] \\ &= \frac{2 - \theta}{\theta^3}. \end{aligned}$$

223. a) Para la distribución  $N(\theta, \sigma^2)$  puede demostrarse que  $I(\theta)$  es la función constante  $1/\sigma^2$ . Por lo tanto, la información de Fisher de la muestra es  $I_{X_1, \dots, X_n}(\theta) = n \cdot I_{X_1}(\theta) = n/\sigma^2$ .
- b) Para la distribución  $\text{gama}(\gamma, \theta)$  puede demostrarse que  $I(\theta) = \gamma/\theta^2$ . Por lo tanto, la información de Fisher de la muestra es  $I_{X_1, \dots, X_n}(\theta) = n \cdot I_{X_1}(\theta) = n\gamma/\theta^2$ .
224. a) Puede comprobarse que si  $X$  tiene la distribución Rayleigh indicada, entonces  $X^2$  tiene distribución  $\exp(1/\theta)$ . Por lo tanto,  $T$  tiene distribución  $\text{gama}(n, 1/\theta)$ . Su información de Fisher es entonces  $I_T(\theta) = n/\theta^2$ . Observe que aquí se aplica el resultado del ejercicio 215 sobre la reparametrización de una distribución. Por otro lado, puede comprobarse que la información de Fisher de una v.a. con la distribución Rayleigh indicada es  $I_{X_1}(\theta) = 1/\theta^2$ . Por lo tanto, la información de Fisher de la muestra aleatoria es  $I_{X_1, \dots, X_n}(\theta) = n/\theta^2$ . Como  $I_T(\theta) = I_{X_1, \dots, X_n}(\theta)$ , se concluye que  $T$  es suficiente para  $\theta$ .
- b) Recordemos que la información de Fisher de una v.a. con distribución  $N(\theta, \sigma^2)$  es  $I(\theta) = 1/\sigma^2$ . Puede comprobarse que  $T$  tiene distribución  $N(3\theta, 5)$ . Por lo tanto,  $I_T(\theta) = 1/5$ . Por otro lado, la información de Fisher de la muestra aleatoria es  $I_{X_1, X_2}(\theta) = 2 \cdot I_{X_1}(\theta) = 2$ . Como  $I_T(\theta) \neq I_{X_1, X_2}(\theta)$ , se concluye que  $T$  no es suficiente para  $\theta$ .

c) Solución omitida.

225. Por simplicidad en la escritura consideremos el caso discreto. Sea  $t = (t_1, \dots, t_n)$  un valor cualquiera del vector de estadísticas  $T = (X_1, \dots, X_n)$ . Entonces

$$\begin{aligned} P(X_1 = x_1, \dots, X_n = x_n | T = t) \\ &= P(X_1 = x_1, \dots, X_n = x_n | X_1 = t_1, \dots, X_n = t_n) \\ &= 1_{\{t_1\}}(x_1) \cdots 1_{\{t_n\}}(x_n). \end{aligned}$$

Esta cantidad no depende de  $\theta$ , por lo tanto  $T$  es suficiente.

226. Por simplicidad en la escritura consideremos el caso discreto. Sea  $t = (t_1, \dots, t_n)$  un valor cualquiera del vector de estadísticas de orden  $T = (X_{(1)}, \dots, X_{(n)})$ . Observe que necesariamente  $t_1 \leq t_2 \leq \dots \leq t_n$ . Entonces

$$\begin{aligned} P(X_1 = x_1, \dots, X_n = x_n | T = t) \\ &= P(X_1 = x_1, \dots, X_n = x_n | X_{(1)} = t_1, \dots, X_{(n)} = t_n) \\ &= 1_{\{t_1\}}(x_{(1)}) \cdots 1_{\{t_n\}}(x_{(n)}). \end{aligned}$$

Esta cantidad no depende de  $\theta$ , por lo tanto  $T$  es suficiente.

227. Se puede usar el teorema de factorización. Omitiendo los factores  $1_{\{0,1\}}(x_i)$  en el lado derecho, tenemos que

$$\begin{aligned} P(X_1 = x_1, \dots, X_n = x_n) &= \theta^{x_1} (1 - \theta)^{1-x_1} \cdots \theta^{x_n} (1 - \theta)^{1-x_n} \\ &= \theta^{x_1 + \cdots + x_n} (1 - \theta)^{n-x_1 - \cdots - x_n} \\ &= \theta^{x_1 + \cdots + x_k} (1 - \theta)^{k-x_1 - \cdots - x_k} \\ &\quad \cdot \theta^{x_{k+1} + \cdots + x_n} (1 - \theta)^{(n-k)-x_{k+1} - \cdots - x_n} \\ &= g(T_1, T_2, \theta) h(x_1, \dots, x_n). \end{aligned}$$

228. Demostración omitida.

229. Usando el teorema de factorización, la función  $g$  es constante respecto de la estadística añadida  $T_{k+1}$ .

230. a) No lo es.  
 b) No lo es.  
 c) No lo es.  
 d) Lo es.  
 e) No lo es.  
 f) Lo es.

- g) Lo es.  
 h) No lo es.  
 i) Lo es.  
 j) No lo es.
231. a) Sean  $x_1, \dots, x_n$  y  $y_1, \dots, y_n$  dos valores de la m.a. tales que  $S(x_1, \dots, x_n) = S(y_1, \dots, y_n)$ . Como  $T$  es función de  $S$ ,  $T(x_1, \dots, x_n) = T(y_1, \dots, y_n)$ . Y como  $U$  es función de  $T$ ,  $U(x_1, \dots, x_n) = U(y_1, \dots, y_n)$ .  
 b) Evidente.  
 c)  $T = X_1 + X_2$  es función de  $S = X_1$ , pero  $S$  no es función de  $T$ .

232. Se puede usar el teorema de factorización. Como  $T$  es suficiente para  $\theta$ ,

$$f(x_1, \dots, x_n; \theta) = g(T(x_1, \dots, x_n); \theta) h(x_1, \dots, x_n).$$

Por lo tanto, el valor  $\hat{\theta}$  que maximiza a la función  $f(x_1, \dots, x_n; \theta)$  también maximiza a  $g(T(x_1, \dots, x_n); \theta)$ . Y de esta manera  $\hat{\theta}$  depende de  $x_1, \dots, x_n$  sólo a través de  $T(x_1, \dots, x_n)$ .

233. Solución omitida.

234. Solución omitida.

235. a) Se comprueba que  $S(0, 0, 1) = S(0, 0, 0)$  y sin embargo,  $T(0, 0, 1) \neq T(0, 0, 0)$ . Por lo tanto,  $T$  no es función de  $S$ , y en consecuencia,  $S$  no es suficiente.  
 b) Se comprueba que  $S(0, 0, 1) = S(1, 1, 0)$  y sin embargo,  $T(0, 0, 1) \neq T(1, 1, 0)$ . Por lo tanto,  $T$  no es función de  $S$ , y en consecuencia,  $S$  no es suficiente.  
 c) Se comprueba que  $S(0, 1, 0) = S(1, 0, 1)$  y sin embargo,  $T(0, 1, 0) \neq T(1, 0, 1)$ . Por lo tanto,  $T$  no es función de  $S$ , y en consecuencia,  $S$  no es suficiente.

236. Por contradicción. Supongamos que  $S$  es suficiente. Como  $T$  es suficiente minimal,  $T$  es función de  $S$ . Sin embargo,  $S(0, 1, 1, 0) = S(0, 0, 0, 0)$  y  $T(0, 1, 1, 0) \neq T(0, 0, 0, 0)$ .

237. Omitiendo el soporte de la distribución, puede verificarse que

$$\frac{f(x_1, \dots, x_n; \theta)}{f(y_1, \dots, y_n; \theta)} = (1 - \theta)^{(x_1 + \dots + x_n) - (y_1 + \dots + y_n)}.$$

De modo que esta expresión no depende de  $\theta \Leftrightarrow T(x_1, \dots, x_n) = T(y_1, \dots, y_n)$ . Esta es la condición del teorema 2.6 para concluir que  $T$  es suficiente minimal.

238. Por contradicción. Supongamos que  $S$  es suficiente. Como  $T$  es suficiente minimal,  $T$  es función de  $S$ . Sin embargo,  $S(0,0) = S(1,1)$  y  $T(0,0) \neq T(1,1)$ .

239. Puede verificarse que

$$\frac{f(x_1, \dots, x_n; \theta)}{f(y_1, \dots, y_n; \theta)} = \frac{1_{(0,\theta)}(x_1) \cdots 1_{(0,\theta)}(x_n)}{1_{(0,\theta)}(y_1) \cdots 1_{(0,\theta)}(y_n)} = \frac{1_{(x_{(n)}, \infty)}(\theta)}{1_{(y_{(n)}, \infty)}(\theta)}.$$

Esta expresión es idénticamente 1 y no depende de  $\theta \Leftrightarrow x_{(n)} = y_{(n)}$ , es decir,  $T(x_1, \dots, x_n) = T(y_1, \dots, y_n)$ . Esta es la condición del teorema 2.6 para concluir que  $T$  es suficiente minimal.

240. Solución omitida.

241. Solución omitida.

242. Por contradicción. Supongamos que  $S$  es suficiente. Como  $T$  es suficiente minimal,  $T$  es función de  $S$ . Sin embargo,  $S(0,0) = S(2,-1)$  y  $T(0,0) \neq T(2,-1)$ .

243. a) Se comprueba que  $f_{X_1, \dots, X_n | T_1, T_2}(x_1, \dots, x_n | t_1, t_2)$  no depende de  $\theta$ . Esta función es igual a

$$\frac{f_{X_1, \dots, X_n}(x_1, \dots, x_n)}{f_{T_1, T_2}(t_1, t_2)},$$

cuando  $x_1, \dots, x_n$  son tales que  $T_1 = t_1$  y  $T_2 = t_2$ . El numerador es

$$\left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\sum_{i=1}^n (x_i - \theta)^2 / 2\sigma^2\right\}.$$

El denominador es

$$\frac{1}{\pi n \sigma^2} \exp\left\{-\frac{(t_1 - (n/2)\theta)^2}{n\sigma^2} - \frac{(t_2 - (n/2)\theta)^2}{n\sigma^2}\right\}.$$

Se desarrollan estos exponentes tanto en el numerador como en el denominador y se observa que los coeficientes de  $\theta$  y de  $\theta^2$  coinciden arriba y abajo. De esta manera tales términos desaparecen y se demuestra así la no dependencia de  $\theta$  de esta función.

b) Puesto que  $(T_1, T_2)$  no es función de la estadística suficiente  $T = X_1 + \cdots + X_n$ , pues un valor de  $T$  no puede determinar los valores de  $T_1$  y  $T_2$ , se concluye que  $(T_1, T_2)$  no puede ser suficiente minimal.

244. a) Tómesese  $G = \Omega$  en la tercera condición de la definición de esperanza condicional.  
 b) Compruebe que  $X$  satisface las tres condiciones de la definición de esperanza condicional

245. a)  $E(c | X) = c$ .  
 b)  $E(X | c) = E(X)$ .  
 c)  $E(cX | X) = cX$ .  
 d)  $E(X | cX) = X$  ( $c \neq 0$ ).  
 e)  $E(X + c | X) = X + c$ .  
 f)  $E(X | X + c) = X$ .

246. En este caso la esperanza condicional  $E(X | Y)$  es discreta y adquiere la expresión

$$E(X | Y) = \sum_y E(X | Y = y) \cdot 1_{(Y=y)}.$$

Por lo tanto,

$$E(E(X | Y)) = \sum_y E(X | Y = y) P(Y = y) = E(X).$$

247.

$$E(X | Y) = \sum_y E(X | Y = y) \cdot 1_{(Y=y)} = \sum_y E(X) \cdot 1_{(Y=y)} = E(X).$$

248. a)

$$\begin{aligned} E(X | Y) &= E(X | Y = 0) \cdot 1_{(Y=0)} + E(X | Y = 1) \cdot 1_{(Y=1)} \\ &= (2/3) \cdot 1_{(Y=0)} + (3/5) \cdot 1_{(Y=1)} \end{aligned}$$

$$\begin{aligned} E(Y | X) &= E(Y | X = 0) \cdot 1_{(X=0)} + E(Y | X = 1) \cdot 1_{(X=1)} \\ &= (2/3) \cdot 1_{(X=0)} + (3/5) \cdot 1_{(X=1)} \end{aligned}$$

b)

$$\begin{aligned} E(X | Y) &= E(X | Y = -1) \cdot 1_{(Y=-1)} + E(X | Y = 1) \cdot 1_{(Y=1)} \\ &= (9/4) \cdot 1_{(Y=-1)} + (7/4) \cdot 1_{(Y=1)} \end{aligned}$$

$$\begin{aligned} E(Y | X) &= E(Y | X = 1) \cdot 1_{(X=1)} + E(Y | X = 2) \cdot 1_{(X=2)} \\ &\quad + E(Y | X = 3) \cdot 1_{(X=3)} \\ &= (1/3) \cdot 1_{(X=1)} + (-1/3) \cdot 1_{(X=3)} \end{aligned}$$

c) No es necesario hacer muchos cálculos. Observe que  $X$  y  $Y$  son independientes. Entonces  $E(X | Y) = E(X) = 2$  y  $E(Y | X) = E(Y) = 0$ .

249. Por la propiedad de linealidad y la hipótesis de idéntica distribución,

$$\begin{aligned} E(X_1 | X_1 + \cdots + X_n) &= \frac{1}{n} E(X_1 + \cdots + X_n | X_1 + \cdots + X_n) \\ &= \frac{1}{n} (X_1 + \cdots + X_n). \end{aligned}$$

250. Algunos de los siguientes cálculos pueden simplificarse observando que  $X_1 \cdot X_2$  tiene distribución  $\text{Ber}(\theta^2)$ .

- a)  $E(T) = E(X_1 \cdot X_2) = E(X_1) \cdot E(X_2) = \theta^2 = \tau(\theta)$ .
- b)  $\text{Var}(T) = \text{Var}(X_1 \cdot X_2) = E(X_1^2 \cdot X_2^2) - E^2(X_1 \cdot X_2) = \theta^2 - \theta^4 = \theta^2(1 - \theta^2)$ .
- c)  $E(T | U) = (U(U - 1)/(n(n - 1)))$ .
- d) Encuentre  $\text{Var}(E(T | U)) = \dots$
- e) Compruebe que  $\text{Var}(E(T | U)) \leq \text{Var}(T)$ .

251. En ambos casos la transformación es una biyección.

252. a) Observe que  $T \sim \text{bin}(nk, \theta)$ . Entonces

$$E(h(T)) = \sum_{t=0}^{nk} h(t) \binom{nk}{t} \theta^t (1-\theta)^{nk-t} = (1-\theta)^{nk} \sum_{t=0}^{nk} h(t) \binom{nk}{t} (\theta/(1-\theta))^t.$$

La suma corresponde a un polinomio de la variable  $x = \theta/(1 - \theta)$ . Para que este polinomio en  $x$  sea cero para cualquier posible valor de  $x$ , sus coeficientes deben ser forzosamente cero, esto es,  $h(t) \binom{nk}{t} = 0$  para cualquier  $t = 0, 1, \dots, nk$ . Por lo tanto,  $h(t) = 0$  para cualquier  $t = 0, 1, \dots, nk$ . Es decir,  $h(T) = 0$ .

b) Observe que  $T \sim \text{Poisson}(n\theta)$ . Entonces

$$E(h(T)) = \sum_{t=0}^{\infty} h(t) e^{-n\theta} \frac{(n\theta)^t}{t!} = e^{-n\theta} \sum_{t=0}^{\infty} h(t) \frac{n^t}{t!} \theta^t.$$

Para que este polinomio en  $\theta$  sea cero para cualquier valor de  $\theta > 0$ , sus coeficientes deben ser forzosamente cero, esto es,  $h(t) \frac{n^t}{t!} = 0$  para cualquier  $t = 0, 1, \dots$ . Por lo tanto,  $h(t) = 0$  para  $t = 0, 1, \dots$ . Es decir,  $h(T) = 0$ .

c) Observe que  $T \sim \text{bin neg}(n, \theta)$ . Entonces

$$E(h(T)) = \sum_{t=0}^{\infty} h(t) \binom{n+t-1}{t} \theta^n (1-\theta)^t = \theta^n \sum_{t=0}^{\infty} h(t) \binom{n+t-1}{t} (1-\theta)^t.$$

Para que este polinomio en  $1 - \theta$  sea cero para cualquier valor de  $\theta \in (0, 1)$ , sus coeficientes deben ser forzosamente cero, esto es,  $h(t) \binom{n+t-1}{t} = 0$  para cualquier  $t = 0, 1, \dots$ . Por lo tanto,  $h(t) = 0$  para  $t = 0, 1, \dots$ . Es decir,  $h(T) = 0$ .

d) Solución omitida.

e) Solución omitida.

253. Solución omitida.

254. a) Observe que  $T$  tiene distribución Poisson( $k\theta$ ). Sea  $h$  una función tal que

$$0 = E[h(T)] = \sum_{t=0}^{\infty} h(t) e^{-k\theta} \frac{(k\theta)^t}{t!} = e^{-k\theta} \sum_{t=0}^{\infty} \frac{h(t)}{t!} \theta^t.$$

Esta suma es una serie de potencias en  $\theta$  que se anula para todo valor  $\theta > 0$ . Esto sólo ocurre cuando sus coeficientes son todos cero. Es decir,  $h(t)/t! = 0$ , lo que implica que  $h(t) = 0$  para  $t = 0, 1, \dots$ . Esto significa que  $h(T) = 0$ .

b) Sea la función  $h(x_1, \dots, x_k) = x_1 - x_2$  definida en  $\{0, 1, \dots\} \times \{0, 1, \dots\}$ , que es distinta de cero. Se comprueba que  $T$  no es completa pues se cumple la condición  $E[h(T)] = 0$  sin que  $h(T)$  sea cero. En efecto,

$$E[h(T)] = \sum_{x_1, x_2=0}^{\infty} (x_1 - x_2) \left( e^{-\theta} \frac{\theta^{x_1}}{x_1!} \right) \left( e^{-\theta} \frac{\theta^{x_2}}{x_2!} \right) = \theta - \theta = 0.$$

255. Tome la función  $h(x) = x$ , que es distinta de cero. Sea  $T$  una variable aleatoria con distribución unif( $-\theta, \theta$ ), cuya media es cero. Es inmediato comprobar que  $E[h(T)] = 0$  para todo  $\theta > 0$  sin que  $h(T)$  sea cero.

256. Tome la función  $h(x) = x$ , que es distinta de cero. Sea  $T$  una variable aleatoria con distribución  $N(0, \theta)$ , cuya media es cero y cuya varianza es  $\theta$ . Es inmediato comprobar que  $E[h(T)] = 0$  para todo  $\theta > 0$  sin que  $h(T)$  sea cero.

257. a) Se propone  $T = 1_{\{1\}}(X_1) + e^2 1_{\{0\}}(X_1)$ . Este estimador es insesgado pues

$$E(T) = P(X_1 = 1) + e^2 P(X_1 = 0) = \theta + (1 - \theta)e^2.$$

b) Sea  $u \in \{0, 1, \dots, n\}$  un posible valor de  $U = X_1 + \dots + X_n$ . Entonces

$$\begin{aligned} E(T | U = u) &= E(1_{\{1\}}(X_1) + e^2 1_{\{0\}}(X_1) | X_1 + \dots + X_n = u) \\ &= P(X_1 = 1 | X_1 + \dots + X_n = u) \\ &\quad + e^2 P(X_1 = 0 | X_1 + \dots + X_n = u) \\ &= \frac{P(X_1 = 1)P(X_2 + \dots + X_n = u - 1)}{P(X_1 + \dots + X_n = u)} \\ &\quad + e^2 \frac{P(X_1 = 0)P(X_2 + \dots + X_n = u)}{P(X_1 + \dots + X_n = u)} \\ &\quad \vdots \\ &= \frac{u}{n} + e^2 \frac{n - u}{n}. \end{aligned}$$

Por lo tanto,  $E(T | U) = U/n + e^2(n - U)/n$ .

c) Como  $U$  es suficiente y completa, y  $T$  es insesgado, por el teorema de Lehmann-Scheffé,  $E(T | U)$  es el UMVUE para  $\tau(\theta) = \theta + (1 - \theta)e^2$ .

258. Solución omitida.

259. a) Evidente.

b) Evidente del hecho de que  $\text{Var}(X) = \theta$ .

c)

$$\begin{aligned} E(T | U) &= E\left(\frac{1}{2}(X_1 + X_2) | X_1 + \dots + X_n\right) \\ &= E(X_1 | X_1 + \dots + X_n) \\ &= \bar{X}. \end{aligned}$$

d) Evidente del hecho de que  $\text{Var}(X) = \theta$ .

e) Este es el Ejercicio 192. Vea su solución.

f) Evidente de las expresiones anteriores.

260. a)  $E(T) = E(1_{\{0\}}(X_1)) = P(X_1 = 0) = e^{-\theta}$ .

b) Para cualesquiera números reales  $x_1, \dots, x_n$ ,

$$\begin{aligned} f(x_1, \dots, x_n; \theta) &= e^{-\theta} \frac{\theta^{x_1}}{x_1!} \cdots e^{-\theta} \frac{\theta^{x_n}}{x_n!} 1_{\{0,1,\dots\}}(x_1) \cdots 1_{\{0,1,\dots\}}(x_n) \\ &= e^{-n\theta} \theta^{x_1 + \dots + x_n} \cdot \frac{1}{x_1! \cdots x_n!} 1_{\{0,1,\dots\}}(x_1) \cdots 1_{\{0,1,\dots\}}(x_n) \\ &= g(U(x_1, \dots, x_n); \theta) \cdot h(x_1, \dots, x_n). \end{aligned}$$

El teorema de factorización garantiza que  $U$  es suficiente para  $\theta$ .

c) Sean  $x_1, \dots, x_n$  y  $y_1, \dots, y_n$  cualesquiera números reales. Entonces

$$\frac{f(x_1, \dots, x_n; \theta)}{f(y_1, \dots, y_n; \theta)} = \frac{y_1! \cdots y_n! \theta^{x_1 + \dots + x_n} 1_{\{0,1,\dots\}}(x_1) \cdots 1_{\{0,1,\dots\}}(x_n)}{x_1! \cdots x_n! \theta^{y_1 + \dots + y_n} 1_{\{0,1,\dots\}}(y_1) \cdots 1_{\{0,1,\dots\}}(y_n)}$$

Esta cantidad no depende  $\theta \Leftrightarrow U(x_1, \dots, x_n) = U(y_1, \dots, y_n)$ . Esto comprueba que  $U$  es suficiente minimal.

d) Para cada valor  $u = 0, 1, \dots$

$$\begin{aligned} E(T | U = u) &= E(1_{\{0\}}(X_1) | X_1 + \dots + X_n = u) \\ &= P(X_1 = 0 | X_1 + \dots + X_n = u) \\ &= P(X_1 = 0, X_1 + \dots + X_n = u) / P(X_1 + \dots + X_n = u) \\ &= P(X_1 = 0, X_2 + \dots + X_n = u) / P(X_1 + \dots + X_n = u) \\ &= P(X_1 = 0) P(X_2 + \dots + X_n = u) / P(X_1 + \dots + X_n = u) \\ &= (e^{-\theta} e^{-(n-1)\theta} ((n-1)\theta)^u / u!) / (e^{-n\theta} (n\theta)^u / u!) \\ &= ((n-1)/n)^u. \end{aligned}$$

De donde se concluye que  $E(T | U) = ((n-1)/n)^U$ .

e) El resultado se obtiene al observar que  $T$  tiene distribución Bernoulli de parámetro  $e^{-\theta}$ .

f) Usando la expresión de la f.g.p. para la distribución Poisson( $n\theta$ ),

$$\begin{aligned} \text{Var}(E(T | U)) &= E(E^2(T | U)) - E^2(E(T | U)) \\ &= E(((n-1)/n)^{2U}) - E^2(((n-1)/n)^U) \\ &= \exp\{n\theta(((n-1)/n)^2 - 1)\} - \exp\{2n\theta((n-1)/n - 1)\} \\ &= e^{-2\theta}(e^{\theta/n} - 1). \end{aligned}$$

g) Haciendo las operaciones indicadas en la definición para la cota inferior de Cramér-Rao se encuentra que

$$\text{CICR}(\theta) = \frac{(\tau'(\theta))^2}{nE[(\frac{\partial}{\partial \theta} \ln f(X; \theta))^2]} = e^{-2\theta}(\theta/n).$$

- h) Para la primera desigualdad use  $e^x > 1 + x$ . La segunda desigualdad es estricta para  $n \geq 2$ .
- i) No se puede garantizar que el estimador  $E(T | U)$  sea el UMVUE para  $\tau(\theta)$ , pues podría existir otro estimador insesgado que alcance la cota inferior de Cramér-Rao.
- j) Supongamos que  $E[h(U)] = 0$  para cada  $\theta > 0$ . Entonces, como  $U$  tiene distribución Poisson( $n\theta$ ),

$$0 = \sum_{u=0}^{\infty} h(u) e^{-n\theta} \frac{(n\theta)^u}{u!} = e^{-n\theta} \sum_{u=0}^{\infty} \frac{h(u)}{u!} (n\theta)^u$$

Esta es una serie de potencias en  $\theta$  que se anula para todo valor de  $\theta > 0$ . En consecuencia, sus coeficientes son cero. Esto implica que  $h(u) = 0$  para  $u = 0, 1, \dots$

- k) Por el teorema de Lehmann-Scheffé, puede concluirse que  $E(T | U)$  es el UMVUE para  $\tau(\theta)$ . Este es un ejemplo de un UMVUE que no alcanza la cota inferior de Cramér-Rao.

261. a)  $E(T) = P(X_1 = 1) = \theta e^{-\theta}$ .

- b) El resultado es evidente a partir de la observación de que  $T$  tiene distribución  $\text{Ber}(\theta e^{-\theta})$ .

- c) Para  $u \geq 1$  entero,

$$\begin{aligned} E(T | U = u) &= P(X_1 = 1 | X_1 + \dots + X_n = u) \\ &= \frac{P(X_1 = 1, X_2 + \dots + X_n = u - 1)}{P(X_1 + \dots + X_n = u)} \\ &= e^{-\theta} \theta \cdot e^{-(n-1)\theta} \frac{((n-1)\theta)^{u-1}}{(u-1)!} \cdot e^{n\theta} \frac{u!}{(n\theta)^u} \\ &= \left( \frac{n-1}{n} \right)^{u-1} \frac{u}{n}. \end{aligned}$$

d)

$$\begin{aligned}
\text{Var}(E(T|U)) &= E\left(\left(\frac{n-1}{n}\right)^{n\bar{X}-1} \bar{X}\right)^2 - E^2\left(\left(\frac{n-1}{n}\right)^{n\bar{X}-1} \bar{X}\right) \\
&= \left(\frac{n}{n-1}\right)^2 \sum_{x=1}^{\infty} \left(\frac{n-1}{n}\right)^{2x} \frac{x^2}{n^2} e^{-n\theta} \frac{(n\theta)^x}{x!} - \theta^2 e^{-2\theta} \\
&= \left(\frac{n}{n-1}\right)^2 \sum_{x=1}^{\infty} \frac{x(x-1) + x}{n^2} e^{-n\theta} \frac{\left(\left(\frac{n-1}{n}\right)^2 n\theta\right)^x}{x!} - \theta^2 e^{-2\theta} \\
&\quad \vdots \\
&= e^{-2\theta + \theta/n} \frac{\theta}{n} \left(1 + (n-1)^2 \frac{\theta}{n}\right) - e^{-2\theta} \theta^2.
\end{aligned}$$

e) Sabemos que  $\text{CICR}(\theta) = \theta/n$  para la varianza de cualquier estimador insesgado para el parámetro  $\theta$  de la distribución Poisson. Si ahora consideramos la función parametral  $\tau(\theta) = \theta e^{-\theta}$ , entonces  $\text{CICR}(\theta) = (\theta/n)(\tau'(\theta))^2 = e^{-2\theta}(1-\theta)^2\theta/n$ .

f) Solución omitida.

262. a) Puede comprobarse que la función de densidad de  $X_{(1)}$  es

$$f_{X_{(1)}}(x) = \begin{cases} ne^{-n(x-\theta)} & \text{si } x > \theta, \\ 0 & \text{en otro caso.} \end{cases}$$

De donde puede encontrarse que  $E(X_{(1)}) = \theta + 1/n$ . Para la suficiencia tenemos la factorización

$$\begin{aligned}
f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) &= e^{-\sum_{i=1}^n (x_i - \theta)} \cdot 1_{(\theta, \infty)}(x_1) \cdots 1_{(\theta, \infty)}(x_n) \\
&= e^{n\theta} e^{-n\bar{x}} 1_{(0, x_{(1)})}(\theta) \\
&= e^{n\theta} 1_{(0, x_{(1)})}(\theta) \cdot e^{-n\bar{x}} \\
&= g(x_{(1)}; \theta) \cdot h(x_1, \dots, x_n).
\end{aligned}$$

Por lo tanto,  $X_{(1)} - 1/n$  es también suficiente. Para la completez, sea  $T = X_{(1)} - 1/n$  y suponga que  $h$  es una función tal que  $E(h(T)) = 0$ . Para cualquier valor de  $\theta > 0$  se cumple que

$$0 = E(h(T)) = \int_{\theta}^{\infty} h(t) n e^{-n(t+1/n-\theta)} dt = n e^{n\theta-1} \int_{\theta}^{\infty} h(t) e^{-nt} dt.$$

Esto implica que la última integral es cero. Derivando esta integral respecto de  $\theta$ , y suponiendo continuidad de la función  $h$ , se obtiene que para casi cualquier  $t > \theta$ ,  $h(t) = 0$ .

- b)  $X_{(1)} - 1/n$  es el UMVUE para  $\theta$ .
263. a)  $E(T) = E(X_1) = \theta$ .  
 b)  $\text{Var}(T) = \text{Var}(X_1) = \sigma^2$ .  
 c)  $E(T|U) = E(X_1 | X_1 + \dots + X_n) = (X_1 + \dots + X_n)/n = \bar{X}$ .  
 d)  $\text{Var}(E(T|U)) = \text{Var}(\bar{X}) = \sigma^2/n$ .  
 e) Este es el Ejercicio 193. Resulta que  $\text{CICR}(\theta)$  es la constante  $\sigma^2/n$ . Es constante pues no es función del parámetro  $\theta$ .  
 f) Los resultados anteriores demuestran que, efectivamente,

$$\frac{\sigma^2}{n} = \text{CICR}(\theta) = \text{Var}(E(T|U)) \leq \text{Var}(T) = \sigma^2.$$

264. Solución omitida.

265. Solución omitida.

266. Solución omitida.

267. Puede comprobarse que la primera estadística de orden  $X_{(1)}$  tiene función de densidad

$$f(x) = \begin{cases} n e^{-n(x-\theta)} & \text{si } x > \theta, \\ 0 & \text{en otro caso.} \end{cases}$$

Puede verificarse entonces que  $E(X_{(1)}) = 1/n + \theta$ , y de aquí se desprende la propiedad de insesgamiento del estimador. Para demostrar la propiedad de suficiencia puede usarse el teorema de factorización de Neyman. Tenemos que

$$\begin{aligned} L(x_1, \dots, x_n; \theta) &= \prod_{i=1}^n e^{-(x_i-\theta)} \cdot 1_{(\theta, \infty)}(x_i) \\ &= e^{n\bar{x}} \cdot e^{n\theta} \cdot 1_{(\theta, \infty)}(x_{(1)}) \\ &= h(x_1, \dots, x_n) \cdot g(x_{(1)}; \theta). \end{aligned}$$

Esto demuestra la suficiencia de  $X_{(1)}$ , y en consecuencia, también la suficiencia de  $X_{(1)} - 1/n$ . Alternativamente, puede modificarse ligeramente el factor  $g(x_{(1)}; \theta)$  en las ecuaciones anteriores para escribirlo como  $g(x_{(1)} - 1/n; \theta)$ . Veamos ahora la completez. Es suficiente demostrar esta propiedad para  $X_{(1)}$ . Sea  $h(t)$  una función definida en el intervalo  $(\theta, \infty)$  y tal que  $E(h(T)) = 0$ , con  $T = X_{(1)}$ . Esta condición es equivalente a la identidad

$$\int_{\theta}^{\infty} h(t) e^{-nt} dt = 0.$$

La integral corresponde a una transformada de Laplace de la función  $h(t)$  en el intervalo  $(\theta, \infty)$ . Por la propiedad de unicidad de esta transformada, la

integral es cero si, y sólo si, la función  $h(t)$  es cero. La completez de  $X_{(1)}$  es equivalente a la completez de  $X_{(1)} - 1/n$ . Por el teorema de Lehmann-Scheffé, se concluye que  $X_{(1)} - 1/n$  es el UMVUE para  $\theta$ .

268. a) Veamos primero la suficiencia. Esta propiedad se puede demostrar usando el teorema de factorización. Tenemos que

$$\begin{aligned} L(x_1, \dots, x_n; \theta) &= \theta^n (x_1 \cdots x_n)^{\theta-1} 1_{(0,1)}(x_1) \cdots 1_{(0,1)}(x_n) \\ &= \theta^n ((x_1 \cdots x_n)^{1/n})^{n(\theta-1)} \cdot 1_{(0,1)}(x_1) \cdots 1_{(0,1)}(x_n) \\ &= g(U(x_1, \dots, x_n); \theta) \cdot h(x_1, \dots, x_n). \end{aligned}$$

Ahora veamos la completez. Sea  $h$  una función tal que  $E[h(U)] = 0$ , para cualquier  $\theta > 0$ , y en donde es suficiente considerar la estadística  $U = X_1 \cdots X_n$ . Esto es,

$$\int_0^1 \cdots \int_0^1 h(x_1 \cdots x_n) \theta^n (x_1 \cdots x_n)^{\theta-1} dx_1 \cdots dx_n = 0.$$

Sin pérdida de generalidad puede considerarse el caso cuando sólo hay una variable involucrada. La identidad anterior se reduce entonces a la condición: para cualquier  $\theta > -1$ ,

$$\int_0^1 h(x) x^\theta dx = 0.$$

Esto implica que  $h(x) = 0$  c.s.

- b) Se escribe  $T = -(n-1)/\ln U$  y se observa que  $T$  es  $U$ -medible. Esto implica que  $E(T|U) = T$ , y por el teorema de Lehmann-Scheffé, se concluye que  $T$  es el UMVUE para  $\theta$ .
269. Se debe observar primeramente que la distribución en cuestión es gama(2,  $\theta$ ), de modo que  $X_1 + \cdots + X_n$  tiene distribución gama( $2n$ ,  $\theta$ ).

- a) Se puede demostrar la suficiencia usando el teorema de factorización. Tenemos que

$$\begin{aligned} f_{X_1, \dots, X_n}(x_1, \dots, x_n) &= \theta^{2n} e^{-\theta(x_1 + \cdots + x_n)} \cdot \left( \prod_{i=1}^n x_i \right) 1_{(0, \infty)}(x_1) \cdots 1_{(0, \infty)}(x_n) \\ &= g(T(x_1, \dots, x_n); \theta) \cdot h(x_1, \dots, x_n). \end{aligned}$$

Para la completez, supongamos que  $h$  es una función tal que para cualquier  $\theta > 0$ ,

$$0 = E[h(T)] = \int_0^\infty h(t) \cdot \frac{(\theta t)^{n-1}}{(2n-1)!} \theta e^{-\theta t} dt.$$

Esto es equivalente a

$$0 = \int_0^{\infty} h(t) \cdot t^{n-1} e^{-\theta t} dt,$$

lo cual corresponde a una transformada de Laplace de la función  $h(t) \cdot t^{n-1}$ . Por la propiedad de unicidad, esta transformada es cero si, y sólo si, la función  $h(t) \cdot t^{n-1}$  es cero. De aquí se obtiene que  $h(t)$  debe ser cero.

b) Tenemos que

$$\begin{aligned} E\left(\frac{1}{T}\right) &= \int_0^{\infty} \frac{1}{t} \frac{(\theta t)^{2n-1}}{(2n-1)!} \theta e^{-\theta t} dt \\ &= \frac{\theta}{2n-1} \int_0^{\infty} \frac{(\theta t)^{2n-2}}{(2n-2)!} \theta e^{-\theta t} dt \\ &= \frac{\theta}{2n-1}. \end{aligned}$$

c) De los incisos anteriores se obtiene que  $(2n-1)/T$  es una estadística insesgada, suficiente y completa para  $\theta$ . Recordemos que las dos últimas propiedades se preservan bajo transformaciones biyectivas. Por el teorema de Lehmann-Scheffé, este estimador es el UMVUE para  $\theta$ .

270. a) Sea  $(t_1, \dots, t_k)$  un posible valor de  $T = (T_1, \dots, T_k)$ . Entonces

$$f_{X_1, \dots, X_n | T}(x_1, \dots, x_n | t_1, \dots, t_k) = \frac{f_{X_1, \dots, X_n}(x_1, \dots, x_n)}{f_T(t_1, \dots, t_k)},$$

en donde se deben cumplir las relaciones

$$\begin{aligned} T_1(x_1, \dots, x_n) &= \sum_{i=1}^n d_1(x_i) = t_1, \\ T_2(x_1, \dots, x_n) &= \sum_{i=1}^n d_2(x_i) = t_2, \\ &\vdots \\ T_k(x_1, \dots, x_n) &= \sum_{i=1}^n d_k(x_i) = t_k. \end{aligned}$$

Suponiendo el caso continuo y definiendo la imagen inversa  $D = \{(y_1, \dots, y_n) : T(y_1, \dots, y_n) = (t_1, \dots, t_k)\}$ , se puede escribir

$$f_T(t_1, \dots, t_k) = \int_D f_{X_1, \dots, X_n}(y_1, \dots, y_n) dy_1 \cdots dy_n.$$

Por lo tanto,

$$\begin{aligned} & f_{X_1, \dots, X_n | T}(x_1, \dots, x_n | t_1, \dots, t_k) \\ &= \left( \int_D \frac{f_{X_1}(y_1) \cdots f_{X_n}(y_n)}{f_{X_1}(x_1) \cdots f_{X_n}(x_n)} dy_1 \cdots dy_n \right)^{-1} \\ &= \left( \int_D \left( \prod_{i=1}^n \frac{b(y_i)}{b(x_i)} \right) \exp(c(\theta) \left[ \sum_{i=1}^n d(y_i) - \sum_{i=1}^n d(x_i) \right]) dy_1 \cdots dy_n \right)^{-1}, \end{aligned}$$

pero  $\sum_{i=1}^n d(y_i) = \sum_{i=1}^n d(x_i) = (t_1, \dots, t_k)$ , por lo tanto el exponente es nulo y la expresión no depende de  $\theta$ .

b) La factorización es

$$\prod_{i=1}^n f(x_i; \theta) = \left[ a^n(\theta) \exp\left(c(\theta) \sum_{i=1}^n d(x_i)\right) \right] \left[ \prod_{i=1}^n b(x_i) \right].$$

271. Puede comprobarse que la función de densidad de la variable  $X_1/\theta$  no depende de  $\theta$  y está dada por

$$f(u) = \begin{cases} 2(1-u) & \text{si } 0 < u < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

Dado un valor de  $\alpha$ , pueden encontrarse dos valores  $0 < a < b$  tales que  $P(a < X_1/\theta < b) = 1 - \alpha$ . Por ejemplo,  $a = 1 - \sqrt{1 - \alpha/2}$  y  $b = 1 - \sqrt{\alpha/2}$ , para  $\alpha > 1/2$ . De aquí se obtiene el intervalo

$$P\left(\frac{X_1}{1 - \sqrt{\alpha/2}} < \theta < \frac{X_1}{1 - \sqrt{1 - \alpha/2}}\right) = 1 - \alpha.$$

272. La variable aleatoria  $X_{(n)}/\theta$  tiene función de distribución

$$F(u) = \begin{cases} u^n & \text{si } 0 < u < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

El intervalo  $(a, b)$  de longitud mínima tal que  $F(b) - F(a) = 1 - \alpha$  está dado por  $(\alpha^{1/n}, 1)$ . Por lo tanto,  $P(\alpha^{1/n} < X_{(n)}/\theta < 1) = 1 - \alpha$ . De aquí se obtiene el intervalo

$$P(X_{(n)} < \theta < X_{(n)}/\alpha^{1/n}) = 1 - \alpha.$$

273. La variable aleatoria  $\max_{1 \leq i \leq n} |X_i|/\theta$  tiene función de distribución

$$F(u) = \begin{cases} u^n & \text{si } 0 < u < 1, \\ 0 & \text{en otro caso,} \end{cases}$$

la cual no depende de parámetros desconocidos. De manera que, dado un valor de  $\alpha$ , pueden encontrarse dos valores  $0 < a < b$  tales que  $P(a < \max_{1 \leq i \leq n} |X_i|/\theta < b) = 1 - \alpha$ . Por ejemplo,  $a = \sqrt[n]{\alpha/2}$  y  $b = \sqrt[n]{1 - \alpha/2}$ . De aquí se obtiene el siguiente intervalo de confianza, suponiendo  $0 < \alpha < 1/2$ ,

$$P\left(\frac{\max_{1 \leq i \leq n} |X_i|}{\sqrt{1 - \alpha/2}} < \theta < \frac{\max_{1 \leq i \leq n} |X_i|}{\sqrt{\alpha/2}}\right) = 1 - \alpha.$$

274. Sea  $X$  una variable aleatoria con la distribución indicada. Observe que  $X - a \sim \exp(1/\theta)$ . Se procede como en el caso exponencial. Tenemos la muestra aleatoria  $X_1 - a, \dots, X_n - a$  de la distribución  $\exp(1/\theta)$ . Por lo tanto,  $(X_1 + \dots + X_n) - an$  tiene distribución  $\text{gama}(n, 1/\theta)$ . Entonces

$$\frac{1}{\theta} [n\bar{X} - an] \sim \text{gama}(n, 1).$$

Pueden encontrarse dos valores  $0 < \gamma_{1-\alpha/2} < \gamma_{\alpha/2}$  tales que  $P(\gamma_{1-\alpha/2} < [n\bar{X} - an]/\theta < \gamma_{\alpha/2}) = 1 - \alpha$ , de donde se obtiene el intervalo de confianza

$$P\left(\frac{n(\bar{X} - a)}{\gamma_{\alpha/2}} < \theta < \frac{n(\bar{X} - a)}{\gamma_{1-\alpha/2}}\right) = 1 - \alpha.$$

275. La media  $\theta$  es desconocida y la varianza  $\sigma^2$  es conocida. En este caso un intervalo para  $\theta$  con confianza del  $(1 - \alpha)100\%$  está dado por

$$(\bar{X} - z_{\alpha/2} \cdot \sigma/\sqrt{n}, \bar{X} + z_{\alpha/2} \cdot \sigma/\sqrt{n}),$$

cuya longitud es  $\ell = 2 \cdot z_{\alpha/2} \cdot \sigma/\sqrt{n}$ . Como la confianza debe ser  $0.95 = 1 - \alpha$ , se tiene que  $\alpha = 0.05$ . Y por lo tanto,  $z_{\alpha/2} = z_{0.025} = 1.96$ . Se requiere que la longitud  $\ell$  del intervalo sea de 2 cm. Entonces  $2 \cdot 1.96 \cdot \sigma/\sqrt{n} = 2$  cm. De aquí se obtiene  $n = (1.96)^2 \cdot \sigma^2$ , o el entero más pequeño que exceda esta cantidad.

276. a)  $\mathcal{C} = \{0\}$ ,  $\alpha = 0$ ,  $\beta = 31/32$ .

b)  $\mathcal{C} = \{0, 1, 2, 3, 4, 5\}$ ,  $\alpha = 5/6$ ,  $\beta = 0$ .

277. Use el teorema de probabilidad total. Si  $D$  denota el evento de lanzar el dado y  $M$  el evento de lanzar la moneda, entonces para  $x = 0, 1, 2, 3, 4, 5, 6$ ,

$$P(X = x) = P(X = x | D) P(D) + P(X = x | M) P(M).$$

278. Las mejores regiones de rechazo de tamaño  $\alpha = 1/6$  son:

Región de rechazo	$\alpha$	$\beta$
$\mathcal{C} = \{0, 1\}$	1/6	26/32
$\mathcal{C} = \{0, 2\}$	1/6	21/32
$\mathcal{C} = \{0, 3\}$	1/6	21/32
$\mathcal{C} = \{0, 4\}$	1/6	26/32
$\mathcal{C} = \{0, 5\}$	1/6	30/32
$\mathcal{C} = \{0, 6\}$	1/6	31/32

De las cuales la segunda y la tercera son mejores pues  $\beta$  es menor.

279. Sea  $X$  el resultado de lanzar la moneda. Entonces  $X$  tiene distribución  $\text{Ber}(\theta)$ . Si  $X_1, \dots, X_n$  es una muestra aleatoria de esta distribución, entonces

$$\begin{aligned}
 \alpha &= P(\bar{X} \geq 13/24 \mid \theta = 1/2) \\
 &= P\left(\frac{\bar{X} - 1/2}{\sqrt{(1/2)(1 - 1/2)/n}} \geq \frac{13/24 - 1/2}{\sqrt{(1/2)(1 - 1/2)/n}} \mid \theta = 1/2\right) \\
 &\approx P(Z \geq \sqrt{n}/12) \\
 &= 1 - \Phi(\sqrt{n}/12).
 \end{aligned}$$

Por otro lado,

$$\begin{aligned}
 \beta &= P(\bar{X} < 13/24 \mid \theta = 7/12) \\
 &= P\left(\frac{\bar{X} - 7/12}{\sqrt{(7/12)(1 - 7/12)/n}} < \frac{13/24 - 7/12}{\sqrt{(7/12)(1 - 7/12)/n}} \mid \theta = 7/12\right) \\
 &\approx P(Z < -(1/2)\sqrt{n/35}) \\
 &= 1 - \Phi(-(1/2)\sqrt{n/35}).
 \end{aligned}$$

280. Las siguientes probabilidades fueron calculadas en el paquete R usando la función `pnorm(x, 0, 1)`.

$$\begin{aligned}
 a) \quad \alpha &= P(Z > (4.7 - 2)/2) = 0.08850799, \\
 \beta &= P(Z \leq (4.7 - 5)/2) = 0.4403823.
 \end{aligned}$$

$$\begin{aligned}
 b) \quad \alpha &= P(Z > (4.5 - 2)/(\sqrt{20}/3)) = 0.04676626, \\
 \beta &= P(Z \leq (4.5 - 5)/(\sqrt{20}/3)) = 0.3686578.
 \end{aligned}$$

- c)  $\alpha = P(Z > (4.2 - 2)/\sqrt{2}) = 0.05989747,$   
 $\beta = P(Z \leq (4.2 - 5)/\sqrt{2}) = 0.2858038.$
- d)  $\alpha = P(Z > (4.1 - 2)/\sqrt{4/3}) = 0.03448217,$   
 $\beta = P(Z \leq (4.1 - 5)/\sqrt{4/3}) = 0.2178653.$

281. a)  $\pi(\theta) = 1 - \Phi(\sqrt{n}(\frac{\theta}{\theta_0} - 1)),$  para  $0 < \theta < \infty.$

b)  $\sup_{\theta \in (0, \theta_0]} \pi(\theta) = 1 - \Phi(-\sqrt{n}).$

c)  $\sup_{\theta \in (\theta_0, \infty)} (1 - \pi(\theta)) = 1.$

282.  $c = \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1 - \alpha/2),$   $\pi(\theta) = 1 - \Phi(\frac{\theta_0 - \theta + c}{\sigma/\sqrt{n}}) + \Phi(\frac{\theta_0 - \theta - c}{\sigma/\sqrt{n}}),$   $-\infty < \theta < \infty.$

283. Supondremos un modelo normal para el consumo de agua con media desconocida  $\theta$  y varianza conocida  $\sigma^2 = (2000)^2$ . Llevaremos a cabo la prueba de hipótesis

$$H_0 : \theta = 20,000 \quad vs \quad H_1 : \theta \neq 20,000.$$

Los datos proporcionados corresponden a los valores de una muestra aleatoria de tamaño  $n = 15$ , y haciendo el promedio de estos valores se obtiene una media muestral  $\bar{x} = 20546.2$ . La estadística de prueba toma entonces el valor

$$z_0 = \frac{\bar{x} - \theta_0}{\sigma/\sqrt{n}} = \frac{20546.2 - 20000}{2000/\sqrt{15}} = 1.0577.$$

Por otro lado, tomando  $\alpha = 0.1$ , de la tabla de probabilidades de la distribución normal se encuentra que  $z_{\alpha/2} = 1.65$ . Como no se cumple la condición  $|z_0| \geq z_{\alpha/2}$ , la estadística de prueba  $Z_0$  cae fuera de la región de rechazo y, por lo tanto, no se rechaza la hipótesis  $H_0$ , es decir, no existen evidencias para afirmar que el consumo de agua por casa en la zona de estudio haya cambiado.

284. Para contestar a esta pregunta podemos llevar a cabo la prueba de hipótesis

$$H_0 : \theta_1 - \theta_2 = 0 \quad vs \quad H_1 : \theta_1 - \theta_2 \neq 0,$$

en donde  $\theta_1$  corresponde a la media de la población de mujeres, y  $\theta_2$  a la media de la población de hombres. Con los datos recabados la estadística de la prueba toma el valor

$$z = \frac{\bar{x}_1 - \bar{x}_2 - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = -8.11.$$

Con  $\alpha = 0.10$  se tiene que  $z_{\alpha/2} = 1.65$ . Entonces  $|z| \geq z_{\alpha/2}$  y por lo tanto se rechaza la hipótesis nula, es decir, las poblaciones de hombres y mujeres muestran tiempos promedios diferentes para terminar la prueba escrita.

285. La región de rechazo es  $\mathcal{C} = \{(x_1, \dots, x_n) : x_1 + \dots + x_n \geq c\}$ , en donde  $c$  es el entero más pequeño tal que  $P(Z_0 \geq c) \leq \alpha$ , con  $Z_0 \sim \text{bin}(n, \theta_0)$ . La probabilidad del error tipo II es  $\beta = P(Z_1 < c)$  con  $Z_1 \sim \text{bin}(n, \theta_1)$ .
286. La región de rechazo es  $\mathcal{C} = \{(x_1, \dots, x_n) : x_1 + \dots + x_n \geq c\}$ , en donde  $c$  es el entero más pequeño tal que  $P(Z_0 \geq c) \leq \alpha$ , con  $Z_0 \sim \text{bin}(nk, \theta_0)$ . La probabilidad del error tipo II es  $\beta = P(Z_1 < c)$  con  $Z_1 \sim \text{bin}(nk, \theta_1)$ .
287. La región de rechazo es  $\mathcal{C} = \{(x_1, \dots, x_n) : x_1 + \dots + x_n \geq c\}$ , en donde  $c$  es el entero más pequeño tal que  $P(Z_0 \geq c) \leq \alpha$ , con  $Z_0 \sim \text{bin neg}(n, \theta_0)$ . La probabilidad del error tipo II es  $\beta = P(Z_1 < c)$  con  $Z_1 \sim \text{bin neg}(n, \theta_1)$ .
288. La región de rechazo es  $\mathcal{C} = \{(x_1, \dots, x_n) : x_1 + \dots + x_n \leq c\}$ , en donde  $c$  es tal que  $P(Z_0 \leq c) = \alpha$ , con  $Z_0 \sim \text{gama}(n, \theta_0)$ . La probabilidad del error tipo II es  $\beta = P(Z_1 > c)$  con  $Z_1 \sim \text{gama}(n, \theta_1)$ .
289. El procedimiento es análogo al caso cuando  $\sigma^2$  es conocida. La región de rechazo es nuevamente  $\mathcal{C} = \{(x_1, \dots, x_n) : \bar{x} \geq c\}$ , en donde  $c$  es tal que

$$\begin{aligned} \alpha &= P(\bar{X} \geq c | \theta = \theta_0) \\ &= P\left(\frac{\bar{X} - \theta_0}{S/\sqrt{n}} \geq \frac{c - \theta_0}{S/\sqrt{n}}\right) \\ &= P\left(Z_0 \geq \frac{c - \theta_0}{S/\sqrt{n}}\right), \end{aligned}$$

en donde  $Z_0$  tiene distribución  $t(n-1)$ . La probabilidad de cometer el error tipo II es

$$\begin{aligned} \beta &= P(\bar{X} < c | \theta = \theta_1) \\ &= P\left(\frac{\bar{X} - \theta_1}{S/\sqrt{n}} < \frac{c - \theta_1}{S/\sqrt{n}}\right) \\ &= P\left(Z_1 < \frac{c - \theta_1}{S/\sqrt{n}}\right), \end{aligned}$$

en donde  $Z_1$  tiene distribución  $t(n-1)$ .

290. La región de rechazo es  $\mathcal{C} = \{(x_1, \dots, x_n) : \sum_{i=1}^n (x_i - \mu)^2 \geq c\}$ , en donde  $c$  es tal que  $P(Z_0 \geq c/\sigma_0^2) = \alpha$ , con  $Z_0 \sim \chi^2(n)$ . La probabilidad del error tipo II es  $\beta = P(Z_0 < c/\sigma_1^2)$ .

# Bibliografía

- [1] Aguirre V. *et al.* *Fundamentos de probabilidad y estadística*. Jit Press, 2003.
- [2] Alonso Reyes Ma. del P., Flores Díaz J. A. *Estadística descriptiva para bachillerato*. Instituto de Matemáticas, UNAM, 2004.
- [3] Bernardo J. M., Smith A. F. *Bayesian theory*. Wiley, 1994.
- [4] Brychkov Yu. A., Glaeske H.-J., Prudnikov A. P., Tuan V. K. *Multi-dimensional integral transformations*. Gordon and Breach Science Publishers, 1992.
- [5] Casella G., Berger R. L. *Statistical inference*. Thomson Press, 2008.
- [6] Clarke L. E. *Random variables*. Longman, 1975.
- [7] Devore J. *Probability and statistics for the engineering and the sciences*. Duxbury Press, 2008.
- [8] Flanders H. Differentiation under the integral sign. *The American Mathematical Monthly*, Vol. 80, No. 6 (1973), pp. 615-627.
- [9] Fleming W. *Functions of several variables*. Addison-Wesley, 1965.
- [10] Hoel P. G., Port S. C., Stone C. J. *Introduction to statistical theory*. Houghton Mifflin, 1971.
- [11] Hogg R. V., McKean J., Craig A. T. *Introduction to mathematical statistics*. Pearson, 2013.
- [12] Karr A. F. *Probability*. Springer-Verlag, 1993.

- [13] Lehmann E. L., Casella G. *Theory of point estimation*. Springer, 1998.
- [14] Loomis L. H., Sternberg S. *Advanced calculus*. Addison-Wesley, 1968.
- [15] Marsden J. E., Tromba A. J. *Vector calculus*. W. H. Freeman and Company, 1976.
- [16] Mendenhall W., Sincich T. *Probabilidad y estadística para ingeniería y ciencias*. Prentice Hall, 1997.
- [17] Migon H. S., Gamerman D, Louzada F. *Statistical inference. An integrated approach*. CRC Press, 2015.
- [18] Mood A. M., Graybill F. A., Boes D. C. *Introduction to the theory of statistics*. McGraw Hill, 1983.
- [19] Mukhopadhyay N. *Introductory statistical inference*. Chapman & Hall/CRC, 2006.
- [20] Miller I., Miller M. *John E. Freund's mathematical statistics*. Prentice Hall, 1999.
- [21] Rincón L. *Curso intermedio de probabilidad*. Las Prensas de Ciencias, Facultad de Ciencias, UNAM, 2007.
- [22] Rincón L. *Estadística descriptiva*. Las Prensas de Ciencias, Facultad de Ciencias, UNAM, 2017.
- [23] Shunichi A., Hiroshi, N. *Methods of information geometry*. Translations of Mathematical Monographs, v. 191, American Mathematical Society, 2000.
- [24] Sanabria Brenes G. *Comprendiendo la estadística inferencial*. Editorial Tecnológica de Costa Rica, 2011.
- [25] Tukey J. W. *Exploratory data analysis*. Addison-Wesley, 1977.
- [26] Weiss N. A. *Introductory statistics*. Addison-Wesley, 1999.
- [27] Valencia G., Mendoza M., Aranda F. *Introducción a la inferencia estadística*. Comunicación Interna No. 42, Departamento de Matemáticas, Facultad de Ciencias, UNAM, 1978.

- [28] Williams D. *Probability with martingales*. Cambridge University Press, 1991.
- [29] O'Connor J. J., Robertson E. F. *Maclutor history of mathematics archive*. Disponible en <http://www-history.mcs.st-and.ac.uk/> Consultado el 24 de septiembre de 2019.

# Índice analítico

- Agrupamiento de valores, 17
- Asimetría
  - coeficiente de, 51
- Binomio
  - teorema del, 313
- Boxplots, 74
- Browniano
  - movimiento, 127, 138
- Caja y brazos
  - diagrama de, 74
- Cantidad pivotal, 248
- Censo, 5
- Chebyshev
  - desigualdad de, 148
- CICR, 154
- Clase modal, 25
- Clases, 17
  - marcas de, 18
- Coefficiente
  - binomial generalizado, 313
  - de asimetría, 51
  - de una v.a., 84
  - de variación, 39
  - de una v.a., 84
- Completez, 224
- Condiciones de regularidad, 162
- Conjunto
  - bimodal, 25
  - multimodal, 25
  - unimodal, 25
- Consistencia, 142
- Convergencia
  - casi segura, 317
  - débil, 318
  - de v.a.s, 317
  - en distribución, 318
  - en media, 318
  - en media cuadrática, 318
  - en probabilidad, 317
  - puntual, 317
- Cota inferior
  - de Cramér-Rao, 154
- Cramér-Rao
  - cota inferior, 154
- Cuantiles, 48
  - de una v.a., 84
- Cuartiles, 50
- Curtosis, 54
  - curva leptocúrtica, 55
  - curva mesocúrtica, 55
  - curva platicúrtica, 55
  - de una v.a., 84
- Datos, 5, 6

- Datos agrupados
  - descripciones numéricas, 56
- Deciles, 50
- Descripciones
  - gráficas, 60
  - numéricas, 18
- Descripciones numéricas
  - para datos agrupados, 56
- Desigualdad
  - de Chebyshev, 148
- Desviación
  - estándar, 34
  - de una v.a., 84
  - para datos agrupados, 35
  - media, 36
  - de una v.a., 84
  - típica, 34
- Diagrama
  - de caja y brazos, 74
  - de tallo y hojas, 71
- Distribución
  - doble exponencial, 105
  - empírica, 76
  - Rayleigh, 105, 125
  - tipo exponencial, 238
- ECM, 150
- Eficiencia, 165
  - relativa, 166
- Error
  - cuadrático medio, 150
  - tipo I, 276
  - tipo II, 276
- Escala de medición, 12
  - de intervalo, 14
  - de razón, 14
  - nominal, 12
  - ordinal, 12
- Espacio parametral, 86
- Esperanza condicional, 211
  - propiedades, 212
- Estadística, 89
  - s suficientes conjuntamente, 194
  - completa, 225
  - de orden, 90
  - función de otra, 199
  - suficiente, 170
  - suficiente minimal, 200
- Estimación
  - por intervalos, 245
  - puntual, 85
- Estimador
  - asintóticamente eficiente, 165
  - asintóticamente insesgado, 138
  - consistente, 142
  - de máxima verosimilitud, 109
  - eficiencia de un, 165
  - eficiente, 165
  - error cuadrático medio de un, 151
  - insesgado, 127
  - máximo verosímil, 109
  - puntual, 91
  - sesgado, 138
  - sesgo de un, 138, 151
- Euler
  - Fórmula de, 313
- Fórmula
  - s de derivación, 313
  - s de integración, 313
  - s para exponentes, 311
  - s para logaritmos, 312
  - s para sumas, 312

- de Euler, 313
  - de integración por partes, 315
  - de Leibnitz, 315, 318
  - de Stirling, 315
- Factorización
  - teorema de, 195
- Familia de distribuciones
  - completa, 225
- Familia exponencial, 238
  - momentos, 321
- Fisher
  - información de, 182
- Frecuencias, 43
  - absolutas, 43
  - absolutas acumuladas, 44
  - acumuladas, 44
  - relativas, 46
  - relativas acumuladas, 47
  - relativas porcentuales, 46
- Función
  - beta, 316
  - de distribución, 82
    - empírica, 76
  - de verosimilitud, 108
  - gama, 316
  - parametral, 119
  - potencia, 278
- Gráfica
  - de barras, 60
  - de pastel, 68
  - de tallo y hojas, 71
- Grado de confianza, 246
- Hipótesis
  - compuesta, 274
  - estadística, 273
  - simple, 274
- Histograma, 63
- Identidades trigonométricas, 312
- Información de Fisher, 182
  - reparametrización, 191
- Insesgamiento, 127
  - asintótico, 138
- Intervalo
  - s de confianza conjuntos, 264
  - de confianza, 246
  - grado de confianza, 246
  - lim inferior, 246
  - lim superior, 246
- Intervalo modal, 25
- Intervalos
  - estimación por, 245
- Invarianza
  - principio de, 122
- Lehmann-Scheffé
  - teorema de, 229
- Leibnitz
  - fórmula de, 315, 318
- Lema
  - de Neyman-Pearson, 302
- Método
  - de máxima verosimilitud, 108
  - de momentos, 94
  - pivotal, 247
- Marca de clase, 18
- Matriz
  - hessiana, 320
- Media, 19
  - aritmética, 19
  - armónica, 24

- de una v.a., 84
- geométrica, 23
- muestral, 89
- para datos agrupados, 20
- Mediana, 26
  - de una v.a., 84
- Medición
  - escalas de, 12
- Medidas
  - de dispersión, 30
  - de localización, 19
  - de tendencia central, 19
- Menores principales, 320
- Moda, 24
  - de una v.a., 84
- Momentos, 40
  - centrales, 40
  - de una v.a., 84
  - método de, 94
  - muestrales, 91, 95
  - poblacionales, 94
- Movimiento browniano, 127, 138
- Muestra, 5
  - tamaño de la, 5
- Muestra aleatoria, 88
  - tamaño de una, 88
- Neyman
  - teorema de factorización de, 173
- Neyman-Pearson
  - lema de, 302
- Nivel de significancia, 275
- Observación
  - unidad de, 4
- Ojiva, 67
- Outliers, 75
- Percentiles, 50
- Población, 3
- Poisson
  - proceso de, 126, 138
- Polígono
  - de frecuencias, 66
  - de frecuencias acumuladas, 66
- Porcentajes, 46
- Potencia
  - función, 278
- Principio
  - de invarianza, 122
- Proceso de Poisson, 126, 138
- Prueba de hipótesis, 267, 275
  - nivel de significancia, 275
  - para la varianza, 299
  - región crítica, 275
- Puntos críticos, 319
- Rango, 38
  - de una v.a., 84
  - intercuartil, 75
- Rao-Blackwell
  - teorema de, 216
- Región crítica, 275
  - tamaño de la, 275
- Región de rechazo, 269, 275
- Regla de la cadena, 314
- Regularidad
  - condiciones de, 162
- RIC, 75
- Sesgo, 138, 150
- Skewness, 51
- Stirling, 315
- Suficiencia, 169
  - conjunta, 193

- métodos para probar —, 210
- minimal, 198
- minimal, teorema para, 201
- función de, 108
- Whiskers, 74
- Tallo y hojas
  - gráfica de, 71
- Teorema
  - de convergencia dominada, 319
  - de convergencia monótona, 319
  - de factorización, 173, 195
  - de Lehmann-Scheffé, 229
  - de Rao-Blackwell, 216
  - del binomio, 313
  - para la suficiencia minimal, 201
- UMVUE, 158, 229
- Unidad de observación, 4
- Valores
  - agrupamiento de, 17
  - atípicos, 75
- VARIABLES, 6
  - aleatorias, 81
  - categorías, 8
  - clasificación de, 8
  - continuas, 10
  - cualitativas, 8
  - cuantitativas, 8
  - dicotómicas, 10
  - discretas, 9
  - mixtas, 10
- Varianza, 31
  - de una v.a., 84
  - muestral, 90
  - para datos agrupados, 32
  - prueba de hipótesis, 299
- Verosimilitud





**Una introducción a la estadística inferencial**  
editado por la Facultad de Ciencias  
de la Universidad Nacional Autónoma de México,  
se terminó de imprimir el 20 de octubre de 2019  
en los talleres de Amy Soluciones Gráficas, S. A. de C. V.  
Corregidora 79, Santa Anita, Iztacalco.  
C.P. 8300. Ciudad de México

El tiraje fue de 540 ejemplares.  
Está impreso en papel book creamy de 60 g.  
En su composición se utilizó tipografía  
computer modern de 11/13 puntos de pica.

Tipo de impresión: offset

El cuidado de la edición estuvo a cargo de  
*Mercedes Perelló Valls*



La estadística es un área muy amplia y diversa de las matemáticas. Sus aplicaciones han abarcado prácticamente todas las disciplinas del quehacer humano. En este trabajo se proporciona una introducción a tres grandes temas clásicos de la estadística inferencial relativos al problema de la estimación de parámetros: la estimación puntual, la estimación por intervalos y las pruebas de hipótesis.

En todos los casos el énfasis principal ha sido puesto en la estimación de parámetros de las distribuciones de probabilidad; sin embargo los métodos y las ideas aquí expuestas también son aplicables para tratar otros problemas similares.

El enfoque con el que se tratan los temas es principalmente matemático, buscando proveer las demostraciones de casi todos los resultados que se estudian.

ISBN: 978-607-30-2432-7



9 786073 024327

