ESTADÍSTICA MATEMÁTICA

ESTADÍSTICA MATEMÁTICA

Edilberto Cepeda-Cuervo

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE BOGOTÁ
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA

Bogotá, D. C., 2015

ESTADÍSTICA MATEMÁTICA

©Universidad Nacional de Colombia, sede Bogotá
Facultad de Ciencias
Departamento de Estadística

©Editorial Universidad Nacional de Colombia

©Edilberto Cepeda

Primera edición: marzo de 2015

Primera reimpresión: octubre de 2015

Segunda reimpresión: mayo de 2017

ISBN: 978-958-775-249-6

Diagramación en LaTeX

Edilberto Cepeda Cuervo

Edición

Coordinación de Publicaciones - Facultad de Ciencias

Corrección de textos, diseño de cubierta, finalización de arte e impresión

Proceditor, teléfonos: 757 9200, fax: ext. 102. correo electrónico: proceditor@yahoo.es

Prohibida la reproducción total o parcial por cualquier medio sin la autorización escrita del titular de los derechos patrimoniales

Impreso y hecho en Bogotá, D. C., Colombia

Catalogación en la publicación Universidad Nacional de Colombia

Cepeda Cuervo, Edilberto, 1959-
Estadística matemática / Edilberto Cepeda Cuervo. – Primera edición. - Bogotá :
Universidad Nacional de Colombia (Sede Bogotá). Facultad de Ciencias. Departamento de Estadística, 2015
16,5 x 24 cm, xiv, 184 páginas : ilustraciones - (Colección textos)

Incluye referencias bibliográficas e índices

ISBN: 978-958-775-249-6

1. Estadística matemática 2. Probabilidades 3. Variables aleatorias 4. Teoría de la estimación I. Título II. Serie

CDD-21 519.5 / 2015

A la memoria de mis padres:
Ana J. Cuervo y Tito Pablo Cepeda,
y de mi querido hermano Carlos Gonzalo.

Contenido

Introducción xiii

1 Funciones de distribución de variables aleatorias **1**

1.1 Distribuciones discretas 2

1.1.1 Distribución Uniforme Discreta 2

1.1.2 Distribución Bernoulli 3

1.1.3 Distribución Geométrica 4

1.1.4 Distribución Poisson 5

1.1.5 Distribución Binomial 6

1.1.6 Distribución Binomial Negativa 8

1.1.7 Distribución Hipergeométrica 9

1.2 Distribuciones continuas 11

1.2.1 Distribución Normal 12

1.2.2 Distribución *Ji*-cuadrado (χ^2) 16

1.2.3 Distribución Gamma 19

1.2.4 Distribución exponencial 20

1.2.5 Distribución Weibull 22

1.2.6 Distribución Beta 23

1.2.7 Distribución t 25

1.2.8 Distribución F 28

1.2.9 Otras funciones de variables aleatorias 31
1.3 Teorema del límite central 33

2 Estimadores 39
2.1 Conceptos básicos 39
2.2 Método de los momentos 43
2.3 Estimaciones de máxima verosimilitud 46
2.4 Error de estimación 53
2.5 Eficiencia 55
2.6 Consistencia 58

3 Modelos estadísticos 63
3.1 Estadístico suficiente 63
3.2 Familia exponencial uniparamétrica 70
3.3 Forma canónica de la familia exponencial uniparamétrica . . . 72
3.4 Familia exponencial biparamétrica 76
3.5 Forma canónica de la familia exponencial biparamétrica 78

4 Propiedades de los estimadores 81
4.1 Errores cuadrático medio y medio absoluto 81
4.2 Completitud 83
4.3 Estadísticos insesgados uniformes de mínima varianza 86
4.3.1 Desigualdad de información 90

5 Intervalos de confianza 97
5.1 Para la media de una distribución Normal 98
5.1.1 con varianza conocida 98
5.1.2 con varianza desconocida 103

5.2 Intervalos de confianza aproximados para la probabilidad de éxito 104
5.3 Otros intervalos para una proporción 108
5.4 Probabilidad de cobertura 109
5.5 Intervalos de confianza para la diferencia de medias 112
5.6 Intervalos de confianza aproximados para la diferencia de proporciones 114
5.7 Intervalos de confianza para σ^2 117
5.8 Intervalos de confianza para σ_1^2/σ_2^2 118

6 Prueba de hipótesis 121

6.1 Prueba de hipótesis simple 121
6.2 Prueba de hipótesis compuesta 122
6.3 Probabilidad de error 123
6.4 Pruebas de hipótesis 124
6.4.1 para el parámetro de la distribución binomial 125
6.5 Prueba de hipótesis referente a la media 127
6.5.1 de una distribución Normal con varianza conocida . . . 127
6.5.2 de una distribución Normal con varianza desconocida . 129
6.6 Prueba de hipótesis para una diferencia de medias 132
6.6.1 usando estadísticos normalmente distribuidos 132
6.6.2 usando estadísticos con distribución t 133
6.7 Pruebas de hipótesis referentes a proporciones 135
6.7.1 referente a una proporción 135
6.7.2 referente a la diferencia de dos proporciones 136
6.8 Pruebas de hipótesis referentes a varianzas 138
6.8.1 de una población 138
6.8.2 de dos poblaciones 139

6.9 Potencia de una prueba 141
6.9.1 Función de potencia 142
6.9.2 El p-valor 146
6.10 Intervalos de confianza y pruebas de hipótesis 148
6.11 Hipótesis compuestas 149

7 Razón de verosimilitud 153
7.1 Lema de Neyman-Pearson 153
7.2 Prueba de razón de verosimilitud 162

Referencias bibliográficas 168

A Software estadístico en R 171
A.1 Distribuciones de variables aleatorias 171
A.1.1 Tablas estadísticas 171
A.1.2 Gráficas de las funciones de densidad y distribución 173
A.1.3 Aproximación de la distribución Binomial a la Normal 174
A.1.4 Estudio de la distribución de los datos 175
A.2 Estimaciones de máxima verosimilitud 177
A.3 Intervalos de confianza y pruebas de hipótesis 179
A.3.1 Inferencia para la media poblacional 179
A.3.2 Inferencia para la varianza poblacional 182
A.3.3 Inferencia sobre la probabilidad de éxito 184

B Tablas estadísticas 187

Índice alfabético 195

Introducción

En este libro el autor presenta un conjunto de tópicos fundamentales de la Estadística Matemática, cuyo estudio requiere conocimientos de probabilidad y cálculo diferencial, integral y vectorial. Está diseñado para desarrollar un curso de Inferencia Estadística, de un semestre, en programas de matemática, física, ingeniería y estadística, y puede ser utilizado como libro-texto en cursos de maestría en Estadística, Economía o Ingeniería Industrial, especialmente si se incluye el capítulo 4 titulado "Propiedades de los estimadores".

En el primer capítulo se estudian las funciones de variables aleatorias independientes, con distribución común. Está conformado por tres secciones: las dos primeras relacionadas con funciones de distribución de variables aleatorias discretas y continuas, y una tercera, titulada "Teorema del límite central". El objetivo de este capítulo es construir funciones de variables aleatorias a partir de una muestra aleatoria y determinar sus funciones de distribución.

En el capítulo 2 se definen algunas funciones de variables aleatorias, llamadas estimadores o estadísticos, que se utilizan para hacer estimación puntual de parámetros como la media y la varianza de una población, y se presenta el método de los momentos y el método de máxima verosimilitud para la obtención de estos. Este capítulo también incluye criterios para la evaluación y comparación de estimadores.

En el capítulo 3 se trata el concepto de estadístico suficiente y la familia exponencial uniparamétrica de distribuciones, determinando su representación en forma natural. Usando el teorema de factorización se determinan, en forma general, los estadísticos suficientes para el parámetro de una función de distribución perteneciente a la familia exponencial uniparamétrica y, finalmente, se incluyen algunas extensiones a la familia exponencial biparamétrica de distribuciones.

El capítulo 4 considera propiedades de los estimadores, que permiten determinar aquellos que tienen un buen desempeño. En el capítulo 5 se da el concepto de intervalo de confianza y de probabilidad de cobertura, y se hallan intervalos de confianza para los parámetros de una distribución Normal y una distribución Binomial. En el capítulo 6 se consideran algunos tópicos de prueba de hipótesis y se establece la relación entre estas y los intervalos de confianza. Finalmente, en el capítulo 7 se presenta el lema de Neyman-Pearson, el concepto de pruebas uniformemente más potentes y pruebas de razón de verosimilitudes.

Este libro contiene un apéndice en el que se incluyen funciones computacionales en el software libre R, asociadas a los contenidos del libro, que permiten hacer gráficas de funciones de densidad y de distribución, obtener muestras aleatorias de una población y hacer inferencias estadísticas a partir de ciertos conjuntos de datos.

La elaboración de este libro contó con la colaboración de muchos de mis estudiantes de los cursos de Inferencia Estadística y de Estadística Matemática, ofrecidos por el Departamento de Estadística de la Universidad Nacional de Colombia, especialmente en su versión inicial correspondiente a las notas de clase. Varios de los problemas que se incluyen fueron propuestos en el proceso de preparación de las clases y otros pensados, propuestos y desarrollados durante ellas.

Otras personas colaboraron en el proceso de este libro, pero la colaboración de María Fernanda Zárate Jiménez y del profesor Fernando Ruiz Guzmán fue muy importante. Ella, por su colaboración en el desarrollo del apéndice titulado “Software estadístico en R”, que incluye elementos de computación estadística, relacionados con tablas estadísticas, gráficas de funciones de densidad y distribución y elementos computacionales de estimación, intervalos de confianza y pruebas de hipótesis. Él, por la lectura cuidadosa y detallada de este libro y sus contribuciones en el uso del lenguaje, dando coherencia y armonía al texto, para que este sea más agradable y de más fácil comprensión para los lectores. A los dos muchas gracias.

Capítulo 1

Funciones de distribución de variables aleatorias

Este capítulo tiene como objetivo estudiar las funciones de distribución de variables aleatorias. Una variable aleatoria es una función a valor real definida sobre un espacio muestral. Por ejemplo, en el lanzamiento de dos monedas, el espacio muestral es el conjunto $\Omega = \{HH, HT, TH, TT\}$, conformado por todos los resultados posibles (eventos) de este experimento. Una variable aleatoria X, definida sobre Ω, está dada por $X(HH) = 2$, $X(HT) = X(TH) = 1$ y $X(TT) = 0$, el número de caras (*heads*) obtenidas en un lanzamiento de dos monedas. Si los lanzamientos son independientes y en el lanzamiento de cada una de las monedas, la probabilidad de obtener H o T es la misma, entonces $P(X = 0) = 1/4$, $P(X = 1) = 1/2$ y $P(X = 2) = 1/4$.

La función de distribución de una variable aleatoria X está definida por $F_X(x) = P[X \leq x]$, para todo $x \in \mathbb{R}$. El anterior ejemplo es de variables aleatorias discretas, es decir, variables aleatorias cuyo recorrido es un conjunto discreto.

Si una función de distribución es continua, entonces $P[X = x] = 0$ para todo $x \in \mathbb{R}$. Una variable aleatoria con función de distribución continua se llama variable aleatoria continua. Por ejemplo, una variable aleatoria normalmente distribuida es una variable continua.

Como una variable aleatoria X es una función a valor real, su función de distribución $F_X(x) = P(X \leq x)$ es continua si es continua en el conjunto

de los números reales. Así, del concepto de continuidad de una función, se concluye que $P[X = x_0] = 0$, para todo $x_0 \in \mathbb{R}$. Un ejemplo de estas funciones de distribución es la distribución Normal.

Este capítulo está dividido en tres secciones tituladas: distribuciones discretas, distribuciones continuas y teorema del límite central.

1.1 Distribuciones discretas

En esta sección se estudian algunas funciones de distribución de variables aleatorias discretas. Está dividida en siete subsecciones tituladas: distribución Uniforme Discreta, distribución Bernoulli, distribución Geométrica, distribución Poisson, distribución Binomial, distribución Binomial Negativa y distribución Hipergeométrica.

1.1.1 Distribución Uniforme Discreta

Una variable aleatoria X tiene distribución Uniforme Discreta en $A = \{1, 2, \ldots, n\}$, si su función de probabilidad está dada por

$$f(x) = \frac{1}{n} I_A(x), \tag{1.1}$$

donde I_A es la función indicadora.

Esto significa que la variable aleatoria X puede tomar cualquier valor de A con probabilidad $1/n$; y que para todo número real x, que no pertenezca a A, $f(x) = 0$.

Esta distribución aparece en un conjunto amplio de experimentos. Por ejemplo, en el lanzamiento de un dado, la variable aleatoria "número de puntos obtenidos" tiene una distribución Uniforme en el conjunto $A = \{1, 2, 3, 4, 5, 6\}$, con $f(x) = 1/6$, para toda x que pertenece a A y $f(x) = 0$ en caso contrario.

La función de distribución de una variable aleatoria X, uniformemente distribuida en A, está dada por

$$F(x) = P[X \leq x] = \sum_{i=1}^{n_0} f(i), \tag{1.2}$$

donde n_0 es el mayor entero menor o igual que x.

La media, la varianza y la función generadora de momentos de una variable aleatoria X que tiene una función de probabilidad dada por la ecuación (1.1) son

$$E(X) = \frac{n+1}{2}$$
$$Var(X) = \frac{n^2-1}{12}$$
$$M_X(t) = E[\exp(tX)] = \frac{e^t(1-e^{nt})}{n(1-e^t)} = \frac{1}{n}\sum_{j=1}^{n} e^{tj}, \quad t \in \mathbb{R}.$$

1.1.2 Distribución Bernoulli

Una variable aleatoria X tiene función de distribución Bernoulli si su función de probabilidad está definida por

$$f(x) = \theta^x(1-\theta)^{1-x} I_A(x), \quad 0 \leq \theta \leq 1, \tag{1.3}$$

donde $A = \{0, 1\}$.

Esto significa que la variable aleatoria X puede tomar los valores 0 y 1 con probabilidades $P(X = 1) = \theta$ y $P(X = 0) = 1 - \theta$, respectivamente. En general, estas variables aleatorias aparecen vinculadas a experimentos en los cuales los resultados pueden asociarse a dos categorías, denominadas éxito y fracaso.

La media, la varianza y la función generadora de momento de una variable aleatoria con distribución Bernoulli son

$$E(X) = \theta$$
$$Var(X) = \theta(1-\theta)$$
$$M_R(t) = E[\exp(tX)] = 1 - \theta + \theta e^t, t \in \mathbb{R}.$$

La función de distribución Bernoulli es

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1-\theta & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x \geq 1 \end{cases}$$

La siguiente expresión, $X \sim Ber(\theta)$, indica que la variable aleatoria X tiene distribución Bernoulli con probabilidad de éxito $P(X = 1) = \theta$.

1.1.3 Distribución Geométrica

Una variable aleatoria X tiene función de distribución Geométrica si su función de probabilidad está definida por

$$f(x) = \theta(1-\theta)^{x-1} I_A(x), \quad 0 \leq \theta \leq 1, \tag{1.4}$$

donde $A = \{1, 2, 3, \dots\}$.

En la ecuación (1.4), θ denota la probabilidad de éxito en un experimento. Así, $f(1)$ es la probabilidad de que el primer éxito ocurra en el primer experimento, $f(2)$ es la probabilidad de que el primer éxito ocurra en el segundo experimento y así sucesivamente.

La media, la varianza y la función generadora de momentos de una variable aleatoria con esta distribución son

$$\begin{aligned} E(X) &= \frac{1}{\theta} \\ Var(X) &= \frac{1-\theta}{\theta^2} \\ M_R(t) &= E[\exp(tX)] = \frac{\theta e^t}{1-(1-\theta)e^t}, \quad t < -ln(1-\theta). \end{aligned}$$

La función de distribución Geométrica está definida por

$$\begin{aligned} F(x) &= P[X \leq x] \\ &= \sum_{i=1}^{[x]} f(i) \\ &= 1 - P[X > x] \\ &= 1 - P([x] \textit{ primeros eventos son fallas}), \end{aligned} \tag{1.5}$$

donde $[x]$ representa la parte entera de x.

En la figura 1.1, se muestran las gráficas de la distribución Geométrica y de su función de probabilidad para $\theta = 0.3$ y $\theta = 0.6$.

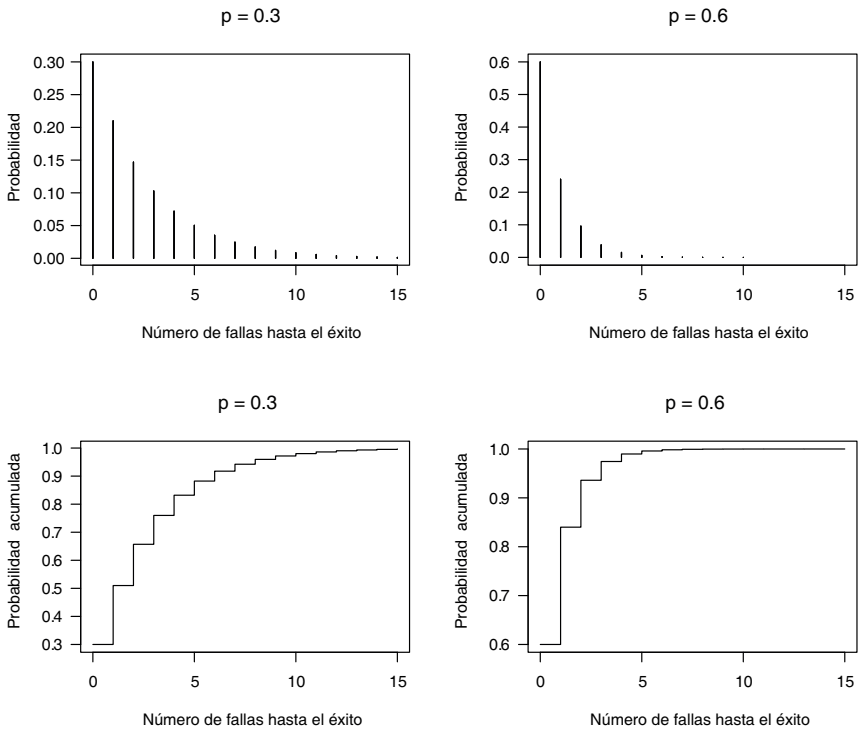

Figura 1.1. Función de probabilidad y función de distribución Geométrica para $\theta = 0.3$ y $\theta = 0.6$.

1.1.4 Distribución Poisson

Una variable aleatoria X tiene función de distribución Poisson, con parámetro $\lambda > 0$, si su función de probabilidad está dada por

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!} I_A(x), \tag{1.6}$$

donde $A = \{0, 1, 2, \dots\}$.

La distribución Poisson constituye, en algunos casos, un buen modelo para explicar el número de eventos que ocurren en un intervalo de tiempo o en una región geográfica. Por ejemplo, cuando se considera el número de llamadas que recibe un operador telefónico diariamente, asumiendo que estas son independientes. Esta distribución también es útil en el análisis de variables aleatorias como el número de personas atendidas diariamente por un banco o el número de accidentes de tránsito que ocurren anualmente en una ciudad.

Si una variable aleatoria X tiene distribución Poisson con parámetro λ, lo que se nota $X \sim P(\lambda)$, por tanto,

$$\begin{aligned} E(X) &= \lambda \\ Var(X) &= \lambda \\ M_X(t) &= E[\exp(tX)] = \exp(\lambda(e^t - 1)). \end{aligned}$$

La suma de variables aleatorias independientes, con distribución Poisson, tiene distribución Poisson. Esto es, si $X_i \sim P(\lambda_i), i = 1, 2, \ldots, n$; son n variables aleatorias independientes, entonces, la variable aleatoria $X = X_1 + X_2 + ... + X_n$ tiene distribución Poisson, con parámetro $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_n$.

1.1.5 Distribución Binomial

Si X_i, $i = 1, 2, \ldots, n$; son n variables aleatorias independientes que tienen distribución Bernoulli con parámetro θ, entonces, la variable aleatoria

$$X = \sum_{i=1}^{n} X_i \tag{1.7}$$

tiene distribución Binomial con parámetros n y θ y se nota $X \sim Bin(n, \theta)$.

La función de probabilidad de una variable aleatoria que tiene una distribución Binomial, con parámetros n y θ, es

$$f(x) = \binom{n}{x} \theta^x (1-\theta)^{n-x} I_{\{0,1,\ldots,n\}}(x), \tag{1.8}$$

donde $0 \leq \theta \leq 1$ y $\binom{n}{x} = \frac{n!}{x!(n-x)!}$.

Esto significa que una variable aleatoria Binomial X toma valores en el conjunto $A = \{0, 1, 2, 3, \ldots, n\}$ con probabilidades determinadas por (1.8). La figura 1.2 muestra la gráfica de la función de probabilidad y de densidad de la distribución Binomial $Bin(15, 0.3)$ y $Bin(15, 0.6)$.

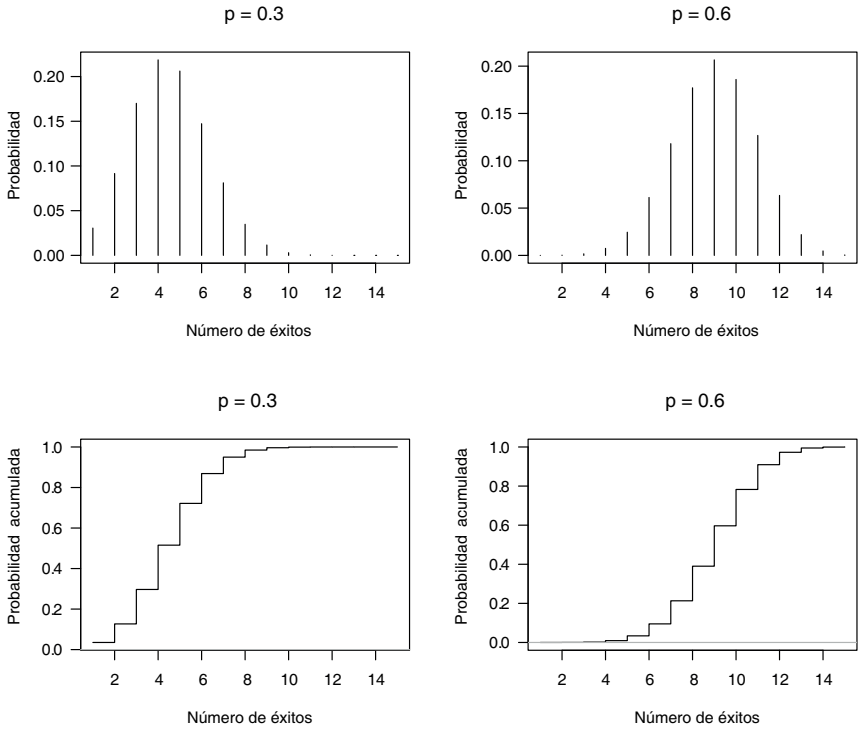

Figura 1.2. Distribución Binomial: funciones de probabilidad y distribución.

La media, la varianza y la función generadora de momento de una variable aleatoria con distribución Binomial son

$$E(X) = n\theta$$
$$Var(X) = \sum_{i=1}^{n} V(X_i) = n\theta(1-\theta)$$
$$M_X(t) = E[\exp(tX)] = (1-\theta+\theta e^t)^n.$$

Si una variable aleatoria X representa el número de éxitos que ocurren en n eventos con probabilidad de éxito θ, se dice que la variable aleatoria X tiene distribución Binomial con parámetros n y θ.

Esta distribución aparece en múltiples casos, donde ocurren n eventos independientes, cada uno con probabilidad de éxito θ. Por ejemplo, en el lanzamiento de una moneda, si el éxito es obtener cara y $P(\text{éxito}) = 0.5$, la variable aleatoria "número de caras en 10 lanzamientos" tiene distribución

Binomial con parámetros $n = 10$ y $\theta = 0.5$. En el lanzamiento de un dado, si el éxito es obtener 1 o 6, la variable aleatoria "número de éxitos en 20 lanzamientos independientes" tiene distribución Binomial con parámetros $n = 20$ y $\theta = 1/3$.

En otro ejemplo considere un experimento en el que se seleccionan aleatoriamente elementos de un conjunto $\{a_1, a_2, \ldots, a_N\}$, para determinar si tienen o no una propiedad P. En este caso, si el muestreo se hace con reemplazo, las observaciones son independientes y la variable aleatoria "número de elementos con la propiedad P"tiene distribución Binomial. La propiedad P puede ser defectuosa, poseer un virus o apoyar a un candidato, entre otras.

1.1.6 Distribución Binomial Negativa

Una variable aleatoria Binomial Negativa, con parámetros n y θ, hace referencia al número de fallas, x, necesarias para que ocurran n éxitos. Así, en una sucesión de experimentos Bernoulli, con probabilidad de éxito θ, la probabilidad de que el n-ésimo éxito ocurra en el $(n + x)$-ésimo evento es igual al producto de la probabilidad de éxito en el $(n + x)$-ésimo evento y la probabilidad de x fracasos en los primeros $(n + x - 1)$ eventos. Así, la distribución Binomial Negativa está definida por

$$f(x) = \binom{n+x-1}{x} \theta^n (1-\theta)^x I_{\{0,1,\ldots\}}(x), \quad 0 \leq \theta \leq 1. \tag{1.9}$$

Esto significa que una variable aleatoria X toma valores en el conjunto de los números naturales con probabilidades determinadas por 1.9.

La media, la varianza y la función generadora de momento de una variable aleatoria con distribución Binomial Negativa son

$$\begin{aligned}
E(X) &= \frac{n(1-\theta)}{\theta} \\
Var(X) &= \frac{n(1-\theta)}{\theta^2} \\
M_X(t) &= E[\exp(tX)] \\
&= \left(\frac{\theta}{1-(1-\theta)e^t}\right)^n, \quad t < -\ln(1-\theta).
\end{aligned}$$

Por ejemplo, la distribución Binomial Negativa hace referencia a la probabilidad de que el décimo carro producido por *Carros S.A.* sea el tercer carro defectuoso de la producción, dado que se ha observado que la producción de carros defectuosos es del 3 %. En este caso, $X = 7$ y los parámetros son $n = 3$ y $\theta = 0.03$, considerando que producir un solo carro defectuoso corresponde a un éxito.

1.1.7 Distribución Hipergeométrica

Sea S una población conformada por N elementos, por ejemplo, una producción de N tornillos o N resistencias eléctricas. Al mismo tiempo, se asume que existe un número no conocido de elementos defectuosos en la población. En estas condiciones, el número de elementos defectuosos en una muestra aleatoria de tamaño n, es una variable aleatoria X, que toma valores sobre el conjunto $A = \{0, 1, 2, \ldots, n\}$ y que tiene la siguiente función de probabilidad:

$$P(x) = \frac{\binom{N\theta}{x}\binom{N-N\theta}{n-x}}{\binom{N}{n}} I_A(x), \tag{1.10}$$

donde θ es la proporción de elementos defectuosos en la población, $N\theta$ es el número de elementos defectuosos en la población y $\text{máx}(n - N(1-\theta), 0) \leq x \leq \text{mín}(N\theta, n)$.

En esta función de probabilidad θ y, por consiguiente, $N\theta$ son desconocidos. La ecuación (1.10) resulta del cociente entre el número de maneras posibles de obtener x elementos defectuosos y $n - x$ elementos no defectuosos de una población de tamaño N, y el número de maneras posibles de obtener n elementos en dicha población. Para notar que una variable aleatoria X tiene distribución Hipergeométrica se escribe $X \sim H(N\theta, N, n)$.

La media y la varianza de una variable aleatoria con distribución Hipergeométrica están dadas por:

$$E(Y) = n\theta \tag{1.11}$$

$$V(Y) = \frac{n\theta(N - N\theta)(N - n)}{N(N-1)}. \tag{1.12}$$

Las demostraciones de (1.11) y (1.12) pueden verse en Mood *et al.* (1969). La función generadora de momentos de la distribución hipergeométrica no tiene forma explicita.

Ejercicios 1.1.

1. Sea $A = \{1, 2, \ldots, n\}$ un conjunto de números reales. Halle la media y la varianza de la función de probabilidad $f(x_i) = \frac{1}{n} I_A(x_i)$.

   Ayuda: $1 + 2 + \cdots + n = \frac{n(n+1)}{2}$

2. Sean $X_i \sim Geom(\theta)$, $i = 1, 2, \ldots, n$; n variables aleatorias independientes. Determine la función de densidad y la función generadora de momentos de $Y = \sum_{i=1}^{n} X_i$.

3. Halle la función de probabilidad de $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$, asumiendo que X_i, $i = 1, 2, \ldots, n$, es una muestra aleatoria de una distribución:

   a) Bernoulli.

   b) Poisson.

   c) Binomial.

4. Para cada una de las funciones de probabilidad de $\bar{X}$, determinadas en el ejercicio anterior, halle la función generadora de momentos, la media y la varianza.

5. Determine el valor esperado de la variable aleatoria $I_{\{X \leq 1\}}$, si X es una variable aleatoria con distribución:

   a) Binomial $B(n, \theta)$.

   b) Poisson de parámetro θ.

6. Si $X \sim Bin(10, \theta)$, $p(2) = 0.1$ y $p(3) = 0.2$, halle $E(X)$ y $V(X)$.

7. Si $X \sim P(\lambda)$ y $p(2) = 2e^{-2}$, hallar $P(X > 3)$, $E(X)$ y $V(X)$.

8. Si X tiene distribución Geométrica, con parámetro θ, halle $E(X)$, $E(X^2)$ y $M_X(t)$, aplicando la definición de valor esperado de una variable aleatoria.

1.2 Distribuciones continuas

Una variable aleatoria X tiene función de distribución continua, si existe una función f no negativa, definida en los números reales, tal que

$$P[X \in A] = \int_A f(x)dx \tag{1.13}$$

para todo $A \subseteq \mathbb{R}$, que cumple las siguientes propiedades:

1. $f(x) \geq 0$, para todo $x \in \mathbb{R}$
2. $\int_{\mathbb{R}} f(x)dx = 1$

La función f se denomina función de densidad.

La función de distribución Uniforme, en un intervalo (a, b), es un ejemplo de distribución continua y está definida por la ecuación

$$F(x) = \begin{cases} 0 & \text{si } x \leq a \\ \frac{x-a}{b-a} & \text{si } a < x < b \\ 1 & \text{si } x \geq b \end{cases} \tag{1.14}$$

En este caso, como en el de todas las funciones de distribución continua, la función de densidad no es única. Por ejemplo, si

$$f_1(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in (a, b) \\ 0 & \text{si } x \notin (a, b) \end{cases} \tag{1.15}$$

y $f_2(x) = f_1(x)$, excepto para un número finito de puntos en el intervalo (a, b), para los que $f_2(x) = 0$, $\int_A f_1(x)dx = \int_A f_2(x)dx$. Así,

$$F(x) = \int_{-\infty}^{x} f_1(x)dx = \int_{-\infty}^{x} f_2(x)dx. \tag{1.16}$$

En esta sección se presentan algunas funciones de distribución continua y el teorema del límite central. Se encuentra dividida en 8 subsecciones tituladas: distribución Normal, distribución Ji-cuadrado, distribución Gamma, distribución Weibull, distribución Beta, distribución t, distribución F y distribuciones de otras funciones de variables aleatorias.

1.2.1 Distribución Normal

Una variable aleatoria X tiene distribución Normal, si y solo si su función de densidad es

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-\theta)^2\right], \tag{1.17}$$

donde $-\infty < x < \infty$, $\sigma > 0$ y $-\infty < \theta < \infty$.

Los parámetros $\theta = E(X)$ y $\sigma^2 = Var(X)$ corresponden a la media y la varianza de la variable aleatoria X, respectivamente, y si $\theta = 0$ y $\sigma^2 = 1$, se dice que X tiene distribución Normal estándar.

Para notar que una variable aleatoria X tiene distribución Normal, con media θ y varianza σ^2, se escribe $X \sim N(\theta, \sigma^2)$.

La gráfica de la función de densidad de una variable aleatoria que tiene distribución Normal estándar es simétrica con respecto a $X = 0$, como se muestra en la figura 1.3.

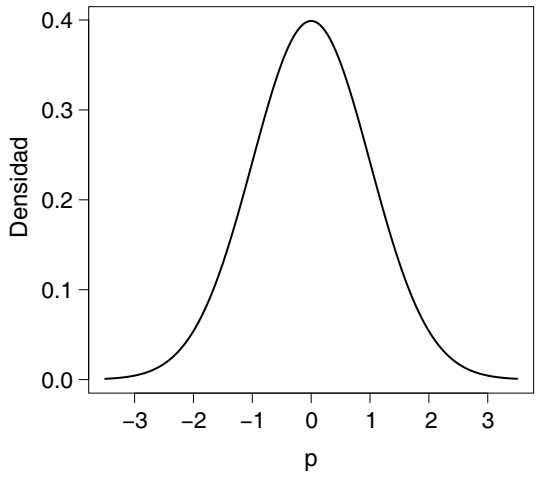

Figura 1.3. Función de densidad normal estándar.

Para demostrar que (1.17) define una función de densidad, basta con verificar que $\int_{-\infty}^{\infty} f_X(x)dx = 1$, pues $f_X(x) \geq 0$. Este resultado se sigue de la igualdad:

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}dxdy = 1. \tag{1.18}$$

La integral de la ecuación (1.18) se resuelve usando coordenadas polares. Haciendo $x = rcos(\theta)$ y $y = rsen(\theta)$, se encuentra que

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}dxdy = \int_0^{2\pi}\int_0^{\infty} e^{-r^2/2}rdrd\theta = 2\pi.$$

La función generadora de momentos de una variable aleatoria $X \sim N(\theta, \sigma^2)$, normalmente distribuida, con media θ y varianza σ^2, es

$$\begin{aligned} M_X(t) = E[\exp(tX)] &= \int_{-\infty}^{\infty} \exp(tx)\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2\sigma^2}(x-\theta)^2\right]dx \\ &= \exp\left(\theta t + \frac{\sigma^2\, t^2}{2}\right). \end{aligned}$$

A partir de la función generadora de momentos, calculando sus derivadas con respecto a t y evaluándolas en $t = 0$, se obtienen los momentos de Y con respecto al origen. Por ejemplo:

$$\begin{aligned} E(X) &= M'_X(t)|_0 = (\theta + \sigma^2 t)\exp\left(\theta t + \frac{1}{2}\,\sigma^2 t^2\right)|_0 = \theta \\ E(X^2) &= M''_X(t)|_0 \\ &= \sigma^2 \exp\left(\theta t + \frac{1}{2}\,\sigma^2 t^2\right)|_0 + (\theta + \sigma^2 t)^2 \exp\left(\theta t + \frac{1}{2}\,\sigma^2 t^2\right)|_0 \\ &= \sigma^2 + \theta^2. \end{aligned}$$

Obsérvese que $Var(X) = E(X^2) - E^2(X)$.

Teorema 1.1. Sean X_i, $i = 1, 2, \ldots, n$; n variables aleatorias independientes normalmente distribuidas, con media θ_i y varianza σ_i^2, respectivamente. Si $a_1, a_2, \ldots, a_n$ son números reales, la variable aleatoria $R = \sum_{i=1}^n a_i X_i$ tiene distribución Normal con media $\theta = \sum_{i=1}^n a_i\theta_i$ y varianza $\sigma^2 = \sum_{i=1}^n a_i^2\sigma_i^2$.

Demostración. La función generadora de momentos de la variable aleatoria R, es

$$\begin{aligned} M_R(t) = E[\exp(tR)] = E\left[\exp(t\sum_{i=1}^n a_i X_i)\right] &= \prod_{i=1}^n E\left[\exp(ta_i X_i)\right] \\ &= \exp\left[t\sum_{i=1}^n a_i\,\theta_i \;+\; \frac{t^2}{2}\sum_{i=1}^n a_i^2\sigma_i^2\right]. \end{aligned}$$

$M_R(t)$ tiene la forma de la función generadora de momentos de una variable aleatoria distribuida normalmente, con media $\theta = \sum_{i=1}^{n} a_i\theta_i$ y varianza $\sigma^2 = \sum_{i=1}^{n} a_i^2\, \sigma_i^2$.

Finalmente, dado que si la función generadora de momentos existe, esta determina unívocamente a la distribución de probabilidad (Shao, 2003), entonces, R tiene distribución Normal con media θ y varianza σ^2. □

A partir del teorema 1.1 se demuestra que

a) Si $X \sim N(\theta, \sigma^2)$, entonces, $Z = \frac{X-\theta}{\sigma}$ tiene distribución Normal estándar.

b) Si $X_i \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$; son n variables aleatorias independientes, entonces, $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ tiene distribución Normal con media θ y varianza σ^2/n.

Ejercicios 1.2.

1. Si X_i, $i = 1, 2, \ldots, n$, es una muestra aleatoria, de tamaño n, de $X \sim N(0, 4)$, ¿cuál es el menor valor de n tal que la media muestral $\bar{X}$ se encuentre a menos de 0.2 unidades de la media poblacional, con una probabilidad del 95 %?

   Nota: Una muestra aleatoria de tamaño n, es un conjunto de n variables aleatorias independientes idénticamente distribuidas.

2. Asuma que la temperatura de una ciudad tiene distribución Normal con media de 18 °C y desviación estándar de 3 °C. Si durante el último mes se realizaron 125 mediciones que tienen una media de 19.5 °C, ¿qué se puede afirmar de la temperatura en el último mes?

   Sugerencia: Calcule $P[\bar{X} \geq 19.5]$.

3. Sean $X_i \sim N(\theta_1, \sigma_1^2)$, $i = 1, 2, \ldots, n_1$, y $Y_i \sim N(\theta_2, \sigma_2^2)$, $i = 1, 2, \ldots, n_2$, dos muestras aleatorias independientes, de tamaño n_1 y n_2, respectivamente; demuestre que la variable aleatoria

$$R = \frac{1}{n_1}\, \Sigma_{i=1}^{n_1} X_i \; - \; \frac{1}{n_2}\, \Sigma_{i=1}^{n_2} Y_i$$

tiene distribución Normal con la siguiente media y varianza:

$$E(R) = \theta_1 - \theta_2 \quad \text{y} \quad V(R) = \frac{1}{n_1}\,\sigma_1^2 \;+\; \frac{1}{n_2}\,\sigma_2^2.$$

¿Qué hipótesis adicionales son necesarias para ello?

4. Dos atletas A y B se preparan para la gran final de 100 metros planos. Si ambos se consideran igualmente rápidos y $\sigma_A^2 = \sigma_B^2 = 4$, ¿cuál es la probabilidad de que el corredor A salga seleccionado, después de 10 carreras selectivas, si se considera que los tiempos siguen una distribución Normal y se exige que la diferencia entre las medias de los tiempos de los atletas sea de, por lo menos, 0.4 segundos?

5. Si $X_i \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$, son n variables aleatorias independientes, determine la función de densidad de la variable aleatoria

$$Y = \frac{\sqrt{n}(\bar{X} - \theta)}{\sigma}$$

   a) Aplicando el teorema (1.1).

   b) Utilizando la función generadora de momentos.

6. Si $Z_i \sim N(0, 1)$, $i = 1, 2$, son dos variables aleatorias independientes, determine la función de densidad de $Y = Z_1 - Z_2$ a partir de la función de distribución y a partir de la función generadora de momentos.

7. Si $Y \sim N(\theta, \sigma^2)$, muestre que $X = \exp(Y)$ tiene función de distribución logarítmica Normal con función de densidad

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}}\;\exp\left[-\frac{1}{2}\left(\frac{ln(x) - \theta}{\sigma}\right)^2\right] I_{(0,\infty)}(x).$$

   Halle $E(X)$ y $Var(X)$.

8. Demuestre que la función de densidad normal es simétrica con respecto a la media. Analice esta función con referencia a valores extremos, concavidades y puntos de inflexión.

9. Si $X \sim N(2, 4)$, $i = 1, 2, \ldots, 9$; hallar $P(|\bar{X} - 2| \leq 2)$.

1.2.2 Distribución *Ji*-cuadrado (χ^2)

En estadística y en probabilidad, es necesario definir funciones de variables aleatorias y conocer su función de distribución. Por ejemplo, si $Z \sim N(0,1)$, entonces, la función de distribución de la variable aleatoria $Y = Z^2$ es

$$\begin{aligned} F(y) = P(Y \leq y) &= P(Z^2 \leq y) \\ &= P(|Z| \leq \sqrt{y}) \\ &= P(-\sqrt{y} \leq Z \leq \sqrt{y}) \\ &= 2\int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\, z^2\right)\, dz. \end{aligned}$$

Aplicando el teorema fundamental del cálculo (Apostol, 1967), se encuentra que la función de densidad de Y es

$$f(y) = \frac{1}{\sqrt{2\pi y}} \exp\left(-\frac{1}{2}y\right) I_{(0,\infty)}(y). \tag{1.19}$$

Si una variable aleatoria tiene la función de densidad definida por la ecuación (1.19), se dice que tiene distribución Ji-cuadrado, con 1 grado de libertad, y se nota $Y \sim \chi^2_{(1)}$.

Definición 1.1. Sean $Z_i \sim N(0,1)$, $i = 1, 2, \ldots, n$, n variables aleatorias independientes. Si $R = \sum_{i=1}^n Z_i^2$, entonces, R tiene distribución Ji-cuadrado con n grados de libertad.

Si una variable aleatoria Y tiene distribución Ji-cuadrado con n grados de libertad, entonces, su función de densidad está dada por

$$f(y) = \frac{y^{(n/2)-1}\, e^{-y/2}}{2^{n/2}\, \Gamma(n/2)}\, I_{(0,\infty)}(y), \tag{1.20}$$

donde $\Gamma(.)$ nota la función Gamma, que está definida por

$$\Gamma(\alpha) = \int_0^\infty e^{-t} t^{\alpha-1} dt \quad \alpha > 0.$$

Integrando por partes se demuestra que $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$.

La expresión $Y \sim \chi^2_{(n)}$ da a entender que la variable aleatoria Y tiene distribución Ji-cuadrado con n grados de libertad.

La función generadora de momentos de una variable $Y \sim \chi^2_{(n)}$ está dada por

$$\begin{aligned} m(t) = E(e^{Yt}) &= \int_0^\infty \frac{y^{(n/2)-1}\, e^{-(y/2)(1-2t)}}{2^{n/2}\,\Gamma(n/2)}\, dy \qquad (1.21) \\ &= (1-2t)^{-n/2}; \quad t < \frac{1}{2}. \end{aligned}$$

La integral en la igualdad (1.21) se calcula por el método de sustitución, haciendo $u = y(1-2t)$.

Ahora se puede demostrar, directamente y por el método de los momentos, que $E(Y) = n$ y $Var(Y) = 2n$.

Teorema 1.2. Sean Y_i, $i = 1, 2, \ldots, n$; n variables aleatorias independientes y normalmente distribuidas con media θ y varianza σ^2 desconocidas. Si $S^2 = \dfrac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$, entonces,

$$\frac{1}{\sigma^2}\, \Sigma_{i=1}^n\, (Y_i - \bar{Y})^2 = \frac{(n-1)\, S^2}{\sigma^2} \qquad (1.22)$$

tiene distribución χ^2 con $(n-1)$ grados de libertad.

Nota: La variable S^2 se denomina varianza muestral, y se utiliza para hacer inferencias relacionadas con la varianza de una población a partir de los datos observados de una muestra aleatoria.

Ejemplo 1.1. Si $Y_i \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, 11$, es una muestra aleatoria, de tamaño once, con media y varianza desconocidas, entonces,

1. $P(S^2 \leq 0.394\sigma^2) = P(\chi^2_{10} \leq 3.94) = 0.05$.
2. $P(0.394\sigma^2 \leq S^2 \leq 1.831\sigma^2) = P(3.94 \leq \chi^2_{10} \leq 18.31) = 0.90$.

Ejercicios 1.3.

1. Considere que la variable aleatoria X tiene distribución χ^2 con 29 grados de libertad. Si $\alpha = 0.05$, determine dos números reales x_i y x_s, tales que
$$P(x_i < X < x_s) = 1 - 2\alpha.$$

2. Si $X_i \sim N(\theta, 36)$, $i = 1, \ldots, 11$, son once variables aleatorias independientes, determine $P(2 < S^2 < 6)$.

3. Una empresa productora de tornillos garantiza que los diámetros de sus tornillos tienen una desviación estándar no mayor que 0.01. Para determinar si la producción satisface las condiciones de calidad establecidas por la compañía, se mide el diámetro de 100 unidades. Si la varianza de estas mediciones es 0.09, ¿qué se puede concluir?

   Sugerencia: calcule $P(S^2 > 0.09)$ y diga ¿qué se puede concluir si $(n-1)s^2/0.0001 = 1$?

4. Una vez determinada la cantidad de zinc que contendrá un medicamento, es importante garantizar una varianza inferior a 0.02 gramos. En un estudio de control de calidad se determina la cantidad de zinc presente en 10 tabletas de la producción, elegidas aleatoriamente, obteniéndose los siguientes resultados:

Tabletas	1	2	3	4	5	6	7	8	9	10
gramos de zinc	0.015	0.017	0.014	0.015	0.018	0.013	0.015	0.016	0.013	0.018

   De acuerdo con estos datos, ¿usted solicitaría detener y revisar el proceso de producción? ¿Qué se debe hacer si se desea garantizar una desviación estándar muestral menor de 0.0001, con una probabilidad del 95 %?

5. Demuestre que una variable aleatoria con distribución Ji-cuadrado con 1 grado de libertad tiene esperanza 1 y varianza 2. Aplique la definición 1.1 para demostrar que si $Y \sim \chi^2_{(n)}$, entonces, $E(Y) = n$ y $Var(Y) = 2n$.

6. Halle $E(Y)$, $E(Y^2)$ y $V(Y)$, asumiendo que $Y \sim \chi^2_{(n)}$.

7. Halle la función generadora de momentos de $Y \sim \chi^2_{(n)}$.

8. Sea $X_i \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$, una muestra aleatoria, de tamaño n, con media θ y varianza σ^2 conocida. Asumiendo que θ_0 es el verdadero

valor de θ, determine la función de distribución de $-2\log\lambda(X)$ donde $X = (X_1, X_2, \ldots, X_n)$, $\lambda(X) = p(X, \bar{X})/p(X, \theta_0)$,

$$p\ (X, \bar{X}) = \frac{1}{(2\pi\sigma^2)^{n/2}}\ \exp\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2\right]$$

y $p(X, \theta)$ nota la función de densidad de la distribución Normal, con media θ y varianza σ^2.

9. Suponga que se tienen dos muestras aleatorias independientes $X_1 = (X_{11}, \ldots, X_{1n_1})$ y $X_2 = (X_{21}, \ldots, X_{2n_2})$ provenientes de distribuciones normales $N(\theta_1, \sigma^2)$ y $N(\theta_2, \sigma^2)$, de tamaños n_1 y n_2, respectivamente, con $\theta_1 \in \mathbb{R}$, $\theta_2 \in \mathbb{R}$ y $\sigma^2 > 0$, desconocidos. Si

$$S^2 = \frac{1}{n_1 + n_2 - 2}\ \left[\Sigma_{i=1}^{n_1}\ (X_{1i} - \bar{X}_1)^2\ +\ \Sigma_{i=1}^{n_2}\ (X_{2i} - \bar{X}_2)^2\right]$$

demuestre que

$$\frac{(n_1 + n_2 - 2)\ S^2}{\sigma^2} \sim \chi^2_{(n_1+n_2-2)}.$$

10. Demuestre el teorema 1.2 para $n = 2$.

11. Halle la función de densidad de $Y = X^{1/2}$, asumiendo que X tiene distribución ji-cuadrado con n grados de libertad.

1.2.3 Distribución Gamma

Se dice que una variable aleatoria X tiene distribución Gamma, con parámetros $p > 0$ y $\lambda > 0$, lo que se nota $X \sim \Gamma(p, \lambda)$, si su función de densidad es

$$f_{\lambda,p}\ (x) = \frac{\lambda^p\ x^{p-1}\ e^{-\lambda x}}{\Gamma(p)}\ I_{(0,\infty)}\ (x), \tag{1.23}$$

donde $\Gamma(.)$ nota la función Gamma y $\lambda > 0$.

La media, la varianza y la función generadora de momentos de una variable aleatoria X que tiene distribución Gamma, con parámetros p y λ, son:

$$\begin{aligned} E(X) &= \frac{p}{\lambda} \\ Var(X) &= \frac{p}{\lambda^2} \\ M_X(t) &= (1 - t/\lambda)^{-p}, \quad t < \lambda \end{aligned}$$

La figura 1.4 muestra la gráfica de la función de densidad de la distribución Gamma para $\theta = (2, 3)$, en línea continua; para $\theta = (6, 2)$, en línea a trazos, y para $\theta = (4, 0.5)$, en línea punteada.

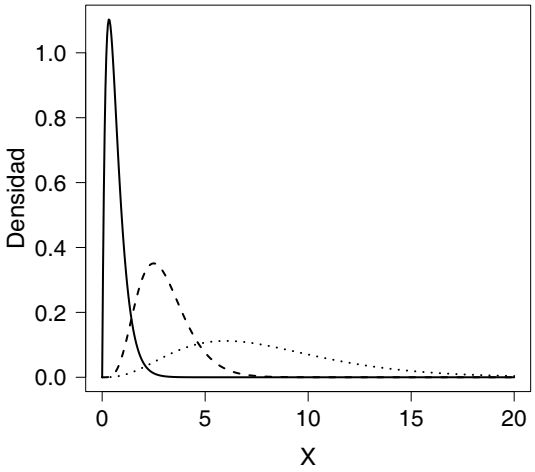

Figura 1.4. Función de densidad Gamma.

Proposición 1.1. Sean $X_i, i = 1, 2, \ldots, n$; n variables aleatorias independientes con distribución $\Gamma(p_i, \lambda)$, entonces, $Y = \Sigma_{i=1}^{n} X_i$ tiene distribución $\Gamma(\Sigma_{i=1}^{n} p_i, \lambda)$.

Demostración. Si $X_i \sim \Gamma(p_i, \lambda)$, $i = 1, 2, \ldots, n$; entonces, la función generadora de momentos es $M_{X_i}(t) = \left(1 - \frac{t}{\lambda}\right)^{-p_i}$, donde $t < \lambda$. Así, $M_Y(t) = (1 - t/\lambda)^{-\Sigma_{i=1}^{n} p_i}$. Esto concluye la demostración, ya que $M_Y(t)$ es la función generadora de momentos correspondiente a la distribución $\Gamma(\Sigma_{i=1}^{n} p_i, \lambda)$. $\square$

1.2.4 Distribución exponencial

Si $p = 1$, la expresión (1.23) toma la forma

$$f_\lambda(x) = \lambda e^{-\lambda x} I_{(0,\infty)}(x). \tag{1.24}$$

Este caso particular, de la función de densidad Gamma, se denomina función de densidad exponencial y se aplica en el estudio de tiempo de vida y en el análisis de los tiempos empleados en hacer una fila india de espera.

Si $X \sim \text{Exp}(\lambda), Y = 2\lambda X$ tiene distribución χ^2 con 2 grados de libertad. La función de densidad de Y está dada por

$$f_Y(y) = \lambda e^{-\lambda(y/2\lambda)} \left(\frac{1}{2\lambda}\right) = \frac{1}{2} e^{-y/2},$$

ya que

$$f_Y(y) = f_X(x) \left|\frac{dx}{dy}\right|.$$

Obsérvese que si $X_i \sim \text{Exp}(\lambda)$, $i = 1, 2, \ldots, n$, son n variables aleatorias independientes, a partir de la definición de una distribución Ji-cuadrado se se concluye que

$$2\lambda \, \Sigma_{i=1}^n \, X_i \sim \chi^2_{(2n)}.$$

Expresando (1.20) en la forma de la ecuación (1.23), se concluye que la distribución $\chi^2_{(n)}$ es igual a una distribución $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$. Obsérvese que $\chi^2_{(1)} = \Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$.

Ejercicios 1.4.

1. Demuestre que $\Gamma(\frac{1}{2}) = 2\Gamma(\frac{3}{2}) = \sqrt{\pi}$ y que $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$, $\alpha > 1$.

2. Si X es una variable aleatoria con distribución Gamma y función de densidad dada por la ecuación (1.23), determine $E(X)$, $V(X)$ y $M_X(t)$.

3. Si X es una variable aleatoria con distribución exponencial y $E(X) = \theta$, halle la distribución de la variable aleatoria $Y = X/\theta$.

4. Sea $X_i \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una población normal, con media conocida y varianza desconocida. Demuestre que

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{(n)}, \quad \text{donde} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_1^n (X_i - \theta)^2.$$

5. Determine la función generadora de momentos de la distribución $\Gamma(p, \lambda)$. ¿Cuál es la función generadora de momentos de la distribución exponencial?

6. Si $X_i \sim \Gamma(p_i, \lambda)$, $i = 1, 2, \ldots, n$, son n variables aleatorias independientes, demuestre que $\Sigma_{i=1}^n X_i$ tiene distribución $\Gamma(\Sigma_{i=1}^n p_i, \lambda)$.

7. Si $X_i, i = 1, 2, \ldots, n$, es una muestra aleatoria, de tamaño n, de una distribución exponencial con parámetro λ, definida por (**??**), demuestre que $\bar{X} \sim \Gamma(n, n\lambda)$. Halle la función generadora de momentos de $\bar{X}$.

8. Escriba la función de densidad gamma en términos de:

   a) Media y parámetro de forma.

   b) Media y varianza.

   Ayuda: Ver Cepeda-Cuervo (2001).

9. Escriba la función de densidad de la distribución ji-cuadrado, ecuación 1.20, en la forma de la función de densidad de la distribución gamma.

1.2.5 Distribución Weibull

Una variable aleatoria X tiene distribución Weibull, con parámetros α y β, lo que se nota $X \sim W(\alpha, \beta)$, si su función de densidad es

$$f(x) = \begin{cases} \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left(-\left(\frac{x}{\beta}\right)^{\alpha}\right) & \text{si } x \geq 0 \\ 0 & \text{si } x < 0, \end{cases} \tag{1.25}$$

donde $\alpha > 0$ es el parámetro de forma y $\beta > 0$ parámetro de escala.

La media y la varianza de X están dadas por

$$\begin{aligned} E(X) &= \beta\Gamma(1 + \alpha^{-1}) \\ V(X) &= \beta^2(\Gamma(1 + 2\alpha^{-1}) - \Gamma^2(1 + \alpha^{-1})), \end{aligned}$$

donde $\Gamma(.)$ es la función Gamma.

Si $\alpha \leq 1$, entonces, la función de distribución Weibull es monótona decreciente; si $\alpha = 1$, se reduce a la distribución exponencial con media β; y si $\alpha > 1$, es una función monótona creciente hasta la moda y luego monótona decreciente.

La distribución Weibull se utiliza, por ejemplo, en análisis de sobrevida y para modelar procesos estocásticos relacionados con el tiempo de fabricación de un producto.

Ejercicios 1.5.

1. A partir de la función (1.25), halle la función de distribución Weibull.
2. Si X tiene distribución $W(\alpha, \beta)$, demuestre que la función generadora de momentos de la variable aleatoria $Y = \log(X)$ es $E[e^{t\log(X)}] = \beta^t \Gamma(\frac{t}{\alpha} + 1)$.
3. Si X tiene distribución $W(\alpha, \beta)$, halle la distribución de la variable aleatoria $Y = (X/\beta)^\alpha$.
4. Sea X una variable aleatoria con función de distribución Weibull y con función de densidad

$$f(x, \lambda) = \lambda\alpha\, x^{\alpha-1}\, e^{-\lambda x^\alpha}\, I_{(0,\infty)}(x), \quad \alpha > 0,$$

demuestre que $y = x_i^\alpha$ tiene distribución $\exp(\lambda)$.

1.2.6 Distribución Beta

Una variable aleatoria Y tiene distribución Beta si su función de densidad está dada por

$$f(y|p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1}(1-y)^{q-1} I_{(0,1)}(y) \tag{1.26}$$

donde $p > 0$, $q > 0$ y $\Gamma(.)$ denota la función Gamma.

La media y la varianza de una variable aleatoria con esta distribución son

$$\mu = \frac{p}{p+q} \tag{1.27}$$

$$\sigma^2 = \frac{pq}{(p+q)^2(p+q+1)}. \tag{1.28}$$

La distribución Beta se emplea, por ejemplo, en el estudio de la humedad relativa del aire o de la proporción del ingreso familiar dedicada a educación. También se puede aplicar en el estudio de variables aleatorias que toman valores en un intervalo acotado. En este caso, si una variable aleatoria continua X toma valores en un intervalo (a, b), se puede asumir que la variable aleatoria $Y = (X - a)/(b - a)$ tiene distribución Beta.

Muchas veces algunas reparametrizaciones de la función de distribución Beta, dada en (1.26), son convenientes. La primera de ellas se obtiene haciendo $\phi = p + q$, $p = \mu\phi$, $q = \phi(1-\mu)$ y $\sigma^2 = \frac{\mu(1-\mu)}{\phi+1}$. En este caso, ϕ se interpreta como un parámetro de precisión en el sentido que, para valores fijos de μ, valores grandes de ϕ corresponden a valores pequeños de la varianza de Y. Esta reparametrización que fue presentada en Ferrari y Cribari-Neto (2004), apareció antes en la literatura, por ejemplo en Jorgensen (1997) o en Cepeda-Cuervo (2001). Con esta reparametrización, la función de densidad de la distribución Beta (1.26) puede escribirse como

$$f(y|\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1} I_{(0,1)}(y) \tag{1.29}$$

La distribución Beta, definida en (1.26), también puede reparametrizarse en función de la media y la varianza, haciendo

$$p = \frac{(1-\mu)\mu^2 - \mu\sigma^2}{\sigma^2} \tag{1.30}$$

$$q = \frac{(1-\mu)[\mu - \mu^2 - \sigma^2]}{\sigma^2}. \tag{1.31}$$

Aunque al escribir (1.26), en términos de μ y σ^2, es complejo, esta parametrización resulta bastante útil especialmente en regresión beta.

La figura 1.5 muestra la gráfica de la función de densidad Beta para $\theta = (2,4)$, en línea continua; para $\theta = (4,2)$, en línea a trozos, y para $\theta = (4,3)$, en línea punteada; siendo α la primera componente de θ y β la segunda.

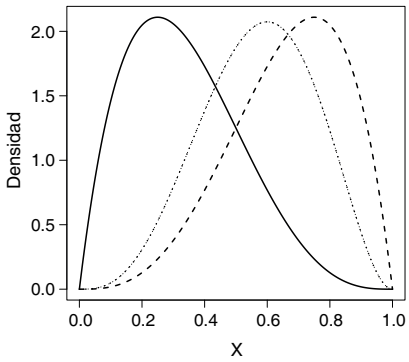

Figura 1.5. Función de densidad Beta.

1.2.7 Distribución t

Definición 1.2. Sea Z una variable aleatoria con distribución Normal estándar y X una variable aleatoria con distribución Ji-cuadrado con n grados de libertad (g.l.), independientes. La variable aleatoria

$$T = \frac{Z}{\sqrt{X/n}} \tag{1.32}$$

tiene distribución t con n grados de libertad, lo que se nota $T \sim t_{(n)}$. La media y la varianza de T están definidas respectivamente para $n > 1$ y $n > 2$. Si $n > 1$, $E(T) = 0$ y, si $n > 2$, $Var(T) = \frac{n}{n-2}$. La función generadora de momentos no está definida.

La función de densidad de la variable aleatoria T está dada por

$$p_n(t) = \frac{\Gamma[(n+1)/2]}{\Gamma(n/2)\sqrt{\pi n}} \left[1 + \frac{t^2}{n}\right]^{-(n+1)/2}, \quad t \in \mathbb{R} \tag{1.33}$$

Obsérvese que la función de densidad t es simétrica con respecto a $T = 0$. La figura 1.6 muestra la función de densidad t para 10 g.l. en línea continua; para 4 g.l. en línea a trazos y para 2 g.l. en línea punteada. En esta secuencia gráfica, se observa que a medida que disminuyen los grados de libertad, las colas de la función de densidad t se hacen más pesadas.

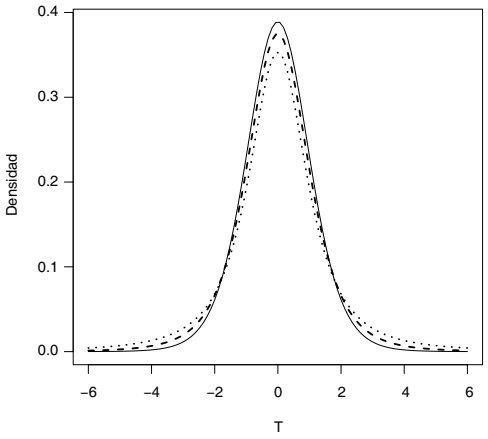

Figura 1.6. Distribución t con 10, 4 y 2 grados de libertad.

Ejemplo 1.2. Sea $X_i \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$; n variables aleatorias independientes con $\theta \in \mathbb{R}$ y $\sigma^2 > 0$ desconocidos. Si $X = (X_1, X_2, \ldots, X_n)$, entonces,

$$T(X) = \frac{\sqrt{n}\,(\bar{X} - \theta)}{S} \tag{1.34}$$

donde S es definida como en 1.22, tiene distribución $t_{(n-1)}$.

Dado que, por el teorema de Basu (Shao, 2003), S^2 y $\bar{X}$ son independientes, este resultado se demuestra haciendo

$$\begin{aligned} Z &= \frac{\sqrt{n}\,(\bar{X} - \theta)}{\sigma} \sim N(0,1) \\ X &= \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)} \end{aligned}$$

y aplicando la definición (1.2).

La variable aleatoria $T(X)$, definida en la ecuación (1.34), se utiliza para hacer inferencias estadísticas respecto a la media de una distribución Normal, con media y varianza desconocidas, cuando el tamaño de la muestra no es suficientemente grande para considerar que su distribución se aproxima a una distribución Normal.

Ejemplo 1.3. Suponga que $X_1 = (X_{11}, \ldots, X_{1n_1})$ y $X_2 = (X_{21}, \ldots, X_{2n_2})$ son dos muestras aleatorias independientes, de tamaño n_1 y n_2, respectivamente, provenientes de distribuciones normales $N(\theta_1, \sigma^2)$ y $N(\theta_2, \sigma^2)$, respectivamente; con $\theta_1 \in \mathbb{R}$, $\theta_2 \in \mathbb{R}$ y $\sigma^2 > 0$, desconocidos. Dado que

$$Z = \frac{[(\bar{X}_1 - \bar{X}_2) - (\theta_1 - \theta_2)]}{\sigma\sqrt{n_1^{-1} + n_2^{-1}}} \sim N(0,1)$$

y $\frac{(n_1+n_2-2)\,S^2}{\sigma^2} \sim \chi^2_{(n_1+n_2-2)}$ son independientes, aplicando la definición (1.2) se concluye que

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\theta_1 - \theta_2)}{S\sqrt{n_1^{-1} + n_2^{-1}}} \sim t_{(n_1+n_2-2)} \tag{1.35}$$

donde $S^2 = \frac{1}{n_1+n_2-2}\,\left[\Sigma_{i=1}^{n_1}\,(X_{1i} - \bar{X}_1)^2 \;+\; \Sigma_{i=1}^{n_2}\,(X_{2i} - \bar{X}_2)^2\right]$.

La variable aleatoria definida por la ecuación (1.35) se utiliza para hacer inferencias respecto a la diferencia de las medias de dos distribuciones normales, con medias y varianzas desconocidas, cuando el tamaño de las muestras no es suficientemente grande para aproximar su distribución a una distribución Normal.

Ejemplo 1.4. Sea $X_i, i = 1, 2, \ldots, n$, una muestra aleatoria, de tamaño n, de una población normal con media y varianza desconocidas. Si se sabe que la varianza muestral es $s^2 = 4$, ¿cuál es el menor valor de n tal que la media muestral $\bar{X}$ se encuentre a menos de 2 unidades de la media poblacional, con una probabilidad mayor de 0.95?

Para solucionar este interrogante, se debe encontrar n tal que $P(|\bar{X} - \theta| < 2) > 0.95$. Esto equivale a encontrar n tal que $P\left(\frac{\sqrt{n}|\bar{X}-\theta|}{S} < \sqrt{n}\right) > 0.95$. A partir de la tabla de la distribución t, se encuentra que esta desigualdad es válida para $n \geq 7$.

Ejemplo 1.5. Una empresa afirma que el diámetro de los tornillos que produce es de 12 mm. Si en una muestra aleatoria de tamaño 9, los diámetros de sus tornillos tienen una media de $\bar{x} = 9.84$ y una desviación estándar de $s = 2$, entonces, $P(\bar{X} < 9.84) = P\left(\frac{\sqrt{n}\,(\bar{X}-\theta)}{S} < \frac{3\,(9.84-12)}{2}\right) = 0.005$. Dado que esta probabilidad es bastante pequeña, es muy posible que la empresa esté produciendo tornillos de menor diámetro que el indicado.

Ejercicios 1.6.

1. Si T tiene distribución t con 14 grados de libertad, determine

$$P(-1.34 < T < 2.145).$$

2. Si T tiene distribución t con 9 grados de libertad y $\alpha = 0.05$, determine un número real t_0 tal que

   a) $P(T > t_0) = 1 - \alpha$.

   b) $P(T < t_0) = 1 - \dfrac{\alpha}{2}$.

   c) $P(-t_0 < T < t_0) = 1 - \alpha$.

3. Un medicamento debe contener 0.011 gramos de zinc. Si en una muestra aleatoria se obtienen los datos del ejercicio 4 de la subsección 1.2.2, ¿qué puede afirmarse sobre la cantidad de zinc que contiene el medicamento? Hallar $P_{0.011}(\bar{X} \geq \bar{x})$, donde $\bar{x}$ es la media muestral.

4. Se afirma que, en cierta ciudad, el consumo medio de agua por familia es 5 m^3 mensuales. Si se toma una muestra aleatoria de 10 observaciones y se obtiene $T(x) = 2.8$, ¿qué se puede afirmar sobre el consumo de agua, asumiendo que la muestra proviene de una distribución Normal? ¿Se está ahorrando agua? ¿Qué se podría afirmar si $T(x) = -2.8$? ¿Si $T(x) = 0$?

5. Demuestre que la distribución de la variable aleatoria $T = \frac{Z}{\sqrt{X/n}}$, definida en la ecuación (1.32), está dada por la ecuación (1.33).

   Sugerencia: Determine la función de densidad del vector (Z, X). Luego considere el vector aleatorio (Y_1, Y_2), donde $Y_1 = \frac{Z}{\sqrt{X/n}}$ y $Y_2 = X$ y finalmente recuerde que

   $$f_{Y_1,Y_2}(y_1, y_2) = f_{Z,X}(y_1, y_2)\,|J|,$$

   donde

   $$J = \begin{pmatrix} \frac{\partial z}{\partial y_1} & \frac{\partial z}{\partial y_2} \\ \frac{\partial x}{\partial y_1} & \frac{\partial x}{\partial y_2} \end{pmatrix}.$$

   Finalmente, halle la función de distribución marginal de Y_1.

6. Demuestre que $E(S^2) = \sigma^2$, donde S^2 se define como en el ejemplo 1.3.

1.2.8 Distribución *F*

Definición 1.3. Sean X_1 y X_2 dos variables aleatorias que tienen distribución de probabilidad Ji-cuadrado, con n_1 y n_2 grados de libertad, respectivamente. Si X_1 y X_2 son variables aleatorias independientes, entonces, la variable aleatoria

$$F = \frac{X_1/n_1}{X_2/n_2} \tag{1.36}$$

tiene una distribución F con n_1 grados de libertad en el numerador y n_2 grados de libertad en el denominador, lo que se nota $F_{n_2}^{n_1}$.

En esta sección α representa la probabilidad de que F tome valores mayores que un valor específico F_α. Por ejemplo, si $\alpha = 0.05$ y F tiene 10 grados de libertad en el numerador y 7 en el denominador, $f_\alpha = 3.64$, ya que $P\left(\mathfrak{F}_7^{10} > 3.64\right) = 0.05$, donde $\mathfrak{F}_7^{10}$ denota una variable aleatoria con distribución F, con 7 grados de libertad en el denominador y 10 en el numerador.

De la definición 1.3 se tiene que si una variable aleatoria tiene distribución F, con n_1 grados de libertad en el numerador y n_2 grados de libertad en el denominador, entonces,

$$\frac{1}{F} = \frac{X_2/n_2}{X_1/n_1}$$

tiene distribución F con n_2 grados de libertad en el numerador y n_1 grados de libertad en el denominador. Este hecho es de gran utilidad cuando se desea determinar, por ejemplo, un número real F_0 tal que $P\left(F_7^{10} < F_0\right) = 0.05$, ya que

$$P\left(\mathfrak{F}_7^{10} < F_0\right) = P\left(\frac{1}{\mathfrak{F}_7^{10}} > \frac{1}{F_0}\right) = P\left(\mathfrak{F}_{10}^{7} > \frac{1}{F_0}\right) = 0.05$$

Observando la tabla de la distribución F con 7 grados de libertad en el numerador y 10 grados de libertad en el denominador, se halla que $\frac{1}{f_\alpha} = 3.14$ siendo $F_0 = 0.318$.

Ejemplo 1.6. Sean $X_{1i} \sim N(\theta_1, \sigma_1^2)$, $i = 1, 2, \ldots, n_1$, y $X_{2i} \sim N(\theta_2, \sigma_2^2)$, $i = 1, 2, \ldots, n_2$; dos muestras aleatorias de tamaños n_1 y n_2, respectivamente, independientes y normalmente distribuidas. Si S_1^2 nota la varianza de la primera muestra y S_2^2 la de la segunda, las variables aleatorias

$$X_1^2 = \frac{(n_1 - 1)S_1^2}{\sigma_1^2} \quad \text{y} \quad X_2^2 = \frac{(n_2 - 1)S_2^2}{\sigma_2^2} \tag{1.37}$$

son independientes y tienen distribución Ji-cuadrado con $n_1 - 1$ y $n_2 - 1$ grados de libertad, respectivamente.

En consecuencia, a partir de la definición (1.36) se concluye que la variable aleatoria

$$F = \frac{\sigma_2^2\, S_1^2}{\sigma_1^2\, S_2^2} \tag{1.38}$$

tiene una distribución F con $n_1 - 1$ grados de libertad en el numerador y $n_2 - 1$ grados de libertad en el denominador.

Esta variable aleatoria se usa, por ejemplo, para hacer inferencias respecto a la igualdad o no de las varianzas de las dos distribuciones normales, con medias y varianzas desconocidas.

Ejemplo 1.7. Si se tienen dos muestras aleatorias del pH de un río, una tomada en verano (7.8, 7.3, 6.8, 7.0, 6.5, 7.0) y otra en invierno (7.5, 7.0, 6.5, 8.0, 7.9, 7.5, 8.0, 6.5), se pueden determinar dos números reales a y b tales que

$$P\left(\frac{a\ s_2^2}{s_1^2} < \mathfrak{F}_5^7 < \frac{b\ s_2^2}{s_1^2}\right) = P\left(1.98a < \mathfrak{F}_5^7 < 1.98b\right) = 0.9$$

donde s_1^2 y s_2^2 denota los valores de las varianzas muestrales. Un par de valores posibles de a y b se obtiene haciendo $P\left(F_5^7 < 1.98a\right) = P\left(F_5^7 > 1.98b\right) = 0.05$. Esto es equivalente a determinar a y b tales que $P\left(F_7^5 > 1/1.98a\right) = P\left(F_5^7 > 1.98b\right) = 0.05$. En consecuencia, $1/1.98a = 3.97$ y $1.98b = 4.88$, siendo $a = 0.127$ y $b = 2.46$.

Ejercicios 1.7.

1. Demuestre que la variable aleatoria F definida en (1.38), tiene distribución F con $n_1 - 1$ grados de libertad en el numerador y $n_2 - 1$ en el denominador.

   Sugerencia: Aplique la definición de distribución F y el teorema (1.2).

2. Sean $X_i \sim N(\theta, \sigma_X^2),\ i = 1, 2, \ldots, 5$, y $Y_j \sim N(\theta, \sigma_Y^2),\ j = 1, 2, \ldots, 10$, variables aleatorias independientes con media y varianza desconocidas. Si $\sigma_X^2 = 2\sigma_Y^2$, halle $P(S_X^2/S_Y^2 \geq 0.7435)$.

3. Sean $X_i \sim N(\theta_1, \sigma_X^2)$, $i = 1, 2, \ldots, 5$, y $Y_j \sim N(\theta_2, \sigma_Y^2)$, $j = 1, 2, \ldots, 10$, variables aleatorias independientes. Si $\sigma_X^2 = 2\sigma_Y^2$, halle

$$P\left(0.5 < S_X^2/S_Y^2 < 2.5\right).$$

4. Desarrolle los ejercicios 2 y 3, asumiendo que la media es conocida.

5. Si $X \sim t_{(p)}$, muestre que $Y = X^2 \sim F_p^1$. Aplique las definiciones de las distribuciones T, χ^2 y F.

1.2.9 Otras funciones de variables aleatorias

Sean $X_1, X_2, \ldots, X_n$; n variables aleatorias continuas e independientes con función de distribución F y función de densidad f. Se definen las variables

$$\begin{aligned} X_{(1)} &= \text{mín}\{X_1, X_2, \ldots, X_n\} \\ X_{(n)} &= \text{máx}\{X_1, X_2, \ldots, X_n\}, \end{aligned}$$

donde $X_{(1)}$ es el valor mínimo de $X_i,\ i = 1, 2, \ldots, n$, y $X_{(n)}$ es el valor máximo de $X_i,\ i = 1, 2, \ldots, n$.

Las funciones de densidad de $X_{(1)}$ y $X_{(n)}$ se obtienen a partir de las funciones de distribución de los X_i. Dado que F es la función de distribución de X_i, $i = 1, 2, \ldots, n$, la función de distribución de $X_{(n)}$ es

$$G_{X_{(n)}}(x) = P[X_{(n)} \leq x] = P[X_1 \leq x,\ X_2 \leq x, \ldots,\ X_n \leq x] = [F(x)]^n$$

y su función de densidad

$$g_{X_{(n)}}(x) = n[F(x)]^{n-1} f(x); \quad \text{donde} \quad f(x) = F^{'}(x).$$

Si X_i, $i = 1, 2, \ldots, n$, tiene distribución F, la función de distribución de $X_{(1)}$ es

$$\begin{aligned} G_{X_{(1)}}(x) &= P[X_{(1)} \leq x] = 1 - P[X_{(1)} \geq x] \\ &= 1 - P[X_1 \geq x, X_2 \geq x, \ldots, X_n \geq x] \\ &= 1 - [1 - F(x)]^n. \end{aligned}$$

Así, la función de densidad de $X_{(1)}$ es

$$g_{X_{(1)}}(x) = n[1 - F(x)]^{n-1} f(x), \quad \text{donde} \quad f(x) = F^{'}(x).$$

Ejemplo 1.8. Si $X_i \sim U(0, \theta)$, $i = 1, 2, 3, \ldots, n$; son n variables aleatorias independientes, entonces, la función de distribución $X_{(n)}$ es

$$G_{X_{(n)}}(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ (\frac{x}{\theta})^n & \text{si } 0 < x < \theta \\ 1 & \text{si } x \geq \theta \end{cases}$$

y su función de densidad $g_{X_{(n)}}(x) = \dfrac{n}{\theta^n} x^{n-1} I_{(0,\theta)}(x)$.

Ejemplo 1.9. Si $X_i \sim U(0,\theta)$, $i = 1,2,3,\ldots,n$, son n variables aleatorias independientes, la función de distribución $X_{(1)}$ es

$$G_{X_{(1)}}(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ 1-(1-\frac{x}{\theta})^n & \text{si } 0 < x < \theta \\ 1 & \text{si } x \geq \theta \end{cases}$$

y su función de densidad $g_{X_{(1)}}(x) = \dfrac{n}{\theta^n}(\theta - x)^{n-1} I_{(0,\theta)}(x)$.

Ejercicios 1.8.

1. Sean X_i, $i = 1,2,\ldots,n$, n variables aleatorias independientes uniformemente distribuidas en el intervalo $(0,\theta)$. Halle $E(X_{(1)})$ y $E(X_{(n)})$.

2. Sean X_i, $i = 1,2,\ldots,n$; n variables aleatorias independientes uniformemente distribuidas en el intervalo $(0,\theta)$.

   a) Halle las funciones de distribución de $Y_1 = \dfrac{X_{(1)}}{\theta}$ y $Y_2 = \frac{X_{(n)}}{\theta}$.

   b) Halle $E(Y_1)$ y $E(Y_2)$.

   ¿Qué se puede concluir?

3. Sean $X_i \sim F_\theta$, $i = 1,2$, dos variables aleatorias independientes. Si $x_1 < x_2$, calcule
$$P(X_{(1)} \leq x_1,\ X_{(2)} \leq x_2).$$

4. Sean $X_i \sim \text{Exp}\,(\theta)$, $i = 1,2,\ldots,n$; n variables aleatorias independientes, con $E(X_i) = 1/\theta$. Muestre que $X_{(1)}$ tiene distribución exponencial Exp $(n\theta)$. Determine $E(X_{(1)})$.

5. Determine la función de densidad de $X_{(1)}$ y $X_{(n)}$, si X_i, $i = 1,2,\ldots,n$; son variables aleatorias independientes con distribución:

   a) Gamma.

   b) Beta.

6. Sean X_i, $i = 1,2,\ldots,n$; n variables aleatorias independientes con función de densidad $f(x) = e^{-(x-\theta)} I_{(\theta,\infty)}(x)$.

a) Determine las funciones de densidad de $X_{(i)}$, $i = 1, n$.

b) Calcule $E(X_{(i)})$ y $V(X_{(i)})$ para $i = 1, n$.

7. Sean $X_i \sim \text{Exp}\,(\theta)$, $i = 1, 2, \ldots, n$; n variables aleatorias independientes, con $E(X_i) = \theta$. Demuestre que

$$Y = 2\theta \sum_{i=1}^{n} (X_i - X_{(1)}) \text{ tiene distribución } \chi^2_{2(n-1)}.$$

8. Sea X_i, $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una variable continua. Si $X_{(i)}$ es una variable aleatoria que, para cada conjunto de valores observados de la muestra, asume como valor observado el i-ésimo menor valor de la muestra, determine su función de densidad.

   Aplique el resultado obtenido para encontrar la función de distribución de X_i, dado que F_θ es la función de distribución:

   a) Uniforme.

   b) Exponencial.

9. Sea X_i, $i = 1, 2, \ldots, n$; una muestra aleatoria de una distribución $U(\theta_1, \theta_2)$. Halle:

   a) La función de densidad de $X_{(1)}$ y $X_{(n)}$.

   b) $E(X_{(1)})$ y $E(X_{(n)})$.

1.3 Teorema del límite central

Sea $\{X_i\}$ una sucesión de variables aleatorias con funciones de distribución F_i, $i = 1, 2, \ldots$, respectivamente. Si X es una variable aleatoria con función de distribución F, se dice que la sucesión $\{X_i\}$ converge en distribución a X, y se nota $X_i \xrightarrow{d} X$, si y solo si, $\{F_n(x)\}$ converge a $F(x)$ en todos los puntos x donde F es continua.

Ejemplo 1.10. Sea X una variable aleatoria con función de distribución F, exponencial de parámetro θ. Si X_i, $i = 1, 2, \ldots$, tiene distribución F_{X_i}, exponencial con parámetro $\theta_i = \theta + 1/i$, entonces $X_i \xrightarrow{d} X$.

Obsérvese que, para todo $x \in \mathbb{R}$,

$$F_{X_i}(x) = \int_0^x (\theta + \frac{1}{i}) e^{(\theta+\frac{1}{i})t} dt = 1 - e^{(\theta+\frac{1}{i})x}.$$

Y, así, $\lim_{i\to\infty} F_{X_i}(x) = 1 - e^{\theta x} = F(x)$.

Ejemplo 1.11. Si X tiene distribución uniforme $U(0,1)$ y X_i, $i = 1, 2, \ldots$, tiene función de densidad

$$f_{X_i}(x) = \begin{cases} 1 + \frac{1}{i} & \text{si } 0 < x < \frac{1}{2} \\ 1 - \frac{1}{i} & \text{si } \frac{1}{2} \leq x < 1, \end{cases}$$

entonces, $\{X_i\}$ converge en distribución a X. Obsérvese que:

$$F_{X_i}(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ x(1 + \frac{1}{i}) & \text{si } 0 < x < \frac{1}{2} \\ \frac{1}{2}(1 + \frac{1}{i}) + (x - \frac{1}{2})(1 - \frac{1}{i}) & \text{si } \frac{1}{2} \leq x < 1, \\ 1 & \text{si } x \geq 0 \end{cases}$$

y

$$\lim_{i\to\infty} F_{X_i}(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ x & \text{si } 0 < x < 1 \\ 1 & \text{si } x \geq 1. \end{cases}$$

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

Teorema 1.3 (del límite central)**.** Sean $X_1, X_2, \ldots, X_n$; n variables aleatorias independientes distribuidas idénticamente con $E(X_i) = \mu$ y $Var(X_i) = \sigma^2$, $\sigma^2 < \infty$. Entonces,

$$Z_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma}$$

converge en distribución a una variable Z normal estándar.

Este teorema indica, por ejemplo, que para valores grandes de n, la distribución Binomial se puede aproximar por una distribución Normal. En particular, si Y representa el número de éxitos en $n = 10.000$ experimentos de Bernoulli en los que la probabilidad de éxito es 0.7, entonces, la probabilidad que el número de éxitos esté entre 6910 y 7000 es

$$P\,(6910 < Y < 7000) \approx P\left(\frac{6910 - np}{\sqrt{npq}} < \frac{Y - np}{\sqrt{npq}} < \frac{7000 - np}{\sqrt{npq}}\right)$$
$$= P(-1.96 < Z < 0) = 0.475$$

En el ejercicio 1, de esta sección, se observa que la aproximación de la distribución Binomial a la Normal es poco "precisa" para valores pequeños de n. En estos casos es conveniente determinar el valor exacto de la probabilidad. Si se desea hacer el cálculo aproximado, es conveniente aplicar un factor de corrección de 0.5. Por ejemplo, si $Y \sim B(20, 0.3)$, $P(Y \leq 8) \approx P(W \leq 8.5)$ y $P(Y < 8) \approx P(W \leq 7.5)$, donde $W \sim N(6, 4.2)$.

Demostración 1.1. (del límite central). Supóngase que las n variables aleatorias X_i, $i = 1, 2, \ldots, n$, tienen una función de distribución para la cual existe la función generadora de momentos. Se demuestra que la función generadora de momentos de

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma}$$

converge a la función generadora de momentos de una variable Z con distribución Normal estándar. Para ello se demuestra que la función generadora de momentos de Z_n converge a la función generadora de momentos de Z.

Dado que las n variables aleatorias X_i, $i = 1, 2, \ldots, n$, son independientes, la función generadora de momentos de Z_n, Ψ_{Z_n}, es el producto de las funciones generadoras de momentos de las variables aleatorias $Y_i = (X_i - \mu)/\sigma$. Esto es,

$$\Psi_{Z_n}(t) = (\Psi_{Y_i}(t/\sqrt{n}\,))^n.$$

Dado que la expansión de Taylor de Ψ_{Y_i} alrededor de $t = 0$ es

$$\Psi_{Y_i}(\frac{t}{\sqrt{n}}) = \Psi_{Y_i}(0) + \frac{t}{\sqrt{n}}\Psi'_{Y_i}(0) + \frac{t^2}{2!n}\Psi''_{Y_i}(0) + \frac{t^3}{3!n^{3/2}}\Psi'''_{Y_i}(0) + ...$$

y que $E(Y_i) = 0$ y $Var(Y_i) = 1$, entonces,

$$\Psi_{Z_n}(\frac{t}{\sqrt{n}}) = \Big(1 + \frac{1}{n}\Big(\frac{t^2}{2} + \frac{t^3}{3!n^{1/2}}\Psi'''_{Y_i}(0) + ...\Big)\Big)^n.$$

Finalmente, dado que

$$\lim_{n\to\infty}\Big(\frac{t^2}{2} + \frac{t^3}{3!n^{1/2}}\Psi'''_{Y_i}(0) + ...\Big) = \frac{t^2}{2},$$

entonces, $\lim_{n\to\infty}\Psi_{Z_n} = \exp(\frac{1}{2}t^2)$, que es la función generadora de momentos de una distribución Normal estándar.

El anterior resultado se sigue, dado que si $\lim_{n\to\infty} a_n$ existe, entonces, $\lim_{n\to\infty}$ $(1 + \frac{a_n}{n})^n = \exp\,(\lim_{n\to\infty} a_n)$.

Para completar la demostración se debe probar que si $\lim_{n\to\infty}\Psi_{Z_n} = \Psi_Z$, entonces, Z_n converge a Z en distribución. Este enunciado, demostrado en textos más avanzados, como por ejemplo Dudewicz & Mishra (1988), completa esta demostración.

Ejercicios 1.9.

1. Una fábrica de zapatillas sabe que el 10 % de su producción es defectuosa. Supóngase que se desea seleccionar una muestra aleatoria de 20 unidades. Halle la probabilidad de que la muestra contenga al menos 17 zapatillas no defectuosas, aplicando la distribución:

   a) Normal, con y sin factor de corrección.

   b) Binomial.

2. Un encuestador considera que el 30 % de los colombianos son fumadores. Si se seleccionan aleatoriamente 70 colombianos, ¿cuál es la probabilidad de que la fracción de fumadores en la muestra difiera en más de 0.02 del porcentaje de fumadores considerado por el encuestador?

3. Suponga que $X_i \sim Ber(0.5),\ i = 1, \ldots, 20$, son veinte variables aleatorias independientes. Halle c tal que $P(|\bar{X} - 0.5| < c) > 0.5$.

   a) Utilizando el teorema del límite central.

b) Aplicando el teorema de Tchebysheff:

Teorema. Si X es una variable aleatoria con media finita θ y varianza σ^2, entonces,

$$P(|X-\theta| < k\sigma) > 1 - \frac{1}{k^2}, \tag{1.39}$$

donde k es una constante positiva.

4. La probabilidad de ocurrencia de muerte por ataque cardíaco es 0.7 para los fumadores. Si se seleccionan en forma aleatoria 100 historias clínicas de personas fumadoras, que han sufrido algún ataque cardíaco, ¿cuál es la probabilidad de que el número de muertes:
   a) exceda a 65?
   b) esté entre 65 y 85?

5. Sea p la proporción de tornillos defectuosos en la producción de una empresa. Determine el tamaño n de la muestra tal que, para todo p, la proporción de tornillos defectuosos en la muestra se aleje a lo más 0.1 de la proporción de unidades defectuosas en la producción, con una probabilidad del 95 %:
   a) Aplicando el teorema del límite central.
   b) Aplicando el teorema de Tchebysheff. Véase ecuación (1.39).
   c) Compare los resultados obtenidos en los literales a y b.

6. Sean $X_i \sim Ber(p_1)$, $i = 1, 2, \ldots, n_1$, y $Y_j \sim Ber(p_2)$, $j = 1, \ldots, n_2$; $n_1 + n_2$ variables aleatorias independientes. Determine
   a) $E(\bar{X} - \bar{Y})$.
   b) $Var(\bar{X} - \bar{Y})$.
   c) La distribución de $\bar{X} - \bar{Y}$.

Capítulo 2

Estimadores

En este capítulo se utilizan algunas funciones de variables aleatorias independientes para hacer estimaciones puntuales de ciertos parámetros de interés. Por ejemplo, se utiliza $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ para estimar la media de una población y $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ para estimar su varianza. A las funciones de variables aleatorias como $\bar{X}$ y S^2 se les denomina estimadores y a cada una de sus realizaciones, $\bar{x}$ y s^2, estimaciones.

Después de una corta introducción, que incluye algunos conceptos básicos, en las secciones 2.2 y 2.3 se presenta el método de los momentos y el método de máxima verosimilitud para la obtención de estimadores. Luego, en la sección 2.4, se presenta el concepto de error de estimación y, finalmente, en la sección 2.5, se comparan las varianzas de los estimadores y, en la sección 2.6, se presenta el concepto de consistencia.

2.1 Conceptos básicos

Un estimador es una regla que establece cómo calcular estimaciones de un parámetro, basada en una muestra aleatoria. Uno de los ejemplos más sencillos corresponde al estimador de la media de una población. Así, si $X_i, i = 1, 2, \ldots, n$; es una muestra aleatoria, de tamaño n, de una distribución Normal con media θ y varianza σ^2, entonces,

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{2.1}$$

es un estimador puntual de θ.

Si $\hat{\theta}$ un estimador de un parámetro θ y si $E(\hat{\theta}) = \theta$, entonces, se dice que $\hat{\theta}$ es un estimador insesgado de θ. Si $E(\hat{\theta}) \neq \theta$, se dice que $\hat{\theta}$ es un estimador sesgado de θ y a $B = E(\hat{\theta}) - \theta$ se le denomina sesgo de $\hat{\theta}$. Obsérvese que (2.1) es un estimador insesgado de la media de una población, ya que si X_i, $i = 1, 2, \ldots, n$, es una muestra aleatoria, de tamaño n, de una población con media θ,

$$\begin{aligned} E(\hat{\theta}) &= \frac{1}{n}\sum_{i=1}^{n} E(X_i) \\ &= \frac{1}{n}(n\theta) = \theta. \end{aligned}$$

De igual forma puede demostrarse que si $X_1, X_2, \ldots, X_{n_1}$ y $Y_1, Y_2, \ldots, Y_{n_2}$ son dos muestras aleatorias, de tamaño n_1 y n_2, respectivamente, independientes de distribuciones normales con medias θ_1 y θ_2, respectivamente, entonces, $\hat{\theta}$ es un estimador insesgado de $\theta_1 - \theta_2$, donde

$$\hat{\theta} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i - \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i.$$

Ejemplo 2.1. Si Y_i, $i = 1, 2, \ldots, n$; es una muestra aleatoria, de tamaño n, de una distribución exponencial con función de densidad $f_\theta(y) = \frac{1}{\theta}e^{-y/\theta} I_{\{y:\, y>0\}}(y)$, entonces, $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ es un estimador insesgado de θ.

Obsérvese que

$$E(Y_i) = \int_{-\infty}^{\infty} y f(y) dy = \frac{1}{\theta}\int_{0}^{\infty} y e^{-\frac{y}{\theta}} dy = \theta$$

y así

$$E(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n} E(Y_i) = \theta.$$

Ejemplo 2.2. Sea Y_i, $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una distribución Normal con media θ y varianza σ^2, desconocidas. Entonces,

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

es un estimador insesgado de σ^2.

Obsérvese que

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

dado que $E(Y_i^2) = \sigma^2 + \theta^2$ y $E(\bar{Y}^2) = \frac{\sigma^2}{n} + \theta^2$.

Definición 2.1. Un estimador, también llamado estadístico, es una función de variables aleatorias independientes e idénticamente distribuidas, que no contiene parámetros desconocidos.

Ejercicios 2.1.

1. Sea Y_i, $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una distribución Normal con media μ y varianza σ^2 desconocidas. Demuestre que $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$ es un estimador sesgado de σ^2.

2. Sean $Y_{11}, Y_{12}, \ldots, Y_{1n}$ y $Y_{21}, Y_{22}, \ldots, Y_{2m}$ dos muestras aleatorias, de tamaño n y m, respectivamente, independientes de dos poblaciones normales con varianza σ^2. Demuestre que

$$\begin{aligned} S^2 &= \frac{\sum_{i=1}^{n}(Y_{1i} - \bar{Y}_1)^2 + \sum_{j=1}^{m}(Y_{2j} - \bar{Y}_2)^2}{n+m-2} \\ &= \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2} \end{aligned}$$

   es un estimador insesgado de σ^2.

3. Sea X_i, $i = 1 \ldots, n$; una muestra aleatoria, de tamaño n, de una distribución $U(0, \theta)$. Demuestre que $\frac{n+1}{n}X_{(n)}$ es un estimador insesgado de θ.

4. Supóngase que el tiempo, antes de la falla de un equipo, tiene distribución exponencial con media $\mu = 1/\lambda$. Si se toman n equipos aleatoriamente y se observan los tiempos de falla $Y_1, Y_2, \ldots, Y_n$, demuestre que

$X = I_{[Y_1<x]}$, donde x es un número real positivo, es un estimador insesgado de la probabilidad del tiempo, antes de la falla, de los equipos, $P_\lambda[Y_1 < x] = 1 - e^{-\lambda x}$.

5. Sean Y_1, Y_2, Y_3 una muestra aleatoria, de tamaño tres, de una distribución uniforme en el intervalo $(\theta, \theta + 2)$. Demuestre que $\hat{\theta} = Y_{(3)}$ es un estimador sesgado de θ. Determine el sesgo y defina un estimador insesgado.

6. Sea $\hat{\theta}$ un estimador de un parámetro θ. Demuestre que $E(\hat{\theta} - \theta)^2 = Var(\hat{\theta}) + B^2$, donde B^2 es el cuadrado del sesgo, y calcule la media del cuadrado del error para $\hat{\theta}$ definido como en el ejercicio 5.

7. La media del cuadrado del error de un estimador puntual $\hat{\theta}$ se define como $E(\hat{\theta}-\theta)^2$. Suponga que Y_i, $i = 1, \ldots, n$; es una muestra aleatoria, de tamaño n, de una distribución Normal con media θ y varianza σ^2 desconocidas. Determine la media del cuadrado del error de

   a) $\hat{\theta} = \bar{Y}$ como estimador de la media.

   b) $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$ como estimador de la varianza.

8. Sea Y_i, $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una distribución normal con media θ y varianza σ^2. Demuestre que el estimador
$$\hat{\theta} = \frac{\sum_{i=k}^{n-(k+1)} Y_i}{n - 2k}, \quad 0 < k < \frac{n}{2}; \quad k \text{ entero},$$
es un estimador insesgado de θ. Determine $Var(\hat{\theta})$ y compárela con $Var(\bar{Y})$.

9. Sea Y_i, $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una distribución Poisson con parámetro θ. Calcule $\lim_{n\to\infty} E(e^{-\bar{Y}})$.

10. Si Y_i, $i = 1, 2, \ldots, n$; es una muestra aleatoria, de tamaño n, de una distribución exponencial con función de densidad $f_\theta(y) = \frac{1}{\theta}e^{-y/\theta}I_{\{y:\, y>0\}}(y)$, halle la función generadora de momentos de la variable aleatoria $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$.

2.2 Método de los momentos

En esta sección se introduce un método para obtener estimadores de un parámetro, llamado método de los momentos. Este método considera como estimadores de los momentos poblacionales, los momentos muestrales correspondientes, y si un parámetro se puede escribir como una función de los momentos poblacionales, su estimador se obtiene reemplazando en la función los momentos poblacionales por sus respectivos momentos muestrales.

En este método, por ejemplo, el estimador de la media de una población es la media muestral y el estimador de la varianza poblacional se obtiene reemplazando el primer y segundo momento poblacional por el primer y segundo momento muestral en $V(X) = E(X^2) - E^2(X)$.

Supóngase que X_i, $i = 1, 2, \ldots, n$; es una muestra aleatoria, de tamaño n, de una distribución F, que depende de θ. Se define el j-ésimo momento de F como

$$m_j(\theta) = E_\theta(X_1^j), j = 1, 2, \ldots$$

y sus estimadores muestrales por

$$\hat{m}_j = \frac{1}{n}\sum_{i=1}^{n} X_i^j, j = 1, 2, \ldots$$

Si una cantidad $q(\theta)$ puede expresarse como una función continua de los primeros k momentos poblacionales, $q(\theta) = h(m_1, m_2, \ldots, m_k)$, entonces, el método de los momentos indica que un estimador de $q(\theta)$ se obtiene al sustituir en h cada uno de los momentos poblacionales por sus respectivos momentos muestrales. Así, $T(X) = h(\hat{m}_1, \hat{m}_2, \ldots, \hat{m}_k)$ es un estimador de $q(\theta)$.

Ejemplo 2.3. Dado que $Var(X) = m_2(\theta) - m_1^2(\theta)$, $\hat{\sigma}^2 = \hat{m}_2(\theta) - \hat{m}_1^2(\theta)$ es un estimador de la varianza . En particular, dada una muestra aleatoria, de tamaño n, $X_i \sim Ber(\theta)$, $i = 1, 2, \ldots, n$, los estimadores de θ y σ^2, obtenidos por el método de los momentos, son $\hat{\theta} = \bar{X}$ y $\hat{\sigma}^2 = \bar{X}(1-\bar{X})$, respectivamente. Se puede demostrar que $\hat{\sigma}^2$ no es un estimador insesgado de la varianza.

Ejemplo 2.4. Sea $Y_i \sim G(\alpha, \beta)$, $i = 1, 2, \ldots, n$, una muestra aleatoria, de tamaño n, de una distribución Gamma, con función de densidad dada por

$$f(y) = \left[\frac{1}{\Gamma(\alpha)\beta^\alpha}\right] \quad y^{\alpha-1}e^{-y/\beta}I_{(0,\,\infty)}(y),$$

donde α y β son números reales positivos y $\Gamma(\alpha)$ denota la función Gamma, definida por

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}\, e^{-t}\, dt.$$

Dado que

$$\begin{aligned} m_1(\alpha,\beta) &= \alpha\beta \\ m_2(\alpha,\beta) &= \alpha\beta^2 + [m_1(\alpha,\beta)]^2 \\ &= m_1(\alpha,\beta)\beta + [m_1(\alpha,\beta)]^2, \end{aligned}$$

$$\hat{\beta} = \frac{\hat{m}_2(\alpha,\beta) - [\hat{m}_1(\alpha,\beta)]^2}{\hat{m}_1(\alpha,\beta)}. \tag{2.2}$$

Y, por otra parte,

$$\hat{\alpha} = \frac{(\hat{m}_1(\alpha,\beta))^2}{\hat{m}_2(\alpha,\beta) - [\hat{m}_1(\alpha,\beta)]^2}. \tag{2.3}$$

Finalmente, sustituyendo $\hat{m}_1$ y $\hat{m}_2$ en (2.2) y (2.3), en términos de X_i, $i = 1, 2, \ldots, n$; se obtiene que

$$\begin{aligned} \hat{\alpha} &= \bar{X}^2/\hat{\sigma}^2 \\ \hat{\beta} &= \hat{\sigma}^2/\bar{X}, \end{aligned}$$

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

Ejercicios 2.2.

1. Supóngase que $X_i \sim U(\theta_1, \theta_2)$, $i = 1, 2, \ldots, n$; es una muestra aleatoria, de tamaño n, de una distribución uniforme en el intervalo (θ_1, θ_2). Determine los estimadores de θ_1 y θ_2 por el método de los momentos.

2. La distribución Geométrica tiene función de densidad dada por

$$f(x) = (1-\theta)^{x-1}\theta, \quad x = 1, 2, \ldots$$

   Determine los estimadores de θ, $E(X)$ y $Var(X)$, aplicando el método de los momentos.

3. Sea X_i, $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una población con función de densidad dada por

$$f(x;\theta) = \frac{2(\theta - x)}{\theta^2} I_{(0,\theta)}(x). \tag{2.4}$$

Halle los estimadores de θ y $Var(X)$ por el método de los momentos.

4. La función de distribución Beta tiene función de densidad dada por

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} I_{(0,1)}(x),$$

donde $\alpha > 0$ y $\beta > 0$. Si $\alpha + \beta = 5$, halle los estimadores de α y β por el método de los momentos.

5. La función de densidad de la distribución log-normal es

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, x^{-1} e^{-(\log x - \theta)^2/2\sigma^2} I_{(0,\,\infty)}(x),$$

donde $\theta \in \mathbb{R}$ y $\sigma > 0$. Asumiendo que σ^2 es conocido, halle un estimador de θ por el método de los momentos.

6. La función de densidad de la distribución Weibull es

$$f(x) = \frac{\alpha}{\theta} x^{\alpha-1} e^{-x^\alpha/\theta} I_{(0,\,\infty)}(x),$$

donde $\theta > 0$ y $\alpha > 0$. Si α es conocido, halle un estimador de θ por el método de los momentos.

Sugerencia: Véase la definición de la función Gamma.

7. Si X_i, $i = 1, 2, \ldots, n$; son n variables aleatorias independientes con función de densidad

$$f(x) = \frac{1}{2\theta}\, e^{-|x-\mu|/\theta},$$

donde $-\infty < x < \infty$, $\mu \in \mathbb{R}$ y $\theta > 0$, halle estimadores insesgados de μ y θ por el método de los momentos.

2.3 Estimaciones de máxima verosimilitud

En esta sección se presenta un segundo método para hallar estimadores de un parámetro, llamado método de máxima verosimilitud. Este método permite encontrar estimadores, funciones de variables aleatorias que, calculadas en un conjunto de datos, permiten hallar el valor del parámetro que los hacen más probables.

Sea Y_i, $i = 1, 2, \ldots, n$; n variables aleatorias independientes con función de probabilidad o de densidad p_θ, donde θ toma valores en un espacio paramétrico Θ. Si y_i, $i = 1, 2, \ldots, n$; es un conjunto de n valores observados de Y_i, $i = 1, 2, \ldots, n$; una pregunta natural en estadística es ¿cuál es la función de probabilidad o la función de densidad de la cual provienen estos datos? Esto, en el caso en que el espacio paramétrico sea un conjunto contable, equivale a preguntar, ¿cuál es el valor de θ que hace más probable la observación de estos datos? y, en el caso en que el espacio paramétrico sea un intervalo del conjunto de los números reales, equivale a preguntar, ¿cuál es el valor de θ que maximiza la función $p_\theta(y_1, y_2, \ldots, y_n) = \prod_{i=1}^{n} p_\theta(y_i)$? Este valor de θ se obtiene maximizando la función $L(\theta) = p_\theta(y_1, y_2, \ldots, y_n)$, llamada función de verosimilitud, sobre el conjunto de valores posibles de θ, notado Θ.

Si la función de verosimilitud $L(\theta)$ es diferenciable en Θ o en un subconjunto abierto de Θ, el valor $\hat{\theta}$ de θ, donde la función de verosimilitud alcanza su valor máximo, se halla aplicando conceptos básicos de cálculo diferencial. Si Y_i, $i = 1, 2, \ldots, n$; son n variables aleatorias independientes con función de probabilidad p_θ, donde θ toma valores en un espacio paramétrico finito $\Theta = (\theta_1, \ldots, \theta_J)$, la estimación de máxima verosimilitud de θ se obtiene calculando y comparando $p_{\theta_j}(y_1, y_2, \ldots, y_n)$, para $j = 1, 2, \ldots, J$.

Ejemplo 2.5. Sea Y_i, $i = 1, 2, \ldots, n$; n variables aleatorias independientes Bernoulli, con $P(Y_i = 1) = \theta$ y $P(Y_i = 0) = 1 - \theta$ desconocidos. Si y_i, $i = 1, 2, \ldots, n$, es un conjunto de n valores observados de Y_i, $i = 1, 2, \ldots, n$; entonces, la función de verosimilitud es una función de θ definida sobre el espacio paramétrico $\Theta = (0, 1)$ por la ecuación (2.5), donde

$$P_\theta(Y_i = y_i) = \theta^{y_i} (1 - \theta)^{1 - y_i} I_{\{0,1\}}(y_i), \quad i = 1, 2, \ldots, n.$$

$$L(\theta) = \prod_{i=1}^{n} P_\theta(Y_i = y_i) \tag{2.5}$$

Dado que hallar el valor de θ que maximiza $L(\theta)$ es equivalente a encontrar el valor de θ que maximiza

$$\ln L(\theta) = \sum_{i=1}^{n} \left[y_i \ln \theta + (1 - y_i) \ln (1 - \theta) \right], \tag{2.6}$$

entonces, derivando (2.6) con respecto a θ se encuentra que, el valor de θ que maximiza $L(\theta)$, es $\hat{\theta} = \bar{Y}$. Calculando la segunda derivada de $\ln L$, con respecto a θ, se encuentra que $\hat{\theta}$ es el punto donde L alcanza su máximo valor.

Ejemplo 2.6. Sean Y_i, $i = 1, 2, \ldots, n$; n variables aleatorias independientes normalmente distribuidas, con media θ desconocida, y varianza σ^2 conocida. Si y_i, $i = 1, 2, \ldots, n$; es un conjunto de n valores observados de Y_i, $i = 1, 2, \ldots, n$, entonces, la función de verosimilitud es una función de θ definida sobre el espacio paramétrico $\Theta = \mathbb{R}$ por la ecuación (2.7).

$$\begin{aligned} L(\theta) &= f_\theta(y_1, y_2, \ldots, y_n) = \prod_{i=1}^{n} f_\theta(y_i) \\ &= (2\pi\sigma^2)^{-n/2} \exp\Big\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \theta)^2 \Big\}. \end{aligned} \tag{2.7}$$

Así, el logaritmo de la función de verosimilitud es

$$\ell(\theta) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \theta)^2 \tag{2.8}$$

y la primera y segunda derivada, del logaritmo de la función de verosimilitud, es $\ell^{'}(\theta) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \theta)$ y $\ell^{''}(\theta) = -\frac{n}{\sigma^2}$, respectivamente. Finalmente, igualando $\ell^{'}(\theta)$ a 0 y resolviendo la ecuación resultante se encuentra que

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} Y_i \tag{2.9}$$

es un estimador de máxima verosimilitud de θ, dado que $\ell^{''}(\theta) < 0$. El número real $\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} y_i$ es una estimación de máxima verosimilitud de θ.

Ejemplo 2.7. Sea Y_i, $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una distribución Uniforme en el intervalo $(0, 2\theta]$. La función de máxima verosimilitud es

$$\begin{aligned} L(\theta) = f_\theta(y_1, y_2, \ldots, y_n) &= \prod_{i=1}^{n} f_\theta(y_i) \\ &= \frac{1}{(2\theta)^n} I_{(0,\, 2\theta]}(Y_{(n)}) \\ &= \frac{1}{(2\theta)^n} I_{[\frac{1}{2}Y_{(n)},\, \infty)}(\theta). \end{aligned}$$

Como L es una función decreciente de θ, L se hace mayor a medida que θ se hace menor. Dado que $Y_i \leq 2\theta$, para todo i, el menor valor posible de 2θ es igual al mayor de los valores observados en la muestra. En consecuencia, $\hat{\theta} = \frac{1}{2}Y_{(n)}$ es un estimador de máxima verosimilitud de θ.

Ejemplo 2.8. Sea $X_i \sim N(\mu, \sigma^2)$, $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una distribución Normal con media y varianza desconocidas. El logaritmo de la función de verosimilitud es

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

y, en consecuencia,

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

Igualando a cero estas derivadas parciales, se halla como punto crítico $\hat{\theta} = (\hat{\mu}, \hat{\sigma^2})$, donde $\hat{\mu} = \bar{x}$ y $\hat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\mu})^2$.

Finalmente, se encuentra que $\hat{\theta}$ corresponde a un punto de máximo absoluto de L, ya que

$$\begin{aligned} &\frac{\partial^2 \ell}{\partial \mu^2}(\hat{\mu}, \hat{\sigma}^2)\, \frac{\partial^2 \ell}{\partial \sigma^{2\,2}}(\hat{\mu}, \hat{\sigma}^2) - \left[\frac{\partial^2 \ell}{\partial \mu\, \partial \sigma^2}(\hat{\mu}, \hat{\sigma}^2) \right]^2 > 0 \quad \text{y} \\ &\frac{\partial^2 \ell}{\partial \mu^2}(\hat{\mu}, \hat{\sigma}^2) = -\frac{n}{\sigma^2} < 0. \end{aligned}$$

Por tanto, $\hat{\theta}$ es un estimador de máxima verosimilitud de $\theta = (\mu, \sigma^2)$.

Ejemplo 2.9. Sea $Y_i \sim N(\mu_i, \sigma^2)$, $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, con media $\mu_i = \beta_0 + \beta_1 x_i$, donde x_i, $i = 1, 2, \ldots, n$, y σ^2 son números reales conocidos. Si β_0 y β_1 se asumen como parámetros, el logaritmo de la función de máxima verosimilitud está dado por

$$\begin{aligned}\ell(\beta_0, \beta_1) &= -\frac{n}{2} log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu_i)^2. \\ &= -\frac{n}{2} log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2.\end{aligned}$$

Ahora, derivando parcialmente a ℓ con respecto a β_0 y β_1, se obtiene

$$\frac{\partial \ell}{\partial \beta_0} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) \tag{2.10}$$

$$\frac{\partial \ell}{\partial \beta_1} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)x_i \tag{2.11}$$

e igualando (2.10) y (2.11) a cero y resolviendo el sistema de ecuaciones, se encuentra que

$$\begin{aligned}\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 &= \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}.\end{aligned}$$

Finalmente, como

$$\begin{aligned}&\frac{\partial^2 \ell}{\partial \beta_0^2}(\hat{\beta})\frac{\partial^2 \ell}{\partial \beta_1}(\hat{\beta}) - \left[\frac{\partial^2 \ell}{\partial \beta_0\, \partial \sigma^2}(\hat{\beta})\right]^2 > 0 \quad \text{y} \\ &\frac{\partial^2 \ell}{\partial \beta_0^2}(\hat{\beta}) < 0,\end{aligned}$$

se encuentra que $\hat{\beta}_0$ y $\hat{\beta}_1$ son los valores de β_0 y β_1 donde la función de verosimilitud alcanza su valor máximo. $\hat{\beta}_0$ y $\hat{\beta}_1$ son las estimaciones de máxima verosimilitud de β_0 y β_1.

Ejemplo 2.10. Suponga que se desea estimar la proporción de tornillos defectuosos en una caja que contiene 5 tornillos, a partir de una muestra aleatoria con rempazo, de tamaño 3. Dado que la caja contiene 5 tornillos, si θ indica la proporción de tornillos defectuosos, el conjunto de valores posibles de θ es $\Theta = \{0, 1/5, 2/5, 3/5, 4/5, 1\}$. Si la muestra aleatoria obtenida es $(x_1, x_2, x_3) = (1, 1, 0)$, donde $x_i = 1$ si el tornillo examinado es defectuoso y $x_i = 0$ si no lo es, 0 y 1 no son valores posibles de θ pues en la muestra se encontraron tornillos defectuosos y no defectuosos. Así, si $L(.)$ denota la función de verosimilitud, $L(0) = L(1) = 0$.

Por otra parte, como se tienen tres observaciones independientes de una distribución Bernulli con parámetro θ, la función de verosimilitud está dada por:

$$L(\theta) = \theta^2(1-\theta), \text{ para } \theta = \frac{1}{5}, \frac{2}{5}, \frac{3}{5} \text{ y } \frac{4}{5} \tag{2.12}$$

Así, evaluando $L(.)$ en cada uno de los valores posibles de θ se encuentra el valor máximo de la función de verosimilitud es $L(\frac{3}{5}) = \frac{18}{53}$, siendo $\theta = \frac{3}{5}$ la estimación de máxima verosimilitud de θ.

Ejemplo 2.11. Si X es una variable aleatoria con distribución Hipergeométrica $X \sim H(5\theta, 5, 3)$; θ toma valores en el conjunto $\Theta = \{0, 1/5, 2/5, 3/5, 4/5, 1\}$ y X toma valores en el conjunto $A = \{0, 1, 2, 3\}$. Si L nota la función de verosimilitud y $x = 1$ es el valor observado de X, entonces, θ es diferente de cero y, en consecuencia, $L(0) = 0$.

Dado que el tamaño de la muestra es $n = 3$ y el valor observado de X es $x = 1$, existen dos elementos no defectuosos en la muestra y, en consecuencia, la proporción de elementos defectuosos en la población no puede ser mayor que $3/5$, siendo $L(4/5) = L(1) = 0$, donde $L(.)$ denota la función de verosimilitud.

Además, $L(4/5) = L(1) = 0$, ya que hay dos elementos no defectuosos en la muestra y así la proporción de defectuosos en la población no puede ser mayor que $3/5$. Finalmente, dado que para $\theta = 1/5$, $2/5$, $3/5$, la función de verosimilitud está dada por

$$L(\theta) = \frac{\binom{5\theta}{1}\binom{5-5\theta}{3-1}}{\binom{5}{3}},$$

$L(1/5) = 3/5$, $L(2/5) = 3/5$ y $L(3/5) = 3/10$. Así, la estimación de máxima verosimilitud de θ es $\hat{\theta} = 1/5$ o $\hat{\theta} = 2/5$.

Ejemplo 2.12. Si $X \sim H(m, N, n)$, su función de probabilidad está dada por

$$P(x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}} I_A(x), \tag{2.13}$$

donde $A = \{0, 1, 2, \ldots, n\}$, m es el número de elementos defectuosos en la población (parámetro de la distribución) y $\text{máx}(n - (N - m), 0) \leq x \leq \text{mín}(m, n)$.

En esta función de probabilidad, si se asume que N y n son conocidos y que m es el parámetro, la estimación de máxima verosimilitud de m se obtiene comparando las funciones de verosimilitud $L(m + 1, x)$ y $L(m, x)$, para lo cual se desarrolla el siguiente cociente:

$$\begin{aligned} \frac{L(m+1,x)}{L(m,x)} &= \frac{\binom{m+1}{x}\binom{N-m-1}{n-x}}{\binom{m}{x}\binom{N-m}{n-x}} \\ &= \frac{(N-m-n+x)(m+1)}{(N-m)(m+1-x)}. \end{aligned}$$

En consecuencia, $L(m + 1, x) > L(m, x)$, si y solo si $m < \frac{x}{n}(N + 1) - 1$. Así, para determinar los puntos donde L alcanza su valor máximo, se debe considerar los siguientes casos:

1. Si $\frac{x}{n}(N+1)$ no es un número entero, entonces, la función de verosimilitud L alcanza su valor máximo en el menor número entero mayor que $\frac{x}{n}(N+1)-1$, es decir, en el mayor número entero menor que $\frac{x}{n}(N+1)$, lo cual se nota $\hat{m} = [\frac{x}{n}(N + 1)]$.

2. Si $\frac{x}{n}(N+1)$ es un número entero, entonces, $L(m+1, x) > L(m, x)$ para valores de m menores que $\frac{x}{n}(N + 1) - 1$ y $L(m + 1, x) < L(m, x)$ para valores de m mayores que $\frac{x}{n}(N + 1) - 1$. En consecuencia, el estimador de máxima verosimilitud de m es $\hat{m}_1 = \frac{x}{n}(N + 1) - 1$ o $\hat{m}_2 = \frac{x}{n}(N + 1)$.

Ejercicios 2.3.

1. Sea $M_1 = \{9.6, 2.2, 4.6\}$, $M_2 = \{-6.5, 0.9, 8.3\}$ y $M_3 = \{0.5, 8.9, 8.9\}$ tres muestras aleatorias provenientes de distribuciones normales diferentes. De cuál de las siguientes distribuciones considera usted que proviene cada una de las muestras: $N(5, 9)$, $N(1, 20)$ y $N(6, 16)$.

2. Sea X_i, $i = 1, 2, \ldots, n$, una muestra aleatoria, de tamaño n, de una distribución con parámetro θ. Halle un estimador de máxima verosimilitud para θ dado que

   a) $X_i \sim \text{Exp}(\theta)$.

   b) $X_i \sim P(\theta)$.

   c) $X_i \sim N(\mu, \theta)$, μ conocido.

   d) $X_i \sim N(\theta, \sigma^2)$, σ^2 conocido.

3. En el ejemplo 2.11, si el valor observado de X es $x = 1$, determine la estimación de máxima verosimilitud de θ. Desarrolle el ejemplo 2.11 si $X \sim H(7\theta, 7, 4)$.

4. Sea Y_i, $i = 1, 2, \ldots, n$, una muestra aleatoria, de tamaño n, de una distribución Uniforme en el intervalo $(0, \theta)$. Halle un estimador de máxima verosimilitud para θ.

5. Sea Y_i, $i = 1, 2, \ldots, n$, una muestra aleatoria, de tamaño n, de una distribución Uniforme en el intervalo $(\theta, \theta + 2)$. Halle un estimador de máxima verosimilitud para θ.

6. Suponga que X tiene función de probabilidad

$$f_\theta(x) = (e^\theta - 1)^{-1} \frac{\theta^x}{x!}, \quad x = 1, 2 \ldots; \ \theta > 0.$$

   Entonces, calcule:

   a) $E(X)$.

   b) $E\left[\left(1 - \dfrac{1}{e^X}\right) X\right]$.

   c) ¿Existe un estimador de máxima verosimilitud de θ?

7. Sea Y_i, $i = 1, 2, \ldots, n$, una muestra aleatoria, de tamaño n, de una población cuya función de densidad está dada por

$$f_\theta(y) = \frac{2}{\sqrt{\pi}} \exp\left[-(y - \theta)^2\right] I_{(-\infty, \theta]}(y).$$

   Demuestre que f_θ define una función de densidad y halle el estimador de máxima verosimilitud de θ.

8. Sea Y_i, $i = 1, 2, \ldots, n$, una muestra aleatoria, de tamaño n, de una población normal con media θ desconocida, y varianza σ^2 conocida. Encuentre un estimador de máxima verosimilitud de θ bajo la condición $\theta \geq 0$.

9. Desarrolle el ejercicio 7, de la sección 2.2, por el método de máxima verosimilitud.

10. Sean $X_i, i = 1, 2, \ldots, n$; n variables aleatorias independientes con función de densidad Rayleigh dada por

$$f(x_i, \theta) = (x_i/\theta^2) \exp\{-x_i^2/2\theta^2\}, x_i > 0, \theta > 0:$$

    a) Halle $E(X_i)$, $E(X_i^2)$ y $V(X_i)$.

    b) Encuentre el estimador de máxima verosimilitud de θ.

11. Sea $\hat{\theta}$ un estimador de máxima verosimilitud de $\theta \in \mathbb{R}$. Si g es una función inyectiva de $\mathbb{R}$ en $\mathbb{R}$; demuestre que $g(\hat{\theta})$ es un estimador de máxima verosimilitud de $g(\theta)$.

12. Supóngase que x_i, $i = 1, 2, \ldots, n$; son n valores posibles de n variables aleatorias independientes $X_i \sim Ber(\theta_1)$, $i = 1, 2, \ldots, n$. Si y_i, $i = 1, 2, \ldots, n$; son n valores posibles de las n variables aleatorias independientes $Y_i \sim Ber(\theta_2)$, $i = 1, 2, \ldots, n$, halle una estimación de máxima verosimilitud para $\theta_1 - \theta_2$. Asuma que X_i y Y_j son independientes, para todo $i, j = 1, 2, \ldots, n$.

13. Si X es una variable aleatoria con distribución Hipergeométrica $X \sim H(\theta, 5, 3)$, y $x = 2$ es un valor observado de X, halle la estimación de máxima verosimilitud de θ.

2.4 Error de estimación

En una muestra aleatoria de $n = 250$ estudiantes universitarios se encontró que 50 practican deporte con frecuencia. Si se utiliza esta proporción $\frac{50}{250} = 0.2$ como una estimación de la proporción de estudiantes universitarios que hacen deporte, entonces, ¿cuál es el error de estimación?

Dado que se desconoce la proporción θ de estudiantes que practican deporte, no se puede determinar el error de estimación $\epsilon = |\hat{\theta} - \theta|$ para una estimación particular $\hat{\theta}$ de θ. Sin embargo, se puede preguntar, por ejemplo, ¿cuál es la probabilidad que $\epsilon = |\hat{\theta} - \theta|$ sea menor que 0.01? y ¿cuál es el número real b tal que $P(|\hat{\theta} - \theta| < b) = 0.95$? Dado que

$$\begin{aligned} P(|\hat{\theta} - \theta| < 0.01) &\approx P\left(|Z| < \frac{0.01}{\sqrt{(0.2)(0.8)/250}}\right) \\ &= 1 - 2P(Z > 0.395) \\ &\approx 1 - 2(0.345) = 0.31, \end{aligned}$$

se tiene una confianza del 31 % que la proporción $\hat{\theta} = 0.2$ difiere de θ en una cantidad que no excede 0.01.

Por otra parte,

$$P(|\hat{\theta} - \theta| < b) \approx P\left(|Z| < \frac{b}{\sqrt{(0.2)(0.8)/250}}\right) = 0.95.$$

y, en consecuencia, $b = 1.96\sqrt{(0.2)(0.8)/250} = 0.049$. Esto significa que existe una probabilidad del 95 % de que el error de estimación no sea mayor que 0.049, cuando la estimación de la proporción de estudiantes universitarios que practican deporte es de 20 %.

Ejercicios 2.4.

1. Sean $Y_i \sim Ber(\theta)$, $i = 1, 2, \ldots, n$; n variables aleatorias independientes. Demuestre que $T_n(Y) = \bar{Y}(1 - \bar{Y})$ no es un estimador insesgado de $\theta(1 - \theta)$.

2. Se desea determinar la proporción de colombianos que no consumen cigarrillo. ¿Cuál es el tamaño de la muestra requerida para garantizar una precisión de 0.01, con un nivel de confianza del 95 %?

3. Sean Y_i, $i = 1, \ldots, n$; n variables aleatorias dicotómicas independientes, con probabilidad de éxito $\theta = 0.7$. Para $n = 40, 60, 100, 150$ y 200, halle $P\left(\left|\frac{Y}{n} - \theta\right| < 0.1\right)$, donde $Y = \sum_{i=1}^{n} Y_i$. ¿Qué se puede concluir?

4. Suponga que Y_i, $i = 1, 2, \ldots, n$; son n variables dicotómicas independientes, con probabilidad de éxito $P(Y_i = 1) = \theta$. Si ϵ es un número real positivo y $Y = \sum_{i=1}^{n} Y_i$; determine $\lim_{n\to\infty} P\left(\left|\frac{Y}{n} - \theta\right| < \epsilon\right)$.

2.5 Eficiencia

En esta sección se comparan las varianzas de estimadores insesgados de un parámetro, con el objetivo de encontrar, en un conjunto de estimadores insesgados, el estimador de menor varianza.

Sea Y_i, $i = 1, 2, \ldots, n$, una muestra aleatoria, de tamaño n, de una distribución normal con media θ y varianza σ^2. Entonces,

$$\hat{\theta}_1 = \bar{Y} \quad \text{y} \quad \hat{\theta}_2 = \frac{\sum_{i=k}^{n-(k+1)} Y_i}{n-2k}, \quad 0 < k < \frac{n}{2},$$

son estimadores insesgados de θ, con $Var(\hat{\theta}_1) = \frac{\sigma^2}{n}$ y $Var(\hat{\theta}_2) = \frac{\sigma^2}{n-2k}$.

Ahora bien, dado que $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$, se dice que $\hat{\theta}_1$ es más eficiente que $\hat{\theta}_2$.

Definición 2.2. Si $\hat{\theta}_1$ y $\hat{\theta}_2$ son estimadores insesgados de θ, se dice que $\hat{\theta}_1$ es más eficiente que $\hat{\theta}_2$ si $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$. Asimismo, se dice que $\hat{\theta}_1$ y $\hat{\theta}_2$ son igualmente eficientes, si $Var(\hat{\theta}_1) = Var(\hat{\theta}_2)$.

De la definición (2.2) se deduce que lo ideal es tener estimadores insesgados con la menor varianza posible.

Dado un estimador insesgado, se puede determinar si su varianza tiene el menor valor posible a partir del siguiente teorema.

Teorema 2.1. *Sea X_i, $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una población con función de distribución P_θ asociada a un parámetro θ. Si $\hat{\theta}$ es un estimador insesgado de θ y*

$$Var(\hat{\theta}) = \frac{1}{n\, E\left[\left(\frac{\partial \ln p_\theta(x)}{\partial \theta}\right)^2\right]}, \tag{2.14}$$

donde $p_\theta(x)$ es la función de densidad, entonces, $\hat{\theta}$ es un estimador insesgado de θ de mínima varianza.

Ejemplo 2.13. Sea $X_i, i = 1, 2, \ldots, n$, una muestra aleatoria, de tamaño n, de una distribución exponencial con parámetro $\theta = E(X_i)$. Si $\hat{\theta} = \bar{X}$, entonces, $\hat{\theta}$ es un estimador insesgado de θ y $Var(\hat{\theta}) = \frac{\theta^2}{n}$.

Para demostrar que $\hat{\theta}$ es un estimador insesgado de mínima varianza de θ, se debe demostrar que cumple con la ecuación (2.14) para $f_{\theta(x)} = \frac{1}{\theta}\exp(-\frac{1}{\theta}x)$ $I_{(0,\infty)}(x)$.

Este resultado se sigue dado que

$$E\left[\left(\frac{\partial \ln f_\theta(x)}{\partial\theta}\right)^2\right] = \frac{1}{\theta^4}E\left[(X-\theta)^2\right] \tag{2.15}$$

$$= \frac{1}{\theta^2}. \tag{2.16}$$

Ejemplo 2.14. Sea $X_i \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$, una muestra aleatoria, de tamaño n, de una distribución Normal con media conocida y varianza desconocida. Si

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\theta)^2, \tag{2.17}$$

entonces, $\hat{\sigma}^2$ es un estimador insesgado de mínima varianza de σ^2.

Obsérvese que

1. $E(\hat{\sigma}^2) = \frac{\sigma^2}{n}E\left[\sum_{i=1}^{n}(\frac{X_i-\theta}{\sigma})^2\right] = \sigma^2$.
2. $Var(\hat{\sigma}^2) = \frac{\sigma^4}{n^2}Var\left[\sum_{i=1}^{n}(\frac{X_i-\theta}{\sigma})^2\right] = \frac{2\sigma^4}{n}$.
3.
$$\begin{aligned} E\left[\left(\frac{\partial \ln f_\theta(x)}{\partial\sigma^2}\right)^2\right] &= \frac{1}{4\sigma^8}E[((x-\theta)^2-\sigma^2)^2] \\ &= \frac{1}{4\sigma^8}Var[(x-\theta)^2] \\ &= \frac{1}{4\sigma^4}Var\left[\left(\frac{x-\theta}{\sigma}\right)^2\right] \\ &= \frac{1}{2\sigma^4}. \end{aligned}$$

De los numerales 2 y 3 se sigue que

$$Var(\hat{\sigma}^2) = \frac{1}{n[\frac{1}{2\sigma^4}]} = \frac{2\sigma^4}{n},$$

lo cual, conjuntamente con el numeral 1, demuestran que $\hat{\sigma}^2$ es un estimador insesgado de mínima varianza de σ^2.

Ejercicios 2.5.

1. Supóngase que Y_i, $i = 1, \ldots, n$; es una muestra aleatoria, de tamaño n, de una distribución exponencial con función de densidad

$$f(y) = \begin{cases} \lambda e^{-\lambda y} & \text{si} \quad y > 0 \\ 0 & \text{si} \quad y \leq 0. \end{cases}$$

   Demuestre que $\hat{\theta}_1 = \bar{Y}$ y $\hat{\theta}_2 = \frac{2Y_1+4Y_2+\ldots+2nY_n}{n(n+1)}$ son estimadores insesgados de la media. ¿Cuál de ellos es el estimador más eficiente?

2. Sea $Y_i, i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una distribución normal con media θ desconocida y varianza σ^2 conocida. Demuestre que $\bar{Y}$ es un estimador de θ, insesgado y de mínima varianza.

3. Sea $Y_i, i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una distribución Normal con media θ, conocida, y varianza σ^2 desconocida. ¿$S_1^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \theta)^2$ es un estimador insesgado de mínima varianza de σ^2?

4. Sea $Y_i, i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una distribución normal con media θ y varianza σ^2, desconocidas. ¿$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$ es un estimador insesgado de mínima varianza de σ^2?

5. Sea Y_i, $i = 1, 2, \ldots, n$, una muestra aleatoria, de tamaño n, de una distribución Uniforme definida en el intervalo $(0, \theta)$. Demuestre que $\hat{\theta} = \left(\frac{n+1}{n}\right) Y_{(n)}$ no satisface la igualdad 2.14. Obsérvese que el conjunto $A = \{x \in \mathbb{R} : P_\theta(x) > 0\}$ depende de θ.

6. Sea Y_i, $i = 1, 2, \ldots, n$, una muestra aleatoria, de tamaño n, de una distribución Gamma (p, λ). Demuestre que $\bar{Y}$ es un estimador insesgado de varianza mínima de $E(Y_i)$.

2.6 Consistencia

La consistencia es una propiedad asintótica, es decir, indica que el valor de los errores de estimación serán más cercanos a cero, en probabilidad, a medida que el tamaño de la muestra tiende a infinito. En consecuencia, si un estimador no es insesgado, pero es consistente, su sesgo decrecerá a medida que crece el tamaño de la muestra. A continuación se ilustra esta propiedad, de los estimadores, a partir de una muestra aleatoria de una distribución Binomial.

Sea $Y_i \sim Ber(\theta)$, $i = 1, 2, 3, \ldots$; una sucesión de variables aleatorias independientes. Se define una sucesión $\{T_n\}_{n \in N}$ de estimadores de θ, haciendo $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$. Dado que $E(\bar{Y}_n) = \theta$ y $Var(\bar{Y}_n) = \sigma_n^2 = \frac{\sigma^2}{n}$, aplicando el teorema de Tchebysheff (inecuación 1.39), con $k^2 = \frac{\epsilon^2}{\sigma_n^2}$, se obtiene que

$$P(|\bar{Y}_n - \theta| < \epsilon) \geq 1 - \frac{\sigma_n^2}{\epsilon^2}. \tag{2.18}$$

Dado que $\sigma_n^2 = \frac{\sigma^2}{n}$ tiende a 0, cuando n tiende a infinito, si se toma el límite cuando n tiende a infinito en (2.18), resulta:

$$\lim_{n \to \infty} P(|\bar{Y}_n - \theta| < \epsilon) = 1.$$

Esto significa que a medida que n crece, se hace mayor la probabilidad de que las observaciones de $\bar{Y}_n$ se encuentren cerca de θ. Lo anterior hace que $\bar{Y}_n$ sea un estimador consistente de θ.

Definición 2.3. Sea $\{\hat{\theta}_n\}$ una sucesión de estimadores de θ. Se dice que $\hat{\theta}_n = T_n(X_1, X_2, \ldots, X_n)$ es un estimador consistente de θ si, para cualquier número positivo ϵ, $\lim_{n \to \infty} P(\mid \hat{\theta}_n - \theta \mid \leq \epsilon) = 1$.

Ejemplo 2.15. Sea $X_i \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$, una muestra aleatoria, de tamaño n, de una distribución normal con media y varianza desconocidas. Dado que $\bar{X}_n \sim N(\theta, \sigma^2/n)$, para todo $\epsilon > 0$, entonces,

$$\begin{aligned} P[|\bar{X}_n - \theta| \leq \epsilon] &= P\Big[\Big|\frac{\bar{X}_n - \theta}{\sigma/\sqrt{n}}\Big| \leq \frac{\sqrt{n}\epsilon}{\sigma}\Big] \\ &= 1 - 2P\Big(Z \geq \frac{\sqrt{n}\epsilon}{\sigma}\Big) \xrightarrow{n} 1, \end{aligned}$$

pues, si Z nota una variable normal estándar, $P\Big(Z \geq \frac{\sqrt{n}\epsilon}{\sigma}\Big)$ tiende a 0 cuando n tiende a infinito. Así, $\bar{X}_n$ es un estimador consistente de θ.

Ejemplo 2.16. Sean $X_i \sim U(0,\theta)$, $i = 1, 2, \ldots, n$; n variables aleatorias independientes. Dado ϵ tal que $0 < \epsilon < \theta$,

$$\begin{aligned} P[|X_{(n)} - \theta| \leq \epsilon] &= 1 - P[X_{(n)} \leq \theta - \epsilon] \\ &= 1 - \Big(1 - \frac{\epsilon}{\theta}\Big)^n. \end{aligned}$$

Dado que $(1-\epsilon/\theta)^n$ tiende a 0 cuando n tiende a infinito; $X_{(n)}$ es un estimador consistente de θ.

Teorema 2.2. *Si $\hat{\theta}_n$ es un estimador insesgado de θ y $\lim_{n\to\infty} Var(\hat{\theta}_n) = 0$, entonces, $\hat{\theta}_n$ es un estimador consistente de θ.*

Demostración. Del teorema de Tchebysheff se tiene que

$$P\Big(\Big|\hat{\theta}_n - E(\hat{\theta}_n)\Big| < k\sqrt{Var(\hat{\theta}_n)}\,\Big) \geq 1 - \frac{1}{k^2}, \tag{2.19}$$

haciendo

$$k = \frac{\epsilon}{\sqrt{Var(\hat{\theta}_n)}},$$

la inecuación (2.19) toma la forma

$$P(|\hat{\theta}_n - E(\hat{\theta}_n)| < \epsilon) \geq 1 - \frac{Var(\hat{\theta}_n)}{\epsilon^2} \tag{2.20}$$

y tomando el límite, cuando n tiende a infinito, en (2.20) se tiene lo que se quería demostrar. □

Ejemplo 2.17. Sea $Y_i, i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una distribución Uniforme en el intervalo $(0,\theta)$. Entonces, $\hat{\theta}_n = 2\bar{Y}$ es un estimador insesgado de θ, cuya varianza

$$Var(\hat{\theta}_n) = \frac{\theta^2}{3n}$$

tiende a 0 cuando n tiende a infinito. En consecuencia, $\hat{\theta}_n$ es un estimador consistente de θ.

Ejemplo 2.18. En las secciones anteriores se observa que si Y_i, $i = 1, \ldots, n$; es una muestra aleatoria, de tamaño n, de una distribución Normal con media μ y varianza $\sigma^2 < \infty$, entonces,

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

es un estimador insesgado de σ^2.

Dado que

$$\begin{aligned} Var(S^2) &= \frac{\sigma^4}{(n-1)^2} \, Var\left[\frac{(n-1)S^2}{\sigma^2}\right] \\ &= \frac{\sigma^4}{(n-1)^2} \, Var\left(\chi^2_{(n-1)}\right) \\ &= \frac{2\sigma^4}{(n-1)} \xrightarrow{n} 0, \end{aligned}$$

entonces, S^2 es un estimador consistente de σ^2.

Ejercicios 2.6.

1. Sean Y_i, $i = 1, 2, \ldots, n$; n variables aleatorias positivas, independientes e idénticamente distribuidas. Demuestre que $T_n(Y_1, \ldots, Y_n) = 1/\bar{Y}$ es un estimador consistente de $1/E(Y_1)$.

2. Sea $Y_1, i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una distribución Normal con media θ y varianza σ^2. Si k es un número entero menor que n,

   a) Demuestre que $S_k^2 = \dfrac{1}{k-1} \sum_{i=1}^{k} (Y_i - \bar{Y})^2$ es un estimador insesgado de σ^2.

   b) ¿S_k^2 es un estimador consistente de σ^2?

3. Sea $Y_i, i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una distribución Uniforme en el intervalo $(0, \theta)$.

   a) Si $\hat{\theta} = 2Y_1$, ¿$\hat{\theta}$ es un estimador insesgado de θ, de mínima varianza?

   b) Si $\hat{\theta} = 2Y_1$, ¿$\hat{\theta}$ es un estimador consistente de θ?

c) Si $\hat{\theta} = X_{(n)}$, ¿$\hat{\theta}$ es un estimador consistente de θ?

4. Sea $Y_i, i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, con una función de densidad

$$f(y) = \begin{cases} \frac{y}{2}, & 0 < y < 2; \\ 0, & \text{en otros puntos.} \end{cases}$$

Calcule:

a) $E(Y_i)$, $\mathrm{Var}(Y_i)$.

b) $P(|Y_1 - E(Y_1)| < 0.1)$.

c) $P(|\bar{Y} - E(\bar{Y})| < 0.1)$, para $n = 1, 10, 100$.

d) $\lim\limits_{n\to\infty} P(|\bar{Y} - E(\bar{Y})| < 0.1)$.

5. Sea $\hat{\theta}$ un estimador consistente de $\theta \in \mathbb{R}$. Si g es una función monótona diferenciable de $\mathbb{R}$ en $\mathbb{R}$, demuestre que $g(\hat{\theta})$ es un estimador consistente de $g(\theta)$. ¿$\sqrt{S^2}$ es un estimador consistente de σ?

Capítulo 3

Modelos estadísticos

En este capítulo se hace una exposición de lo que es un estadístico suficiente y la familia exponencial de distribuciones. La sección 3.1 incluye la definición de estadístico suficiente y el teorema de factorización, el cual permite determinar la suficiencia de un estadístico. En la sección 3.2 se define la familia exponencial uniparamétrica. En la sección 3.3 se presenta la familia exponencial unipararamétrica en forma natural. En la sección 3.4 se hace una introducción a la familia exponencial biparamétrica, a través de la distribución Normal, con media y varianza desconocidas. En la sección 3.5 se define la familia exponencial biparamétrica y se estudia su reparametrización en la forma natural y, finalmente, se incluyen algunos ejercicios que permiten afianzar los conceptos presentados.

3.1 Estadístico suficiente

Sea $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ una muestra aleatoria, de tamaño n, donde cada componente X_i, $i = 1, 2, \ldots, n$, es una variable aleatoria con función de distribución F_θ asociada a un parámetro desconocido θ. Un estadístico $\boldsymbol{T} = T(\boldsymbol{X})$, donde T es una función de $\boldsymbol{X}$, tiene sentido, cuando la variable aleatoria $\boldsymbol{T}$ contiene la misma información que $\boldsymbol{X}$, acerca del parámetro θ. Los estadísticos que tienen esta propiedad se denominan estadísticos suficientes.

Definición 3.1. Sea X_i, $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una función de distribución F_θ, con parámetro θ desconocido. Un

estadístico $T(\boldsymbol{X})$ es suficiente para θ, si y solo si, la distribución condicional de $\boldsymbol{X}$ dado $T(\boldsymbol{X}) = t$ no depende de θ.

Los análisis estadísticos, para obtener información acerca de θ, resultan más simples y sin pérdida de información a partir de un estadístico suficiente ya que, además de contener la misma información que $\boldsymbol{X}$ acerca de θ, el rango de un estadístico suficiente no trivial $T(\boldsymbol{X})$ es, en general, más simple que el rango de $\boldsymbol{X}$.

Ejemplo 3.1. Se desea obtener información sobre la proporción de tornillos defectuosos producidos por una empresa. De su producción se obtiene una muestra aleatoria con reemplazo, de tamaño n, asignando 1 a la variable aleatoria X_i, $i = 1, 2, \ldots, n$; si el tornillo es defectuoso y 0 si no lo es.

En este caso, $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$ es un estadístico suficiente para θ. En efecto, obsérvese que si $T(\mathbf{x}) = t$, entonces, $P_\theta(\boldsymbol{X} = \mathbf{x}, T(\mathbf{x}) = t) = P_\theta(\boldsymbol{X} = \mathbf{x}) = \theta^t(1-\theta)^{n-t}$; ya que $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ es un vector de n realizaciones de X_i y $X_i \sim Ber(\theta)$.

Así, dado que $T \sim Bin(n, \theta)$,

$$\begin{aligned} P_\theta(\boldsymbol{X} = \mathbf{x}|T = t) &= \frac{P_\theta(\boldsymbol{X} = \mathbf{x})}{P_\theta(T = t)} \\ &= \frac{\theta^t(1-\theta)^{n-t}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} \\ &= \frac{1}{\binom{n}{t}}. \end{aligned}$$

Finalmente, si $T(\mathbf{x}) \neq t$, entonces, $P_\theta(\boldsymbol{X} = \mathbf{x}, T = t) = 0$ y $P_\theta(\boldsymbol{X} = \mathbf{x}|T = t) = 0$. En consecuencia, $P_\theta(\boldsymbol{X} = \mathbf{x}|T = t)$ no depende de θ y, por tanto, $T(\boldsymbol{X})$ es un estadístico suficiente para θ.

Ejemplo 3.2. Se desea obtener información sobre la proporción de tornillos defectuosos producidos por una empresa. De su producción se obtiene una muestra aleatoria sin reemplazo, de tamaño n, asignando 1 a la variable aleatoria X_i, $i = 1, 2, \ldots, n$; si el i-ésimo tornillo es defectuoso y 0 si no lo es. Si el tamaño de la población es N y el muestreo se hace sin reemplazo, la probabilidad $P(\boldsymbol{X} = \mathbf{x}, T = t)$, donde $\boldsymbol{X} = (X_1, \ldots, X_n)'$ y $\mathbf{x} = (x_1, \ldots, x_n)'$, depende únicamente de t y no del orden en que los 1's aparecen en $\mathbf{x}$. En

consecuencia, se puede considerar que $x_i = 1$ para $i = 1, 2, \ldots, t$ y $x_i = 0$ para $i = t+1, t+2, \ldots, n$. Así, la probabilidad de que el primer tornillo obtenido en la muestra sea defectuoso es $N\theta/N$; si el primer tornillo es defectuoso, la probabilidad de que el segundo tornillo sea defectuoso es $(N\theta-1)/(N-1)$, y así sucesivamente. Después de obtener t tornillos defectuosos, la probabilidad de obtener un tornillo no defectuoso es $(N - N\theta)/(N - t)$ y continuando de esta manera se tiene que:

$$\begin{aligned} P_\theta(\boldsymbol{X} = \mathbf{x}, T = t) = \\ \left(\frac{N\theta}{N}\right)\left(\frac{N\theta-1}{N-1}\right)\cdots\left(\frac{N\theta-t+1}{N-t+1}\right)\times \\ \left(\frac{N-N\theta}{N-t}\right)\left(\frac{N-N\theta-1}{N-t-1}\right)\cdots\left(\frac{N-N\theta-n+t+1}{N-n+1}\right) \\ = \frac{(N\theta)!\,(N-N\theta)!(N-n)!}{N!(N\theta-t)!(N-N\theta-n+t)!} \end{aligned} \tag{3.1}$$

Finalmente, dado que T tiene distribución Hipergeométrica,

$$P_\theta(T = t) = \frac{(N\theta)!\,(N-N\theta)!}{t!N!(N\theta-t)!(N-N\theta-n+t)!}\frac{n!(N-n)!}{(n-t)!}, \tag{3.2}$$

y el cociente entre (3.1) y (3.2) no depende de θ. Esto indica que $T(X) = \sum_{i=1}^n X_i$ es un estadístico suficiente para θ.

El teorema siguiente se refiere a un conjunto de funciones de probabilidad denominadas modelos regulares, conformadas por las funciones de densidad continuas y por las funciones de probabilidad discretas $P_\theta(X)$ para las cuales existe un conjunto contable de valores posibles de X, digamos $A = \{x_1, x_2, \ldots\}$, tal que $P_\theta(A) = 1$.

Ejemplo 3.3. Sea $X_i \sim Exp(\lambda)$, $i = 1, \ldots, n$, una muestra aleatoria de tamaño n, de una distribución exponencial con media $\mu = 1/\lambda$. A partir de la definición 3.1 se puede demostrar que $T(\mathbf{X}) = \sum_{i=1}^n X_i$, donde $\mathbf{X} = (X_1, \ldots, X_n)$ es un estadístico suficiente para λ.

Dado $\mathbf{x} = (x_1, \ldots x_n)$ y $T(\mathbf{x}) = t$, se tiene que la función de densidad conjunta de $\mathbf{x}$ y t es:

$$f(\mathbf{x},t) = \begin{cases} \lambda^n e^{-\lambda \sum X_i} I_{(0,\infty)^n}(\mathbf{x}) & \text{si } T(\mathbf{x}) = t \\ 0 & \text{si } T(\mathbf{x}) \neq t \end{cases} \tag{3.3}$$

Así, dado que $Y = 2\lambda \sum X_i \sim \chi^2_{2n}$, la función de densidad de $T = \sum X_i$ es:

$$f(t) = \frac{t^{n-1}e^{-\lambda t}}{\Gamma(n)} I_{(0,\infty)}(t) \tag{3.4}$$

Finalmente, de 3.3 y 3.4 se concluye que la función de distribución condicional no depende de λ:

$$f(\mathbf{x}|t) = \begin{cases} \frac{\Gamma(n)}{t^{n-i}} I_{(0,\infty)^n}(\mathbf{x}) & \text{si } T(\mathbf{x}) = t \\ 0 & \text{si } T(\mathbf{x}) \neq t \end{cases}$$

Algunos ejemplos de modelos regulares son, entre otros, los modelos Binomial, Poisson, Normal y Gamma.

Teorema 3.1. En un modelo regular un estadístico $T(\boldsymbol{X})$, con rango I, es suficiente para θ, si y únicamente si, existe una función $g(t,\theta)$, definida para todo t en I y θ en Θ, y una función h, definida en $\mathbb{R}^n$, tal que

$$p(\mathbf{x},\theta) = g(t,\theta)h(\mathbf{x}). \tag{3.5}$$

donde $p(\mathbf{x},\theta)$ es la función de probabilidad de $\boldsymbol{X}$, en el caso discreto, o la función de densidad de $\boldsymbol{X}$, en el caso continuo.

Ejemplo 3.4. Sea $X_i \sim$ Exp (λ), $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una distribución exponencial con media $1/\lambda$. Si $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$, entonces, la densidad conjunta de $\boldsymbol{X}$ está dada por

$$\begin{aligned} f(\mathbf{x},\lambda) &= \lambda^n e^{-\lambda \sum_{i=1}^n x_i} I_{(0,\infty)^n}(\mathbf{x}) \\ &= g(t,\lambda)h(\mathbf{x}), \end{aligned}$$

donde $t(\mathbf{x}) = \sum_{i=1}^n x_i$, $g(t,\lambda) = \lambda^n e^{-\lambda t}$ y $h(\mathbf{x}) = I_{(0,\infty)^n}(\mathbf{x})$.

Así, por el teorema (3.1), $T(\boldsymbol{X}) = \sum_{i=1}^n X_i$ es un estadístico suficiente para λ.

Ejemplo 3.5. Si una función de densidad se puede expresar de la forma $p(x,\theta) = \exp[c(\theta)T(x) + d(\theta) + S(x)]I_A(x)$, entonces, $T(X)$ es un estadístico suficiente para θ.

Obsérvese que $p(x,\theta)$ se puede expresar de la forma (3.5) del teorema (3.1), haciendo $g(x,\theta) = \exp[c(\theta)T(x) + d(\theta)]$ y $h(x) = \exp[S(x)]I_A(x)$ y, por lo tanto, $T(\boldsymbol{X})$ es un estadístico suficiente para θ.

Ejemplo 3.6. Sea $X_i \sim U(0,\theta)$, $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una función de distribución Uniforme. Si $\boldsymbol{X} = (X_1, \ldots, X_n)$, entonces,

$$\begin{aligned} f(\mathbf{x},\theta) &= \frac{1}{\theta^n} I_{(x_{(n)},\,\infty)}(\theta) \\ &= g(x_{(n)},\theta)h(\boldsymbol{x}), \end{aligned}$$

donde $g(x_{(n)},\theta) = \frac{1}{\theta^n} I_{(x_{(n)},\,\infty)}(\theta)$ y $h(\mathbf{x}) = 1$ y, por lo tanto, $T(\boldsymbol{X}) = X_{(n)}$ es un estadístico suficiente para θ.

Para demostrar el teorema 3.1, en el caso discreto, es necesario verificar las siguientes proposiciones:

1. Si T es un estadístico suficiente para θ, entonces,

$$P(\mathbf{x},\theta) = g(T(\mathbf{x}),\theta)h(\mathbf{x}).$$

Demostración: Sea $\{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$, el conjunto de posibles realizaciones de $\boldsymbol{X}$ y $t_j = T(\mathbf{x}_j)$, $j = 1, 2 \ldots$ Por hipótesis, $P_\theta(\boldsymbol{X} = \mathbf{x}_i | T = t_j)$ no depende de θ, ya que T es un estadístico suficiente. Así, $P_\theta(\boldsymbol{X} = \mathbf{x}_i | T = t_j)$ es una función $h(\mathbf{x}_i)$. Además, si $T(\mathbf{x}_i) = t_j$, entonces, $P_\theta(\boldsymbol{X} = \mathbf{x}_i, T = t_j) = P_\theta(\boldsymbol{X} = \mathbf{x}_i)$ y si $T(\mathbf{x}_i) \neq t_j$, $P_\theta(\boldsymbol{X} = \mathbf{x}_i, T = t_j) = 0$. En los dos casos, la ecuación (3.5) se sigue a partir de la igualdad

$$P_\theta(\boldsymbol{X} = \mathbf{x}_i, T = t_j) = P_\theta(\boldsymbol{X} = \mathbf{x} | T = t_j) P_\theta(T = t_j),$$

haciendo

$$g(T(\mathbf{x}_i),\theta) = P_\theta(T = t_j).$$

2. Si $P(\mathbf{x}, \theta) = g(T(\mathbf{x}), \theta)h(\mathbf{x})$, entonces, T es un estadístico suficiente para θ.

   Demostración: Dado que $P_\theta(T = t_j) = \sum_{\{\mathbf{x}_k | T(\mathbf{x}_k) = t_j\}} P_\theta(\mathbf{x}_k)$, aplicando la hipótesis dada por la ecuación $P(\mathbf{x}, \theta) = g(T(\boldsymbol{x}), \theta)h(\mathbf{x})$, se tiene que

$$\begin{aligned} P_\theta(\boldsymbol{X} = \mathbf{x}_i | T = t_j) &= \frac{P_\theta(\boldsymbol{X} = \mathbf{x}_i, T = t_j)}{P_\theta(T = t_j)} \\ &= \frac{P_\theta(\boldsymbol{X} = \mathbf{x}_i)}{\sum_{\{\mathbf{x}_k | T(\mathbf{x}_k) = t_j\}} P_\theta(\mathbf{x}_k)} \\ &= \frac{h(\mathbf{x}_i)}{\sum_{\{\mathbf{x}_k | T(\mathbf{x}_k) = t_j\}} h(\mathbf{x}_k)}. \end{aligned}$$

   La segunda igualdad se sigue si $T(\mathbf{x}_i) = t_j$. Así, se encuentra que $P_\theta(\boldsymbol{X} = \mathbf{x}_i | T = t_j)$ no depende de θ, lo cual concluye la demostración.

Ejemplo 3.7. Estadísticos suficientes para la media y la varianza de una distribución normal.

Sea $X_i \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una distribución normal con parámetros $\boldsymbol{\theta} = (\theta, \sigma^2)$. La densidad conjunta de $\boldsymbol{X} = (X_1, \ldots, X_n)$ está dada por

$$\begin{aligned} f(\mathbf{x} | \theta, \sigma^2) &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\Big[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2 \Big] \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\Big[-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 + \frac{\theta}{\sigma^2} \sum_{i=1}^n x_i - \frac{n\theta^2}{2\sigma^2} \Big] \\ &= g(\boldsymbol{t}, \boldsymbol{\theta}) h(\mathbf{x}), \end{aligned}$$

donde $\boldsymbol{t}(\mathbf{x}) = (t_1(\boldsymbol{x}), t_2(\mathbf{x})) = (\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2)$ y $h(\mathbf{x}) = 1$.

Así, por el teorema (3.1), $T(\boldsymbol{X}) = (T_1(\boldsymbol{X}), T_2(\boldsymbol{X}))$, donde $T_1(\boldsymbol{X}) = \sum_{i=1}^n X_i$ y $T_2(\boldsymbol{X}) = \sum_{i=1}^n X_i^2$, es un estadístico suficiente para $\boldsymbol{\theta} = (\theta, \sigma^2)$.

Ejercicios 3.1.

1. De una producción de N tornillos se obtiene una muestra aleatoria, sin reemplazamiento, de tamaño n, y se le asigna a la variable X_i, $i = 1, 2, \ldots, n$; 1 si el tornillo es defectuoso y 0 si no. Si $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ y $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$, demuestre que

   a) Si $n = 3$ y $T = 2$, entonces, $p[\boldsymbol{X} = (1, 1, 0), T = 2] = p[\boldsymbol{X} = (1, 0, 1), T = 2]$.

   b) Si $n = 4$ y $T = 3$, entonces, $p[\boldsymbol{X} = (1, 1, 0, 1), T = 3] = p[\boldsymbol{X} = (1, 1, 1, 0), T = 3]$.

   c) $p[\boldsymbol{X} = \mathbf{x}, T = t]$ depende de t y no del orden en que los unos aparecen en $\mathbf{x}$.

2. Sea X_i, $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una distribución Exponencial con parámetro $\lambda = E(X_i)$, $i = 1, 2, \ldots, n$. Demuestre que $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$ es un estadístico suficiente para λ.

   a) Aplicando la definición 3.1.

   b) Aplicando el teorema 3.1.

3. Desarrolle el ejercicio anterior asumiendo que la función de densidad es:

   a) $\mathrm{Ber}(\theta)$.

   b) $\mathrm{N}(\theta, \sigma^2)$; σ^2 conocido.

4. Sea X_i, $i = 1, 2, \ldots, n$; es una muestra aleatoria, de tamaño n, de una distribución Gamma, con parámetro de forma conocido. Demuestre que $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$ es un estadístico suficiente para la media.

   a) Aplicando la definición 3.1.

   b) Aplicando el teorema 3.1.

5. Sea $X_i \sim B(p, q)$, $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una distribución Beta. Escriba la función de densidad Beta en términos de $\theta = E(X_i)$ y $\phi = p + q$. ¿$T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$ es un estadístico suficiente para θ?

6. Sea X_i, $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una distribución uniforme $U(0, \theta)$. Aplicando la definición 3.1, demuestre que $T(\boldsymbol{X}) = X_{(n)}$, es un estadístico suficiente para θ.

7. Si X_i, $i = 1, 2, \ldots, n$; es una muestra aleatoria, de tamaño n, de una distribución uniforme $U(\theta_1, \theta_2)$; $\theta_1 < \theta_2$, demuestre que $T(\boldsymbol{X}) = (X_{(1)}, X_{(n)})$, donde $\boldsymbol{X} = (X_1, \ldots, X_n)$, es un estadístico suficiente para $\boldsymbol{\theta} = (\theta_1, \theta_2)$.

3.2 Familia exponencial uniparamétrica

Se dice que una función de distribución pertenece a la familia exponencial uniparamétrica si su función de densidad o de probabilidad $p(x, \theta)$ puede expresarse de la forma

$$p(x, \theta) = \exp[c(\theta)T(x) + d(\theta) + S(x)]I_A(x), \tag{3.6}$$

donde c y d son funciones a valor real definidas en el conjunto de parámetros Θ; S y T son funciones a valor real, definidas sobre R y $A \subset R$ no depende de θ.

Ejemplo 3.8. La distribución Binomial $\text{Bin}(n, \theta)$ pertenece a la familia exponencial uniparamétrica de distribuciones. En efecto, ya que su función de densidad puede escribirse en la forma (3.6), donde $c(\theta) = \ln\left(\frac{\theta}{1-\theta}\right)$, $d(\theta) = n\ln(1-\theta)$, $T(x) = x$ y $S(x) = \ln\binom{n}{x}$, para $A = \{0, 1, 2, \ldots, n\}$.

Ejemplo 3.9. La distribución Normal, con media θ conocida y varianza σ^2 desconocida, pertenece a la familia exponencial uniparamétrica. En efecto, ya que su función de densidad puede expresarse en la forma (3.6), donde $c(\sigma^2) = -\frac{1}{2\sigma^2}$, $d(\sigma^2) = -\frac{1}{2}\ln(2\pi\sigma^2)$, $T(x) = (x-\theta)^2$ y $S(x) = 0$, para todo x en los números reales.

Si X_i, $i = 1, 2, \ldots, n$; son n variables aleatorias independientes e idénticamente distribuidas, con función de densidad en la familia exponencial uniparamétrica de distribuciones, entonces, la función de densidad conjunta $P(\mathbf{x}, \theta)$ está dada por

$$\begin{aligned} p(\mathbf{x},\theta) &= \prod_{i=1}^{n} \exp[c(\theta)T(x_i) + d(\theta) + S(x_i)]I_A(x_i) \\ &= \exp\left[c(\theta)\sum_{i=1}^{n} T(x_i) + nd(\theta) + \sum_{i=1}^{n} S(x_i)\right] I_{A^n}(\mathbf{x}) \\ &= \exp[c(\theta)T(\mathbf{x}) + d_0(\theta) + S(\mathbf{x})]I_{A^n}(\mathbf{x}), \end{aligned}$$

donde $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, $T(\mathbf{x}) = \sum_{i=1}^{n} T(x_i)$, $d_0(\theta) = nd(\theta)$, $S(\mathbf{x}) = \sum_{i=1}^{n} S(x_i)$ y A está determinado por la distribución de X_i, $i = 1, 2, \ldots, n$. En consecuencia, la función de densidad conjunta $p(\mathbf{x}, \theta)$ puede expresarse de la forma indicada en (3.6) y, por tanto, pertenece a la familia exponencial uniparamétrica.

Dado que la función de densidad $p(\mathbf{x}, \theta)$ puede escribirse como en la ecuación (3.5), donde $g(t, \theta) = \exp[c(\theta)T(\mathbf{x}) + d_0(\theta)]$ y $h(\mathbf{x}) = \exp[S(\mathbf{x})]I_{A^n}(\mathbf{x})$, se concluye que $T(\mathbf{x}) = \sum_{i=1}^{n} T(x_i)$ es un estadístico suficiente para θ.

En los ejemplos siguientes, relacionados con distribuciones de la familia exponencial uniparamétrica, se muestra que $T(\mathbf{x}) = \sum_{i=1}^{n} x_i$ es un estadístico suficiente para la media.

Ejemplo 3.10. Si X_i, $i = 1, 2, \ldots, m$; es una muestra aleatoria, de tamaño m, de una distribución $Bin(n, \theta)$, entonces,

$$P(\mathbf{x}, \theta) = \exp\left[\ln\left(\frac{\theta}{1-\theta}\right)\sum_{i=1}^{m} x_i + mn\ln(1-\theta) + \sum_{i=1}^{m} \ln\binom{n}{x_i}\right] I_{A^n}(\mathbf{x}),$$

donde $A = \{0, 1, 2, \ldots, n\}$ y $\mathbf{x} = (x_1, x_2, \ldots, x_m)$, y por lo tanto, la función de distribución de $\boldsymbol{X} = (X_1, X_2, \ldots, X_m)$ pertenece a la familia exponencial uniparamétrica.

Ejemplo 3.11. Si $X_i \sim N(\theta, \sigma^2)$, $i = 1, 2, 3, \ldots, n$; es una muestra aleatoria, de tamaño n, de una distribución normal con media desconocida y varianza conocida, entonces

$$\begin{aligned} P(\mathbf{x}, \theta) &= \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \theta)^2 - \frac{n}{2}\ln(2\pi\sigma^2)\right] \\ &= \exp\left[\frac{\theta}{\sigma^2}\sum_{i=1}^{n} x_i - \frac{n\theta^2}{2\sigma^2} - \frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2\right], \end{aligned}$$

pertenece a la familia exponencial uniparamétrica.

Nota: La familia de distribuciones uniformes $U(0,\theta)$, con θ desconocido, no pertenece a la familia exponencial (Shao, 2003).

3.3 Forma canónica de la familia exponencial uniparamétrica

Si una función de densidad o de probabilidad pertenece a la familia exponencial uniparamétrica (3.6) y si c es una función inyectiva, haciendo $\eta = c(\theta)$, puede escribirse como

$$p(x,\eta) = \exp[\eta T(x) + d_0(\eta) + S(x)]I_A(x), \tag{3.7}$$

donde $d_0(\eta) = d(c^{-1}(\eta))$.

Esta forma de escribir la familia exponencial uniparamétrica de distribuciones se conoce como forma canónica de la familia exponencial. η se conoce como parámetro natural de la distribución.

Si c no es una función inyectiva, $d_0(\eta)$ se puede determinar integrando simultáneamente los dos lados de la ecuación (3.7). Ya que $\int_A p(x,\eta)dx = 1$, se obtiene que

$$1 = \int_A \exp[\eta T(x) + d_0(\eta) + S(x)]dx,$$

de donde $d_0(\eta) = -\log \int_A \exp[\eta T(x) + S(x)]dx$.

Ejemplo 3.12. La distribución $Bin(n,\theta)$, que pertenece a la familia exponencial uniparamétrica de distribuciones, puede expresarse de la forma dada en (3.7) con $\eta = \ln(\frac{\theta}{1-\theta})$, $d_0(\eta) = -n\log(1+\exp(\eta))$ y $S(x) = \log\binom{n}{x}$, para $A = \{0,1,2,\ldots,n\}$.

Ejemplo 3.13. La distribución normal con media θ conocida, y varianza σ^2 desconocida, puede expresarse de la forma dada en (3.7), con $\eta = -\frac{1}{2\sigma^2}$ y $d_0(\eta) = -\frac{1}{2}\ln(-\frac{1}{\eta})$.

La forma canónica de esta distribución es

$$f_\eta(x) = \exp\left[\eta(x-\theta)^2 + \frac{1}{2}\ln(-\eta) - \frac{1}{2}\ln(\pi)\right].$$

Ejemplo 3.14. Suponga que X_i, $i = 1, 2, \ldots, n$, es una muestra aleatoria, de tamaño n, de una distribución Geométrica $G(\theta)$. La función de densidad conjunta de $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ es

$$f_\theta(\mathbf{x}) = \exp\left[\log(1-\theta)\sum_{i=1}^{n} x_i + n\log\left(\frac{\theta}{1-\theta}\right)\right] I_{N^n}(\mathbf{x}),$$

donde $N = \{1, 2, \ldots\}$, y su forma canónica es

$$f_\eta(\mathbf{x}) = \exp\left[\eta\sum_{i=1}^{n} x_i + n\log\left(\frac{1-e^\eta}{e^\eta}\right)\right] I_{N^n}(\mathbf{x}),$$

donde $\eta = \ln(1-\theta)$.

Teorema 3.2. Si X tiene distribución en la familia exponencial uniparamétrica (3.7) y η es un punto interior de $H = \{\eta\colon d_0(\eta)\, es\ \ finito\}$, la función generadora de momentos de $T(X)$ está dada por

$$\Psi(s) = \exp[d_0(\eta) - d_0(\eta + s)]. \tag{3.8}$$

Demostración 3.1. Asumiendo que X es una variable aleatoria continua, su función generadora de momentos está dada por

$$\begin{aligned}
\Psi(s) &= E(\exp(sT(X))) \\
&= \int_A \{\exp[(s+\eta)T(x) + d_o(\eta) + S(x)]\}dx \\
&= \int_A \{\exp[(s+\eta)T(x) + d_o(s+\eta) - d_o(s+\eta) + d_o(\eta) + S(x)]\}dx \\
&= \exp[d_o(\eta) - d_o(s+\eta)] \int_A \{\exp[(s+\eta)T(x) + d_o(s+\eta) + S(x)]\}dx \\
&= \exp[d_o(\eta) - d_o(s+\eta)].
\end{aligned}$$

Además de las hipótesis del teorema, en esta demostración se asume que s es suficientemente pequeño, tal que $\eta + s$ pertenece a H.

De (3.8) se sigue que $E[T(X)] = -d_0'(\eta)$ y $Var[T(X)] = -d_0''(\eta)$. Estos resultados se obtienen evaluando en $s = 0$ la primera y segunda derivada de $\Psi(s)$, como se indica a continuación:

a) $E[T(X)] = \psi^{'}(s)|_0 = -d_0^{'}(\eta)$.

b) Dado que

$$\begin{aligned} E[T^2(X)] &= \psi^{''}(s)|_0 \\ &= \exp[d_0(\eta) - d_0(\eta+s)][(d_0^{'}(\eta+s))^2 - d_0^{''}(\eta)]|_0 \\ &= [d_0^{'}(\eta)]^2 - d_0^{''}(\eta), \end{aligned}$$

teniendo en cuenta el literal a) se tiene que

$$\begin{aligned} Var[T(X)] &= E[T^2(X)] - (E[T(X)])^2 \\ &= -d_0^{''}(\eta). \end{aligned}$$

Ejemplo 3.15. Sea X_i, $i = 1, 2, \ldots n$; una muestra aleatoria, de tamaño n, de una distribución Bernoulli con parámetro θ. Dado que la función de probabilidad de $\boldsymbol{X} = (X_1, \ldots, X_n)$ puede expresarse en la forma

$$f_\eta(\mathbf{x}) = \exp\left[\eta \sum_{i=1}^{n} X_i - n\log(1+\exp(\eta))\right] I_{A^n}(\mathbf{x}),$$

donde $\eta = \log(\theta/(1-\theta))$ y $A = \{0, 1\}$, entonces, $d_0(\eta) = -n\log(1+\exp(\eta))$ y

$$\begin{aligned} E[T(X)] &= -d_0{}'(\eta) = n\left[\frac{\exp(\eta)}{1+\exp(\eta)}\right] = n\theta \\ Var[T(X)] &= -d_0{}''(\eta) = n\left[\frac{\exp(\eta)}{(1+\exp(\eta))^2}\right] = n\theta(1-\theta). \end{aligned}$$

Ejemplo 3.16. Supóngase que X_i, $i = 1, 2, \ldots, n$; es una muestra aleatoria, de tamaño n, de una distribución normal $N(\theta, \sigma^2)$, con media conocida. Dado que la función de densidad de $\boldsymbol{X} = (X_1, \ldots, X_n)$ puede escribirse en la forma

$$f_\eta(\mathbf{x}) = \exp\left[\eta \sum_{i=1}^{n} (x_i - \theta)^2 + \frac{n}{2}\ln(-\eta) - \frac{n}{2}\ln(\pi)\right],$$

entonces,

$$\begin{aligned} E[T(\boldsymbol{X})] &= -d_0{}'(\eta) = -\frac{n}{2\eta} = n\sigma^2 \\ Var[T(\boldsymbol{X})] &= -d_0{}''(\eta) = \frac{n}{2\eta^2} = 2n\sigma^4, \end{aligned}$$

donde $T(\boldsymbol{X}) = \sum_{i=1}^{n}(x_i - \theta)^2$ y $d_0(\eta) = \frac{n}{2}\ln(-\eta)$.

Ejemplo 3.17. Suponga que X_i, $i = 1, 2, \ldots, n$; es una muestra aleatoria, de tamaño n, de una distribución Geométrica $G(\theta)$. Si $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$, entonces, la función generadora de momentos de $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$, del ejemplo (3.14), está dada por

$$\begin{aligned}
\Psi(t) &= \exp[d_0(\eta) - d_0(\eta + s)] \\
&= \exp\left[n \log\left(\frac{1 - e^{\eta}}{e^{\eta}}\right) - n \log\left(\frac{1 - e^{\eta+s}}{e^{\eta+s}}\right)\right] \\
&= \left[\frac{(1 - e^{\eta})e^{s}}{1 - e^{\eta+s}}\right]^{n} \\
&= \left[\frac{\theta e^{s}}{1 - (1 - \theta)e^{s}}\right]^{n}.
\end{aligned}$$

Esta es la función generadora de momentos de una variable aleatoria con función de distribución Binomial Negativa. Esto indica que $\sum_{i=1}^{n} X_i \sim BN(\theta, n)$.

Ejercicios 3.2.

1. Demuestre que la distribución $P_\theta(x) = (\theta^x e^{-\theta}/x!) I_{\{0,1,2,\ldots\}}(x)$, $\theta > 0$, pertenece a la familia exponencial. Encuentre su forma canónica y el conjunto de valores posibles del parámetro natural. Aplique el teorema (3.2) para encontrar su función generadora de momentos y exprésela como una función del parámetro natural y como una función de θ.

2. Demuestre que si r es fijo, entonces, la función de densidad binomial negativa pertenece a la familia exponencial. Encuentre la forma canónica, el conjunto de valores posibles del parámetro natural y su función generadora de momentos, como en el ejercicio anterior.

3. Desarrolle el ejercicio 2, asumiendo que P_θ es la función de densidad:

   a) Ber(θ).

   b) Exp(θ).

   c) N(θ, σ^2); σ^2 conocido.

   d) $\Gamma(\theta, \alpha)$; con parámetro de forma α conocido y media θ desconocida.

4. En el caso de la distribución Normal, con media conocida y varianza desconocida, por ejemplo (3.13), demuestre que

$$d_0(\eta) = -\log \int_R \exp[\eta(x-\theta)^2 - \frac{1}{2}\log(\pi)]dx.$$

5. Si X tiene distribución Geométrica con parámetro θ, demuestre que

$$d_0(\eta) = -\log\Big(\sum_{x=1,2,\ldots} \exp(\eta x)\Big).$$

6. Desarrolle el ejercicio 3 considerando una muestra aleatoria de tamaño n de cada una de las poblaciones dadas allí.

7. En cada uno de los ejercicios anteriores, halle $E[T(X)]$ y $Var[T(X)]$ aplicando el teorema 3.2.

8. Escriba cada una de las funciones de densidad, del ejercicio 3, en la forma canónica o natural de la familia exponencial uniparamétrica y halle la estimación de máxima verosimilitud del parámetro natural.

9. Determine las condiciones necesarias para que exista la estimación de máxima verosimilitud de η en la familia exponencial (3.7).

10. Supóngase que X_i, $i = 1, \ldots, n$; es una muestra aleatoria, de tamaño n, de una distribución de la familia exponencial uniparamétrica de la forma $f_\eta(x_i) = \exp[\eta x_i + d_0(\eta) + S(x_i)]I_A(x_i)$. Demuestre que el estimador de máxima verosimilitud de $E(X_1)$ es $-\hat{d}_0'(\eta) = \overline{X}$.

3.4 Familia exponencial biparamétrica

Sea $\boldsymbol{X}$ una variable aleatoria n-dimensional con componentes independientes $X_i \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$. Si θ y σ^2 son desconocidos, la función de densidad de $\boldsymbol{X}$ se puede escribir como

$$
\begin{aligned}
f_\theta(\mathbf{x}) &= (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2\right] \qquad (3.9)\\
&= (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i^2-2\theta x_i+\theta^2\right)\right]\\
&= \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2+\frac{\theta}{\sigma^2}\sum_{i=1}^{n}x_i-\frac{n\theta^2}{2\sigma^2}-\frac{n}{2}\log(2\pi\sigma^2)\right]\\
&= \exp\left[\eta' T(x)+\xi(\eta)\right],
\end{aligned}
$$

donde $\eta' = (\eta_1, \eta_2), T' = (T_1, T_2),\ \eta_1 = \frac{\theta}{\sigma^2},\ \eta_2 = -\frac{1}{2\sigma^2},\ T_1(\mathbf{x}) = \sum_{i=1}^{n} x_i$, $T_2(\mathbf{x}) = \sum_{i=1}^{n} x_i^2,\ y\ \xi(\eta) = -\frac{n\theta^2}{2\sigma^2} - \frac{n}{2}\ln(2\pi\sigma^2)$.

En consecuencia, la función de distribución dada por la ecuación (3.9), puede escribirse en la forma

$$
f_\eta(\mathbf{x}) = \exp\left[\eta' T(x)+\xi(\eta)+n\sum_{i=1}^{n}S(x_i)\right] I_A(x_1,\ldots,x_n).
$$

Esta función caracteriza a la familia exponencial biparamétrica de distribuciones, con $S(x_i) = 0$ y $\xi(\eta) = \frac{n\eta_2^2}{4\eta_1} - \frac{n}{2}\ln(-\frac{\pi}{\eta_1})$.

Obsérvese que

$$
\begin{aligned}
\frac{\partial\xi}{\partial\eta_1} &= -\frac{n\eta_2}{2\eta_1} = -n\theta = -E[T_2(X)]\\
\frac{\partial\xi}{\partial\eta_2} &= -\frac{n}{4}\frac{\eta_2^2}{\eta_1^2}+\frac{n}{2\eta_1} = -n(\theta^2+\sigma^2) = -E[T_1(X)]\\
\frac{\partial^2\xi}{\partial\eta_1^2} &= -\frac{n}{2\eta_1} = -n\sigma^2 = -Var[T_1(X)]\\
\frac{\partial^2\xi}{\partial\eta_2^2} &= \frac{n}{2}\frac{\eta_2^2}{\eta_1^3}+\frac{n}{2\eta_1^2} = -2n\sigma^2(\theta^2+\sigma^2) = -Var[T_2(X)].
\end{aligned}
$$

Nota: De expresar la distribución Normal, con media y varianza desconocidas, en la forma de la familia exponencial biparamétrica, se concluye que

$$
T'(\mathbf{x}) = \Big(\sum_{i=1}^{n}x_i,\ \sum_{i=1}^{n}x_i^2\Big)
$$

es un estadístico suficiente para $\theta = (\mu, \sigma^2)$.

3.5 Forma canónica de la familia exponencial biparamétrica

Una función de densidad que pertenece a la familia exponencial biparamétrica tiene función de densidad o de probabilidad que puede escribirse de la forma

$$P(x,\theta) = \exp[(c(\theta))^t T(x) + d(\theta) + S(x)]I_A(x), \tag{3.10}$$

donde $[c(\theta)]^t = (c_1(\theta), c_2(\theta))$ es una función del espacio paramétrico Θ en $\mathbb{R}^2$; $T'(x) = (T_1(x), T_2(x))$ es una función de $\mathbb{R}^n$ en $\mathbb{R}^2$; $S(x)$ es una función de $\mathbb{R}^n$ en $\mathbb{R}$; y $d(\theta)$ es una función de Θ en $\mathbb{R}$.

Si $c = (c_1, c_2)$ es una función inyectiva, entonces, haciendo $\eta_1 = c_1(\theta)$ y $\eta_2 = c_2(\theta)$, la función de densidad dada en (3.10) puede escribirse de la forma

$$P(x,\eta) = \exp[\eta^t T(x) + d_0(\eta) + S(x)]I_A(x), \tag{3.11}$$

donde $\eta = (\eta_1, \eta_2)^t$ y $d_0(\eta) = d(c^{-1}(\eta))$.

Esta forma de escribir la familia exponencial biparamétrica de distribuciones se conoce como forma natural o canónica de la familia exponencial biparamétrica, donde η se conoce como parámetro natural de la distribución.

Si c no es una función inyectiva, entonces,

$$d_0(\eta) = -\log \int_A \exp[\eta T(x) + S(x)]dx,$$

donde $A = \{x \in \mathbb{R}^n : p_\theta(x) > 0\}$.

Las estimaciones de máxima verosimilitud de $\eta = (\eta_1, \eta_2)$ se pueden obtener solucionando el siguiente sistema de ecuaciones:

$$\frac{\partial d_0}{\partial \eta_1} = -T_1(x) \tag{3.12}$$

$$\frac{\partial d_0}{\partial \eta_2} = -T_2(x) \tag{3.13}$$

En el siguiente ejemplo, dada una muestra aleatoria de una distribución Normal, con media y varianza desconocidas, se obtienen las estimaciones de máxima verosimilitud del parámetro natural η.

Ejemplo 3.18. Si X_i, $i = 1, 2, \ldots, n$, es una muestra aleatoria, de tamaño n, de una distribución Normal con media y varianza desconocidas, entonces,

$$p_\theta(x) = \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2 + \frac{\theta}{\sigma^2}\sum_{i=1}^{n} x_i + d(\theta) - \frac{n}{2}\ln(2\pi)\right], \tag{3.14}$$

donde $d(\theta) = -\dfrac{n\theta^2}{2\sigma^2} - \dfrac{n}{2}\ln(\sigma^2)$.

Ahora, si $\eta = (\eta_1, \eta_2)$, $\eta_1 = -\dfrac{1}{2\sigma^2}$, $\eta_2 = \dfrac{\theta}{\sigma^2}$ y $d_0(\eta) = \dfrac{n}{2}\ln(-2\eta_1) + \dfrac{n}{4}\dfrac{\eta_2^2}{\eta_1}$, entonces, la función de densidad (3.14) puede escribirse de la forma natural. Si x_i es una realización de X_i, $i = 1, \ldots, n$, entonces, el logaritmo de la función de verosimilitud es

$$l = \eta_1 T_1(x) + \eta_2 T_2(x) + d_0(\eta) - \frac{n}{2}\ln(2\pi).$$

Dado que

$$\frac{\partial d_0}{\partial \eta_1} = \frac{n}{2\eta_1} - \frac{n}{4}\frac{\eta_2^2}{\eta_1^2}$$

y

$$\frac{\partial d_0}{\partial \eta_2} = \frac{n}{2}\frac{\eta_2}{\eta_1},$$

las soluciones de las ecuaciones (3.12) y (3.13) son:

$$\hat{\eta}_1 = \frac{n^2}{2[T_2^2(x) - nT_1(x)]}$$

$$\hat{\eta}_2 = \frac{-T_2(x)}{T_2^2(x) - nT_1(x)}.$$

Así, $(\hat{\eta}_1, \hat{\eta}_2)^t$ es la estimación de máxima verosimilitud de $(\eta_1, \eta_2)^t$, ya que

$$\frac{\partial^2 d_0}{\partial \eta_2^2} = \frac{n}{2}\frac{1}{\eta_1} < 0 \qquad \text{y} \qquad \frac{\partial^2 d_0}{\partial \eta_2^2}\frac{\partial^2 d_0}{\partial \eta_1^2} - \left(\frac{\partial^2 d_0}{\partial \eta_1 \eta_2}\right)^2 = -\frac{n^2}{4\eta_1^3} > 0.$$

Dado que $\eta = (\eta_1, \eta_2)$ es una función biyectiva de Θ, en el espacio paramétrico Γ de η, entonces, $\hat{\theta} = \frac{1}{n}T_2(X)$ y $\hat{\sigma}^2 = \frac{1}{n}T_1(X) - \frac{1}{n^2}T_2^2(X)$ son los estimadores de máxima verosimilitud de θ y σ^2, respectivamente.

Ejercicios 3.3.

1. Muestre que las siguientes familias de distribuciones pertenecen a la familia exponencial biparamétrica:

   a) Gamma.

   b) Beta.

   a) Normal.

2. Escriba las funciones de distribución del ejercicio anterior en forma natural y determine los estimadores de máxima verosimilitud del parámetro natural $\eta = (\eta_1, \eta_2)$.

Capítulo 4

Propiedades de los estimadores

En este capítulo se presentan algunos procedimientos para encontrar estimadores insesgados óptimos. Inicialmente se incluyen los conceptos de error cuadrático medio y de error medio absoluto, como medidas de "qué tan bueno es un estimador". Luego se incluyen algunos teoremas, siguiendo a Bickel & Doksum (2001) cuya aplicación permite encontrar estimadores insesgados uniformes de mínima varianza y finalmente se incluye un estudio sobre el número de información de Fisher, que puede interpretarse como la cantidad de información que contienen las observaciones acerca del parámetro, mediante el cual se expresa una cota inferior para la varianza de los estimadores.

4.1 Errores cuadrático medio y medio absoluto

Sea $T(\boldsymbol{X})$ un estimador de $q(\theta)$. Para determinar si $T(\boldsymbol{X})$ es un buen estimador de $q(\theta)$, usualmente se usa el error cuadrático medio, notado $R(\theta, T)$ y definido como

$$R(\theta, T) = E_\theta[T(\boldsymbol{X}) - q(\theta)]^2. \tag{4.1}$$

El error cuadrático medio se puede expresar como la suma de la varianza y el cuadrado del sesgo de $T(\boldsymbol{X})$. Esto es,

$$\begin{aligned} R(\theta, T) &= Var_\theta(T(\boldsymbol{X})) + [E_\theta(T(\boldsymbol{X})) - q(\theta)]^2 \\ &= Var_\theta(T(\boldsymbol{X})) + b^2(\theta, T), \end{aligned}$$

donde $b(\theta, T)$ es el sesgo de $T(\boldsymbol{X})$ como estimador de $q(\theta)$. Según este criterio, un buen estimador $T(\boldsymbol{X})$ de $q(\theta)$ debe ser insesgado y de mínima varianza, para todo θ en el espacio paramétrico Θ.

El error medio absoluto, $R_1(\theta, T) = E_\theta[|T(\boldsymbol{X}) - q(\theta)|]$, es otra medida de qué tan bueno es $T(\boldsymbol{X})$ como estimador de $q(\theta)$. Si $T(\boldsymbol{X})$ es un estimador insesgado de $q(\theta)$, normalmente distribuido, la comparación del error medio absoluto con el error cuadrático medio es sencilla. Obsérvese que

$$\begin{aligned} E_\theta[|T(\boldsymbol{X}) - q(\theta)|] &= \sigma_T E\left[\left|\frac{T(\boldsymbol{X}) - q(\theta)}{\sigma_T}\right|\right] \\ &= \sqrt{\frac{2}{\pi}}\sigma_T \int_0^\infty z e^{-\frac{1}{2}z^2} dz \\ &= \sqrt{2/\pi}\sqrt{R(\theta, T)}. \end{aligned}$$

Otros resultados relacionados con el error cuadrático medio son de gran interés. Por ejemplo, si $T(\boldsymbol{X})$ es un estimador insesgado de $q(\theta)$, de la desigualdad (1.39) resulta

$$P_\theta[|T(\boldsymbol{X}) - q(\theta)| < \epsilon] \geq 1 - \frac{E[(T(\boldsymbol{X}) - q(\theta))^2]}{\epsilon^2},$$

para todo $\epsilon > 0$. En consecuencia, si $E_\theta[T(\boldsymbol{X}) - q(\theta)]^2 \to 0$ cuando $n \to \infty$, entonces, $P_\theta[|T(\boldsymbol{X}) - q(\theta)| \leq \epsilon] \to 1$. Esto significa que si $R(\theta, T) \to 0$, $T(\boldsymbol{X}) \to q(\theta)$ en probabilidad.

Ejemplo 4.1. Si $Y_i \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$; es una muestra aleatoria, de tamaño n, de una distribución normal con media y varianza desconocidas, entonces,

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

es un estimador insesgado de σ^2.

Por otra parte, aplicando el teorema (1.2) se tiene que

$$Var(S^2) = \frac{\sigma^4}{(n-1)^2} Var\Big(\frac{(n-1)S^2}{\sigma^2}\Big) = \frac{2\sigma^4}{n-1}$$

y así

$$R(\sigma^2, S^2) = Var(S^2) + b^2(\sigma^2, S^2) = \frac{2\sigma^4}{n-1}. \tag{4.2}$$

Ejercicios 4.1.

1. Sea X una variable aleatoria con función de densidad

$$f(x,\theta) = \frac{2(\theta - x)}{\theta^2} I_{(0,\theta)}(x). \tag{4.3}$$

   a) Halle el valor de a que minimiza $E[(X - \frac{1}{a}\theta)^2]$.

   b) Halle el error cuadrático medio de $T(X) = X$, como estimador de $q(\theta) = \theta$.

   c) Asumiendo que X_i, $i = 1, \ldots, n$, es una muestra aleatoria de tamaño n de (4.3), halle el error cuadrático medio de $T(X) = \bar{X}$ como estimador de $q(\theta) = \theta/3$.

2. En el ejemplo (4.1), halle el error cudrático medio de $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (Y_i - \bar{Y})^2$ como estimador de σ^2. Es $R(\sigma^2, S^2) < R(\sigma^2, \hat{\sigma}^2)$?

3. Sea $X_1, \ldots, X_n$ una muestra aleatoria, de tamaño n, de una distribución $U(0,\theta)$. ¿Es $R(\theta, \frac{n+1}{n}X_{(n)}) < R(\theta, 2\bar{X})$?

4.2 Completitud

Un estadístico T es completo, si y solo si, para cada función de valor real U, definida en el rango de T, que satisface $E_\theta(U(T)) = 0$; para todo $\theta \in \Theta$; se tiene que $P_\theta(U(T) = 0) = 1$, para todo $\theta \in \Theta$.

Ejemplo 4.2. Una variable aleatoria X tiene distribución en la familia de distribución Hipergeométrica, si su función de densidad está dada por

$$P_\theta(X = x) = \frac{\binom{N\theta}{x}\binom{N-N\theta}{n-x}}{\binom{N}{n}}, \tag{4.4}$$

donde x es un número real tal que $\text{máx}(n - N(1-\theta), 0) \leq x \leq \text{mín}(N\theta, n)$.

Del teorema 3.1 se sigue que $T(X) = X$ es un estadístico suficiente para θ. Para demostrar que es completo se debe probar que si $E_\theta(U(X)) = 0$; para todo θ, $\theta = 0, 1/N, \ldots, 1$, entonces, $U(x) = 0$; para todo x, tal que $P_\theta(X = x) > 0$.

Si $\theta = 0$, entonces, $P_\theta(X = 0) = 1$ y, en consecuencia, $E_\theta(U(X)) = 0$ implica $U(0) = 0$. Si $\theta = 1/N$, entonces, $P_\theta(X = x) > 0$ para $x = 0, 1$, y $P_\theta(X = x) = 0$; para $2 \leq x \leq n$. Así,

$$\begin{aligned} E_\theta(U(X)) &= U(0)P_\theta(X = 0) + U(1)P_\theta(X = 1) \\ &= U(1)P_\theta(X = 1) \\ &= 0. \end{aligned}$$

Por tanto, dado que $U(0) = 0$ y que $P_\theta(X = 1) \neq 0$, entonces, $U(1) = 0$.

Finalmente, procediendo por inducción matemática, se concluye que si $\theta = 1$, $U(0) = U(1) = \ldots = U(n) = 0$, es decir, que $U(x) = 0$; para todo x, tal que $P_\theta(X = x) > 0$.

Ejemplo 4.3. Sean $X_i \sim U(0, \theta)$, $i = 1, 2, \ldots, n$; n variables aleatorias independientes. Dado que la función de densidad conjunta de la variable $X = (X_1, \ldots, X_n)$ es $f_\theta(\boldsymbol{x}) = \frac{1}{\theta^n} I_{(0,\theta)}(x_{(n)})$, del teorema 3.1 se sigue que $T = X_{(n)}$ es un estadístico suficiente para θ.

Dado que

$$\begin{aligned} F(t) &= P(X_{(n)} \leq t) \\ &= P(X_1 \leq t, \ldots, X_n \leq t) \\ &= \frac{t^n}{\theta^n} I_{(0,\theta)}(t), \end{aligned}$$

la función de densidad de T es

$$f_T(t) = \frac{nt^{n-1}}{\theta^n} I_{(0,\theta)}(t).$$

Así, $E_\theta(U(T)) = \frac{n}{\theta^n} \int_0^\theta u(t)t^{n-1}dt = 0$, si y solo si, $\int_0^\theta u(t)t^{n-1}dt = 0$. Finalmente, si u es una función continua, aplicando la diferenciación de una integral se encuentra que $u(\theta)\theta^{n-1} \overset{\text{cs}}{=} 0$; $\theta \in R^+$. Así, $u(t) \overset{\text{cs}}{=} 0$; para $t > 0$ y, en consecuencia, $X_{(n)}$ es un estadístico completo y suficiente.

Teorema 4.1. Sea P_θ; $\theta \in \Theta$, una familia exponencial uniparamétrica dada por $p_\theta(\mathbf{x}) = \exp\{c(\theta)T(\mathbf{x}) + d(\theta) + S(\boldsymbol{x})\}I_A(\boldsymbol{x})$. Si el rango de c tiene interior no vacío, entonces, $T(\boldsymbol{X})$ es completo y suficiente para $\theta \in \Theta$.

Ejemplo 4.4. Sean $X_i \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$; n variables aleatorias independientes. Si la media es desconocida y la varianza conocida, por el teorema (4.1), $T_1(\boldsymbol{X}) = \bar{X}$ es un estimador completo y suficiente de θ.

Ahora, si la media es conocida y la varianza desconocida, por el mismo teorema, $T_2(\boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \theta)^2$ es un estadístico completo y suficiente de σ^2.

Ejercicios 4.2.

1. Sean X_i, $i = 1, 2, \ldots, n$; n variables aleatorias independientes con función de densidad $f_\theta(x) = \frac{1}{\theta}e^{-\frac{1}{\theta}x}I_{(0,\infty)}(x)$; $\theta > 0$. Calcule el error cuadrático medio de $T(\mathbf{X}) = \bar{X}$ como estimador de θ.

2. Sean X_i, $i = 1, 2, \ldots, n$; n variables aleatorias independientes con función de densidad $f_\theta(x) = \theta e^{-\theta x}I_{(0,\infty)}(x)$; $\theta > 0$. Calcule el error cuadrático medio de $T(\boldsymbol{X}) = 1/\bar{X}$ como un estimador de θ.

3. Sean $X_i \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$; n variables aleatorias independientes, con media conocida y varianza desconocida. Determine los errores cuadráticos medios de $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \theta)^2$ y de $\hat{\sigma}^2 = \frac{n-1}{n}S^2$ como estimadores de σ^2.

4. Sean $X_i \sim U(0, \theta)$, $i = 1, 2, \ldots, n$; n variables aleatorias independientes. Determine el error cuadrático medio de $T_1(\boldsymbol{X}) = X_{(n)}$ y $T_2(\boldsymbol{X}) = 2\bar{X}$ como estimadores de θ.

5. Sean X_i, $i = 1, 2, \ldots, n$; n variables aleatorias independientes con función de densidad $f_\theta(x) = \frac{1}{\theta}e^{-\frac{1}{\theta}x}I_{(0,\infty)}(x)$; $\theta > 0$. Calcule el error cuadrático medio de $T(\boldsymbol{X}) = \bar{X}$ como estimador de θ.

6. Sean $X_i \sim Ber(\theta)$, $i = 1, 2, \ldots, n$; n variables aleatorias independientes. Demuestre que $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$ es un estadístico completo y suficiente para θ.

7. Sean X_i, $i = 1, 2, \ldots, n$; n variables aleatorias independientes con función de densidad $f_{\alpha,\theta}(x) = \frac{\alpha^\alpha}{\Gamma(\alpha)\theta^\alpha}x^{\alpha-1}e^{-\frac{\alpha}{\theta}x}I_{(0,\infty)}(x)$; $\theta > 0$, $\alpha > 0$. Si α es conocido, demuestre que $T(\boldsymbol{X}) = \bar{X}$ es completo y suficiente para θ.

4.3 Estadísticos insesgados uniformes de mínima varianza

Sea Γ el conjunto de todos los estimadores insesgados de $q(\theta)$. Si un estimador T de $q(\theta)$, $T \in \Gamma$, tiene la menor varianza para todos los valores de θ, se dice que T es un estimador insesgado uniforme de mínima varianza (UMVU, por su nombre en inglés) de $q(\theta)$.

A continuación se introduce el teorema de Rao Blackwell. Sea $X = (X_1, \ldots, X_n)$ una muestra aleatoria, de tamaño n, de una distribución P_θ; $\theta \in \Theta$. Si $T(\boldsymbol{X})$ es una estadística suficiente para θ y $S(\boldsymbol{X})$ es un estimador de $q(\theta)$, entonces, $T^*(\boldsymbol{X}) = E_\theta(S(\boldsymbol{X})|T(\boldsymbol{X}))$ es un estimador de $q(\theta)$ y $b(\theta, T^*) = b(\theta, S)$.

Este resultado se sigue de si $T(\boldsymbol{X})$ es un estadístico suficiente para θ, entonces, $P(\boldsymbol{X}|T(\boldsymbol{X}) = t)$ no depende de θ. En consecuencia, $E_\theta(S(\boldsymbol{X})|T(\boldsymbol{X}))$ no depende de θ y, por lo tanto, $T^*(\boldsymbol{X})$ es un estimador de $q(\theta)$.

Además,

$$\begin{aligned} b(\theta, T^*) &= E_\theta[E_\theta(S(\boldsymbol{X})|T(\boldsymbol{X})] - q(\theta) \\ &= E_\theta(S(\boldsymbol{X})) - q(\theta) \\ &= b(\theta, S). \end{aligned}$$

Teorema 4.2 (teorema de Rao Blackwell)**.** Sea $\boldsymbol{X} = (X_1, \ldots, X_n)$ una muestra aleatoria, de tamaño n, de una distribución P_θ; $\theta \in \Theta$. Si $T(\boldsymbol{X})$ es un estadístico suficiente de θ, y $S(\boldsymbol{X})$ es un estimador insesgado de $q(\theta)$, tal que $Var_\theta(S(\boldsymbol{X})) < \infty$ para todo θ, entonces, $T^*(\boldsymbol{X})$ es un estimador de $q(\theta)$ y $Var_\theta(T^*(\boldsymbol{X})) \leq Var_\theta(S(\boldsymbol{X}))$ para todo θ.

Demostración. Obsérvese que

$$\begin{aligned} Var_\theta(S(\boldsymbol{X})) &= E_\theta[S(\boldsymbol{X}) - q(\theta)]^2 \\ &= E_\theta[S(\boldsymbol{X}) - E_\theta(S(\boldsymbol{X})|T(\boldsymbol{X})) + E_\theta(S(\boldsymbol{X})|T(\boldsymbol{X})) - q(\theta)]^2 \\ &= E_\theta[S(\boldsymbol{X}) - E_\theta(S(\boldsymbol{X})|T(\boldsymbol{X}))]^2 + E_\theta[E_\theta(S(\boldsymbol{X})|T(\boldsymbol{X})) - q(\theta)]^2 \\ &\geq E_\theta[E_\theta(S(\boldsymbol{X})|T(\boldsymbol{X})) - q(\theta)]^2 \\ &= Var_\theta(T^*(\boldsymbol{X})). \end{aligned}$$

La tercera igualdad se sigue dado que

$$\begin{aligned}\int_R (S(\boldsymbol{X}) - T^*(\boldsymbol{X}))f_1(\mathbf{x}|\ t)d\mathbf{x} &= E_\theta(S(\boldsymbol{X})|T(\boldsymbol{X})) \\ &\quad - E_\theta[E_\theta(S(\boldsymbol{X})|T(\boldsymbol{X}))|T(\boldsymbol{X})] \\ &= 0\end{aligned}$$

y, en consecuencia,

$$\begin{aligned}E_\theta[(S(\boldsymbol{X})-T^*(\boldsymbol{X}))(T^*(\boldsymbol{X}) - q(\theta))] &= \int_R \int_{R^n} (S(\mathbf{x}) - T^*(\mathbf{x}))(T^*(\mathbf{x}) - q(\theta))f(\mathbf{x}, t)d\mathbf{x}dt \\ &= \int_R (T^*(\mathbf{x}) - q(\theta))\Big[\int_{R^n} (S(\mathbf{x}) - T^*(\mathbf{x}))f_1(\mathbf{x}|t)d\boldsymbol{x}\Big]f_2(t)dt \\ &= 0.\end{aligned}$$

La igualdad entre las varianzas de $T^*(\boldsymbol{X})$ y $S(\boldsymbol{X})$ se tiene si

$$E_\theta[E_\theta(S(\boldsymbol{X})|T(\boldsymbol{X})) - S(\boldsymbol{X})]^2 = 0,$$

es decir, si $P[E_\theta(S(\boldsymbol{X})|T(\boldsymbol{X})) = S(\boldsymbol{X})] = 1$ □.

Teorema 4.3 (teorema de Lehmann-Scheffé)**.** Si $T(\boldsymbol{X})$ es un estadístico completo y suficiente de θ y si $S(\boldsymbol{X})$ es un estimador insesgado de $q(\theta)$ de varianza finita, entonces, $T^*(\boldsymbol{X}) = E_\theta(S(\boldsymbol{X})|T(\boldsymbol{X}))$ es un estimador insesgado de mínima varianza de $q(\theta)$ y es único.

Demostración. Primero se demuestra la unicidad. Supóngase que $S_1(\boldsymbol{X})$ y $S_2(\boldsymbol{X})$ son dos estimadores insesgados de $q(\theta)$ y que $T(\boldsymbol{X})$ es un estadístico completo y suficiente de θ. Si $T_i^*(\boldsymbol{X}) = E_\theta(S_i(\boldsymbol{X})|T(\boldsymbol{X}))$, $i = 1, 2$, aplicando propiedades de la esperanza condicional se tiene que

$$\begin{aligned}E_\theta[T_1^*(\boldsymbol{X}) - T_2^*(\boldsymbol{X})] &= E_\theta[E_\theta(S_1(\boldsymbol{X})|T(\boldsymbol{X})) - E_\theta(S_2(\boldsymbol{X})|T(\boldsymbol{X}))] \\ &= E_\theta(S_1(\boldsymbol{X})) - E_\theta(S_2(\boldsymbol{X})) \\ &= 0.\end{aligned}$$

Dado que $T_1^* - T_2^*$ es una función de un estadístico completo T, $T_1^*(\boldsymbol{X}) = T_2^*(\boldsymbol{X})$. Finalmente, del teorema (4.2), se sigue que $T^*(\boldsymbol{X})$ es de mínima varianza.□

Como una consecuencia del teorema (4.3) se tiene que si T es un estadístico completo y suficiente de θ y $h(T)$ es un estimador insesgado de $q(\theta)$, entonces, $h(T)$ es un estimador UMVU de $q(\theta)$.

Ejemplo 4.5. En el ejemplo (4.3) se demuestra que si $X_i \sim U(0,\theta)$, $i = 1,2,\ldots,n$; son n variables aleatorias independientes, entonces, $X_{(n)}$ es completo y suficiente para θ.

Así pues, dado que

$$E_\theta\Big(\frac{n+1}{n}X_{(n)}\Big) = \frac{n+1}{n}\frac{n}{\theta^n}\int_0^\theta t^n dt = \theta,$$

del teorema (4.3) se sigue que $S(\boldsymbol{X}) = \frac{n+1}{n}X_{(n)}$ es un estadístico UMVU para θ, ya que $E_\theta(S(\boldsymbol{X})|X_{(n)} = x) = S(\mathbf{x})$.

En el ejemplo 4.5, se puede hallar un estimador UMVU de $g(\theta)$, donde g es una función diferenciable en $(0,\infty)$.

En efecto, dado que $X_{(n)}$ es un estadístico suficiente y completo de θ, basta determinar una función a valor real h tal que $E_\theta[h(X_{(n)})] = g(\theta)$. Para lo cual se debe encontrar h tal que

$$\begin{aligned} E_\theta[h(X_{(n)})] &= \int_0^\theta h(x) f_{X_{(n)}}(x)dx \\ &= n\theta^{-n}\int_0^\theta h(x)x^{n-1}dx, \quad \theta > 0, \end{aligned} \tag{4.5}$$

dado que la función de densidad de $X_{(n)}$ está dada por

$$f_{X_{(n)}}(x) = n\theta^{-n}x^{n-1}I_{[0,\theta]}(x).$$

De (4.5) resulta que $\theta^n g(\theta) = n\int_0^\theta h(x)x^{n-1}dx$ y derivando con respecto a θ se encuentra que $h(\theta) = g(\theta) + n^{-1}\theta g^{'}(\theta)$. En consecuencia, por el teorema 4.3, $h(X_{(n)}) = g(X_{(n)}) + n^{-1}X_{(n)}g^{'}(X_{(n)})$ es un estadístico UMVU de $g(\theta)$.

Ejemplo 4.6. Sean $X_i \sim U(0,\theta)$, $i = 1,2,\ldots,n$; n variables aleatorias independientes. Dado que $X_{(n)}$ es completo y suficiente para θ, por el resultado anterior $h(X_{(n)}) = (1+\frac{2}{n})X_{(n)}^2$ es un estadístico UMVU para θ^2.

Ejemplo 4.7. Sean $X_i \sim \text{Pois}(\theta)$, $i = 1,2,\ldots,n$; n variables aleatorias independientes. Dado que $T(X) = \sum_{i=1}^n X_i$ es un estadístico completo y suficiente para θ, por el teorema (4.3), un estadístico UMVU para $q(\theta) = e^{-\theta}$ está determinado por $T^*(\boldsymbol{X}) = E_\theta(S(\boldsymbol{X})|T(\boldsymbol{X}))$, donde $S(\boldsymbol{X}) = I_{\{X_1=0\}}$ es un estimador insesgado de $q(\theta)$.

Obsérvese que como

$$\begin{aligned} E_\theta\Big(I_{\{X_1=0\}}|\sum_{i=1}^n X_i = t\Big) &= P_\theta\Big(X_1 = 0|\sum_{i=1}^n X_i = t\Big) \\ &= \frac{P(X_1=0)P_\theta\Big(\sum_{i=2}^n X_i = t\Big)}{P_\theta\Big(\sum_{i=1}^n X_i = t\Big)} \\ &= \left(1-\frac{1}{n}\right)^t, \end{aligned}$$

entonces, $T^*(\boldsymbol{X}) = \left(1-\frac{1}{n}\right)^T$.

La última igualdad se obtiene teniendo en cuenta que la suma de las n variables aleatorias Poisson, de parámetro θ, es una variable aleatoria que tiene distribución Poisson, con parámetro $n\theta$. Es decir, si X_i, $i = 1, \ldots, n$; es una muestra aleatoria, de tamaño n, de una distribución Poisson con media θ, entonces,

$$P_\theta\Big(\sum_{i=1}^n X_i = t\Big) = \frac{e^{-n\theta}(n\theta)^t}{t!},$$

donde $t = 1, 2, \ldots$

Ejercicios 4.3.

1. Sea X_i, $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una distribución Poisson $P(\theta)$.

   a) Hallar $q(\theta) = P_\theta[X_1 = 0]$.

   b) ¿$T(\boldsymbol{X}) = \bar{X}$ es un estadístico completo y suficiente de $q(\theta)$?

   c) Hallar t, tal que $E\Big[\exp(t\sum_{i=1}^n X_i)\Big] = q(\theta)$.

   d) Para el t encontrado en 1*c*, es ¿$S(\boldsymbol{X}) = \exp(t\sum_{i=1}^n X_i)$ es un estimador UMVU de $q(\theta)$?

2. Sea X_i, $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una distribución Poisson $P(\theta)$.

a) ¿$T(\boldsymbol{X}) = I_{\{0\}}(X_1)$, donde I es la función indicadora, es un estimador insesgado de $q(\theta) = P_\theta[X_1 = 0]$?

b) Hallar $E[T(\boldsymbol{X})| \sum_{i=1}^n X_i = t]$.

c) Defina un estimador UMVU de $q(\theta)$ y compare sus resultados con los obtenidos en el primer ejercicio.

3. En el ejemplo 4.4 de la sección 4.2, demuestre que

   a) $T_1(\boldsymbol{X}) = \bar{X}$ es un estadístico UMVU de θ.

   b) $T_2(X) = \frac{1}{n}\sum_{i=1}^n (X_i - \theta)^2$ es un estadístico UMVU de σ^2.

4. En el ejercicio 5, de la sección 4.2, demuestre que $T(\boldsymbol{X}) = \bar{X}$ es un estadístico UMVU de θ.

5. Si $X_1, \ldots, X_n$ es una muestra aleatoria, de tamaño n, de $U(\theta_1, \theta_2)$, donde θ_1 y θ_2 son desconocidos, entonces,

   a) Halle $E(X_{(1)} + X_{(n)})$ y proponga un estimador $T(X)$ tal que $E(T(X)) = (\theta_1 + \theta_2)/2$.

   b) ¿Qué condición debe tener $T(X)$ para ser un estimador UMVU de $(\theta_1 + \theta_2)/2$?

6. Sea X_i, $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una población $\Gamma(p, \lambda)$, definida como en (1.23), donde p y λ son desconocidos. Encuentre un estimador UMVU para p/λ.

4.3.1 Desigualdad de información

En esta sección se estudian las funciones P_θ, pertenecientes a la familia de distribuciones $\{P_\theta : \theta \in \Theta\}$, que satisfacen las siguientes propiedades de regularidad:

a) El conjunto $A = \{\mathbf{x} \in \mathbb{R} : p_\theta(\mathbf{x}) > 0\}$ no depende de θ.

b) Para todo $\mathbf{x} \in A$ y todo $\theta \in \Theta$, $\frac{\partial}{\partial\theta} \log p_\theta(\mathbf{x})$ existe y es finito.

c) Si T es un estadístico tal que $E_\theta(|T|) < \infty$; para todo $\theta \in \Theta$, entonces,

$$\frac{\partial}{\partial\theta}\left[\int_{\mathbb{R}^n} T(\mathbf{x}) p_\theta(\mathbf{x}) d\mathbf{x}\right] = \int_{\mathbb{R}^n} T(\mathbf{x}) \frac{\partial}{\partial\theta} p_\theta(\mathbf{x}) d\mathbf{x}. \tag{4.6}$$

En los siguientes ejemplos se muestra que, para $T(X) = X$, las familias de distribuciones Exponencial y Normal, con varianza conocida, cumplen las propiedades dadas en los literales a, b y c. En estos casos es evidente que $E_\theta(|T(X)|) < \infty$.

Ejemplo 4.8. En la distribución Exponencial, el conjunto $A = \mathbb{R}^+$ no depende del parámetro. Para todo $x \in A$ y todo $\lambda \in \mathbb{R}^+$,

$$\begin{aligned}\frac{\partial}{\partial\lambda}\log p_\lambda(x) &= \frac{\partial}{\partial\lambda}(\log\lambda - \lambda x)\\ &= \frac{1}{\lambda} - x,\end{aligned}$$

existe y es finito.

Y, finalmente, dado que

$$\begin{aligned}\int_{\mathbb{R}^+} x\frac{\partial}{\partial\lambda}\lambda e^{-\lambda x}dx &= \int_{\mathbb{R}^+} x(e^{-\lambda x} - x\lambda e^{-\lambda x})dx\\ &= \frac{1}{\lambda}E(X) - Var(X) - E^2(X)\\ &= -\frac{1}{\lambda^2}\end{aligned}$$

se tiene que

$$\frac{\partial}{\partial\lambda}\int_{\mathbb{R}^+} x\lambda e^{-\lambda x}dx = -\frac{1}{\lambda^2}.$$

Así, la función de distribución Exponencial cumple con las propiedades de regularidad dadas en los literales a, b y c.

Ejemplo 4.9. En la distribución Normal, con media θ desconocida y varianza conocida, el conjunto $A = \mathbb{R}$ no depende de θ y, para todo $x \in A$ y todo $\theta \in \mathbb{R}$, se tiene que $\frac{\partial}{\partial\theta}\log p_\theta(x) = \frac{1}{\sigma^2}(x-\theta)$ existe y es finito.

Finalmente, dado que

$$\begin{aligned}\int_{\mathbb{R}} x\frac{\partial}{\partial\theta}p_\theta(x)dx &= \frac{1}{\sigma^2}[E(X^2) - \theta E(X)]\\ &= \frac{1}{\sigma^2}[Var(X) + E^2(X) - \theta E(X)]\\ &= 1\\ &= \frac{\partial}{\partial\theta}E(X).\end{aligned}$$

Así, la distribución Normal, con media desconocida y varianza conocida, cumple con las propiedades de regularidad dadas en los literales a, b y c.

Definición 4.1. Sea $X \sim P_\theta$ tal que las propiedades dadas en los literales a y b son válidas. Se define el número de información de Fisher, notado $I(\theta)$, por medio de la siguiente expresión:

$$I(\theta) = E_\theta \left[\frac{\partial}{\partial \theta} \log p\left(X, \theta\right) \right]^2 .$$

$I(\theta)$ se interpreta como la cantidad de información acerca de θ que contiene una observación x de X.

Asumiendo que $\{P_\theta : \theta \in \Theta\}$ satisface las propiedades dadas en los literales a, b y c algunos resultados de interés, asociados con el número de información de Fisher, son los siguientes:

1. $E_\theta \left[\frac{\partial}{\partial \theta} \log P(X, \theta) \right] = 0$ y, en consecuencia, $I(\theta) = Var_\theta[\frac{\partial}{\partial \theta} \log P(X, \theta)]$.

2. Si $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ es una muestra aleatoria, de tamaño n, de una población con función de densidad $p(x, \theta)$, entonces,

$$Var_\theta \left(\frac{\partial}{\partial \theta} \log p(\boldsymbol{X}, \theta) \right) = \sum_{i=1}^{n} Var_\theta \left[\frac{\partial}{\partial \theta} \log p(X_i, \theta) \right]$$

   y $I(\theta) = nI_1(\theta)$.

3. Supóngase que $I(\theta) < \infty$ y que $\frac{\partial^2}{\partial \theta^2} \log p(x, \theta)$ existe para todo $x \in A$ y para todo $\theta \in \Theta$. Si

$$\int_{\mathbb{R}_n} \frac{\partial^2}{\partial \theta^2} p(x, \theta) dx = \frac{\partial^2}{\partial \theta^2} \int_{\mathbb{R}_n} p(x, \theta) dx,$$

   entonces, $I(\theta) = -E_\theta \left[\frac{\partial^2}{\partial \theta^2} \log p(X, \theta) \right]$.

Observe que

$$\begin{aligned}
E_\theta\left[\frac{\partial^2}{\partial\theta^2}\log p(x,\theta)\right] &= \int_A \left[\frac{\partial^2}{\partial\theta^2}\log p(x,\theta)\right] p(x,\theta)dx \\
&= -\int_A \frac{1}{p(x,\theta)}\left(\frac{\partial}{\partial\theta}p(x,\theta)\right)^2 dx + \int_A \frac{\partial^2}{\partial\theta^2}p(x,\theta)dx \\
&= -E_\theta\left[\left(\frac{\partial}{\partial\theta}\log p(X,\theta)\right)^2\right] + \frac{\partial^2}{\partial\theta^2}\int_A p(x,\theta)dx \\
&= -E_\theta\left[\left(\frac{\partial}{\partial\theta}\log p(X,\theta)\right)^2\right] \\
&= -I(\theta).
\end{aligned}$$

Ejemplo 4.10. Si X tiene distribución Poisson, con parámetro θ, entonces:

$$\begin{aligned}
I(\theta) &= E_\theta\left[\frac{\partial}{\partial\theta}\log\left(\frac{\theta^X e^{-\theta}}{X!}\right)\right]^2 \\
&= E_\theta\left[\frac{\partial}{\partial\theta}(X\log(\theta)-\theta-\log(X!))\right]^2 \\
&= E\left[\left(\frac{X-\theta}{\theta}\right)^2\right] \\
&= \frac{1}{\theta^2}Var_\theta(X) \\
&= \frac{1}{\theta}.
\end{aligned}$$

Otra forma de calcular $I_1(\theta)$ es la siguiente:

$$\begin{aligned}
I(\theta) &= -E_\theta\left[\frac{\partial^2}{\partial\theta^2}(X\log(\theta)-\theta-\log(X!))\right] \\
&= \frac{1}{\theta^2}E_\theta(X) \\
&= \frac{1}{\theta}.
\end{aligned}$$

Ejemplo 4.11. Si X tiene una distribución Normal, con media conocida y varianza desconocida, entonces,

$$\begin{aligned} Var_{\sigma^2}\Big(\frac{\partial}{\partial\sigma^2}\log p(X,\sigma^2)\Big) &= Var_{\sigma^2}\Big(-\frac{1}{2\sigma^2}+\frac{1}{2\sigma^4}(X-\theta)^2\Big) \\ &= \frac{1}{4\sigma^4}Var\Big(\frac{X-\theta}{\sigma}\Big)^2. \\ &= \frac{1}{2\sigma^4} \end{aligned}$$

$$\begin{aligned} E_{\sigma^2}\Big(\frac{\partial^2}{\partial\sigma^{2^2}}\log p(X,\sigma^2)\Big) &= E_{\sigma^2}\Big(\frac{1}{2\sigma^4}-\frac{1}{\sigma^6}(X-\theta)^2\Big) \\ &= -\frac{1}{2\sigma^4} = -I(\sigma^2). \end{aligned}$$

El siguiente teorema establece una cota inferior para la varianza, denominada cota de Crámer-Rao. En él, X denota el vector aleatorio con n componentes, independientes e idénticamente distribuidas, e $I(\theta)$ la información de Fisher dada por X.

Teorema 4.4 (teorema de desigualdad de información)**.** Sea $T(\boldsymbol{X})$ un estadístico tal que $Var_\theta T(\boldsymbol{X}) < \infty$; para todo θ. Bajo las propiedades de regularidad a, b y c, si $0 < I(\theta) < \infty$, entonces,

$$Var_\theta[T(\boldsymbol{X})] \geq \frac{[\Psi'(\theta)]^2}{I(\theta)};$$

donde $\Psi(\theta) = E_\theta[T(\boldsymbol{X})]$ and $I(\theta) = E_\theta\left[\frac{\partial}{\partial\theta}\log P(X,\theta)\right]^2$.

Demostración. Dado que $E_\theta\left[\frac{\partial}{\partial\theta}\log P(\boldsymbol{X},\theta)\right] = 0$, entonces, aplicando la propiedad (4.6), se encuentra que

$$\begin{aligned} Cov_\theta\left[T(\boldsymbol{X}),\frac{\partial}{\partial\theta}\log P(\boldsymbol{X},\theta)\right] &= E_\theta\left[T(\boldsymbol{X})\frac{\partial}{\partial\theta}\log P(\boldsymbol{X},\theta)\right] \\ &= \int_{R^n} T(\mathbf{x})\frac{\partial}{\partial\theta}P(\mathbf{x},\theta)d\mathbf{x} \\ &= \frac{\partial}{\partial\theta}\int_{R^n} T(\mathbf{x})P(\mathbf{x},\theta)d\mathbf{x} \\ &= \frac{\partial}{\partial\theta}E_\theta[T(\boldsymbol{X})]. \end{aligned}$$

Así, a partir del concepto de correlación y teniendo en cuenta que $I(\theta) = Var_\theta\left[\frac{\partial}{\partial\theta}\log P(\boldsymbol{X},\theta)\right]$, se encuentra que

$$\begin{aligned} 1 &\geq \frac{\left|Cov_\theta\left[T(\boldsymbol{X}), \frac{\partial}{\partial\theta}\log P(\boldsymbol{X},\theta)\right]\right|}{\sqrt{Var_\theta[T(\boldsymbol{X})]Var_\theta\left[\frac{\partial}{\partial\theta}\log P(\boldsymbol{X},\theta)\right]}} \\ &= \frac{\left|\frac{\partial}{\partial\theta}E_\theta[T(\boldsymbol{X})]\right|}{\sqrt{Var_\theta[T(\boldsymbol{X})]I(\theta)}} \end{aligned}$$

Y, finalmente, haciendo las operaciones pertinentes, el teorema queda demostrado. □

Ejemplo 4.12. La familia normal de distribuciones, con varianza conocida y media desconocida, satisface las propiedades de regularidad dadas en los literales a, b y c. En consecuencia, si $T(\mathbf{X}) = \sum_{i=1}^n X_i$,

1. $\Psi(\theta) = E(T(\mathbf{X}) = n\theta$
2. $Var(T(\mathbf{X}) = n\sigma^2$
3. $\frac{\partial}{\partial\theta}\log P(\mathbf{X},\theta) = -\frac{1}{\sigma^2}\sum_{i=1}^n(x_i-\theta)$
4. $E\Big(\frac{\partial}{\partial\theta}\log P(\mathbf{X},\theta)\Big) = -E\Big(\frac{1}{\sigma^2}\sum_{i=1}^n(x_i-\theta)\Big) = 0$
5. $Var\Big(\frac{\partial}{\partial\theta}\log P(\mathbf{X},\theta)\Big) = Var\Big(-\frac{1}{\sigma^2}\sum_{i=1}^n(x_i-\theta)\Big) = \frac{n}{\sigma^2}$

Finalmente, de 1 y 5 se obtiene:

$$\frac{[\Psi'(\theta)]^2}{I(\theta)} = n\sigma^2$$

y, por lo tanto, se concluye que $T(\mathbf{X})$ es un estimador de mínima varianza de $n\theta$.

Ejercicios 4.4.

1. Para $T(X) = X$, determine si las familias de distribuciones Binomial, Geométrica, Gamma (con media desconocida y parámetro de forma conocido) y Beta (con media desconocida y parámetro de "precisión" conocido) cumplen con las propiedades de regularidad dadas en los literales *a*, *b* y *c*.

2. Sea X_i, $i = 1, 2, \ldots, n$; n variables aleatorias independientes con función de densidad $f_\theta(x) = \frac{1}{\theta} e^{-\frac{1}{\theta}x} I_{(0,\infty)}(x)$, $\theta > 0$. Calcule la información de Fisher $I(\theta)$.

3. En el ejercicio anterior, calcule la información de Fisher $I(\lambda)$, para $\lambda = 1/\theta$ e interprete los resultados.

4. Desarrolle el segundo ejercicio asumiendo la distribución Gamma:

   a) Con parámetros p y λ, como en (1.23), con p desconocido y λ conocido.

   b) Reparametrizada, como en el ejercicio 5 de la sección 4.2.

5. Si $X_i \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$; son n variables aleatorias independientes, encuentre un estimador insesgado de θ^2. Determine $Var(\bar{X}^2)$. Halle la cota de Crámer-Rao, $1/I(\theta)$.

6. Si $X \sim \text{Pois}(\theta)$, $Y = I_{\{0\}}(X)$ tiene distribución $\text{Ber}(e^{-\theta})$. Verifique la desigualdad de información considerando θ como parámetro y $T(X) = I_{\{0\}}(X)$ como estimador de $e^{-\theta}$.

7. Sea X_i, $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una población P_θ. Halle la cota de Crámer-Rao para la varianza del estimador de la media dado por:

   a) $T(X) = \bar{X}$, asumiendo que $X_i \sim \text{Exp}(\theta)$.

   b) $T(X) = X_1$, asumiendo que $X_i \sim N(\mu, \sigma^2/n)$, σ^2 conocido.

   c) $T(X) = \bar{X}$, asumiendo que $X_i \sim G(p, \lambda)$, p conocido.

8. Demuestre el resultado enunciado en el numeral 1 de la subsección 4.3.1.

Capítulo 5

Intervalos de confianza

Como una introducción a los intervalos de confianza, para los parámetros de una distribución P_θ, se considera el caso de una variable aleatoria X con distribución uniforme en el intervalo $(0, \theta)$. Si $X \sim U(0, \theta)$, entonces, la variable aleatoria $Y = \frac{X}{\theta}$ tiene distribución uniforme en el intervalo $(0, 1)$. Así, si α es un número real positivo menor que 1, entonces,

$$P\left(\frac{\alpha}{2} \leq \frac{X}{\theta} \leq 1 - \alpha/2\right) = P\left(\frac{X}{1-\alpha/2} \leq \theta \leq \frac{X}{\alpha/2}\right) = 1 - \alpha.$$

Así, se encuentra un límite inferior $\underline{L} = \frac{X}{1-\alpha/2}$ y un límite superior $\bar{L} = \frac{X}{\alpha/2}$, que determinan un intervalo con probabilidad $1-\alpha$ de contener el parámetro θ, con un nivel de confianza de $1 - \alpha$.

A este intervalo, notado

$$\left[\frac{X}{1-\alpha/2}, \frac{X}{\alpha/2}\right], \tag{5.1}$$

se le denomina intervalo de confianza de θ del $(1 - \alpha)100\,\%$. Este es un intervalo aleatorio y para cada valor x, de la variable X, le corresponde un intervalo $[\frac{x}{1-\alpha/2}, \frac{x}{\alpha/2}]$, el cual no necesariamente contiene a θ.

Ejemplo 5.1. Sea X una variable aleatoria con distribución $U(0, \theta)$ y $x = 9.8$ un valor observado de esta. Obsérvese, por ejemplo, que si $\theta = 10$ y $\alpha = 0.05$, entonces, el intervalo de confianza observado es $(10.05, 392)$ y este no contiene a θ.

Asumiendo $\alpha = 0.2$, si x es uno de los valores posibles de X y $x < 1$, el intervalo de confianza dado en (5.1) no contiene a θ pues $\frac{x}{\alpha/2} < 10$. De igual manera, si $x > 9$, el intervalo en referencia no contiene a θ, pues $\frac{x}{1-\alpha/2} > 10$. Así, $\theta \in [\frac{x}{1-\alpha/2}, \frac{x}{\alpha/2}]$, si y solo si $1 \leq x \leq 9$. Observe que si $\alpha = 0.2$,

$$P\left[\frac{X}{1-\alpha/2} \leq \theta \leq \frac{X}{\alpha/2}\right] = P\left[\frac{10}{9}X \leq \theta \leq 10X\right] = 0.8. \tag{5.2}$$

Este resultado puede generalizarse para otros valores de α y otros valores de θ.

5.1 Intervalos de confianza para la media de una distribución Normal

En esta sección se consideran intervalos de confianza para la media de una distribución Normal con varianza conocida, en la sección (5.1.1), y con varianza desconocida, en la sección (5.1.2).

5.1.1 Intervalos de confianza para la media de una distribución Normal, con varianza conocida

Sean Y_i, $i = 1, \ldots, n$; n variables aleatorias independientes normalmente distribuidas, con media desconocida θ y varianza conocida σ^2. Si $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$, entonces,

$$Z = \frac{\sqrt{n}(\bar{Y} - \theta)}{\sigma}$$

tiene distribución Normal estándar.

Así, para todo número real $\alpha \in (0, 1)$, existe un número real positivo $z_{\alpha/2}$ tal que

$$P\left(-z_{\frac{\alpha}{2}} \leq \frac{\sqrt{n}(\bar{Y} - \theta)}{\sigma} \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha.$$

En consecuencia,

$$P\left(\bar{Y} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \theta \leq \bar{Y} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha, \tag{5.3}$$

donde $\bar{Y} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ y $\bar{Y} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ determinan un intervalo de confianza para θ. En este intervalo, los estadísticos $\underline{L} = \bar{Y} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ y $\bar{L} = \bar{Y} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ se denominan límite inferior y límite superior del intervalo, respectivamente.

Estos límites, $\underline{L}$ y $\bar{L}$, son variables aleatorias que determinan un intervalo que tiene probabilidad $1-\alpha$ de contener el parámetro θ. A la cantidad $1-\alpha$ se denomina coeficiente de confianza y es la probabilidad de que el intervalo de confianza $[\underline{L}, \bar{L}]$ contenga a θ.

Si $\bar{y}$ es una observación de $\bar{Y}$, el intervalo $[\underline{L}(\bar{y}), \bar{L}(\bar{y})]$ puede contener o no a θ. En este caso, $1-\alpha$ corresponde a la incertidumbre que se tiene sobre la pertenencia de θ al intervalo $[\underline{L}(\bar{y}), \bar{L}(\bar{y})]$ y no debe interpretarse como la probabilidad de que θ pertenezca a él. Además, $1-\alpha$ también puede interpretarse como la probabilidad de que al observar $\bar{Y}$, el error de observación sea menor que $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. Obsérvese que el error de observación y la longitud del intervalo pueden hacerse tan pequeñas como se quiera, siempre que n sea suficientemente grande.

Nota: $\underline{L}(Y) = \bar{Y} - z_{\alpha}\frac{\sigma}{\sqrt{n}}$ se llama una cota inferior de nivel α para θ. Obsérvese que $P(\underline{L}(Y) \leq \theta) = 1-\alpha$. De igual forma, se puede definir una cota superior $\bar{L}(Y)$ como $\bar{Y} + z_{\alpha}\frac{\sigma}{\sqrt{n}}$. En este caso, $P(\overline{L}(Y) \geq \theta) = 1-\alpha$.

En esta sección, se considera el intervalo de confianza determinado por $\underline{L}$ y $\bar{L}$, dado en la ecuación (5.3), ya que este es el intervalo más pequeño de nivel de confianza $1-\alpha$; $0 < \alpha < 0.5$. Este resultado se demuestra en la siguiente proposición.

Proposición 5.1. Sea $X_i \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una distribución Normal con varianza conocida. El intervalo de confianza determinado por $\underline{L}$ y $\bar{L}$, dado en la ecuación (5.3), es el intervalo más pequeño de nivel de confianza $1-\alpha$, $0 < \alpha < 0.5$, para θ.

Demostración 1. Para demostrar esta proposición, se expresa la longitud del intervalo en función de una variable t, con $0 < t < 1$. Así, dado que, en este caso, todo intervalo de nivel $1-\alpha$ puede expresarse de forma

$$\left(\bar{X} - z_{(1-t)\alpha}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{t\alpha}\frac{\sigma}{\sqrt{n}}\right), \tag{5.4}$$

su longitud está determinada por

$$h(t) \quad = z_{t\alpha} + z_{(1-t)\alpha} = \Phi^{-1}(1-t\alpha) + \Phi^{-1}(1-(1-t)\alpha), \tag{5.5}$$

donde

$$\Phi(t) = \frac{1}{2\pi}\int_{-\infty}^{t} e^{-\frac{1}{2}x^2}dx$$

y Φ^{-1} es la función inversa de Φ.

Ahora bien, derivando (5.5) con respecto a t, se tiene:

$$h'(t) = -[\Phi^{-1}(1-t\alpha)]'\alpha + [\Phi^{-1}(1-(1-t)\alpha)]'\alpha.$$

Por lo tanto, $h(t)$ alcanza su valor mínimo en t tal que

$$[\Phi^{-1}(1-t\alpha)]' - [\Phi^{-1}(1-(1-t)\alpha)]' = 0.$$

Dado que $[\Phi^{-1}(x)]'$ es una función monótona decreciente, para $x < 0.5$, y monótona creciente, para $x > 0.5$, la anterior igualdad se cumple únicamente para $t\alpha = (1-t)\alpha$, es decir, para $t = \frac{1}{2}$.$\square$

A continuación se presenta otra forma de demostrar este resultado.

Demostración 2. Dado que la función de densidad de la distribución Normal es simétrica, con respecto a cero, el problema consiste en minimizar $z_1 + z_2$, sujeto a la restricción $\int_0^{z_1} f(x)dx + \int_0^{z_2} f(x)dx = 1-\alpha$, donde f es la función de densidad de la distribución Normal estándar. Esto equivale a minimizar

$$H(z_1, z_2, \lambda) = z_1 + z_2 + \lambda\left(\int_0^{z_1} f(x)dx + \int_0^{z_2} f(x)dx - 1 + \alpha\right).$$

Dado que f es continua, derivando con respecto a z_1 y a z_2 y aplicando el teorema fundamental del cálculo (Apostol, 1967), se obtiene que

$$H'_{z_1}(z_1, z_2, \lambda) = 1 + \lambda f(z_1) \quad \text{y} \quad H'_{z_2}(z_1, z_2, \lambda) = 1 + \lambda f(z_2).$$

Ahora, igualando a cero estas derivadas se concluye que $f(z_1) = f(z_2)$ y, por tanto, $z_1 = z_2$. A partir de la restricción dada anteriormente, se encuentra que

$$2\int_0^{z_1} f(x)dx = 1-\alpha \quad \text{y} \quad \int_{z_1}^{\infty} f(x)dx = \frac{\alpha}{2}.$$

Así pues, $z_1 = z_2 = z_{\alpha/2}$.□

Este resultado responde a la pregunta siguiente: ¿Por qué, con un nivel de confianza del $100(1-\alpha)\,\%$, se debe considerar como intervalo de confianza para la media de una distribución Normal, el intervalo dado por:

$$\left[\bar{Y} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{Y} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$$

y no, por ejemplo, el intervalo:

$$\left[\bar{Y} - z_{\frac{1}{3}\alpha}\frac{\sigma}{\sqrt{n}}, \bar{Y} + z_{\frac{2}{3}\alpha}\frac{\sigma}{\sqrt{n}}\right]?$$

Como un ejemplo de estimación de intervalos de confianza, cuando la varianza es conocida, supóngase que una muestra aleatoria, de tamaño $n = 16$, se obtiene de una población normal con varianza $\sigma^2 = 4$. Si la media muestral es $\bar{y} = 11.5$, el intervalo de confianza del $95\,\%$, para la media, es

$$\left[11.5 - 1.96\frac{2}{\sqrt{16}}, 11.5 + 1.96\frac{2}{\sqrt{16}}\right] = [10.52, 12.48].$$

Ejercicios 5.1.

1. Suponga que $P(\bar{X} < 5) = 0.95$ y $P(\bar{X} > 3) = 0.90$. ¿Cuál es la probabilidad de que $\bar{X}$ se encuentre entre 3 y 5?

2. Se toman N muestras aleatorias independientes de tamaño n, de una población normal con media desconocida θ y varianza conocida σ^2. Si $\bar{y}_i$ denota la media de la i-ésima muestra, con relación a los intervalos de confianza $I_i = \bar{y}_i \pm z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, i = 1, 2, 3, \ldots, N$, ¿cuáles de las siguientes afirmaciones son verdaderas? Justifique la respuesta.

   a) Cada intervalo contiene a θ.

   b) El $100(1-\alpha)\,\%$ de los intervalos contiene a θ.

c) El $100\alpha\,\%$ de los intervalos no contiene a θ.

d) Todos los intervalos son iguales.

e) Todos los intervalos tienen la misma longitud.

3. La varianza de la resistencia de ciertos componentes es $\sigma^2 = 0.16$. Si las medidas (en ohms) de una muestra aleatoria de 10 unidades son 5.3, 5, 5.2, 4.9, 4.8, 5.1, 5.3, 4.7, 5.2, y 4.7, halle intervalos de confianza del 90 % y del 95 % para la resistencia media; asumiendo que las observaciones provienen de una distribución Normal.

4. Si X es una variable con función de densidad $f_\theta(x) = \frac{1}{\theta}e^{-\frac{1}{\theta}x}I_{(0,\infty)}(x)$; $\theta > 0$, halle los límites inferior y superior $\underline{L}(X)$ y $\bar{L}(X)$ tales que $P_\theta(\underline{L}(X) \leq \theta) = P_\theta(\bar{L}(X) \geq \theta) = 1 - \alpha$, para $\alpha < 0.5$.

5. Sean X_i, $i = 1, \ldots, n$; n variables aleatorias independientes con función de densidad $f_\theta(x) = \frac{1}{\theta}\, e^{-\frac{1}{\theta}x}I_{(0,\infty)}(x)$, $\theta > 0$, halle los límites aleatorios $\underline{L}(X)$ y $\bar{L}(X)$, donde $X = (X_1, \ldots, X_n)$, tales que $P_\theta(\underline{L}(X) \leq \theta) = P_\theta(\bar{L}(X) \geq \theta) = \alpha$, para $\alpha < 0.5$.

6. Sean $X_i \sim U(0, \theta), i = 1, 2, \ldots, n$; n variables aleatorias independientes. Halle un intervalo de confianza para θ, en función de $X_{(n)}$.

7. Sean $X_i, i = 1, 2, \ldots, n$; n variables aleatorias independientes con función de densidad $f_\theta(x) = \frac{1}{\theta}e^{-\frac{1}{\theta}x}$, $x > 0$, $\theta > 0$.

   a) Construya un intervalo de confianza para θ.

      1) A partir de la distribución Ji-cuadrado.

      2) Aplicando el teorema del límite central.

   b) Halle intervalos de confianza para $e^{-\frac{1}{2}\theta}$ y para $\lambda = \frac{1}{\theta}$.

8. Sean $X_i, i = 1, 2, \ldots, n$; n variables aleatorias independientes normalmente distribuidas $N(\theta, \theta^2)$. Determine un intervalo de confianza para θ usando la distribución Normal y Ji-cuadrado. Compare sus resultados.

9. Sean $X_i, i = 1, 2, \ldots, n$; n variables aleatorias independientes normalmente distribuidas $N(\theta, \theta)$, $\theta > 0$. Determine un intervalo de confianza para θ usando la distribución

   a) Normal.

   b) Ji-cuadrado.

5.1.2 Intervalos de confianza para la media de una distribución Normal, con varianza desconocida

Sea Y_i, $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, seleccionada de una población Normal con media θ y varianza σ^2, desconocidas. Para determinar un intervalo de confianza para θ, es razonable reemplazar σ por el valor de la estadística:

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

Dado que $(n-1)S^2/\sigma^2$ tiene distribución $\chi^2_{(n-1)}$, a partir de la definición de la distribución t se tiene que

$$T = \frac{\bar{Y} - \theta}{S/\sqrt{n}}, \tag{5.6}$$

tiene distribución τ con $n-1$ grados de libertad.

En consecuencia, existe un número real positivo $t_{n-1,\frac{\alpha}{2}}$ tal que $P(-t_{n-1,\frac{\alpha}{2}} \leq T \leq t_{n-1,\frac{\alpha}{2}}) = 1 - \alpha$, y así,

$$P\left(\bar{Y} - t_{n-1,\frac{\alpha}{2}}\left(\frac{S}{\sqrt{n}}\right) \leq \theta \leq \bar{Y} + t_{n-1,\frac{\alpha}{2}}\left(\frac{S}{\sqrt{n}}\right)\right) = 1 - \alpha.$$

En este caso,

$$\underline{L}(Y) = \bar{Y} - t_{n-1,\frac{\alpha}{2}}\left(\frac{S}{\sqrt{n}}\right)$$

y

$$\bar{L}(Y) = \bar{Y} + t_{n-1,\frac{\alpha}{2}}\left(\frac{S}{\sqrt{n}}\right)$$

son los límites inferior y superior, respectivamente, del intervalo de confianza que tiene probabilidad $1 - \alpha$ de contener a θ. Para determinar $t_{n-1,\frac{\alpha}{2}}$ en forma aproximada, se utiliza la tabla de distribución t. Por ejemplo, si $n = 9$ y $\alpha = 0.05$, entonces, $t_{8,\,0.025} = 2.306$.

Dado que la variable aleatoria T, definida en la ecuación (5.6) depende del tamaño de la muestra, se puede definir una sucesión de variables aleatorias, $T_2, T_3, \ldots, T_n, \ldots$, cuyo n-ésimo elemento está dado por la expresión (5.6). Esta sucesión de variables aleatorias converge en distribución a una variable normal estándar y, en consecuencia, esta distribución puede utilizarse para encontrar los intervalos de confianza cuando n sea "grande".

Ejercicios 5.2.

1. Una empresa desea que el diámetro de sus tornillos sea 0.53 cm. Al tomar una muestra aleatoria de su producción se encuentran tornillos de 0.51, 0.48, 0.53, 0.49, 0.52, 0.49 y 0.50 cm de diámetro. A partir de esta muestra y asumiento normalidad, ¿qué se puede afirmar con un nivel de confianza del 98 %?

2. Al medir la cantidad de azúcar contenida en cada uno de los ocho sobres de una muestra aleatoria obtenida de la producción de una empresa empacadora, se tuvo los siguientes pesos (en gramos): 10.2, 11.1, 9.8, 9.7, 10.5, 10.4, 10.1 y 11.2. Haga una estimación del contenido promedio de los sobres de azúcar, con una nivel de confianza del 98 %. Si el contenido deseado es 11.1 gramos, ¿qué se puede decir del proceso de empacado?

3. Una empresa afirma que las bombillas que produce tienen una duración media de 2000 horas. Al tomar una muestra aleatoria de 100 bombillas se encuentra una media muestral de $\bar{X} = 2015$ horas y una varianza $s^2 = 260$. Asumiendo que la variable "duración media de las bombillas" tiene distribución Normal, ¿qué afirmaciones se pueden hacer respecto de esta producción?

5.2 Intervalos de confianza aproximados para la probabilidad de éxito

Sea Y_i, $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una distribución Bernoulli con probabilidad de éxito θ. Dado que para n "grande"

$$Z = \frac{\sqrt{n}(\bar{Y} - \theta)}{\sqrt{\theta(1-\theta)}} \tag{5.7}$$

tiene, aproximadamente, distribución Normal estándar (véase teorema 1.3), entonces,

$$\begin{aligned} P\left(\left|\frac{\sqrt{n}(\bar{Y}-\theta)}{\sqrt{\theta(1-\theta)}}\right| \leq z_{\alpha/2}\right) &= P\left(n\left(\bar{Y}-\theta\right)^2 \leq z_{\alpha/2}^2\theta(1-\theta)\right) \\ &= P\left(\theta^2(n+z_{\alpha/2}^2)-\theta(2n\bar{Y}+z_{\alpha/2}^2)+n\bar{Y}^2 \leq 0\right) \\ &= 1-\alpha, \end{aligned}$$

y resolviendo, para θ, la desigualdad $\theta^2(n+z_{\alpha/2}^2)-\theta(2n\bar{Y}+z_{\alpha/2}^2)+n\bar{Y}^2 < 0$, se obtiene:

$$\bar{L}(Y) = \frac{(2n\bar{Y}+z_{\alpha/2}^2)+z_{\alpha/2}\sqrt{4n\bar{Y}(1-\bar{Y})+z_{\alpha/2}^2}}{2(n+z_{\alpha/2}^2)}. \tag{5.8}$$

$$\underline{L}(Y) = \frac{(2n\bar{Y}+z_{\alpha/2}^2)-z_{\alpha/2}\sqrt{4n\bar{Y}(1-\bar{Y})+z_{\alpha/2}^2}}{2(n+z_{\alpha/2}^2)}. \tag{5.9}$$

Dado que el coeficiente de θ^2 es positivo, la desigualdad se cumple para $\underline{L} \leq \theta \leq \bar{L}$; por lo tanto, $[\underline{L}\,,\bar{L}]$ es un intervalo de confianza aproximado de nivel $1-\alpha$ para θ.

Ahora, si n es pequeño, se debe utilizar intervalos de confianza exactos, obtenidos a partir de la distribución Binomial. Agresti & Coull (1998) muestran que el intervalo de confianza, definido por las ecuaciones (5.8) y (5.9), puede ser recomendado para casi todos los tamaños de muestra y valores de θ. Este intervalo se conoce como intervalo *Score* y fue propuesto por Wilson (1927).

Otra forma común de hallar intervalos de confianza para θ, es utilizar $\hat{\theta} = \bar{Y}$ para estimar la varianza. Entonces, dado que para valores "grandes" de n, la variable aleatoria Z, definida en la sección (5.7), tiene aproximadamente distribución Normal estándar, se tiene que:

$$\begin{aligned} &P\left(-z_{\alpha/2} \leq \frac{\sqrt{n}(\bar{Y}-\theta)}{\sqrt{\theta(1-\theta)}} \leq z_{\alpha/2}\right) = \\ &\qquad P\left(\bar{Y}-z_{\alpha/2}\sqrt{\frac{\theta(1-\theta)}{n}} \leq \theta \leq \bar{Y}+z_{\alpha/2}\sqrt{\frac{\theta(1-\theta)}{n}}\right) \approx 1-\alpha. \end{aligned}$$

Así, reemplazando θ por $\bar{Y}$, para estimar la varianza, se encuentra que el intervalo de confianza del $100(1-\alpha)\,\%$ para θ es

$$\bar{Y} - z_{\alpha/2}\sqrt{\frac{\bar{Y}(1-\bar{Y})}{n}} < \theta < \bar{Y} + z_{\alpha/2}\sqrt{\frac{\bar{Y}(1-\bar{Y})}{n}}. \tag{5.10}$$

A pesar de lo simple que resulta el cálculo de este intervalo, llamado intevalo de Wald, y de su gran difusión, muchos estudios muestran que tiene un mal desempeño (Agresti & Coull, 1998; Brown, Cai & DasGupta, 2002; Newcombe & Merino, 2006). Agresti & Coull (1998) muestran que las estimaciones dadas por (5.10) son especialmente inadecuadas, dado que para valores de θ cercanos a 0 o a 1, la distribución Binomial es altamente sesgada y las estimaciones de los intervalos de confianza esperados, definidas por (5.10), tienen a $\hat{\theta}$ como punto medio (véase, por ejemplo, Cepeda-Cuervo *et al.*, 2008).

La figura 5.1 muestra diferencias entre estimaciones de intervalos esperados de Score y de Wald, a medida que aumenta el tamaño de muestra. En la parte superior de esta figura se representa intervalos de confianza para $\theta = 0.1$, cuando el tamaño de la muestra varía entre 1 y 30 y entre 100 y 130. Las estimaciones del intervalo de Score presentan un desplazamiento hacia 1/2, lo que concuerda con el sesgo de la distribución Binomial para valores "pequeños del parámetro". En la parte inferior de la figura 5.1 se muestra el comportamiento de estos intervalos para $\theta = 0.8$, asumiendo los mismos tamaños de muestra. En este caso, el intervalo de Score también presenta un desplazamiento hacia $\theta = 1/2$, lo que concuerda con el sesgo de la distribución Binomial para valores "grandes"del parámetro. El intervalo de Wald es simétrico con respecto al parámetro, en el primer caso, centrado en $\theta = 0.1$ y, en el segundo, centrado en $\theta = 0.8$.

La figura 5.2 muestra estimaciones de los intervalos esperados de Score y de Wald. En cada figura, el eje de las abscisas corresponde al espacio parámetrico y el eje de las ordenadas a los límites del intervalo. Para cada uno de los valores de θ, los límites del intervalo de confianza están determinados por los puntos de intersección entre la línea vertical que pasa por el punto $(0, \theta)$ y las curvas superior e inferior, correspondientes. Se observa en esta figura, grandes diferencias entre los intervalos esperados de Score y de Wald, para valores pequeños del tamaño de la muestra. Estas diferencias se hacen menores a medida que el tamaño de la muestra aumenta.

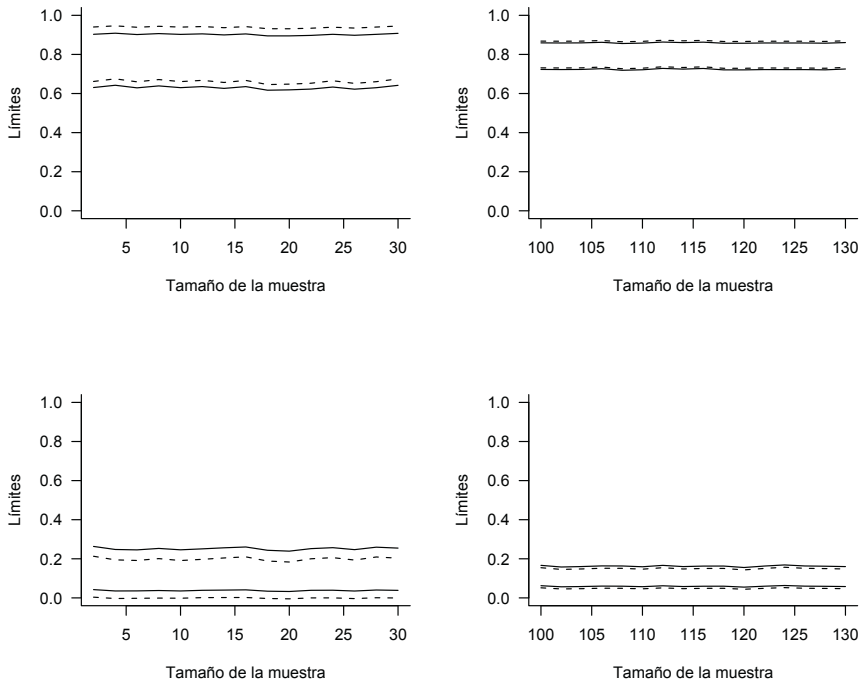

Figura 5.1. Intervalos de confianza esperados. Intervalo de Score (línea continua). Intervalo de Wald (línea segmentada).

Ejemplo 5.2. En una muestra aleatoria de 400 estudiantes de una universidad colombiana, 80 son fumadores. El intervalo de confianza del 95 % para la verdadera proporción de fumadores, de esta universidad, es $0.161 < \theta < 0.239$. Este resultado se obtiene al reemplazar, en (5.10), $\hat{\theta}$ y $\bar{Y}$ por 0.2. Cuando se utilizan las ecuaciones (5.8) y (5.9) para calcular los límites del intervalo de confianza del 95 % para θ, se encuentra que $0.163 < \theta < 0.242$.

Agresti & Coull (1998) presentan una nueva propuesta para obtener intervalos de confianza para una proporción, basada en una corrección del intervalo de confianza de Wald. En esta propuesta si $X = \sum_{i=1}^{n} X_i \sim Bin(n, \theta)$, entonces, el intervalo de confianza para θ está dado por

$$\tilde{\theta} \pm z_{\alpha/2}\sqrt{\frac{\tilde{\theta}(1-\tilde{\theta})}{n+4}}, \tag{5.11}$$

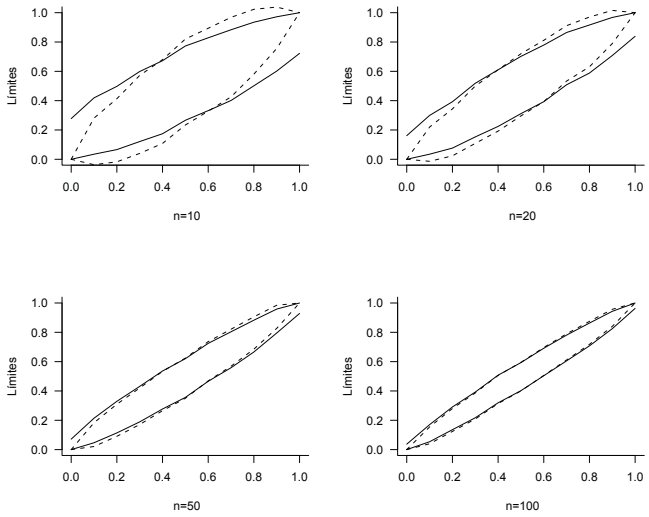

Figura 5.2. Estimaciones de intervalos de confianza esperados para una proporción. Intervalo de Score (línea continua). Intervalo de Wald (línea segmentada).

donde $\tilde{\theta} = (X+2)/(n+4)$. La expresión (5.11), conocida como intervalo de Wald ajustado, resulta bastante apropiada para la estimación de los intervalos de confianza (Wei, 2002).

5.3 Otros intervalos para una proporción

Sea $X \sim B(n,p)$ una variable aleatoria con distribución Binomial. El intervalo *logit* para el parámetro p está definido por

$$IC_{Logit} = \Big[\frac{e^{\lambda_I}}{1+e^{\lambda_I}}, \frac{e^{\lambda_S}}{1+e^{\lambda_S}}\Big],$$

con $\lambda_I = \hat{\lambda} - z_{\alpha/2}\hat{V}^{\frac{1}{2}}$, $\lambda_s = \hat{\lambda} + z_{\alpha/2}\hat{V}^{\frac{1}{2}}$, donde

$$\begin{aligned} \hat{\lambda} &= log\Big(\frac{\hat{p}}{1-\hat{p}}\Big) = log\Big(\frac{X}{n-X}\Big). \\ \hat{V} &= Var(\hat{\lambda}) = \frac{n}{X(n-X)}. \end{aligned}$$

El intervalo *logit*, IC_{Logit}, se obtiene invirtiendo el intervalo de confianza IC_λ del logaritmo de la razón $p/(1-p)$, $\lambda = log\left(\frac{p}{1-p}\right)$, definido por:

$$IC_\lambda = [\hat{\lambda} - z_{\alpha/2}\hat{V}^{\frac{1}{2}}, \hat{\lambda} + z_{\alpha/2}\hat{V}^{\frac{1}{2}}]$$

Este intervalo se comporta muy bien en términos de cobertura de p lejos de 0 y de 1, pero su longitud esperada es aún mayor que la longitud del intervalo de Clopper-Pearson (Stone, 1995). Para más información, relacionada con intervalos de confianza para una proporción, pueden verse Brown *et al.* (2002) o Cepeda-Cuervo *et al.* (2008).

Otros intervalos de confianza para una proporción también aparecen en la literatura estadística. Por ejemplo, el intervalo basado en la verosimilitud relativa. Brown *et al.* (2002) muestran que este intervalo tiene buenas propiedades. Sin embargo, no es fácil de calcular.

Ejercicios 5.3.

1. En Colombia, de una muestra aleatoria de 500 estudiantes que terminaron ingeniería en el año 2015, 50 iniciaron estudios de posgrado. Asumiendo un modelo binomial halle intervalos de confianza (Score, Wald, Wald ajustado y logit) del 95 % para la proporción de estudiantes de ingeniería que iniciaron estudios de posgrado.

2. Genere 10 muestras aleatoria, de tamaño 20, de una distribución $Ber(0.2)$. Halle los intervalos de confianza (Score, Wald, Wald ajustado y logit) del 95 % y determine, en cada caso, el número de intervalos que contienen a 0.2. Repita el ejercicio con $\theta = 0.45$.

3. Genere 10 muestras aleatoria, de tamaño 1, de una distribución Binomial $B(50, 0.2)$. Halle los intervalos de confianza (Score, Wald, Wald ajustado y logit) del 95 % y determine, en cada caso, el número de intervalos que contienen a 0.2. Repita el ejercicio con $\theta = 0.55$.

5.4 Probabilidad de cobertura

Al evaluar intervalos de confianza existen dos conceptos importantes para determinar cuáles métodos son más eficaces: la longitud del intervalo, $LI =$

$\overline{L} - \underline{L}$, que indica su precisión, y la probabilidad de cobertura (PC), definida como $P(\underline{L} < \theta < \overline{L})$, donde $\underline{L}$ y $\overline{L}$ son las variables aleatorias que indican el límite inferior y el límite superior del intervalo de confianza.

Se debe tener presente que tanto la probabilidad de cobertura como la longitud del intervalo de confianza son funciones del número de éxitos observados en la muestra, a través de los límites inferior y superior del intervalo. En la distribución binomial $B(n, \theta)$, dado α, la probabilidad de cobertura para un valor fijo θ es

$$PC = \sum_{x=0}^{n} I(x, \theta) P_\theta(x), \tag{5.12}$$

con $I(x, \theta)$ igual a 1 si el intervalo contiene a θ cuando $X = x$, e igual a 0 si el intervalo no contiene a θ. La probabilidad de cobertura hace referencia al cociente entre el número de intervalos que contienen el verdadero valor del parámetro y el número de intervalos estimados, cuando el número de estos últimos tiende a infinito.

En la tabla 5.1 se ilustra la forma en que se determina la probabilidad de cobertura del intervalo de Wald del 95 % de confianza, cuando los datos provienen de una distribución $Bin(6, 0.1)$. En ella, las columnas corresponden al número de éxitos observados x; a la probabilidad de cada uno de los posibles valores de X, $p_\theta(x)$; a los límites $\underline{L}$ y $\overline{L}$ del intervalo; a los valores de la función indicadora, $I(x, \theta)$, y al producto $I(x, \theta)p_\theta(x)$. Sumando las cantidades de la última columna se encuentra que, si el tamaño de la muestra es 6 y la probabilidad de éxito es 0.1, la probabilidad de cobertura es de 0.467. Esta tabla muestra que, en este ejemplo, el intervalo de confianza contiene al parámetro $\theta = 0.1$ únicamente si x es igual a 1, a 2 o a 3.

En otros ejemplos, como el ilustrado en la tabla 5.1, se encuentra que si los datos provienen de una distribución Binomial $Bin(6, 0.3)$, la probabilidad de cobertura es de 0.871 y que si los datos provienen de una distribución $Bin(6, 0.4)$, la probabilidad de cobertura es de 0.912. En realidad, la probabilidad de cobertura de un intervalo de confianza, para una proporción, depende tanto del valor del parámetro como del tamaño de la muestra considerada. En general, como en todos los casos considerados, la probabilidad de cobertura del intervalo de Wald siempre es menor que el nivel de confianza $100(1 - \alpha)$ % utilizado para la estimación del intervalo, especialmente cuando el tamaño de la muestra es pequeño o cuando la probabilidad de éxito es próxima a 0 o a 1.

x	$p_\theta(x)$	$\underline{L}$	$\overline{L}$	$I(x,\theta)$	$I(x,\theta)p_\theta(x)$
0	0.531	0	0	0	0
1	0.354	-0.132	0.465	1	0.354
2	0.098	-0.044	0.711	1	0.098
3	0.014	0.099	0.900	1	0.015
4	0.001	0.289	1.044	0	0
5	0	0.535	1.132	0	0
6	0	1	1	0	0
					0.467

Tabla 5.1. Cálculo de la probabilidad de cobertura.

Si la interpretación deseada, para el grado de confianza del intervalo, es que en promedio se tenga una cobertura de $100(1-\alpha)\,\%$, entonces, la alternativa más robusta está dada por el intervalo de *Score*, definido por las ecuaciones (5.8) y (5.9), ya que es el que presenta mejor comportamiento en longitud esperada del intervalo, en su varianza y en probabilidad de cobertura, la cual fluctúa alrededor del valor nominal sin importar los valores de n y θ.

La figura 5.3 muestra el nivel de cobertura del intervalo de Wald del 95 %, para cada uno de los valores posibles de θ. Para un tamaño de muestra de $n = 10$, se observa que el nivel de cobertura del intervalo es inferior al nivel de confianza establecido, indicado por la línea horizontal. Para un tamaño de muestra de $n = 100$, esta figura muestra que el intervalo de Wald presenta un mejor nivel de cobertura, especialmente para valores del parámetro cercanos a 1/2. Sin embargo, para valores del parámetro cercanos a cero (o a uno) el intervalo presenta un mal desempeño.

La figura 5.4 muestra el nivel de cobertura del intervalo de Wald ajustado, para cada uno de los valores posibles del parámetro. Para los dos tamaños de muestra considerados, $n = 10$ y $n = 100$, se observa que el intervalo de Wald ajustado presenta un mejor desempeño que el intervalo de Wald. Los niveles de cobertura están cerca del nivel de confianza considerado, que es del 95 %. Sin embargo, este intervalo muestra cierto nivel de sobre-cobertura para valores del parámetro cercanos a cero o a uno.

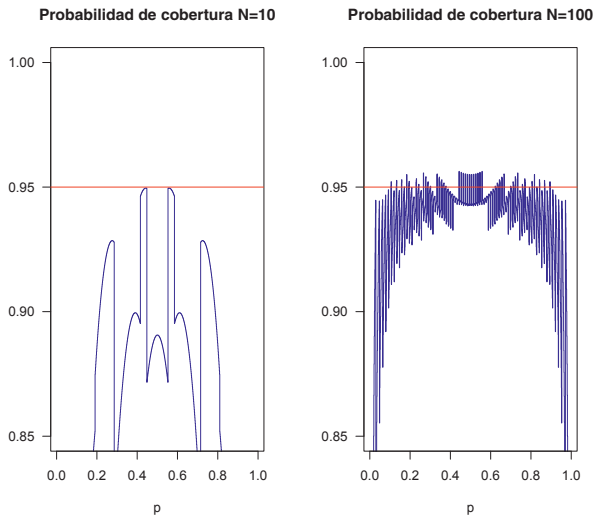

Figura 5.3. Cobertura del intervalo de Wald.

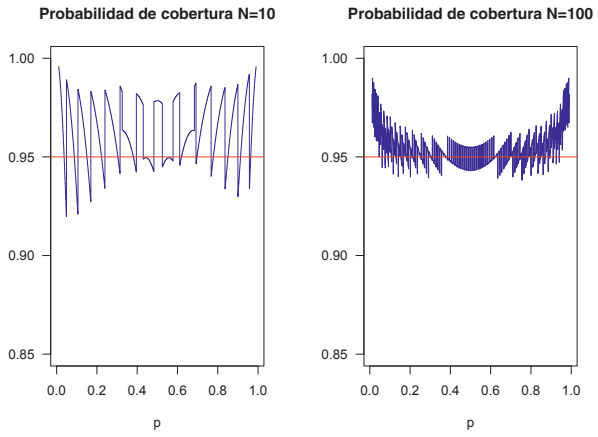

Figura 5.4. Cobertura del intervalo de Wald ajustado.

5.5 Intervalos de confianza para la diferencia de medias

Sean $Y_{11}, Y_{12}, \ldots, Y_{1n}$ una muestra aleatoria, de tamaño n, de una población normal con media θ_1 y varianza σ_1^2, y sea $Y_{21}, Y_{22}, \ldots, Y_{2m}$ una muestra

aleatoria, de tamaño m, de una población normal con media θ_2 y varianza σ_2^2. Si σ_1^2 y σ_2^2 son conocidas, de la subsección 5.1.1, se concluye que

$$Z = \frac{\left(\bar{Y_1} - \bar{Y_2}\right) - (\theta_1 - \theta_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \tag{5.13}$$

tiene distribución Normal estándar. Por tanto, procediendo como se hizo en la subsección 5.1.1, el intervalo de confianza para la diferencia de medias, $\theta_1 - \theta_2$, tiene la forma $(\underline{L}, \bar{L})$, donde

$$\underline{L}(Y) = \left(\bar{Y_1} - \bar{Y_2}\right) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$$

y

$$\bar{L}(Y) = \left(\bar{Y_1} - \bar{Y_2}\right) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}.$$

Si se desconocen las varianzas y se desea hallar un intervalo de confianza para la diferencia de las medias poblacionales, $\theta_1 - \theta_2$, se debe estimar la varianza a partir de los datos. Bajo la hipótesis $H : \sigma_1^2 = \sigma_2^2 = \sigma^2$, un estimador apropiado de σ^2 es

$$S^2 = \frac{(n-1)\, S_1^2 + (m-1)\, S_2^2}{n+m-2}, \tag{5.14}$$

donde

$$S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_{1i} - \bar{Y_1})^2$$

y

$$S_2^2 = \frac{1}{m-1}\sum_{i=1}^{m}(Y_{2i} - \bar{Y_2})^2.$$

Obsérvese que bajo la hipótesis $H : \sigma_1^2 = \sigma_2^2 = \sigma^2$,

$$\frac{(n+m-2)\, S^2}{\sigma^2} = \frac{(n-1)S_1^2}{\sigma^2} + \frac{(m-1)S_2^2}{\sigma^2} \tag{5.15}$$

tiene distribución t con $(n+m-2)$ grados de libertad. Así, de (5.13) y (5.15) y aplicando la ecuación (1.32), en la definición de la distribución t, se concluye que

$$T = \frac{\left(\bar{Y}_1 - \bar{Y}_2\right) - (\theta_1 - \theta_2)}{S\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

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

En consecuencia, el $100(1-\alpha)\,\%$ intervalo de confianza para $(\theta_1 - \theta_2)$ tiene la forma $[\underline{L}, \bar{L}]$, donde

$$\underline{L} = \left(\bar{Y}_1 - \bar{Y}_2\right) - t_{n+m-2,\,\alpha/2} S\sqrt{\frac{1}{n} + \frac{1}{m}}$$

y

$$\bar{L} = \left(\bar{Y}_1 - \bar{Y}_2\right) + t_{n+m-2,\,\alpha/2} S\sqrt{\frac{1}{n} + \frac{1}{m}}.$$

5.6 Intervalos de confianza aproximados para la diferencia de proporciones

Sean $X_{1i} \sim Ber(p_1)$, $i = 1, 2, 3, \ldots, n_1$, y $X_{2j} \sim Ber(p_2)$, $j = 1, 2, 3, \ldots, n_2$; $n_1 + n_2$ variables aleatorias independientes. Si $\hat{p}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_{1i}$ y $\hat{p}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} X_{2i}$, entonces,

$$\frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \xrightarrow{d} N(0,1), \tag{5.16}$$

donde $q_k = 1 - p_k$; $i = 1, 2$. Esto significa que, para valores grandes de n_1 y n_2, la distribución de la variable aleatoria dada por el cociente en (5.16), es muy aproximada a una distribución Normal estándar y que la aproximación es mejor a medida que el tamaño de las muestras aumente. Así, para valores grandes de n_1 y n_2,

$$P\left(-z_{\alpha/2} \leq \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \leq z_{\alpha/2}\right) \approx 1 - \alpha.$$

Finalmente, sustituyendo p_k y q_k; $k = 1, 2$, por $\hat{p}_k$ y $\hat{q}_k$ se encuentra un intervalo de confianza para $p_1 - p_2$, con límites determinados por

$$\underline{L}(X, Y) = (\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} \tag{5.17}$$

y

$$\bar{L}(X, Y) = (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}, \tag{5.18}$$

tiene un nivel de confianza aproximada del $100(1 - \alpha)\,\%$.

Al igual que el intervalo de confianza dado por (5.10), el intervalo de confianza dado por (5.17) y (5.18), aunque aparece en múltiples textos de estadística como Freund, Miller y Miller (2000) y Mendenhall, Wackerly y Scheaffer, (1994), no es recomendable para la estimación de intervalos de confianza de la diferencia de dos proporciones.

Agresti & Caffo (2000) generalizaron el intervalo de confianza (5.11) para la estimación de intervalos de confianza de la diferencia de dos proporciones. Su propuesta está dada por las ecuaciones (5.19) y (5.20), y en general se recomienda para todos los valores de θ y todos los tamaños de muestra (Wei 2002).

$$\underline{L}(X, Y) = (\tilde{p}_1 - \tilde{p}_2) - z_{\alpha/2}\sqrt{\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \frac{\tilde{p}_2\tilde{q}_2}{n_2}} \tag{5.19}$$

y

$$\bar{L}(X, Y) = (\tilde{p}_1 - \tilde{p}_2) + z_{\alpha/2}\sqrt{\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \frac{\tilde{p}_2\tilde{q}_2}{n_2}}. \tag{5.20}$$

En estas ecuaciones, $\tilde{p}_k = \frac{1}{n_k+2}\left(1 + \sum_{i=1}^{n_k} X_{ki}\right)$ y $\tilde{q}_k = 1 - \tilde{p}_k$, $k = 1, 2$

Ejercicios 5.4.

1. Dos muestras aleatorias independientes, A y B, obtenidas de distribuciones Normales, con medias desconocidas y varianzas 4 y 9, respectivamente, están dadas por las siguientes observaciones:

a) Muestra A: -1.98, 0.79, 4.53, -3.48, -1.15, -1.49, 6.55, -0.48.

b) Muestra B: 3.48, 7.59, -14.83, 19.08, 5.33, 0.80, 6.80, 6.36, 14.77, 19.63, 3.18, 5.21.

Considere $\alpha = 0.05$ y halle intervalos de confianza para cada una de las medias de las distribuciones y para su diferencia. ¿Qué se puede concluir?

2. Considere dos muestras aleatorias independientes, A y B, obtenidas de distribuciones Normales, con medias y varianzas desconocidas, que están dadas por las siguientes observaciones:

   a) Muestra A: -0.93, 5.27, 5.67, 1.69, 8.49, 6.13, 4.90, 3.93.

   b) Muestra B: 6.60, 2.30, 9.66, 3.01, 10.04, 14.47, 8.81, 3.84, 0.70, 8.94.

   Asumiendo que $\sigma_A = \sigma_B$ y $\alpha = 0.05$, halle intervalos de confianza para cada una de las medias de las distribuciones y para su diferencia. ¿Qué se puede concluir?

3. En una muestra de estudiantes de una universidad colombiana, 30 de 150 hombres y 50 de 200 mujeres presentan niveles altos de colesterol. Construya intervalos de confianza para

   a) Las proporciones de hombres y de mujeres que tienen niveles altos de colesterol.

   b) La diferencia entre las proporciones de hombres y de mujeres que tienen niveles altos de colesterol.

4. Dos máquinas A y B producen el mismo tipo de artículos y por políticas de la empresa se debe parar una de ellas. Al obtener muestras aleatorias independientes de 150 artículos, para cada una, se encuentran 12 artículos defectuosos en la muestra obtenida de la máquina A y 24 de la B. ¿Existe alguna evidencia estadística suficiente para parar la máquina B? Concluya con niveles de significación del 90 % y del 95 %.

5.7 Intervalos de confianza para σ^2

Si Y_i, $i = 1, 2, \ldots, n$; son n variables aleatorias independientes, de tamaño n_1 y n_2, respectivamente, normalmente distribuidas con media θ y varianza σ^2 desconocidas, entonces, la variable aleatoria $X^2 = \frac{(n-1)S^2}{\sigma^2}$ tiene distribución χ^2 con $n-1$ grados de libertad. En consecuencia, existen números reales positivos x_1^2 y x_2^2 tales que

$$P\left[x_1^2 \leq \frac{(n-1)\,S^2}{\sigma^2} \leq x_2^2\right] = 1 - \alpha.$$

Así, con un nivel de $1-\alpha$, un intervalo de confianza para σ^2 es

$$\left[\frac{(n-1)\,S^2}{x_2^2}, \frac{(n-1)\,S^2}{x_1^2}\right].$$

Aunque interesa escoger el intervalo más pequeño que corresponda a una probabilidad de $1-\alpha$, se acostumbra escoger los números reales positivos x_1^2 y x_2^2 tales que $P(\chi^2_{n-1} \leq x_1^2) = P(\chi^2_{n-1} \geq x_2^2) = \alpha/2$.

Ejercicios 5.5.

1. Se desea comprar un cultivo de pinos para producir cierto tipo de muebles. La media del diámetro de los pinos es de 50 cm y existe acuerdo en cuanto al precio. El comprador toma una muestra aleatoria de 7 árboles, obteniendo los resultados dados en la tabla 5.2.

Medición	1	2	3	4	5	6	7
Diámetro (cm)	50	53	55	47	48	53	54

Tabla 5.2. Diámetro de pinos.

   Considera a $\alpha = 0.05$ y basándose en estos datos, decide estadísticamente que no debe hacer la compra. ¿Qué se puede concluir con relación a la varianza requerida por el comprador?

2. De la medición de la cantidad de azúcar contenida en cada uno de los sobres de una muestra de tamaño 8, tomados aleatoriamente de la producción de una empresa empacadora, se obtuvo los datos que se muestran en la tabla 5.3. Halle un intervalo de confianza para σ^2 y estime el error en este proceso de producción. Considere $\alpha = 0.05$.

Número de sobre	1	2	3	4	5	6	7	8
Peso (gr)	10.2	11.1	9.8	9.8	10.5	10.4	10.1	11.2

Tabla 5.3. Peso en gramos de sobres de azúcar.

3. En una muestra aleatoria de una distribución Normal $N(\theta, \theta)$, se obtuvo las siguientes observaciones: 10.52, 1.31, 3.83, 2.07, 2.69, -3.11, 10.39, 10.70, 8.21, 3.38. Halle intervalos de confianza para θ y compare sus resultados.

4. Sea X_i, $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una distribución Normal con media θ y varianza σ^2 desconocidas. Determine un límite inferior de confianza $\underline{L}$ y un límite superior de confianza $\bar{L}$ tal que $\sigma^2 \in (\underline{L}, \infty)$ con probabilidad $1-\alpha_1$ y $\sigma^2 \in (0, \bar{L})$ con probabilidad $1 - \alpha_2$. Demuestre que $\sigma^2 \in (\underline{L}, \bar{L})$ con probabilidad $1 - \alpha_1 - \alpha_2$.

5. Calcule $E(S^2)$ y $Var(S^2)$, donde S^2 está definido en (5.14).

6. Sean Y_i, $i = 1, 2, \ldots, n$; n variables aleatorias con distribución Normal con media θ conocida y varianza σ^2 desconocida. Determine un intervalo de confianza de nivel $(1 - \alpha)$ para σ^2.

7. Sean $X_i, i = 1, 2, \ldots, n$; n variables aleatorias independientes con función de densidad Rayleigh, dada por

$$f(x_i, \theta) = (x_i/\theta^2) \exp\{-x_i^2/2\theta^2\}, x_i > 0, \theta > 0.$$

   Demuestre que $Y_i = (X_i/\theta)^2$ tiene distribución $\chi^2_{(2)}$ y determine un intervalo de confianza para θ.

5.8 Intervalos de confianza para σ_1^2/σ_2^2

Si $X_i \sim N(\theta, \sigma_1^2)$, $i = 1, 2, \ldots, n_1$, y $Y_j \sim N(\theta, \sigma_2^2)$, $j = 1, 2, \ldots, n_2$, son dos muestras aleatorias, de tamaño n_1 y n_2, respectivamente, independientes con medias y varianzas desconocidas, la variable aleatoria $F = \sigma_2^2 S_1^2/\sigma_1^2 S_2^2$ tiene distribución F con $n_1 - 1$ grados de libertad en el numerador y $n_2 - 1$ grados de libertad en el denominador (definición 1.3). Por tanto, existen números reales a y b tales que

$$P\left(a \leq \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \leq b\right) = 1 - \alpha. \tag{5.21}$$

a y b se eligen de modo que $P(F \leq a) = P(F \geq b) = \alpha/2$. A partir de la ecuación (5.22) se encuentra que

$$P\left(\frac{S_1^2}{bS_2^2} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{S_1^2}{aS_2^2}\right) = 1 - \alpha.$$

Así, los límites superior e inferior del intervalo de confianza para σ_1^2/σ_2^2 están dados por

$$\underline{L} = \frac{S_1^2}{bS_2^2} \quad \text{y} \quad \bar{L} = \frac{S_1^2}{aS_2^2}. \tag{5.22}$$

Ejemplo 5.3. Se observan dos procesos de producción de tornillos, A y B. Una muestra aleatoria de 8 tornillos del proceso A tiene varianza $S^2 = 3.6$ y una muestra aleatoria de 10 tornillos del proceso B, varianza $S^2 = 4.7$. Para hallar un intervalo de confianza del 90 % para σ_A/σ_B, inicialmente se eligen números reales a y b, como se indica en la subsección 1.2.8, y luego, reemplazando en (5.22), se obtiene el intervalo de confianza $(0.23, 2.82)$.

Ejercicios 5.6.

1. Suponga que $(\underline{l}, \bar{l})$ es un intervalo de confianza para el cociente de varianzas σ_A^2/σ_B^2. ¿Qué se puede concluir si

   a) $\underline{l} > 1$?

   b) $\bar{l} < 1$?

   c) $\underline{l} < 1$ y $\bar{l} > 1$?

2. De la medición de la cantidad de azúcar contenida en cada uno de los sobres de dos muestras, tomados aleatoriamente de la producción de dos empresas empacadoras A y B, se obtuvo los datos que se muestran en la tabla 5.4. Asumiendo normalidad y $\alpha = 0.05$,

   a) Halle un intervalo de confianza para la varianza de la producción en cada una de las empresas. ¿Qué se puede concluir?

   b) Halle un intervalo de confianza para σ_A^2/σ_B^2 y compare sus resultados con los obtenidos en el literal anterior.

	Número de sobre	1	2	3	4	5	6	7	8
A	Peso (gr)	10.2	11.1	9.8	9.8	10.5	10.4	10.1	11.2
B	Peso (gr)	11.2	8.10	8.8	9.8	10.5	10.0	10.5	11.0

Tabla 5.4. Peso en gramos de sobres de azúcar.

c) Halle el máximo valor de α, α_m, tal que el intervalo de confianza para el cociente de las varianzas contenga a uno.

3. Dos muestras aleatorias independientes, obtenidas de distribuciones Normales, con media 3 y varianza desconocida, están dadas por las siguientes observaciones:

   a) Muestra 1: 6.44, 0.23, 5.43, 7.24, 3.17, 7.64, 8.30, 3.34, 1.27, 0.16.

   b) Muestra 2: -5.66, 4.53, 8.93, 1.75, -4.76, 9.26, 17.37, 12.70.

   Halle intervalos de confianza para cada una de las varianzas y para el cociente de las mismas, con $\alpha = 0.05$.

4. Dos muestras aleatorias independientes obtenidas de distribuciones Normales $N(\theta_i, \theta_i)$, $\theta_i > 0$, $i = 1, 2$; están dadas por las siguientes observaciones:

   a) Muestra 1: 10.52, 1.31, 3.83, 2.07, 2.69, -3.11, 10.39, 10.70, 8.21, 3.38.

   b) Muestra 2: 5.84, 3.44, 4.11, -0.63, 4.17, 5.55, 10.81, 1.72, 11.58, -0.82.

   Halle intervalos de confianza aproximado para la diferencia de medias y para el cociente de las varianzas, con $\alpha = 0.05$.

Capítulo 6

Prueba de hipótesis

Este capítulo está dividido en siete secciones. En las dos primeras se hace una introducción a las pruebas de hipótesis. Luego, se incluyen tópicos de probabilidad de error, regiones de rechazo, intervalos de confianza y pruebas de hipótesis y, finalmente, algunos ejemplos de planteamiento de hipótesis compuestas y ejercicios.

6.1 Prueba de hipótesis simple

Para garantizar la continuidad de un contrato, una empresa de publicidad afirma que durante el primer mes de la campaña en contra del cigarrillo ha disminuido la proporción de fumadores en una ciudad. El Ministerio de Salud considera que la publicidad no ha surtido ningún efecto y propone un experimento para determinar cuál de las siguientes hipótesis es verdadera:

$$H : \text{la publicidad no es efectiva.}$$
$$K : \text{la publicidad es efectiva.}$$

Si "se acepta" H, se concluye que la publicidad no ha tenido éxito (la proporción de fumadores no disminuye) y se debe suspender el contrato. Si "se acepta" K, y hay interés en reducir el número de fumadores, se debe prorrogar el contrato, pues la publicidad ha contribuido a reducir el número de fumadores.

Para tomar una decisión sobre la continuidad o no del contrato se obtiene un conjunto de valores observados, $\mathbf{x} = (x_1, \ldots, x_n)$, de una muestra aleatoria, de tamaño n, de potenciales fumadores, asignando a x_i uno si el individuo es fumador y cero si no, y se determina el número $T(\mathbf{x}) = \Sigma_{i=1}^{n} x_i$ de fumadores en la muestra. Si este número es menor o igual que algún número k, $k = 1, 2, \ldots, n$; es decir, si la proporción de fumadores en la muestra es menor que algún número real θ_0, $0 \leq \theta_0 \leq 1$, se rechaza la hipótesis H y, en consecuencia, se prorroga el contrato.

Determinada una "región de rechazo" (RR) de la hipótesis H, dada por un conjunto $RR = \{0, 1, \ldots, k\}$, se rechaza la hipótesis H siempre que $T(\mathbf{x}) \in RR$. En caso contrario, es decir, si $T(\mathbf{x}) \notin RR$ (o si la proporción de fumadores en la muestra es mayor que θ_0), entonces, no se debe prorrogar el contrato y se afirma que no existe evidencia estadística para certificar que la publicidad ha sido efectiva.

El Ministerio de Salud debe suspender el contrato si la publicidad no es efectiva. Sin embargo, sabe que puede cometer dos tipos de error en su decisión. El primero consiste en prorrogar el contrato, dado que la publicidad no es efectiva; y el segundo, suspender el contrato, dado que la publicidad ha dado los resultados esperados. Si el Ministerio desea reducir la probabilidad de cometer estos dos tipos de error en su decisión, debe obtener mayor información de la población, incrementando el tamaño de la muestra, dado que las probabilidades de error pueden hacerse tan pequeñas como se quiera, siempre que el tamaño de la muestra sea suficientemente grande.

Sin embargo, una vez establecido el tamaño de la muestra, no es posible controlar simultáneamente los dos tipos de error. Así, si el interés del Ministerio es velar por la adecuada inversión de los recursos públicos, debe plantear un procedimiento estadístico que así lo garantice. Es decir, proceder de modo que sea poco probable prorrogar un contrato publicitario, si la publicidad no es efectiva. Esto significa que debe hacer lo más "pequeña" posible la probabilidad de aceptar K, dado que la hipótesis H es verdadera.

6.2 Prueba de hipótesis compuesta

Se tiene información para afirmar que el porcentaje de personas no fumadoras que se recuperan de un ataque cardíaco es 77 %. Dado que el cigarrillo no

tiene ningún efecto positivo sobre la salud, es razonable confrontar hipótesis como las siguientes:

H: El porcentaje de fumadores que se recupera de un ataque cardíaco es mayor o igual a 77 %.

K: El porcentaje de fumadores que se recupera de un ataque cardíaco es menor de 77 %.

En este caso, se tienen las siguientes hipótesis compuestas:

$$H : \theta \in \Theta_1$$
$$K : \theta \in \Theta_2$$

donde $\Theta_1 = [0.77, 1), \Theta_2 = (0, 0.77), \Theta_1 \cap \Theta_2 = \emptyset, \Theta_1 \cup \Theta_2 = (0, 1)$ y $\Theta_2 = \Theta_1^c$.

Definición 6.1. Si Θ_1 contiene más de un punto, H se denomina hipótesis compuesta. Si Θ_1 contiene solo un punto, H se denomina hipótesis simple.

Nota: La misma convención es válida para K.

La decisión de aceptar o rechazar la hipótesis H (hipóstesis nula) se basa en los datos obtenidos a través de un muestreo. Intuitivamente, si de n fumadores, el número de personas que sobrevive a un ataque cardíaco es mayor o igual que algún entero k_0, $0 \leq k_0 \leq n$, se decide en favor de H y si el número de fumadores que sobrevive es menor que k_0, se decide en favor de K. Es evidente que, en este caso, el estadístico de prueba es $T(\mathbf{X}) = \Sigma_{i=1}^n X_i$. Si $T(\mathbf{x}) \geq k_0$, no existen argumentos para rechazar H y, por otra parte, si $T(\mathbf{x}) < k_0$, se rechaza H.

6.3 Probabilidad de error

Retomando el ejemplo de la proporción de fumadores, pero ahora considerando la hipótesis simple se tiene que si $H : \theta = \theta_0$ es la "hipótesis nula" y $K : \theta = \theta_1$, $\theta_1 < \theta_0$, la "hipótesis alternativa", entonces, se presentan dos tipos de error, cuando se hace una prueba de hipótesis. Un error de tipo I se comete cuando se rechaza la hipótesis H siendo esta verdadera; un error de tipo II, cuando se acepta la hipótesis H siendo esta falsa. La probabilidad de error de tipo I se nota por α y la probabilidad de cometer el error de tipo II, por β. Así, en este ejemplo, dado que $RR = \{0, 1, \ldots, k_0\}$,

$$\begin{aligned}\alpha &= P_{\theta_0}\{\text{rechazar } H\}\\ &= P_{\theta_0}[T(\mathbf{X}) \leq k_0]\\ &= \sum_{i=0}^{k_0} \binom{n}{i} \theta_0^i (1-\theta_0)^{n-i}. \end{aligned} \tag{6.1}$$

El error de tipo II depende de la hipótesis alternativa, $\theta_1 \in \Theta_1$, que debe considerarse. Así, para cada θ_1,

$$\beta = P_{\theta_1}\{\text{aceptar } H\} = 1 - P_{\theta_1}\{\text{rechazar } H\}.$$

En el ejemplo en cuestión,

$$\beta = P_{\theta_1}(T(\mathbf{X}) > k_0) = \sum_{i=k_0+1}^{n} \binom{n}{i} \theta_1^i (1-\theta_1)^{n-i} \tag{6.2}$$

y, en consecuencia, comparando las expresiones (6.1) y (6.2) se observa que si el tamaño de la muestra es constante, entonces, al disminuir el error de tipo I aumenta el error de tipo II.

Como se afirma en la sección 6.1, no es fácil controlar los errores de tipo I y II. En consecuencia, se deben plantear las hipótesis de modo que el error tipo I sea no solo el más importante para la situación estudiada, sino también el más fácil de controlar.

6.4 Pruebas de hipótesis

Para las pruebas de hipótesis, primero se debe fijar un α de tal manera que la probabilidad de cometer el error de tipo I, mayor que α, sea inadmisible. En la práctica, $\alpha = 0.01$ y $\alpha = 0.05$ son los usados comúnmente. Una vez se fija α, se centra la atención en las pruebas que tengan probabilidad de rechazo "menor que α" o "igual a α", y se determina la región de rechazo (RR).

6.4.1 Prueba de hipótesis para el parámetro de la distribución Binomial

Supóngase que se desea probar las hipótesis $H : \theta = \theta_0$ versus $K : \theta < \theta_0$. Si dado α existe un número entero positivo k_0 tal que

$$\begin{aligned} \alpha &= P_{\theta_0}(\text{rechazar } H) \qquad (6.3) \\ &= \sum_{i=0}^{k_0} \binom{n}{i} \theta_0^i (1-\theta_0)^{n-i}, \end{aligned}$$

la región de rechazo es $RR = \{0, 1, 2, \ldots, k_0\}$.

Si no existe un número entero k_0 tal que cumpla (6.3), la región de rechazo está determinada por el mayor número entero positivo n_0 tal que

$$\alpha > \sum_{i=0}^{n_0} \binom{n}{i} \theta_0^i (1-\theta_0)^{n-i}.$$

En consecuencia, la región de rechazo queda determinada por el conjunto $RR = \{0, 1, 2, \ldots, n_0\}$. Obsérvese que para este nivel de significación, no se puede considerar el conjunto $\{\mathbf{x} \in \mathbb{R}^n : T(\mathbf{x}) \leq n_0 + 1\}$ como región de rechazo, ya que como $P[T(\mathbf{X}) \leq n_0 + 1] > \alpha$, entonces, el tamaño de esta será mayor que α. Si se considera este conjunto como región de rechazo, entonces, no se puede afirmar que la probabilidad de cometer el error de tipo I es α, pues en realidad se está cometiendo un error de tipo I igual a $P[T(\mathbf{X}) \leq n_0 + 1]$, que es mayor que α.

Si no existe k_0 tal que

$$\alpha = \sum_{i=0}^{k_0} \binom{n}{i} \theta_0^i (1-\theta_0)^{n-i}$$

y si $\text{mín}\{n\theta_0, n(1-\theta_0)\} \geq 5$, se puede determinar la región de rechazo mediante la aproximación por la distribución Normal de la distribución Binomial

$$\begin{aligned} P_{\theta_0}[T(X) \leq k_0] &\approx P\left[\frac{T(X) - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}} \leq \frac{k_0 - n\theta_0 + \frac{1}{2}}{\sqrt{n\theta_0(1-\theta_0)}}\right] \\ &= \Phi\left(\frac{k_0 - n\theta_0 + \frac{1}{2}}{\sqrt{n\theta_0(1-\theta_0)}}\right). \end{aligned}$$

Ya que se espera un error de tipo I "menor que α" o "igual a α"; haciendo

$$\frac{(k_0 - n\theta_0) + \dfrac{1}{2}}{\sqrt{n\theta_0(1-\theta_0)}} \leq z_\alpha,$$

se encuentra que la región de rechazo está determinada por el mayor número entero positivo k_0 tal que $k_0 \leq n\theta_0 - \frac{1}{2} + z_\alpha\sqrt{n\theta_0(1-\theta_0)}$.

Ejercicios 6.1.

1. En una muestra aleatoria de veinte estudiantes de una universidad se encuentra que dos hacen deporte. Si θ denota la proporción de los estudiantes de la universidad que hacen deporte y se desea probar la hipótesis $H : \theta = 0.15$ frente a la alternativa $K : \theta = 0.30$, encuentre:

   a) El nivel de significación α y la probabilidad de error de tipo II; asumiendo que la región de rechazo es $RR = \{x \in \mathbb{R} : x > 3\}$, donde x denota el número de estudiantes de la muestra que hacen deporte.

   b) La región de rechazo y la probabilidad de error de tipo II, para un nivel de significación $\alpha = 0.05$.

   c) La región de rechazo y el error de tipo II, aplicando la aproximación normal de la distribución Binomial, para $\alpha = 0.05$.

   d) Comparar la región de rechazo obtenida en 1*b*, con la región de rechazo obtenida en 1*c*.

2. Desarrolle el primer ejercicio, asumiendo que en una muestra aleatoria de diez estudiantes siete hacen deporte y que se desea probar la hipótesis $H : \theta = 0.8$ frente a la alternativa $K : \theta = 0.5$. En el literal *a*) asuma que $RR = \{x \in \mathbb{R} : x < 7\}$.

3. Sean $X_i \sim Ber(\theta)$, $i = 1, 2, \ldots, n$; n variables aleatórias independientes. Dado α, hallar la región de rechazo y el error de tipo II para la prueba de hipótesis $H : \theta = \theta_0$ versus $K : \theta = \theta_1$, $\theta_1 > \theta_0$.

6.5 Prueba de hipótesis referente a la media de una distribución Normal

6.5.1 Prueba de hipótesis para la media de una distribución Normal con varianza conocida

Sea X_i, $i = 1, 2, 3, \ldots, n$; una muestra aleatoria, de tamaño n, de una distribución Normal con media desconocida θ y varianza conocida σ^2. Si se desea probar $H : \theta = \theta_0$ versus $K : \theta > \theta_0$, es claro que el estadístico de prueba es $T(\mathbf{X}) = \bar{X}$ y que dado un nivel de significación α, la región de rechazo está determinada por un número real c tal que

$$\begin{aligned} \alpha &= P_{\theta_0}[T(\mathbf{X}) \geq c] \\ &= P_{\theta_0}\Big[\frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} \geq \frac{c - \theta_0}{\sigma/\sqrt{n}}\Big]. \end{aligned}$$

En consecuencia, dado que bajo la hipótesis H, la variable aleatoria $Z = \frac{\bar{X}-\theta_0}{\sigma/\sqrt{n}}$ tiene distribución Normal estándar, entonces, $\alpha = P_{\theta_0}\left(Z \geq \frac{c-\theta_0}{\sigma/\sqrt{n}}\right)$ para $\frac{c-\theta_0}{\sigma/\sqrt{n}} = z_\alpha$. Así, $c = \theta_0 + z_\alpha \sigma/\sqrt{n}$ y, por tanto, la región de rechazo está dada por el conjunto

$$RR = \{\mathbf{x} \in \mathbb{R}^n : T(\mathbf{x}) \geq \theta_0 + z_\alpha \sigma/\sqrt{n}\}. \tag{6.4}$$

Con las mismas observaciones, en una prueba de hipótesis de $H : \theta = \theta_0$ versus $K : \theta \neq \theta_0$, se puede considerar como estadístico de prueba $T(\mathbf{X}) = \sqrt{n}(\bar{X} - \theta_0)/\sigma$, el cual bajo la hipótesis H tiene distribución Normal estándar. En este caso, la región de rechazo se establece para valores grandes de $|T|$. Más exactamente, $RR = \{\mathbf{x} \in \mathbb{R}^n : |T(\mathbf{x})| \geq z_{\alpha/2}\}$.

En general, cuando se hacen pruebas de hipótesis como estas, no se conoce la varianza; sin embargo, si el tamaño, n, de la muestra es mayor que 30, entonces, se puede utilizar la distribución Normal, reemplazando σ por su valor estimado en la muestra. Si $n < 30$, se debe utilizar la distribución t, como se indica en la subsección 6.5.2.

Ejemplo 6.1. Una muestra de azúcar empacados por *Azucarco S.A.*, tienen un peso promedio de 5.2 gramos y una desviación estándar de 0.4 gramos. Si el gerente afirma que el contenido medio de estos sobres es de 5 gramos,

¿qué se puede afirmar, con un nivel de significación del 5 %, con respecto a la posibilidad de que la media sea mayor?

Para solucionar este interrogante, se debe plantear como hipótesis nula $H : \theta = 5$ y como hipótesis alternativa $K : \theta > 5$. Dado que la región de rechazo está determinada por el conjunto:

$$\begin{aligned} RR &= \{\mathbf{x} \in \mathbb{R}^{100} : \bar{x} \geq \theta_0 + z_\alpha \sigma/\sqrt{n}\} \\ &= \{\mathbf{x} \in \mathbb{R}^{100} : \bar{x} \geq 5 + 1.645\,(0.4/10)\} \\ &= \{\mathbf{x} \in \mathbb{R}^{100} : \bar{x} \geq 5.066\}, \end{aligned}$$

se tiene que $\bar{x} = 5.2$ pertenece a la región de rechazo. Esto significa que, con un nivel de significación $\alpha = 0.05$, se rechaza la hipótesis H en favor de la hipótesis K, es decir, el contenido de los sobres de azúcar es mayor que 5 gramos.

Si el contenido medio de los cien sobres de azúcar fuera de 4.8 gramos y la desviación estándar 0.4 gramos, para la misma prueba de hipótesis, la región de rechazo continuará siendo $RR = \{\mathbf{x} \in \mathbb{R}^{100} : \bar{x} \geq 5.066\}$. Pero en este caso, no se rechaza H, pues 4.8 no pertenece a la región de rechazo. Entonces, no hay evidencia estadística para afirmar que el contenido medio de los sobres de azúcar es mayor que 5 gramos. Sin embargo, esto no significa que se acepte H, pues no se conoce el error de tipo II, que solo se puede calcular para valores específicos de la hipótesis alternativa.

Por ejemplo, si $K : \theta_1 = 5.1$,

$$\begin{aligned} \beta &= P_{\theta_1}(\bar{X} < 5.066) \\ &= P_{\theta_1}\left(Z < \frac{5.066 - 5.1}{0.04}\right) \\ &= P_{\theta_1}(Z < -0.85) \\ &= 0.1977 \end{aligned}$$

y si $K : \theta_1 = 5.5$,

$$\begin{aligned} \beta &= P_{\theta_1}(\bar{X} < 5.066) \\ &= P_{\theta_1}\left(Z < \frac{5.066 - 5.5}{0.04}\right) \\ &= P_{\theta_1}(Z < -10.85) \\ &\approx 0. \end{aligned}$$

En las pruebas de hipótesis de la forma $H : \theta = \theta_0$ versus $K : \theta \neq \theta_0$, se puede considerar $T(\mathbf{X}) = |\bar{X} - \theta_0|$ como un estadístico de prueba. Si la varianza es conocida, para un nivel de significación α, la hipótesis H se rechaza para valores de $T(\mathbf{X})$ mayores o iguales que c, donde $c = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. Esto es equivalente a rechazar H para valores de $\bar{X}$ pertenecientes a $(\underline{L}, \bar{L})^c$, donde

$$\underline{L} = \theta_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \quad \text{y} \quad \bar{L} = \theta_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

En estos intervalos, si la varianza es desconocida, pero el tamaño de la muestra es mayor que 30, en general, la desviación estándar poblacional puede reemplazarse por su estimación muestral.

Ejemplo 6.2. Supóngase que el contenido medio de los cien sobres de azúcar, considerados en el ejemplo 6.1, es de 4.8 gramos y la desviación estándar 0.4 gramos, pero que se desea probar las hipótesis $H : \theta = 5$ versus $K : \theta \neq 5$. Si $\alpha = 0.05$, considerando como un estadístico de prueba

$$T(\mathbf{X}) = \frac{\sqrt{n}(\bar{X} - 5)}{\sigma}, \tag{6.5}$$

se encuentra que la región de rechazo está determinada por el conjunto $RR = (-1.96, 1.96)^c$. Dado que $\bar{x} = 4.8$ gramos, entonces, $T(\mathbf{x}) = -5 \in (-1.96, 1.96)^c$ y por ende se rechaza H. Así se concluye: "existe evidencia estadística para afirmar que el contenido medio de azúcar en los empaques es diferente de 5 gramos".

6.5.2 Prueba de hipótesis para la media de una distribución Normal con varianza desconocida

Supóngase que X_i, $i = 1, 2, \ldots, n$; es una muestra aleatoria, de tamaño n, que proviene de una distribución Normal con media θ y varianza σ^2 desconocidas, y que se desea probar $H : \theta = \theta_0$ versus $K : \theta > \theta_0$. En este caso, el estadístico de prueba es $T(\mathbf{X}) = \bar{X}$, y dado un nivel de significación α, la región de rechazo está determinada por un número real c tal que $\alpha = P_{\theta_0}[\bar{X} \geq c]$. Así pues,

$$\alpha = P_{\theta_0}\left(T_{n-1}(\mathbf{X}) \geq \frac{c - \theta_0}{s/\sqrt{n}}\right),$$

donde $T_{n-1}(\mathbf{X}) = \frac{\bar{X}-\theta_0}{S/\sqrt{n}}$.

Dado que $T_{n-1}(\mathbf{X})$ tiene distribución t con $n-1$ grados de libertad, existe un número real $t_{n-1,\,\alpha}$ tal que

$$\frac{c-\theta_0}{s/\sqrt{n}} = t_{n-1,\alpha}$$

y, en consecuencia, la región de rechazo está determinada por el conjunto $RR = \{\mathbf{x} \in \mathbb{R}^n : T(\mathbf{x}) \geq c\}$, donde $c = \theta_0 + t_{n-1,\,\alpha} s/\sqrt{n}$.

En una prueba de hipótesis $H : \theta = \theta_0$ versus $K : \theta \neq \theta_0$, también se considera a $T_{n-1}(\mathbf{X})$ como un estadístico de prueba. En este caso, se rechaza la hipótesis H para valores grandes de $|T_{n-1}(\mathbf{X})|$. Dado que $T_{n-1}(\mathbf{X})$ tiene distribución t con $n-1$ grados de libertad, la región de rechazo está dada por

$$RR = \{\mathbf{x} \in \mathbb{R}^n : |T_{n-1}(\mathbf{x})| \geq t_{n-1,\,\alpha/2}\}.$$

En las pruebas de hipótesis de la forma $H : \theta = \theta_0$ versus $K : \theta \neq \theta_0$, si la varianza es desconocida y el tamaño de la muestra es menor que 30, se puede utilizar el mismo estadístico de prueba $T_{n-1}(\mathbf{X})$ que bajo H tiene distribución t con $n-1$ grados de libertad. En este caso, la región de rechazo queda determinada por $(\underline{L}, \bar{L})^c$, donde

$$\underline{L} = \theta_0 - t_{n-1,\,\alpha/2}\frac{S}{\sqrt{n}} \quad \text{y} \quad \bar{L} = \theta_0 + t_{n-1,\,\alpha/2}\frac{S}{\sqrt{n}}.$$

Ejemplo 6.3. Supóngase que, en el ejemplo anterior, solo se observan nueve sobres de azúcar y que se desea probar $H : \theta = 5$ versus $K : \theta \neq 5$. Si el contenido medio de esta muestra es de 4.8 gramos y su desviación estándar 0.4 gramos, considerando un nivel de significación $\alpha = 0.05$, la región de rechazo queda determinada por

$$\left(5 - 2.306\frac{0.4}{3}, \quad 5 + 2.306\frac{0.4}{3}\right)^c = [4.692, 5.307]^c.$$

Dado que $\bar{x} = 4.8$ no pertenece a la región de rechazo, no existe evidencia estadística para rechazar la hipótesis H.

Ejercicios 6.2.

1. Determine el error de tipo II, en la prueba de hipótesis del ejemplo 6.2, para $\theta_1 = 4$, $\theta_1 = 4.5$ y $\theta_1 = 5.3$.

2. Sea $X_i \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una distribución Normal. En la prueba de hipótesis de $H : \theta = \theta_0$ versus $K : \theta < \theta_0$, determine la región de rechazo de nivel de significación α, asumiendo que

   a) σ^2 es conocida.

   b) σ^2 es desconocida.

   En cada caso, halle el error de tipo II para θ_1, $\theta_1 < \theta_0$.

3. Sean $X_i \sim U(0, \theta)$, $i = 1, 2, \ldots, n$; n variables aleatorias independientes. En la prueba de hipótesis de $H : \theta = \theta_0$ versus $K : \theta > \theta_0$, determine la región de rechazo, de nivel de significación α, y el error de tipo II para:

   a) $\theta_1 < \theta_0$.

   b) $\theta_1 > \theta_0$.

4. Un medicamento debe contener 15 mg de Escina. Para probar si la producción del medicamento cumple con esta condición se determina la cantidad de Escina contenida en cada una de las tabletas de una muestra aleatoria de tamaño 100, obteniendo una media de 15.1 mg y una desviación estándar de 0.3 mg. Si el nivel de significación es $\alpha = 0.05$, ¿qué se puede concluir en una prueba de hipótesis de

   a) $H : \theta = 15$ versus $H : \theta \neq 15$?

   b) $H : \theta = 15$ versus $H : \theta > 15$?

   Determine el error de tipo II, para $\theta = 20$, en cada una de las pruebas.

6.6 Prueba de hipótesis para una diferencia de medias

6.6.1 Prueba de hipótesis para una diferencia de medias, usando estadísticos normalmente distribuidos

Sean $X_i \sim N(\theta_1, \sigma_1^2)$, $i = 1, 2, \ldots, n_1$, y $Y_i \sim N(\theta_2, \sigma_2^2)$, $i = 1, 2, \ldots, n_2$, dos muestras aleatorias, de tamaños n_1 y n_2, respectivamente, independientes de dos poblaciones normales con medias desconocidas y varianzas conocidas. Si se desea probar $H : \theta_2 - \theta_1 = \theta_0$ versus $K : \theta_2 - \theta_1 > \theta_0$, entonces es conveniente considerar el siguiente estadístico de prueba:

$$T(\mathbf{X}, \mathbf{Y}) = \frac{\bar{Y} - \bar{X} - \theta_0}{\sqrt{\frac{\sigma_2^2}{n_2} + \frac{\sigma_1^2}{n_1}}}, \tag{6.6}$$

donde $\mathbf{X} = (X_1, \ldots, X_{n_1})$ y $\mathbf{Y} = (Y_1, \ldots, X_{n_2})$.

Dado que, bajo la hipótesis H, el estadístico $T(\mathbf{X}, \mathbf{Y})$ tiene distribución Normal estándar, entonces, para un nivel de significación α, la región de rechazo está dada por $RR = \{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} : T(\mathbf{x}, \mathbf{y}) \geq z_\alpha\}$. Esta región es equivalente a

$$RR = \left\{ (\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} : T_1(\mathbf{x}, \mathbf{y}) \geq \theta_0 + z_\alpha \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2} \right\}, \tag{6.7}$$

donde $T_1(\mathbf{X}, \mathbf{Y}) = \bar{Y} - \bar{X}$.

Si se cambia la hipótesis alternativa por $K : \theta_2 - \theta_1 \neq \theta_0$, el estadístico de prueba es el mismo. Pero, en este caso, la región de rechazo está determinada por $RR = \{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} : |T(\mathbf{x}, \mathbf{y})| \geq z_{\alpha/2}\}$.

Si las varianzas σ_1^2 y σ_2^2 son desconocidas, se puede utilizar el estadístico (6.6), reemplazando las varianzas por sus estimaciones, siempre que $n_1 > 30$ y $n_2 > 30$.

6.6.2 Prueba de hipótesis para una diferencia de medias, usando estadísticos con distribución t

Sean $X_i \sim N(\theta_1, \sigma_1^2)$, $i = 1, 2, \ldots, n_1$, y $Y_i \sim N(\theta_2, \sigma_2^2)$, $i = 1, 2, \ldots, n_2$, dos muestras aleatorias independientes de tamaño n_1 y n_2, respectivamente, de dos poblaciones normales con medias y varianzas desconocidas y con $\sigma_1^2 = \sigma_2^2$. Si se desea probar $H : \theta_2 - \theta_1 = \theta_0$ versus $K : \theta_2 - \theta_1 > \theta_0$, es conveniente considerar como estadístico de prueba

$$T(\mathbf{X}, \mathbf{Y}) = \frac{\bar{Y} - \bar{X} - \theta_0}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \tag{6.8}$$

donde $\mathbf{X} = (X_1, \ldots, X_{n_1})$, $\mathbf{Y} = (Y_1, \ldots, Y_{n_2})$ y

$$S^2 = \frac{1}{n_1 + n_2 - 2} \left[\Sigma_{i=1}^{n_1} (X_i - \bar{X})^2 \; + \; \Sigma_{i=1}^{n_2} (Y_i - \bar{Y})^2 \right].$$

Dado que, bajo la hipótesis H, el estadístico $T(\mathbf{X}, \mathbf{Y})$ tiene distribución $\mathbf{t}_{(n_1+n_2-2)}$, entonces, para un nivel de significación α, la región de rechazo está dada por

$$RR = \{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} : T(\mathbf{x}, \mathbf{y}) \geq t_{\alpha; n_1+n_2-2}\}.$$

Esta región es equivalente a

$$RR = \left\{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} : T_1(\mathbf{x}, \mathbf{y}) \geq \theta_0 + t_{\alpha, n_1+n_2-2} s \sqrt{1/n_1 + 1/n_2}\right\}$$

donde $T_1(\mathbf{X}, \mathbf{Y}) = \bar{Y} - \bar{X}$.

Si se cambia la hipótesis alternativa por $K : \theta_2 - \theta_1 \neq \theta_0$, el estadístico de prueba es el mismo. Pero, en este caso, la región de rechazo está determinada por $RR = \{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} : |T(\mathbf{x}, \mathbf{y})| \geq t_{\frac{\alpha}{2}, n_1+n_2-2}\}$.

La figura 6.1 muestra la distribución de una muestra aleatoria de tamaño diez mil, del estadístico $T(\mathbf{X}, \mathbf{Y})$, cuando este se calcula a partir de muestras aleatorias de tamaño quince, obtenidas de una distribución $N(10, 4)$ y de una distribución exponencial con media diez.

En esta figura, la línea más gruesa corresponde a la distribución de T y la más delgada a la distribución t con 28 grados de libertad. Obsérvese que,

en este caso, la distribución real de $T(\mathbf{X}, \mathbf{Y})$ es sesgada a la derecha, lo cual tiene incidencia directa sobre los p-valores en las pruebas de hipótesis. Este cambio en la distribución de $T(\mathbf{X}, \mathbf{Y})$, generado por la violación del supuesto de normalidad, puede observarse usando simulación de Monte Carlo.

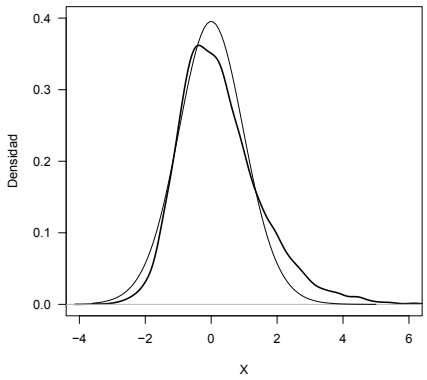

Figura 6.1. Distribución de la estadística T.

Ejercicios 6.3.

1. Dos muestras aleatorias A y B fueron generadas de distribuciones Normales, con medias θ_A y θ_B, respectivamente, encontrándose los siguientes datos.

   Muestra A: 2.2, 3.1, 4.2, 2.1, 2.9, 6.2, 4.1, 1.1
   Muestra B: 8.4, 6.6, 6.1, 6.6, 6.8, 5.2, 7.0, 5.6, 5.7.

   Asumiendo un nivel de significación $\alpha = 0.001$, pruebe la hipótesis que las medias son iguales, $H : \theta_A = \theta_B$, frente a cada una de las hipótesis alternativas $K_1 : \theta_A < \theta_B$ y $K_2 : \theta_A \neq \theta_B$, asumiendo que estas muestras aleatorias se obtienen de poblaciones normales con

   a) Medias desconocidas y varianzas $\sigma_A^2 = 16$ y $\sigma_B^2 = 1$.

   b) Medias y varianzas desconocidas.

2. De las poblaciones del ejercicio anterior, se generaron dos muestras A y B de 100 observaciones, cada una, encontrándose los siguientes resultados:

Muestra A: $\bar{X}_A = 2.81$, $s_A = 2.08$.
Muestra B: $\bar{X}_B = 6.03$, $s_B = 0.97$.

Pruebe las hipótesis planteadas en el primer ejercicio.

6.7 Pruebas de hipótesis referentes a proporciones

6.7.1 Prueba de hipótesis referente a una proporción

Sean $X_i \sim Ber(\theta)$, $i = 1, 2, 3, \ldots, n$; n variables aleatorias independientes. Si el tamaño de las muestras es grande y se desea probar $H : \theta = \theta_0$ versus $K : \theta > \theta_0$, entonces se puede usar como estadístico de prueba $S(\mathbf{X}) = \sum_{i=1}^{n} X_i$ y rechazar H para valores grandes de $S(\mathbf{X})$, de acuerdo con el nivel de significación considerado, como se presentó al comienzo de este capítulo.

Sin embargo, teniendo en cuenta que

$$T(\mathbf{X}) = \frac{\sqrt{n}(\hat{\theta} - \theta)}{\sqrt{\theta(1-\theta)}} \xrightarrow{d} N(0,1), \tag{6.9}$$

donde $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$, para valores grandes de n, se puede considerar como estadístico de prueba a

$$T(\mathbf{X}) = \frac{\sqrt{n}(\hat{\theta} - \theta_0)}{\sqrt{\theta_0(1-\theta_0)}}, \tag{6.10}$$

dado que, bajo la hipótesis H, tiene "aproximadamente" distribución Normal estándar.

Así, para un nivel de significación α, se rechaza la hipótesis H, para valores de $T(\mathbf{X})$ mayores que z_α, donde z_α es tal que $\alpha = P(Z > z_\alpha)$, siendo Z una variable normal estándar.

Ejemplo 6.4. En un proceso de control de calidad de la producción de tornillos de una empresa, se observan 100 tornillos de los cuales 9 son defectuosos. Para probar si existe un incremento en el número de tornillos defectuosos, de esta producción, se plantea la siguiente prueba de hipótesis: $H : \theta = 0.05$ versus $K : \theta > 0.05$, donde θ representa la proporción de tornillos defectuosos. Entonces:

1. Usando el estadístico de prueba, definido en la ecuación (6.10), se encuentra que $T(\mathbf{x}) = 1.8353$. Dado que $P(T(\mathbf{X}) > 1.8353) = 0.0332$, entonces, la hipótesis H se rechaza para niveles de significación mayores que 0.0332 (por ejemplo, para $\alpha = 0.05$), pero no se rechaza para niveles de significación menores que 0.0332 (por ejemplo, para $\alpha = 0.01$).

2. Usando el mismo estadístico de prueba, pero considerando el factor de corrección en la aproximación de la distribución Normal a la distribución Binomial, se encuentra que $T(\mathbf{x}) = 1.6059$. Así, $P(T(\mathbf{X}) > 1.6059) = 0.05415$ y se rechaza la hipótesis H para niveles de significación mayores que 0.05415.

3. Considerando como estadístico de prueba $S(\mathbf{X}) = \sum_{i=1}^{n} X_i$, de la distribución Binomial se encuentra que $P_{0.05}(S(\mathbf{X}) \geq 9) = 0.06309$ y, en consecuencia, la hipótesis $H : \theta = 0.05$ no se rechaza para valores de α menores que 0.06309.

De estas tres pruebas se infiere que, a pesar de tener una muestra "grande", las conclusiones de la prueba desarrollada, bajo normalidad, no concuerdan con la prueba desarrollada usando la distribución Binomial.

Nota: Dada la existencia de software estadísticos libre, como R, la recomendación es hacer este tipo de pruebas usando la distribución Binomial. Por otra parte, si se usa la aproximación Normal de la distribución Binomial, es conveniente incluir el factor de corrección.

6.7.2 Prueba de hipótesis referente a la diferencia de dos proporciones

Sean $X_i \sim Ber(\theta_1)$, $i = 1, 2, \ldots, n_1$, y $Y_j \sim Ber(\theta_2)$, $j = 1, 2, \ldots, n_2$, dos muestras aleatorias independientes de tamaño n_1 y n_2, respectivamente. Si el tamaño de las muestras es grande y se desea probar la hipótesis $H : \theta_2 - \theta_1 = \theta_0$ versus $K : \theta_2 - \theta_1 > \theta_0$, se puede usar el siguiente estadístico de prueba:

$$T(\mathbf{X}, \mathbf{Y}) = \frac{(\hat{\theta}_2 - \hat{\theta}_1) - \theta_0}{\sqrt{\frac{\hat{\theta}_1(1-\hat{\theta}_1)}{n_1} + \frac{\hat{\theta}_2(1-\hat{\theta}_2)}{n_2}}}, \tag{6.11}$$

donde $\hat{\theta}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$, $\hat{\theta}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_j$, $\mathbf{X} = (X_1, \ldots, X_{n_1})$ y $\mathbf{Y} = (Y_1, \ldots, Y_{n_2})$. En este caso, se rechaza la hipótesis H para valores de $T(\mathbf{X}, \mathbf{Y})$ mayores que z_α; donde z_α es tal que $\alpha = P(Z > z_\alpha)$ y Z es una variable Normal estándar. Esto es posible dado que, bajo la hipótesis H, $T(\mathbf{X}, \mathbf{Y}) \xrightarrow{d} N(0, 1)$.

Ejemplo 6.5. Para determinar el comportamiento de la prueba de hipótesis de diferencia de proporciones, $H : \theta_A - \theta_B = 0$ versus $K : \theta_A - \theta_B < 0$, basada en la estadística (6.11), se genera una muestra aleatoria de cada una de las siguientes poblaciones: $Bin(50, 0.3)$ y $Bin(40, 0.5)$. Dado que los valores observados fueron $y_A = 18$ y $y_B = 22$, respectivamente, el valor del estadístico de prueba (6.11) es

$$t = \frac{\hat{\theta}_A - \hat{\theta}_B}{\sqrt{\frac{\hat{\theta}_A(1-\hat{\theta}_A)}{50} + \frac{\hat{\theta}_B(1-\hat{\theta}_B)}{40}}} = -1.8287 \tag{6.12}$$

Finalmente, asumiendo normalidad del estadístico $T(\mathbf{X}, \mathbf{Y})$, definido por la ecuación (6.11), se encuentra que $P(T(\mathbf{X}, \mathbf{Y}) < -1.8287) = 0.0337$ y, en consecuencia, se rechaza la hipótesis H para niveles de significación α mayores que 0.0337.

Ejercicios 6.4.

1. Para determinar el comportamiento de la prueba de hipótesis de diferencia de proporciones, $H : \theta_A - \theta_B = 0$ versus $K : \theta_A - \theta_B < 0$, basada en la estadística (6.11), se genera una muestra aleatoria de cada una de las siguientes poblaciones: $Bin(50, 0.3)$ y $Bin(40, 0.7)$. Dado que los valores observados fueron $y_A = 10$ y $y_B = 25$, respectivamente, halle el valor del estadístico de prueba y determine los valores de α para los cuales se rechaza la hipótesis H.

2. Suponga que, con los datos del ejemplo 6.5, se desea probar la hipótesis $H : \theta_A - \theta_B = 0$ versus $K : \theta_A - \theta_B \neq 0$, asumiendo que el estadístico

$$\chi^2 = \sum_{k=1}^{2} \frac{(Y_i - n_i\hat{\theta})^2}{n_i\hat{\theta}(1-\hat{\theta})} \tag{6.13}$$

tiene distribución Ji-cuadrado con 2 grados de libertad, bajo la hipótesis H, donde $Y_i \sim Bin(n_i, \theta_i)$, $i = 1, 2$; y $\hat{\theta} = \frac{Y_1+Y_2}{n_1+n_2}$ es un estimador de θ. Encuentre los valores de α para los cuales se rechaza H.

6.8 Pruebas de hipótesis referentes a varianzas

6.8.1 Pruebas de hipótesis referentes a las varianzas de una población

Supóngase que Y_i, $i = 1, \ldots, n$; es una muestra aleatoria, de tamaño n, de una distribución Normal con media θ y varianza σ^2, desconocidas. Para una prueba de hipótesis, de nivel de significancia α, de $H : \sigma^2 = \sigma_0^2$ versus $K : \sigma^2 > \sigma_0^2$, es conveniente considerar el siguiente estadístico de prueba:

$$T(\mathbf{X}) = \frac{(n-1)S^2}{\sigma_0^2}, \tag{6.14}$$

dado que, bajo la hipótesis H, este estadístico tiene distribución χ^2 con $n-1$ grados de libertad.

En consecuencia, la región de rechazo está determinada por $RR = \{\mathbf{x} \in \mathbb{R}^n : T(\mathbf{x}) \geq x^2_{n-1,\alpha}\}$, donde $x^2_{n-1,\alpha}$ es un número real tal que $\alpha = P[\chi^2_{n-1} \geq x^2_{n-1,\alpha}]$.

Si la hipótesis alternativa es $K : \sigma^2 \neq \sigma_0^2$, se puede utilizar el mismo estadístico de prueba y la región de rechazo queda determinada por el conjunto

$$RR = \{\mathbf{x} \in \mathbb{R}^n : T(\mathbf{x}) \leq x^2_{n-1,\,\alpha/2}\} \cup \{\mathbf{x} \in \mathbb{R}^n : T(\mathbf{x}) \geq x^2_{n-1,1-\alpha/2}\}.$$

Ejemplo 6.6. Un distribuidor de tornillos desea establecer, con un nivel confianza de 0.95, si la varianza del diámetro de cierto tipo de tornillos, distribuidos por su empresa, es mayor que 0.02 mm^2; para lo cual propone la siguiente prueba de hipótesis: $H : \sigma^2 = 0.02$ versus $K : \sigma^2 > 0.02$.

1. Si obtiene una muestra aleatoria del diámetro de 9 tornillos y observa una varianza muestral $s^2 = 0.025$, entonces,

   $$T(\mathbf{x}) = \frac{(n-1)s^2}{\sigma_0^2} = \frac{8 \times 0.025}{0.02} = 10$$

   y no rechazaría la hipótesis H, pues $T(\mathbf{x}) \notin RR = (x^2_{0.05,8}, \infty) = (15.5073, \infty)$.

2. Si obtiene una muestra del diámetro de 31 tornillos y observa una varianza muestral $s^2 = 0.035$, entonces,

$$T(\mathbf{x}) = \frac{(n-1)s^2}{\sigma_0^2} = \frac{30 \times 0.035}{0.02} = 52.5.$$

En este caso, se rechaza la hipótesis H en favor de la hipótesis alternativa K; dado que $T(\mathbf{x}) \in RR = (x^2_{0.05,30}, \infty) = (43.7729, \infty)$.

6.8.2 Pruebas de hipótesis referentes a las varianzas de dos poblaciones

Otro caso de prueba de hipótesis, referente a las varianzas, se presenta cuando se tienen muestras aleatorias $X_i \sim N(\theta_1, \sigma_1^2)$, $i = 1, 2, \ldots, n_1$, y $Y_i \sim N(\theta_2, \sigma_2^2)$, $i = 1, 2, \ldots, n_2$, con medias y varianzas desconocidas. En este caso, es común encontrar dos tipos de pruebas de hipótesis dadas por $H : \sigma_1^2 = \sigma_2^2$, como hipótesis nula, y cada una de las siguientes alternativas $K_1 : \sigma_1^2 > \sigma_2^2$ o $K_2 : \sigma_1^2 \neq \sigma_2^2$. En este caso se considera como estadístico de prueba el cociente

$$T(\mathbf{X}, \mathbf{Y}) = S_1^2/S_2^2, \tag{6.15}$$

el cual, bajo la hipótesis H, tiene distribución F con $n_1 - 1$ grados de libertad en el numerador y $n_2 - 1$ grados de libertad en el denominador.

Así, si la hipótesis alternativa es $K : \sigma_1^2 > \sigma_2^2$, con un nivel de significancia α, se rechaza la hipótesis H para valores de $T(\mathbf{X}, \mathbf{Y}) \geq f_{\alpha, n_1-1, n_2-1}$, siendo f_{α, n_1-1, n_2-1} un número real tal que $\alpha = P(F_{n_2-1}^{n_1-1} \geq f_{\alpha, n_1-1, n_2-1})$.

Si la hipótesis alternativa es $K : \sigma_1^2 \neq \sigma_2^2$, entonces, con un nivel de significación α, se rechaza la hipótesis H para valores de $T(\mathbf{X}, \mathbf{Y}) \geq f_{\frac{\alpha}{2}, n_1-1, n_2-1}$, y para valores de $T(\mathbf{X}, \mathbf{Y}) \leq f_{1-\frac{\alpha}{2}, n_1-1, n_2-1}$, siendo $f_{\frac{\alpha}{2}, n_1-1, n_2-1}$ y $f_{1-\frac{\alpha}{2}, n_1-1, n_2-1}$ números reales tales que

$$\frac{\alpha}{2} = P(F_{n_2-1}^{n_1-1} \geq f_{\frac{\alpha}{2}, n_1-1, n_2-1}) = P[F_{n_1-1}^{n_2-1} \geq (f_{\frac{\alpha}{2}, n_1-1, n_2-1})^{-1}].$$

Ejemplo 6.7. Un distribuidor de tornillos desea establecer, con una confianza de 0.95, si la varianza del diámetro de los tornillos, producidos por dos empresas, A y B, es diferente. Para ello, plantea la siguiente prueba de hipótesis:

$$H : \sigma_A^2 = \sigma_B^2 \quad \text{versus} \quad \sigma_A^2 \neq \sigma_B^2. \tag{6.16}$$

El distribuidor de tornillos observa una varianza muestral $s_A^2 = 1.05$ en una muestra aleatoria del diámetro de 13 tornillos de la empresa A y una varianza muestral $s_B^2 = 0.85$ en una muestra de los diámetros de 10 tornillos de la empresa B. Dado que

$$T(\mathbf{X}, \mathbf{Y}) = S_A^2/S_B^2 \sim F_9^{12}, \tag{6.17}$$

entonces, la región de rechazo está determinada por el conjunto $RR = (0, \frac{1}{3.44}] \cup [3.87, \infty)$. Así, $T(\mathbf{x}, \mathbf{y}) = \frac{1.05}{0.85} = 1.235 \notin RR$, lo cual indica que no existe evidencia estadística para afirmar que la varianza del diámetro de los tornillos producidos por las empresas es diferente.

Ejercicios 6.5.

1. Dos muestras aleatorias A y B fueron generadas de distribuciones Normales, obteniéndose los siguientes datos:

   Muestra A: 2.2, 3.1, 4.2, 2.1, 2.9, 6.2, 4.1, 1.1
   Muestra B: 8.4, 6.6, 6.1, 6.6, 6.8, 5.2, 7.0, 5.6, 5.7

   Considerando $\alpha = 0.05$,

   a) Pruebe la hipótesis $H : \sigma_A^2 = 16$ versus $K_1 : \sigma_A^2 < 16$.
   b) Pruebe la hipótesis $H : \sigma_B^2 = 1$ versus $K_1 : \sigma_B^2 > 1$.
   c) Asuma que $\mu_A = 3$ y $\mu_B = 6$ y pruebe las hipótesis formuladas en los literales anteriores.

2. Con las muestras dadas en el primer ejercicio y el mismo nivel de significación, desarrolle las pruebas de hipótesis de

   a) $H : \sigma_A^2 = \sigma_B^2$ versus $K_1 : \sigma_A^2 > \sigma_B^2$.
   b) $H : \sigma_A^2 = \sigma_B^2$ versus $K_2 : \sigma_A^2 \neq \sigma_B^2$.

3. De las poblaciones dadas en el primer ejercicio se generaron dos muestras aleatorias A y B, cada una de cien observaciones, encontrándose los siguientes resultados:

   Muestra A: $\bar{X}_A = 2.81$, $s_A = 2.08$.
   Muestra B: $\bar{X}_B = 6.03$, $s_B = 0.97$.

   Pruebe las hipótesis planteadas en el primer ejercicio.

4. Si las muestras A y B, consideradas en esta sección, fueron generadas de distribuciones normales $N(3, 16)$ y $N(6, 1)$, ¿qué se puede concluir en cada uno de los ejercicios anteriores?

6.9 Potencia de una prueba

La potencia de la prueba de una hipótesis $H : \theta = \theta_0$ versus $K : \theta = \theta_1$ es la probabilidad de rechazar H cuando K es verdadera. Así, si δ denota la función indicadora

$$\delta(\mathbf{X}, \theta) = \begin{cases} 1 & \text{si } T(\mathbf{X}) \in RR \\ 0 & \text{si } T(\mathbf{X}) \notin RR \end{cases}$$

la función de potencia está definida por

$$\begin{aligned} \beta(\theta_1, \delta) &= P_{\theta_1}[\text{rechazar } H] \\ &= P_{\theta_1}[\delta = 1]. \end{aligned}$$

Ejemplo 6.8. Sea X_i, $i = 1, 2, \ldots, n$, una muestra aleatoria, de tamaño n, de una distribución normal con media θ desconocida y varianza σ^2 conocida. En la prueba de hipótesis de $H : \theta = 0$ versus $K : \theta = \theta_1$, $\theta_1 > 0$, es natural rechazar la hipótesis H para valores grandes de $\bar{X}$. Sin embargo, dado que una función estrictamente creciente de un estadístico suficiente, es a su vez, un estadístico suficiente que genera la misma familia de regiones de rechazo, es conveniente considerar el estadístico $T(\mathbf{X}) = \frac{\sqrt{n}\bar{X}}{\sigma}$, el cual, bajo la hipótesis H, tiene distribución Normal estándar; dado que para un coeficiente de confianza de $1 - \alpha$, se rechaza H para valores mayores que z_α.

En este caso, la función de potencia de la prueba toma una forma simple:

$$\begin{aligned} \beta(\theta_1, \delta) &= P_{\theta_1}\left(\frac{\sqrt{n}\,\bar{X}}{\sigma} \geq z_\alpha\right) \\ &= P_{\theta_1}\left(\frac{\sqrt{n}\,(\bar{X} - \theta_1)}{\sigma} \geq z_\alpha - \frac{\sqrt{n}\,\theta_1}{\sigma}\right) \end{aligned}$$

$$
\begin{aligned}
&= 1 - P_{\theta_1}\left(Z < z_\alpha - \frac{\sqrt{n}\,\theta_1}{\sigma}\right) \\
&= 1 - \phi\left(z_{(\alpha)} - \frac{\sqrt{n}\,\theta_1}{\sigma}\right) \\
&= \phi\left(\frac{\sqrt{n}\,\theta_1}{\sigma} - z_\alpha\right).
\end{aligned}
$$

Si la hipótesis alternativa es $K : \theta > 0$, la potencia de la prueba de hipótesis es una función de θ, definida en $[0, \infty)$ por la ecuación

$$
\beta(\theta, \delta) = \phi\Big(\frac{\sqrt{n}\theta}{\sigma} - z_\alpha\Big), \tag{6.18}
$$

donde $\beta(\theta, \delta)$ es una función de $\sqrt{n}\,\theta/\sigma$, continua y creciente, que tiende a α cuando $\sqrt{n}\theta/\sigma$ tiende a cero, y a uno cuando $\sqrt{n}\theta/\sigma$ tiende a más infinito. Así, dado θ, si se desea una potencia β mayor que algún valor β_0, el tamaño n de la muestra debe ser tal que $\frac{\sqrt{n}\theta}{\sigma} - z_\alpha > z_{\beta_0}$. Ahora, si se desea que esta desigualdad se cumpla para todo θ mayor que algún número Δ, real positivo, entonces, n es el menor número entero que satisface esta desigualdad:

$$
\frac{\sqrt{n}\Delta}{\sigma} - z_\alpha > z_{\beta_0},
$$

es decir, $n > \frac{\sigma^2}{\Delta^2}(z_{\beta_0} + z_\alpha)^2$.

6.9.1 Función de potencia

La *función de potencia* de una prueba de hipótesis $H : \theta \in \Theta_0$ versus $K : \theta \in \Theta_1$, denotada por $\beta(\theta, \delta)$, está definida por

$$
\begin{aligned}
\beta(\theta, \delta) &= P_\theta[\text{rechazar } H] \\
&= P_\theta[\delta = 1]
\end{aligned} \tag{6.19}
$$

para todo θ en el espacio paramétrico Θ.

Si $\theta \in \Theta_0$, entonces, $\beta(\theta, \delta)$ es la probabilidad de error tipo I. Si $\theta \in \Theta_1$, entonces, $\beta(\theta, \delta)$ es la probabilidad de aceptar K, dado que K es verdadero.

En este caso, $\beta(\theta, \delta)$ es uno menos la probabilidad de error tipo II, notada $\beta(\theta)$; es decir,

$$\beta(\theta, \delta) = \begin{cases} \alpha(\theta) & \text{si } \theta \in \Theta_0 \\ 1 - \beta(\theta) & \text{si } \theta \in \Theta_1 \end{cases}$$

Ejemplo 6.9. Suponga que se observa una muestra aleatoria $X_i \sim N(\theta, \sigma^2)$, $i = 1, \ldots, n$; de tamaño n, con θ desconocido y σ^2 conocido. Se desea probar la hipótesis $H : \theta = 0$ versus $K : \theta = \theta_1$; $\theta_1 > 0$. Si la prueba se realiza con un nivel de significancia α, considerando como estadístico de prueba $T(X) = \sqrt{n}\bar{X}/\sigma$, entonces,

$$\begin{aligned} \beta(\theta_1, \delta) &= P_{\theta_1}\left(\frac{\sqrt{n}\bar{X}}{\sigma} > z_\alpha\right) \\ &= 1 - \Phi\left(z_\alpha - \frac{\sqrt{n}\theta_1}{\sigma}\right). \end{aligned}$$

para $\theta_1 > 0$.

Así, la probabilidad de error tipo II es $\Phi\left(z_\alpha - \frac{\sqrt{n}\theta_1}{\sigma}\right)$. Si se desea probabilidades de error de tipo I y II iguales a α, el tamaño de la muestra debe ser tal que

$$z_\alpha - \frac{\sqrt{n}\theta_1}{\sigma} = -z_\alpha.$$

Así, los errores de tipo I y II son menores que α, siempre que:

$$n > \frac{4\sigma^2}{\theta_1^2} z_\alpha^2.$$

La figura 6.2 muestra la función de potencia $\beta(\theta_1, \delta)$, de la prueba de hipótesis propuesta en el ejemplo 6.9, para $\alpha = 0.05$, $n = 9$ y $\sigma^2 = 9$.

Ejemplo 6.10. Supóngase que X_i, $i = 1, 2, \ldots, n$; denota el tiempo de vida de n bombillas, que estos son independientes y que X_i tiene distribución exponencial, con función de densidad dada por $f_\lambda(x) = \lambda e^{-\lambda x} I_{(0,\infty)}(x)$, para todo i.

Dado que $2\lambda \sum_{i=1}^n X_i \sim \chi^2_{(2n)}$, la región de rechazo, de nivel de significancia α, para la prueba de hipótesis $H : \theta \leq \theta_0$ versus $K : \theta > \theta_0$, relacionada con

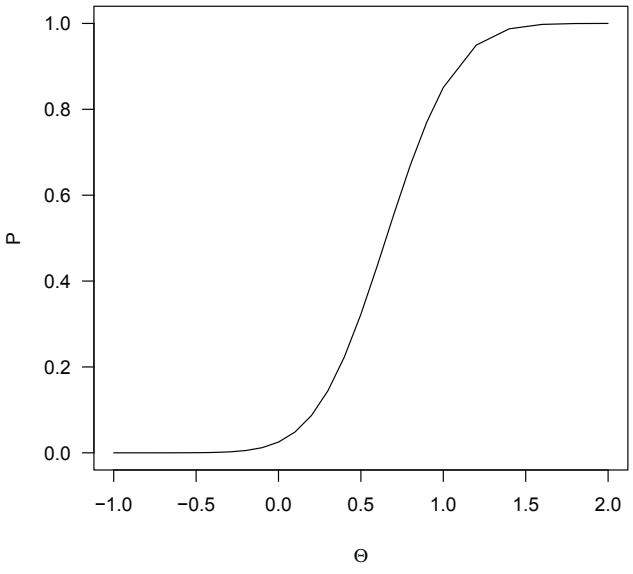

Figura 6.2. Función de potencia.

el tiempo de vida medio, $\theta = 1/\lambda$, de las bombillas, está determinada por

$$\alpha = P_{\theta_0}\left[2\lambda_0 \sum_{i=1}^{n} X_i \geq \chi^2_{\alpha,2n}\right] = P_{\theta_0}\left[\bar{X} \geq \frac{\theta_0}{2n}\chi^2_{\alpha,2n}\right].$$

Así, la función de potencia queda determinada por

$$\begin{aligned}
\beta(\theta,\delta) &= P_\theta\left[\bar{X} \geq \frac{\theta_0}{2n}\,\chi^2_{\alpha,2n}\right] \\
&= P_\theta\left[\frac{2n\bar{X}}{\theta} \geq \frac{\theta_0}{\theta}\,\chi^2_{\alpha,2n}\right] \\
&= P_\theta\left[\chi^2_{(2n)} \geq \frac{\theta_0}{\theta}\,\chi^2_{\alpha,2n}\right].
\end{aligned}$$

Obsérvese que si:

1. $\theta = \theta_0$, entonces $\beta(\theta,\delta) = \alpha$.

2. $\frac{\theta_0}{\theta}\chi^2_{\alpha,2n} \to 0$ y, por lo tanto, $\beta(\theta,\delta) \to 1$.

Ejemplo 6.11. Supóngase que $X_1, \ldots, X_n$ son n variables aleatorias independientes y uniformemente distribuidas, en un intervalo $(0,\theta)$. Como la función de distribución del estadístico $M_{(n)} = \text{Max}\{X_1, \ldots, X_n\}$ está dada por

$$F(y) = \begin{cases} 0 \quad \text{si} \quad y \leq 0 \\ P_\theta[X_1 \leq y]^n \quad \text{si} \quad y \in (0,\theta) \\ 1 \quad \text{si} \quad y \geq \theta \end{cases}$$

su función de densidad es $f(y) = n[F_1(y)]^{n-1}F_1'(y)$, donde

$$F_1(y) = \begin{cases} 0 \quad \text{si} \quad y \leq 0 \\ \frac{y}{\theta} \quad \text{si} \quad y \in (0,\theta) \\ 1 \quad \text{si} \quad y \geq \theta \end{cases}$$

Así, $f(y) = \frac{ny^{n-1}}{\theta^n}$ en el intervalo $(0,\theta)$ y $f(y) = 0$ si $y \notin (0,\theta)$.

Dado un nivel de significación α, la región de rechazo, para la prueba de hipótesis de $H : \theta \leq \theta_0$ versus $K : \theta > \theta_0$, está determinada por

$$\alpha = P_{\theta_0}\{M_{(n)} \geq c\} = \int_c^{\theta_0} \frac{ny^{n-1}}{\theta_0^n} dy = 1 - \left(\frac{c}{\theta_0}\right)^n, \quad \text{con} \quad c \in (0,\theta_0).$$

En consecuencia, $c = \theta_0(1-\alpha)^{1/n}$ y la función de potencia, para esta prueba, está definida por

$$\beta(\theta,\delta) = P_\theta\{M_{(n)} \geq c\} = \int_c^{\theta} \frac{ny^{n-1}}{\theta^n} dy = 1 - \left(\frac{c}{\theta}\right)^n.$$

Reemplazando c por su equivalente, en términos de α, se halla que

$$\beta(\theta,\delta) = 1 - \left(\frac{\theta_0}{\theta}\right)^n (1-\alpha), \quad \theta > \theta_0.$$

Finalmente, a partir de la expresión anterior, se halla que $\beta(\theta_0,\delta) = \alpha$.

6.9.2 El p-valor

En la presentación de informes, un analista de datos puede afirmar que rechaza la hipótesis H con un nivel de significación $\alpha = 0.05$. Sin embargo, si los lectores de este informe están interesados en saber qué se puede concluir para otros niveles de significación como $\alpha = 0.01$ o $\alpha = 0.001$, no les sería posible. En realidad, los lectores del informe están impedidos para decidir si aceptan o rechazan H, para valores de α menores que 0.05. Por esta razón, es conveniente presentar los informes de modo que el lector quede informado del nivel de significación real de la prueba. Esto es posible mediante el p-valor (p_v) que, en el caso de pruebas de hipótesis de la forma:

1. $H : \theta = \theta_0$ y $K : \theta > \theta_0$, se define como $p_v = P_{\theta_0}[T(\boldsymbol{X}) \geq T(\mathbf{x}_0)]$, donde $T(\boldsymbol{X})$ es el estadístico de prueba, x_0 es el vector de valores observados en la muestra y $T(x_0)$ el valor del estadístico de prueba calculado en $X = x_0$.

2. $H : \theta = \theta_0$ y $K : \theta < \theta_0$, se define como $p_v = P_{\theta_0}[T(\boldsymbol{X}) \leq T(\mathbf{x}_0)]$, donde $T(\boldsymbol{X})$, x_0 y $T(x_0)$ están definidos como en la prueba anterior.

En el caso de pruebas de hipótesis de la forma $H : \theta = \theta_0$ versus $K : \theta \neq \theta_0$, si bajo la hipótesis H el estadístico de prueba tiene distribución simétrica (Normal o t de Student), entonces,

$$p_v = 2P_{\theta_0}[T(\boldsymbol{X}) \geq |T(\mathbf{x}_0)|\ ],$$

donde $T(\mathbf{X})$ es el estadístico de prueba y $\mathbf{x}_0$ los valores observados en la muestra.

Ejemplo 6.12. En el ejemplo 6.1, una muestra de cien sobres de azúcar empacados por *Azucarco S.A*, tienen un peso promedio de 5.2 gramos y una desviación estándar de 0.4 gramos. En la prueba de hipótesis de $H : \theta = 5.13$ versus $K : \theta > 5.13$,

$$\begin{aligned} p_v &= P(\bar{X} \geq 5.2) \\ &= P\Big(Z \geq \frac{10(5.2 - 5.13)}{0.4}\Big) \\ &= P(Z \geq 1.75) = 0.040 \end{aligned}$$

En consecuencia, se rechaza la hipótesis H_0 para niveles de significación (α) mayores que 0.040.

Ejemplo 6.13. En el ejemplo 6.12, para la prueba de hipótesis determinada por $H : \theta = 5.13$ versus $K : \theta \neq 5.13$, $p_v = 2P(\bar{X} \geq 5.2) = 0.08$. En este caso, para $\alpha = 0.05$, no se rechaza la hipótesis H_0, dado que $\bar{x} = 5.2$ no cae en la región de rechazo.

Ejemplo 6.14. Una empresa de tornillos desea establecer, con una confiabilidad de 0.95, si la varianza del diámetro de cierto tipo de tornillos de su producción es mayor que 0.02 mm^2, propone la siguiente prueba de hipótesis: $H : \sigma^2 = 0.02$ versus $K : \sigma^2 > 0.02$. Si en una muestra aleatoria del diámetro de 9 tornillos se observa una varianza $s^2 = 0.025$, el valor del estadístico de prueba es:

$$T(\mathrm{x}) = \frac{(n-1)s^2}{\sigma_0^2} = 10 \tag{6.20}$$

y, el p-valor correspondiente es $P_v = P(T(X) \geq 10) = P(\chi_8^2 \geq 10) = 0.265$. En consecuencia, para valores de α como 0.05, 0.025, 0.001 se rechaza $H : \sigma^2 = 0.02$.

Ejercicios 6.6.

1. Asuma que se observa una muestra aleatoria de tamaño n, $X_i \sim N(\theta, 4)$, $i = 1, \ldots, n$; con θ desconocido.

   a) Si $n = 10$, en la prueba hipótesis $H : \theta = 0$ versus $K : \theta > 0$,

      1) Hallar el valor de la función de potencia en $\theta = 1,\ 2,\ 5$.
      2) Calcular el error de tipo II en $\theta = 2,\ 3,\ 5$.

   b) Si la hipótesis alternativa es $K : \theta = 2$, determine los valores de n para los cuales la función de potencia es igual a 0.1, 0.5 y 0.8.

2. En el ejemplo 6.10, asuma que, con un nivel de significación del 95 %, se desea probar la hipótesis $H : \theta \leq 2$ versus $K : \theta > 2$. Encuentre la gráfica de

   a) La función de potencia.

   b) El error de tipo II.

3. En el ejemplo 6.11, si $n = 10$ y $\alpha = 0.05$, determine:

   a) La región de rechazo en la prueba de $H : \theta \leq 5$ versus $\theta > 5$.

b) La función de potencia en la prueba de $H : \theta \leq 5$ versus $\theta > 5$.

4. Determine la región de rechazo, de nivel de significancion α, para la prueba de hipótesis $H : \theta = \theta_0$ versus $K : \theta > \theta_0$, y las funciones de potencia y de error de tipo II, asumiendo una muestra aleatoria, de tamaño uno, de una distribución:

   a) Exponencial con media θ.

   b) Uniforme en un intervalo $(0, \theta)$.

6.10 Intervalos de confianza y pruebas de hipótesis

Considérese el caso de una prueba de hipótesis $H : \theta = \theta_0$ versus $k : \theta \neq \theta_0$ para la media de una distribución normal con varianza conocida, definida en la sección 6.5.2. Para todo θ_0, en el espacio paramétrico Θ, se puede determinar un conjunto de valores del estadístico $\bar{X}$ para los cuales no se rechaza la hipótesis $H : \theta = \theta_0$, en un nivel de significancia α. Dado que para este nivel, la región de rechazo está dada por $RR(\bar{X}) = \left[\bar{x} \in \mathbb{R} : \frac{\sqrt{n}|\bar{x}-\theta_0|}{\sigma} \geq z_{\alpha/2}\right]$, el conjunto de los $\bar{x}$ para los cuales no se rechaza la hipótesis H es

$$A(\bar{X}) = \left[\bar{x} \in \mathbb{R} : |\bar{x} - \theta_0| < \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right].$$

Ahora se puede preguntar, dado $\bar{x}$, ¿cuál es el conjunto de parámetros $C(\bar{x})$ para los cuales, a un nivel de significación α, se acepta la hipótesis $H : \theta = \theta_0$? A partir de la definición de la región de rechazo, aquí especificada, se concluye que no se rechaza la hipótesis $H : \theta = \theta_0$, si y solo si

$$\bar{x} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \leq \theta_0 \leq \bar{x} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}. \tag{6.21}$$

Así, dado $\bar{x}$, el conjunto de parámetros para los cuales no se rechaza la hipótesis H, está dado por el conjunto $C(\bar{x}) = \{\theta \in \mathbb{R} : |\theta - \bar{x}| \leq \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\}$.

Para finalizar esta sección, el lector puede representar gráficamente la relación definida por $C(\bar{x})$ y $A(\bar{X})$ para $\Theta = R$.

6.11 Hipótesis compuestas

En esta sección se dan tres ejemplos del planteamiento de las hipótesis compuestas.

Ejemplo 6.15. Supóngase que X es el número de artículos defectuosos en una muestra aleatoria de N objetos, tomados sin reemplazo, de un lote que contiene $b = N\theta$ artículos defectuosos, con θ desconocido. Un comprador de estos lotes considera que $b_0 = N\theta_0$ artículos defectuosos o más, en cada lote, es inaceptable y considera las hipótesis:

$$H : b \geq b_0 \text{ versus } K : b < b_0. \tag{6.22}$$

$$H : b \leq b_0 \text{ versus } K : b > b_0. \tag{6.23}$$

En (6.22), si el número de artículos defectuosos en la muestra es pequeño, entonces se rechaza H y se acepta el lote. En (6.23), si el número de artículos defectuosos en la muestra es pequeño, entonces no se rechaza H y se acepta el lote de artículos.

Es natural que el comprador se decida por la hipótesis (6.22), puesto que $\alpha = P_{\theta_0}[\text{rechazar } H]$ es la probabilidad de aceptar un lote que tenga más de b_0 artículos defectuosos. En (6.23), $\alpha = P_{\theta_0}[\text{rechazar } H]$ es la probabilidad de no aceptar un lote, dado que tiene menos de b_0 artículos defectuosos.

Ejemplo 6.16. Supóngase que se desea determinar si la vida media de las bombillas producidas por una fábrica es mayor que $1/\lambda_0$. Si X_i, $i = 1, \ldots, n$, es una muestra aleatoria, de tamaño n, de tiempos de vida de las bombillas, entonces, asumiendo que los tiempos tienen una distribución exponencial con media $1/\lambda$, un comprador de bombillas quiere que sea pequeña la probabilidad α de aceptar que dichas bombillas tienen una duración mayor que $1/\lambda_0$, cuando en realidad tienen una duración menor. En consecuencia, se plantean las siguientes hipótesis:

$$H : \frac{1}{\lambda} \leq \frac{1}{\lambda_0} \quad \text{versus} \quad K : \frac{1}{\lambda} > \frac{1}{\lambda_0}.$$

Se puede determinar un valor de $\alpha = P_{\frac{1}{\lambda_0}}[\text{rechazar } H]$ como la probabilidad de aceptar que las bombillas tienen un tiempo medio de vida mayor que $1/\lambda_0$, cuando en realidad es menor.

Ejemplo 6.17. El interés de un comprador está en que la probabilidad de recibir tornillos, que tengan alta variabilidad de su diámetro, sea lo más pequeña posible. En consecuencia, las hipótesis para plantear son

$$H : \sigma^2 \geq \sigma_0^2$$
$$K : \sigma^2 < \sigma_0^2$$

Así, se podrá controlar la probabilidad, $\alpha = P_H(\text{rechazar}\, H)$, de aceptar que la producción de tornillos tiene una varianza en su diámetro menor que σ_0^2, cuando en realidad tiene una varianza mayor.

Ejercicios 6.7.

1. Se desea probar la hipótesis $H : \theta_1 - \theta_2 = 0$ frente a la alternativa $K : \theta_1 - \theta_2 \neq 0$. Si el p-valor del estadístico de prueba está entre 0.005 y 0.01, ¿para cuáles de los siguientes valores de α se rechaza la hipótesis H?

   $\alpha_1 = 0.005$, $\alpha_2 = 0.0025$, $\alpha_3 = 0.01$, $\alpha_4 = 0.02$.

2. Sea $X_i \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$, una muestra aleatoria, de tamaño n, de una distribución normal con varianza conocida. Si se desea probar la hipótesis $H : \theta = \theta_0$ versus $K : \theta > \theta_0$ y el valor observado del estadístico $T(X) = \frac{\sqrt{n}(\bar{X}-\theta_0)}{\sigma}$ de prueba es de 2.17, ¿para qué valores de α no se rechaza la hipótesis nula?

3. En un estudio comparativo del pH medio de dos ríos A y B, se hallan los siguientes datos: $n_A = 50$, $n_B = 40$, $\bar{x}_A = 8.3$, $\bar{x}_B = 7.9$, $s_A^2 = 12.9$ y $s_B^2 = 9.7$. ¿Para qué valores de α se rechaza $H : \theta_A = \theta_B$ si la hipótesis alternativa es $K : \theta_A \neq \theta_B$, y para qué valores de α, si la hipótesis alternativa es $K : \theta_A > \theta_B$?

4. Una empresa de maquinaria produce un 10 % de artículos defectuosos. Para establecer un proceso de control de calidad, la gerencia establece que si en 350 artículos, 45 resultan defectuosos, la máquina está produciendo más artículos defectuosos que lo establecido y que, por tanto, debe ser reparada. ¿En qué nivel de significación se ha tomado esta decisión?

5. Cuatro estudiantes están interesados en demostrar que el promedio de las notas de su facultad es de 4.0. Cada uno de ellos obtiene una muestra aleatoria y determina un p-valor; hallando $p_1 = 0.045$, $p_2 = 0.035$, $p_3 = 0.011$ y $p_4 = 0.0001$, ¿qué podrá concluir cada uno de los estudiantes, si $\alpha = 0.01, 0.02, 0.0004$?

6. Supóngase que se tiene una observación de una variable aleatoria X con función de densidad $f_\theta(x) = \frac{1}{\theta}\exp(-\frac{1}{\theta}x)I_{(0,\infty)}(x)$. Si se desea probar la hipótesis $H : \theta = 2$ versus $K : \theta = \theta_1$, $\theta_1 > 2$, encuentre el número real c que determina una región de rechazo de nivel de significación α. Halle la función de potencia. Si el valor observado de la variable aleatoria es $x = 3$, entonces, encuentre el p-valor de la prueba. ¿Se rechaza H a un nivel de significación α de 0.05?

7. Supóngase que se desea probar la hipótesis $H : \sigma_1^2 = \sigma_2^2$ versus $K : \sigma_1^2 < \sigma_2^2$ y que el valor observado de $F = s_1^2/s_2^2$ es de 0.065. Si F tiene 5 grados de libertad en el numerador y 4 en el denominador, ¿qué se puede afirmar del valor p? ¿Para qué valores de α se rechaza la hipótesis?

8. En un experimento, para determinar diferencias en el pH del agua de dos ríos, se obtuvieron los siguientes datos:

Río A	8.33	8.19	7.72	7.82	7.83	8.26	7.62	8.07	7.55	7.81	7.51
Río B	7.62	7.64	7.81	7.25	7.90	7.76	7.60	7.53	7.69		

   a) Hallar intervalos de confianza para las varianzas poblacionales de cada uno de los ríos, con una confiabilidad del 99 %.

   b) Probar la hipótesis: $H : \sigma_a^2 = \sigma_b^2$ versus $K : \sigma_a^2 \neq \sigma_b^2$. ¿Qué puede concluir para $\alpha = 0.005$ y para $\alpha = 0.01$.

9. En un proceso de producción de tornillos, si una máquina produce más del 3 % de artículos defectuosos se debe parar la producción. Si se desea una confiabilidad del 95 %, ¿cuántos tornillos defectuosos deben aparecer en una muestra de trescientos para detener la producción?

10. Sea X_i, $i = 1, ..., n$, una muestra aleatoria, de tamaño n, de una distribución $N(\theta, \sigma^2)$, con media y varianza desconocidas. Para la prueba de hipótesis $H : \sigma^2 = \sigma_0^2$ versus $K : \sigma^2 \neq \sigma_0^2$, de nivel de significación α, halle la función de potencia y grafíquela.

11. Una muestra aleatoria conformada por la longitud de nueve tornillos de la producción de una empresa, tiene una media $\bar{x} = 7.6$. Con $\alpha = 0.05$ y asumiendo que la muestra proviene de una distribución Normal con $\sigma^2 = 4$, efectúe una prueba de hipótesis para $H : \theta = 6$ cm frente a la hipótesis alternativa $H : \theta \neq 6$ cm, utilizando la distribución Ji-cuadrado. Halle la función de potencia.

Capítulo 7

Razón de verosimilitud

Este capítulo incluye dos secciones. La primera, titulada "Lema de Neyman-Pearson", incluye el concepto de pruebas uniformemente más potentes, teoremas fundamentales y algunos ejercicios. La segunda, titulada "Pruebas de razón de verosimilitudes", incluye algunos ejemplos que ilustran con algún detalle el proceso de pruebas de hipótesis a partir del cociente de las funciones de verosimilitud determinadas por las hipótesis bajo estudio.

7.1 Lema de Neyman-Pearson

Supóngase que se desea probar $H : \theta = \theta_0$ versus $K : \theta = \theta_1$; $\theta_1 > \theta_0$, de tamaño n, basados en una muestra aleatoria X_i, $i = 1, 2, \ldots, n$, proveniente de una función de densidad (o de probabilidad) p, con parámetro θ. Para un nivel de significación α, la prueba de máxima potencia en $\theta = \theta_1$ tiene región de rechazo dada por $L(\mathbf{x}, \theta_0, \theta_1) = p(\mathbf{x}, \theta_1)/p(\mathbf{x}, \theta_0)$, donde $p(\mathbf{x}, \theta_i)$, $i = 0, 1$; es la función de densidad (o de probabilidad) de $\mathbf{X} = (X_1, \ldots, X_n)$ dado $\theta = \theta_i$. Esta prueba tiene una región de rechazo de la forma $\{\mathbf{x} \in \mathbb{R} : T(\mathbf{x}) \geq c\}$, para algún estadístico de prueba T determinado por

$$\frac{p(\mathbf{x}, \theta_1)}{p(\mathbf{x}, \theta_0)} \geq k; \ k \text{ constante.} \tag{7.1}$$

Dado que el error de tipo II es uno menos la potencia, esta es la prueba de nivel de significación α, con menor probabilidad de error de tipo II.

Supóngase que se desea probar la hipótesis $H : \theta = \theta_0$ versus $K : \theta > \theta_0$. Una prueba δ^*, de nivel de significación α, es uniformemente más potente para probar H versus K, si para todo θ_a, mayor que θ_0, δ^* es la prueba más potente de nivel de significación α para $H : \theta = \theta_0$ versus $K : \theta = \theta_a$.

Definición 7.1. Una prueba δ^*, de nivel de significación α, es uniformemente más potente para $H : \theta \in \Theta_0$ versus $K : \theta \in \Theta_1$, si y solo si, $\beta(\theta, \delta^*) \geq \beta(\theta, \delta)$ para todo $\theta \in \Theta_1$ y para toda prueba δ de nivel de significación α.

Ejemplo 7.1. Supóngase que se tiene una muestra aleatoria $X_i \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$, de tamaño n, con media desconocida y varianza conocida. La prueba de nivel de significación α, uniformemente más potente para $H : \theta = \theta_0$ versus $K : \theta > \theta_0$, se obtiene a partir de (7.1) considerando $\theta_1 > \theta_0$ arbitrario:

$$\begin{aligned} L(\mathbf{x}, \theta_0, \theta_1) &= \frac{p(\mathbf{x}, \theta_1)}{p(\mathbf{x}, \theta_0)} \\ &= \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i - \theta_1)^2 - \sum_{i=1}^{n}(x_i - \theta_0)^2\right)\right] \\ &= \exp\left[-\frac{1}{2\sigma^2}\left(2(\theta_0 - \theta_1)\sum_{i=1}^{n} x_i - n(\theta_0^2 - \theta_1^2)\right)\right]. \end{aligned} \tag{7.2}$$

Así pues, se rechaza H si el vector de valores de la muestra, $\mathbf{x} = (x_1, \ldots, x_n)$, es más probable bajo la hipótesis K que bajo la hipótesis H, es decir, si $L(\mathbf{x}, \theta_0, \theta_1) \geq k$, para algún $k > 1$.

Despejando $\bar{x}$ en $L(\mathbf{x}, \theta_0, \theta_1) \geq k$, se encuentra que la región de rechazo está determinada por el conjunto

$$RR = \left\{\mathbf{x} \in \mathbb{R}^n | \bar{x} \geq \frac{-2\sigma^2 \ln k - n\theta_1^2 + n\theta_0^2}{2n(\theta_0 - \theta_1)} = k'\right\}.$$

En consecuencia, L y $T(\mathbf{X}) = \bar{X}$ son estadísticos equivalentes en el sentido que $T(\mathbf{x}) \geq k'$, si y solo si, $L(\mathbf{x}, \theta_0, \theta_1) \geq k$. Este resultado se deriva directamente de (7.2), dado que $L(\mathbf{x}, \theta_0, \theta_1)$ es una función estrictamente creciente de $T(\mathbf{x}) = \bar{x}$. Así, la prueba uniformemente más potente está determinada por el conjunto $RR = \{\mathbf{x} \in \mathbb{R}^n : \bar{x} \geq k'\}$. En consecuencia, la prueba uniformemente más potente, de nivel de significación α, tiene región de rechazo:

$$RR = \left\{\mathbf{x} \in \mathbb{R}^n : \frac{\sqrt{n}(\bar{x} - \theta_0)}{\sigma} \geq z_\alpha\right\}; \quad \text{donde} \quad z_\alpha = \frac{\sqrt{n}(k' - \theta_0)}{\sigma}.$$

Esto indica que se rechaza H siempre que $\bar{x} > \theta_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$.

Finalmente, dado que estos resultados son válidos para todo $\theta_1 \in \mathbb{R}$, mayor que θ_0, queda demostrado que L es la prueba uniformemente más potente para la prueba de hipótesis antes mencionada, pues θ_1 es un número real arbitrario en el intervalo (θ_0, ∞).

Ejemplo 7.2. Supóngase ahora que, a partir de la muestra del ejemplo 7.1, se desea desarrollar la prueba uniformemente más potente, de nivel de significación α, para $H : \theta = \theta_0$ versus $K : \theta < \theta_0$. A partir de (7.2) se sigue que para todo $\theta_1 < \theta_0$, $L(\mathbf{x}, \theta_0, \theta_1)$ es una función estrictamente decreciente de $T(\mathbf{x}) = \bar{x}$. Por tanto, para todo número real positivo k, existe un número real k' tal que $T(\mathbf{x}) \leq k'$, si y solo si, $L(\mathbf{x}, \theta_0, \theta_1) \geq k$. En consecuencia, la prueba uniformemente más potente, de nivel de significación α, está determinada por la región de rechazo $RR = \{\mathbf{x} \in \mathbb{R}^n : \frac{\sqrt{n}(\bar{x}-\theta_0)}{\sigma} \leq -z_\alpha\}$.

Teorema 7.1. Sea δ_k la función crítica de la prueba de hipótesis que rechaza $H : \theta = \theta_0$, si y solo si, la razón de verosimilitud es al menos k, donde $0 \leq k \leq \infty$. Sea δ la función crítica de una prueba cuyo tamaño no es mayor que el tamaño de δ_k, es decir, tal que

$$\beta(\theta_0, \delta) \leq \beta(\theta_0, \delta_k). \tag{7.3}$$

Entonces,

$$\beta(\theta_1, \delta) \leq \beta(\theta_1, \delta_k). \tag{7.4}$$

Demostración. Si $\delta_k(\mathbf{x}) = 1$, entonces, $\delta_k(\mathbf{x}) - \delta(\mathbf{x}) \geq 0$. En consecuencia, $p(\mathbf{x}, \theta_1)/p(\mathbf{x}, \theta_0) \geq k$ y

$$p(\mathbf{x}, \theta_1)(\delta_k(\mathbf{x}) - \delta(\mathbf{x})) \geq kp(\mathbf{x}, \theta_0)(\delta_k(\mathbf{x}) - \delta(\mathbf{x})). \tag{7.5}$$

Además, si $\delta_k(\mathbf{x}) = 0$, entonces, $p(\mathbf{x}, \theta_1)/p(\mathbf{x}, \theta_0) < k$, se obtiene nuevamente la desigualdad (7.5).

Ahora bien, integrando a ambos lados de (7.5), con respecto a $\mathbf{x}$, se obtiene que:

$$E_{\theta_1}(\delta_k(\mathbf{X})) - E_{\theta_1}(\delta(\mathbf{X})) \geq k[E_{\theta_0}(\delta_k(\mathbf{X})) - E_{\theta_0}(\delta(\mathbf{X}))].$$

Finalmente, teniendo en cuenta que $\beta(\theta, \delta) = E_\theta(\delta(\mathbf{X}))$, a partir de (7.3) se concluye (7.4), ya que $E_{\theta_0}(\delta_k(\mathbf{X})) - E_{\theta_0}(\delta(\mathbf{X})) \geq 0$. $\square$

Ejemplo 7.3. Sea $X_i \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, de una distribución Normal con media conocida y varianza desconocida. Se desea probar que

$$H : \sigma^2 = \sigma_0^2 \quad \text{versus} \quad K : \sigma^2 = \sigma_1^2; \;\; \sigma_1^2 > \sigma_0^2,$$

donde σ_0^2 representa la máxima varianza permitida en un proceso de producción. En este caso, la razón de verosimilitud, dada por (7.6), es una función estrictamente creciente de $\sum_{i=1}^n (x_i - \theta)^2$. En consecuencia, para todo $k > 0$ existe $k' > 0$ tal que $\sum_{i=1}^n (x_i - \theta)^2 \geq k'$ siempre que $L(\mathbf{x}, \sigma_0^2, \sigma_1^2) \geq k$.

$$\begin{aligned} L(\mathbf{x}, \sigma_0^2, \sigma_1^2) &= \frac{p(\mathbf{x}, \sigma_1^2)}{p(\mathbf{x}, \sigma_0^2)} \\ &= \frac{\sigma_0^n}{\sigma_1^n} \exp\left\{\frac{1}{2}\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right) \sum_{i=1}^n (x_i - \theta)^2\right\}. \end{aligned} \tag{7.6}$$

Así, dado que, bajo la hipótesis nula H, $T(\mathbf{x}) = \frac{1}{\sigma_0^2} \sum_{i=1}^n (x_i - \theta)^2 \sim \chi^2_{(n)}$, la región de rechazo de la prueba uniformemente más potente de nivel de significación α, está determinada por el conjunto

$$RR = \{\mathbf{x} \in \mathbb{R}^n : \frac{1}{\sigma_0^2} \sum_{i=1}^n (x_i - \theta)^2 \geq \chi^2_{n,\alpha}\},$$

donde $\mathbf{x} = (x_1, \ldots, x_n)$.

Obsérvese que $L(\mathbf{X}, \sigma_0^2, \sigma_1^2)$ y $T(\mathbf{X})$ son estadísticos equivalentes: $T(\mathbf{x}) \geq k'/\sigma_0^2$, si y solo si, $L(\mathbf{x}, \theta_0, \theta_1) \geq k$.

El error de tipo II, en esta prueba de hipótesis, está determinado por la expresión:

$$\begin{aligned} \beta &= P_{\sigma_1^2}\left(\frac{1}{\sigma_0^2} \sum_{i=1}^n (x_i - \theta)^2 < \chi^2_{n,\alpha}\right) \\ &= P_{\sigma_1^2}\left(\frac{1}{\sigma_1^2} \sum_{i=1}^n (x_i - \theta)^2 < \frac{\sigma_0^2}{\sigma_1^2} \chi^2_{n,\alpha}\right). \end{aligned}$$

Ejemplo 7.4. Sea $X_i \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, proveniente de una distribución normal, con media θ conocida

y varianza σ^2 desconocida. Supóngase que se desea desarrollar la prueba de hipótesis, uniformemente más potente, de nivel de significación α, para $H : \sigma^2 = \sigma_0^2$ versus $K : \sigma^2 = \sigma_1^2$; $\sigma_1^2 < \sigma_0^2$. En este caso, $L(\mathbf{x}, \sigma_0^2, \sigma_1^2)$ es una función estrictamente decreciente de $\sum_{i=1}^n (x_i - \theta)^2$ y la región de rechazo, de la prueba uniformemente más potente de nivel de significación α, está determinada por el conjunto

$$RR = \{\mathbf{x} \in \mathbb{R}^n : \frac{1}{\sigma_0^2} \sum_{i=1}^n (x_i - \theta)^2 \leq \chi^2_{n,1-\alpha}\}.$$

Ejemplo 7.5. Supóngase que $X \sim Geo(\theta)$. Para determinar la prueba uniformemente más potente, de nivel de significación α, para probar $H : \theta = \theta_0$ versus $K : \theta = \theta_1$; $\theta_0 < \theta_1$, se determina el conjunto de los $x \in Z^+$ tales que

$$L(x, \theta_0, \theta_1) = \left(\frac{\theta_1}{\theta_0}\right) \left(\frac{1-\theta_1}{1-\theta_0}\right)^{x-1} \geq k. \tag{7.7}$$

En efecto, dado que $\theta_1 > \theta_0$, entonces, L es una función estrictamente decreciente de $x \in Z^+$ y, por ende, la región de rechazo está determinada por el conjunto $RR = \{x \in Z^+ : x \leq k'\}$. Ahora bien, si k' es entero, esta es la región de rechazo, de la prueba uniformemente más potente de nivel de significación $\alpha = \sum_{j=1}^{k'} p_{\theta_0}(j)$, para la prueba de hipótesis de H versus K, donde p es la función de probabilidad Geométrica. Nótese que $0 < \alpha < 1$, si y solo si $0 < k \leq \frac{\theta_1}{\theta_0}$. Si $k > \frac{\theta_1}{\theta_0}$, $\alpha = 0$.

Teorema 7.2. Asuma que $0 \leq k < \infty$ y que δ_k está definida como en el teorema 7.1. Supóngase que δ es la función crítica de una prueba de hipótesis definida arbitrariamente en el conjunto $A = \{\mathbf{x} \in \mathbb{R}^n : L(\mathbf{x}, \theta_0, \theta_1) = k\}$ y tal que $\delta(\mathbf{x}) = \delta_k(\mathbf{x})$ en el conjunto $B = \{\mathbf{x} \in \mathbb{R}^n : L(\mathbf{x}, \theta_0, \theta_1) \neq k\}$. Entonces, δ es la prueba más potente, de nivel de significación $\alpha = P_{\theta_0}[\delta(\mathbf{x}) = 1]$, para probar $H : \theta = \theta_0$ versus $K : \theta = \theta_1$.

Demostración. Dado que $\delta(\mathbf{x}) = 1$ en un subconjunto de $\{\mathbf{x} \in \mathbb{R}^n | \delta_k(\mathbf{x}) = 1\}$, entonces, $\beta(\theta_0, \delta) \leq \beta(\theta_0, \delta_k)$ y

$$\delta_k(\mathbf{x})(p(\mathbf{x}, \theta_1) - kp(\mathbf{x}, \theta_0)) \geq \delta(\mathbf{x})(p(\mathbf{x}, \theta_1) - kp(\mathbf{x}, \theta_0)). \tag{7.8}$$

En A, $p(\mathbf{x}, \theta_1) = kp(\mathbf{x}, \theta_0)$. Así, para todo $\mathbf{x}$ en A, ambos lados de (7.8) se reducen a cero. En B, $\delta = \delta_k$ y, en consecuencia, (7.8) se reduce a

una igualdad. Por lo anterior, integrando sobre $A \cup B$, a ambos lados de esta igualdad, se obtiene

$$\beta(\theta_1, \delta_k) - \beta(\theta_1, \delta) = k \int_{A \cup B} (\delta_k(\mathbf{x}) - \delta(\mathbf{x}))p(\mathbf{x}, \theta_0)d\mathbf{x} \geq 0,$$

de donde $\beta(\theta_1, \delta_k) \geq \beta(\theta_1, \delta)$.

Sea δ^* una prueba para $H : \theta = \theta_0$ versus $K : \theta = \theta_1$, $\theta_0 < \theta_1$, tal que $\beta(\theta_0, \delta^*) \leq \beta(\theta_0, \delta) \leq \beta(\theta_0, \delta_k)$. A partir del teorema (7.1), se tiene que

$$\begin{aligned} \beta(\theta_1, \delta_k) - \beta(\theta_1, \delta^*) &\geq k(\beta(\theta_0, \delta_k) - \beta(\theta_0, \delta^*)) \\ &\geq k(\beta(\theta_0, \delta_k) - \beta(\theta_0, \delta)) \\ &= \beta(\theta_1, \delta_k) - \beta(\theta_1, \delta), \end{aligned}$$

donde la primera desigualdad se sigue a partir de (7.5) y la igualdad, a partir de (7.8). Esto significa que $\beta(\theta_1, \delta^*) \leq \beta(\theta_1, \delta)$, es decir, que δ es la prueba uniformemente más potente de un nivel de significación $\alpha = P_{\theta_0}[\delta = 1]$ $\square$.

Ejemplo 7.6. Dada una observación de la variable aleatoria $X \sim U(0, \theta)$, se desea encontrar la prueba uniformemente más potente para probar $H : \theta = \theta_0$ versus $K : \theta = \theta_1$, con $\theta_0 < \theta_1$.

Entonces,

$$L(x, \theta_0, \theta_1) = \begin{cases} \frac{\theta_0}{\theta_1} & \text{si } 0 < x < \theta_0 \\ \infty & \text{si } \theta_0 \leq x < \theta_1. \end{cases}$$

Si $k = \dfrac{\theta_0}{\theta_1}$, la prueba

$$\delta_k(x) = \begin{cases} 1 & \text{si } L(x, \theta_0, \theta_1) \geq k \\ 0 & \text{en caso contrario} \end{cases}$$

tiene como región de rechazo el conjunto $RR = \{x \in \mathbb{R} : 0 < x < \theta_1\}$ y probabilidad de error de tipo I, $\alpha = P_{\theta_0}[x \in RR] = 1$.

Si $k > \dfrac{\theta_0}{\theta_1}$, la prueba δ_k tiene como región de rechazo el conjunto $RR = \{x \in \mathbb{R} : \theta_0 \leq x < \theta_1\}$ y probabilidad de error de tipo I, $\alpha = P_{\theta_0}[x \in RR] = 0$.

La prueba de nivel de significación α, $0 < \alpha < 1$, se define considerando l, $0 < l < \theta_0$, tal que $\alpha = P_{\theta_0}[X \geq l] = \frac{\theta_0 - l}{\theta_0}$. Así, de acuerdo con el teorema

7.2, la prueba uniformemente más potente, de nivel de significación α, tiene como función crítica

$$\delta_l(x) = \begin{cases} 1 & \text{si } x \geq \theta_0(1-\alpha) \\ 0 & \text{en caso contrario.} \end{cases}$$

La función de potencia es $\beta(\theta, \delta) = 1 - (1-\alpha)\frac{\theta_0}{\theta}$.

Ejemplo 7.7. Suponga que los integrantes de una población se encuentran marcados con números consecutivos de 1 hasta θ, siendo θ el tamaño de la población. Si se obtiene una muestra aleatoria con remplazo, de tamaño n, $X_1, \ldots, X_n$, en la que se determina el número de cada miembro de la muestra, entonces, la función de probabilidad conjunta está dada por $P(x_1, \ldots, x_n) = \theta^{-n}$.

Así, para determinar la prueba uniformemente más potente, de nivel de significación α, para probar $H : \theta = \theta_0$ versus $K : \theta = \theta_1$; con $\theta_1 > \theta_0$, se encuentra que

$$L(\mathbf{x}, \theta_0, \theta_1) = \begin{cases} \left(\dfrac{\theta_0}{\theta_1}\right)^n & \text{si } 1 \leq \text{máx}(x_1, \ldots, x_n) \leq \theta_0 \\ \infty & \text{si } \theta_0 < \text{máx}(x_1, \ldots, x_n) \leq \theta_1. \end{cases}$$

Si $L(\theta_0, \theta_1) = \infty$, se rechaza H, ya que $\text{máx}(x_1, \ldots, x_n) > \theta_0$. Para un $j \leq \theta_0$, la prueba que rechaza H, si y solo si, $\text{máx}(x_1, \ldots, x_n) > j$, es la prueba uniformemente más potente de nivel de significación α_j, ya que

$$\begin{aligned} \alpha_j &= P_{\theta_0}(\text{máx}(x_1, \ldots, x_n) \geq j) \\ &= 1 - P_{\theta_0}(X_1 < j, \ldots, X_n < j) \\ &= 1 - \left(\frac{j-1}{\theta_0}\right)^n. \end{aligned}$$

Esto define la familia de pruebas uniformemente más potentes, de nivel de significación α_j, $j = 1, 2, \ldots, \theta_0$, sobre el conjunto $A = \{\mathbf{x} \in \mathbb{Z}^{+n} : L(\mathbf{x}, \theta_0, \theta_1) = (\frac{\theta_0}{\theta_1})^n\}$, donde $\mathbf{x} = (x_1, \ldots, x_n)$.

Ejemplo 7.8. Sea $X_i \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$, una muestra aleatoria de tamaño n, proveniente de una distribución Normal, con media θ desconocida

y varianza σ^2 conocida. Para obtener la prueba uniformemente más potente, que permita probar la hipótesis $H : \theta = \theta_0$ versus $K : \theta \neq \theta_0$, se considera la estadística

$$S(\mathbf{X}) = \frac{\sqrt{n}(\bar{X} - \theta_0)}{\sigma}, \tag{7.9}$$

la cual, bajo la hipótesis H, tiene una distribución Normal estándar, y bajo K, con $\theta = \theta_1$, una distribución Normal con media $\Delta = \frac{\sqrt{n}(\theta_1 - \theta_0)}{\sigma}$ y varianza 1.

A partir de la estadística S, se define la estadística $T(X) = |S(X)|$ la cual, bajo la hipótesis K, tiene como función de distribución:

$$\begin{aligned} F_\Delta(z) &= P_\Delta(T(\mathbf{X}) \leq z) \\ &= P_\Delta(|S(\mathbf{X})| \leq z) \\ &= P_\Delta\Big(- z - \Delta \leq \frac{\sqrt{n}(\bar{X} - \theta_0)}{\sigma} - \Delta \leq z - \Delta \Big) \\ &= \int_{-z-\Delta}^{z-\Delta} f(x)dx, \end{aligned} \tag{7.10}$$

donde f corresponde a la función de densidad de la distribución Normal estándar. Aplicando el teorema fundamental del cálculo en (7.10), se obtiene la función de densidad de T:

$$\begin{aligned} g(z, \Delta) &= F'_\Delta(z) \\ &= f(z - \Delta) + f(-z - \Delta) \\ &= f(z)e^{-\Delta^2/2}(e^{\Delta z} + e^{-\Delta z}). \end{aligned}$$

Dado que el sistema de hipótesis $H : \theta = \theta_0$ versus $K : \theta \neq \theta_0$, puede expresarse en términos de Δ como $H : \Delta = 0$ versus $K : \Delta = \Delta_1$, para $z > 0$, entonces,

$$\begin{aligned} L(z, 0, \Delta) &= \frac{g(z, \Delta_1)}{g(z, 0)} \\ &= e^{-\Delta_1^2/2} \left(\frac{e^{\Delta_1 z} + e^{-\Delta_1 z}}{2} \right) \end{aligned}$$

es una función estrictamente creciente de z. En consecuencia, dado que $\alpha = P_0(T \geq z_{\alpha/2})$, entonces, δ_α es la prueba uniformemente más potente de nivel de significación α para $H : \Delta = 0$ versus $K : \Delta = \Delta_1$.

Así, la función de potencia es

$$\begin{aligned}\beta(\delta,\Delta) &= 1 - P_\Delta(T(X) < z_{\alpha/2})\\ &= 1 - P_\Delta(-z_{\alpha/2} < S(X) < z_{\alpha/2})\\ &= 1 - P_\Delta(-z_{\alpha/2} - \Delta < S(X) - \Delta < z_{\alpha/2} - \Delta)\\ &= 1 - \Phi(z_{\alpha/2} - \Delta) + \Phi(-z_{\alpha/2} - \Delta).\end{aligned}$$

Ejercicios 7.1.

1. Sea $X_i, i = 1, 2, 3, \ldots, n$, una muestra aleatoria, de tamaño n, de una función de distribución exponencial con función de densidad $f(x) = \theta e^{-\theta x} I_{\{x:x>0\}}$.

   a) Encuentre la región de rechazo, de la prueba uniformemente más potente de nivel de significación α, para $H : \theta = \theta_0$ versus $K : \theta = \theta_1$, $\theta_1 > \theta_0$.

   b) Encuentre n tal que la región de rechazo, de la prueba uniformemente más potente de nivel de significación α, para $H : \theta = \theta_0$ versus $K : \theta = \theta_1$; $\theta_1 > \theta_0$, sea $RR = \{\bar{x} \in \mathbb{R} : \bar{x} < 0.2\}$.

2. Resuelva el primer ejercicio si la muestra aleatoria proviene de una distribución exponencial con función de densidad $f(x) = \frac{1}{\theta} e^{-\frac{1}{\theta}x} I_{\{x:x>0\}}$.

3. Sea X_i, $i = 1, 2, \ldots, n$, una muestra aleatoria, de tamaño n, de una función de distribución exponencial con función de densidad $f(x) = \frac{1}{\theta} e^{-\frac{1}{\theta}x} I_{\{x:x>0\}}$. Halle la prueba uniformemente más potente, de nivel de significación α, para $H : \theta = \theta_0$ versus $K : \theta < \theta_0$.

4. Supóngase que se tiene una muestra aleatoria, de tamaño n, proveniente de una distribución $Bin(n, \theta)$. Encuentre la región de rechazo de la prueba uniformemente más potente de nivel de significación α para probar

   a) $H : \theta = \theta_0$ versus $K : \theta = \theta_1$; $\theta_1 < \theta_0$.

   b) $H : \theta = \theta_0$ versus $K : \theta = \theta_1$; $\theta_1 > \theta_0$.

5. Si $X_i, i = 1, 2, \ldots, n$, es una muestra aleatoria de tamaño n, que proviene de una distribución Geométrica, determine la región de rechazo de la prueba uniformemente más potente, de nivel de significancia α, para la prueba de hipótesis de $H : \theta = \theta_0$ versus $K : \theta = \theta_1$, $\theta_1 > \theta_0$, y grafique la función de potencia.

6. Suponga que se tiene una muestra aleatoria, de tamaño uno, proveniente de una distribución Hipergeométrica $H(N\theta, N, n)$. Halle la región de rechazo de la prueba uniformemente más potente, de nivel de significacion α, para probar $H : \theta = \theta_0$ versus $K : \theta = \theta_1$, con $\theta_1 > \theta_0$.

7. Sea $X_i \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$, una muestra aleatoria, de tamaño n, de una distribución Normal, con media y varianza desconocidas. Como en el ejemplo 7.8, obtenga la prueba uniformemente más potente, de nivel de significación α, para probar la hipótesis $H : \theta = \theta_0$ versus $K : \theta \neq \theta_0$.

8. Sea x una observación de una variable aleatoria $X_{(n)}$ con función de densidad $g_{X_{(n)}}(x) = \dfrac{n}{\theta^n} x^{n-1} I_{(0,\theta)}(x)$. Determine la prueba uniformemente más potente, de nivel de significación α, para probar $H : \theta = \theta_0$ versus $K : \theta = \theta_1$, con $\theta_1 > \theta_0$.

7.2 Prueba de razón de verosimilitud

Suponga que $\mathbf{X} = (X_1, \ldots, X_n)$ tiene una función de densidad o función de probabilidad $P(\mathbf{x}, \theta)$, siendo $\mathbf{x}$ una posible realización de $\mathbf{X}$, y que se desea probar la hipótesis $H : \theta \in \Theta_0$ versus $K : \theta \in \Theta_1$, donde $\Theta_0 \subseteq \mathbb{R}^p$ y $\Theta_1 \subseteq \mathbb{R}^p$, $p \geq 1$. Y, como estadístico de prueba, se considera la razón de verosimilitud dada por

$$L(\mathbf{x}, \theta_0, \theta_1) = \frac{\sup\{p(\mathbf{x}, \theta) : \theta \in \Theta_1\}}{\sup\{p(\mathbf{x}, \theta) : \theta \in \Theta_0\}}. \tag{7.11}$$

Esta prueba rechaza H para valores grandes de L. En consecuencia, se rechaza H, si y solo si, $\mathbf{x}$ es más probable para algún θ en Θ_1.

Si $p(\mathbf{x}, \theta)$ es una función continua de θ y Θ_0 es de menor dimensión que $\Theta = \Theta_0 \cup \Theta_1$, la estadística (7.11) es igual a

$$\lambda(\mathbf{x}) = \frac{\sup\{p(\mathbf{x}, \theta) : \theta \in \Theta\}}{\sup\{p(\mathbf{x}, \theta) : \theta \in \Theta_0\}}. \tag{7.12}$$

Esta estadística rechaza H, para valores grandes de $\lambda(\mathbf{x})$.

El siguiente ejemplo ilustra los pasos a seguir en pruebas de hipótesis en las que se aplica (7.12).

Ejemplo 7.9. Sea X_i, $i = 1, 2, \ldots, n$, una muestra aleatoria, de tamaño n, de una distribución exponencial, con función de densidad $f_\theta(x) = \frac{1}{\theta} e^{-\frac{1}{\theta}x} I_{(0,\,\infty)}(x)$. Dado que la función de verosimilitud alcanza su valor máximo en θ igual a $\bar{X}$, si se desea probar $H : \theta = \theta_0$ versus $K : \theta \neq \theta_0$,

$$\lambda(\mathbf{x}) = \frac{\frac{1}{\bar{x}^n} \exp(-\frac{1}{\bar{x}} \sum_{i=1}^n x_i)}{\frac{1}{\theta_0^n} \exp(-\frac{1}{\theta_0} \sum_{i=1}^n x_i)} = \left(\frac{\theta_0}{\bar{x}}\right)^n \exp\left(\frac{n\bar{x}}{\theta_0} - n\right).$$

Entonces, derivando L con respecto a $\bar{x}$ se obtiene:

$$\frac{\partial \lambda}{\partial \bar{x}} = n \left(\frac{\theta_0}{\bar{x}}\right)^n \left(\frac{1}{\theta_0} - \frac{1}{\bar{x}}\right) \exp\left(\frac{n\bar{x}}{\theta_0} - n\right),$$

de donde se concluye que si $\bar{x} > \theta_0$, entonces, L es una función estrictamente creciente de $\bar{x}$ y, además, que si $0 < \bar{x} < \theta_0$, entonces, L es una función estrictamente decreciente de $\bar{x}$. En consecuencia, en una prueba, de nivel de significación α, se rechaza la hipótesis H para valores de $\sum_{i=1}^n x_i \in (0, c_1] \cup [c_2, \infty)$, donde c_1, c_2 son números reales tales que

$$\frac{\alpha}{2} = P_{\theta_0}\left[\sum_{i=1}^n X_i < c_1\right] \quad \text{o} \quad \frac{\alpha}{2} = P_{\theta_0}\left[\sum_{i=1}^n X_i > c_2\right].$$

Finalmente, dado que $\frac{2}{\theta_0} \sum_{i=1}^n X_i \sim \chi^2_{(2n)}$; bajo $H : \theta = \theta_0$, entonces,

$$\begin{aligned} c_1 &= \frac{1}{2}\theta_0 x^2_{2n,1-\alpha/2} \\ c_2 &= \frac{1}{2}\theta_0 x^2_{2n,\alpha/2}. \end{aligned}$$

Ejemplo 7.10. Supóngase que se tiene una muestra aleatoria de una distribución Normal $N(\theta, \sigma^2)$, con media y varianza desconocidas y que se desea probar $H : \theta = \theta_0$ versus $K : \theta \neq \theta_0$. El espacio paramétrico asociado a la hipótesis H es $\Theta_0 = \{(\theta, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+ |\ \theta = \theta_0, \sigma^2 > 0\}$. Haciendo $\Theta = \{(\theta, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+ | \theta \in R, \sigma^2 > 0\}$, se observa que Θ_0 es de menor dimensión que Θ.

En consecuencia, dado que $p(\mathbf{x}, \theta)$ es una función continua de θ, se aplica la estadística (7.12) al desarrollo de la prueba de hipótesis, teniendo en cuenta los siguientes pasos:

1. Determine las estimaciones de máxima verosimilitud $(\hat{\theta}, \hat{\sigma}^2)$ de (θ, σ^2) sobre el espacio Θ. En el capítulo 2 se encontró que estas estimaciones son $\hat{\theta} = \bar{X}$ y $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

2. Calcule las estimaciones de máxima verosimilitud de θ sobre Θ_0. Dado que $\theta = \theta_0$, únicamente se debe calcular $\hat{\sigma}_0^2$. Obsérvese que $\hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \theta_0)^2$.

3. Forme el cociente que define el $\lambda(\mathbf{X})$.

$$\begin{aligned}
\lambda(\mathbf{X}) &= \frac{p(\mathbf{X}|\hat{\theta}, \hat{\sigma}^2)}{p(\mathbf{X}|\theta_0, \hat{\sigma}_0^2)} \\
&= \frac{(2\pi\hat{\sigma}^2)^{-n/2} \exp\left[-\dfrac{1}{2\hat{\sigma}^2}\sum(X_i - \bar{X})^2\right]}{(2\pi\hat{\sigma}_0^2)^{-n/2} \exp\left[-\dfrac{1}{2\hat{\sigma}_0^2}\sum(X_i - \theta_0)^2\right]} \\
&= \left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2}\right)^{n/2}. \qquad (7.13)
\end{aligned}$$

   Dado que $\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} = \frac{\sum(X_i - \bar{X})^2 + n(\bar{X} - \theta_0)^2}{\sum(X_i - \bar{X})^2}$, (7.13) puede escribirse como

$$\lambda(\mathbf{X}) = (1 + \frac{1}{n-1}T^2)^{n/2}. \qquad (7.14)$$

   donde $T = \frac{\sqrt{n}(\bar{X} - \theta_0)}{S}$ y $S^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$.

4. Encuentre una función $h(\lambda(\mathbf{X}))$, estrictamente creciente, definida en el rango de λ, tal que $h(\lambda(\mathbf{X}))$ tenga una distribución conocida.

Dado que λ es una función estrictamente creciente de T^2 y que la región de rechazo de H corresponde a $\lambda(\mathbf{X}) = (1 + \frac{1}{n-1}T^2)^{n/2} > c$, la prueba del cociente de máxima verosimilitud rechaza H si $T^2(\mathbf{X}) > c_1$ o, equivalentemente, si $|T(\mathbf{X})| > c'$, donde c, c_1 y c' son números reales positivos. Así, dado que $T|H \sim t_{n-1}$, el valor c' para la prueba, de nivel de significación α, es $t_{\alpha/2,n-1}$.

Y finalmente, asumiendo un $\theta_1 \neq \theta_0$, se puede determinar la función de potencia de la prueba:

$$\begin{aligned}
\beta(\theta_1, \delta) &= P_{\theta_1}\left(\left|\frac{\sqrt{n}(\bar{X} - \theta_0)}{S}\right| > t_{\frac{\alpha}{2},n-1}\right) \\
&= 1 - P_{\theta_1}\left(\frac{\sqrt{n}\theta_0}{S} - t_{\frac{\alpha}{2},n-1} < \frac{\sqrt{n}\bar{X}}{S} < \frac{\sqrt{n}\theta_0}{S} + t_{\frac{\alpha}{2},n-1}\right) \\
&= 1 - P_{\theta_1}\left(\frac{\sqrt{n}(\theta_0 - \theta_1)}{S} - t_{\frac{\alpha}{2},n-1} < \frac{\sqrt{n}(\bar{X} - \theta_1)}{S} < \frac{\sqrt{n}(\theta_0 - \theta_1)}{S} + t_{\frac{\alpha}{2},n-1}\right).
\end{aligned}$$

Así, $\beta(.)$ es una función de θ_1, es estrictamente decreciente para $\theta_1 < \theta_0$, y estrictamente creciente para $\theta_1 > \theta_0$.

Obsérvese que si $\theta_1 < \theta_0$, entonces, la probabilidad

$$P_{\theta_1}\left(\frac{\sqrt{n}(\theta_0 - \theta_1)}{S} - t_{\frac{\alpha}{2},n-1} < \frac{\sqrt{n}(\bar{X} - \theta_1)}{S} < \frac{\sqrt{n}(\theta_0 - \theta_1)}{S} + t_{\frac{\alpha}{2},n-1}\right)$$

decrece a medida que θ_1 decrece, y si $\theta_1 > \theta_0$, entonces, la probabilidad decrece a medida que θ_1 crece.

Ejemplo 7.11. Supóngase que se tiene dos muestras aleatorias independientes $\mathbf{X}_1 = (X_{11}, \ldots, X_{1n_1})$ y $\mathbf{X}_2 = (X_{21}, \ldots, X_{2n_2})$, de tamaño n_1 y n_2, respectivamente, provenientes de distribuciones $N(\theta_1, \sigma^2)$ y $N(\theta_2, \sigma^2)$, respectivamente. Asúmase que se desea probar $H : \theta_1 = \theta_2$ versus $K : \theta_1 \neq \theta_2$. Entonces, la hipótesis de interés H está asociada al espacio de parámetros:

$$\Theta_0 = \{(\theta_1, \theta_2, \sigma^2) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^+ : \theta_1 = \theta_2 = \theta, \theta \in \mathbb{R}, \sigma^2 > 0\},$$

que es subespacio del espacio paramétrico $\Theta = \{(\theta_1, \theta_2, \sigma^2) : \theta_1 \in \mathbb{R}, \theta_2 \in \mathbb{R}, \sigma^2 > 0\}$. En consecuencia, si $\hat{\theta}$ y $\hat{\sigma}_0^2$ son las estimaciones de máxima verosimilitud de θ y σ^2, bajo la hipótesis H, y $\bar{x}_1$, $\bar{x}_2$ y $\hat{\sigma}^2$ las estimaciones de máxima verosimilitud de θ_1, θ_2 y σ^2, respectivamente, bajo la hipótesis K, y

siguiendo el procedimiento indicado en el ejemplo anterior, en el paso 3, se encuentra que

$$\lambda(\mathbf{X}_1, \mathbf{X}_2) = (\hat{\sigma}_0^2/\hat{\sigma}^2)^{(n_1+n_2)/2}, \tag{7.15}$$

donde

$$\begin{aligned} n\hat{\sigma}_0^2 &= \sum_{i=1}^{n_1}(x_{1i} - \hat{\theta})^2 + \sum_{i=1}^{n_2}(x_{2i} - \hat{\theta})^2 \\ n\hat{\sigma}^2 &= \sum_{i=1}^{n_1}(x_{1i} - \bar{x}_1)^2 + \sum_{i=1}^{n_2}(x_{2i} - \bar{x}_2)^2 \end{aligned}$$

y $n = n_1 + n_2$. Así, dado que

$$\begin{aligned} \sum_{i=1}^{n_1}(x_{1i} - \hat{\theta})^2 &= \sum_{i=1}^{n_1}(x_{1i} - \bar{x}_1)^2 + n_1(\bar{x_1} - \hat{\theta})^2 \\ &= \sum_{i=1}^{n_1}(x_{1i} - \bar{x}_1)^2 + n_1\left(\bar{x_1} - \frac{n_1\bar{x_1} + n_2\bar{x_2}}{n_1 + n_2}\right)^2 \\ &= \sum_{i=1}^{n_1}(x_{1i} - \bar{x}_1)^2 + \frac{n_1 n_2^2}{n^2}(\bar{x_1} - \bar{x_2})^2 \end{aligned}$$

y

$$\sum_{i=1}^{n_2}(x_{2i} - \hat{\theta})^2 = \sum_{i=1}^{n_2}(x_{2i} - \bar{x}_2)^2 + \frac{n_1^2 n_2}{n^2}(\bar{x_1} - \bar{x_2})^2,$$

reemplazando en el cociente de máxima verosimilitud se obtiene:

$$\lambda(\mathbf{X}_1, \mathbf{X}_2) = (\hat{\sigma}_0^2/\hat{\sigma}^2)^{(n_1+n_2)/2} = \left(1 + \frac{T^2}{n_1 + n_2 - 2}\right)^{(n_1+n_2)/2},$$

donde $T = \frac{\bar{X}_1 - \bar{X}_2}{S\sqrt{n_1^{-1} + n_2^{-1}}}$ con $s^2 = \frac{n_1+n_2}{n_1+n_2-2}\hat{\sigma}^2$.

Dado que la región de rechazo corresponde a

$$\lambda(\mathbf{X}) = \left(1 + \frac{T^2}{n_1 + n_2 - 2}\right)^{(n_1+n_2)/2} > c,$$

la prueba del cociente de máxima verosimilitud, rechaza H_0 si $T^2(\mathbf{X}) > c_1$, es decir, si $|T(\mathbf{X})| > c$, donde $c = \sqrt{c_1}$. Dado que $T(\mathbf{X})|H \sim t_{(n_1+n_2-2)}$, el valor c para la prueba, de nivel de significación α, es $t_{\alpha/2,n_1+n_2-2}$.

La función de potencia puede determinarse como en el ejercicio anterior.

Ejercicios 7.2.

1. Supóngase que se tienen observaciones provenientes de una distribución $N(\theta, \sigma^2)$, con θ y σ^2 desconocidas, y que se desea probar $H : \theta = \theta_0$ versus $K : \theta > \theta_0$. Use la prueba de razón de verosimilitud para la prueba de hipótesis. Determine la región de rechazo y la función de potencia.

2. Sea X_i, $i = 1, 2, \ldots, n$, una muestra aleatoria, de tamaño n, proveniente de una distribución uniforme en el intervalo $(0, \theta)$. Encuentre la región de rechazo de la prueba uniformemente más potente, de nivel de significación α, para la prueba de hipótesis de $H : \theta = \theta_0$ versus $K : \theta \neq \theta_0$.

3. Sea $X_i \sim N(\mu, \gamma\mu)$, $i = 1, 2, \ldots, n$; una muestra aleatoria, de tamaño n, proveniente de una distribución Normal. Si $\mu > 0$ y $\gamma > 0$, encuentre el cociente de razón de verosimilitud para probar $H_0 : \gamma = \gamma_0$ versus $K : \gamma \neq \gamma_0$.

4. Supóngase que se tiene una observación de una distribución Exponencial con media θ. Si se desea probar $H : \theta = \theta_0$ versus $K : \theta = \theta_1$, $\theta_0 < \theta_1$, encuentre el número real c_α, que determina la región de rechazo de nivel de significación α. Determine la función de potencia y si $x = 3$ y $\theta_0 = 1$, encuentre el p-valor de la prueba.

5. Sean X_i, $i = 1, 2, \ldots, n$; n variables aleatorias independientes con función de densidad:

$$f_{\alpha,\theta}(x) = \frac{\alpha^\alpha}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-\frac{\alpha}{\theta}x} I_{(0,\infty)}(x), \tag{7.16}$$

   $\theta > 0$, $\alpha > 0$.

   Demuestre que si el nivel de significación α es conocido, entonces, $T(X) = \bar{X}$ es completo y suficiente para θ.

Referencias bibliográficas

Agresti, A. & Caffo, B. (2000), 'Simple and effective confidence intervals for proportion and differences of proportions result from adding two successes and two failures', *The American Statistician* **54**(4), 280–288.

Agresti, A. & Coull, B. A. (1998), 'Approximate is better than exact for interval estimation of binomial proportions', *The American Statistician* **52**(2), 119–126.

Apostol, T. (1967), *Calculus*, Jon Wiley & Sons, New York.

Atkinson, A. (1973), 'Testing transformations to normality', *Journal of the Royal Statistical Society, Series B: Methodological* **35**, 473–479.

Bickel, P. & Doksum, K. (2001), *Mathematical Statistics*, Prentice Hall, New Jersey.

Brown, L., Cai, D. & DasGupta, A. (2002), 'Confidence intervals for a binomial proportion and asymptotic expansions', *Ann. Statist.* **30**, 160–201.

Cepeda-Cuervo, E. (2001), *Modelagem da variabilidade em modelos lineares generalizados*, Unpublished Ph.D. Thesis. Mathematics Institute, Universidade Federal do Rio de Janeiro.

Cepeda-Cuervo, E., Aguilar, W., Cervantes, V., Corrales, M., Díaz, I. & Rodríguez, D. (2008), 'Intervalos de confianza e intervalos de credibilidad para una proporción', *Revista Colombiana de Estadística* **31**, 211–228.

Dudewicz, E. & Mishra, S. (1988), *Modern mathematical statistics*, John Wiley and Sons, New York.

Ferrari, S. & Cribari-Neto, F. (2004), 'Beta Regression for Modelling Rates and Proportions', *Journal of Applied Statistics* **31**(7), 799–815.

Freund, J., Miller, I. & Miller, M. (2000), *Estadística matemática con aplicaciones*, Pearson Educación, México.

Jørgensen, B. (1997), 'Proper dispersion models (with discussion)', *Brazilian Journal of Probability and Statistics* **11**, 89–140.

Mendenhall, W., Scheaffer, R. & Wackerly, D. (1994), *Estadística matemática con aplicaciones*, Grupo Editorial Iberoamérica, México, D. F.

Mood, A. M. & Graybill, F. A. (1969), *Introducción a la teoría de la estadística*, Aguilar, Madrid.

Newcombe, R. & Merino, C. (2006), 'Intervalos de confianza para las estimaciones de proporciones y las diferencias entre ellas', *Interdisciplinaria* **23**, 141–154.

Santamaría, M. S. & Malla, M. S. (2010), 'Algoritmos en R para la enseñanza de la aproximación de la distribución binomial a la normal', *Investigación Operacional* **31**(1), 80–92.

Shao, J. (2003), *Mathematical statistics*, Springer, New York.

Stone, C. J. (1995), *A course in probability and statistics*, Duxbury, Belmont, C. A.

Wei, P. (2002), 'Approximate confidence intervals for one proportion and dierence of two proportions', *Computational Statistics & Data Analysis* **40**, 143–157.

Wilson, E. B. (1927), 'Probable inference, the law of succession, and statistical inference', *Journal of the American Statistical Association* **22**, 209–212.

Apéndice A

Software estadístico en R

Este apéndice incluye programas computacionales, desarrollados en el software libre R, útiles en el estudio de algunos de los temas tratados en este libro.

A.1 Distribuciones de variables aleatorias

A.1.1 Tablas estadísticas

En esta sección se indica la forma de obtener tablas estadísticas por medio del software libre `R`. Incluye funciones computacionales que permiten evaluar: la función de distribución, $P(X \leq x)$; la función de densidad y la función cuantil (el menor número x tal que $P(X \leq x) > q$). Incluye, también, la función computacional que permite obtener muestras pseudoaleatorias de una distribución.

En `R` se usan los prefijos 'd' para la densidad, 'p' para la función de distribución acumulada, 'q' para la función cuantil y 'r' para simulación de números pseudoaleatorios. La estructura de la función computacional es `prefijoNombreDistribucion`. Las cuatro funciones tienen los mismos argumentos; los cuales, a excepción del primero, que depende de lo que se desee hallar, corresponden a los parámetros que definen la distribución. En el caso de las funciones de densidad y de distribución, el primer argumento corresponde a un vector de cuantiles; en el caso de la función cuantil, a un

vector de probabilidades y, para la generación de números pseudoaleatorios, a un número entero positivo que indica cuantos números se desean generar.

Si $X \sim B(\texttt{shape1}, \texttt{shape2})$, las siguientes funciones permiten evaluar: la función de densidad beta en cada una de las componentes de un vector $\mathbf{x} \in \mathbb{R}^n$; evaluar $P(X \leq x)$ o $P(X > x)$ en cada una de las componentes de un vector $\mathbf{x} \in \mathbb{R}^n$, según la especificación que se haga del argumento `lower.tail`; el número real x tal que $P(X \leq x) = p$, para un p conocido y obtener una muestra aleatoria de tamaño n.

```
# Densidad
dbeta(x, shape1, shape2, ncp = 0, log = FALSE)
# Distribución acumulada
pbeta(q, shape1, shape2, ncp = 0, lower.tail = TRUE, log.p = FALSE)
# Función cuantil
qbeta(p, shape1, shape2, ncp = 0, lower.tail = TRUE, log.p = FALSE)
# Generación de números pseudo-aleatorios
rbeta(n, shape1, shape2, ncp = 0)
```

Los siguientes ejemplos ilustran el uso de estas funciones de `R` para obtener tablas estadísticas.

Ejemplo A.1. Si X tiene distribución Binomial con $n = 12$ y $p = 0.4$,

```
> qbinom(c(0.05), size=12, prob=.4, lower.tail=TRUE)
[1] 2
> pbinom(c(4), size=12, prob=.4, lower.tail=TRUE)
[1] 0.4381782
> dbinom(c(4), size=12, prob=.4)
[1] 0.2128409
> rbinom(6, size=12, prob=.4)
[1] 3 3 7 4 3 5
```

Ejemplo A.2. Si $X \sim N(2, 16)$,

```
> qnorm(c(0.05), mean=2, sd=4, lower.tail=TRUE)
[1] -4.579415
> qnorm(c(0.05, 0.4, 0.7, 0.8), mean=2, sd=4, lower.tail=TRUE)
[1] -4.5794145  0.9866116  4.0976021  5.3664849
```

```
> pnorm(c(2.54), mean=0, sd=1, lower.tail=TRUE)
[1] 0.9944574
> pnorm(c(-1, 0,1, 1.5, 2.54), mean=0, sd=1, lower.tail=TRUE)
[1] 0.1586553 0.5000000 0.8413447 0.9331928 0.9944574
> dnorm(c(2.54), mean=0, sd=1)
[1] 0.01584758
> dnorm(c(-2.54,-1, 0,1,2.54), mean=0, sd=1)
[1] 0.01584758 0.24197072 0.39894228 0.24197072 0.01584758
> rnorm(5,2,4)
[1]  0.4960678  7.1252148  5.6871103  0.5773360 -0.8119615
```

A.1.2 Gráficas de las funciones de densidad y distribución

Existen varias alternativas para hacer gráficas de las funciones de densidad y distribución. A continuación se presentan algunas:

1. Función de probabilidad de la distribución Binomial.

   ```
   > x <- 0:12
   > plot(x,dbinom(x,size=12,prob=.4),type="h")
   ```

2. Función de densidad de la distribución Normal.

   ```
   > x <- seq(-11,15.5,0.1)
   # Construcción de la gráfica con la función plot
   > plot(x, dnorm(x,2,4),type="l")
   # Construcción de la gráfica con la función curve
   > curve(dnorm(x,2,4), from=-11, to=15.5)
   ```

   Al hacer la gráfica de una función de densidad, se debe tener en cuenta que la secuencia escogida corresponda al conjunto de valores de la variable donde la función de densidad (probabilidad) no es nula. Para graficar variables continuas el argumento `type` de la función `plot` debe ser `l`, ya que este traza una línea continua. Para graficar funciones de probabilidad el argumento es `h`, correspondiente a líneas verticales (figura A.1).

3. Las funciones de distribución acumulada se obtienen reemplazando el segundo argumento de `plot` por la función de distribución acumulada *pnorm*, en

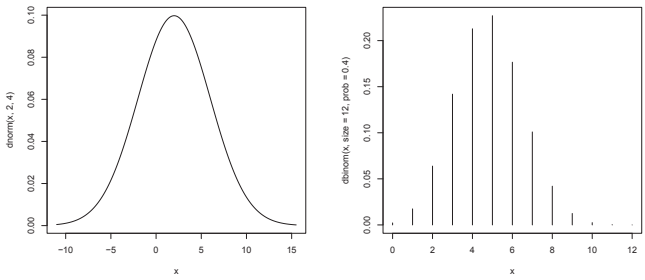

Figura A.1. Gráfica de las funciones de (probabilidad) densidad.

el caso de la distribución Normal, o graficando el resultado de la función `ecdf`, la cual calcula funciones de distribución empíricas.

Para hacer la gráfica de una función de distribución se debe simular una muestra "suficientemente grande". En el siguiente ejemplo se simularon 1000 observaciones de una distribución Binomial con $n = 12$ y $p = 0.4$ y luego, mediante la función `ecdf`, se obtuvo su distribución acumulada empírica, como se muestra a continuación:

```
# Grafica distribución acumulada
> x <- seq(-11,15.5,0.1)
> plot(x, pnorm(x, mean=2, sd=4),type="l")
# Grafica distribución empírica
> x <- rbinom(1000,12,0.4)
> plot(ecdf(x))
```

A.1.3 Aproximación de la distribución Binomial a la Normal

Santamaría & Malla (2010) presentan una propuesta didáctica consistente en dos algoritmos de fácil manejo, desarrollados en el paquete R. Estos algoritmos permiten ilustrar, mediante representaciones gráficas y tablas, las ideas básicas de la aproximación de la distribución Binomial a la Normal.

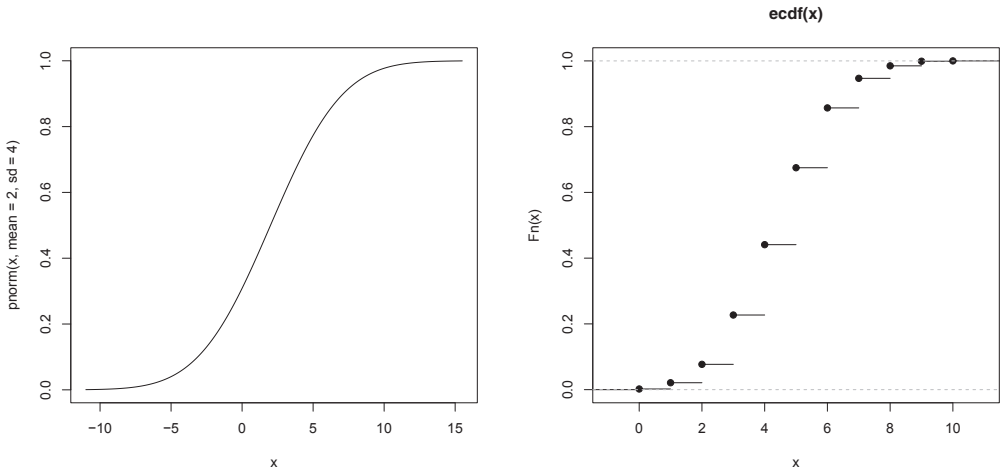

Figura A.2. Gráfica de las funciones de distribución acumulada.

A.1.4 Estudio de la distribución de los datos

A continuación se presentan varias herramientas estadísticas que permiten asumir que un conjunto de datos sigue el comportamiento probabilístico de una variable aleatoria. Una primera idea acerca de la distribución de un conjunto de datos, se puede obtener mediante el uso de medidas de resumen y gráficas descriptivas.

Las siguientes funciones permiten calcular la media, varianza, desviación estándar y cuantiles de un conjunto de datos:

```
A = c(3.03,3.19,3.46,5.88,6.43,5.53,4.26,5.22,6.74,9.97,5.60,4.47,
5.69,6.90,10.39,12.95,5.78,10.19,8.59, 16.81,15.21,6.79,10.71,9.80,
18.39,16.04,9.37,12.58,12.28,20.84, 16.84,12.75,13.41,25.46,21.51)
> hist(A)
>  mean(A)
[1] 10.37314
>  var(A)
[1] 32.7724
>  sd(A)
[1] 5.724718
> summary(A)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  3.030   5.735   9.800  10.370  13.180  25.460
```

A continuación se incluyen funciones computacionales que permiten hacer el gráfico de cajas, el histograma y la densidad empírica a partir de un conjunto de datos (figura A.3):

```
# Histograma
> Hist(A,scale="density")
# Densidad
> plot(density(A), col="red")
```

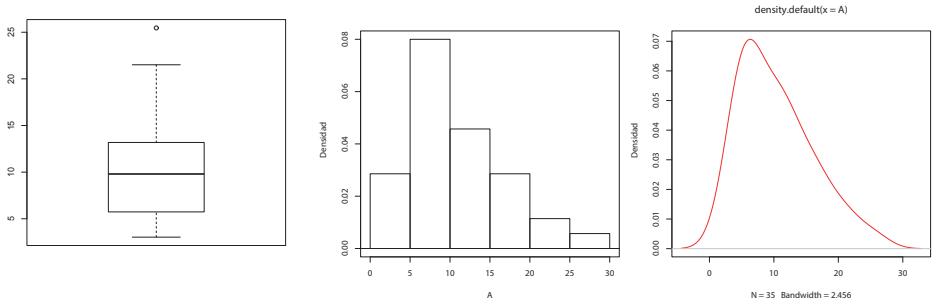

Figura A.3. Diagrama de cajas, histograma y densidad empírica.

A partir de un supuesto inicial, acerca de qué distribución es apropiada para el análisis estadístico de un conjunto de datos, y mediante la prueba de Kolmogorov-Smirnov se determina si tal supuesto es correcto o no. Esto se hace aplicando la función `ks.test`, la cual tiene como argumentos el vector de datos; un objeto tipo `character` correspondiente al nombre dado por `R` para la distribución que se quiere contrastar, precedida del prefijo 'p'; argumentos correspondientes al valor de los parámetros que determinan la distribución y la especificación de la hipótesis alternativa.

```
>  ks.test(A,"pgamma",3.4554458, 0.3331147,alternative="two.sided")

        One-sample Kolmogorov-Smirnov test

data:  A
D = 0.1284, p-value = 0.5671
alternative hypothesis: two-sided
```

Los intervalos de confianza y las pruebas de hipótesis, vistos en los capítulos 5 y 6, asumen normalidad de la variable respuesta. Este supuesto se valida mediante el

test de Shapiro, el cual en R puede desarrollarse mediante la función `shapiro.test`, que tiene como único argumento el vector de observaciones:

```
> shapiro.test(A)

  Shapiro-Wilk normality test

data:  A
W = 0.927, p-value = 0.02282
```

Un gráfico de envelope permite no solo verificar si una muestra aleatoria proviene de una distribución determinada, sino también observar qué tan erróneo es el supuesto inicial respecto a la función de distribución y si existen puntos aberrantes de la muestra que hagan que las conclusiones de las pruebas de Kolmogorov-Smirnov no sean válidas (Atkinson, 1973). El paquete `car` contiene la función `qqPlot`, la cual permite crear envelopes asumiendo las distribuciones listadas en la tabla A.3.3. Los argumentos de esta función son similares a los de la función `ks.test.`, la diferencia está en que la distribución especificada no está precedida por la letra 'p'.

A continuación se usa este comando para hacer envelopes tomando como argumento el conjunto de datos A.

```
> library(car)
> qqPlot(A,"gamma",3.4554458, 0.3331147)
> qqPlot(A,"norm",mean(A),sd(A))
```

A.2 Estimaciones de máxima verosimilitud

En esta sección se describe un procedimiento de maximización que permite obtener estimaciones de máxima verosimilitud de los parámetros de una distribución. Se usa la función `maxLik` contenida en el paquete del mismo nombre. Para lo cual se crea un objeto tipo `est`, correspondiente a la función de log-verosimilitud, cuyo argumento es el vector de parámetros a estimar. En este procedimiento se proporcionan valores iniciales de los parámetros.

A continuación se hallan las estimaciones de máxima verosimilitud de los parámetros de una distribución Gamma, a partir del conjunto de datos introducidos en la sección (A.1.4).

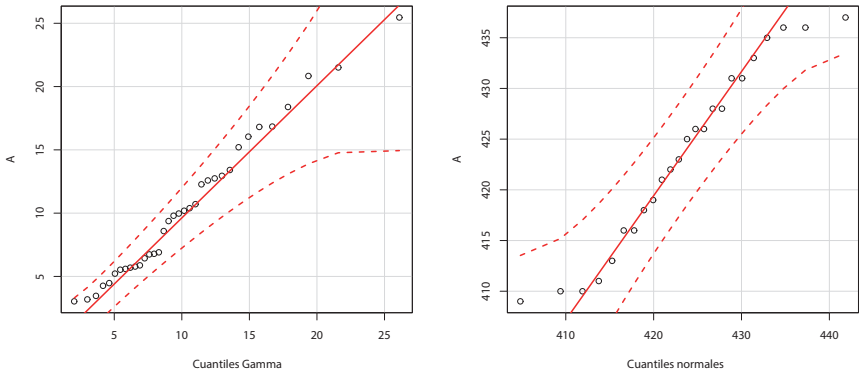

Figura A.4. Envelopes.

```
# Longitud de la muestra
n=length(A)

# Función de verosimilitud de la distribución gamma
log.ver.gamma <- function(theta) {
 k = theta[1]
 lambda = theta[2]
 n*k*log(lambda) - n*log(gamma(k)) - lambda*sum(A)
 + (k-1)*sum(log(A))
}

# Estimación máximo verosímil
est <- maxLik(log.ver.gamma, start=c(1,1), print.level=2)

# Resultados de la estimación de parámetros
> est$estimate
[1] 3.4554458 0.3331147
```

A.3 Intervalos de confianza y pruebas de hipótesis basados en normalidad

En esta sección se consideran funciones computacionales que permiten obtener estimaciones puntuales y por intervalos de la media, varianza y probabilidad de éxito para una y dos poblaciones y de las cantidades necesarias para el juzgamiento de hipótesis relacionadas con estos parámetros.

A.3.1 Inferencia para la media poblacional

R tiene la ventaja de que con una única salida se consiguen las estimaciones puntuales, las estimaciones por intervalos y las cantidades de interés para el juzgamiento de hipótesis. Por otra parte, al variar los argumentos de la función computacional, se obtienen todos los escenarios posibles dependiendo de los supuestos hechos sobre las varianzas poblacionales.

- **Varianza conocida**

  Las siguientes funciones permiten obtener el intervalo de confianza, el estadístico de prueba y el p-valor, para concluir acerca de si la resistencia media puede considerarse igual a 0.3 (véase ejercicio 3, de la sección 5.1).

```
resistencia <- c(5.3,5,5.2,4.9,4.8,5.1,5.3,4.7,5.2,4.7)
# Función que genera el intervalo de confianza
> IC.mu <- function(x,sigma,confidence) {
  li = mean(x) - qnorm((1-confidence)/2,lower.tail=F)
  *sigma/sqrt(length(x))
  ls = mean(x) + qnorm((1-confidence)/2,lower.tail=T)
  *sigma/sqrt(length(x))
  ic = cbind(li,ls)
  list(cat(paste(confidence, "percent confidence interval \n")))
  return(ic)
}
> IC.mu(resistencia,0.16,0.95)
0.95 percent confidence interval
          li       ls
[1,] 4.920833 5.119167

# Función que genera la prueba de hipótesis
```

```
> resistencia <- c(5.3,5,5.2,4.9,4.8,5.1,5.3,4.7,5.2,4.7)
hip.mu <- function(x,mu,sigma,confidence) {
 z = sqrt(length(x)) * (mean(x)-mu)/sigma
 df = length(x)-1
 p.value = 2*pnorm(abs(z),lower.tail=F)
 if(p.value<(1-confidence)){salida="Se rechaza la hipótesis
 nula"}
 else{salida="No se rechaza la hipótesis nula"}
 list(paste("z=",z,", df=",df,", p-value=",p.value),salida)
}

> hip.mu(resistencia,mu=0.3,sigma=0.16,confidence=0.95)
[[1]]
[1] "z= 93.2871909749672 , df= 9 , p-value= 0"

[[2]] [1] "Se rechaza la hipótesis nula"
```

- **Varianza desconocida**

  La función `t.test` permite desarrollar pruebas `t`, para una y dos muestras, bajo el escenario de varianzas desconocidas. Para ilustrar el uso de la función `t.test` considérese el ejercicio 1 de la sección (5.2).

```
diametro <- c(0.51,0.48,0.53,0.49,0.52,0.49,0.50)
t = t.test(x=diametro,y=NULL,mu=0.53,
alternative="two.sided",conf.level=0.98)
```

  El primer argumento de esta función corresponde al vector de datos. Teniendo en cuenta que se trata de hacer inferencias para una población, el segundo argumento, o bien, no se especifica o se le asigna `NULL`; el argumento `mu` corresponde al valor de la media asumido bajo la hipótesis nula; en el argumento `alternative` se especifica la hipótesis alternativa, siendo posibles las opciones `two.sided` `less` o `greater` y, por último, se especifica el nivel de confianza en el argumento `conf.level`, expresado como una probabilidad.

  La salida de esta función, para la prueba de hipótesis con alternativa de dos colas, es:

```
> t
```

```
        One Sample t-test

data:  diametro
t = -3.9908, df = 6, p-value = 0.007194
alternative hypothesis: true mean is not equal to 0.53
95 percent confidence interval:
 0.4862148 0.5194995
sample estimates:
mean of x
0.5028571
```

En la primera línea de la salida, `t` corresponde al valor del estadístico de prueba, `df` a los grados de libertad de su distribución y `p-value` al p-valor. En la segunda línea se verifica la hipótesis alternativa, planteada de acuerdo con la especificación hecha en el argumento `alternative`. La quinta línea muestra el intervalo de confianza para la media poblacional, y en la última, la estimación puntual de la media.

Cuando se trabaja con un número grande de hipótesis, puede necesitarse solo una parte de la salida, por ejemplo, una lista con los p-valores. Ahora bien, para la extracción del p-valor se usa la siguiente instrucción:

```
> t$p.value
[1] 0.007194035
```

Se puede extraer cualquier valor retornado por la función `t.test` usando la instrucción `names`.

```
> names(t)
[1] "statistic"  "parameter"  "p.value" "conf.int" "estimate"
[6] "null.value" "alternative" "method" "data.name"
```

La estructura de esta salida se conserva igual para todas las otras inferencias.

- **Diferencia de medias poblacionales, asumiendo varianzas iguales**

  Considérense los datos del ejercicio 8 de la sección (6.7). En este caso, el segundo vector, notado `y`, corresponde a los datos de la segunda muestra y `mu` a la diferencia entre las medias poblacionales, asumida bajo la hipótesis nula. En esta función debe incluirse un nuevo argumento en donde se declare si se asume o no varianzas poblacionales iguales, este argumento es `var.equal`. En el siguiente ejemplo se asume igualdad entre las varianzas:

```
>  rioA <- c(8.33,8.19,7.72,7.82,7.83,8.26,7.62,8.07,7.55,
7.81,7.51)
>  rioB <- c(7.62,7.64,7.81,7.25,7.90,7.76,7.60,7.53,7.69)
>  t.test(x=rioA, y = rioB, alternative = c("two.sided",
 "less",
+ "greater"), mu = 0, var.equal = TRUE,conf.level = 0.95)

        Two Sample t-test

data:  rioA and rioB
t = 2.1364, df = 18, p-value = 0.04664
alternative hypothesis: true difference in means is not equal
to 0 95 percent confidence interval:
 0.00395317 0.47261249
sample estimates:
mean of x mean of y
 7.882727  7.644444
```

A.3.2 Inferencia para la varianza poblacional

Para hacer inferencias sobre la varianza poblacional se utiliza la función `var.test`, la cual tiene una estructura semejante a la función `t.test`. Todos sus argumentos son iguales con excepción del parámetro `mu`, el cual se reemplaza por el argumento `ratio`, que corresponde al valor del cociente de varianzas poblacionales, asumido bajo la hipótesis nula.

Nuevamente se recurre al uso de funciones creadas por el usuario para hacer inferencias acerca de la varianza poblacional.

```
# Intervalo de confianza
IC.sigma <- function(x,confidence) {
 n  = length(x)
 ls = (n-1)*var(x)/qchisq((1-confidence)/2,n-1)
 li = (n-1)*var(x)/qchisq((1+confidence)/2,n-1)
 ic = cbind(li,ls)
 list(cat(paste(confidence, "percent confidence interval \n")))
 return(ic)
}

# Prueba de hipótesis H: sigma^2= sigma^2_0 versus k:sigma^2>sigma^2_0.
```

```
hip.sigma <- function(x,sigma,confidence) {
 df = length(x)-1
 x = df * var(x)/sigma
 p.value = pchisq(x,df,lower.tail=F)
 if(p.value<(1-confidence)){salida="Se rechaza la hipótesis nula"}
 else{salida="No se rechaza la hipótesis nula"}
 list(paste("xhi=",x,", df=",df,", p-value=",p.value),salida)
}
```

Para ilustrar el uso de estas funciones, se retoma el conjunto de datos de la sección (A.3.1), en el cual se asumió que el valor de la varianza poblacional es igual a 0.16

```
> hip.sigma(resistencia,0.4,0.95)
[[1]]
[1] "xhi= 1.24 , df= 9 , p-value= 0.9986544478"

[[2]] [1] "No se rechaza la hipótesis nula"

> IC.sigma(resistencia,0.95)
0.95 percent confidence interval
             li        ls
[1,] 0.02607402 0.1836772
```

Ahora, mediante el uso de la función `var.test`, se determina si el supuesto de varianzas poblacionales iguales se acepta en el ejercicio del pH de dos ríos, con un nivel de confianza del 95 %:

```
> var.test(x=rioA,y=rioB,alternative=c("two.sided"),ratio=1,
conf.level=0.95)

        F test to compare two variances

data:  rioA and rioB
F = 2.3735, num df = 10, denom df = 8, p-value = 0.2334
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.5526079 9.1496649
sample estimates:
ratio of variances
          2.373521
```

A.3.3 Inferencia sobre la probabilidad de éxito

En esta sección se presentan los códigos que permiten hacer inferencias respecto a la probabilidad de éxito. La función `prop.test` calcula intervalos de confianza y pruebas de hipótesis aproximados para la probabilidad de éxito de k poblaciones. Los argumentos de la función son un vector `x` que contiene el número de éxitos, un vector `n` que contiene el número de ensayos, y un vector `p` que contiene la probabilidad de éxito asumida bajo la hipótesis nula. El intervalo de confianza, para valores grandes de n, se obtiene especificando `correct=F`. A continuación se ilustra esta función asumiendo 400 ensayos, 80 éxitos y $p = 0.5$. La salida sigue el esquema de las funciones `t.test` y `var.test`.

```
> prop.test(x=80,n=400,p=0.5,alternative="two.sided",conf.level=0.95)

        1-sample proportions test with continuity correction

data:  80 out of 400, null probability 0.5
X-squared = 142.8025, df = 1, p-value < 2.2e-16
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.1625896 0.2432985
sample estimates:
  p
0.2

> prop.test(x=80,n=400,p=0.5,alternative="two.sided",
conf.level=0.95,
correct=F)

        1-sample proportions test without continuity correction

data:  80 out of 400, null probability 0.5
X-squared = 144, df = 1, p-value < 2.2e-16
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.1637371 0.2419703
sample estimates:
  p
0.2
```

```
> prop.test(x=c(12,24),n=c(150,150),alternative="less",
conf.level=0.95,
correct=T)

        2-sample test for equality of proportions with continuity
        correction

data:  c(12, 24) out of c(150, 150)
X-squared = 3.8194, df = 1, p-value = 0.02533
alternative hypothesis: less
95 percent confidence interval:
 -1.00000000 -0.01208232
sample estimates:
prop 1 prop 2
  0.08   0.16
```

Inferencias no aproximadas, para una única muestra, pueden hacerse usando la función `binom.test`, la cual tiene los mismos argumentos que la función `prop.test`:

```
> binom.test(x=80,n=400,p=0.5,alternative="two.sided",
conf.level=0.95)

        Exact binomial test

data:  80 and 400
number of successes = 80, number of trials = 400, p-value < 2.2e-16
alternative hypothesis: true probability of success is not equal
to 0.5 95 percent confidence interval:
 0.1618952 0.2426100
sample estimates:
probability of success
                   0.2
```

Distribución	Nombre en R	Argumentos adicionales	Significado
Beta	`beta`	`shape1`	Parámetro de forma 1
		`shape2`	Parámetro de forma 2
		`ncp`	Parámetro de no centralidad
Binomial	`binom`	`size`	Número de ensayos
		`prob`	Probabilidad de éxito en cada ensayo
Ji-cuadrado	`chisq`	`df`	Grados de libertad
		`ncp`	Parámetro de no centralidad
Exponencial	`exp`	`rate`	Tasa de falla/éxito
F de Snedecor	`f`	`df1`	Grados de libertad en el numerador
		`df2`	Grados de libertad en el denominador
		`ncp`	Parámetro de no centralidad
Gamma	`gamma`	`shape`	Parámetro de forma
		`scale`	Parámetro de escala
Geométrica	`geom`	`prob`	Probabilidad de éxito
Hipergeométrica	`hyper`	`m`	Número de bolas blancas
		`n`	Número de bolas negras
		`k`	Número de extracciones
Log-normal	`lnorm`	`meanlog`	Media de la población en escala logarítmica
		`sdlog`	Desviación estándar de la población en escala logarítmica
Logística	`logis`	`location`	Parámetro de localización
		`scale`	Parámetro de escala
Binomial negativa	`nbinom`	`size`	Número de ensayos positivos
		`prob`	Probabilidad de éxito
Normal	`norm`	`mean`	Media de la población
		`sd`	Desviación estándar de la población
Poisson	`pois`	`lambda`	Media de la población
t de Student	`t`	`df`	Grados de libertad
		`ncp`	Parámetro de no centralidad
Uniforme	`unif`	`min`	Mínimo valor posible
		`max`	Máximo valor posible

Tabla A.1. Tablas estadísticas.

Apéndice B

Tablas estadísticas

Tabla B.1: Distribución Binomial acumulada: $P(Y \leq a)$.

n	x	0,01	0,05	0,10	0,15	0,20	0,25	0,30	0,33	0,35	0,40	0,45	0,50
2	0	0,9801	0,9025	0,8100	0,7225	0,6400	0,5625	0,4900	0,4489	0,4225	0,3600	0,3025	0,2500
	1	0,9999	0,9975	0,9900	0,9775	0,9600	0,9375	0,9100	0,8911	0,8775	0,8400	0,7975	0,7500
	2	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
3	0	0,9703	0,8574	0,7290	0,6141	0,5120	0,4219	0,3430	0,3008	0,2746	0,2160	0,1664	0,1250
	1	0,9997	0,9928	0,9720	0,9392	0,8960	0,8438	0,7840	0,7452	0,7183	0,6480	0,5748	0,5000
	2	1,0000	0,9999	0,9990	0,9966	0,9920	0,9844	0,9730	0,9641	0,9571	0,9360	0,9089	0,8750
	3	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
4	0	0,9606	0,8145	0,6561	0,5220	0,4096	0,3164	0,2401	0,2015	0,1785	0,1296	0,0915	0,0625
	1	0,9994	0,9860	0,9477	0,8905	0,8192	0,7383	0,6517	0,5985	0,5630	0,4752	0,3910	0,3125
	2	1,0000	0,9995	0,9963	0,9880	0,9728	0,9492	0,9163	0,8918	0,8735	0,8208	0,7585	0,6875
	3	1,0000	1,0000	0,9999	0,9995	0,9984	0,9961	0,9919	0,9881	0,9850	0,9744	0,9590	0,9375
	4	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
5	0	0,9510	0,7738	0,5905	0,4437	0,3277	0,2373	0,1681	0,1350	0,1160	0,0778	0,0503	0,0312
	1	0,9990	0,9774	0,9185	0,8352	0,7373	0,6328	0,5282	0,4675	0,4284	0,3370	0,2562	0,1875
	2	1,0000	0,9988	0,9914	0,9734	0,9421	0,8965	0,8369	0,7950	0,7648	0,6826	0,5931	0,5000
	3	1,0000	1,0000	0,9995	0,9978	0,9933	0,9844	0,9692	0,9564	0,9460	0,9130	0,8688	0,8125
	4	1,0000	1,0000	1,0000	0,9999	0,9997	0,9990	0,9976	0,9961	0,9947	0,9898	0,9815	0,9688
	5	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
6	0	0,9415	0,7351	0,5314	0,3771	0,2621	0,1780	0,1176	0,0905	0,0754	0,0467	0,0277	0,0156
	1	0,9985	0,9672	0,8857	0,7765	0,6554	0,5339	0,4202	0,3578	0,3191	0,2333	0,1636	0,1094
	2	1,0000	0,9978	0,9842	0,9527	0,9011	0,8306	0,7443	0,6870	0,6471	0,5443	0,4415	0,3437
	3	1,0000	0,9999	0,9987	0,9941	0,9830	0,9624	0,9295	0,9031	0,8826	0,8208	0,7447	0,6562
	4	1,0000	1,0000	0,9999	0,9996	0,9984	0,9954	0,9891	0,9830	0,9777	0,9590	0,9308	0,8906
	5	1,0000	1,0000	1,0000	1,0000	0,9999	0,9998	0,9993	0,9987	0,9982	0,9959	0,9917	0,9844
	6	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
7	0	0,9321	0,6983	0,4783	0,3206	0,2097	0,1335	0,0824	0,0606	0,0490	0,0280	0,0152	0,0078
	1	0,9980	0,9556	0,8503	0,7166	0,5767	0,4449	0,3294	0,2696	0,2338	0,1586	0,1024	0,0625
	2	1,0000	0,9962	0,9743	0,9262	0,8520	0,7564	0,6471	0,5783	0,5323	0,4199	0,3164	0,2266
	3	1,0000	0,9998	0,9973	0,9879	0,9667	0,9294	0,8740	0,8318	0,8002	0,7102	0,6083	0,5000
	4	1,0000	1,0000	0,9998	0,9988	0,9953	0,9871	0,9712	0,9566	0,9444	0,9037	0,8471	0,7734
	5	1,0000	1,0000	1,0000	0,9999	0,9996	0,9987	0,9962	0,9935	0,9910	0,9812	0,9643	0,9375
	6	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9998	0,9996	0,9994	0,9984	0,9963	0,9922
	7	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
8	0	0,9227	0,6634	0,4305	0,2725	0,1678	0,1001	0,0576	0,0406	0,0319	0,0168	0,0084	0,0039
	1	0,9973	0,9428	0,8131	0,6572	0,5033	0,3671	0,2553	0,2006	0,1691	0,1064	0,0632	0,0352
	2	0,9999	0,9942	0,9619	0,8948	0,7969	0,6785	0,5518	0,4764	0,4278	0,3154	0,2201	0,1445
	3	1,0000	0,9996	0,9950	0,9786	0,9437	0,8862	0,8059	0,7481	0,7064	0,5941	0,4770	0,3633
	4	1,0000	1,0000	0,9996	0,9971	0,9896	0,9727	0,9420	0,9154	0,8939	0,8263	0,7396	0,6367
	5	1,0000	1,0000	1,0000	0,9998	0,9988	0,9958	0,9887	0,9813	0,9747	0,9502	0,9115	0,8555
	6	1,0000	1,0000	1,0000	1,0000	0,9999	0,9996	0,9987	0,9976	0,9964	0,9915	0,9819	0,9648
	7	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999	0,9998	0,9993	0,9983	0,9961

Tabla B.1: continua

n	x	0,01	0,05	0,10	0,15	0,20	0,25	0,30	0,33	0,35	0,40	0,45	0,50
	8	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
9	0	0,9135	0,6302	0,3874	0,2316	0,1342	0,0751	0,0404	0,0272	0,0207	0,0101	0,0046	0,0020
	1	0,9966	0,9288	0,7748	0,5995	0,4362	0,3003	0,1960	0,1478	0,1211	0,0705	0,0385	0,0195
	2	0,9999	0,9916	0,9470	0,8591	0,7382	0,6007	0,4628	0,3854	0,3373	0,2318	0,1495	0,0898
	3	1,0000	0,9994	0,9917	0,9661	0,9144	0,8343	0,7297	0,6585	0,6089	0,4826	0,3614	0,2539
	4	1,0000	1,0000	0,9991	0,9944	0,9804	0,9511	0,9012	0,8602	0,8283	0,7334	0,6214	0,5000
	5	1,0000	1,0000	0,9999	0,9994	0,9969	0,9900	0,9747	0,9596	0,9464	0,9006	0,8342	0,7461
	6	1,0000	1,0000	1,0000	1,0000	0,9997	0,9987	0,9957	0,9922	0,9888	0,9750	0,9502	0,9102
	7	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9996	0,9991	0,9986	0,9962	0,9909	0,9805
	8	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9997	0,9992	0,9980
	9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
10	0	0,9044	0,5987	0,3487	0,1969	0,1074	0,0563	0,0282	0,0182	0,0135	0,0060	0,0025	0,0010
	1	0,9957	0,9139	0,7361	0,5443	0,3758	0,2440	0,1493	0,1080	0,0860	0,0464	0,0233	0,0107
	2	0,9999	0,9885	0,9298	0,8202	0,6778	0,5256	0,3828	0,3070	0,2616	0,1673	0,0996	0,0547
	3	1,0000	0,9990	0,9872	0,9500	0,8791	0,7759	0,6496	0,5684	0,5138	0,3823	0,2660	0,1719
	4	1,0000	0,9999	0,9984	0,9901	0,9672	0,9219	0,8497	0,7936	0,7515	0,6331	0,5044	0,3770
	5	1,0000	1,0000	0,9999	0,9986	0,9936	0,9803	0,9527	0,9268	0,9051	0,8338	0,7384	0,6230
	6	1,0000	1,0000	1,0000	0,9999	0,9991	0,9965	0,9894	0,9815	0,9740	0,9452	0,8980	0,8281
	7	1,0000	1,0000	1,0000	1,0000	0,9999	0,9996	0,9984	0,9968	0,9952	0,9877	0,9726	0,9453
	8	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9997	0,9995	0,9983	0,9955	0,9893
	9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9997	0,9990
	10	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Tabla B.2. Distribución Normal Estándar: $P(0 \leq z \leq a)$, donde a es un número real positivo.

	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

Tabla B.3. Puntos porcentuales de la distribución t: $P(Y \leq a)$.

ν	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	0.975	0.99	0.995
1	0.1584	0.3249	0.5095	0.7265	1.0000	1.3764	1.9626	3.0777	6.3138	12.7062	31.8205	63.6567
2	0.1421	0.2887	0.4447	0.6172	0.8165	1.0607	1.3862	1.8856	2.9200	4.3027	6.9646	9.9248
3	0.1366	0.2767	0.4242	0.5844	0.7649	0.9785	1.2498	1.6377	2.3534	3.1824	4.5407	5.8409
4	0.1338	0.2707	0.4142	0.5686	0.7407	0.9410	1.1896	1.5332	2.1318	2.7764	3.7469	4.6041
5	0.1322	0.2672	0.4082	0.5594	0.7267	0.9195	1.1558	1.4759	2.0150	2.5706	3.3649	4.0321
6	0.1311	0.2648	0.4043	0.5534	0.7176	0.9057	1.1342	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.1303	0.2632	0.4015	0.5491	0.7111	0.8960	1.1192	1.4149	1.8946	2.3646	2.9980	3.4995
8	0.1297	0.2619	0.3995	0.5459	0.7064	0.8889	1.1081	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.1293	0.2610	0.3979	0.5435	0.7027	0.8834	1.0997	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.1289	0.2602	0.3966	0.5415	0.6998	0.8791	1.0931	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.1286	0.2596	0.3956	0.5399	0.6974	0.8755	1.0877	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.1283	0.2590	0.3947	0.5386	0.6955	0.8726	1.0832	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.1281	0.2586	0.3940	0.5375	0.6938	0.8702	1.0795	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.1280	0.2582	0.3933	0.5366	0.6924	0.8681	1.0763	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.1278	0.2579	0.3928	0.5357	0.6912	0.8662	1.0735	1.3406	1.7531	2.1314	2.6025	2.9467
16	0.1277	0.2576	0.3923	0.5350	0.6901	0.8647	1.0711	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.1276	0.2573	0.3919	0.5344	0.6892	0.8633	1.0690	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.1274	0.2571	0.3915	0.5338	0.6884	0.8620	1.0672	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.1274	0.2569	0.3912	0.5333	0.6876	0.8610	1.0655	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.1273	0.2567	0.3909	0.5329	0.6870	0.8600	1.0640	1.3253	1.7247	2.0860	2.5280	2.8453
21	0.1272	0.2566	0.3906	0.5325	0.6864	0.8591	1.0627	1.3232	1.7207	2.0796	2.5176	2.8314
22	0.1271	0.2564	0.3904	0.5321	0.6858	0.8583	1.0614	1.3212	1.7171	2.0739	2.5083	2.8188
23	0.1271	0.2563	0.3902	0.5317	0.6853	0.8575	1.0603	1.3195	1.7139	2.0687	2.4999	2.8073
24	0.1270	0.2562	0.3900	0.5314	0.6848	0.8569	1.0593	1.3178	1.7109	2.0639	2.4922	2.7969
25	0.1269	0.2561	0.3898	0.5312	0.6844	0.8562	1.0584	1.3163	1.7081	2.0595	2.4851	2.7874
26	0.1269	0.2560	0.3896	0.5309	0.6840	0.8557	1.0575	1.3150	1.7056	2.0555	2.4786	2.7787
27	0.1268	0.2559	0.3894	0.5306	0.6837	0.8551	1.0567	1.3137	1.7033	2.0518	2.4727	2.7707
28	0.1268	0.2558	0.3893	0.5304	0.6834	0.8546	1.0560	1.3125	1.7011	2.0484	2.4671	2.7633
29	0.1268	0.2557	0.3892	0.5302	0.6830	0.8542	1.0553	1.3114	1.6991	2.0452	2.4620	2.7564
30	0.1267	0.2556	0.3890	0.5300	0.6828	0.8538	1.0547	1.3104	1.6973	2.0423	2.4573	2.7500
40	0.1265	0.2550	0.3881	0.5286	0.6807	0.8507	1.0500	1.3031	1.6839	2.0211	2.4233	2.7045
50	0.1263	0.2547	0.3875	0.5278	0.6794	0.8489	1.0473	1.2987	1.6759	2.0086	2.4033	2.6778
60	0.1262	0.2545	0.3872	0.5272	0.6786	0.8477	1.0455	1.2958	1.6706	2.0003	2.3901	2.6603
70	0.1261	0.2543	0.3869	0.5268	0.6780	0.8468	1.0442	1.2938	1.6669	1.9944	2.3808	2.6479
80	0.1261	0.2542	0.3867	0.5265	0.6776	0.8461	1.0432	1.2922	1.6641	1.9901	2.3739	2.6387
90	0.1260	0.2541	0.3866	0.5263	0.6772	0.8456	1.0424	1.2910	1.6620	1.9867	2.3685	2.6316
100	0.1260	0.2540	0.3864	0.5261	0.6770	0.8452	1.0418	1.2901	1.6602	1.9840	2.3642	2.6259
110	0.1260	0.2540	0.3863	0.5259	0.6767	0.8449	1.0413	1.2893	1.6588	1.9818	2.3607	2.6213
120	0.1259	0.2539	0.3862	0.5258	0.6765	0.8446	1.0409	1.2886	1.6577	1.9799	2.3578	2.6174
∞	0.1257	0.2534	0.3853	0.5244	0.6745	0.8417	1.0365	1.2816	1.6450	1.9602	2.3267	2.5763

Tabla B.4. Puntos porcentuales de la distribución χ^2.

ν	0.5	0.25	0.1	0.05	0.025	0.01	0.005	0.001
1	0.4549	1.3233	2.7055	3.8415	5.0239	6.6349	7.8794	10.8276
2	1.3863	2.7726	4.6052	5.9915	7.3778	9.2103	10.5966	13.8155
3	2.3660	4.1083	6.2514	7.8147	9.3484	11.3449	12.8382	16.2662
4	3.3567	5.3853	7.7794	9.4877	11.1433	13.2767	14.8603	18.4668
5	4.3515	6.6257	9.2364	11.0705	12.8325	15.0863	16.7496	20.5150
6	5.3481	7.8408	10.6446	12.5916	14.4494	16.8119	18.5476	22.4577
7	6.3458	9.0371	12.0170	14.0671	16.0128	18.4753	20.2777	24.3219
8	7.3441	10.2189	13.3616	15.5073	17.5345	20.0902	21.9550	26.1245
9	8.3428	11.3888	14.6837	16.9190	19.0228	21.6660	23.5894	27.8772
10	9.3418	12.5489	15.9872	18.3070	20.4832	23.2093	25.1882	29.5883
11	10.3410	13.7007	17.2750	19.6751	21.9200	24.7250	26.7568	31.2641
12	11.3403	14.8454	18.5493	21.0261	23.3367	26.2170	28.2995	32.9095
13	12.3398	15.9839	19.8119	22.3620	24.7356	27.6882	29.8195	34.5282
14	13.3393	17.1169	21.0641	23.6848	26.1189	29.1412	31.3193	36.1233
15	14.3389	18.2451	22.3071	24.9958	27.4884	30.5779	32.8013	37.6973
16	15.3385	19.3689	23.5418	26.2962	28.8454	31.9999	34.2672	39.2524
17	16.3382	20.4887	24.7690	27.5871	30.1910	33.4087	35.7185	40.7902
18	17.3379	21.6049	25.9894	28.8693	31.5264	34.8053	37.1565	42.3124
19	18.3377	22.7178	27.2036	30.1435	32.8523	36.1909	38.5823	43.8202
20	19.3374	23.8277	28.4120	31.4104	34.1696	37.5662	39.9968	45.3147
21	20.3372	24.9348	29.6151	32.6706	35.4789	38.9322	41.4011	46.7970
22	21.3370	26.0393	30.8133	33.9244	36.7807	40.2894	42.7957	48.2679
23	22.3369	27.1413	32.0069	35.1725	38.0756	41.6384	44.1813	49.7282
24	23.3367	28.2412	33.1962	36.4150	39.3641	42.9798	45.5585	51.1786
25	24.3366	29.3389	34.3816	37.6525	40.6465	44.3141	46.9279	52.6197
26	25.3365	30.4346	35.5632	38.8851	41.9232	45.6417	48.2899	54.0520
27	26.3363	31.5284	36.7412	40.1133	43.1945	46.9629	49.6449	55.4760
28	27.3362	32.6205	37.9159	41.3371	44.4608	48.2782	50.9934	56.8923
29	28.3361	33.7109	39.0875	42.5570	45.7223	49.5879	52.3356	58.3012
30	29.3360	34.7997	40.2560	43.7730	46.9792	50.8922	53.6720	59.7031
40	39.3353	45.6160	51.8051	55.7585	59.3417	63.6907	66.7660	73.4020
50	49.3349	56.3336	63.1671	67.5048	71.4202	76.1539	79.4900	86.6608
60	59.3347	66.9815	74.3970	79.0819	83.2977	88.3794	91.9517	99.6072
70	69.3345	77.5767	85.5270	90.5312	95.0232	100.4252	104.2149	112.3169
80	79.3343	88.1303	96.5782	101.8795	106.6286	112.3288	116.3211	124.8392
90	89.3342	98.6499	107.5650	113.1453	118.1359	124.1163	128.2989	137.2084
100	99.3341	109.1412	118.4980	124.3421	129.5612	135.8067	140.1695	149.4493

Tabla B.5. Distribución F, para $\alpha = 0,01$ ($P(Y \geq a)$). Grados de libertad del numerador en la horizontal y grados de libertad del denominador en la vertical.

	1	2	3	4	5	6	7	8	9	10
1	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47	6055.85
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63
70	7.01	4.92	4.07	3.60	3.29	3.07	2.91	2.78	2.67	2.59
80	6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64	2.55
90	6.93	4.85	4.01	3.53	3.23	3.01	2.84	2.72	2.61	2.52
100	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50
110	6.87	4.80	3.96	3.49	3.19	2.97	2.81	2.68	2.57	2.49
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47

Tabla B.6. Distribución F, para $\alpha = 0,05$ ($P(Y \geq a)$). Grados de libertad del numerador en la horizontal y grados de libertad del denominador en la vertical.

	1	2	3	4	5	6	7	8	9	10
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02	1.97
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93
110	3.93	3.08	2.69	2.45	2.30	2.18	2.09	2.02	1.97	1.92
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91

Índice alfabético

Completitud
- de un estadístico, 83
- de una familia de funciones, 83

Desigualdad de información, 90
Distribuciones, 1
- continuas, 11
  - t de Student, 25
  - Beta, 23
  - F, 28
  - Gamma, 19
  - Normal, 12
  - otras funciones, 31
  - Weibull, 22
- discretas, 2
  - Bernoulli, 3
  - Binomial, 6
  - Binomial Negativa, 8
  - Geométrica, 4
  - Hipergeométrica, 9
  - Poisson, 5
  - Uniforme, 2

Error
- cuadrático medio, 81
- medio absoluto, 82

Estadísticos
- de mínima varianza, 86
- insesgados, 86
- propiedades, 81

Estimadores, 39
- consistencia, 58
- eficiencia, 55
- error de estimación, 53
- insesgado de mínima varianza, 56
- insesgados, 55
- máxima verosimilitud, 46
- método de los momentos, 43

Funciones en software R, 171

Información de Fisher, 92
Intervalos de confianza, 97
- diferencia de medias, 112
- diferencia de proporciones, 114
- media de la distribución normal con varianza desconocida, 103
- media de la distribución Normal con varianza conocida, 98
- para el cociente de varianzas, 118
- para la varianza, 117
- para una proporción, 104
  - intervalo *logit*, 108
- probabilidad de cobertura, 109

Lema
- Neyman-Pearson, 153

Pruebas de hipótesis
- p-valor, 146
- compuestas, 122, 149
- diferencia de medias, 132
- diferencia de proporción, 135
- función de potencia, 142
- intervalos de confianza, 148

para diferencia de proporciones, 136
para la media de una distribución normal, 127
para la varianza, 138
para una proporción, 125
potencia de una prueba, 141
probabilidad de error, 123
simples, 121

Razón de verosimilitud, 153
Razón de verosimilitudes
Prueba de, 162

Teorema
del límite central, 34
desigualdad de información, 94
Lehmann-Scheffé, 87
Rao Blackwell, 86
Tchebysheff, 37

Estadística matemática
se terminó de reimprimir y encuadernar
en Proceditor, en mayo de 2017,
con un tiraje de 300 ejemplares,
sobre papel bond blanco bahía de 75 g.
Bogotá, D. C., Colombia.