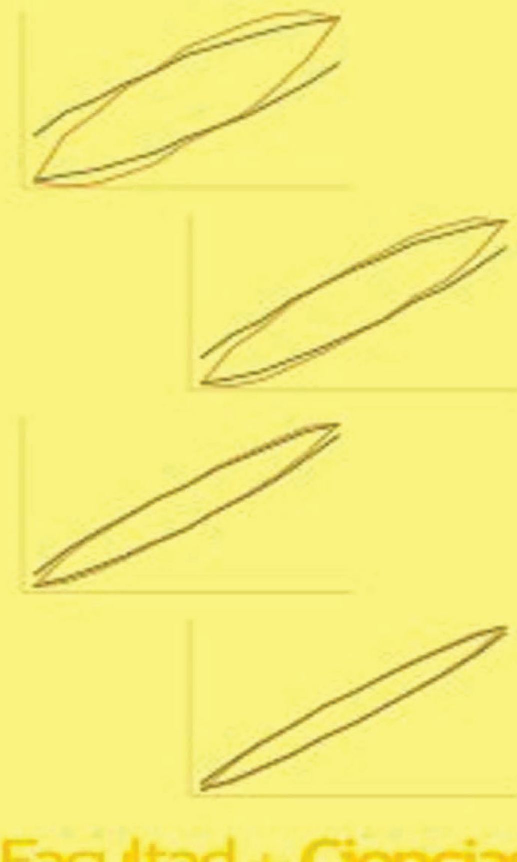
ESTADÍSTICA MATEMÁTICA

notas

Edilberto Cepeda Cuervo



UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE BOGOTÁ
FACULTAD DE CIENCIAS

Facultad de Ciencias
Saber más y formar mejor

Estadística Matemática

Estadística Matemática

Edilberto Cepeda Cuervo

Departamento de Estadística
Facultad de Ciencias
Universidad Nacional de Colombia
Sede Bogotá

Estadística Matemática

Departamento de Estadística
Facultad de Ciencias
Universidad Nacional de Colombia

Edición, 2008
Bogotá, Colombia

ISBN 978-958-719-070-0

Impresión: Proceditor Ltda.
proceditor@etb.net.co
Bogotá, Colombia

Diagramación en LaTeX : Ana María Rojas
Diseño de carátula: Andrea Kratzer

Catalogación en la publicación Universidad Nacional de Colombia

Cepeda Cuervo, Edilberto, 1959 –
Estadística Matemática / Edilberto Cepeda Cuervo. – Bogotá : Universidad Nacional de Colombia. Facultad de Ciencias. Departamento de Estadística, 2008
ix, 107 p.: 13il.

Incluye referencias bibliográficas
ISBN 978-958-719-070-0

1. Estadística matemática 2. Probabilidades 3. Variables aleatorias 4. Teorema del límite central
CDD-21 519 / 2008

Índice general

Introducción

Este libro presenta un conjunto de tópicos fundamentales de la Estadística Matemática, cuyo estudio requiere conocimientos de probabilidad, cálculo diferencial e integral y cálculo vectorial. Está diseñado de tal forma que sirva de base para desarrollar cursos de inferencia estadística de un semestre en programas de matemática, física, ingeniería y estadística.

Su principal objetivo es facilitar la apropiación de conceptos estadísticos y, de promover y facilitar el desarrollo de competencias comunicativas, matemáticas y estadísticas, a través de la lectura y análisis de cada una de sus partes. Por esta razón, se recomienda que el estudiante haga la lectura de cada uno de los temas antes de que sean abordados en clase, desarrollando con claridad cada uno de los procesos necesarios para la comprensión total de los temas considerados. Así, las clases podrán ser utilizadas para la discusión de conceptos y procedimientos, para la exposición y discusión de ejercicios, y para el desarrollo de propuestas, métodos alternativos y teorías no incluidas en este documento.

En el primer capítulo de este libro se estudian funciones de variables aleatorias independientes, con distribución común. En el capítulo 2 se utilizan algunas funciones de variables aleatorias para hacer estimación puntual de parámetros como la media y la varianza de una población, y se incluyen métodos para la obtención y comparación de estimadores. En el capítulo 3, se hace un tratado de estadísticos suficientes, de familia exponencial y de su reparametrizaciones en la forma natural. En el capítulo 4 se introducen procedimientos para encontrar estimadores no sesgados óptimos. El capítulo 5 incluye intervalos de confianza. En el capítulo 6 se hacen pruebas de hipótesis y se establece la relación entre estas e intervalos de confianza. El libro concluye, en el capítulo 7, incluyendo el Lema de Neyman-Pearson, el concepto de pruebas uniformemente más potentes y pruebas de razón de verosimilitudes.

Capítulo 1

Distribuciones de funciones de variables aleatorias

Este capítulo tiene como objetivo el estudio de funciones de distribución. Esta dividido en 3 secciones tituladas: distribuciones discretas, funciones de variables aleatorias, y otras funciones de variables aleatorias.

1.1. Distribuciones discretas

En esta sección se estudian algunas funciones de distribución de variables aleatorias discretas. Se encuentra dividida en 6 sub-secciones tituladas: distribución uniforme discreta, distribución Bernoulli, distribución Geométrica, distribución Poisson, distribución Binomial y Distribución Binomial negativa.

1.1.1. Distribución Uniforme Discreta

Una variable X tiene distribución uniforme discreta en $A = \{1, 2, 3, \ldots, n\}$ si su función de probabilidad está definida por

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

donde I_A es la función indicadora. Esto significa que la variable X puede tomar los valores $1, 2 \ldots, n$ con probabilidad $1/n$, y que, para todo número real x que no pertenece a A, $f(x) = 0$.

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

La función de distribución de una variable X uniformemente distribuida en A es:

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

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 X que tiene función de densidad dada por 1.1, son $E(X) = \frac{n+1}{2}$, $Var(X) = \frac{n^2-1}{2}$ y $M_R(t) = E[\exp(tR)] = \frac{e^{\theta}(1-e^{n\theta})}{n(1-e^{\theta}}$, respectivamente.

1.1.2. Distribución Bernoulli

Una variable X tiene 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 X puede tomar los valores 0 y 1 con probabilidades $P(X = 1) = \theta$ y $P(X = 0) = 1-\theta$. En general, estas variables aparecen vinculadas a experimentos en los cuales los resultados pueden ser asociados a dos categorías, generalmente denominadas *éxito* y *fracaso*. La media, la varianza y la función generadora de momento de una variable que tiene distribución Bernoulli son:

$$\begin{aligned}
&E(X) = \theta, \\
&Var(X) = \theta(1-\theta) \\
&M_R(t) = E[\exp(tX)] = 1 - \theta + \theta e^t
\end{aligned}$$

La función de distribución Bernoulli es:

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

$X \sim Ber(\theta)$ 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

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

donde $A = \{1, 2, 3, \dots\}$. En la ecuación 1.4, θ denota la probabilidad de éxito en un experimento. Así, f(1) la probabilidad de que el primer éxito ocurra en el primer experimento, f(2) la probabilidad de que el primer éxito ocurra en el segundo experimento, y así sucesivamente. La media, la varianza y la función generadora de momento de una variable que tiene esta distribución son: $E(X) = \frac{1-\theta}{\theta}$, $Var(X) = \frac{1}{\theta^2}$ y $M_R(t) = E[\exp(tX)] = \frac{\theta e^t}{1-(1-\theta)e^t}$.

La función de distribución Geométrica está definida por:

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

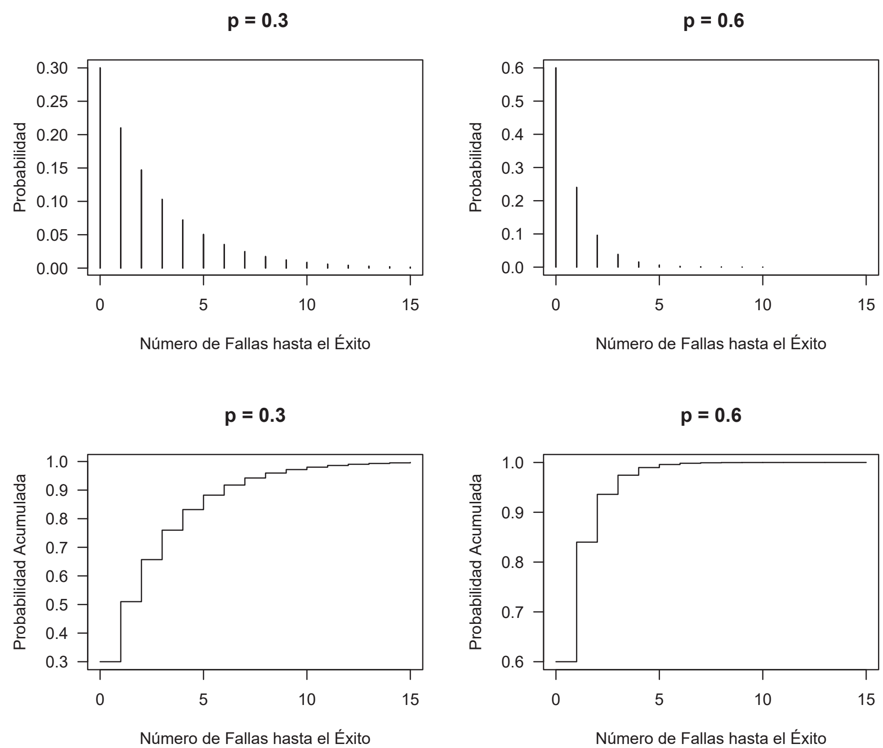
En la figura 1.1 se muestran las gráficas de la distribución Geométrica y de su función de probabilidad para $\theta = 0.3$ y $\theta = 0.6$.

1.1.4. Distribución Poisson

Una variable aleatoria X tiene función de distribución de 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 de Poisson constituye en algunos casos, un buen modelo para explicar el número X 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 diariamente un operador*



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

telefónico, asumiendo que las llamadas son independientes. Esta distribución también puede ser útil en el análisis de variables como *el número de personas atendidas diariamente por un Banco* o *el número de accidentes de tránsito que ocurren anualmente en Bogotá*.

Si X tiene distribución de Poisson con parámetro λ, lo que se denota $X \sim P(\lambda)$, entonces $E(X) = \lambda$, $Var(X) = \lambda$ y $M_R(t) = E[\exp(tR)] = \exp(\lambda(e^t - 1))$.

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

1.1.5. Distribución Binomial

Si $X_i \sim Ber(\theta)$, i=1,2,..., n, son variables aleatorias independientes,

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

tiene distribución Binomial con parámetros n y θ, lo que se denota $X \sim B(n, \theta)$. La función de probabilidad de una variable que tiene 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 $0 \leq \theta \leq 1$,

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

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

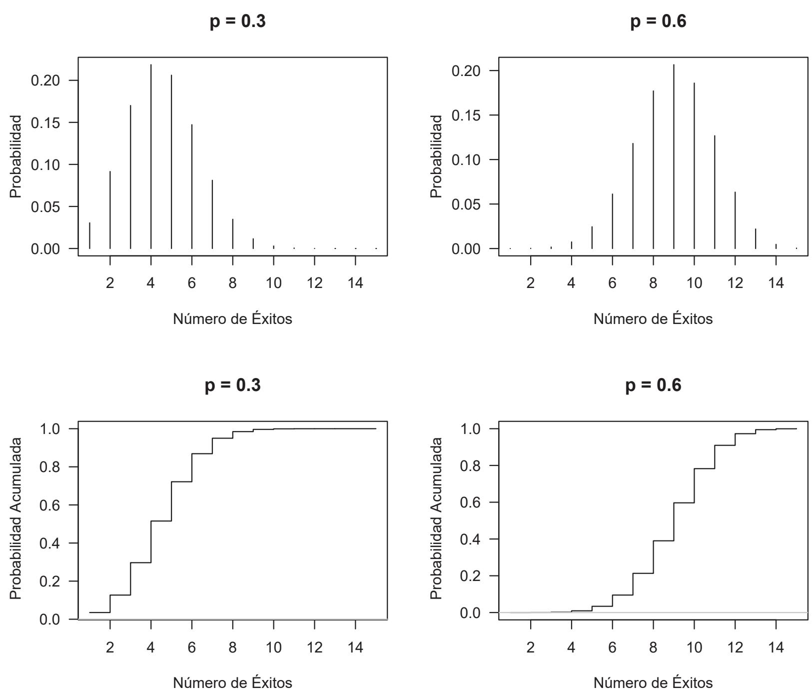
La media, la varianza y la función generadora de momento de una variable que tiene distribución binomial son:

$$\begin{aligned}
&E(X) = n\theta \\
&Var(X) = \sum_{i=1}^{n} V(X_i) = n\theta(1-\theta) \text{ por independencia} \\
&M_R(t) = E[\exp(tX)] = (1 - \theta + \theta e^t)^n
\end{aligned}$$

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

Esta distribución aparece en múltiples aplicaciones, donde ocurre 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 número de caras en 10 lanzamientos tiene distribución binomial con parámetros $n = 10, \theta = 0.5$. 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 $\theta = 0.3$.

Como tercer ejemplo se considera un experimento en el que se seleccionan elementos de un conjunto $\{a_1, a_2, \dots, a_N\}$ 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 *número de elementos con la propiedad P* tiene distribución Binomial. La propiedad P puede ser *ser defectuoso*, *poseer un virus* o *apoyar a un candidato*.



Gráfica 1.2. Distribución binomial: Funciones de probabilidad y distribución.

1.1.6. Distribución Binomial negativa

La distribución Binomial está definida por

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

Esto significa que una variable binomial 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 que tiene distribución binomial negativa son:

$$
\begin{aligned}
E(X) &= \frac{n(1-\theta)}{\theta}, \\
Var(X) &= \frac{n(1-\theta)}{\theta^2} \text{ y} \\
M_R(t) &= E[\exp(tX)] = \Big(\frac{\theta}{1-(1-\theta)e^{\theta}}\Big)^n
\end{aligned}
$$

Una variable binomial negativa con parámetros n y θ hace referencia al número de ensayos en que ocurre el n-ésimo éxito. Por ejemplo, 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, $X = 3$ y los parámetros son n=3 y $\theta = 0.03$.

1.2. Distribuciones continuas

En esta sección se estudian funciones de variables aleatorias independientes, con distribución común. Estas funciones serán utilizadas para estimar, por ejemplo, media y varianza de una población. La sección se encuentra dividida en 6 sub-secciones tituladas: distribución normal, distribución *ji*-cuadrado, distribución gamma, distribución t, distribución F y teorema del límite central.

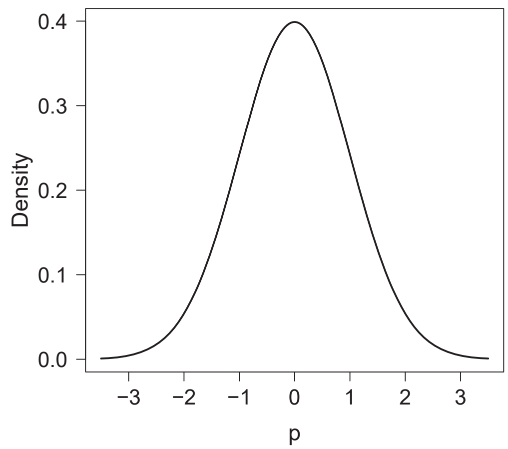
1.2.1. Distribución normal

Una variable X tiene distribución normal si y solo si su función de densidad es

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-\theta)^2\right] \tag{1.10}$$

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

Los parámetros $\theta = E(X)$ y $\sigma^2 = Var(X)$ corresponden a media y varianza de la variable X, respectivamente. Si $\theta = 0$ y $\sigma^2 = 1$, se dice que X tiene distribución normal estándar. La gráfica de la función de densidad de una variable que tiene distribución normal estándar es simétrica con



Gráfica 1.3. Función de densidad Normal Estándar.

respecto a $X = 0$ como se indica en la figura 1.3. Para demostrar que (1.10) 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 fácilmente dado que

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

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

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

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,

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

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

Teorema 1.1. *Sean X_i, $i = 1, 2, \ldots, n$, variables aleatorias independientes normalmente distribuidas con media θ_i y varianza σ_i^2. 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

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

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

Finalmente, dado que existe una relación 1-1 entre funciones de densidad y funciones generadoras de momentos (Shao 2003), R tiene distribución $N(\theta, \sigma^2)$. □

A partir del teorema 1.1 es fácil demostrar que:

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

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

Ejercicios 1.1.

1. Si $X_i,\ i = 1, 2, \ldots, n,$ es una muestra aleatoria 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 %?

2. Asuma que la temperatura de una ciudad tiene distribución normal con media 18°C y desviación estándar de 3ºC. Si durante el último mes se realizaron 125 mediciones que tienen una media de 19.5ºC, ¿qué se puede afirmar de la temperatura del 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 $Y_i \sim N(\theta_2, \sigma_2^2),\ i = 1, 2, \ldots, n_2$.

   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 media
   $$E(R) = \theta_1 - \theta_2 \quad \text{y} \quad V(R) = \frac{1}{n_1}\,\sigma_1^2 \ + \ \frac{1}{n_2}\,\sigma_2^2$$
   ¿Qué hipótesis adicional es necesaria?

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

5. Si $X_i \sim N(\theta, \sigma^2),\ i = 1, 2, \ldots, n$, son variables aleatorias independientes, determine la función de distribución de
   $$Y = \frac{\sqrt{n}(\bar{X} - \theta)}{\sigma}$$
   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 variables aleatorias independientes, determine la función de densidad de $Y = Z_1 - Z_2$ a partir de la función de distribución y 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

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

1.2.2. Distribución χ^2

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

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

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

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

Si una variable tiene esta función de densidad, se dice que tiene distribución Ji-cuadrado con 1 grado de libertad, y se denota $Y \sim \chi^2_{(1)}$.

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

Si una variable Y tiene distribución χ^2 con n grados de libertad, su función de densidad es

$$f(y) = \frac{y^{(n/2)-1}\, e^{-y/2}}{2^{n/2}\,\Gamma(n/2)}\, I_{(0,\infty)}(y) \tag{1.11}$$

$Y \sim \chi^2_{(n)}$ denota 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

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

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

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

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

$$\frac{1}{\sigma^2}\ \Sigma_{i=1}^n\ (Y_i - \bar{Y})^2 = \frac{(n-1)\ S^2}{\sigma^2} \tag{1.12}$$

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

Ejemplo 1.1. Una muestra aleatoria $Y_i \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, 10$, con media y varianza desconocidas, tiene varianza muestral $s^2 = 4.46$. Dado que $9S^2/\sigma^2$ tiene distribución χ^2 con 9 grados de libertad,

$$P\left(\frac{9s^2}{3} < \frac{9S^2}{\sigma^2} < \frac{9s^2}{2}\right) = P\left(13.38 < \frac{9S^2}{\sigma^2} < 20.07\right) = 0.09$$

Esto indica que es muy poco probable tener una varianza muestral de 4.46, si $2 < \sigma^2 < 3$.

Ejercicios 1.2.

1. Una variable 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. Once mediciones aleatorias del diámetro (en centímetros) de las plantas de café, vendidas por un invernadero, tienen una varianza $s^2 = 1.7$. Determine la probabilidad que la muestra provenga de una población con varianza igual a 1, 1.7 y 5.63.

3. Una empresa productora de tornillos garantiza que sus diámetros 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 $P(S^2 > 0.09)$. ¿Qué se puede concluir si $(n-1)s^2/0.0001 = 1$?

4. Una vez determinada la cantidad de zinc que contendrá una droga, 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 ¿solicitaría usted detener y revisar el proceso de producción? ¿Qué se debe hacer si se desea garantizar una desviación estándar menor de 0.01, con una probabilidad del 95 %?

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

6. Sea $X_i \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$, una muestra aleatoria con media θ y varianza σ^2 conocida. Si θ_0 es conocido, determine la función de distribución de $2 \log \lambda(X)$ donde $X = (X_1, X_2, \ldots, X_n)$, $\lambda(X) = p(X, \bar{X})/p(X, \theta_0)$ y

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

7. Suponga 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 $N(\theta_2, \sigma^2)$, 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

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

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

1.2.3. Distribución gamma

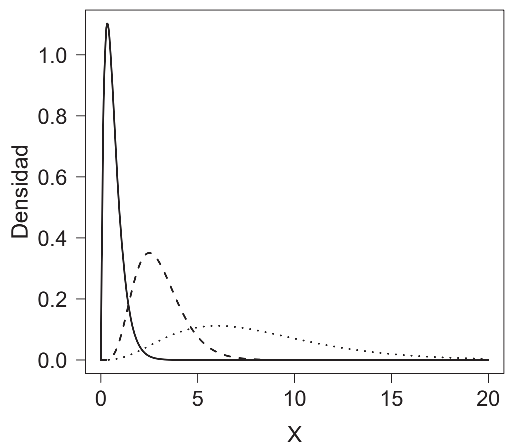
Una variable aleatoria X tiene distribución gamma, con parámetros $p > 0$ y $\lambda > 0$, lo que se denota $X \sim \Gamma(p, \lambda)$, si su función de densidad es

$$f_{\lambda,p}\,(x) = \frac{\lambda^p\; x^{p-1}\; e^{-\lambda x}}{\Gamma(p)}\; I_{(0,\infty)}\,(x) \tag{1.13}$$

donde

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

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. Si $p = 1$, la expresión (1.13) toma la



Gráfica 1.4. Función de densidad Gamma.

forma

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

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 de espera. Si $X \sim \exp(\lambda), Y = 2\lambda X$ 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

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

Observe que si $X_i \sim \exp(\lambda)$, $i = 1, 2, \ldots, n$, a partir de la definición de una distribución χ^2, se concluye que $2\lambda\,\Sigma_{i=1}^n X_i \sim \chi^2_{(2n)}$.

Expresando (1.11) en la forma de la ecuación (1.13), se concluye que la distribución χ^2_n es igual a una distribución $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$. Observe que $\chi^2_{(1)} = \Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$.

Proposición 1.1. *Sean $X_i, i = 1, 2, \ldots, n$, 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 $X_i \sim \Gamma(p_i, \lambda)$, $\quad m_{X_i}(t) = \left(1 - \frac{t}{\lambda}\right)^{-p_i}$, 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 $\Gamma(\Sigma_{i=1}^n p_i, \lambda)$.

□

Ejercicios 1.3.

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

2. Sea X una variable aleatoria con función de distribución Weibull, con densidad
$$f(x, \lambda) = \lambda\alpha\, x^{\alpha-1}\, e^{-\lambda x^\alpha}\, I_{(0,\infty)}(x), \quad \alpha > 0$$
Muestre que $y = x_i^c$ tiene distribución $\exp(\lambda)$.

3. Sea $X_i \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$, una muestra aleatoria de una población normal con media conocida y varianza desconocida. Demuestre que
$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_n, \quad \text{donde} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_1^n (X_i - \theta)^2$$

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

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

6. Si $X_i, i = 1, 2, \ldots, n$, es una muestra aleatoria de una distribución exponencial con parámetro λ, muestre que $\bar{X} \sim G(n, n\lambda)$. Halle la función generadora de momentos de $\bar{X}$.

1.2.4. Distribución t

Definición 1.2. Sean Z una variable aleatoria normal estándar y X una variable aleatoria con distribución Ji-cuadrado con n grados de libertad, independientes. La variable aleatoria

$$T = \frac{Z}{\sqrt{X/n}} \tag{1.14}$$

tiene distribución t con n grados de libertad, lo que se denota $T \sim t_{(n)}$.

La función de densidad T está dada por

$$p_n(t) = \frac{\Gamma[(n+1)/2]}{\Gamma(n/2)\sqrt{\pi n}} \left[1 + \frac{t^2}{n}\right]^{-(n+1)/2}, \quad t \in \mathbb{R} \tag{1.15}$$

Observe que la función de densidad t es simétrica con respecto a $t = 0$. La figura 1.5 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.

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

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

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 fácilmente aplicando la definición (1.14) haciendo $Z = \frac{\sqrt{n}\,(\bar{X}-\theta)}{\sigma} \sim N(0,1)$ y $X = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$.

Ejemplo 1.3. Suponga 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 $N(\theta_2, \sigma^2)$, respectivamente, con $\theta_1 \in \mathbb{R}$, $\theta_2 \in \mathbb{R}$ y $\sigma^2 > 0$, desconocidos. Dado que

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

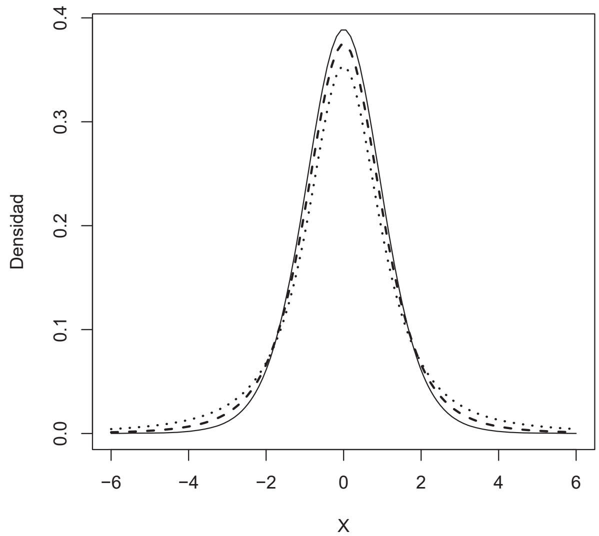
y

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

son independientes, de (1.14) se concluye que

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\theta_1 - \theta_2)}{S\sqrt{n_1^{-1} + n_2^{-1}}} \sim t_{n_1+n_2-2} \tag{1.16}$$

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



Gráfica 1.5. Distribución t con 10, 4 y 2 grados de libertad.

Ejemplo 1.4. Sea $Y_i, i = 1, 2, \ldots, n$, una muestra aleatoria 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 $\bar{Y}$ 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 $P(|\bar{Y} - \theta| < 2) > 0.95$. Esto es equivalente 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_6, 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 12 mm. Si en una muestra aleatoria de sus diámetros, 9 tornillos tienen una media de $\bar{x} = 9.84$ y una desviación estándar de $s = 0.4$, 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 1.4.

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. Una droga debe contener 0.011 gramos de zinc. Si en una muestra aleatoria se obtienen los datos del ejercicio 4 de la sección (1.2.2) ¿qué puede afirmar sobre la cantidad de zinc que contiene la droga?

4. Se afirma que el consumo medio de agua por familia en cierta ciudad es 5 m^3. Si se toma una muestra aleatoria de 10 observaciones y se obtiene $t = 2.8$ ¿qué puede afirmar sobre el consumo de agua? ¿Se está ahorrando agua? ¿Qué podría afirmar si $t = -2.8$? ¿Si $t = 0$?

5. Demuestre que la distribución de la variable aleatoria $T = \frac{Z}{\sqrt{X/v}}$ está dada por la ecuación (1.15).

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

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

   donde

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

1.2.5. Distribución *F*

Definición 1.3. Sean X_1 y X_2 variables aleatorias Ji-cuadrado, con n_1 y n_2 grados de libertad, respectivamente. Si X_1 y X_2 son independientes, la variable

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

tiene distribución F con n_1 grados de libertad en el numerador y n_2 grados de libertad en el denominador.

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 $P\left(F_7^{10} > 3.64\right) = 0.05$.

De la definición de una variable con distribución F, con n_1 grados de libertad en el numerador y n_2 grados de libertad en el denominador, se concluye que

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

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, F_0 tal que $P\left(F_7^{10} < F_0\right) = 0.05$, ya que

$$P\left(F_7^{10} < F_0\right) = P\left(\frac{1}{F_7^{10}} > \frac{1}{F_0}\right) = P\left(F_{10}^{7} > \frac{1}{F_0}\right) = 0.05$$

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 $\frac{1}{F_0} = 3.14$ siendo $F_0 = 0.318$.

A partir de muestras aleatorias de dos poblaciones normales de tamaños n_1 y n_2, se pueden comparar sus varianzas poblacionales σ_1^2 y σ_2^2 mediante la variable aleatoria

$$F = \frac{\sigma_2^2\ S_1^2}{\sigma_1^2\ S_2^2} \tag{1.17}$$

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

Por ejemplo, si se tienen dos muestras aleatorias del pH de un río, una para el verano (7.8, 7.3, 6.8, 7.0, 6.5, 7.0) y otra para el invierno (7.5, 7.0, 6.5, 8.0, 7.9, 7.5, 8.0, 6.5), podemos determinar a y b tales que

$$P\left(\frac{a\ s_2^2}{s_1^2} < \frac{\sigma_1^2\ S_2^2}{\sigma_2^2\ S_1^2} < \frac{b\ s_2^2}{s_1^2}\right) = P\left(1.98a < F_5^7 < 1.98b\right) = 0.9$$

Un par de valores posibles de a y b se obtienen 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 1.5.

1. Demuestre que la variable aleatoria F definida en (1.17) 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 (1.12).

2. Sean $X_i \sim N(\theta, \sigma_1^2),\ i = 1, 2, \ldots, 10$, y $Y_j \sim N(\theta, \sigma_2^2),\ j = 1, 2, \ldots, 15$, variables aleatorias independientes. Si $\sigma_1^2 = 2\sigma_2^2$, halle $P(S_X^2/S_Y^2 > 2.5)$.

3. Sean $X_i \sim N(\theta_1, \sigma_1^2), i = 1, 2, \ldots, 5$, y $Y_j \sim N(\theta_2, \sigma_2^2), j = 1, 2, \ldots, 10$, variables aleatorias independientes. Si $\sigma_1^2 = 2\sigma_2^2$, halle $P(\ 0.5 < S_X^2/S_Y^2 < 2.5\ )$.

4. 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.

1.2.6. 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, lo que se denota $X_i \xrightarrow{d} X$, si y solo si $\{F_n(x)\}$ converge a $F(x)$ en todos los puntos x donde F es continua.

Teorema 1.3. *(Del límite central). Sean $X_1, X_2, \ldots, X_n$ 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, 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) En esta demostración se asume que las variables X_i, $i = 1, 2, \ldots, n, \ldots$ 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, demostrando que la función generadora de momentos de Z_n converge converge a la función generadora de momentos de Z.

Dado que las variables aleatorias X_i, $i = 1, 2, \ldots, n, \ldots$ 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} al rededor de t=0 es

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

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

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

Finalmente, dado que

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

$\lim_{n\to\infty} \Psi_{Z_n} = exp(\frac{1}{2}t)$, que es la función generadora de momentos de una distribución normal estándar. Este resultado sigue dado que si $lim_{n\to\infty}a_n$ existe, $\lim_{n\to\infty}(1 + \frac{a_n}{n})^n = exp\ (lim_{n\to\infty}a_n)$.

Para completar la demostración se debe probar que si $\lim_{n\to\infty} \Psi_{Z_n} = \Psi_Z$, Z_n converge a Z en distribución. Este enunciado, demostrado en textos más avanzados, completa esta demostración.

Ejercicios 1.6.

1. Una fábrica de zapatillas sabe que el 10 % de su producción es defectuosa. Suponga que se desea seleccionar una muestra aleatoria de 20 unidades. Halle la probabilidad de que la muestra contenga al menos 17 no defectuosas, aplicando la distribución:
   a) normal, con y sin aproximación.
   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 variables aleatorias independientes.
   a) Utilizando el teorema del límite central, halle c tal que $P(|\bar{X}-0.5| < c) = 0.5$.
   b) Determine c aplicando el siguiente teorema.

   Teorema 1.4. *(De Tchebysheff). Si X es una variable aleatoria con media finita θ y varianza σ^2,*

   $$P(|X - \theta| < k\sigma) > 1 - \frac{1}{k^2} \tag{1.18}$$

   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 que han sufrido 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.
   c) compare los resultados de a y b.

6. Suponga que $X_i \sim Ber(p_1)$, $i = 1, 2, \ldots, n$, y $Y_j \sim Ber(p_2)$, $j = 1, 2, 3, \ldots, n_2$, independientes. Determine $E(\bar{X} - \bar{Y})$ y $Var(\bar{X} - \bar{Y})$. ¿Cuál es la distribución de $\bar{X} - \bar{Y}$?

1.3. Otras funciones de variables aleatorias

Sean $X_1, X_2, \ldots, X_n$ variables aleatorias continuas independientes con función de distribución F y función de densidad f. Definimos las variables

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

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 fácilmente 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 1.6. Si $X_i \sim U(0, \theta),\ i = 1, 2, 3, \ldots, n$, 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 1.7. Si $X_i \sim U(0, \theta),\ i = 1, 2, 3, \ldots, n$, entonces 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 1.7.

1. Sean $X_i,\ i = 1, 2, \ldots, n,$ variables aleatorias uniformemente distribuidas en el intervalo $(0, \theta)$. Determine la función de distribución de $\dfrac{X_{(1)}}{\theta}$ y de $X_{(n)}/\theta$. ¿Qué concluye?

2. Sean $X_i \sim F_\theta,\ i = 1, 2$, independientes. Si $x_1 < x_2$, calcule

$$P(X_{(1)} \leq x_1,\ X_{(2)} \leq x_2)$$

3. Sean $X_i \sim \exp\ (\theta),\ i = 1, 2, \ldots, n$, independientes. Muestre que $X_{(1)}$ tiene distribución exponencial $\exp\ (n\theta)$. Determine $E(X_{(1)})$.

4. Determine la función de densidad de $X_{(1)}$ y $X_{(n)}$, si $X_i,\ i = 1, 2, \ldots, n$ tiene distribución

   a) gamma.

   a) beta.

5. Sean X_i, $i = 1, 2, \ldots, n$, 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 $i = 1, n$.

6. Sean $X_i \sim \exp(\theta)$, $i = 1, 2, \ldots, n$, independientes. Muestre que

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

   Asuma independencia entre $\sum_{i=1}^{n} X_i$ y $X_{(1)}$.

7. Sea X_i el i-ésimo menor valor de $X_1,\ X_2, \ldots,\ X_n$. Determine la función de densidad de $X_{(i)}$, dado que $X_i \sim F_\theta$, $i = 1, 2, \ldots, n$, son independientes. 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.

Capítulo 2

Estimadores

En este capítulo se utilizan algunas funciones de variables aleatorias para hacer estimaciones puntuales de ciertos parámetros de interés. Por ejemplo, se utiliza el valor de $\bar{X}$ para estimar la media de una población y el valor de S^2 para estimar su varianza. A 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 la sección (2.2) y (2.3) se presenta el método de los momentos y el método de máxima verosimilitud para la obtención de estimadores. Luego, se estudia el error de estimación en la sección (2.4). Finalmente, en la sección (2.5) se comparan las varianzas de los estimadores y, en la sección (2.6), se estudia el concepto de consistencia.

2.1. Conceptos básicos

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

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

es un estimador puntual de θ.

Sea $\hat{\theta}$ un estimador de un parámetro θ. Si $E(\hat{\theta}) = \theta$, se dice que $\hat{\theta}$ es un estimador no sesgado de θ. Si $E(\hat{\theta}) \neq \theta$, que $\hat{\theta}$ es un estimador sesgado de $\hat{\theta}$ y

$B = E(\hat{\theta}) - \theta$ se denomina sesgo de θ. Obsérvese que (2.1) es un estimador insesgado de θ. De igual forma, si $X_1, X_2, \ldots, X_{n_1}$ y $Y_1, Y_2, \ldots, Y_{n_2}$ son muestras aleatorias independientes de distribuciones normales con medias θ_1, θ_2, respectivamente,

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

es un estimador insesgado de $\theta = \theta_1 - \theta_2$.

Ejemplo 2.1. Si Y_i, $i = 1, 2, \ldots, n$, es una muestra aleatoria de una distribución exponencial $f_\theta(y) = \frac{1}{\theta}e^{-y/\theta}I_{\{x:x>0\}}(x), \hat{\theta} = \bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$ es un estimador insesgado de θ. Obsérvese que

$$E(Y_i) = \int_{-\infty}^{\infty} y f(y)dy = \int_0^{\infty} y\frac{1}{\theta}e^{-\frac{y}{\theta}}dy = \theta$$

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 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. Observe que

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

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

Definición 2.1. Un estadístico es una función de variables aleatoria que no contiene parámetros desconocidos.

Ejercicios 2.1.

1. Sea Y_i, $i = 1, 2, \ldots, n$, una muestra aleatoria 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 $Y_{21}, Y_{22}, \ldots, Y_{2m}$ muestras aleatorias 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 $X_1, \ldots, X_n$ una muestra aleatoria de una distribución $U(0, \theta)$. Demuestre que $\frac{n+1}{n}X_{(n)}$ es un estimador insesgado de θ.

4. Suponga que el tiempo antes de falla de un equipo tiene distribución exponencial $E(\lambda)$. 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 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 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. La media del cuadrado del error de un estimador puntual $\hat{\theta}$ se define como el valor esperado de $(\hat{\theta} - \theta)^2$. Suponga que $Y_1, Y_2, \ldots, Y_n$ es una muestra aleatoria 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.

7. Demuestre que $E(\hat{\theta} - \theta)^2 = Var(\hat{\theta}) + B^2$ y calcule la media del cuadrado del error para $\hat{\theta}$ definido como en el ejercicio (5).

8. Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria 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 compare con $Var(\bar{X})$.

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

2.2. Método de los momentos

Suponga que X_i, $i = 1, 2, \ldots, n$, es una muestra aleatoria de una distribución F que depende de θ. Definimos el j-ésimo momento de F por

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

y sus estimadores muestrales por

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

Si $q(\theta) = h(m_1, m_2, \ldots, m_k)$, donde h es una función continua, el método de los momentos indica que un estimador $q(\theta)$ se obtienen 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 $Var(X) = m_2(\theta) - m_1^2(\theta)$, un estimador de la varianza es $\hat{\sigma}^2 = \hat{m}_2(\theta) - \hat{m}_1^2(\theta)$. En particular, dada una muestra aleatoria $X_i \sim Ber(\theta)$, $i = 1, 2, \ldots, n$, 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.

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

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

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

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

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

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

Ejercicios 2.2.

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

2. La distribución geométrica tiene función de densidad

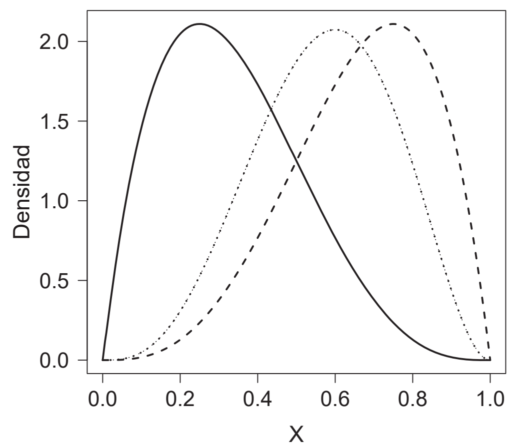
$$f(x) = (1-\theta)^{x-1}\theta, \quad x = 1, 2, \ldots$$

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

3. La distribución beta tiene función de densidad

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

   donde $\alpha > 0$ y $\beta > 0$. Si $\alpha + \beta = 5$, halle estimadores de α y β por el método de los momentos. La figura 2.1 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.



Gráfica 2.1. Función de densidad Beta.

4. 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.

5. 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.

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

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

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

2.3. Estimaciones de máxima verosimilitud

Sean Y_i, $i = 1, 2, \ldots, n$, variables aleatorias Bernoulli, con $\theta = P(Y_i = 1)$ y $P(Y_i = 0) = 1 - \theta$ desconocidos. Si y_i, $i = 1, 2, \ldots, n$, es un conjunto de valores observables de Y_i, $i = 1, 2, \ldots, n$, una pregunta de interés en estadística es cuál es la función de probabilidad P_θ de la cual provienen los datos. En este caso, equivale a preguntar cuál es el valor de θ que hace más probable la observación de los datos. Si el espacio paramétrico Θ es finito, bastaría calcular y comparar

$$L(\theta_i) = P_{\theta_i}(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n) = \prod_{i=1}^{n} P_{\theta_i}(Y_i = y_i)$$

para todo $\theta_i \in \Theta$.

Si $L(\theta)$ es diferenciable en Θ°, el interior de Θ, entonces el número real $\hat{\theta}$ que hace mayor la probabilidad de obtener una muestra determinada de valores observables de Y_i, $i = 1, 2, \ldots, n$, se encuentra maximizando la función de verosimilitud, $L(\theta)$, sobre el espacio paramétrico Θ.

En el caso de la distribución de Bernoulli, dado que

$$P(Y_i = y_i) = \theta^{y_i} (1-\theta)^{1-y_i}, \quad i = 1, 2, \ldots, n$$

la función de verosimilitud es $L = \prod_{i=1}^{n} \theta^{y_i}(1-\theta)^{1-y_i}$, hallar el valor de θ que maximiza L es equivalente a encontrar el valor de θ que maximiza

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

Derivando con respecto a θ se obtiene $\hat{\theta} = \bar{Y}$. A partir de 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 2.5. Sea Y_i, $i = 1, 2, \ldots, n$, una muestra aleatoria de una de distribución uniforme en el intervalo $(0, 2\theta)$. La función de máxima verosimilitud es

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

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

Ejemplo 2.6. Si x_i, $i = 1, 2, \ldots, n$, son valores observados de $X \sim N(\mu, \sigma^2)$, el logaritmo de la función de verosimilitud es

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

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

Ejemplo 2.7. Suponga que $Y_i \sim N(\mu_i, \sigma^2)$, $i = 1, 2, \ldots, n$, con $\mu_i = \beta_0 + \beta_1 x_i$, donde x_i, $i = 1, 2, \ldots, n$, y σ^2 son números reales conocidos. El logaritmo de la función de máxima verosimilitud está dado por

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

Derivando ℓ con respecto a β_0 y β_1, se obtiene

$$\frac{\partial \ell}{\partial \beta_0} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) \tag{2.4}$$

$$\frac{\partial \ell}{\partial \beta_1} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) x_i \tag{2.5}$$

Igualando (2.4) y (2.5) a cero se halla

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

Dado que

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

$\hat{\beta}_0$ y $\hat{\beta}_1$, son las estimaciones de máxima verosimilitud de β_0 y β_1.

Ejercicios 2.3.

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

    a) $X_i \sim \exp(\theta)$
    b) $X_i \sim Pois(\theta)$
    c) $X_i \sim N(\mu, \theta)$, μ conocido
    d) $X_i \sim N(\theta, \sigma^2)$, σ^2 conocido

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

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

4. Suponga que X tiene función de densidad

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

    Calcule

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

c) Halle un estimador de máxima verosimilitud de θ.

5. Sea Y_i, $i = 1, 2, \ldots, n$, una muestra aleatoria de la población

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

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

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

7. Desarrolle el ejercicio 6 de la sección 2.2 por el método de máxima verosimilitud.

8. Sea $\hat{\theta}$ un estimador de máxima verosimilitud de $\theta \in \mathbb{R}$. Si g es una función uno a uno de $\mathbb{R}$ en $\mathbb{R}$, demuestre que $g(\hat{\theta})$ es un estimador de maxima verosimilitud de $g(\theta)$.

9. Suponga que x_i, $i = 1, 2, \ldots, n$, son valores posibles de variables Bernoulli X_i, $i = 1, 2, \ldots, n$, con probabilidad de éxito $\theta_1 = P(X_i = 1)$. Si y_i, $i = 1, 2, \ldots, n$, son valores posibles de variables Bernoulli Y_i, $i = 1, 2, \ldots, n$, con probabilidad de éxito $\theta_2 = (Y_i = 1)$, halle una estimación de máxima verosimilitud para $\theta_1 - \theta_2$.

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 $\hat{\theta} = \frac{50}{250} = 0.2$ como una estimación de la proporción de estudiantes universitarios que hacen deporte, ¿cuál es el error de estimación? Dado que se desconoce la proporción θ de estudiantes que practican deporte, no podemos determinar el error de estimación $\epsilon = |\hat{\theta} - \theta|$ para una estimación particular $\hat{\theta}$ de θ. Sin embargo, podemos preguntarnos, por ejemplo, ¿cuál es la probabilidad que $\epsilon = |\hat{\theta} - \theta|$ sea menor que 0.01 y cuál es el límite probabilístico b tal que $P(|\hat{\theta} - \theta| < b) = 0.95$. Dado que

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

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,

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

En consecuencia, $\frac{b}{\sqrt{(0.2)(0.8)/250}} = 1.96$ y $b = 0.049$. 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 20 %.

Ejercicios 2.4.

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

2. Para determinar cuál de los candidatos A y B a la alcaldía de Bogotá ganará las elecciones, se entrevistaron 900 ciudadanos, elegidos aleatoriamente. Los resultados muestran que el 47 % de los entrevistados votarán por el candidato A y que el 5 % de los entrevistados no votarán. ¿Qué se puede concluir con una confiabilidad del 95 %?

3. 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 y una confiabilidad del 95 %?

4. Sean $Y_1, Y_2, \ldots, Y_n$ variables 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 puede concluir?

5. Suponga que Y_i, $i = 1, 2, \ldots, n$, son 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

Sea Y_i, $i = 1, 2, \ldots, n$, una muestra aleatoria 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 $Var(\hat{\theta}_2) = \frac{\sigma^2}{n-2k}$. 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, $\hat{\theta}_1$ es más eficiente que $\hat{\theta}_2$ si $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$. Se dice que $\hat{\theta}_1$ y $\hat{\theta}_2$ son igualmente eficientes, si $Var(\hat{\theta}_1) = Var(\hat{\theta}_2)$.

A partir de la definición 2.2 se observa que lo ideal es tener estimadores insesgados con la menor varianza posible. Dado un estimador insesgado, podemos determinar si su varianza tiene el menor valor posible a partir del siguiente teorema.

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

$$Var(\hat{\theta}) = \frac{1}{n\ E\left[\left(\frac{\partial \ln f_\theta(x)}{\partial \theta}\right)^2\right]} \tag{2.6}$$

entonces $\hat{\theta}$ es un estimador insesgado de θ de mínima varianza.

Ejemplo 2.8. Sean $X_i, i = 1, 2, \ldots, n$, variables aleatorias con función de distribución exponencial con parámetro $\theta = E(X_i)$. Dado que $Var(\bar{X}) = \frac{\theta^2}{n}$ y $E\left[\left(\frac{\partial \ln f_\theta(x)}{\partial \theta}\right)^2\right] = \frac{1}{\theta^2}$, $\bar{X}$ es un estimador insesgado de θ de mínima varianza.

Ejercicios 2.5.

1. Suponga que $Y_1, Y_2, \ldots, Y_n$ es una muestra aleatoria 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 estimador es más eficiente?

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

4. Sea Y_i, $i = 1, 2, \ldots, n$, una muestra aleatoria de una distribución uniforme en el intervalo $(0, \theta)$. Muestre que $\hat{\theta} = \left(\frac{n+1}{n}\right) Y_{(n)}$ no satisface la igualdad 2.6. Observe que el conjunto $A = \{x : P_\theta(x) > 0\}$ depende de θ.

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

2.6. Consistencia

Sea $Y_i \sim Ber(\theta)$, $i = 1, 2, 3, \ldots$, una sucesión de variables aleatorias independientes. Definimos 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, con $k^2 = \frac{\epsilon^2}{\sigma_n^2}$, se obtiene

$$P(|\bar{Y}_n - \theta| < \epsilon) \geq 1 - \frac{\sigma_n^2}{\epsilon^2} \tag{2.7}$$

donde $\sigma_n^2 = \frac{\sigma^2}{n}$ tiende a 0 cuando n tiende a infinito. Así, tomando el límite cuando n tiende a infinito en 2.7, 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 observaciones de $\bar{Y}_n$ se encuentren cerca de θ. Lo anterior indica que $\bar{Y}_n$ es un estimador consistente de θ de acuerdo con la siguiente definición.

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, $\lim_{n\to\infty} P(|\ \hat{\theta_n} - \theta\ | \leq \epsilon) = 1$.

Ejemplo 2.9. Sea $X_i \sim U(0, \theta)$, $i = 1, 2, \ldots, n$. Dado ϵ tal que $0 < \epsilon < \theta$,

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

Dado que $(1 - \epsilon/\theta)^n$ tiende a 0 cuando n tiende a infinito, $X_{(n)}$ es un estimador consistente de θ. Este resultado se puede generalizar fácilmente para el caso en que X_i, $i = 1, 2, \ldots, n$, tienen distribución creciente en el intervalo $(0, \theta)$.

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

Demostración. Del teorema de Tchebysheff

$$P\left(\left|\hat{\theta}_n - E(\hat{\theta}_n)\right| < k\sqrt{Var(\hat{\theta}_n)}\right) \geq 1 - \frac{1}{k^2} \tag{2.8}$$

haciendo

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

la desigualdad 2.8 toma la forma

$$P(|\hat{\theta}_n - E(\hat{\theta}_n)| < \epsilon) \geq 1 - \frac{Var(\hat{\theta}_n)}{\epsilon^2} \tag{2.9}$$

Tomando el límite cuando n tiende a infinito en 2.9 concluye la demostración. □

Ejemplo 2.10. Sea $Y_i, i = 1, 2, \ldots, n$, una muestra aleatoria 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 tiene a infinito. En consecuencia, $\hat{\theta_n}$ es un estimador consistente de θ.

Ejemplo 2.11. En las secciones anteriores se observó que si $Y_1, Y_2, \ldots, Y_n$ es una muestra aleatoria de una distribución normal con media μ y varianza $\sigma^2 < \infty$,

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

es un estimador insesgado de σ^2. Dado que

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

tiende a 0 cuando n tiene a infinito, S^2 es un estimador consistente de σ^2.

Ejercicios 2.6.

1. Sean Y_i, $i = 1, 2, \ldots, n$, variables aleatorias independientes idénticamente distribuidas, con $E(Y_1) \neq 0$. Muestre 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 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) ¿Es S_k^2 un estimador consistente de σ^2?

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

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

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

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

4. Sea $Y_i, i = 1, 2, \ldots, n$ una muestra aleatoria 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) $\lim_{n \to \infty} P(|\bar{Y} - E(\bar{Y})| < 0.1)$

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 $g(\theta)$. ¿Es $\sqrt{S^2}$ un estimador consistente de σ?

Capítulo 3

Modelos estadísticos

En este capítulo se hace un tratado de estadísticos suficientes y familia exponencial. La sección 1 incluye la definición y el teorema de factorización, que permite determinar fácilmente si un estadístico es suficiente, seguido de algunos ejemplos ilustrativos. En la sección 2 se estudia la familia exponencial uniparamétrica. En la sección 3 se presenta la familia exponencial uniparamétrica en forma natural. La sección 4 introduce la familia exponencial biparamétrica a través de la distribución normal con media y varianza desconocidas. En la sección 5 se define familia exponencial biparamétrica y se estudia su reparametrización en la forma natural. Finalmente, se incluyen algunos ejercicios que permitirán afianzar los conceptos presentados.

3.1. Estadístico suficiente

Sea $X = (X_1, X_2, \ldots, X_n)$, donde X_i, $i = 1, 2, \ldots, n$, tiene función de distribución F_θ asociada a un parámetro desconocido θ. Un estadístico $T(X)$ tiene sentido cuando contiene la misma información que X acerca de 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 una función de distribución F_θ, con θ desconocido. Un estadístico $T(X)$ es suficiente para θ si y solo si $P_\theta(X = X|T = t)$ no depende de θ.

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

Ejemplo 3.1. Se desea obtener información sobre la proporción de tornillos defectuosos θ, producidos por una empresa. Se obtiene una muestra aleatoria con remplazo de la producción, 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. Intuitivamente vemos que dado el número t de tornillos defectuosos, el orden en que aparecen los 1 en la muestra no es relavante para obtener información acerca de θ.

Observe que $T = \sum_{i=1}^{n} X_i$ es un estadístico suficiente para θ. Si $T(X) = t$, $P_\theta(X = X, T = t) = P_\theta(X = X) = \theta^t(1-\theta)^{n-t}$, ya que X es un vector de n realizaciones de $X_i \sim Ber(\theta)$. Así, dado que $T \sim Bin(n, \theta)$,

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

Finalmente, si $T(X) \neq t$, $P_\theta(X = x, T = t) = 0$ y $P_\theta(X = x|T = t) = 0$. En consecuencia, $P_\theta(X = x|T = t)$ no depende de θ.

Ejemplo 3.2. Si en el ejemplo anterior, el muestreo se hace sin reemplazo, la probabilidad $P(X = x, T = t)$ depende únicamente de t y no del orden en que los 1 aparecen en x. En consecuencia, podemos considerar que $x_i = 1$ para $i = 1, 2, \ldots, t$ y $x_i = 0$ para $i = t+1, t+2, \ldots, n$. Así,

$$\begin{aligned} P_\theta(x_1, x_2, \ldots, x_n, T = t) = &\\ \left(\frac{N\theta}{N}\right)\left(\frac{N\theta-1}{N-1}\right)\cdots\left(\frac{N\theta-t+1}{N-t+1}\right) \times &\\ \left(\frac{N-N\theta}{N-t}\right)\left(\frac{N-N\theta-1}{N-t-1}\right)\cdots\left(\frac{N-N\theta-n+t+1}{N-n+1}\right) & \qquad (3.1) \\ = \frac{(N\theta)!(N-N\theta)!(N-n)!}{N!(N\theta-t)!(N-N\theta-n+t)!} & \end{aligned}$$

Finalmente, dado que T tiene distribución hipergeométrica,

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

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.

En el siguiente teorema nos referimos a un conjunto de funciones de probabilidad denominadas modelos regulares, conformados 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, $A = \{x_1, x_2, \ldots\}$, tal que $P_\theta(A) = 1$. Algunos ejemplos de modelos regulares son el modelo binomial, poisson, normal y gamma.

Teorema 3.1. *En un modelo regular, un estadístico $T(X)$ 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 en R^n, tal que*

$$P(X,\theta) = g(t,\theta)h(X) \tag{3.3}$$

Para la demostración del teorema 3.1, es necesario verificar las siguientes proposiciones:

1. *Si T es un estadístico suficiente para θ, $P(X,\theta) = g[T(X),\theta]h(X)$.*

   Demostración 3.1. Sea $X_1, X_2, \ldots$, el conjunto de posibles realizaciones de X y $t_j = T(X_j)$, $j = 1, 2\ldots$ Por hipótesis, $P_\theta(X = X_i|T = t_j)$ no depende de θ, ya que T es un estadístico suficiente. Así, $P_\theta(X = X_i|T = t_j)$ es una función $h(X_i)$. Además, si $T(X_i) = t_j$, entonces $P_\theta(X = X_i, T = t_j) = P_\theta(X = X_i)$, y si $T(X_i) \neq t_j$, $P_\theta(X = X_i, T = t_j) = 0$. En los dos casos, (3.3) se sigue a partir de la igualdad

   $$P_\theta(X = X_i, T = t_j) = P_\theta(X = X_i|T = t_j)P_\theta(T = t_j)$$

   haciendo

   $$g(T(X_i),\theta) = P_\theta(T = t_j)$$

2. *Si $P(X,\theta) = g[T(X),\theta]h(X)$, T es un estadístico suficiente para θ.*

   Demostración 3.2. Dado que $P_\theta(T = t_j) = \sum_{\{X_k|T(X_k)=t\}} P_\theta(X_k)$, aplicando la hipótesis dada por la ecuación $P(X,\theta) = g(T(X),\theta)h(X)$, tenemos que

   $$\begin{aligned} P_\theta(X = X_i|T = t_j) &= \frac{P_\theta(X = X_i, T = t_j)}{P_\theta(T = t_j)} \\ &= \frac{P_\theta(X = X_i)}{\sum_{\{X_k|T(X_k)=t\}} P_\theta(X_k)} \\ &= \frac{h(X_i)}{\sum_{\{X_k|T(X_k)=t\}} h(X_k)} \end{aligned}$$

La segunda igualdad se sigue, si $T(x_i) = t_j$. Así, $P_\theta(X = X_i|T = t_j)$ no depende de θ, lo cual concluye la demostración.

Ejemplo 3.3. Si una función de densidad se puede expresar en la forma $P(x,\theta) = \exp[c(\theta)T(x) + d(\theta) + S(x)]I_A(x)$, $T(X)$ es un estadístico suficiente. Observe que $p(x,\theta)$ se puede expresar en la forma (3.3) del teorema (3.1), con $g(x,\theta) = \exp[c(\theta)T(x) + d(\theta)]$ y $h(x) = \exp[S(x)]I_A(x)$.

Ejemplo 3.4. Sea $X_i \sim U(0,\theta)$, $i = 1, 2, \ldots, n$, una muestra aleatoria de una distribución uniforme. Si $X = (X_1, \ldots, X_n)$,

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

donde $g(x_{(n)},\theta) = \frac{1}{\theta^n} I_{[x_{(n)},\infty)^n}(\theta)$ y $h(x) = 1$.

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 en la forma

$$P(x,\theta) = \exp[c(\theta)T(x) + d(\theta) + S(x)]I_A(x) \tag{3.4}$$

donde c y d son funciones de valor real definidas en el conjunto de parámetros Θ, S y T son funciones de valor real definidas sobre R^n y A no depende de θ.

Ejemplo 3.5. La distribución binomial pertenece a la familia exponencial uniparamétrica de distribuciones. Su función de densidad puede escribirse en la forma (3.4), 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 3.6. La distribución normal con media μ_0 conocida y varianza σ^2 desconocida pertenece a la familia exponencial uniparamétrica. En este caso, la función de densidad puede expresarse en la forma (3.4), donde $c(\sigma^2) = -\frac{1}{2\sigma^2}$, $d(\sigma^2) = -\frac{1}{2}\ln(2\pi\sigma^2)$, $T(x) = (x-\mu_0)^2$ y $S(x) = 0$, para todo x en los números reales.

Si $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes, idénticamente distribuidas, con función de densidad en la familia exponencial uniparamétrica de distribuciones, la función de densidad conjunta $P(x, \theta)$ está dada por

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

donde $x = (x_1, x_2, \ldots, x_n)$ y A está dado por la distribución de X_i, $i = 1, 2, \ldots, n$. Así la distribución conjunta $P(x, \theta)$ puede expresarse en la forma indicada en (3.4), y por tanto pertenece a la familia exponencial uniparamétrica.

Ejemplo 3.7. Observe que si $X_1, X_2, \ldots, X_n$ tienen distribución $Bin(n, \theta)$,

$$\begin{aligned} &P(x_1,\ldots,x_n,\theta) \\ &\quad = \exp\left[\ln\left(\frac{\theta}{1-\theta}\right)\sum_{i=1}^{n} x_i + n^2\ln(1-\theta) + \sum_{i=1}^{n}\ln\binom{n}{x_i}\right] I_{A^n}(x) \end{aligned}$$

Si $X_i \sim N(\theta, \sigma^2)$, $i = 1, 2, 3, \ldots, n$, con media conocida y varianza desconocida, $P(x_1, \ldots, x_n, \theta) = \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu_0)^2 - \frac{n}{2}\ln(2\pi\sigma^2)\right]$.

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

3.3. Familia exponencial uniparamétrica en forma natural

Si una función de densidad pertenece a la familia exponencial uniparamétrica, es decir, si su función de densidad o de probabilidad es de la forma (3.4) y si c es una función 1-1, haciendo $\eta = c(\theta)$, $\theta = c^{-1}(\eta)$. En consecuencia, (3.4) puede ser escrita como

$$P(x, \eta) = \exp[\eta T(x) + d_0(\eta) + S(x)]I_A(x) \tag{3.5}$$

donde $d_0(\eta) = d(c^{-1}(\eta))$. Esta forma de escribir la familia exponencial uniparamétrica de distribuciones es conocida como forma natural de la familia exponencial. η es conocido como parámetro natural de la distribución.

Si c no es una función 1-1, $d_0(\eta)$ se puede determinar fácilmente integrando los dos lados de la ecuación (3.5). Dado que $\int_A P(x,\eta)dx = 1$, se obtiene

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

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

Ejemplo 3.8. La distribución $Bin(n,\theta)$, que pertenece a la familia exponencial uniparamétrica de distribuciones, puede expresarse en la forma (3.5) con $\eta = \ln(\frac{\theta}{1-\theta})$, $d_0(\eta) = n\log\left(\frac{1}{1+\exp(\eta)}\right)$ y $S(x) = \log\binom{n}{x}$, para $A = \{0,1,2,\ldots,n\}$.

Ejemplo 3.9. Si X tiene distribución normal con media θ conocida y varianza σ^2 desconocida, $\eta = -\frac{1}{2\sigma^2}$. Así, $d_0(\eta) = -\frac{1}{2}\ln(-\frac{1}{\eta})$ y

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

Ejemplo 3.10. Suponga que X_i, $i = 1,2,\ldots,n$, es una muestra aleatoria de una distribución geométrica $G(\theta)$. Si $X = (X_1, X_2, \ldots, X_n)$,

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

donde $N = \{1,2,\ldots\}$.

Dado que $\eta = \ln(1-\theta)$,

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

Teorema 3.2. *Si X tiene distribución en la familia exponencial uniparamétrica* (3.5) *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*

$$\Psi(s) = \exp[d_0(\eta) - d_0(\eta+s)] \tag{3.6}$$

Demostración 3.3. Dado que la función de densidad de X está dada por (3.5), para el caso continuo

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

De 3.6 se sigue que $E[T(X)] = -d_0'(\eta)$ y $Var[T(X)] = -d_0''(\eta)$. Estos resultados se obtienen evaluando en 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

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

y

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

Ejemplo 3.11. Sea X_i, $i = 1, 2, \ldots n$, una muestra aleatoria de una distribución Bernoulli con parámetro θ. Dado que $T(X) = \sum_{i=1}^n X_i \sim B(n, \theta)$, por el ejemplo (3.8), $d_0(\eta) = n \log\left(\frac{1}{1+\exp(\eta)}\right)$ 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 3.12. Suponga que X_i, $i = 1, 2, \ldots, n$, es una muestra aleatoria de una distribución normal $N(\theta, \sigma^2)$. Dado que la función de densidad de X puede ser escrita en la forma

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

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

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

Ejemplo 3.13. Suponga que X_i, $i = 1, 2, \ldots, n$, es una muestra aleatoria de una distribución geométrica $G(\theta)$. Si $X = (X_1, X_2, \ldots, X_n)$, la función generadora de momentos de $T(X) = \sum_{i=1}^n X_i$ en el ejemplo (3.10) está dada por

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

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

3.4. Familia exponencial biparamétrica

Sea 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 desconocidas, la función de distribución de X se puede escribir en la forma

$$\begin{aligned}
f_\theta(x) &= (2\pi\sigma^2)^{n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \theta)^2\right] \qquad (3.7) \\
&= (2\pi\sigma^2)^{n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n \left(x_i^2 - 2\theta x_i + \theta^2\right)\right] \\
&= \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2 + \frac{\theta}{\sigma^2}\sum_{i=1}^n x_i - \frac{n\theta^2}{2\sigma^2} + \frac{n}{2}\log(2\pi\sigma^2)\right] \\
&= \exp\left[\eta' T(x) + \xi(\eta)\right]
\end{aligned}$$

donde $\eta' = (\eta_1, \eta_2), T' = (T_1, T_2)$, $\eta_1 = -\frac{1}{2\sigma^2}$, $\eta_2 = \frac{\theta}{\sigma^2}$, $T_1(x) = \sum_{i=1}^n x_i^2$, $T_2(x) = \sum_{i=1}^n x_i$ *y* $\xi(\eta) = -\frac{n}{\sigma^2}\theta^2 - \frac{n}{2}\ln(2\pi\sigma^2)$ En consecuencia, (3.7) puede escribirse en la forma

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

que caracteriza a la familia exponencial biparamétrica, con $S(x_i) = 0$ y $\xi(\eta) = \frac{n\eta_2^2}{4\eta_1} + \frac{n}{2}\ln(\frac{\pi}{\eta_1})$. Así,

$$\frac{\partial \xi}{\partial \eta_2} = -\frac{n\eta_2}{2\eta_1} = -n\theta = -E[T_2(X)]$$
$$\frac{\partial \xi}{\partial \eta_1} = -\frac{n}{4}\frac{\eta_2^2}{\eta_1^2} + \frac{n}{2\eta_1} = -n(\theta^2 + \sigma^2) = -E[T_1(X)]$$

Además

$$\frac{\partial^2 \xi}{\partial \eta_2^2} = -\frac{n}{2\eta_1} = -n\sigma^2 = -Var[T_1(X)]$$
$$\frac{\partial^2 \xi}{\partial \eta_1^2} = \frac{n}{2}\frac{\eta_2^2}{\eta_1^3} + \frac{n}{2\eta_1^2} = -2n\sigma^2(\theta^2 + \sigma^2) = -Var[T_2(X)]$$

Finalmente, a partir de las derivadas de segundo orden, vemos que

$$\frac{\partial^2 \xi}{\partial \eta_1 \partial \eta_2} = \frac{n\eta_2}{2\eta_1^2} = 2n\theta\sigma^2$$

3.5. Familia exponencial biparamétrica en forma natural

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

$$P(x, \eta) = \exp[(c(\theta))^t T(x) + d(\theta) + S(x)] I_A(x) \tag{3.8}$$

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 1-1, haciendo $\eta_1 = c_1(\theta)$ y $\eta_2 = c_2(\theta)$, la función (3.8) puede escribirse en la forma

$$P(x, \eta) = \exp[\eta^t T(x) + d_0(\eta) + S(x)] I_A(x) \tag{3.9}$$

donde $d_0(\eta) = d(c^{-1}(\eta))$. Esta forma de escribir la familia exponencial uniparamétrica de distribuciones es conocida como forma natural de la familia exponencial biparamétrica, donde η es conocido como parámetro natural de la distribución. Si c no es una función 1-1,

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

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:

$$\frac{\partial d_0}{\partial \eta_1} = -T_1(x) \tag{3.10}$$

$$\frac{\partial d_0}{\partial \eta_2} = -T_2(x) \tag{3.11}$$

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.

Ejemplo 3.14. Si X_i, $i = 1, 2, \ldots, n$, es una muestra aleatoria de una distribución normal con media y varianza desconocidas,

$$p_\theta(x) = \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2 + \frac{\theta}{\sigma^2}\sum_{i=1}^{n} x_i + d(\theta) - \frac{n}{2}\ln(2\pi)\right] \tag{3.12}$$

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

Si $\eta = (\eta_1, \eta_2)$, $\eta_1 = -\dfrac{1}{2\sigma^2}$, $\eta_2 = \dfrac{\theta}{\sigma^2}$ y $d_0(\eta) = \dfrac{n}{2}\ln(-2\eta_1) + \dfrac{n}{4}\dfrac{\eta_2^2}{2\eta_1}$, entonces (3.12) puede escribirse en la forma natural y 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 (3.10) son

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

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

$(\hat{\eta}_1, \hat{\eta}_2)^t$ son las estimaciones 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, $\hat{\theta} = \frac{1}{n}T_2(X)$ 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.

Ejercicios 3.1.

1. Muestre que la distribución $P_\theta(x) = (\theta^x e^{-\theta}/x!)I_{\{0,1,2,\ldots\}}(x)$ 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 la función generadora de momentos, expresada como una función del parámetro natural y como una función de θ.

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

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

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

5. Desarrolle el ejercicio 1, si P_θ es la función de densidad

    a) Ber(θ)

    b) Exp(θ)

    c) N(θ, σ^2), σ^2 conocido.

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

6. Desarrolle el ejercicio 5 considerando una muestra aleatoria de tamaño n de cada una de las poblaciones dadas.

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

8. Sea $Y_i \sim F$, $i = 1, 2, \ldots, n$ una muestra aleatoria de una distribución F con parámetro θ. Muestre que las siguientes distribuciones pertenecen a la familia exponencial biparámetrica

   a) gamma
   b) beta
   c) normal.

9. Determine las condiciones necesarias para que exista la estimación de máxima verosimilitud de η en la familia exponencial (3.5).

10. Suponga que $X_1, X_2, \ldots, X_n$ es una muestra de una familia exponencial de la forma $f_\eta(x_i) = \exp[\eta x_i + d_0(\eta) + S(x_i)]I_A(x_i)$. Muestre que el estimador de máxima verosimilitud de $E(X_1)$ es $\hat{d'}(\theta) = \overline{X}$.

Capítulo 4

Propiedades de los estimadores

En este capítulo se introducen procedimientos para encontrar estimadores insesgados óptimos. Inicialmente se introduce el error cuadrático medio y el error medio absoluto como medidas de qué tan bueno es un estimador. Luego se incluyen algunos teoremas cuya aplicación permite encontrar estimadores insesgados uniformes de mínima varianza. Finalmente, este capítulo incluye un tratado sobre el número de información de Fisher, que puede interpretarse como la cantidad de información que contienen las observaciones acerca del parámetro, y mediante el cual expresamos una cota inferior para la varianza de los estimadores.

4.1. Comparación de estimadores

Sea $T_n(X)$ un estimador de $q(\theta)$. Para determinar si $T_n(X)$ es un buen estimador de $q(\theta)$, usualmente se usa el error cuadrático medio, denotado $R(\theta, T)$,

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

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

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

donde $b(\theta, T_n)$ es el sesgo de $T_n(X)$ como estimador de $q(\theta)$. Según este criterio, un buen estimador $T_n(X)$ de $q(\theta)$ deber ser insesgado y de mínima varianza para todo θ en el espacio paramétrico Θ. El error medio absoluto, $R_1(\theta, T) = E_\theta[|T_n(X) - q(\theta)|\ ]$, es otra medida de qué tan bueno es $T_n(X)$ como estimador de $q(\theta)$. Si $T_n(X)$ es un estimador insesgado de $q(\theta)$, normalmente distribuida, la comparación del error medio absoluto con el error cuadrático medio es sencilla. Observe que

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

Otros resultados relacionados con el error cuadrático medio son de gran interés. Por ejemplo, si $T(X)$ es un estimador insesgrado de $q(\theta)$, de la desigualdad (1.4) resulta

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

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

4.2. Completez

Una familia de funciones de distribución $F_\theta(x)$ es completa si y solo si para toda función de valor real U y para todo $\theta \in \Theta$, $E_\theta(U(X)) = 0$ implica $U(x) = 0$ para todo x donde $f_\theta(x) > 0$. Observe que la completez es una propiedad de una familia de distribuciones generada cuando θ varía en un espacio paramétrico Θ.

T es un estadístico completo y suficiente, si T es suficiente para θ y $F_\theta(t)$ es una familia de distribuciones completa. Así, T es un estadístico completo si y solo si la única función de valor real U, definida en el rango de T que satisface $E_\theta(U(T)) = 0$, para todo $\theta \in \Theta$, es la función $U(t) \overset{\text{cs}}{=} 0$.

Ejemplo 4.1. Una variable aleatoria X tiene distribución en la familia de distribuciones hipergeométrica, si su función de densidad está dada por:

$$P_\theta(X = x) = \frac{\binom{N\theta}{x}\binom{N-N\theta}{n-x}}{\binom{N}{n}} \tag{4.2}$$

donde x es un número real tal que $\max(n - N(1-\theta), 0) \leq x \leq \min(N\theta, n)$. Del teorema 3.1 se sigue que $T(X) = X$ es un estadístico suficiente. 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 $P_\theta(U(X) = U(0)) = 1$ y, en consecuencia, $E_\theta(U(X)) = 0$ implica $U(0) = 0$. Si $\theta = 1/N$, $P_\theta(X = x) > 0$ 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 $P_\theta(X = 1) \neq 0$, $U(1) = 0$. Finalmente, procediendo por inducción, se concluye que $U(0) = U(1) = \cdots = U(n) = 0$, es decir, que $U(x) = 0$ para todo x.

Ejemplo 4.2. Sean $X_i \sim U(0, \theta)$, $i = 1, 2, \ldots, n$, $\theta > 0$, variables aleatorias independientes. Dado que la función de densidad conjunta de la variable $X = (X_1, \ldots, X_n)$ es $f_\theta(x) = \frac{1}{\theta^n} I_{(X_{(n)}, \infty)}(\theta)$, del teorema 3.1 se sigue que $T = X_{(n)}$ es un estadístico suficiente para θ. Dado que

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

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, aplicando diferenciación de una integral se encuentra que $u(\theta)\theta^{n-1} \stackrel{\text{cs}}{=} 0$, $\theta \in R^+$. Así, $u(t) \stackrel{\text{cs}}{=} 0$ para $t > 0$ y, en consecuencia, $X_{(n)}$ es una estadístico completo y suficiente.

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

Ejemplo 4.3. Sean $X_i \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$, variables aleatorias independientes. Si la media es desconocida y la varianza conocida, por el teorema (4.1), $T_1(X) = \bar{X}$ es un estimador completo y suficiente de θ.

Si la media es conocida y la varianza desconocida, por el mismo teorema, $T_2(X) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \theta)^2$ es un estimador completo y suficiente de σ^2.

Ejercicios 4.1.

1. Sean X_i, $i = 1, 2, \ldots, n$, variables aleatorias independientes con función de densidad $f_\theta(x) = \theta e^{-\theta x} I_{(0,\infty)}(x)$, $\quad \theta > 0$. Calcule el error cuadrático medio de $T(X) = 1/\bar{X}$ como estimador de θ.
2. Sean $X_i \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$, 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.
3. Sean $X_i \sim U(0, \theta)$, $i = 1, 2, \ldots, n$, variables aleatorias independientes. Determine el error cuadrático medio de $T_1(X) = X_{(n)}$ y $T_2(X) = 2\bar{X}$ como estimadores de θ.
4. Sean X_i, $i = 1, 2, \ldots, n$, 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 $T(X) = \frac{1}{n}X_{(n)}$ como estimador de θ.
5. Sean $X_i \sim Ber(\theta)$, $i = 1, 2, \ldots, n$, variables aleatorias independientes. Demuestre que $T(X) = \sum_{i=1}^{n} X_i$ es un estadístico completo y suficiente para θ.
6. Sean X_i, $i = 1, 2, \ldots, n$, 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 $T(X) = \bar{X}$ es completo y suficiente para θ.

4.3. Estadísticos insesgados uniformes de mínima varianza

Suponga que Γ es 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) de $q(\theta)$.

A continuación se introduce el teorema de Rao Blackwel. Sea $X = (X_1, \ldots, X_n)$ una muestra aleatoria de una distribución P_θ, $\theta \in \Theta$. Si $T(X)$ es una estadística suficiente para θ y $S(X)$ es un estimador de $q(\theta)$, $T^*(X) = E(S(X)|T(X))$ es un estimador de $q(\theta)$ y $b(\theta, T^*) = b(\theta, S)$. Este resultado se sigue fácilmente, ya que si $T(X)$ es un estadístico suficiente para θ, $P(X|T(X) = t)$ no depende de θ. Así,

$$E_\theta(S(X)|T(X) = t) = \int_{\mathbf{X}} S(x)P(x|T(x) = t)dx$$

no depende de θ; por tanto $T^*(X)$ es un estimador de $q(\theta)$. Además,

$$\begin{aligned} b(\theta, T^*) &= E_\theta[E_\theta(S(X)|T(X)] - q(\theta) \\ &= E_\theta(S(X)) - q(\theta) \\ &= b(\theta, S) \end{aligned}$$

Teorema 4.2 (Teorema de Rao Blackwell)**.** *Sea $X = (X_1, \ldots, X_n)$ una muestra aleatoria de una distribución P_θ, $\theta \in \Theta$. Si $T(X)$ es un estadístico suficiente de θ, $S(X)$ un estimador insesgado de $q(\theta)$ tal que $Var_\theta(S(X)) < \infty$ para todo θ, entonces $T^*(X)$ es un estimador de $q(\theta)$ y $Var_\theta(T^*(X)) \leq Var_\theta(S(X))$ para todo θ.*

Demostración 4.1. Observe que

$$\begin{aligned} Var_\theta(S(X)) &= E_\theta[S(X) - q(\theta)]^2 \\ &= E_\theta[S(X) - E_\theta(S(X)|T(X)) + E_\theta(S(X)|T(X)) - q(\theta)]^2 \\ &= E_\theta[S(X) - E_\theta(S(X)|T(X))]^2 + E_\theta[E_\theta(S(X)|T(X)) - q(\theta)]^2 \\ &\geq E_\theta[(S(X)|T(X)) - q(\theta)]^2 \\ &= Var_\theta(T^*(X)) \end{aligned}$$

La igualdad cuenta si $E_\theta[E_\theta(S(X)|T(X)) - S(X)]^2 = 0$, es decir, si $P[E_\theta(S(X)|T(X)) = S(X)] = 1$. Para más detalles de esta demostración ver Dudewicz and Mishara, 1997.

Teorema 4.3 (Teorema de Lehmann-Scheffé)**.** *Si $T(X)$ es un estadístico completo y suficiente de θ, y si $S(X)$ un estimador insesgado de $q(\theta)$ de varianza finita, $T^*(X) = E_\theta(S(X)|T(X))$ es único y es un estimador insesgado de mínima varianza.*

Demostración 4.2. Primero demostramos la unicidad. Suponga que $S_1(X)$ y $S_2(X)$ son estimadores insesgados de $q(\theta)$ y que $T(X)$ es un estadístico completo y suficiente de θ. Si $T_i^*(X) = E_\theta(S_i(X)|T(X))$, $i = 1, 2$, aplicando propiedades de la esperanza condicional se encuentra que

$$\begin{aligned} E_\theta[T_1^*(X) - T_2^*(X)] &= E_\theta[E_\theta(S_1(X)|T(X)) - E_\theta(S_2(X)|T(X))] \\ &= E_\theta(S_1(X)) - E_\theta(S_2(X)) \\ &= 0 \end{aligned}$$

Dado que $T_1^* - T_2^*$ es una función de un estadístico completo T, $T_1^*(X) = T_2^*(X)$. Finalmente, del teorema (4.2) se sigue que $T^*(X)$ es de mínima varianza.

Si T es un estadístico completo y suficiente y $h(T)$ es un estimador insesgado de $q(\theta)$, del teorema (4.3) se sigue que $h(T)$ es un estimador UMVU de $q(\theta)$. En el ejemplo (4.2) se demuestra que si $X_i \sim U(0,\theta)$, $i = 1, 2, \ldots, n$, son variables aleatorias independientes, $X_{(n)}$ es completo y suficiente para θ. Entonces, 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 $S(X) = \frac{n+1}{n}X_{(n)}$ es un estadístico UMVU para θ, ya que $E(S(X)|X_{(n)} = x) = S(x)$.

En este ejemplo, podemos calcular un estimador UMVU de $g(\theta)$, donde g es una función diferenciable en $(0,\infty)$. Dado que $X_{(n)}$ es un estadístico suficiente y completo, basta determinar una función de valor real h tal que $E[h(X_{(n)})] = g(\theta)$. Para lo cual debemos encontrar h tal que

$$E[h(X_{(n)})] = n\theta^{-n}\int_0^\theta h(x)x^{n-1}dx, \theta > 0 \tag{4.3}$$

dado que la funcion de densidad de $X_{(n)}$ está dada por

$$f_{X_{(n)}}(x) = n\theta^{-n}x^{n-1}I_{[0,\theta]}(x)$$

De (4.3) resulta que $\theta^n g(\theta) = n\int_0^\theta h(x)x^{n-1}dx$, expresión en la que 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 4.4. Sean $X_i \sim U(0,\theta)$, $i = 1, 2, \ldots, n$, 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 4.5. Sean $X_i \sim \text{Pois}(\theta)$, $i = 1, 2, \ldots, n$, variables aleatorias independientes. Dado que $T(X) = \sum_{i=1}^n X_i$ es un estadístico suficiente para θ. Por el teorema (4.3), un estadístico UMVU para $q(\theta) = e^{-\theta}$ está determinado por $T^*(X) = E_\theta(S(X)|T(X))$, donde $S(X) = I_{\{X_1=0\}}$ es un estimador insesgado de $q(\theta)$. Observe que

$$
\begin{aligned}
E_\theta\Big(I_{\{X_1=0\}}|\sum_{i=1}^n X_i = t\Big) &= P_\theta\Big(X_1 = 0|\sum_{i=1}^n X_i = t\Big) \\
&= \frac{P(X_1 = 0)P_\theta\Big(\sum_{i=2}^n X_i = t\Big)}{P_\theta\Big(\sum_{i=1}^n X_i = t\Big)} \\
&= \left(1 - \frac{1}{n}\right)^t
\end{aligned}
$$

Entonces, $T^*(X) = \left(1 - \frac{1}{n}\right)^T$.

Ejercicios 4.2.

1. En el ejemplo 4.3 de la sección 4.2, demuestre que $T_1(X) = \bar{X}$ y $T_2(X) = \frac{1}{n}\sum_{i=1}^n (X_i - \theta)^2$ son estadísticos UMVU de θ y σ^2, respectivamente.

2. En el ejercicio 6 de la sección 4.2 demuestre que $T(X) = \bar{X}$ es un estadístico UMVU de θ.

3. Sea X_i, $i = 1, 2, \ldots, n$, una muestra aleatoria de una población $\Gamma(p, \lambda)$, definida como en 1.13, donde p y λ son desconocidos. Encuentre un estimador UMVU para p/λ.

4.3.1. Desigualdad de información

En esta sección nos referimos a funciones P_θ en la familia de distribuciones $\{P_\theta : \theta \in \Theta\}$ tales que:

a) el conjunto $A = \{x : P_\theta(x) > 0\}$ no depende de θ.

b) para todo $x \in A$ y todo $\theta \in \Theta$, $\frac{\partial}{\partial\theta} log P_\theta(x)$ existe y es finito.

c) si T es un estadístico tal que $E_\theta(|T|) < \infty$ para todo $\theta \in \Theta$,

$$
\frac{\partial}{\partial\theta}\left[\int_{\mathbb{R}^n} T(x)P_\theta(x)dx\right] = \int_{\mathbb{R}^n} T(x)\frac{\partial}{\partial\theta}P_\theta(x)dx \tag{4.4}
$$

Definición 4.1. Sea $X \sim P_\theta$ tal que las propiedades a y b son válidas. Definimos el número de información de Fisher, $I(\theta)$, por medio de la expresión

$$
I(\theta) = E_\theta\left[\frac{\partial}{\partial\theta}\log P(X, \theta)\right]^2
$$

$I(\theta)$ puede interpretarse como la cantidad de información acerca de θ que contiene una observación x de X.

Algunos resultados de interés asociados con el número de información de Fisher son los siguientes:

1. Si se satisfacen las propiedades a, b y c, $E_\theta\left[\frac{\partial}{\partial\theta}\log P(X,\theta)\right] = 0$ y, en consecuencia, $I(\theta) = Var_\theta[\frac{\partial}{\partial\theta}\log P(X,\theta)]$.

2. Si $X = (X_1, X_2, \ldots, X_n)$ es una muestra aleatoria con función de densidad $P(x,\theta)$, entonces
$$Var_\theta\left(\frac{\partial}{\partial\theta}\log P(X,\theta)\right) = \sum_{i=1}^{n} Var_\theta\left[\frac{\partial}{\partial\theta}\log P(X_i,\theta)\right]$$
y $I(\theta) = nI_1(\theta)$.

3. Suponga que $I(\theta) < \infty$ y que $\frac{\partial^2}{\partial\theta^2}\log P(x,\theta)$ existe para todo x y para todo θ. 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

$$\begin{aligned}
E_\theta\left[\frac{\partial^2}{\partial\theta^2}\log P(x,\theta)\right] &= -\int_A \frac{1}{P(x,\theta)}\left(\frac{\partial}{\partial\theta}P(x,\theta)\right)^2 dx + \int_A \frac{\partial^2}{\partial\theta^2}P(x,\theta)dx \\
&= -E_\theta\left[\left(\frac{\partial}{\partial\theta}logP(x,\theta)\right)^2\right] + \frac{\partial^2}{\partial\theta^2}\int_A P(x,\theta)dx \\
&= -E_\theta\left[\left(\frac{\partial}{\partial\theta}logP(x,\theta)\right)^2\right] = -I(\theta)
\end{aligned}$$

Ejemplo 4.6. Si X tiene distribución Poisson con parámetro θ,
$$\begin{aligned}
I_1(\theta) &= E_\theta\left[\frac{\partial}{\partial\theta}(xlog(\theta) - \theta - log(x!))\right]^2 \\
&= E\left[\left(\frac{x-\theta}{\theta}\right)^2\right] \\
&= \frac{1}{\theta^2}Var_\theta(X) \\
&= \frac{1}{\theta}
\end{aligned}$$

Otra forma de calcular $I_1(\theta)$ es

$$\begin{aligned} I_1(\theta) &= -E_\theta\left[\frac{\partial^2}{\partial\theta^2}(xlog(\theta) - \theta - log(x!))\right] \\ &= \frac{1}{\theta^2}E_\theta(X) \\ &= \frac{1}{\theta} \end{aligned}$$

El siguiente teorema establece una cota inferior para la varianza, denominada cota de Crámer-Rao. En él, X denota el vector con componentes X_i, $i = 1, 2, \ldots, n$, independientes e identicamente distribuidas, e $I(\theta)$ la información de Fisher dada por X.

Teorema 4.4 (Teorema de desigualdad de información)**.** *Sea $T(X)$ un estadístico tal que $Var_\theta T(X) < \infty$ para todo θ. Bajo propiedades de regularidad, si $0 < I(\theta) < \infty$, $Var_\theta[T(X)] \geq \frac{[\Psi'(\theta)]^2}{I(\theta)}$, donde $\Psi(\theta) = E_\theta[T(X)]$.*

Demostración 4.3. Dadas las hipótesis,

$$\begin{aligned} 1 &\geq \frac{\left|Cov\left[T(X), \frac{\partial}{\partial\theta}\log P(X,\theta)\right]\right|}{\sqrt{Var[T(X)]Var\left[\frac{\partial}{\partial\theta}\log P(X,\theta)\right]}} \\ &= \frac{\left|\frac{\partial}{\partial\theta}E[T(X)]\right|}{\sqrt{Var[T(X)]I(\theta)}} \end{aligned}$$

Calculando los cuadrados, el teorema queda demostrado.

Ejemplo 4.7. Si $X \sim N(\theta, \sigma^2)$ y σ^2 es conocido,

1. $A = \mathbb{R}$ y no depende de θ.
2. $\frac{\partial}{\partial\theta}lnP(x,\theta) = \frac{1}{\sigma^2}(x-\theta)$ es finito para todo $(x,\theta) \in \mathbb{R}\times\mathbb{R}$.

En este caso, si $T(X) = X$ se satisface la igualdad (4.4) y $Var_\theta(T(X)) = 1/I(\theta)$, donde $I(\theta) = 1/\sigma^2$.

Ejercicios 4.3.

1. Sea X_i, $i = 1, 2, \ldots, n$, variables aleatorias 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 $I(\theta)$.
2. Desarrolle el ejercicio 1 con la función de distribución

a) Gamma con parámetros p y λ como en 1.13, $\theta = p$ y α fijo.

b) Gamma, reparametrizada como en el ejercicio 6 de la sección 4.2.

3. Si $X_i \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$, encuentre un estimador insesgado de θ^2. Determine $Var(\bar{X}^2)$. Halle la cota de Crámer-Rao.

4. Si $X \sim \text{pois}(\theta)$, $Y = I_{\{0\}}(X)$ tiene distribución Ber$(e^{-\theta})$. Verifique la desigualdad de información considerando a

   a) θ como parámetro y a $T(X) = I_{\{0\}}(X)$ como estimador.

   b) $e^{-\theta}$ como parámetro y a $T(Y) = Y$ como estimador.

5. Sea X_i, $i = 1, 2, \ldots, n$, una muestra aleatoria de una población P_θ. Halle la cota de Crámer-Rao para la varianza del estimador

   a) $T(X) = \bar{X}$, asumiendo que $X_i \sim G(p, \lambda)$.

   b) $T(X) = \bar{X}$, asumiendo que $X_i \sim \exp(\theta)$.

   c) $T(X) = X_1$, asumiendo que $X_i \sim N(\mu, \sigma^2/n)$.

Capítulo 5

Intervalos de confianza

Para incluir una introducción a intervalos de confianza para los parámetros de una distribución P_θ, consideramos el siguiente ejemplo: si $X \sim U(0,\theta)$, $\frac{X}{\theta} \sim U(0,1)$. Entonces, si α es un número real positivo menor que 0.5,

$$P\left(\frac{\alpha}{2} \leq \frac{X}{\theta} \leq 1-\alpha/2\right) = P\left(\frac{X}{1-\alpha/2} \leq \theta \leq \frac{X}{\alpha/2}\right) = 1-\alpha$$

Al intervalo

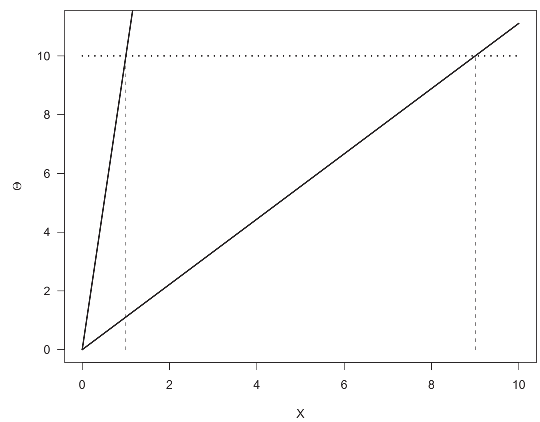
$$\left(\frac{X}{1-\alpha/2}, \frac{X}{\alpha/2}\right) \tag{5.1}$$

se le denomina intervalo de confianza. Este es un intervalo aleatorio y a cada valor x de la variable X le corresponde un valor $(\frac{x}{1-\alpha/2}, \frac{x}{\alpha/2})$ del intervalo, el cual no necesariamente contiene a θ. Por ejemplo, para $\theta = 10$ y $\alpha = 0.05$, si $x = 9.8$, el intervalo de confianza observado es $(10.05, 392)$, el cual no contiene a θ. En realidad, en el límite, cuando el número de intervalos tiende a infinito, el $100(1-\alpha)\,\%$ de ellos contiene a θ.

En figura 5.1 se considera la variable aleatoria $X \sim U(0,10)$ y se determinan los valores de X para los cuales el intervalo 5.1, con $\alpha = 0.2$, contiene a θ. Observe que $\theta \in (\frac{x}{1-\alpha/2}, \frac{x}{\alpha/2})$ si y sólo si $1 < x < 9$.

5.1. Intervalos de confianza para la media de una distribución normal

En esta sección consideramos intervalos de confianza para la media de una distribución normal con varianza conocida, en 5.1.1, y con varianza desconocida, en 5.1.2.



Gráfica 5.1. Valores de X para los cuales $\theta \in (\frac{x}{1-\alpha/2}, \frac{x}{\alpha/2})$.

5.1.1. Intervalos de confianza para la media de una distribución normal con varianza conocida

Sean Y_i, $i = 1, \ldots, n$, variables aleatorias independientes normalmente distribuidas con media desconocida θ, y varianza conocida σ^2. Si $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ y $\sigma_{\bar{Y}}$ denota la desviación estándar de $\bar{Y}$,

$$Z = \frac{\bar{Y} - \theta}{\sigma_{\bar{Y}}}$$

tiene distribución normal estándar. Así, para todo número real $\alpha \in (0,1)$, existe un número real positivo $z_{\alpha/2}$ tal que

$$P\left(-z_{\frac{\alpha}{2}} < \frac{\bar{Y} - \theta}{\sigma_{\bar{Y}}} < z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

En consecuencia,

$$P\left(\bar{Y} - z_{\frac{\alpha}{2}}\sigma_{\bar{Y}} < \theta < \bar{Y} + z_{\frac{\alpha}{2}}\sigma_{\bar{Y}}\right) = 1 - \alpha \tag{5.2}$$

donde $\bar{Y} - z_{\frac{\alpha}{2}}\sigma_{\bar{Y}}$ y $\bar{Y} + z_{\frac{\alpha}{2}}\sigma_{\bar{Y}}$ determinan un intervalo de confianza para θ. En este intervalo, los estadísticos $\underline{L} = \bar{Y} - z_{\frac{\alpha}{2}}\sigma_{\bar{Y}}$ y $\bar{L} = \bar{Y} + z_{\frac{\alpha}{2}}\sigma_{\bar{Y}}$ se llaman límite inferior y límite superior del intervalo, respectivamente.

$\underline{L}$ y $\bar{L}$ son variables aleatorias que determinan un intervalo, que tiene probabilidad $1 - \alpha$ de contener el parámetro θ. $1 - \alpha$ 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 $[\underline{L}(\bar{y}), \bar{L}(\bar{y})]$ contiene o no contiene a θ. En este caso, $1-\alpha$ corresponde a la incertidumbre que tenemos 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 $[\underline{L}(\bar{y}), \bar{L}(\bar{y})]$. Además, $1-\alpha$ puede interpretarse como la probabilidad de que al observar $\bar{Y}$, el error de observación sea menor que $\frac{z_{\alpha/2}\sigma}{\sqrt{n}}$.

Como $\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}}$, la longitud del intervalo $(\bar{Y} - z_{\frac{\alpha}{2}}\sigma_{\bar{Y}} < \theta < \bar{Y} + z_{\frac{\alpha}{2}}\sigma_{\bar{Y}})$ puede hacerse tan pequeña como se quiera, siempre que n sea suficientemente grande.

$\underline{L}(Y) = \bar{Y} - z_\alpha \sigma_{\bar{Y}}$ es llamado una cota inferior de nivel α para θ. Observe que $P(\underline{L}(Y) \leq \theta) = 1 - \alpha$. De igual forma, podemos definir una cota superior $\bar{L}(Y)$ como $\bar{Y} + z_\alpha \sigma_{\bar{Y}}$.

Aquí hemos considerado el intervalo de confianza determinado por $\underline{L}$ y $\bar{L}$, definidos en 5.2, ya que este es el intervalo más corto de nivel de confiabilidad $1-\alpha$, $0 < \alpha < 0.5$. Para demostrar esta afirmación, podemos expresar la longitud del intervalo como una función de una variable t, con $0 < t < 1$

$$h(t) = z_{t\alpha} + z_{(1-t)\alpha} = \Phi^{-1}(1 - t\alpha) + \Phi^{-1}(1 - (1-t)\alpha)$$

Derivando con respecto a t

$$h'(t) = -[\Phi^{-1}(1 - t\alpha)]'\alpha + [\Phi^{-1}(1 - (1-t)\alpha)]'\alpha$$

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 $t = \frac{1}{2}$.

A continuación se presenta otra forma de demostrar este resultado. 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. Esto es equivalente 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

$$H'_{z_1}(z_1, z_2, \lambda) = 1 + \lambda f(z_1) \quad \text{y} \quad H'_{z_2}(z_1, z_2, \lambda) = 1 + \lambda f(z_2)$$

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

$$2\int_0^{z_1} f(x)dx = 1-\alpha \quad \text{y} \quad \int_{z_1}^{\infty} f(x)dx = \frac{\alpha}{2}$$

Este resultado indica por qué, por ejemplo, con una confiabilidad del $100(1-\alpha)\,\%$ se considera como intervalo de confianza para la media

$$\left(\bar{Y} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{Y} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

y no 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 ejemplo de estimación de intervalos de confianza cuando la varianza es conocida, suponga que una muestra aleatoria de tamaño $n = 16$ es obtenida de una población normal con varianza $\sigma^2 = 4$. Si la media muestral es $\bar{x} = 11.5$, 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. 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?

   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, determine intervalos de confianza para 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 $f_\theta(x) = \frac{1}{\theta}e^{-\frac{1}{\theta}x}I_{\{x>0\}}(x)$, $\theta > 0$, determine los límites aleatorios $\underline{L}(X)$ y $\bar{L}(X)$ tales que $P_\theta(\underline{L}(X) < \theta) = P_\theta(\bar{L}(X) > \theta) = \alpha$, para $\alpha < 0.5$.

5. Sean $X_i \sim U(0,\theta), i = 1, 2, \ldots, n$, variables aleatorias independientes. Determine un intervalo de confianza para θ en función de $X_{(n)}$.

6. Sean $X_i, i = 1, 2, \ldots, n$, variables aleatorias independientes con función de densidad $f_\theta(x) = \frac{1}{\theta}e^{-\frac{1}{\theta}x}$, $x > 0$, $\theta > 0$.

   a) Determine un intervalo de confianza para θ a partir de la distribución *ji*-cuadrado, y aplicando el teorema del límite central.
   b) Determine un intervalo de confianza para $e^{-\frac{1}{2}\theta}$.

7. Sean $X_i, i = 1, 2, \ldots, n$, variables aleatorias independientes normalmente distribuidas $N(\theta,\theta)$. Determine un intervalo de confianza para θ. Compare su resultado con el intervalo $\left(\bar{Y} - z_{\alpha/2}\sqrt{\frac{\bar{Y}}{n}}, \bar{Y} + z_{\alpha/2}\sqrt{\frac{\bar{Y}}{n}}\right)$.

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 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

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

Dado que $(n-1)S^2/\sigma^2$ tiene distribución χ^2_{n-1}, a partir de la definición de la distribución t vemos que

$$T = \frac{\bar{Y} - \theta}{S/\sqrt{n}}$$

tiene distribución τ con n-1 grados de libertad. En consecuencia, existe un número real positivo $t_{n-1,\frac{\alpha}{2}}$ tal que $P(-t_{n-1,\frac{\alpha}{2}} < T < t_{\frac{n-1,\alpha}{2}}) = 1 - \alpha$, y así,

$$P\left(\bar{Y} - t_{n-1,\frac{\alpha}{2}}\left(\frac{S}{\sqrt{n}}\right) < \theta < \bar{Y} + t_{n-1,\frac{\alpha}{2}}\left(\frac{S}{\sqrt{n}}\right)\right) = 1 - \alpha$$

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 en forma aproximada $t_{n-1,\frac{\alpha}{2}}$, utilizamos la tabla de distribución τ. Por ejemplo, si $n = 9$ y $\alpha = 0.05$, $t_{8,0.025} = 2.306$.

Cuando n tiende a infinito, T tiende a una distribución normal estándar. En consecuencia, esta distribución podrá utilizarse para encontrar los intervalos de confianza.

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, ¿qué se puede concluir con una confiabilidad del 98 %?

2. Al medir la cantidad de azúcar contenida en cada uno de los ocho sobres de una muestra aleatoria de la producción de una empresa empacadora, se obtuvo 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 confiabilidad del 98 %. Si el contenido deseado es 11.1 gramos, ¿qué se puede afirmar 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 con una varianza $s^2 = 260$. ¿Qué afirmaciones podemos hacer respecto de esta producción?

5.2. Intervalos de confianza aproximados para la probabilidad de éxito

Sea Y_i, $i = 1, 2, \dots, n$, una muestra aleatoria de una distribución binomial con probabilidad de éxito θ. Dado que para n "grande"

$$Z = \frac{\sqrt{n}(\bar{Y} - \theta)}{\sqrt{\theta(1-\theta)}} \tag{5.3}$$

tiene aproximadamente distribución normal estándar (ver teorema 1.3),

$$\begin{aligned} P\left(\left|\frac{\sqrt{n}(\bar{Y}-\theta)}{\sqrt{\theta(1-\theta)}}\right| < z_{\frac{\alpha}{2}}\right) &= P\left(n\left(\bar{Y}-\theta\right)^2 < z^2_{\frac{\alpha}{2}}\theta(1-\theta)\right) \\ &= P\left(\theta^2(n + z^2_{\alpha/2}) - \theta(2n\bar{Y} + z^2_{\alpha/2}) + n\bar{Y}^2 < 0\right) \\ &= 1 - \alpha \end{aligned}$$

Resolviendo en θ la desigualdad $\theta^2(n + z^2_{\alpha/2}) - \theta(2n\bar{Y} + z^2_{\alpha/2}) + n\bar{Y}^2 < 0$, se obtiene:

$$\bar{L}(Y) = \frac{(2n\bar{Y} + z^2_{\alpha/2}) + z_{\alpha/2}\sqrt{4n\bar{Y} + z^2_{\alpha/2} - 4n\bar{Y}^2}}{2(n + z^2_{\alpha/2})} \tag{5.4}$$

$$\underline{L}(Y) = \frac{(2n\bar{Y} + z^2_{\alpha/2}) - z_{\alpha/2}\sqrt{4n\bar{Y} + z^2_{\alpha/2} - 4n\bar{Y}^2}}{2(n + z^2_{\alpha/2})} \tag{5.5}$$

Dado que el coeficiente de θ^2 es positivo, la desigualdad se cumple para $\underline{L} < \theta < \bar{L}$; por tanto $(\underline{L}, \bar{L})$ es un intervalo de confianza aproximado de nivel $1-\alpha$ para θ. Si n es pequeño, se deben utilizar intervalos de confianza exactos, obtenidos a partir de la distribución binomial.

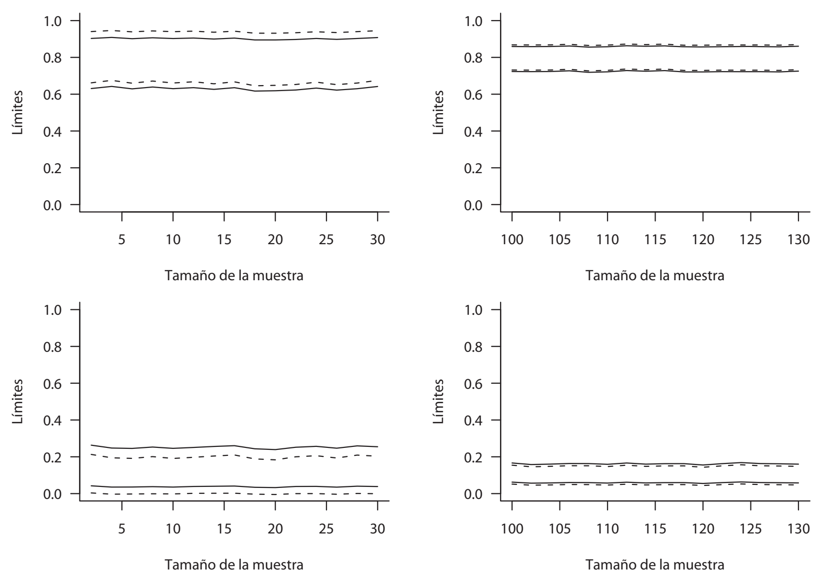
Otra forma común de hallar intervalos de confianza para θ es utilizar $\hat{\theta} = \bar{Y}$ para estimar la varianza. Entonces, dado que para valores "grandes" de n, (5.3) tiene aproximadamente distribución normal estándar,

$$\begin{aligned} &P\left(-z_{\alpha/2} < \frac{\sqrt{n}(\bar{Y}-\theta)}{\sqrt{\theta(1-\theta)}} < z_{\alpha/2}\right) = \\ &\qquad P\left(\bar{Y} - z_{\alpha/2}\sqrt{\frac{\theta(1-\theta)}{n}} < \theta < \bar{Y} + z_{\alpha/2}\sqrt{\frac{\theta(1-\theta)}{n}}\right) = 1 - \alpha \end{aligned}$$

Así, reemplazando θ por $\hat{\theta}$ para estimar la varianza, se encuentra que el intervalo de confianza del $100(1-\alpha)\,\%$ para θ es

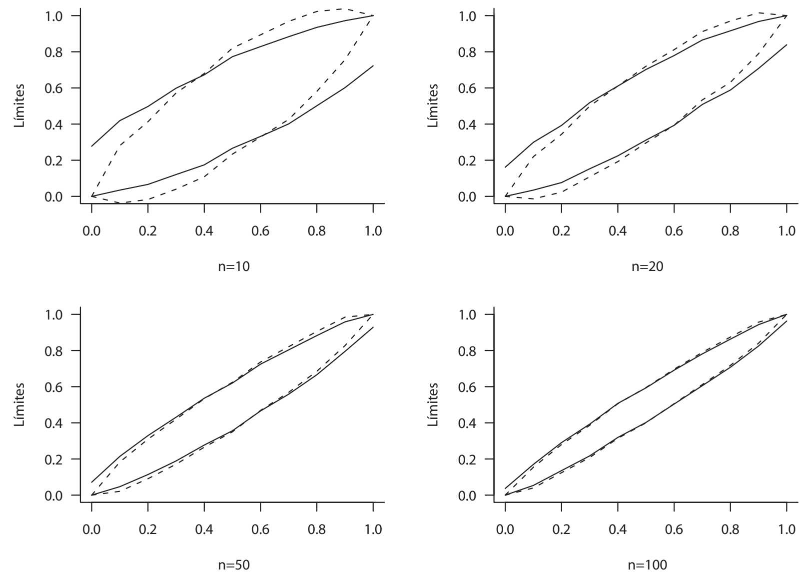
$$\bar{Y} - z_{\alpha/2}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} < \theta < \bar{Y} + z_{\alpha/2}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} \tag{5.6}$$

Agresti & Coull. (1998) muestran que el intervalo de confianza definido por las ecuaciones 5.4 y 5.5 puede ser recomendado para el uso con casi todos los tamaños de muestra y valores de θ. Muestran también que las estimaciones dadas por 5.6 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 5.6 tienen a $\hat{\theta}$ como punto medio.



Gráfica 5.2. La línea continua corresponde a estimaciones de intervalos esperados de confianza dadas por las ecuaciones 5.4 y 5.5. La línea discontinua a estimaciones dadas por 5.6.

La figura 5.2 muestra las diferencias entre las estimaciones de los intervalos de confianza esperados para $\theta = 0.1$ y $\theta = 0.8$ y que las estimaciones dadas por 5.4 y 5.5 no se encuentran centradas en los valores de θ, presentando un desplazamiento hacia $\theta = 1/2$, que concuerda con el sesgo de la distribución binomial para estos valores del parámetro. La figura 5.3 muestra estimaciones de los intervalos de confianza esperados para los valores θ, representados en el eje de las abscisas. Indica grandes diferencias entre las estimaciones dadas por las ecuaciones 5.4 y 5.5 y las obtenidas por 5.6 para valores pequeños del tamaño de muestra. Las diferencias se hacen menores a medida que el tamaño de la muestra aumenta.



Gráfica 5.3. Estimaciones de intervalos esperados de confianza para una proporción. La línea continua corresponde a las estimaciones dadas por las ecuaciones 5.4 y 5.5. La línea discontinua, a estimaciones dadas por 5.6.

Ejemplo 5.1. En una muestra aleatoria de 400 estudiantes de una universidad oficial en Colombia, 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 de reemplazar, en (5.6), $\hat{\theta}$ y $\bar{Y}$ por 0.2. Cuando se utilizan las ecuaciones (5.4) y (5.5) para calcular los límites del intervalo de confianza del 95 % para θ, se encuentra que $0.163 < \theta < 0.242$.

Agresti & Coull. (1998) también proponen una variación de 5.6 para estimar intervalos de confianza. En esta propuesta, si $X = \sum_{i=1}^{n} X_i \sim Bin(n, \theta)$, el intervalo de confianza para θ está dado por

$$\tilde{\theta} \pm z_{\alpha/2}\sqrt{\frac{\tilde{\theta}(1-\tilde{\theta})}{n}} \tag{5.7}$$

donde $\tilde{\theta} = (X + 2)/(n + 4)$. La expresión 5.7 resulta también bastante apropiada para la estimación de los intervalos de confianza (Wei 2002).

5.3. Intervalo de confianza para $\theta_1 - \theta_2$

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 una población normal con media θ_2 y varianza σ_2^2. Si σ_1^2 y σ_2^2 son conocidas, de la sección 5.1.1 se concluye que

$$Z = \frac{\left(\bar{Y}_1 - \bar{Y}_2\right) - (\theta_1 - \theta_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$$

tiene distribución normal estándar. Por tanto, procediendo como en la sección 5.1.1, el intervalo de confianza para θ 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 $\theta_1 - \theta_2$ de las medias poblacionales, 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

$$S^2 = \frac{(n-1)\,S_1^2 + (m-1)\,S_2^2}{n+m-2} \tag{5.8}$$

donde

$$S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_{1i} - \bar{Y}_1)^2$$

y

$$S_2^2 = \frac{1}{m-1}\sum_{i=1}^{m}(Y_{2i} - \bar{Y}_2)^2$$

Observe que bajo la hipótesis $H : \sigma_1^2 = \sigma_2^2 = \sigma^2$, $\frac{(n+m-2)S^2}{\sigma^2} = \frac{(n-1)S_1^2}{\sigma^2} + \frac{(m-1)S_2^2}{\sigma^2}$ tiene distribución χ^2 con $n+m-2$ grados de libertad. Así, de 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 t 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

$$\bar{L} = \left(\bar{Y}_1 - \bar{Y}_2\right) + t_{n+m-2,\,\alpha/2} S\sqrt{\frac{1}{n} + \frac{1}{m}}$$

5.4. Intervalos de confianza aproximados para la diferencia de probabilidades

Sea $X_i \sim Ber(p_1)$, $i = 1, 2, 3, \ldots, n_1$, y $X_j \sim Ber(p_2)$, $j = 1, 2, 3, \ldots, n_2$, variables aleatorias independientes. Si $q_i = 1 - p_i$, $i = 1, 2$, dado que

$$\frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \xrightarrow{d} N(0,1) \tag{5.9}$$

para valores grandes de n_1 y n_2,

$$P\left(-z_{\alpha/2} \leq \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} - \frac{p_2 q_2}{n_2}}} \leq z_{\alpha/2}\right) \approx 1 - \alpha$$

Así, el intervalo con límites determinados por

$$\underline{L}(X,Y) = (\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} \tag{5.10}$$

y

$$\bar{L}(X,Y) = (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} \tag{5.11}$$

tiene una confiabilidad aproximada del $100(1-\alpha)\,\%$.

Al igual que el intervalo de confianza dado por 5.6, el intervalo de confianza dado por 5.10 y 5.11 no es recomendable para la estimación de intervalos de confianza de la diferencia de dos proporciones. Wei (2002) generalizó el intervalo de confianza 5.7 para la estimación de intervalos de confianza de la diferencia de dos proporciones. Su propuesta está dada por las ecuaciones 5.12 y 5.13, y en general es recomendable para todos los valores de θ y todos los tamaños de muestra.

$$\underline{L}(X,Y) = (\tilde{p}_1 - \tilde{p}_2) - z_{\alpha/2}\sqrt{\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \frac{\tilde{p}_2\tilde{q}_2}{n_2}} \tag{5.12}$$

y

$$\bar{L}(X,Y) = (\tilde{p}_1 - \tilde{p}_2) + z_{\alpha/2}\sqrt{\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \frac{\tilde{p}_2\tilde{q}_2}{n_2}} \tag{5.13}$$

En estas ecuaciones, $\tilde{p}_i = (X_i + 1)/(n_i + 2)$, $i = 1, 2$.

Ejercicios 5.3.

1. Dos máquinas A y B producen el mismo tipo de artículos. Se debe parar una de ellas. Al seleccionar muestras aleatorias independientes de 150 artículos, para cada una, se encuentran 12 defectuosos en la muestra de la producción de la máquina A y 24 en la de la máquina B. ¿Existe evidencia suficiente para parar la máquina B? Concluya con niveles de significancia del 90 % y del 95 %.

5.5. Intervalos de confianza para σ^2

Si Y_i, $i = 1, 2, \ldots, n$, son variablea aleatorias independientes normalmente distribuidas con media θ y varianza σ^2 desconocidas, 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 una confiabilidad 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 corto que corresponde a una probabilidad de $1-\alpha$, se acostumbra escoger x_1^2 y x_2^2 tales que $P(\chi^2_{n-1} < x_1^2) = P(\chi^2_{n-1} > x_2^2) = \alpha/2$.

Ejercicios 5.4.

1. Sea Y_i, $i = 1, 2, \ldots, n$, variables aleatorias con distribución normal con media θ, conocida, y varianza σ^2 desconocidas. Determine un intervalo de confianza de nivel $1 - \alpha$ para σ^2.

2. Sea $X_i, i = 1, 2, \ldots, n$, variables aleatorias con función de densidad Rayleigh

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

   Demuestre que $Y_i = (X_i/\theta)^2$ tiene distribución χ_2^2 y determine un intervalo de confianza para θ.

5.6. 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 $Y_i \sim N(\theta, \sigma_2^2)$, $i = 1, 2, \ldots, n_2$, son muestras aleatorias 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

$$P\left(a < \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} < b\right) = 1 - \alpha$$

a y b se eligen de modo que $P(F < a) = P(F > b) = \alpha/2$. A partir de la ecuación (5.14) se encuentra que

$$P\left(\frac{S_1^2}{bS_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{aS_2^2}\right) = 1 - \alpha$$

Así, los límites superior e inferior del intervalo de confianza están dados por

$$\underline{L} = \frac{S_1^2}{bS_2^2} \quad \text{y} \quad \bar{L} = \frac{S_1^2}{aS_2^2} \tag{5.14}$$

Ejemplo 5.2. Se observan dos procesos A y B de producción de tornillos. Una muestra aleatoria de 8 tornillos del proceso A tiene una varianza $S^2 = 3.6$ y una muestra aleatoria de 10 tornillos del proceso B, una varianza

$S^2 = 4.7$. Para hallar un intervalo de confianza del 90 % para σ_A/σ_B, inicialmente se eligen numeros reales a y b, como se indica en la sección 1.2.5. Finalmente, reemplazando en 5.14, se obtiene el intervalo $(0.23, 2.82)$.

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 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 5.1.

Medición	1	2	3	4	5	6	7
Diametro (cm)	50	53	55	47	48	53	54

Tabla 5.1. Diametro de pinos.

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 8, tomados aleatoriamente de la producción de una empresa empacadora, resultaron los que se muestran en la tabla 2. Halle un intervalo de confianza para σ^2 y estime el error en este proceso de producción.

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 5.2. Peso en gramos de sobres de azúcar.

3. Sea X_i, $i = 1, 2, \ldots, n$, una muestra aleatoria 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 $\bar{L}$ tales que $\sigma^2 \in (\underline{L}, \infty)$ con probabilidad $1 - \alpha_1$ y $\sigma^2 \in (-\infty, \bar{L})$ con probabilidad $1 - \alpha_2$. Demuestre que $\sigma^2 \in (\underline{L}, \bar{L})$ con probabilidad $1 - \alpha_1 - \alpha_2$.

4. Calcule $E(S^2)$ y $Var(S^2)$, donde S^2 está definido en (5.8).

Capítulo 6

Prueba de hipótesis

Este capítulo se encuentra dividido en 7 secciones. En las dos primeras se hace una introducción a pruebas de hipótesis. Luego, incluye un tratado de probabilidad de error, regiones de rechazo, intervalos de confianza y pruebas de hipótesis. Finalmente, incluye ejemplos de planteamiento de hipótesis compuestas, y ejercicios.

6.1. Prueba de hipótesis simples

Para garantizar la continuidad de un contrato, una empresa de publicidad afirma que durante el primer mes de 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 "aceptamos" H, concluimos que la publicidad no ha tenido éxito y se debería suspender el contrato. Si "aceptamos" K, y hay interés en reducir el número de fumadores, se debería 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 una muestra aleatoria de potenciales fumadores y se determina el número $T(x)$ de fumadores en la muestra. Si este 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 p_0, $0 \leq p_0 \leq 1$, se rechaza la hipótesis H y, en consecuencia, se prórroga el contrato.

Observe que se ha determinado una región de rechazo de la hipótesis H, conformada por el conjunto $RR = \{0, 1, \ldots, k\}$. Esto significa que se rechaza H siempre que $T(x) \in RR$. En caso contrario, si $T(x) \notin RR$ (o si la proporción de fumadores en la muestra es mayor que p_0), no se debe prorrogar el contrato y se afirmará 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 errores en su decisión. El primero consiste en prorrogar el contrato dado que la publicidad no es efectiva. El segundo, suspender el contrato dado que la publicidad ha dado los resultados esperados. Si el Ministerio desea reducir las 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. El Ministerio debe decidir cuál de los dos tipos de error hace menos probable. Si su interés es velar por la adecuada inversión de los recursos públicos, se 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 debemos hacer lo más "pequeña" posible la probabilidad de aceptar K, dado que la hipótesis H es verdadera.

6.2. Hipótesis compuestas

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 recuperan de un ataque cardíaco es mayor o igual a 77 %.

K: El porcentaje de fumadores que se recuperan de un ataque cardíaco es menor de 77 %.

En este caso, se tienen las hipótesis compuestas

$$H : \theta \in \Theta_1$$
$$K : \theta \in \Theta_2$$

donde $\Theta_1 = [0.77, 1), \Theta_2 = (0, 0.77), \Theta_1 \cap \Theta_2 = \phi, \Theta_1 \cup \Theta_2 = (0, 1)$ y $\Theta_2 = \Theta_1^c$.

Definición 6.1. Si Θ_1 contiene más de un punto, H se denomina una hipótesis compuesta. Si Θ_1 contiene solo un punto, H se denomina hipótesis simple. La misma convención es válida para K.

La decisión de aceptar o rechazar la hipótesis nula H se basa en los datos obtenidos a través de un muestreo. Intuitivamente, si de n fumadores, el número de personas que sobreviven a un ataque cardíaco es mayor o igual que algún entero k_0, $0 \leq k_0 \leq n$, se decide a favor de H, y si el número de fumadores que sobreviven es menor que k_0, se decide a favor de K. Es evidente que, en este caso, el estadístico de prueba es $T(X) = \Sigma_{i=1}^{n} X_i$. Si $T(x) > k_0$, no existen argumentos para rechazar H, y si $T(x) \leq k$, se rechaza H.

6.3. Probabilidad de error

Retomamos nuevamente el ejemplo de la proporción de fumadores, considerando hipótesis simples. Supongamos que $H : \theta = \theta_0$ es la hipótesis nula y $K : \theta = \theta_1$, $\theta_1 < \theta_0$, la hipótesis alternativa. Pueden presentarse dos tipos de error cuando se hace una prueba de hipótesis. Un error de tipo I se comete cuando se rechaza H siendo la hipótesis H verdadera; un error de tipo II, cuando se acepta H siendo K la hipótesis verdadera. La probabilidad de error de tipo I se denota por α y la probabilidad de error de tipo II, por β. Así,

$$\begin{aligned} \alpha &= P_{\theta_0}\{\text{rechazar } H\} \\ &= P_{\theta_0}[T(X) < k] = \sum_{i=0}^{k-1} \binom{n}{i} \theta_0^i (1-\theta_0)^{n-i} \end{aligned}$$

El error de tipo II depende de una alternativa particular $\theta_1 \in \Theta_1$ que debe ser considerada. Así, para cada θ_1,

$$\beta = P_{\theta_1}\{\text{aceptar } H\} = 1 - P_{\theta_1}\{\text{rechazar } H\}$$

En el ejemplo,

$$\beta = P_{\theta_1}(T(X) \geq k) = \sum_{i=k}^{n} \binom{n}{i} \theta_1^i (1-\theta_1)^{n-i} \tag{6.1}$$

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 de tipo I sea el más importante para la situación estudiada y sea fácil de controlar.

6.4. Región de rechazo

Para las pruebas de hipótesis, debemos fijar primero un α de manera que la probabilidad de un error de tipo I mayor que α sea inadmisible. En la práctica, $\alpha = 0.01$ y $\alpha = 0.05$ son comúnmente usados. Una vez fijado α, centramos nuestra atención en pruebas que tengan probabilidad de rechazo menor que o igual a α, y determinamos la región de rechazo.

6.4.1. Distribución binomial

Suponga que deseamos probar las hipótesis $H : \theta = \theta_0$ *versus* $K : \theta < \theta_0$, si dado α existe un número entero positivo k_0 tal que

$$\begin{aligned} \alpha &= P_{\theta_0}(\text{rechazar } H) \qquad (6.2) \\ &= \sum_{i=0}^{k_0-1} \binom{n}{i} \theta_0^i (1-\theta_0)^{n-i}, \end{aligned}$$

la región de rechazo es $RR = \{0, 1, 2, \ldots, k_0 - 1\}$.

Si no existe un número entero k_0 que cumpla 6.2, 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 este caso la región de rechazo queda determinada por el conjunto $RR = \{0, 1, 2, \ldots, n_0\}$. Observe que no se puede considerar el conjunto $RR = \{x : T(x) \leq n_0 + 1\}$ como región de rechazo, ya que $P[T(X) \leq n_0 + 1] > \alpha$, y el tamaño de la región de rechazo sería mayor que α. En este caso, no se podría afirmar que la probabilidad de cometer error de tipo

I es α, pues en realidad se estaría cometiendo un error de tipo I igual a $P[T(X) \leq n_0 + 1]$ 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 $min\{n\theta_0, n(1-\theta_0)\} \geq 5$, podemos determinar la región de rechazo mediante la aproximación por la distribución normal de la distribución binomial

$$\begin{aligned} P_{\theta_0}[T(X) \leq k] &\approx P\left[\frac{T(X) - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}} \leq \frac{k - n\theta_0 + \frac{1}{2}}{\sqrt{n\theta_0(1-\theta_0)}}\right] \\ &= \phi\left(\frac{k - n\theta_0 + \frac{1}{2}}{\sqrt{n\theta_0(1-\theta_0)}}\right) \end{aligned}$$

Ya que esperamos un error de tipo I menor que o igual a α, existe un número real z_α tal que

$$\frac{(k - n\theta_0) + \dfrac{1}{2}}{\sqrt{n\theta_0(1-\theta_0)}} \leq z_\alpha$$

En consecuencia, la región de rechazo está determinada por el mayor número entero positivo k_0 tal que $k_0 \leq n\theta_0 - \frac{1}{2} - z_\alpha\sqrt{n\theta_0(1-\theta_0)}$.

6.4.2. Distribución normal

Sea X_i, $i = 1, 2, 3, \ldots, n$, una muestra aleatoria de una distribución normal con media desconocida θ, y varianza conocida σ^2. Si deseamos probar $H : \theta = \theta_0$ *versus* $K : \theta > \theta_0$, es claro que el estadístico de prueba es $T(X) = \bar{X}$ y que dado un nivel de significancia α, la región de rechazo está determinada por un número real c tal que $\alpha = P_{\theta_0}[T(X) \geq c]$. En consecuencia, $\alpha = P_{\theta_0}\left(Z \geq \frac{c-\theta_0}{\sigma/\sqrt{n}}\right)$ y $\frac{c-\theta_0}{\sigma/\sqrt{n}} = z_\alpha$. Así, $c = \theta_0 + z_{1-\alpha}\sigma/\sqrt{n}$ y, por tanto, la región de rechazo está dada por el conjunto $RR = \{x : T(x) \geq \theta_0 + z_\alpha\sigma/\sqrt{n}\}$.

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 $T(X) = \sqrt{n}(\bar{X} - \theta_0)/\sigma$, el cual bajo H tiene distribución normal estándar. En este caso la región de rechazo se establece para valores grandes de $|T|$. Más exactamente, $RR = \{x : |T(x)| \geq z_{\alpha/2}\}$.

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, se puede utilizar la distribución normal, reemplazando σ por su valor estimado en la muestra. Si $n < 30$, debemos utilizar la distribución t, como se indica en la sección 6.4.3.

Ejemplo 6.1. 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. Si el gerente afirma que el contenido medio de los sobres de azúcar es 5 gramos, ¿qué se puede afirmar con un nivel de significancia del 5 % con respecto a la posibilidad de que la media sea mayor? Para solucionar este interrogante, debemos plantear como hipótesis nula $H : \theta = 5$ y como hipótesis alternativa $K : \theta > 5$. Dado que $5.2 > \theta_0 + z_\alpha s/\sqrt{n} = 5 + 1.645\left(\frac{0.4}{10}\right) = 5.066$, $\bar{x} = 5.2$ pertenece a la región de rechazo. Esto significa que con un nivel de significancia $\alpha = 0.05$ se rechaza H a favor de la hipótesis K: 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 $RR = \{x \in R : \bar{x} > 5.066\}$. 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 conocemos el error de tipo II, que sólo se puede calcular para valores específicos de la hipótesis alternativa. Por ejemplo, si $K : \theta_1 = 5.1$,

$$\begin{aligned}\beta &= P_{\theta_1}(\bar{X} < 5.066)\\ &= P_{\theta_1}\left(Z < \frac{5.066 - 5.1}{0.4}\right)\\ &= P_{\theta_1}(Z < -0,085) = 0.47\end{aligned}$$

y si $K : \theta_1 = 5.5$,

$$\begin{aligned}\beta &= P_{\theta_1}(\bar{X} < 5.066)\\ &= P_{\theta_1}\left(Z < \frac{5.066 - 5.5}{0.4}\right)\\ &= P_{\theta_1}(Z < -1.085) = 0.14\end{aligned}$$

Suponga que el contenido medio de los cien sobres de azúcar fuera de 4.8 gramos y la desviación estándar 0.4 gramos, pero que se desea probar las hipótesis $H : \theta = 5$ *versus* $K : \theta \neq 5$. Considerando como estadístico de prueba $T(X) = \frac{\sqrt{n}(\bar{X}-\theta)}{\sigma}$ y $\alpha = 0.05$, se halla que $T(x) = -5 \in$

$[-1.96, 1.96]^c$ y, por tanto, se rechaza H. Se concluye así que existe evidencia estadística para afirmar que el contenido medio de azúcar en los empaques es diferente de 5 gramos.

6.4.3. Distribución t

Suponga que X_i, $i = 1, 2, \ldots, n$, provienen 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 $T(X) = \bar{X}$, y dado un nivel de significancia α, la región de rechazo está determinada por un número real c tal que $\alpha = P_{\theta_0}[T(X) \geq c]$. Así, $\alpha = P_{\theta_0}\left(T_{n-1} \geq \frac{c-\theta_0}{s/\sqrt{n}}\right)$ donde T_{n-1} tiene distribución t con $n-1$ grados de libertad. De lo anterior, $\frac{c-\theta_0}{s/\sqrt{n}} = t_{n-1,\alpha}$ y, en consecuencia, la región de rechazo está determinada por el conjunto $RR = \{x : |T(x)| \geq c\}$, donde $c = \theta_0 + t_{n-1,\,\alpha} s/\sqrt{n}$.

En una prueba de hipótesis $H : \theta = \theta_0$ *versus* $K : \theta \neq \theta_0$, se considera $T(X) = \sqrt{n}(\bar{X} - \theta_0)/S$ como estadístico de prueba, el cual bajo H tiene distribución t con $n - 1$ grados de libertad. Observe que la región de rechazo se establece para valores grandes de $|T|$. En términos matemáticos, $RR = \{x : |T(x)| \geq t_{n-1,\,\alpha/2}\}$.

En general, en las pruebas de hipótesis de la forma $H : \theta = \theta_0$ *versus* $K : \theta \neq \theta_0$, se puede considerar $T_1(X) = |\bar{X} - \theta_0|$ como estadístico de prueba. Si la varianza es conocida, para un nivel de significancia α, la hipótesis H se rechaza para valores de $T_1(X)$ mayores que c, donde $c = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. Esto es equivalente a rechazar H para valores mayores que $\bar{X}$ pertenecientes a $[\underline{L}, \bar{L}]^c$, donde

$$\underline{L} = \theta_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \quad \text{y} \quad \bar{L} = \theta_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Si la varianza es desconocida, pero el tamaño de la muestra es mayor que 30, σ puede ser reemplazado por su estimación s. Si la varianza es desconocida y el tamaño de la muestra es menor que 30, para esta prueba de hipótesis se puede utilizar el estadístico $T(X)$, 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 6.2. Suponga 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 4.8 gramos y su desviación estándar 0.4 gramos, considerando un nivel de significancia $\alpha = 0.05$, 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.

6.4.4. Diferencia de medias

Sea $X_i \sim N(\theta_1, \sigma_1^2)$, $i = 1, 2, \ldots, n_1$, y $Y_i \sim N(\theta_2, \sigma_2^2)$, $i = 1, 2, \ldots, n_2$, dos muestras aleatorias independientes de dos poblaciones normales con medias desconocidas y varianzas conocidas. Si desea probar $H : \theta_2 - \theta_1 = \theta_0$ *versus* $K : \theta_2 - \theta_1 > \theta_0$, es conveniente considerar como estadístico de prueba

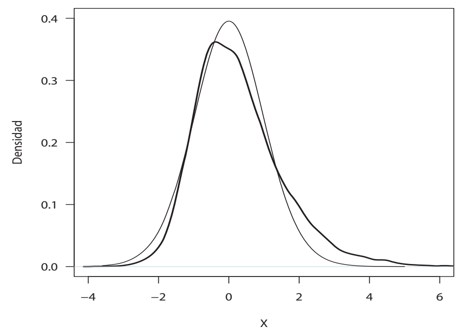
$$T(X, Y) = \frac{\bar{Y} - \bar{X} - \theta_0}{\sqrt{\frac{\sigma_2^2}{n_2} + \frac{\sigma_1^2}{n_1}}} \tag{6.3}$$

donde $X = (X_1, \ldots, X_{n_1})$ y $Y = (Y_1, \ldots, X_{n_2})$. Dado que bajo H, $T(X, Y)$ tiene distribución normal estándar, para un nivel de significancia α, la región de rechazo está dada por $RR = \{(x, y) : T(x, y) \geq z_\alpha\}$. Esta región es equivalente a $RR = \left\{(x, y) : T_1(x, y) \geq \theta_0 + z_\alpha\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}\right\}$, donde $T_1(X, Y) = \bar{Y} - \bar{X}$.

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 $RR = \{(x, y) : |T(x, y)| \geq z_{\alpha/2}\}$.

Si las varianzas σ_1^2 y σ_2^2 son desconocidas, el estadístico de prueba está determinado por (1.16), y la región de rechazo se puede determinar fácilmente usando la distribución t. Si $n_1 > 30$ y $n_2 > 30$, se puede utilizar el estadístico (6.3), reemplazando las varianzas por sus estimaciones.

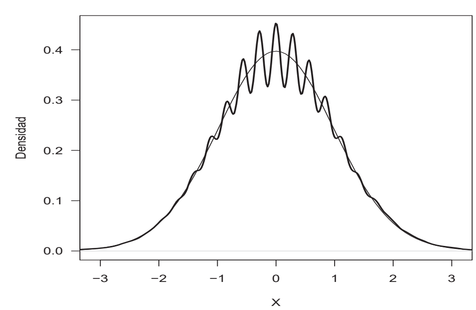
La figura 6.1 muestra la distribución de una muestra de tamaño 10.000 del estadístico T, cuando T es calculado a partir de muestras aleatorias de tamaño 15 obtenidas de una distribución $N(10, 4)$ y de una distribución *Exp* (1/10), com media 10. En esta figura, la línea más gruesa corresponde a la distribución de T y la más delgada a la distribución t con 28 grados de libertad. Observese que la distribución real de T es sesgada a derecha, lo cual tiene incidencia directa sobre los p-valores en las pruebas de hipótesisis.



Gráfica 6.1. Distribución de la estadística T.

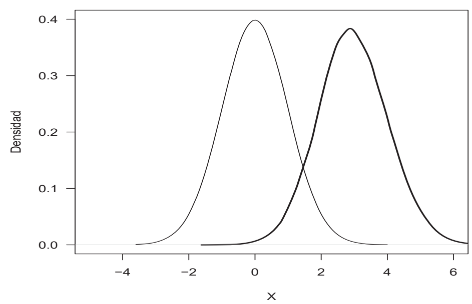
Diferencia de proporciones

La figura 6.2 muestra la distribución de una muestra de tamaño 10.000 del estadístico T, cuando T es calculado a partir de muestras aleatorias de tamaño 15 obtenidas de una distribuciónes $Ber(0.3)$. 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.



Gráfica 6.2. Distribución de la estadística T.

La figura 6.3 muestra la distribución de una muestra de tamaño 10.000 del estadístico T, cuando T es calculado a partir de muestras aleatorias de tamaño 15 obtenidas de una distribuciónes $Ber(0.3)$ y $Ber(0.5)$. En esta figura, la línea más gruesa corresponde a la distribución de T y la más delgada a la distribución t con 198 grados de libertad.



Gráfica 6.3. Distribución de la estadística T.

6.4.5. Pruebas sobre las varianzas

Suponga que Y_i, $i = 1, \ldots, n$, es una muestra aleatoria de una distribución normal con media θ y varianza σ^2, desconocidas. Para una prueba de hipótesis de nivel α de $H : \sigma^2 = \sigma_0^2$ *versus* $K : \sigma^2 > \sigma_0^2$, es conveniente considerar como estadístico de prueba a $T(X) = (n-1)\, S^2/\sigma_0^2$, dado que bajo H este estadístico tiene distribución χ^2 con $n-1$ grados de libertad. En consecuencia, la región de rechazo está determinada por $RR = \{x \in R^n : T(x) > x^2_{n-1,\alpha}\}$.

Si la hipótesis alternativa es $K : \sigma^2 \neq \sigma_0^2$, podemos utilizar el mismo estadístico de prueba, y la región de rechazo está determinada por el conjunto

$$RR = \{x \in R^n : T(x) \leq x^2_{n-1,\, \alpha/2}\} \cup \{x \in R^n : T(x) \geq x^2_{n-1,1-\alpha/2}\}$$

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 hipótesis dadas por $H : \sigma_1^2 = \sigma_2^2$ y 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 consideramos como estadístico de prueba el cociente $T(X, Y) = S_1^2/S_2^2$, el cual bajo 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, rechazamos H para valores de $T(X, Y) \geq f_{\alpha,n_1-1,n_2-1}$, siendo f_{α,n_1-1,n_2-1} un número real tal que $\alpha = P(F^{n_1-1}_{n_2-1} \geq f_{\alpha,n_1-1,n_2-1})$. Si la hipótesis alternativa es $K : \sigma_1^2 \neq \sigma_2^2$, con un nivel de significancia α, rechazamos H para valores de $T(X, Y) \geq f_{\frac{\alpha}{2},n_1-1,n_2-1}$ y para valores de $T(X, Y) \leq$

$f_{1-\frac{\alpha}{2},n_1-1,n_2-1}$, 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 6.3. Un distribuidor de tornillos desea establecer, con una confiabilidad de 0.95, si la variabilidad del diámetro de los los tornillos producidos por dos empresas A y B es diferente. 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 $T(X,Y) = S_A^2/S_B^2 \sim F_9^{12}$, la región de rechazo está determinada por el conjunto $RR = (-\infty, \frac{1}{3.87}) \cup (3.44, \infty)$. Así, $T(x,y) = \frac{1.05}{0.85} = 1.235 \notin RR$, lo cual indica que no existe evidencia estadística para afirmar que las variabilidad del diámetro de los tornillos es diferente.

6.5. Potencia de una prueba

La potencia de la prueba de una hipótesis $H : \theta = \theta_0$ *versus* una alternativa $K : \theta = \theta_1$ es la probabilidad de rechazar H cuando K es verdadera. Si δ denota la función indicadora asociada a la región de rechazo, la potencia está definida por

$$\begin{aligned}\beta(\theta_1, \delta) &= P_{\theta_1}[\text{rechazar } H] \\ &= P_{\theta_1}[\delta = 1]\end{aligned}$$

Ejemplo 6.4. Sea X_i, $i = 1, 2, \ldots, n$, una muestra aleatoria 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 H para valores grandes de $\bar{X}$. Sin embargo, dado que una función creciente de un estadístico suficiente es un estadístico suficiente debido a que generan la misma familia de regiones de rechazo, es conveniente considerar el estadístico $T(X) = \frac{\sqrt{n}\bar{X}}{\sigma}$, que bajo H tiene distribución normal estándar, y 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:

$$\begin{aligned}\beta(\theta_1, \delta) &= P_{\theta_1}\left(\frac{\sqrt{n}\,\bar{X}}{\sigma} > z_\alpha\right) \\ &= P_{\theta_1}\left(\frac{\sqrt{n}\,(\bar{X} - \theta_1)}{\sigma} > z_\alpha - \frac{\sqrt{n}\,\theta_1}{\sigma}\right)\end{aligned}$$

$$= 1 - P_{\theta_1}\left(Z < z_\alpha - \frac{\sqrt{n}\,\theta_1}{\sigma}\right)$$
$$= 1 - \phi\left(z_{(\alpha)} - \frac{\sqrt{n}\,\theta_1}{\sigma}\right) = \phi\left(\frac{\sqrt{n}\,\theta_1}{\sigma} - z_\alpha\right)$$

Si la hipótesis alternativa es $K : \theta > 0$, la potencia es una función de θ, definida en $[0, \infty)$, por la ecuación

$$\beta(\theta, \delta) = \phi\Big(\frac{\sqrt{n}\theta}{\sigma} - z_\alpha\Big) \tag{6.4}$$

$\beta(\theta, \delta)$ es una función continua y creciente de $\sqrt{n}\ \theta/\sigma$, que tiende a α cuando $\sqrt{n}\theta/\sigma$ tiende a cero, y a 1 cuando $\sqrt{n}\theta/\sigma$ tiende a más infinito. Por (6.4), existe z_β tal que $z_\beta = \frac{\sqrt{n}\theta}{\sigma} - z_\alpha$. Así, si deseamos una potencia β mayor que algún valor β_0, el tamaño n de la muestra debe ser tal que $\frac{\sqrt{n}\theta}{\sigma} - z_\alpha > z_{\beta_0}$.

Si deseamos que esta desigualdad se cumpla para todo θ mayor que algún número real positivo Δ, entonces n es el menor número entero que satisface $\dfrac{\sqrt{n}\Delta}{\sigma} - z_\alpha > z_{\beta_0}$ es decir $n > \frac{\sigma^2}{\Delta^2}(z_{\beta_0} + z_\alpha)^2$.

6.5.1. 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 la expresión:

$$\begin{aligned}\beta(\theta, \delta) &= P_\theta[\text{rechazar } H] \\ &= P_\theta[\delta = 1]\end{aligned}$$

Si $\theta \in \Theta_0$, $\beta(\theta, \delta)$ es la probabilidad de error de tipo I. Si $\theta \in \Theta_1$, $\beta(\theta, \delta)$ es la probabilidad de aceptar K, dado que K es verdadero. En este caso, $\beta(\theta, \delta)$ es 1 menos la probabilidad de error de tipo II. Así,

$$\beta(\theta, \delta) = \begin{cases} \alpha(\theta) & \text{si } \theta \in \Theta_0 \\ 1 - \beta(\theta) & \text{si } \theta \in \Theta_1 \end{cases}$$

En el ejemplo de la prueba de hipótesis de la sección 6.4.1, la función de potencia

$$\beta(\theta, \delta) = \sum_{j=0}^{k_0 - 1} \binom{n}{j}\, \theta^j (1 - \theta)^{n-j}$$

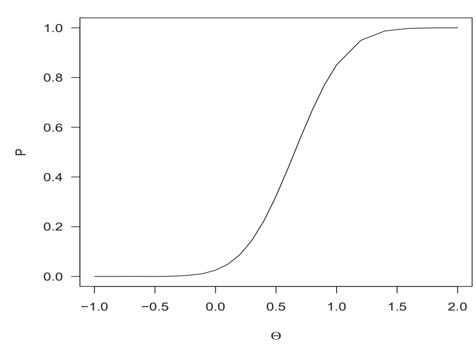
es una función decreciente, definida en $(0, 1)$.

Ejemplo 6.5.

Suponga que observa una muestra aleatoria $X_i \sim N(\theta, \sigma^2)$, $i = 1, \ldots, n$, con θ desconocido y σ^2 conocido. Deseamos probar la hipótesis $H : \theta = 0$ contra la alternativa $K : \theta = \theta_1$, $\theta_1 > 0$. Si la prueba se realiza con un nivel de significancia α, considerando como estadístico de prueba $T(X) = \sqrt{n}\bar{\mathrm{X}}/\sigma$,

$$\begin{aligned}\beta(\theta_1, \delta) &= P_{\theta_1}\left(\frac{\sqrt{n}\bar{X}}{\sigma} > z_\alpha\right)\\ &= 1 - \phi\left(z_\alpha - \frac{\sqrt{n}\theta_1}{\sigma}\right)\end{aligned}$$

Así, la probabilidad de error de tipo II es $\phi\Big(z_\alpha - \frac{\sqrt{n}\theta_1}{\sigma}\Big)$. Si deseamos probabilidades de error de tipo I y II iguales a α, $z_\alpha - \frac{\sqrt{n}\theta_1}{\sigma} = -z_\alpha$. Los errores de tipo I y II son menores que α si $n > \frac{4\sigma^2}{\theta_1^2} z_\alpha^2$.



Gráfica 6.4. Función de Potencia.

La figura 6.4 muestra la función de potencia $\beta(\theta_1, \delta)$ de la prueba de hipotesis propuesta en el ejemplo 6.5, para $\alpha = 0.05$, $n = 9$ y $\sigma^2 = 9$.

Ejemplo 6.6. Suponga que X_i $i = 1, 2, \ldots, n$, denota el tiempo de vida de n bombillas y que, para todo i, X_i, tiene distribución exponencial, con función de densidad dada por $f_\lambda(x) = \lambda e^{-\lambda x} I_{(0,\infty)}(x)$. Dado que $2\lambda \sum_{i=1}^{n} X_i \sim \chi^2_{2n}$, la región de rechazo de nivel α para la prueba de hipótesis $H : 1/\lambda = \theta \leq \theta_0$ *versus* $K : 1/\lambda = \theta > \theta_0$, relacionada con el tiempo de vida medio, $1/\lambda$, de las bombillas, está determinada por

$$\alpha = P_{\frac{1}{\theta_0}}\left[2\lambda_0 \sum_{i=1}^{n} X_i \geq \chi^2_{\alpha,2n}\right] = P_{\frac{1}{\theta_0}}\left[\bar{X} \geq \frac{\theta_0}{2n}\chi^2_{\alpha,2n}\right]$$

Así, la función de potencia queda determinada por

$$\begin{aligned}
\beta(\theta,\delta) &= P_{\frac{1}{\theta}}\left[\bar{X} \geq \frac{\theta_0}{2n}\,\chi^2_{\alpha,2n}\right] \\
&= P_{\frac{1}{\theta}}\left[\frac{2n\bar{X}}{\theta} \geq \frac{\theta_0}{\theta}\,\chi^2_{\alpha,2n}\right] \\
&= P_{\frac{1}{\theta}}\left[\chi^2_{2n} \geq \frac{\theta_0}{\theta}\,\chi^2_{\alpha,2n}\right]
\end{aligned}$$

Ejemplo 6.7. Suponga que $X_1, \ldots, X_n$ son variables independientes uniformemente distribuidas en un intervalo $(0,\theta)$. Como la función de distribución del estadístico $M_n = Max\{X_1, \ldots, X_n\}$ está dada por

$$F(y) = \begin{cases} 0 & \text{si} \quad y < 0 \\ P_\theta[X_1 \leq y]^n & \text{si} \quad y \in [0,\theta] \\ 1 & \text{si} \quad y \geq \theta \end{cases}$$

su función de densidad es $f(y) = n[F_1(y)]^{n-1}F_1'(y)$, donde

$$F_1(y) = \begin{cases} 0 & \text{si} \quad y < 0 \\ \frac{y}{\theta} & \text{si} \quad y \in [0,\theta] \\ 1 & \text{si} \quad y \geq \theta \end{cases}$$

Así, $f(y) = \frac{ny^{n-1}}{\theta^n}$ en el intervalo $(0,\theta)$ y $f(y) = 0$ si $y < 0$ o $y > \theta$.

Dado un nivel de significancia α, la región de rechazo para la prueba de hipótesis de $H : \theta \leq \theta_0$ *versus* $K : \theta > \theta_0$ está determinada por

$$\alpha = P_{\theta_0}\{M_n \geq c\} = \int_c^{\theta_0} \frac{ny^{n-1}}{\theta_0^n} dy = 1 - \left(\frac{c}{\theta_0}\right)^n$$

con $c \in (0,\theta_0)$. 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 $\beta(\theta,\delta) = 1 - \left(\frac{\theta_0}{\theta}\right)^n(1-\alpha)$, $\theta > \theta_0$. Finalmente, a partir de la anterior expresión se halla que $\beta(\theta,\delta) = \alpha$.

6.5.2. Valor p

En la presentación de informes, un analista de datos podría afirmar que rechaza la hipótesis H con un nivel de significancia $\alpha = 0.05$. Sin embargo, si los lectores de este informe están interesados en saber qué se puede concluir para otros niveles de significancia como $\alpha = 0.01$ o $\alpha = 0.001$, no les sería posible. En realidad, los lectores del informe estarían 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 significancia real de la prueba. Esto es posible mediante el valor p, que en el caso de pruebas de hipótesis de la forma $\theta = \theta_0$ *versus* $K : \theta > \theta_0$ se define como sigue:

- Si $H : \theta = \theta_0$ y $K : \theta > \theta_0$, $p_v = P_{\theta_0}[T(X) \geq T(x_0)]$, donde $T(X)$ es el estadístico de prueba y x_0 los valores observados en la muestra.
- Si $H : \theta = \theta_0$ y $K : \theta < \theta_0$, $p_v = P_{\theta_0}[T(X) \leq T(x_0)]$, donde $T(X)$ es el estadístico de prueba y x_0 los valores observados en la muestra.

En el caso de pruebas de hipótesis de la forma $\theta = \theta_0$ *versus* $K : \theta \neq \theta_0$, si bajo H_0 el estadístico de prueba tiene distribución simétrica, normal o t, $p_v = 2P_{\theta_0}[T(X) \geq |T(x_0)|\ ]$, donde $T(X)$ es el estadístico de prueba y x_0 los valores observados en la muestra.

6.6. Intervalos de confianza y pruebas de hipótesis

Consideremos el caso de prueba de hipótesis para la media de una distribución normal con varianza conocida. Para todo θ_0 en el espacio paramétrico Θ, podemos determinar un conjunto de valores del estadístico $\bar{X}$ para los cuales no se rechaza la hipótesis $H : \theta = \theta_0$, en un nivel α. Dado que en un nivel α la región de rechazo está determinada por $RR(\bar{X}) = \left[\bar{x} : \frac{\sqrt{n}|\bar{x}-\theta_0|}{\sigma} \geq z_{\alpha/2}\right]$, el conjunto de los $\bar{x}$ para los cuales no se rechaza H es

$$A(\bar{X}) = \left[\bar{x} : |\bar{x} - \theta_0| \leq \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right]$$

Ahora podemos preguntarnos, dado $\bar{x}$, cuál es el conjunto de parámetros $C(\bar{x})$ para los cuales a un nivel de significancia α se acepta la hipótesis

$H : \theta = \theta_0$, $\theta_0 \in C(\bar{x})$. A partir de la definición de la región de rechazo, especificada anteriormente en esta sección, no se rechaza la hipótesis $H : \theta = \theta_0$ si y solo si $\bar{x} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \leq \theta \leq \bar{x} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}$. Así, el conjunto $C(\bar{x}) = \{\theta : |\theta - \bar{x}| \leq \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\}$.

Para finalizar esta sección, el lector puede representar gráficamente la relación definida por $C(\bar{x})$ y $A(\bar{X})$ para $\Theta = R$. Note que estos intervalos pueden considerarse subconjuntos aleatorios del espacio de parámetros que tienen probabilidad $1 - \alpha$ de contener el verdadero parámetro.

6.7. Hipótesis compuestas

En esta sección se dan tres ejemplos del planteamiento de hipótesis compuestas.

Ejemplo 6.8. Suponga 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 o más en cada lote es insatisfactorio y considera las hipótesis (6.5) y (6.6):

$$H : b \geq b_0 \textit{ versus } K : b < b_0 \tag{6.5}$$
$$H : b \leq b_0 \textit{ versus } K : b > b_0 \tag{6.6}$$

En 6.5, si el número de artículos defectuosos en la muestra es pequeño, se rechaza H y se acepta el lote. En 6.6, si el número de artículos defectuosos en la muestra de un lote es pequeño, no se rechaza H y no se acepta el lote de artículos.

Es natural que el comprador se decida por la hipótesis 6.5, puesto que $\alpha = p_{\theta_0}[\text{rechazo}]$ es la probabilidad de aceptar un lote que tenga más de b_0 artículos defectuosos. En 6.5, $\alpha = p_{\theta_0}[\text{rechazo } H]$ es la probabilidad de no aceptar un lote dado que tiene menos de b_0 artículos defectuosos.

Ejemplo 6.9. Suponga que deseamos determinar si la vida media de las bombillas de una producción es mayor que $1/\lambda_0$. Si X_i, $i = 1, \ldots, n$, es una muestra aleatoria de tiempos de vida y asumiendo que los tiempos tienen una distribución $E(\lambda)$, un comprador de bombillas quiere que sea pequeña la probabilidad α de aceptar que estas tienen una duración mayor que $1/\lambda_0$, cuando en realidad tienen una duración menor. En consecuencia, podemos plantear las siguientes hipótesis:

$$H : \frac{1}{\lambda} \leq \frac{1}{\lambda_0} \quad \textit{versus} \quad K : \frac{1}{\lambda} > \frac{1}{\lambda_0}$$

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 6.10.

El interés de un comprador está en que la probabilidad de recibir tornillos que tengan alta variabilidad sea lo más pequeña posible. En consecuencia, las hipótesis a plantear son:

$$H : \sigma^2 \geq \sigma_0^2$$
$$K : \sigma^2 < \sigma_0^2$$

Así, se podrá controlar la probabilidad $\alpha = P_H(\text{Rechazo})$ de aceptar que la producción de tornillos tiene una varianza menor que σ_0^2, cuando en realidad tiene una probabilidad mayor.

Ejercicios 6.1.

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. Se desea probar la hipótesis $H : \theta = \theta_0$ frente a la alternativa $K : \theta > \theta_0$. Si el valor observado del estadístico $T(X) = \frac{\sqrt{n}(\bar{X}-\theta_0)}{\sigma}$ de prueba es 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 hallaron 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 que valores de α se rechaza $H : \theta_1 = \theta_2$ si la hipótesis alternativa es $K : \theta_1 \neq \theta_2$, y para qué valores de α, si la hipótesis alternativa es $K : \theta_1 > \theta_2$?

4. Una empresa de maquinaria produce un 10 % de artículos defectuosos. Para establecer un proceso de control de calidad, la gerencia ha establecido que si en 350 artículos, 45 resultan defectuosos, la máquina está produciendo una proporción de artículos mayor que el 10 % y que, por tanto, debe ser reparada. ¿En qué nivel de significancia se ha tomado esta decisión?

5. Cuatro estudiantes están interesados en demostrar que el promedio de las notas de su facultad es 4.0. Cada uno de ellos obtiene una muestra

aleatoria y determina un valor p, 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. Suponga que se desea probar la hipótesis $\sigma_1^2 = \sigma_2^2$ frente a la alternativa $\sigma_1^2 < \sigma_2^2$ y que el valor observado de $F = \frac{s_1^2}{s_2^2}$ 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 rechazaría la hipótesis?

7. En un experimento para determinar diferencias en el pH del agua de dos ríos se observaron 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 confiabilidad del 99 %.

b) Probar la hipótesis: $\sigma_a^2 = \sigma_b^2$ frente a la alternativa $\sigma_a^2 \neq \sigma_b^2$. Concluya para $\alpha = 0.005$ y para $\alpha = 0.01$

8. 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 300 para detener la producción?

9. Sea $X_i, i = 1, \ldots, n$, una muestra aleatoria 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 α, halle la función de potencia y represéntela gráficamente.

Capítulo 7

Razón de verosimilitudes

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 en detalle este proceso de pruebas de hipótesis cuando se asume que Θ_0 es de dimensión menor que Θ.

7.1. Lema de Neyman-Pearson

Suponga que se desea probar $H : \theta = \theta_0$ *versus* $K : \theta = \theta_1$, $\theta_1 > \theta_0$, basados en una muestra aleatoria $Y_i, i = 1, 2, \ldots, n$, de una distribución con parámetro θ. Para un nivel de significancia α, la prueba de máxima potencia en $\theta = \theta_1$ tiene región de rechazo determinada por el cociente $L(x, \theta_0, \theta_1) = P(x, \theta_1)/P(x, \theta_0)$, donde $P(x, \theta_i)$ es la función de densidad de X dado $\theta = \theta_i$, $i = 0, 1$. Esta prueba tiene una región de rechazo de la forma $\{x : T(x) \geq c\}$ para algún estadístico de prueba T determinado por

$$\frac{P(x, \theta_1)}{P(x, \theta_0)} \geq k, \ k \text{ constante} \tag{7.1}$$

Dado que el error de tipo II es 1 menos la potencia, esta es la prueba de nivel α con menor probabilidad de error de tipo II.

Suponga que se desea probar $H : \theta = \theta_0$, frente a $K : \theta > \theta_0$. Una prueba δ^* de nivel 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 α para $H : \theta = \theta_0$ *versus* $K : \theta = \theta_a$.

Definición 7.1. Una prueba δ^* de nivel α es uniformemente más potente para $H : \theta \in \Theta_0$ *versus* $K : \theta \in \Theta_1$ si y sólo si $\beta(\theta, \delta^*) \geq \beta(\theta, \delta)$ para todo $\theta \in \Theta_1$ y para toda prueba δ de nivel α.

Ejemplo 7.1. Suponga una muestra aleatoria $X_i \sim N(\theta, \sigma^2), i = 1, 2, \ldots, n$, con media desconocida y varianza conocida. La prueba de nivel 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.

$$\begin{aligned}
L(x, \theta_0, \theta_1) &= \frac{P(x, \theta_1)}{P(x, \theta_0)} \\
&= \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i - \theta_1)^2 - \sum_{i=1}^{n}(x_i - \theta_0)^2\right)\right] \\
&= \exp\left[-\frac{1}{2\sigma^2}\left(2(\theta_0 - \theta_1)\sum_{i=1}^{n} x_i - n(\theta_0^2 - \theta_1^2)\right)\right] \qquad (7.2)
\end{aligned}$$

donde $P(x, \theta_i)$ es la función de densidad con $\theta = \theta_i$, $i = 0, 1$.

Se rechaza H si la muestra $x = (x_1, \ldots, x_n)$ es más probable bajo la hipótesis K que bajo la hipótesis H, es decir, si $L(x, \theta_0, \theta_1) \geq k$ para algún $k > 1$. Despejando $\bar{x}$ en $L(x, \theta_0, \theta_1) \geq k$, se encuentra que la región de rechazo está determinada por el conjunto de los $x \in R^n$ tales que

$$\bar{x} \geq \frac{-2\sigma^2 \ln k - n\theta_1^2 + n\theta_0^2}{2n(\theta_0 - \theta_1)} = k'$$

En consecuencia, L y $T(X) = \bar{X}$ son estadísticos equivalentes en el sentido que $T(x) \geq k'$ si y solo si $L(x, \theta_0, \theta_1) \geq k$. Este resultado se deriva directamente de (7.2), dado que $L(x, \theta_0, \theta_1)$ es una función creciente de $T(x) = \bar{x}$. Así, la prueba uniformemente más potente está determinada por el conjunto $RR = \{x \in R^n : \bar{x} \geq k'\}$. En consecuencia, la prueba uniformemente más potente de nivel α tiene región de rechazo

$$RR = \left\{x \in R^n : \frac{\sqrt{n}(\bar{x} - \theta_0)}{\sigma} \geq z_\alpha\right\}, \quad \text{donde} \quad z_\alpha = \frac{\sqrt{n}(k' - \theta_0)}{\sigma}$$

Esto indica que se rechaza H siempre que $\bar{x} > \theta_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$.

Finalmente, dado que los resultados anteriores son válidos para todo θ_1 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, ∞).

Suponga ahora que deseamos encontrar la prueba uniformemente más potente de nivel α para $H : \theta = \theta_0$ *versus* $K : \theta < \theta_0$. A partir de (7.2) se sigue que $L(x, \theta_0, \theta_1)$ es una función decreciente de $T(x) = \bar{x}$. En consecuencia, un número real k' tal que $T(x) \leq k'$ si y sólo si $L(x, \theta_0, \theta_1) \geq k$. Así, la prueba uniformemente más potente de nivel α está determinada por la región de rechazo $RR = \{x \in R^n : \frac{\sqrt{n}(\bar{x}-\theta_0)}{\sigma} \leq z_\alpha\}$.

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 verosimilitudes 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 7.1. Si $\delta_k(x) = 1$, $\delta_k(x) - \delta(x) \geq 0$. En consecuencia, dado k finito y $P(x, \theta_1)/P(x, \theta_0) \geq k$,

$$P(x, \theta_1)(\delta_k(x) - \delta(x)) \geq kP(x, \theta_0)(\delta_k(x) - \delta(x)) \tag{7.5}$$

Además, si $\delta_k(x) = 0$, $P(x, \theta_1)/P(x, \theta_0) < k$, se obtiene nuevamente la desigualdad (7.5). Integrando a ambos lados de (7.5) con respecto a x, se obtiene

$$E_{\theta_1}(\delta_k(X)) - E_{\theta_1}(\delta(X)) \geq k[E_{\theta_0}(\delta_k(X)) - E_{\theta_0}(\delta(X))]$$

Finalmente, teniendo en cuenta que $\beta(\theta, \delta) = E_\theta(\delta_k(X))$, a partir de (7.3) se sigue (7.4), ya que $E_{\theta_0}(\delta_k(X)) - E_{\theta_0}(\delta(X)) \geq 0$.

Ejemplo 7.2. Sea $X_i \sim N(\theta, \sigma^2)$, θ conocida y σ^2 desconocido, $i = 1, 2, \ldots, n$. Deseamos probar

$$H : \sigma^2 = \sigma_0^2 \quad \textit{versus} \quad K : \sigma^2 = \sigma_1^2, \;\; \sigma_1^2 > \sigma_0^2$$

donde σ_0^2 representa la máxima varianza permitida en un proceso de producción. Dado que $L(x, \sigma_0^2, \sigma_1^2) = \frac{P(x,\sigma_1^2)}{P(x,\sigma_0^2)}$ es una función creciente de $\sum_{i=1}^n (x_i$

$-\ \theta)^2$, a partir de $L(x, \sigma_0^2, \sigma_1^2) \geq k$ se encuentra que la región de rechazo está determinada por la desigualdad $\sum_{i=1}^n (x_i - \theta)^2 \geq k'$. Así, bajo la hipótesis nula H, $T(x) = \frac{1}{\sigma_0^2} \sum_{i=1}^n (x_i - \theta)^2 \sim \chi_n^2$ y la región de rechazo de la prueba uniformemente más potente de nivel α está determinada por el conjunto $RR = \{x \in R^n : \frac{1}{\sigma_0^2} \sum_{i=1}^n (x_i - \theta)^2 \geq \chi^2_{n(\alpha)}\}$. El error de tipo II es

$$\begin{aligned}\beta &= P_{\sigma_1^2}\left(\frac{1}{\sigma_0^2}\sum_{i=1}^n (x_i - \theta)^2 < \chi^2_{n(\alpha)}\right) \\ &= P_{\sigma_1^2}\left(\frac{1}{\sigma_1^2}\sum_{i=1}^n (x_i - \theta)^2 < \frac{\sigma_0^2}{\sigma_1^2}\chi^2_{n(\alpha)}\right)\end{aligned}$$

Suponga ahora que deseamos encontrar la prueba uniformemente más potente de nivel α para $H : \sigma^2 = \sigma_0^2$ *versus* $K : \sigma^2 = \sigma_1^2$, $\sigma_1^2 < \sigma_0^2$. En este caso, $L(x, \sigma_0^2, \sigma_1^2)$ es una función decreciente de $\sum_{i=1}^n (x_i - \theta)^2$. La región de rechazo de la prueba uniformemente más potente de nivel α está determinada por el conjunto $RR = \{x \in R^n : \frac{1}{\sigma_0^2} \sum_{i=1}^n (x_i - \theta)^2 \leq \chi^2_{n(1-\alpha)}\}$.

Ejemplo 7.3. Suponga que $X \sim Geo(\theta)$. Para determinar la prueba uniformemente más potente de nivel α para probar $H : \theta = \theta_0$ *versus* $K : \theta = \theta_1$, $\theta_0 < \theta_1$, determinamos el conjunto de los x tales que

$$L(x, \theta_0, \theta_1) = \left(\frac{\theta_1}{\theta_0}\right)\left(\frac{1-\theta_1}{1-\theta_0}\right)^{x-1} \geq k \tag{7.6}$$

Dado que $\theta_1 > \theta_0$, L es una función decreciente de x y la región de rechazo está determinada por el conjunto $RR = \{x \in Z^+ : x \leq k'\}$. Si k' es entero, esta es la región de rechazo de la prueba uniformemente más potente de nivel $\alpha = \sum_{j=1}^{k'} P_{\theta_0}(j)$ para la prueba de hipótesis de H *versus* K. Note 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 $0 \leq k \leq \infty$ y sea δ_k definida en el teorema 7.1. Suponga que δ es la función crítica de una prueba de hipótesis definida arbitrariamente en el conjunto $A = \{x : L(x, \theta_0, \theta_1) = k\}$ y tal que $\delta(x) = \delta_k(x)$ en el conjunto $B = \{x : L(x, \theta_0, \theta_1) \neq k\}$. Entonces δ es la prueba más potente para probar $H : \theta = \theta_0$ versus $H : \theta = \theta_1$ de nivel $\alpha = P_{\theta_0}[\delta(x) = 1]$.*

Demostración 7.2. Dado que $\delta = 1$ en un subconjunto de $\{x|\delta_k(x) = 1\}$, $\beta(\theta_0, \delta) \leq \beta(\theta_0, \delta_k)$ y

$$\delta_k(x)(P(x, \theta_1) - kP(x, \theta_0)) \geq \delta(x)(P(x, \theta_1) - kP(x, \theta_0)) \tag{7.7}$$

En el conjunto A, $P(x, \theta_1) = kP(x, \theta_0)$. Así, para todo x en A, ambos lados de (7.7) se reducen a 0. En el conjunto B, $\delta = \delta_k$; en consecuencia, (7.7) se reduce a una igualdad. Por lo anterior, integrando sobre $A \cup B$ a ambos lados de esta igualdad se obtiene

$$\beta(\theta_1, \delta_k) - \beta(\theta_1, \delta) = k \int_{A \cup B} (\delta_k(x) - \delta(x)) P(x, \theta_0) dx \geq 0$$

de donde se concluye que $\beta(\theta_1, \delta_k) \geq \beta(\theta_1, \delta)$.

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 sigue la primera desigualdad de

$$\begin{aligned} \beta(\theta_1, \delta_k) - \beta(\theta_1, \delta^*) &\geq k(\beta(\theta_0, \delta_k) - \beta(\theta_0, \delta^*)) \\ &\geq k(\beta(\theta_0, \delta_k) - \beta(\theta_0, \delta)) \\ &= \beta(\theta_1, \delta_k) - \beta(\theta_1, \delta) \end{aligned}$$

donde la igualdad se sigue a partir de (7.7). 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 $\alpha = P_{\theta_0}[\delta = 1]$

Ejemplo 7.4. Dada una observación de la variable aleatoria $X \sim U(0, \theta)$, deseamos encontrar la prueba uniformemente más potente para $\theta = \theta_0$ *versus* $\theta = \theta_1$, con $\theta_0 < \theta_1$. Entonces

$$L(x, \theta_0, \theta_1) = \begin{cases} \dfrac{\theta_0}{\theta_1} & \text{si } 0 \leq x \leq \theta_0 \\ \infty & \text{si } \theta_0 < x \leq \theta_1 \end{cases}$$

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 región de rechazo determinada por el conjunto $\{x : x \geq 0\}$ y probabilidad de error de tipo I, $\alpha = P_{\theta_0}[X \geq 0] = 1$. La prueba de nivel α, $0 \leq \alpha < 1$, se define considerando l, $0 < l \leq \theta_0$, tal que $\alpha = P_{\theta_0}[X \geq l] = \frac{\theta_0 - l}{\theta_0}$. En consecuencia, la prueba uniformemente más potente de nivel 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 $\beta(\theta, \delta) = 1 - (1 - \alpha)\frac{\theta_0}{\theta_1}$. Observe que para $\alpha = 0$, la región de rechazo es $RR = \{x : x \geq \theta_0\}$.

Ejemplo 7.5. Sea θ el tamaño de una población. Para determinar la prueba uniformemente más potente de nivel α para $H : \theta = \theta_0$ *versus* $K : \theta = \theta_1$ con $\theta_1 > \theta_0$

$$L(x, \theta_0, \theta_1) = \begin{cases} \left(\dfrac{\theta_0}{\theta_1}\right)^n & \text{si } 1 \leq \text{máx}(x_1, \ldots, x_n) \leq \theta_0 \\ \infty & \text{si } \theta_0 < \text{máx}(x_1, \ldots, x_n) \leq \theta_1 \end{cases}$$

ya que la función de densidad $f(x|\theta) = 1/\theta$ si $x = 1, 2, \ldots, \theta$.

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 uniformemente más potente de nivel

$$\begin{aligned} \alpha_j &= P_{\theta_0}(\text{máx}(x_1, \ldots, x_n) \geq j) \\ &= 1 - P_{\theta_0}(X_1 < j, \ldots, X_n < j) \\ &= 1 - \left(\frac{j-1}{\theta_0}\right)^n \end{aligned}$$

Observe que se ha definido una familia de pruebas uniformemente más potentes de nivel α_j, $j = 1, 2, \ldots, \theta_0$, definidas sobre el conjunto $A = \{x : L(X, \theta_0, \theta_1) = (\frac{\theta_0}{\theta_1})^n\}$.

Ejemplo 7.6. Sea $X_i \sim N(\theta, \sigma^2)$, $i = 1, 2, \ldots, n$, una muestra aleatoria de una distribución normal, con media θ desconocida y varianza σ^2 conocida. Para obtener la prueba uniformemente más potente para probar la hipótesis $H : \theta = \theta_0$ *versus* $K : \theta \neq \theta_0$, consideramos la estadística

$$S(X) = \frac{\sqrt{n}(\bar{X} - \theta_0)}{\sigma} \tag{7.8}$$

la cual, bajo H, tiene distribución normal estándar y, bajo K, distribución normal con media $\Delta = \frac{\sqrt{n}(\theta_1 - \theta_0)}{\sigma}$ y varianza 1. A partir de la estadística S, definimos la estadística $T(X) = |S(X)|$, la cual bajo la hipótesis K tiene como función de distribución

$$\begin{aligned} F_\Delta(z) &= P_\Delta(T(X) \leq z) \\ &= P_\Delta(|S(X)| \leq z) \\ &= \int_{-z-\Delta}^{z-\Delta} f(x)dx \end{aligned} \tag{7.9}$$

donde f corresponde a la función de densidad de la distribución normal estándar. Aplicando el teorema fundamental del cálculo en (7.9), se obtiene la función de densidad de T.

$$
\begin{aligned}
g(x,\Delta) &= F'_\Delta(z)\\
&= f(z-\Delta)+f(-z-\Delta)\\
&= f(z)e^{-\Delta^2/2}(e^{\Delta z}+e^{-\Delta z})
\end{aligned}
$$

Dado que el sistema de hipótesis puede expresarse en términos de Δ como $H:\Delta=0$ y $K:\Delta=\Delta_1$, para $z>0$,

$$
\begin{aligned}
L &= \frac{g(z,\Delta_1)}{g(z,0)}\\
&= e^{-\Delta_1^2/2}\left(\frac{e^{\Delta_1 z}+e^{-\Delta_1 z}}{2}\right)
\end{aligned}
$$

es una función creciente de z. En consecuencia, dado que $\alpha = P_0(T \geq z_{\alpha/2})$, δ_α es la prueba uniformemente más potente de nivel 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 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 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 α para $H:\theta=\theta_0$ *versus* $K:\theta=\theta_1$, $\theta_1<\theta_0$ sea $RR=\{\bar{x}:\bar{x}<0.2\}$.

2. Resuelva el ejercicio 1 si $X_i, i=1,2,\ldots,n$, 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 $X\sim Bin(n,\theta)$. Encuentre la región de rechazo de la prueba uniformemente más potente de nivel α para probar

   a) $H:\theta=\theta_0$ *versus* $K:\theta=\theta_1$, $\theta_1<\theta_0$.

b) $H : \theta = \theta_0$ *versus* $K : \theta = \theta_1$,$\theta_1 > \theta_0$.

4. Si $X_i, i = 1, 2, \ldots, n$, proviene de una distribución geométrica, determine la región de rechazo de la prueba uniformemente más potente de nivel 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.

5. Suponga que X tiene distribución $H(N\theta, N, n)$. Determine la prueba uniformemente más potente para probar $H : \theta = \theta_0$ *versus* $K : \theta = \theta_1$, con $\theta_1 > \theta_0$.

7.2. Prueba de razón de verosimilitudes

Suponga que $X = (X_1, \ldots, X_n)$ tiene función de densidad o función de probabilidad $P(x, \theta)$, siendo x una posible realización de X, y que deseamos probar la hipótesis $H : \theta \in \Theta_0$ *versus* $K : \theta \in \Theta_1$. Como estadístico, consideramos la razón de verosimilitud dada por

$$L(x, \theta_0, \theta_1) = \frac{\sup\{p(x, \theta) : \theta \in \Theta_1\}}{\sup\{p(x, \theta) : \theta \in \Theta_0\}} \tag{7.10}$$

Esta prueba rechaza H para valores grandes de L. En consecuencia, se rechaza H si y solo si x es más probable para algún θ en Θ_1.

Si $p(x, \theta)$ es una función continua de θ y Θ_0 es de menor dimensión que $\Theta = \Theta_0 \cup \Theta_1$, la estadística (7.10) es igual a

$$\lambda(x) = \frac{\sup\{p(x, \theta) : \theta \in \Theta\}}{\sup\{p(x, \theta) : \theta \in \Theta_0\}} \tag{7.11}$$

Esta estadística rechaza H para valores grandes de $\lambda(x)$. El siguiente ejemplo ilustra los pasos a seguir en pruebas de hipótesis en las que se aplica (7.11).

Ejemplo 7.7. Sea $X \sim \exp(\theta)$, donde $f_\theta(x) = \frac{1}{\theta} e^{-\frac{1}{\theta} x}$. Si se desea probar $H : \theta = \theta_0$ *versus* $K : \theta \neq \theta_0$,

$$L(x, \Theta_0, \Theta_1) = \frac{\frac{1}{\bar{x}^n} e^{-\frac{1}{\bar{x}} \sum_{i=1}^n x_i}}{\frac{1}{\theta_0^n} e^{-\frac{1}{\theta_0} \sum_{i=1}^n x_i}} = \left(\frac{\theta_0}{\bar{x}}\right)^n e^{n\left(\frac{\bar{x}}{\theta_0} - 1\right)}$$

Derivando L con respecto a $\bar{x}$ se obtiene

$$\frac{\partial L}{\partial \bar{x}} = n \left(\frac{\theta_0}{\bar{x}}\right)^n \left(\frac{1}{\theta_0} - \frac{1}{\bar{x}}\right) \exp\left(\frac{n\bar{x}}{\theta_0} - n\right)$$

de donde, si $\bar{x} > \theta_0$, L es una función creciente de $\bar{x}$ y, si $0 < \bar{x} < \theta_0$, L es una función decreciente de $\bar{x}$. En consecuencia, en una prueba de nivel 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

$$\frac{\alpha}{2} = P_{\theta_0}\left[\sum_{i=1}^n x_i < c_1\right] \quad \text{ó} \quad \frac{\alpha}{2} = P_{\theta_0}\left[\sum_{i=1}^n x_i > c_2\right]$$

Dado que $\frac{2}{\theta}\sum_{i=1}^n x_i \sim \chi^2_{2n}$, $c_1 = (\theta_0 x^2_{1-\alpha/2})/2$ y $c_2 = (\theta_0 x^2_{\alpha/2})/2$.

Ejemplo 7.8. Suponga que tenemos observaciones provenientes de una distribución normal $N(\theta, \sigma^2)$, con θ y σ^2 desconocidas, y que deseamos probar $H : \theta = \theta_0$ *versus* $K : \theta \neq \theta_0$. El espacio paramétrico asociado a la hipótesis H es $\Theta_0 = \{(\theta, \sigma^2) | \theta = \theta_0, \sigma^2 > 0\}$. Haciendo $\Theta = \{(\theta, \sigma^2) | \theta \in R, \sigma^2 > 0\}$, se observa que Θ_0 es de menor dimensión que Θ. En consecuencia, dado que $p(x, \theta)$ es una función continua de θ, se aplica la estadística (7.11) al desarrollo de la prueba de hipótesis, teniendo en cuenta los siguientes pasos:

1. Determinar las estimaciones de máxima verosimilitud $(\hat{\theta}, \hat{\sigma}^2)$ de (θ, σ^2) sobre el espacio Θ. En el capítulo 2 hemos encontrado 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. Calcular las estimaciones de máxima verosimilitud de θ sobre Θ_0. Dado que $\theta = \theta_0$, únicamente debemos calcular $\hat{\sigma}_0^2$. Es fácil ver que $\hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \theta_0)^2$.

3. Formar el cociente $\lambda(X)$.

$$\begin{aligned}
\lambda(X) &= \frac{p(X|\hat{\theta}, \hat{\sigma}^2)}{p(X|\hat{\theta}_0, \hat{\sigma}_0^2)} \\
&= \frac{(2\pi\hat{\sigma}^2)^{-n/2} \exp\left[-\frac{1}{2\hat{\sigma}^2}\sum (X_i - \bar{X})\right]}{(2\pi\hat{\sigma}_0^2)^{-n/2} \exp\left[-\frac{1}{2\hat{\sigma}_0^2}\sum (X_i - \theta_0)^2\right]} \\
&= \left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2}\right)^{n/2} \qquad (7.12)
\end{aligned}$$

Dado que $\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} = \frac{\sum (X_i - \bar{X})^2 + n(\bar{X} - \theta_0)^2}{\sum (X_i - \bar{X})^2}$, se halla $\lambda(X) = (1 + \frac{1}{n-1}T^2)^{n/2}$, donde $T = \frac{\sqrt{n}(\bar{X} - \theta_0)}{S}$ y $S^2 = \frac{1}{n-1}\sum (X_i - \bar{X})^2$.

4. Encontrar una función $h(\lambda(X))$ estrictamente creciente definida en el rango de λ, tal que $h(\lambda(X))$ tenga una distribución conocida.

   Dado que λ es una función creciente de T^2 y que la región de rechazo de H_0 corresponde a $\lambda(X) = (1 + \frac{1}{n-1}T^2)^{n/2} > c$, la prueba del cociente de máxima verosimilitud rechaza H si $T^2(X) > c_1$, o equivalentemente si $|T(X)| > c'$, donde c, c_1 y c' son números reales positivos. Dado que $T|H \sim t_{n-1}(0, 1)$, el valor c' para la prueba de nivel α es $t_{\alpha/2,n-1}$.

Finalmente, asumiendo un $\theta_1 \neq \theta_0$, podemos determinar la función de potencia de la prueba:

$$\begin{aligned}
\beta(\theta_1, \delta) &= P_{\theta_1}\left(\left|\frac{\sqrt{n}(\bar{X} - \theta_0)}{S}\right| > t_{\frac{\alpha}{2},n-1}\right) \\
&= 1 - P_{\theta_1}\left(\frac{\sqrt{n}\theta_0}{S} - t_{\frac{\alpha}{2},n-1} < \frac{\sqrt{n}\bar{X}}{S} < \frac{\sqrt{n}\theta_0}{S} + t_{\frac{\alpha}{2},n-1}\right) \\
&= 1 - P_{\theta_1}\left(\frac{\sqrt{n}(\theta_0 - \theta_1)}{S} - t_{\frac{\alpha}{2},n-1} < \frac{\sqrt{n}(\bar{X} - \theta_1)}{S} < \frac{\sqrt{n}(\theta_0 - \theta_1)}{S} + t_{\frac{\alpha}{2},n-1}\right)
\end{aligned}$$

que es una función creciente, dependiente de $\theta_0 - \theta_1$.

Ejemplo 7.9. Suponga 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 $N(\theta_2, \sigma^2)$, respectivamente. Asuma que deseamos 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

$$\Theta_0 = \{(\theta_1, \theta_2, \sigma^2) : \theta_1 = \theta_2 = \theta, \theta \in R, \sigma^2 > 0\}$$

que es subespacio de $\Theta = \{(\theta_1, \theta_2, \sigma^2) : \theta_1 \in R, \theta_2 \in R, \sigma^2 > 0\}$. Si $\hat{\theta}$ es la estimación de máxima verosimilitud de θ, y $\bar{x}_1$ y $\bar{x}_2$ las estimaciones de máxima verosimilitud de θ_1 y θ_2, siguiendo el procedimiento indicado en el ejemplo anterior, en el paso 3 se encuentra $\lambda(X_1, X_2) = (\hat{\sigma}_0^2/\hat{\sigma}^2)^{(n_1+n_2)/2}$, donde

$$n\hat{\sigma}_0^2 = \sum_{i=1}^{n_1}(x_{1i} - \hat{\theta})^2 + \sum_{i=1}^{n_2}(x_{2i} - \hat{\theta})^2$$

y

$$n\hat{\sigma}^2 = \sum_{i=1}^{n_1}(x_{1i} - \bar{x}_1)^2 + \sum_{i=1}^{n_2}(x_{2i} - \bar{x}_2)^2$$

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

$$\sum_{i=1}^{n_2}(x_{2i}-\hat{\theta})^2 = \sum_{i=1}^{n_2}(x_{2i}-\bar{x}_2)^2 + \frac{n_1^2 n_2}{(n_1+n_2)^2}(\bar{x_1}-\bar{x_2})^2$$

reemplazando en el cociente de máxima verosimilitud se obtiene

$$\lambda(X_1,X_2) = (\hat{\sigma}_0^2/\hat{\sigma}^2)^{(n_1+n_2)/2} = \left(1+\frac{T^2}{n_1+n_2-2}\right)^{(n_1+n_2)/2}$$

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

$$\lambda(X) = \left(1+\frac{T^2}{n_1+n_2-2}\right)^{(n_1+n_2)/2} > c$$

la prueba del cociente de máxima verosimilitud rechaza H_0 si $T^2(X) > c_1$, es decir, si $|T(X)| > c$, donde $c = \sqrt{c_1}$. Dado que $T|H \sim t_{n_1+n_2-2}$, el valor c para la prueba de nivel α es $t_{\alpha/2,n_1+n_2-2}$. La función de potencia puede ser determinada como en el ejercicio anterior.

Ejercicios 7.2.

1. Suponga que tenemos observaciones provenientes de una distribución normal $N(\theta,\sigma^2)$, con θ y σ^2 desconocidas, y que deseamos probar $H:\theta \leq \theta_0$ *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 una muestra de una distribución uniforme en el intervalo $(0,\theta)$. Encuentre la región de rechazo de la prueba uniformemente más potente de nivel α para la prueba de hipótesis de $H:\theta=\theta_0$ *versus* $K:\theta \neq \theta_0$.

3. Sea $X_i \sim N(\mu, \gamma\mu)$, $i = 1, 2, \ldots, n$, independientes. Encuentre un test de LR para probar $H_0 : \gamma = \gamma_0$ *versus* $K : \gamma \neq \gamma_0$.

4. Suponga que tenemos una observación de una distribución $\exp(\theta)$. 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 α. Determine la función de potencia. Si el valor observado X=3 y $\theta_0 = 1$, encuentre el p-valor de la prueba.

Bibliografía

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.

Bickel, P. J. & Doksun, K. A. (1977), *Mathematical Statistics*, Prentice-Hall, New York.

Dudewicz, E. J. & Mishra, S.Ñ. (1988), *Modern Mathematical Statistics*, John Wiley & Sons, New York.

Knight, K. (2000), *Mathematical Statistics*, Chapman & Hall/Crc., New York.

Laplace, P. S. (1812), *Théorie Analytique des Probabilitiés*, Courcíer, Paris.

Shao, J. (2003), *Mathematical Statistics*, Springer, New York.

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.