Ejercicios Resueltos y Propuestos
Curso EYP 2113
Tomo I

Primera Edición

Trabajo de Recopilación , Organización y Elaboración
Patricia Jiménez P. & Ricardo Olea O.
Dpto. de Estadística - Facultad de Matemáticas
Pontificia Universidad Católica de Chile

Santiago, Diciembre 2004

Prefacio

Con la intención de apoyar la labor docente que desarrolla el Departamento de Estadística de la Facultad de Matemáticas de la Pontificia Universidad Católica de Chile, se ha realizado un trabajo de recopilación y elaboración de ejercicios resueltos y propuestos para el curso EYP2113, algunos de los cuales fueron desarrollados en ayudantías y han sido parte de interrogaciones en semestres anteriores.

Queremos agradecer muy en especial a FONDEDOC, por haber confiado en este proyecto y habernos entregado todo su apoyo para poder ver realizada esta necesidad tanto para el Departamento de Estadística, como para todos los alumnos y alumnas que son beneficiados de los cursos de servicio que ofrece el mismo.

Este trabajo ha sido fruto de la labor que desarrollaron docentes y ayudantes que dictaron el curso durante los años 2002 y 2003.

Específicamente deseamos agradecer a los profesores

- Ricardo Aravena
- Alejandro Jara
- Ignacio Vidal

Además quisiéramos agradecer el aporte de Jorge González, Mario Tagle y Joaquín Rojas, tanto por el material donado, como por la revisión de este libro.

Atentamente.

Dirección
Departamento de Estadística
Facultad de Matemáticas

Santiago, Diciembre 2004

Índice general

Capítulo 1

Estimación

1.1. Ejercicios Resueltos

EJERCICIO 1

Suponga que se tiene una m.a. de tamaño $2n$ tomada de una población X, con $E(X) = \mu$ y $Var(X) = \sigma^2$. Sean:

$$\overline{X}_1 = \frac{1}{2n}\sum_{i=1}^{2n} x_i \quad y \quad \overline{X}_2 = \frac{1}{n}\sum_{i=1}^{n} x_i$$

dos estimadores de μ.¿Cuál es el mejor estimador de μ?. Explique su elección.

<u>SOLUCIÓN</u>

El mejor estimador será aquel que tenga menor error cuadrático medio E.C.M.. Primero veamos si son insesgados los estimadores.

$$E(\overline{X}_1) = \frac{1}{2n}E\left(\sum_{i=1}^{2n} x_i\right) = \frac{1}{2n}\sum_{i=1}^{2n} E(x_i) = \frac{1}{2n}2n\mu = \mu$$

$$E(\overline{X}_2) = \frac{1}{n}E\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}n\mu = \mu$$

Luego ambos estimadores son insesgados, por lo tanto el mejor estimador de entre los dos, será aquel que tenga menor varianza.

$$Var(\overline{X}_1) = \frac{1}{4n^2}Var\left(\sum_{i=1}^{2n} x_i\right) \overset{ind}{=} \frac{1}{4n^2}\sum_{i=1}^{2n} Var(x_i) = \frac{1}{4n^2}2n\sigma^2 = \frac{\sigma^2}{2n}$$

$$Var(\overline{X}_2) = \frac{1}{n^2}Var\left(\sum_{i=1}^{n} x_i\right) \overset{ind}{=} \frac{1}{n^2}\sum_{i=1}^{n} Var(x_i) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$$

Luego, como el que tiene menor varianza es $\overline{X}_1$, escogemos éste, pues esto produce un menor E.C.M.

EJERCICIO 2

Suponga que $\hat{\Theta}_1$ y $\hat{\Theta}_2$ son estimadores insesgados del parámetro θ. Se sabe que $Var(\hat{\Theta}_1) = 10$ y $Var(\hat{\Theta}_2) = 4$.¿Cuál es el mejor y en que sentido lo es?

SOLUCIÓN

Como ambos son insesgados, el mejor estimador será aquel que tenga menor varianza, lo que, en este caso, conlleva a tener un menor E.C.M.. Luego observando, vemos que $\hat{\Theta}_2$ tiene menor varianza que $\hat{\Theta}_1$, por lo tanto escogemos $\hat{\Theta}_2$ como mejor estimador de θ.

EJERCICIO 3

Suponga que $\hat{\Theta}_1$ y $\hat{\Theta}_2$ son estimadores del parámetro θ. Se sabe que $E(\hat{\Theta}_1) = \theta$, $E(\hat{\Theta}_2) = \frac{\theta}{2}$, $Var(\hat{\Theta}_1) = 10$ y $Var(\hat{\Theta}_2) = 4$. ¿Cuál es el mejor y en qué sentido lo es?

SOLUCIÓN

Si observamos cuidadosamente, vemos que $\hat{\Theta}_1$ es insesgado para θ pero que $\hat{\Theta}_2$ no lo es. Ahora la mejor forma de ver cual es mejor es comparando los E.C.M. de cada uno, ya que esta medida considera el sesgo producido por cada estimador y la varianza que tienen.

$$E.C.M.(\hat{\Theta}_1) = Var(\hat{\Theta}_1) + Sesgo^2(\hat{\Theta}_1) = 10 + 0^2 = 10$$

$$E.C.M.(\hat{\Theta}_2) = Var(\hat{\Theta}_2) + Sesgo^2(\hat{\Theta}_2) = 4 + \left(\frac{\theta}{2} - \theta\right)^2 = 4 + (-\frac{\theta}{2})^2$$

Como se puede ver, el E.C.M. de Θ_2 depende del verdadero valor que tiene θ, luego debemos hacer un análisis más detallado, para saber cuando $\hat{\Theta}_2$ será mejor que $\hat{\Theta}_1$.

Cuando ocurre:

$$E.C.M.(\hat{\Theta}_1) \leq E.C.M.(\hat{\Theta}_2)$$

$$10 \leq 4 + \left(-\frac{\theta}{2}\right)^2$$

$$10 \leq \frac{16 + \theta^2}{4}$$

$$40 \leq 16 + \theta^2$$

$$\theta^2 \geq 40 - 16$$

$$\theta^2 \geq 24$$

Es decir, $\hat{\Theta}_1$ será mejor estimador de θ que $\hat{\Theta}_2$ cuando el verdadero valor de θ sea:

$$\theta \geq \sqrt{24} \quad \text{o cuando} \quad \theta \leq -\sqrt{24}$$

Equivalentemente, $\hat{\Theta}_2$ será mejor estimador de θ que $\hat{\Theta}_1$ cuando

$$-\sqrt{24} < \theta < \sqrt{24}$$

EJERCICIO 4

Sea $X_1, \ldots, X_n$ una muestra aleatoria de tamaño n, de una población $N(\mu, \sigma^2)$.

(a) Demuestre que $\overline{X}^2$ es un estimador sesgado de μ^2.

(b) Determine la magnitud del sesgo en el estimador.

(c) ¿Qué sucede con el sesgo a medida que aumenta el tamaño n de la muestra?

SOLUCIÓN

(a) Como $X \sim N(\mu, \sigma^2)$ entonces se sabe que $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$, luego si queremos demostrar que $\overline{X}^2$ es sesgado para μ^2 ocupamos la siguiente relación:

$$Var(\overline{X}) = E(\overline{X}^2) - E^2(\overline{X})$$

$$\frac{\sigma^2}{n} = E(\overline{X}^2) - \mu^2$$

Luego despejando lo que necesitamos, obtenemos:

$$E(\overline{X}^2) = \frac{\sigma^2}{n} + \mu^2$$

Lo cual es distinto de μ^2, que es el caso donde habría sido insesgado el estimador.

(b) La magnitud del sesgo, no es más que el tamaño de éste, es decir, su valor.

$$Sesgo(\overline{X}^2) = E(\overline{X}^2) - \mu^2$$

$$Sesgo(\overline{X}^2) = \frac{\sigma^2}{n} + \mu^2 - \mu^2 = \frac{\sigma^2}{n}$$

(c) A medida que el tamaño de muestra aumenta, el sesgo $\frac{\sigma^2}{n} \to 0$, es decir, el estimador es asintóticamente (cuando $n \to \infty$) insesgado.

EJERCICIO 5

Una máquina produce artículos defectuosos con probabilidad π. En la inspección de artículos se define la v.a.

$$Y_i = \begin{cases} 1, & \text{si el artículo i es defectuoso;} \\ 0, & \text{si el artículo i no es defectuoso.} \end{cases}$$

En una muestra de tamaño 5 se observan dos artículos defectuosos:

(a) Proponga un modelo apropiado para el problema y estime la proporción de artículos defectuosos usando el método de máxima verosimilitud.

SOLUCIÓN

(a) Dada la definición del problema y la estructura de la variable aleatoria, Y tiene una distribución Bernoulli

$$Y \sim Ber(p) \rightarrow P(Y = y) = p^y(1-p)^{1-y}$$

donde el parámetro p, que es la probabilidad del éxito, es desconocida, por lo que la estimaremos por máxima verosimilitud.

$$L(y;p) = \prod_{i=1}^{5} p^{y_i}(1-p)^{1-y_i}$$

$$L(y;p) = p^{\sum_{i=1}^{5} y_i}(1-p)^{5-\sum_{i=1}^{5} y_i}$$

Aplicando logaritmo natural, obtenemos:

$$\ell(y;p) = \left(\sum_{i=1}^{5} y_i\right)\ln(p) + \left(5 - \sum_{i=1}^{5} y_i\right)\ln(1-p)$$

Luego, para maximizar la función de verosimilitud, derivamos con respecto al parámetro p, que es el que estamos buscando e igualamos a cero para despejar p.

$$\frac{\partial \ell}{\partial p} = \frac{\sum_{i=1}^{5} y_i}{p} + \frac{\left(\sum_{i=1}^{5} y_i\right) - 5}{1-p} = 0$$

$$\hat{p} = \frac{\sum_{i=1}^{5} y_i}{5}$$

Pero como nos dicen que se observaron dos artículos defectuosos, es decir sólo dos de los y_i son 1, la suma de éstos es 2,

$$\therefore \quad \hat{p} = \frac{2}{5}$$

EJERCICIO 6

El número de conexiones mal soldadas por microcircuito integrado en una operación de manufactura electrónica sigue una distribución Binomial(20,p) con p desconocida. El costo de corregir los errores, por microcircuito, es:

$$C = 3X + X^2$$

En base a una muestra aleatoria $X_1, X_2, ..., X_n$ encuentre el EMV del costo total esperado de corregir los errores de estos microcircuitos observados.

<u>SOLUCIÓN</u>

Como p es desconocida, hay que estimarla en una primera instancia, en este caso por máxima verosimilitud.

$$L(x;p) = \prod_{i=1}^{n} \binom{20}{x_i} p^{x_i}(1-p)^{20-x_i}$$

$$L(x;p) = p^{\sum_{i=1}^{n} x_i}(1-p)^{20n-\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} \binom{20}{x_i}$$

Aplicando logaritmo natural, obtenemos:

$$\ell(x;p) = \left(\sum_{i=1}^{n} x_i\right) \ln(p) + \left(20n - \sum_{i=1}^{n} x_i\right) \ln(1-p)$$

Luego, para maximizar la función de verosimilitud, derivamos con respecto al parámetro p, que es el que estamos buscando e igualamos a cero para despejar p.

$$\frac{\partial \ell}{\partial p} = \frac{\sum_{i=1}^{n} x_i}{p} + \frac{\left(\sum_{i=1}^{n} x_i\right) - 20n}{1-p} = 0$$

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i}{20n}$$

Ahora procedemos a calcular el costo esperado por medio de la esperanza.

$$\begin{aligned} E(C) &= 3E(X) + E(X^2) \\ &= 3np + np(1-p) + (np)^2 \\ &= 3np + np - np^2 + n^2p^2 \\ &= 4np + np^2(n-1) \end{aligned}$$

Luego el E.M.V. de $E(C)$ es $\widehat{E(C)} = 4n\widehat{p} + n\widehat{p}^2(n-1)$ por invarianza del E.M.V.

EJERCICIO 7

En encuestas, es difícil obtener respuestas precisas a preguntas delicadas tales como ¿Has usado alguna vez heroína? o ¿Has hecho trampa alguna vez en un examen?. Warner introdujo el método de respuestas aleatorizadas para tratar tales situaciones. El encuestado hace girar una flecha en una rueda o extrae una bola desde una urna que contiene dos bolas de dos colores para determinar cual de las dos afirmaciones contestará: (1) "Tengo la característica A", o (2) "No tengo la característica A". El encuestador no conoce cual afirmación será contestada pero solamente anotará un sí o un no. Se cree que es más probable que el encuestado responda verazmente si él o ella saben que el encuestador no conoce cual afirmación será contestada. Sea R la proporción de una muestra que contesta Sí. Sea p la probabilidad que la afirmación 1 sea contestada (p es conocido desde la estructura del método aleatorizado), y sea q la proporción de la población que tiene la característica A. Sea r la probabilidad que un encuestado responda sí.

(a) Muestre que $r = (2p-1)q + (1-p)$

(b) Si r es conocida, ¿Cómo podría determinarse q?

SOLUCIÓN

Definamos como:

R: Proporción de la muestra que contesta sí.

p: Probabilidad que la afirmación 1 sea contestada.

q: Probabilidad de la población que tiene la característica A.

r: Probabilidad que un encuestado responda si.

(a)

$$
\begin{aligned}
r &= P(\text{responda sí}) \\
&= P(\text{responda sí} \,|\, \text{contesta afirmación 1}) P(\text{contesta afirmación 1}) \\
&\quad + P(\text{responda sí} \,|\, \text{no contesta afirmación 1}) P(\text{no contesta afirmación 1}) \\
&= pq + (1-p)(1-q) \\
&= pq + 1 - p - q + pq \\
&= 2pq + 1 - p - q \\
&= (2pq + 1)q + (1-p)
\end{aligned}
$$

(b) Sería cosa de despejar q, es decir,

$$r - (1-p) = (2p-1)q$$

luego

$$q = \frac{r + p - 1}{2p - 1}$$

EJERCICIO 8

Suponga que $X_1, X_2, ..., X_n$ constituye una m.a. de una distribución cuya función densidad es la siguiente

$$f(x;\theta) = \begin{cases} \theta x^{\theta-1}, & 0 < x < 1; \\ 0, & \text{e.o.c.} \end{cases}$$

Además, suponga que el valor de θ es desconocido ($\theta > 0$).

(a) Determine el EMV de θ.

(b) Determine el EMV de $E(X)$.

<u>**SOLUCIÓN**</u>

(a)

$$\begin{aligned}
L(x_1,\ldots,x_n;\theta) &= \prod_{i=1}^{n} \theta x_i^{\theta-1} \\
&= \theta^n \left(\prod_{i=1}^{n} x_i\right)^{\theta-1} \qquad / \ln \\
l(x_1,\ldots,x_n;\theta) &= n\ln\theta + (\theta-1)\sum_{i=1}^{n}\ln x_i \qquad /\partial\theta \\
\tfrac{\partial l(x_1,\ldots,x_n,\theta)}{\partial\theta} &= \tfrac{n}{\theta} + \sum_{i=1}^{n}\ln x_i \\
\Rightarrow \qquad \widehat{\theta}_{EMV} &= \frac{-n}{\displaystyle\sum_{i=1}^{n}\ln x_i}
\end{aligned}$$

(b)

$$E(X) = \int_0^1 x\theta x^{\theta-1}dx = \int_0^1 \theta x^{\theta}dx = \frac{\theta x^{\theta+1}}{\theta+1}\Big|_0^1 = \frac{\theta}{\theta+1}$$

Luego $\widehat{E(X)} = \frac{\hat{\theta}}{\hat{\theta}+1}$ por la invarianza del E.M.V.

EJERCICIO 9

Sean $X_1, X_2, \ldots, X_n$ variables aleatorias i.i.d. con función densidad dada por

$$f_X(x) = \begin{cases} (\alpha+1)x^{\alpha} & ;0<x<1 \\ 0 & ;\text{e.o.c.} \end{cases}$$

(a) Encuentre el estimador de α por el método de momentos.

(b) Encuentre el estimador de α por el método de máxima verosimilitud.

(c) Evalúe ambos estimadores usando los siguientes datos:

X	0.1 - 0.3	0.3 - 0.6	0.6 - 0.7	0.7 - 0.9
Frecuencia	3	1	2	3

<u>SOLUCIÓN</u>

(a) El método de momentos consiste en igualar el momento muestral con el momento poblacional.

Para el caso $k = 1$ tenemos la siguiente igualdad

$$E(X) = \overline{X}$$

Necesitamos calcular $E(X)$:

$$\begin{aligned} E(X) &= \int_{\text{Rec X}} x f_X(x) dx = \int_{\text{Rec X}} x \cdot (\alpha+1) x^{\alpha} dx \\ &= (\alpha+1) \int_0^1 x^{\alpha+1} dx \\ &= (\alpha+1) \frac{x^{\alpha+2}}{\alpha+2} \Big|_{x=0}^{x=1} \\ &= \frac{\alpha+1}{\alpha+2} \end{aligned}$$

Luego

$$\begin{aligned} E(X) &= \overline{X} \\ \Rightarrow \frac{\alpha+1}{\alpha+2} &= \overline{X} \\ \Rightarrow \alpha + 1 &= \alpha\overline{X} + 2\overline{X} \\ \Rightarrow \alpha(1-\overline{X}) &= 2\overline{X} - 1 \\ \Rightarrow \hat{\alpha}_{MM} &= \frac{2\overline{X}-1}{1-\overline{X}} \end{aligned}$$

(b)

$$L(\mathbf{x}; \alpha) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} (\alpha+1) x_i^{\alpha} = (\alpha+1)^n \left(\prod_{i=1}^{n} x_i \right)^{\alpha} \qquad \backslash \ln$$

$$\ell(\mathbf{x}; \alpha) = n \ln(\alpha+1) + \alpha \sum_{i=1}^{n} \ln(x_i) \qquad \backslash \partial\alpha$$

$$\frac{\partial \ell}{\partial \alpha} = \frac{n}{\alpha+1} + \sum_{i=1}^{n} \ln(x_i) = 0$$

$$\Rightarrow \frac{n}{\alpha+1} = -\sum_{i=1}^{n} \ln(x_i)$$

$$\Rightarrow \alpha + 1 = \frac{-n}{\displaystyle\sum_{i=1}^{n} \ln(x_i)}$$

$$\Rightarrow \hat{\alpha}_{MV} = -\left(\frac{n}{\displaystyle\sum_{i=1}^{n} \ln(x_i)} + 1 \right)$$

(c) Para evaluar los estimadores necesitamos convertir los datos tabulados a un set de datos compuestos por las marcas de cada clases

Marca de Clase X	(0.2) 0.1 - 0.3	(0.45) 0.3 - 0.6	(0.65) 0.6 - 0.7	(0.8) 0.7 - 0.9
Frecuencia	3	1	2	3

Se puede representar el conjunto de valores para X como:

$$[X : 0{,}2; 0{,}2; 0{,}2; 0{,}45; 0{,}65; 0{,}65; 0{,}8; 0{,}8; 0{,}8]$$

Calculando ahora lo necesario para poder evaluar los estimadores con estos datos tabulados

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{k} m_i \cdot f_i = \frac{1}{9} \left(0{,}2 \cdot 3 + 0{,}45 \cdot 1 + 0{,}65 \cdot 2 + 0{,}8 \cdot 3\right) = 0{,}5277$$

$$\sum_{i=1}^{n} \ln(x_i) = \ln\left(\prod_{i=1}^{n} x_i\right) = \ln(0{,}000778752) = -7{,}15781$$

Luego al evaluar estos resultados en los estimadores, estos toman los siguientes valores:

$$\hat{\alpha}_{MM} = \frac{2\overline{X} - 1}{1 - \overline{X}} = \frac{2 \cdot 0{,}52777 - 1}{1 - 0{,}52777} = 0{,}117612$$

$$\hat{\alpha}_{MV} = -\left(\frac{n}{\sum_{i=1}^{n} \ln(x_i)} + 1\right) = -\left(\frac{9}{-7{,}15781} + 1\right) = 0{,}257367$$

EJERCICIO 10

Sean $X_1, ..., X_n, Y_1, ..., Y_n$ v.a. independientes con $X_i \sim Exp(\frac{1}{\alpha})$ e $Y_j \sim Exp(\frac{1}{\beta})$, con $i = 1, ..., n; j = 1, ..., n$. Se define el parámetro $\theta = (\theta_1, \theta_2)$ por $\theta_1 = \alpha$ y $\theta_2 = \frac{\beta}{\alpha}$.

(a) Determine los EMV (estimador máximo verosímil) para θ_1 y θ_2

(b) Encuentre el sesgo y el ECM (error cuadrático medio) de $\hat{\theta}_1$

<u>SOLUCIÓN</u>

(a) Dada la independencia existente entre las variables, tenemos que $f_{X_i,Y_j}(x_i, y_j) = \frac{1}{\alpha} e^{-\frac{x_i}{\alpha}} \frac{1}{\beta} e^{-\frac{y_i}{\beta}}$, luego la verosimilitud conjunta es la siguiente:

$$\begin{aligned}
L(x, y; \boldsymbol{\theta}) &= \tfrac{1}{\alpha^n \beta^n} e^{-\frac{\sum x_i}{\alpha}} e^{-\frac{\sum y_i}{\beta}} && \backslash \ln \\
\ell(x, y; \boldsymbol{\theta}) &= -\tfrac{\sum x_i}{\alpha} - \tfrac{\sum y_i}{\beta} - n\ln(\alpha) - n\ln(\beta) && \backslash \partial\alpha \\
\tfrac{\partial \ell}{\partial \alpha} &= \tfrac{\sum x_i}{\alpha^2} - \tfrac{n}{\alpha} = 0 \longrightarrow \hat{\alpha} = \overline{x} = \hat{\theta}_1 \\
\tfrac{\partial \ell}{\partial \beta} &= \tfrac{\sum y_i}{\beta^2} - \tfrac{n}{\beta} = 0 \longrightarrow \hat{\beta} = \overline{y}
\end{aligned}$$

Tenemos que por la invarianza de los EMV's, $\hat{\theta}_2 = \frac{\hat{\beta}}{\hat{\alpha}}$, luego reemplazando queda que $\hat{\theta}_2 = \frac{\overline{y}}{\overline{x}}$.

(b) Por fórmula el $ECM(\hat{\theta}_1) = Var(\hat{\theta}_1) + Sesgo^2(\hat{\theta}_1)$, donde el $Sesgo(\hat{\theta}_1) = E(\hat{\theta}_1) - \theta_1$. Luego veremos primero si tiene sesgo (sesgado):

$$E(\hat{\theta}_1) = E\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}\sum_{i=1}^{n} \alpha = \alpha$$

Luego como es insesgado (recuerde que $\theta_1 = \alpha$), el $Sesgo(\hat{\theta}_1) = 0$.

Por lo tanto para calcular el $ECM(\hat{\theta}_1)$ basta calcular su varianza.

$$Var(\hat{\theta}_1) = Var\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \frac{1}{n^2}Var\left(\sum_{i=1}^{n} x_i\right)$$

$$\overset{ind}{=} \frac{1}{n^2}\sum_{i=1}^{n} Var(x_i) = \frac{1}{n^2}\sum_{i=1}^{n}\alpha^2 = \frac{\alpha^2}{n}$$

Por lo tanto $ECM(\hat{\theta}_1) = \frac{\alpha^2}{n}$ el cual $\overset{n\to\infty}{\longrightarrow} 0$.

EJERCICIO 11

Sean $X_1, ..., X_n$ iid con densidad $\lambda e^{-\lambda x}, \quad x \geq 0, \quad n \geq 2$. Sea $S_n = \sum_{i=1}^{n} X_i$. Es bien conocido que $Z = \lambda S_n$ tiene densidad:

$$f_Z(z) = \frac{z^{n-1}e^{-z}}{(n-1)!}, \quad z \geq 0$$

Utilice esto para calcular el sesgo y el ECM de $\hat{\lambda} = \frac{n-1}{S_n}$.

<u>SOLUCIÓN</u>

Necesitamos calcular la esperanza y varianza de $\hat{\lambda}$.

$$E(\hat{\lambda}) = E\left(\frac{n-1}{\sum x_i}\right) = (n-1)E\left(\frac{1}{\sum x_i}\right) = (n-1)E\left(\frac{\lambda}{\lambda\sum x_i}\right)$$

$$= (n-1)E\left(\frac{\lambda}{Z}\right) = \lambda(n-1)E\left(\frac{1}{Z}\right)$$

$$= \lambda(n-1)E(Z^{-1})$$

$$= \lambda(n-1)\int_0^\infty z^{-1}\frac{z^{n-1}e^{-z}}{(n-1)!}dz = \lambda(n-1)\int_0^\infty \frac{z^{n-2}e^{-z}}{(n-1)!}dz$$

$$= \frac{\lambda(n-1)}{n-1}\int_0^\infty \frac{z^{n-2}e^{-z}}{(n-2)!}dz = \lambda$$

Por lo tanto $\hat{\lambda}$ es un estimador insesgado, es decir, $Sesgo(\hat{\lambda}) = 0$. Luego queda que:

$$
\begin{aligned}
ECM(\hat{\lambda}) = Var(\hat{\lambda}) &= E(\hat{\lambda}^2) - E^2(\hat{\lambda}) \\
&= E\left(\left[\frac{n-1}{S_n}\right]^2\right) - \lambda^2 \\
&= (n-1)^2 E\left(\frac{\lambda^2}{Z^2}\right) - \lambda^2 \\
&= \lambda^2 (n-1)^2 E(Z^{-2}) - \lambda^2 \\
&= \lambda^2 (n-1)^2 \int_0^\infty z^{-2} \frac{z^{n-1}e^{-z}}{(n-1)!} dz - \lambda^2 \\
&= \frac{\lambda^2 (n-1)^2}{(n-1)(n-2)} \int_0^\infty \frac{z^{n-3}e^{-z}}{(n-3)!} dz - \lambda^2 \\
&= \frac{\lambda^2 (n-1)}{(n-2)} - \lambda^2 \\
&= \frac{\lambda^2}{n-2}
\end{aligned}
$$

Por lo tanto el $ECM(\hat{\lambda}) = \frac{\lambda^2}{n-2} \stackrel{n \to \infty}{\longrightarrow} 0$.

EJERCICIO 12

Sean $Y_1, ..., Y_n \stackrel{iid}{\sim} U(0, \theta)$. Sea $T = Max(Y_1, ..., Y_n)$ y considere los estimadores de θ de la forma $cT, \quad c \geq 0$.

(a) ¿Para qué valor de c, cT es insesgado?

(b) ¿Para qué valor de c, el $ECM(cT)$ es mínimo?

SOLUCIÓN

Si T corresponde al Máximo, entonces su función densidad es de la forma $f_T(t) = n[F_Y(t)]^{n-1} f_Y(t)$, donde $f_Y(t) = \frac{1}{\theta}$ y $F_Y(t) = \frac{t}{\theta}$.

(a) Calcularemos la esperanza para determinar el insesgamiento.

$$E(cT) = cE(T)$$

$$= c\int_0^\theta tn\frac{t^{n-1}}{\theta^{n-1}}\frac{1}{\theta}dt$$

$$= \frac{cn}{\theta^n}\int_0^\theta t^n dt = \frac{cn}{\theta^n}\frac{t^{n+1}}{n+1}\Big|_0^\theta$$

$$= \frac{cn}{\theta^n}\frac{\theta^{n+1}}{n+1} = \frac{cn\theta}{n+1}$$

Luego si $c = \frac{n+1}{n}$, cT es insesgado.

(b) En primer lugar calcularemos lo necesario para obtener el ECM y así después encontrar el c que lo minimice.

$$E(\hat{\theta}^2) = E((cT)^2) = c^2E(T^2)$$

$$= c^2\int_0^\theta t^2 n\frac{t^{n-1}}{\theta^{n-1}}\frac{1}{\theta}dt = \frac{c^2n}{\theta^n}\int_0^\theta t^{n+1}dt$$

$$= \frac{c^2n}{\theta^n}\left(\frac{t^{n+2}}{n+2}\Big|_0^\theta\right) = \frac{c^2n\theta^2}{n+2}$$

Ahora calculemos $Var(\hat{\theta})$:

$$Var(\hat{\theta}) = \frac{c^2n\theta^2}{n+2} - \frac{c^2n^2\theta^2}{(n+1)^2}$$

$$= c^2n\theta^2\left(\frac{1}{n+2} - \frac{n}{(n+1)^2}\right)$$

$$= c^2n\theta^2\left(\frac{(n+1)^2 - n(n+2)}{(n+1)^2(n+2)}\right)$$

$$= c^2n\theta^2\left(\frac{1}{(n+1)^2(n+2)}\right)$$

Además el sesgo queda de la siguiente forma:

$$Sesgo(\hat{\theta}) = \frac{cn\theta}{n+1} - \theta = \theta\left(\frac{cn-n-1}{n+1}\right)$$

y así obtenemos el ECM, resultando:

$$ECM(\hat{\theta}) = \frac{c^2 n\theta^2}{(n+1)^2(n+2)} + \theta^2 \left(\frac{(cn-n-1)^2}{(n+1)^2} \right)$$

Ahora utilizando los métodos matemáticos (1^a Derivada) para minimizar, encontraremos el c correspondiente.

$$\frac{\partial ECM(\hat{\theta})}{\partial c} = \frac{2cn\theta^2}{(n+1)^2(n+2)} + \frac{2\theta^2 n(cn-n-1)}{(n+1)^2} = 0$$

$$\longrightarrow c = \frac{n+2}{n+1}$$

Para verificar si realmente es mínimo, se calcula la segunda derivada.

$$\frac{\partial^2 ECM(\hat{\theta})}{\partial c^2} = \frac{2n\theta^2}{(n+1)^2(n+2)} + \frac{2n^2\theta^2}{(n+1)^2}$$

la cual es positiva $\forall n > 0$, luego cuando $c = \frac{n+2}{n+1}$, el $ECM(\hat{\theta})$ se minimiza.

EJERCICIO 13

Suponga que X sigue una distribución de Pareto, su función de densidad está dada por:

$$f(x;\alpha,\theta) = \theta\alpha^\theta x^{-\theta-1}, x \geq \alpha \quad y \quad \theta \geq 1$$

Asuma que $\alpha > 0$ es conocido y que $X_1, ..., X_n$ son v.a. iid.

(a) Encuentre un estimador de momentos para θ.

(b) Determine el EMV de θ.

SOLUCIÓN

Como los X_i siguen distribución de Pareto, se tiene que su esperanza y varianza son conocidas:

$$E(X) = \frac{\theta\alpha}{\theta-1}, \quad \theta > 1 \qquad Var(X) = \frac{\theta\alpha^2}{(\theta-1)^2(\theta-2)}, \quad \theta > 2$$

(a) Igualando el momento poblacional con el muestral, se obtiene el estimador de momentos:

$$\frac{\theta\alpha}{\theta-1}=\frac{1}{n}\sum_{i=1}^{n}x_i$$

$$\frac{\theta\alpha}{\theta-1}=\overline{x}$$

$$\theta\alpha-\theta\overline{X}=-\overline{x}$$

$$\hat{\theta}=\frac{\overline{x}}{\overline{x}-\alpha}$$

(b) Teniendo que las observaciones distribuyen Pareto, la función de verosimilitud es la siguiente:

$$\begin{aligned}
L(x;\alpha,\theta) &= \prod_{i=1}^{n}\theta\alpha^{\theta}x_i^{-(\theta+1)} \\
&= \theta^{n}\alpha^{n\theta}\left(\prod_{i=1}^{n}x_i\right)^{-(\theta+1)} && \backslash \ln \\
\ell(x;\alpha,\theta) &= n\ln(\theta)+n\theta\ln(\alpha)-(\theta+1)\sum_{i=1}^{n}\ln(x_i) && \backslash \partial\theta \\
\tfrac{\partial\ell}{\partial\theta} &= \tfrac{n}{\theta}+n\ln(\alpha)-\sum_{i=1}^{n}\ln(x_i)=0 \\
\rightarrow\hat{\theta} &= \frac{-n}{n\ln(\alpha)-\sum_{i=1}^{n}\ln(x_i)}
\end{aligned}$$

EJERCICIO 14

Sea $Y_1,...,Y_n$ una muestra aleatoria proveniente de una población $N(\theta,\theta)$, con $\theta>0$ y desconocido. A partir de una muestra aleatoria correspondiente a 25 pesos de circuitos, con $\sum_{i=1}^{n}Y_i=1264$ y con $\sum_{i=1}^{n}Y_i^2=5240$, determine la estimación máximo verosímil de θ.

SOLUCIÓN

$$
\begin{aligned}
L(y;\theta) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta}} \exp\left\{ \frac{1}{2\theta}(y_i-\theta)^2 \right\} \\
&= (2\pi\theta)^{-\frac{n}{2}} \exp\left\{ -\tfrac{1}{2\theta}\sum_{i=1}^{n}(y_i-\theta)^2 \right\} & \backslash \ln \\
\ell(y;\theta) &= -\tfrac{n}{2}\ln(2\pi) - \tfrac{n}{2}\ln(\theta) - \tfrac{1}{2\theta}\sum_{i=1}^{n}(y_i-\theta)^2 \\
&= -\tfrac{n}{2}\ln(2\pi) - \tfrac{n}{2}\ln(\theta) - \tfrac{1}{2\theta}\sum_{i=1}^{n}(y_i^2-2\theta y_i+\theta^2) \\
&= -\tfrac{n}{2}\ln(2\pi) - \tfrac{n}{2}\ln(\theta) - \tfrac{1}{2\theta}\sum_{i=1}^{n}y_i^2 + \sum_{i=1}^{n}y_i - \frac{n\theta}{2} & \backslash \partial\theta \\
\tfrac{\partial\ell}{\partial\theta} &= -\tfrac{n}{2\theta} + \tfrac{1}{2\theta^2}\sum_{i=i}^{n}y_i^2 - \frac{n}{2} = 0 \\
\longrightarrow & \quad -\tfrac{n}{\theta} + \tfrac{1}{\theta^2}\sum_{i=1}^{n}y_i^2 - n = 0 & \backslash \cdot \theta^2 \\
& \quad -n\theta + \sum_{i=1}^{n}y_i^2 - n\theta^2 = 0 & \backslash \cdot -\tfrac{1}{n} \\
& \quad \theta - \sum_{i=1}^{n}y_i^2/n + \theta^2 = 0 \\
& \quad \theta^2 + \theta - \overline{y^2} = 0 \\
& \quad \hat{\theta} = \tfrac{-1\pm\sqrt{1+4\overline{y^2}}}{2} \\
& \quad \hat{\theta}_1 = 13{,}98 \\
& \quad \hat{\theta}_2 = -14{,}98
\end{aligned}
$$

y como en el enunciado nos dicen que $\theta > 0$, nos quedamos con $\hat{\theta}_1 = 13{,}98$.

EJERCICIO 15

Ingenieros eléctricos japoneses han inventado un sistema de radar llamado detector de blancos móviles (MTD, moving target detector), diseñado para rechazar los ecos parásitos provocados por el terreno, la lluvia, las aves y otras fuentes de interferencia. Los investigadores han demostrado que la magnitud X de la frecuencia Doppler de una señal recibida por radar se puede modelar por una distribución Weibull, con parámetro $\alpha = 2$ y $\beta > 0$, tal que:

$$f(x) = \frac{2x}{\beta} \exp\left\{-\frac{1}{\beta}x^2\right\} \qquad x > 0$$

En base a una muestra aleatoria de tamaño n, determine el estimador máximo verosímil de β y obtenga su estimación con las siguientes magnitudes de frecuencias si $\sum_{i=1}^{50} x_i^2 = 51{,}9$.

SOLUCIÓN

$$\begin{aligned}
L(x;\beta) &= \prod_{i=1}^{n} \frac{2x_i}{\beta} \exp\left\{-\frac{1}{\beta}x_i^2\right\} \\
&= \tfrac{2^n}{\beta^n} \left(\prod_{i=1}^{n} x_i\right) \exp\left\{-\tfrac{1}{\beta} \sum_{i=1}^{n} x_i^2\right\} \qquad \backslash \ln \\
\ell(x;\beta) &= n\ln(2) - n\ln(\beta) + \sum_{i=1}^{n} \ln(x_i) - \frac{1}{\beta}\sum_{i=1}^{n} x_i^2 \quad \backslash \partial\beta \\
\tfrac{\partial \ell}{\partial \beta} &= -\tfrac{n}{\beta} + \frac{\sum\limits_{i=1}^{n} x_i^2}{\beta^2} = 0 \\
\rightarrow & \frac{\sum\limits_{i=1}^{n} x_i^2}{n} = \frac{\beta^2}{\beta} \\
\hat{\beta} &= \frac{\sum\limits_{i=1}^{n} x_i^2}{n}
\end{aligned}$$

Reemplazando por los valores dados en el inicio, queda que

$$\hat{\beta} = 1{,}038$$

EJERCICIO 16

En una fábrica se seleccionan diariamente motores y se inspeccionan hasta encontrar el primer motor defectuoso. Sea $(X_1, \ldots, X_n)$ una m.a. de X distribuida geométricamente con p desconocido.

(a) Determine el estimador de momentos para p.

(b) Determine el estimador máximo verosímil de p.

(c) De los registros de 100 días se obtuvo la siguiente información del número de motores inspeccionados.

Nde motores inspeccionados	1	2	3	4	5
Nde días	8	10	15	25	42

Estime la probabilidad de que en un día cualquiera se deban inspeccionar más de dos motores para encontrar uno defectuoso.

SOLUCIÓN

Dado que $X_i \sim Geom(p)$ tenemos que:

$$P(X = x) = p(1-p)^{x-1}; \qquad E(X) = \frac{1}{p} \qquad Var(X) = \frac{1}{p^2}$$

(a) Igualando momento poblacional con el muestral, queda:

$$\frac{1}{p} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\hat{p} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\overline{x}}$$

(b) Haciendo el procedimiento usual tenemos lo siguiente:

$$
\begin{aligned}
L(x;p) &= \prod_{i=1}^{n} p(1-p)^{x_i-1} \\
&= p^n (1-p)^{\sum (x_i-1)} \\
&= p^n (1-p)^{\sum x_i - n} \qquad\qquad \backslash \ln \\
\ell(x;p) &= n\ln(p) + \left(\sum_{i=1}^{n} x_i - n\right)\ln(1-p) \\
&= n\ln(p) + \sum_{i=1}^{n} x_i \ln(1-p) - n\ln(1-p) \qquad\qquad \backslash \partial p
\end{aligned}
$$

$$
\frac{\partial \ell}{\partial p} = \frac{n}{p} - \frac{\displaystyle\sum_{i=1}^{n} x_i}{1-p} + \frac{n}{1-p} = 0
$$

$$
\frac{n}{p} = \frac{\displaystyle\sum_{i=1}^{n} x_i - n}{1-p}
$$

$$
n(1-p) = p\left(\sum_{i=1}^{n} x_i - n\right)
$$

$$
\hat{p} = \frac{n}{\displaystyle\sum_{i=1}^{n} x_i} = 0{,}2610
$$

(c) Recordando la propiedad de invarianza que tienen los estimadores máximo verosímiles, lo que se pide, se puede traducir estadísticamente en:

$$
\begin{aligned}
P(X > 2) &= 1 - P(X \leq 2) \\
&= 1 - P(X = 1) - P(X = 2) \\
&= 1 - \hat{p}(1-\hat{p})^{1-1} - \hat{p}(1-\hat{p})^{2-1} = 1 - \hat{p} - \hat{p}(1-\hat{p}) \\
&= (1-\hat{p})^2 = 0{,}5459
\end{aligned}
$$

EJERCICIO 17

Sean $X_1, \ldots, X_n \overset{iid}{\sim} U(\theta_1, \theta_2)$. Es decir, la densidad de X_i es:

$$f(x) = \frac{1}{\theta_2 - \theta_1} \qquad \theta_1 \leq x \leq \theta_2$$

(a) Encuentre el estimador de momentos para los parámetros de esta distribución.

(b) Encuentre el estimador máximo verosímil para θ_1 y θ_2.

SOLUCIÓN

(a) Se necesita encontrar los estimadores de θ_1 y θ_2, luego por ser dos parámetros, utilizaremos el primer y segundo momento para armar un sistema de ecuaciones.
Momentos poblacionales:

$$E(X) = \frac{\theta_1 + \theta_2}{2}$$

$$
\begin{aligned}
E(X^2) &= \int_{\theta_1}^{\theta_2} x^2 \frac{1}{\theta_2 - \theta_1} dx \\
&= \frac{\theta_2^2 + \theta_2\theta_1 + \theta_1^2}{3}
\end{aligned}
$$

Igualando momentos poblacionales con muestrales, queda el siguiente sistema de ecuaciones:

$$\frac{\theta_1+\theta_2}{2}=\overline{x} \qquad (1)$$

$$\frac{\theta_2^2+\theta_2\theta_1+\theta_1^2}{3}=\overline{x^2} \qquad (2)$$

Despejando tenemos de (1) que $\theta_1 = 2\overline{x} - \theta_2$ y reemplazando en la segunda ecuación se obtiene:

$$(2\overline{x}-\theta_2)^2+(2\overline{x}-\theta_2)\theta_2+\theta_2^2=3\overline{x^2}$$

$$4\overline{x}^2-4\overline{x}\theta_2+\theta_2^2+2\overline{x}\theta_2-\theta_2^2+\theta_2^2=3\overline{x^2}$$

$$\theta_2^2-2\overline{x}\theta_2+4\overline{x}^2-3\overline{x^2}=0$$

Resolviendo esta ecuación de segundo grado con los métodos usuales, se obtiene:

$$\hat{\theta}_2=\frac{2\overline{x}\pm\sqrt{4\overline{x}^2-16\overline{x}^2+12\overline{x^2}}}{2}=\overline{x}\pm\sqrt{3\overline{x^2}-3\overline{x}^2}=\overline{x}\pm\sqrt{3S^2}$$

Luego reemplazando $\hat{\theta}_2$ en (1) se tiene que $\hat{\theta}_1 = 2\overline{x} - \hat{\theta}_2 = 2\overline{x} - \overline{x} \mp \sqrt{3S^2} = \overline{x} \mp \sqrt{3S^2}$.

(b) En el caso de aquellas distribuciones en que su dominio depende de los parámetros a estimar (en este caso la distribución es válida cuando $\theta_1 \leq x \leq \theta_2$), el procedimiento de estimación debe considerar un muy pequeño detalle como se muestra a continuación:

$$L(x;\theta_1,\theta_2)=\frac{1}{(\theta_2-\theta_1)^n}\mathbb{I}_{(\theta_2-\theta_1)}(x_1)\ldots\mathbb{I}_{(\theta_2-\theta_1)}(x_n)$$

$$=\frac{1}{(\theta_2-\theta_1)^n}\prod_{i=1}^{n}\mathbb{I}_{(\theta_2-\theta_1)}(x_i)$$

$$=\frac{1}{(\theta_2-\theta_1)^n}\prod_{i=1}^{n}\mathbb{I}_{(x_i>\theta_1)}(x_i)\mathbb{I}_{(x_i<\theta_2)}(x_i)$$

$$=\frac{1}{(\theta_2-\theta_1)^n}\prod_{i=1}^{n}\mathbb{I}_{(min(x_i)>\theta_1)}(x_i)\mathbb{I}_{(max(x_i)<\theta_2)}(x_i)$$

Como se muestra a continuación tanto el mínimo como el máximo de los x_i son el estimador máximo verosímil de θ_1 y θ_2 respectivamente.

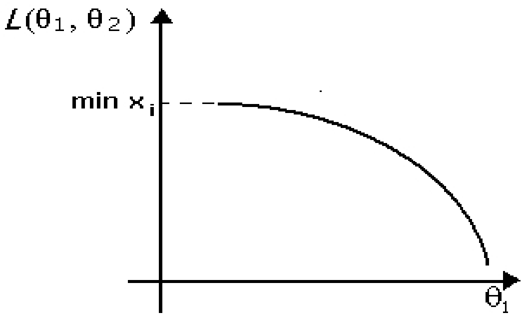

Figura 1.1: Estimación Theta 1

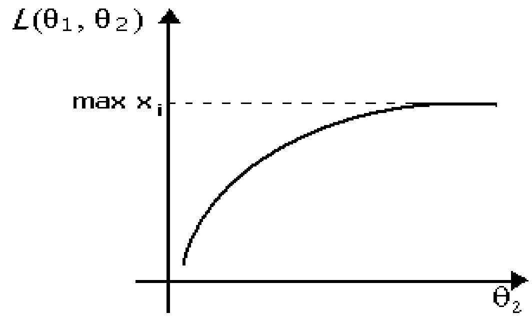

Figura 1.2: Estimación Theta 2

EJERCICIO 18

Sea $X_1, \ldots, X_n$ una m.a. con media μ y varianza σ^2. Se propone estimar μ mediante

$$\overline{X} = \sum_{i=1}^{n} c_i X_i$$

donde los c_i son números fijos. Determine los c_i tal que el estimador sea insesgado y de varianza mínima.

<u>SOLUCIÓN</u>

Se tiene que $\hat{\mu} = \overline{X} = \sum_{i=1}^{n} c_i X_i$. Se determinará inicialmente los c_i para que se cumpla el insesgamiento.

$$E(\hat{\mu}) = E\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i=1}^{n} c_i E(X_i) = \mu \sum_{i=1}^{n} c_i$$

luego para que $\hat{\mu}$ sea insesgado, debe cumplirse que $\displaystyle\sum_{i=1}^{n} c_i = 1$

La varianza es la siguiente:

$$Var(\hat{\mu}) = Var\left(\sum_{i=1}^{n} c_i X_i\right) \stackrel{m.a.}{=} \sum_{i=1}^{n} c_i^2 Var(X_i) = \sum_{i=1}^{n} c_i^2 \sigma^2 = \sigma^2 \sum_{i=1}^{n} c_i^2$$

Para determinar que sea de varianza mínima, se ocuparán procedimientos de cálculo (multiplicadores de Lagrange) para que se cumpla el insesgamiento y la variabilidad mínima simultáneamente.

$$Ł = \min_{c_i} \sigma^2 \sum_{i=1}^{n} c_i^2 - \lambda \left(\sum_{i=1}^{n} c_i - 1 \right)$$

$$\frac{\partial Ł}{\partial c_i} = \sigma^2 2c_i - \lambda = 0 \qquad (1)$$

$$\frac{\partial Ł}{\partial \lambda} = \sum_{i=1}^{n} c_i - 1 = 0 \qquad (2)$$

Para poder ocupar la condición dada en (2), sumaremos las n ecuaciones que se obtienen al derivar con respecto a $c_i, \quad i = 1, \ldots, n$, resultando la siguiente ecuacion:

$$2\sigma^2 \sum_{i=1}^{n} c_i - n\lambda = 0 \qquad (3)$$

Luego despejando (2) y reemplazando en (3), queda lo siguiente:

$$2\sigma^2 = n\lambda$$

$$\lambda = \frac{2\sigma^2}{n} \qquad (4)$$

Finalmente, reemplazando (4) en (1) resulta:

$$2\sigma^2 c_i = \frac{2\sigma^2}{n}$$

$$c_i = \frac{1}{n}$$

Luego $\hat{\mu} = \overline{X}$ es el estimador óptimo.

EJERCICIO 19

Sean $X_1, \ldots, X_n$ una m.a. $N(0, \sigma^2)$.

(a) Encuentre el EMV de σ^2.

(b) Sea $\tilde{\sigma^2} = \frac{X_1^2 + X_n^2}{2}$ otro estimador de σ^2. ¿Son insesgados ambos estimadores?

(c) ¿Cuál de los dos es mejor?. Justifique.

SOLUCIÓN

(a) Usando el procedimiento usual, tenemos:

$$L(x;\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{x_i^2}{2\sigma^2}\right\}$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{\sum x_i^2}{2\sigma^2}\right\} \qquad \backslash \ln$$

$$\ell(x;\sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{\sum x_i^2}{2\sigma^2} \qquad \backslash \partial\sigma^2$$

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{2\pi\sigma^2}2\pi + \frac{\sum x_i^2}{2(\sigma^2)^2} = 0$$

Luego simplificando y despejando se obtiene que:

$$\hat{\sigma^2} = \frac{\sum_{i=1}^{n} x_i^2}{n}$$

(b)

$$E(\hat{\sigma}^2) = E\left(\frac{\sum_{i=1}^{n} x_i^2}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i^2)$$

$$= \frac{1}{n}\sum_{i=1}^{n}(Var(x_i) + E^2(x_i)) = \frac{1}{n}\sum_{i=1}^{n}(\sigma^2 + 0)$$

$$= \sigma^2$$

Por lo tanto el EMV es insesgado. Veremos ahora el propuesto.

$$E(\tilde{\sigma}^2) = E\left(\frac{x_1^2 + x_n^2}{2}\right) = \frac{E(x_1^2) + E(x_n^2)}{2}$$

$$= \frac{[Var(x_1) + E^2(x_1)] + [Var(x_n) + E^2(x_n)]}{2} = \frac{[\sigma^2 + 0] + [\sigma^2 + 0]}{2}$$

$$= \sigma^2$$

Luego el estimador propuesto $\tilde{\sigma^2}$ también es insesgado.

(c) Para determinar cual de ellos es mejor, se calcula el ECM que en este caso, por ser ambos insesgados, se reduce a el cálculo de sus varianzas.

$$Var(\hat{\sigma}^2) = Var\left(\frac{\sum_{i=1}^{n} x_i^2}{n}\right) \overset{m.a.}{=} \frac{1}{n^2}\sum_{i=1}^{n} Var(x_i^2)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}(E(x_i^4) - E^2(x_i^2))$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}(3\sigma^4 - \sigma^4)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} 2\sigma^4$$

$$= \frac{2\sigma^4}{n}$$

$$Var(\tilde{\sigma}^2) = Var\left(\frac{x_1^2 + x_n^2}{2}\right) \overset{ind}{=} \frac{1}{4}(Var(x_1^2) + Var(x_n^2))$$

$$= \frac{1}{4}(2\sigma^4 + 2\sigma^4)$$

$$= \sigma^4$$

Como $\frac{2\sigma^4}{n} < \sigma^4$ para $n > 2$, $\hat{\sigma}^2$ es mejor estimador que $\tilde{\sigma^2}$.

Nota: Observe que si $n = 2$ los estimadores son igualmente eficientes.

EJERCICIO 20

Sean X_1, X_2 ind. con $X_i \sim P(\lambda)$. Se desea estimar $P(X_1 = 0) = e^{-\lambda}$.

(a) Encuentre el EMV de $P(X_1 = 0)$.

(b) Encuentre ECM de:

  i. $\left(\frac{1}{2}\right)^{X_1+X_2}$

  ii. Proporción de $X_i = 0$ (puede ser 0, 0.5 ó 1)

(c) Demuestre que i. tiene menor ECM para todo $\lambda > 0$

(d) Muestre que el estimador $T = \left(\frac{1}{2}\right)^{X_1+X_2}$ no alcanza la cota de cramer rao.

SOLUCIÓN

(a) Para esto primero se debe sacar el EMV de λ y después usar la propiedad de invarianza.

$$L(\mathbf{x};\lambda) = \prod_{i=1}^{2} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

$$= \frac{e^{-2\lambda}\lambda^{x_1+x_2}}{x_1!x_2!} \qquad \backslash \ln$$

$$\ell(\mathbf{x};\lambda) = -2\lambda + (x_1 + x_2)\ln(\lambda) - \ln(x_1!) - \ln(x_2!) \qquad \backslash \partial\lambda$$

$$\frac{\partial \ell}{\partial \lambda} = -2 + \frac{x_1 + x_2}{\lambda} = 0$$

Despejando se obtiene que:

$$\hat{\lambda} = \frac{x_1 + x_2}{2} = \overline{x}$$

Por lo tanto

$$\widehat{P(X = 0)} = \frac{e^{-\hat{\lambda}}\hat{\lambda}^0}{0!} = e^{-\hat{\lambda}} = e^{-\overline{x}} = e^{-\frac{x_1+x_2}{2}}$$

(b-i.) Denominaremos $\hat{\theta}_1 = \left(\frac{1}{2}\right)^Z$, donde $X_1 + X_2 = Z \sim P(2\lambda)$.

En primer lugar calculamos el sesgo.

$$
\begin{aligned}
E\left[\left(\frac{1}{2}\right)^Z\right] &= \sum_{z=0}^{\infty}\left(\frac{1}{2}\right)^z P(Z=z) \\
&= \sum_{z=0}^{\infty}\left(\frac{1}{2}\right)^z \frac{(2\lambda)^z e^{-2\lambda}}{z!} \\
&= e^{-2\lambda}\underbrace{\sum_{z=0}^{\infty}\frac{\lambda^z}{z!}}_{e^{\lambda}\ \longleftarrow\ \text{por Taylor}} \\
&= e^{-2\lambda}e^{\lambda} \\
&= e^{-\lambda}
\end{aligned}
$$

Por lo tanto es un estimador insesgado para $P(X_1 = 0)$, luego el ECM se reduce a la Varianza.

$$
\begin{aligned}
ECM(\hat{\theta}) &= Var\left(\left[\frac{1}{2}\right]^Z\right) \\
&= E\left(\left[\frac{1}{2}\right]^{2Z}\right) - E^2\left(\left[\frac{1}{2}\right]^Z\right) \\
&= \sum_{z=0}^{\infty}\left(\frac{1}{2}\right)^{2z}\frac{(2\lambda)^z e^{-2\lambda}}{z!} - e^{-2\lambda} \\
&= \sum_{z=0}^{\infty}\left(\frac{1}{4}\right)^{z}\frac{2^z\lambda^z e^{-2\lambda}}{z!} - e^{-2\lambda}
\end{aligned}
$$

$$= e^{-2\lambda} \underbrace{\sum_{z=0}^{\infty} \left(\frac{\lambda}{2}\right)^z \frac{1}{z!}}_{e^{\lambda/2}} - e^{-2\lambda}$$

$$= e^{-2\lambda} \cdot e^{\lambda/2} - e^{-2\lambda}$$

$$= e^{-3\lambda/2} - e^{-2\lambda}$$

(b-ii.) Por otra parte definamos $\hat{\theta}_2 = \frac{\sum_{i=1}^{2} \mathbb{I}_{(X_i=0)}}{2}$ que precisamente representa la proporción de $X_i = 0$, entonces $2\hat{\theta}_2 \sim Binomial(2, e^{-\lambda})$ y es claro que $E(2\hat{\theta}_2) = 2e^{-\lambda}$ por lo que $\hat{\theta}_2$ es insesgado, luego el ECM se reduce a al cálculo de la varianza.

$$ECM(\hat{\theta}_2) = Var(\hat{\theta}_2) = \frac{1}{4} Var(2\hat{\theta}_2)$$

$$= \frac{1}{4} 2e^{-\lambda}(1 - e^{-\lambda})$$

$$= \frac{1}{2} e^{-\lambda}(1 - e^{-\lambda})$$

(c) Para tal demostración, ocuparemos la siguiente tautología:

$$0 < (e^{-\frac{1}{2}\lambda} - e^{-\lambda})^2$$

$$< e^{-\lambda} - 2e^{-\frac{3}{2}\lambda} + e^{-2\lambda}$$

$$< \frac{1}{2} e^{-\lambda} - e^{-\frac{3}{2}\lambda} + \frac{1}{2} e^{-2\lambda}$$

$$= \frac{1}{2} e^{-\lambda} - e^{-\frac{3}{2}\lambda} + e^{-2\lambda} - \frac{1}{2} e^{-2\lambda}$$

Luego despejando resulta que $e^{-\frac{3}{2}\lambda} - e^{-2\lambda} < \frac{1}{2} e^{-\lambda}(1 - e^{-\lambda})$.
Por lo tanto $ECM(\hat{\theta}_1) < ECM(\hat{\theta}_2) \quad \forall \lambda > 0$.

(d) Recordemos que la $CCR = \frac{(h(\theta)')^2}{I(\theta)}$ donde $I(\theta) = -E\left(\frac{\partial^2}{\partial\theta^2}\ell(\mathbf{x})\right)$.

De la parte (b-i) tenemos que la $Var(T)$ es

$$Var(T) = Var\left(\left[\frac{1}{2}\right]^{x_1+x_2}\right) = e^{-\frac{3}{2}\lambda} - e^{-2\lambda}$$

Ahora necesitamos:

$$h(\theta) = e^{-\lambda} \qquad \backslash\partial\lambda$$

$$h(\theta)' = -e^{-\lambda} \qquad (\,)^2$$

$$(h(\theta)')^2 = e^{-2\lambda}$$

$$I(\theta) = -E\left(\frac{\partial^2}{\partial\lambda^2}\ell(\mathbf{x})\right)$$

$$= -E\left(-\frac{x_1+x_2}{\lambda^2}\right)$$

$$= \frac{2\lambda}{\lambda^2}$$

$$= \frac{2}{\lambda}$$

Luego se obtiene que

$$CCR = \frac{e^{-2\lambda}}{2/\lambda} = \frac{\lambda e^{-2\lambda}}{2} \neq Var(T)$$

es decir, T no alcanza la CCR.

EJERCICIO 21

Sean $X_1, \ldots, X_n$ v.a. iid con función de densidad dada por:

$$f(x) = \frac{\beta}{x^{\beta+1}}, \quad x > 1, \beta > 1$$

(a) Determine el EMV de β.

(b) Encuentre el EMV de $e^{\frac{1}{\beta}}$.

SOLUCIÓN

(a)

$$L(x;\beta) = \prod_{i=1}^{n} \frac{\beta}{x_i^{\beta+1}}$$

$$= \frac{\beta^n}{\left(\prod_{i=1}^{n} x_i\right)^{\beta+1}} \qquad \backslash \ln$$

$$\ell(x;\beta) = n\ln(\beta) - (\beta+1)\sum_{i=1}^{n}\ln(x_i) \qquad \backslash \partial\beta$$

$$\frac{\partial \ell}{\partial \beta} = \frac{n}{\beta} - \sum_{i=i}^{n}\ln(x_i) = 0$$

Despejando se obtiene que

$$\hat{\beta} = \frac{n}{\sum_{i=1}^{n}\ln x_i}$$

(b) Utilizando la propiedad de invarianza que tienen los EMV, obtenemos que:

$$\widehat{e^{\frac{1}{\beta}}} = e^{\frac{1}{\hat{\beta}}} = e^{\frac{\sum \ln x_i}{n}}$$

EJERCICIO 22

Sean $X_1, \ldots, X_n \overset{iid}{\sim} U(0,\theta)$. Sean c y d dos constantes positivas y considere los estimadores de θ:

$$\widetilde{\theta}_c = c\overline{X} \qquad \widetilde{\theta}_d = dX_{max}$$

(a) Calcule los errores cuadráticos medios de $\widetilde{\theta}_c$ y $\widetilde{\theta}_d$ en función de c, d y θ.

(b) Encuentre los valores de $\hat{c}$ y $\hat{d}$ de c y d, que minimizan los errores cuadráticos medios.

SOLUCIÓN

(a) Tenemos que $ECM(\widetilde{\theta}_c) = Var(\widetilde{\theta}_c) + Sesgo^2(\widetilde{\theta}_c)$.

$$\begin{aligned}
Var(\widetilde{\theta}_c) &= Var(c\overline{X}) = c^2 Var(\overline{X}) \\
&= \frac{c^2}{n^2}\sum_{i=1}^{n} Var(X_i) \\
&= \frac{c^2}{n^2} n \frac{\theta^2}{12} \\
&= \frac{c^2\theta^2}{12n}
\end{aligned}$$

$$\begin{aligned}
Sesgo(\widetilde{\theta}_c) &= E(\widetilde{\theta}_c) - \theta \\
&= cE(\overline{X}) - \theta \\
&= \frac{c}{n}\sum_{i=1}^{n} E(X_i) - \theta \\
&= c\frac{\theta}{2} - \theta \\
&= \theta\frac{c-2}{2}
\end{aligned}$$

Por lo tanto, queda que

$$ECM(\widetilde{\theta}_c) = \frac{c^2\theta^2}{12n} + \theta^2\left(\frac{c-2}{2}\right)^2$$

Tenemos que $ECM(\widetilde{\theta}_d) = Var(\widetilde{\theta}_d) + Sesgo^2(\widetilde{\theta}_d)$.

$$Var(\widetilde{\theta}_d) = d^2 Var(max\{X_i\})$$

Por lo que recordemos que la distribución del máximo(Z) de los X_i es:

$$f_Z(z) = n[F_X(z)]^{n-1} f_X(z)$$

Luego en este caso, haciendo $Z = max\{X_i\}$, la distribución queda de la siguiente forma

$$f_Z(z) = n\left(\frac{z}{\theta}\right)^{n-1}\frac{1}{\theta} = n\frac{z^{n-1}}{\theta^n}, \quad 0 < z < \theta$$

Ahora se calculan las esperanzas correspondientes para obtener la varianza necesaria para el ECM.

$$E(Z) = \int_0^{\theta} znz^{n-1}\theta^{-n}dz$$

$$= n\theta^{-n}\int_0^{\theta} z^n dz$$

$$= n\theta^{-n}\frac{z^{n+1}}{n+1}\bigg|_0^{\theta}$$

$$= n\frac{\theta^{-n}\theta^{n+1}}{n+1}$$

$$= \frac{n}{n+1}\theta$$

$$
\begin{aligned}
E(Z^2) &= \int_0^{\theta} z^2 n z^{n-1} \theta^{-n} dz \\
&= n\theta^{-n} \int_0^{\theta} z^{n+1} dz \\
&= n\theta^{-n} \frac{z^{n+2}}{n+2} \bigg|_0^{\theta} \\
&= n\theta^{-n} \frac{\theta^{n+2}}{n+2} \\
&= \frac{n}{n+2}\theta^2
\end{aligned}
$$

$$
\begin{aligned}
\longrightarrow Var(\widetilde{\theta}_d) &= d^2 \left[E(Z^2) - E^2(Z)\right] \\
&= d^2 \left[\frac{n}{n+2}\theta^2 - \frac{n^2}{(n+1)^2}\theta^2\right] \\
&= d^2\theta^2 \left[\frac{n}{n+2} - \frac{n^2}{(n+1)^2}\right]
\end{aligned}
$$

$$
\begin{aligned}
Sesgo(\widetilde{\theta}_d) &= E(\widetilde{\theta}_d) - \theta \\
&= dE(max\{X_i\}) - \theta \\
&= d\frac{n}{n+1}\theta - \theta \\
&= \theta \left(d\frac{n}{n+1} - 1\right)
\end{aligned}
$$

Por lo tanto, queda que

$$ECM(\widetilde{\theta}_d) = d^2\theta^2 \left[\frac{n}{n+2} - \frac{n^2}{(n+1)^2}\right] + \theta^2 \left(d\frac{n}{n+1} - 1\right)^2$$

(b) Para esto utilizaremos el método de la derivada

$$\frac{\partial}{\partial c} ECM(\widetilde{\theta}_c) = 2c\frac{\theta^2}{12n} + \frac{\theta^2 c}{2} - \theta^2 = 0$$

$$\longrightarrow 2c\frac{\theta^2}{12n} + \frac{c\theta^2}{2} - \theta^2 = 0$$

$$c\left(\frac{1}{6n}\theta^2 + \frac{\theta^2}{2}\right) = \theta^2$$

$$c\left(\frac{\theta^2 + 3n\theta^2}{6n}\right) = \theta^2$$

$$c = \frac{6n}{1+3n}$$

Verificando con el criterio de la segunda derivada, es fácil ver que ésta es positiva $\forall\, \theta > 0$, luego el valor de c encontrado minimiza el $ECM(\widetilde{\theta}_c)$.

$$\frac{\partial}{\partial d} ECM(\widetilde{\theta}_d) = 2\theta^2 d\left(\frac{n}{n+2} - \frac{n^2}{(n+1)^2}\right) + 2\theta^2\left(d\frac{n}{n+1} - 1\right)\frac{n}{n+1} = 0$$

$$\longrightarrow 2\theta^2 d\left(\frac{n}{n+2} - \frac{n^2}{(n+1)^2}\right) + 2\theta^2\left(d\frac{n^2}{(n+1)^2} - \frac{n}{n+1}\right) = 0$$

$$2\theta^2 d\left(\frac{n(n+1)^2 - n^2(n+2)}{(n+2)(n+1)^2}\right) + 2\theta^2 d\frac{n^2}{(n+1)^2} - 2\theta^2\frac{n}{n+1} = 0$$

$$\cancel{2\theta^2}\left(\frac{n(n+1)^2 - n^2(n+2) + n^2(n+2)}{(n+2)(n+1)^2}\right) = \cancel{2\theta^2}\frac{n}{n+1}$$

$$d = \frac{n}{n+1}\left(\frac{(n+2)(n+1)^2}{n(n+1)^2}\right)$$

$$d = \frac{n+2}{n+1}$$

Verificando con el criterio de la segunda derivada, es fácil ver que ésta es positiva $\forall\,\theta > 0$, luego el valor de d encontrado minimiza el $ECM(\widetilde{\theta}_d)$.

EJERCICIO 23

Sean $X_1, X_2, ..., X_n$ variables aleatorias con función de densidad dada por:

$$f(x;\alpha,\beta) = \frac{1}{2\beta}e^{\frac{-|x-\alpha|}{\beta}} \quad x,\alpha \in \mathbb{R} \quad y \quad \beta \geq 0$$

Determine un estimador insesgado y de varianza mínima para β

Hint: $E(X) = \alpha$, $V(X) = 2\beta^2$ y $\mid X - \alpha \mid \sim Exp(\frac{1}{\beta})$.

SOLUCIÓN

Para encontrar tal estimador, se utilizará el método del EMV.

$$L(x;\beta) = \prod_{i=1}^{n} \frac{1}{2\beta} e^{-\frac{|x_i-\alpha|}{\beta}}$$

$$= \frac{1}{(2\beta)^n} \exp\left\{-\frac{1}{\beta}\sum_{i=1}^{n}|x_i-\alpha|\right\} \qquad \backslash \ln$$

$$\ell(x;\beta) = -\frac{1}{\beta}\sum_{i=1}^{n}|x_i-\alpha| - n\ln(2\beta) \qquad \backslash\partial\beta$$

$$\frac{\partial\ell}{\partial\beta} = \frac{\sum_{i=1}^{n}|x_i-\alpha|}{\beta^2} - \frac{n}{\beta} = 0$$

$$\longrightarrow \hat{\beta} = \frac{\sum_{i=1}^{n}|x_i-\alpha|}{n}$$

Para determinar si es insesgado, se ocupa el Hint:

$$E(|X-\alpha|) = \beta \qquad Var(|X-\alpha|) = \beta^2$$

Por lo tanto resulta:

$$E\left(\frac{\sum_{i=1}^{n}|x_i-\alpha|}{n}\right) = \frac{1}{n}\sum_{i=1}^{n}E(|x_i-\alpha|)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\beta$$

$$= \beta \quad \therefore \text{Por lo tanto es Insesgado}$$

Recordemos que la $CCR = \frac{(h(\beta)')^2}{I(\beta)}$ donde $I(\beta) = -E\left(\frac{\partial^2}{\partial\beta^2}\ell(\mathbf{x})\right)$.

Calculamos la varianza, y veremos si alcanza la CCR, de ser así sería de varianza mínima.

$$Var\left(\frac{\sum_{i=1}^{n}|x_i-\alpha|}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n}Var(|x_i-\alpha|)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\beta^2$$

$$= \frac{\beta^2}{n}$$

Como el estimador es insesgado, solo queda por calcular $I(\beta)$

$$I(\beta) = -E\left(\frac{\partial^2}{\partial\beta^2}\ell(\mathbf{x})\right)$$

$$= -E\left(-2\frac{\sum_{i=1}^{n}|x_i-\alpha|}{\beta^3} + \frac{n}{\beta^2}\right)$$

$$= \frac{2}{\beta^3}E\left(\sum_{i=1}^{n}|x_i-\alpha|\right) - \frac{n}{\beta^2}$$

$$= \frac{2}{\beta^3}n\beta - \frac{n}{\beta^2}$$

$$= \frac{n}{\beta^2}$$

Por lo tanto $CCR = \frac{\beta^2}{n}$ la cual corresponde a la $Var(\hat{\beta})$. Por lo tanto es de varianza mínima.

EJERCICIO 24

Sea $X_1, \ldots, X_n$ iid $Ber(p)$.

(a) Demuestre utilizando dos formas distintas que $T = \sum_{i=1}^{n} x_i$ es un estadístico suficiente.

(b) Se quiere estimar $Var(X_1)$. Muestre que $W = X_1(1 - X_2)$ es un estimador insesgado de $Var(X_1)$.

(c) Encuentre un mejor estimador de W.

SOLUCIÓN

(a)

- Por Teorema de Factorización de Neyman.

Teorema 1 *Si existen funciones $g()$ y $h()$, tales que $f(\boldsymbol{x};\theta) = g(T(\boldsymbol{x}),\theta) \cdot h(x)$, entonces $T(\boldsymbol{x})$ es estadístico suficiente.*

Para este caso de la Bernoulli:

$$L(x;\theta) = \prod_{i=1}^{n} f(\mathbf{x};\theta)$$

$$= \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$$

$$= \underbrace{\theta^{\sum x_i}(1-\theta)^{n-\sum x_i}}_{g(T(\mathbf{x}),\theta)} \cdot \underbrace{1}_{h(\mathbf{x})}$$

$$\longrightarrow T(\mathbf{x}) = \sum_{i=1}^{n} x_i$$

- Por familia Exponencial

  Si $f(\mathbf{x},\theta) = \exp\{T(\mathbf{x})c(\theta) + d(\theta) + s(\mathbf{x})\}$ entonces $T(\mathbf{x})$ es estadístico suficiente.

  Para este caso de la Bernoulli:

$$
\begin{aligned}
L(x;\theta) &= \prod_{i=1}^{n} f(\mathbf{x};\theta) \\
&= \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} \\
&= \theta^{\sum x_i}(1-\theta)^{n-\sum x_i} \\
&= \left(\frac{\theta}{1-\theta}\right)^{\sum x_i}(1-\theta)^n \qquad \backslash \ln \qquad \backslash \exp \\
&= \exp\left\{\underbrace{\sum_{i=1}^{n} x_i}_{T(\mathbf{x})} \underbrace{\ln\left(\frac{\theta}{1-\theta}\right)}_{c(\theta)} + \underbrace{n\ln(1-\theta)}_{d(\theta)} + \underbrace{0}_{s(\mathbf{x})}\right\}
\end{aligned}
$$

  Por lo tanto $T(\mathbf{x}) = \sum_{i=1}^{n} x_i$ es estadístico suficiente.

(b) Tenemos que $W = X_1(1-X_2)$ y como $X_1 \sim Ber(p)$ entonces $Var(X_1) = p(1-p)$.

$$
\begin{aligned}
E(W) = E(X_1(1-X_2)) &= E(X_1) - E(X_1X_2) \\
&\overset{ind}{=} E(X_1) - E(X_1)E(X_2) \\
&= p - p^2 \\
&= p(1-p)
\end{aligned}
$$

Por lo tanto W es un estimador insesgado de la varianza de X_1.

(c) Usando el Teorema Rao - Blackwell, el cual dice:

Teorema 2 *Si $T(\boldsymbol{y})$ es un estimador insesgado de $h(\theta)$ y $S(\boldsymbol{y})$ es un estadístico suficiente de θ, entonces $I(\boldsymbol{y}) = E(T(\boldsymbol{y})|S(\boldsymbol{y}))$ es un estimador insesgado de $h(\theta)$ y cumple con $Var(I(\boldsymbol{y})) \leq Var(T(\boldsymbol{y}))$.*

tenemos lo siguiente:

$$\Phi(\mathbf{x}) = E\left(X_1(1-X_2)\middle|\sum_{i=1}^n x_i = u\right)$$

$$= 1 \cdot P\left(X_1(1-X_2) = 1\middle|\sum_{i=1}^n x_i = u\right) + 0 \cdot P\left(X_1(1-X_2) = 0\middle|\sum_{i=1}^n x_i = u\right)$$

$$= P\left(X_1(1-X_2) = 1\middle|\sum_{i=1}^n x_i = u\right)$$

$$= P\left(X_1 = 1, X_2 = 0\middle|\sum_{i=1}^n x_i = u\right)$$

$$= \frac{P\left(X_1 = 1, X_2 = 0, \sum_{i=1}^n x_i = u\right)}{P\left(\sum_{i=1}^n x_i = u\right)}$$

$$= \frac{P\left(X_1 = 1, X_2 = 0, \sum_{i=3}^n x_i = u - 1 - 0\right)}{P\left(\sum_{i=1}^n x_i = u\right)}$$

$$
\overset{ind}{=} \frac{P(X_1 = 1)P(X_2 = 0)P\left(\sum_{i=3}^{n} x_i = u - 1\right)}{P\left(\sum_{i=1}^{n} x_i = u\right)}
$$

$$
= \frac{p(1-p)\binom{n-2}{u-1}p^{u-1}(1-p)^{n-2-u+1}}{\binom{n}{u}p^u(1-p)^{n-u}}
$$

$$
= \frac{\binom{n-2}{u-1}}{\binom{n}{u}}
$$

$$
= \frac{u(n-u)}{n(n-1)}
$$

$$
= \frac{\overline{X}(n-u)}{n-1} \cdot \frac{1/n}{1/n}
$$

$$
= \frac{n\overline{X}(1-\overline{X})}{n-1}
$$

El cual, por Rao-Blackwell, es mejor estimador.

1.2. Ejercicios Propuestos

1. Sean $X_1, \dots, X_n \overset{i.i.d}{\sim} \mathcal{P}(\lambda)$. Se considera el estimador de $\lambda, S_a = a\overline{X}$.

    a) Encuentre el sesgo, varianza y E.C.M. de $S_a(X)$ en función de a y de λ.

    b) Para λ fijo, encuentre el valor de a que minimice el E.C.M.

    c) Para a fijo, determine para qué valores de $\lambda, S_a(X)$ es mejor que $\overline{X}$.

2. Sean $X_1, \dots, X_n \overset{i.i.d}{\sim} N(\mu, \sigma^2)$ y considere la familia de estimadores

$$\sigma_\alpha^2 = \alpha \sum_{i=1}^{n} (X_i - \overline{X})^2$$

    Encuentre el valor de α que minimice el E.C.M.

3. Sean $X_1, \dots, X_n \overset{i.i.d}{\sim} N(\mu, \sigma^2)$. Sea $T_a(X_1, \dots, X_n) = a\overline{X}$. Encuentre el valor de a que minimice el E.C.M. si se quiere estimar μ. Estudie el comportamiento cuando:

    a) $\sigma \to 0$

    b) $\sigma \to \infty$

    c) $n \to \infty$

4. Sean $Y_1, \dots, Y_n$ i.i.d. con distribución Uniforme en el intervalo $[0, \theta]$ y sea $Y_{\max}$ el máximo de los n valores. Alguien sugiere utilizar $Y_{\max}$ para estimar θ. Calcule el error cuadrático medio de este estimador en función de θ.
    Indicación: Verifique que

    a) $\frac{Y_{\max}}{\theta} \sim \text{Beta}(n, 1)$

    b) Si $Z \sim \text{Beta}(\alpha, \beta)$ entonces $V(Z) = \frac{\alpha}{\alpha+\beta} \; \frac{\beta}{\alpha+\beta} \; \frac{1}{\alpha+\beta+1}$

5. Demuestre que $\displaystyle\sum_{i=1}^{n} x_i$ es suficiente minimal cuando

    a) x_i i.i.d. $\mathcal{P}(\lambda)$

    b) x_i i.i.d. $\mathcal{E}(\lambda)$

6. Sea $X_1, \dots, X_n$ una muestra de una población con densidad

$$f(x) = \frac{1}{\sigma} \exp\{-(x - \mu)/\sigma\}, \; \mu < x < \infty, \; 0 < \sigma < \infty$$

    Encuentre un estadístico suficiente bidimensional para (μ, σ).

7. Suponga $X_1, \dots, X_n$ es una muestra de una población con cada una de las siguientes densidades.

a) $f(x|\theta) = \theta x^{\theta-1}, \qquad 0 < x < 1, \theta > 0.$

b) $f(x|\theta) = \theta a x^{a-1} exp(-\theta x^a), \qquad x > 0, \theta > 0, a > 0.$

c) $f(x|\theta) = \theta a^\theta / x^{(\theta+1)}, \qquad x > a, \theta > 0, a > 0.$

Encuentre una estadística suficiente real valorada para θ , con a fijo.

8. Sea $\theta = (\theta_1, \theta_2)$ un parámetro bivariado. Suponga que $T_1(x)$ es suficiente para θ_1 siempre que θ_2 sea fijo y conocido, mientras que $T_2(x)$ es suficiente para θ_2 siempre que θ_1 sea fijo y conocido. Asuma que θ_1 , θ_2 varían independientemente,$\theta_1 \in \Omega_1$, $\theta_2 \in \Omega_2$ y que el conjunto $S = \{x : f(x|\theta) > 0\}$ no depende de θ.

   a) Muestre que si T_1 y T_2 no dependen de θ_1 y θ_2 respectivamente, entonces $(T_1(X), T_2(X))$ es suficiente para θ.

   b) Exhiba un ejemplo en el cual $(T_1(X), T_2(X))$ es suficiente para θ, $T_1(X)$ es suficiente para θ_1 cuando θ_2 es fijo y conocido, pero $T_2(X)$ no es suficiente para θ_2 cuando θ_1 es fijo y conocido.

9. Sea $X_1, \ldots, X_n$ una muestra de una población con densidad $f(x|\theta)$ dada por

$$f(x|\theta) = \frac{1}{\sigma} exp\left\{-\left(\frac{x-\mu}{\sigma}\right)\right\},\ x \leq \mu$$

   Donde $\theta = (\mu, \sigma)$ con $-\infty < \mu < \infty, \sigma > 0$.

   a) Muestre que $\min(X_1, \ldots, X_n)$ es suficiente para μ cuando σ es fijo.

   b) Encuentre un estadístico suficiente minimal para σ cuando μ es fijo.

   c) Encuentre un estadístico suficiente bidimensional para θ.

10. Sea $X_1, \ldots, X_n$ una m.a de una población con densidad

$$f(x) = \theta x^{\theta-1},\ 0 < x < 1,\ \theta > 0$$

   a) ¿Es $\sum X_i$ suficiente para θ?.

11. Sean $X_1, \ldots, X_n \overset{iid}{\sim} Poisson(\theta),\ \theta > 0$. Se quiere estimar $g(\theta) = \exp\{-a\theta\}$ con $a \neq 0$. Se propone el estimador siguiente

$$T_n = \left(1 - \frac{a}{n}\right)^{n\overline{X}}$$

   Estudie y comente las propiedades que tiene el estimador propuesto.

12. Sean $Y_1, \ldots, Y_n$ una m.a. de una población con densidad $Beta(p, q)$. Muestre que el estadístico suficiente para (p, q) es (t_1, t_2), donde

$$t_1 = \left(\prod_{i=1}^n y_i\right)^{1/n}, \qquad t_2 = \left(\prod_{i=1}^n (1 - y_i)\right)^{1/n}$$

   son la media geométrica de los y_i y de $(1 - y_i)$ respectivamente.

13. Muestre que las siguientes distribuciones pertenecen a la Familia Exponencial

    a) Poisson
    b) Exponencial
    c) Gamma
    d) Beta
    e) Weibull

14. Sean $X_1, \ldots, X_n$ i.i.d. con densidad

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

    Encuentre el estimador de θ por el método de sustitución de momentos.

15. Sean $X_1, \ldots, X_n$ i.i.d. $U(\theta_1, \theta_2)$.
    Encuentre el estimador de momentos de $q(\theta_1, \theta_2) = \frac{\theta_1+\theta_2}{2}$.

16. Sean $X_1, \ldots, X_n$ i.i.d con función de densidad

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

    Encuentre un estadístico suficiente.

17. En un experimento de cruce de arvejas, una arveja puede resultar de tipo 1 con probabilidad $(1-\theta)^2$, de tipo 2 con probabilidad $2\theta(1-\theta)$ o de tipo 3 con probabilidad $\theta^2, 0 < \theta < 1, \theta$ desconocido. Se examinan n arvejas y se observa N_j = número de arvejas del tipo $j, j = 1, 2, 3$.

    a) Demuestre que $N_2 + 2N_3$ es un estadístico suficiente.
    b) Encuentre los EMV de las probabilidades que una arveja resulte del tipo j, $j = 1, 2, 3$.

18. El número de errores tipográficos en una página de un diario sigue una $\mathcal{P}(\lambda)$. Se tienen observaciones $X_1, \ldots, X_n$ independientes, donde X_i = número de errores tipográficos en la i-ésima página de un diario, $i = 1, \ldots, n$. Encuentre el EMV de $q(\lambda)$ = los errores de una página exceden 21.

19. Sean $X_1, \ldots, X_n$ i.i.d *Bernoulli*(p) con $0 < p < 1$. Se quiere estimar insesgadamente $g(p) = p$.

    a) Muestre que el estimador X_1 es insesgado para $g(p)$ y mediante el teorema de Rao-Blackwell encuentre un estimador mejor que X_1. Compruebe que es mejor.
    b) Muestre que el estimador $\frac{1}{2}(X_1 + X_2)$ es insesgado para $g(p)$ y al igual que en (a) encuentre un mejor estimador. Compruebe que es mejor.

20. Sean $X_1, \ldots, X_n$ i.i.d con función de densidad

$$f(x;\boldsymbol{\theta}) = \frac{1}{\sigma}\exp\{-(x-\mu)/\sigma\}1_{(x>\mu)}$$

con $\boldsymbol{\theta} = (\mu, \sigma)$

 a) Muestre que $E(x) = \mu + \sigma$.

 b) Muestre que $\boldsymbol{T} = (X_{(1)}, \sum_{i=1}^{n}(X_i - X_{(1)}))$ es suficiente para $\boldsymbol{\theta}$.

21. Suponga que $X_1, \ldots, X_n, X_{n+1}$ son i.i.d $Bernoulli(p)$. Se quiere estimar $g(p) = P(\sum_{i=1}^{n} X_i > X_{n+1})$.

 a) Muestre que $1_{(\sum_{i=1}^{n} X_i > X_{n+1})}$ es un estimador insesgado de $g(p)$.

 b) Encuentre un estimador mejor que el propuesto en (a).

22. Sea $Y_1, ..., Y_n \sim N(\mu, \sigma^2)$. Si μ es conocido, demuestre que $\hat{\sigma}^2 = \frac{\sum(Y_i-\mu)^2}{n}$ alcanza la cota de Cramer–Rao. Comente la significancia de este resultado.

23. Suponga que cuando medimos el radio de un círculo lo hacemos con un error cuya distribución es $N(0, \sigma^2)$ y que realizamos n mediciones independientes. Encuentre un estimador insesgado para el área del círculo y vea si la varianza de este estimador alcanza la cota de Cramer-Rao.

24. Sea Y una v.a. con función de probabilidad

$$P_Y(y;\theta) = \left(\frac{\theta}{2}\right)^{|y|}(1-\theta)^{1-|\theta|}\ ,\ y = -1, 0, 1 \text{ y } 0 \leq \theta \leq 1.$$

 a) Defina el estimador

$$W(Y) = \begin{cases} 2 & \text{si } Y = 1 \\ 0 & \text{en otro caso} \end{cases}$$

 Muestre que $W(Y)$ es un estimador insesgado de θ.

 b) Encuentre un estimador insesgado mejor que $W(Y)$ y muestre que es mejor.

 c) Verifique si el estimador en (b) alcanza la cota de Cramer-Rao.

25. Sean $X_1, \ldots, X_n$ i.i.d $N(\mu, \sigma^2)$. Muestre que

$$T = \frac{\Gamma(\frac{n-1}{2})}{\sqrt{2}\Gamma(\frac{n}{2})}\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

es insesgado para σ^2.

26. Se tienen 2 lotes con artefactos de 2 marcas distintas. Cada artefacto de marca i falla con probabilidad p_i, $i = 1, 2$. Se obtienen m.a de tamaño n_i del lote $i = 1, 2$ y se observa X_i = número de defectuosos en la muestra i.

a) Encuentre un estadístico suficiente para (p_1, p_2).

b) Encuentre estimadores insesgados de p_1 y p_2 que sean funciones del estadístico encontrado en (a).

27. Sea $Y_1, \ldots, Y_n$ una muestra aleatoria de la distribución $Binomial(1, \pi)$, donde π es un parámetro desconocido. Calcule los estimadores máximo verosímil y de momentos de π.

28. Suponga una muestra de tamaño n de la distribución $Normal(\mu,\ \sigma^2)$. Encuentre el estimador máximo verosímil para μ, con σ^2 conocido y el estimador máximo verosímil para σ^2, con μ conocido.

29. Sea $Y_1, ..., Y_n$ una muestra de la distribución $N(\mu, \sigma^2)$.

    a) Grafique la función de verosimilitud y su logaritmo si $\sigma^2 = 1$, $\overline{Y} = 10$.

    b) Grafique el logaritmo de la función verosimilitud si $\overline{Y} = 10, \quad \sum(Y_i - \overline{Y})^2 = 20$.

    c) Repita (a) si el tamaño de la muestra es 100. Compare los resultados.

Capítulo 2

Test de Hipótesis

2.1. Ejercicios Resueltos

EJERCICIO 25

Sean $X_1, ..., X_n$ v.a. i.i.d. $\mathrm{N}(\mu, \sigma^2)$, con σ^2 conocido

(a) Encuentre la region de rechazo para el T.R.V. de

$$H_0 : \mu = \mu_0 \qquad versus \qquad H_1 : \mu \neq \mu_0$$

(b) Encuentre el I.C. de $1 - \alpha\,\%$ resultante de la inversion del test en (a)

SOLUCIÓN

(a) El primer paso a seguir es construir la verosimilitud de los datos y encontrar el EMV del parámetro testeado.

$$L(\boldsymbol{x}; \mu) = \prod_{i=1}^{n} (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}(x_i - \mu)^2\right\}$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right\} \qquad \backslash \ln$$

$$\ell(\boldsymbol{x}; \mu) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 \qquad \backslash \partial\mu$$

$$\frac{\partial \ell(\boldsymbol{x};\mu)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0$$

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

$$\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}$$

Ahora, para construir el TRV encontramos λ de la forma:

$$\lambda = \frac{L_{H_0}(\boldsymbol{x})}{L_{H_1}(\boldsymbol{x})} < c$$

de modo que en $L_{H_0}(\boldsymbol{x})$ se reemplaza el valor de μ bajo H_0 y en $L_{H_1}(\boldsymbol{x})$ se reemplaza el valor de μ bajo H_1, en caso de no ser específico tal valor, se reemplaza por el EMV.

Por lo tanto se tiene:

$$\lambda = \frac{(2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu_0)^2\right\}}{(2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\overline{x})^2\right\}}$$

$$= \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu_0)^2 + \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\overline{x})^2\right\}$$

$$= \exp\left\{\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(x_i-\overline{x})^2 - \sum_{i=1}^{n}(x_i-\mu_0)^2\right]\right\}$$

$$= \exp\left\{-\frac{n}{2\sigma^2}(\overline{x}^2 - 2\mu_0\overline{x} + \mu_0^2)\right\} \qquad \backslash \ln$$

$$\ln(\lambda) = -\frac{n}{2\sigma^2}(\overline{x}^2 - 2\mu_0\overline{x} + \mu_0^2)$$

Debe cumplirse que

$$-\frac{n}{2\sigma^2}(\overline{x}^2 - 2\mu_0\overline{x} + \mu_0^2) < c \qquad \backslash - \frac{2\sigma^2}{n} < 0$$

$$(\overline{x} - \mu_0)^2 > c_1 \qquad \backslash\sqrt{\cdot}$$

$$|\overline{x} - \mu_0| > k$$

Por lo tanto la región de rechazo $R = \{|\overline{x} - \mu_0| > k\}$ donde k dependerá del nivel de confianza y la distribución del resultado.

(b) La inversión de la región anterior, se traduce en una región de confianza

$$A = R^c = \{|\overline{x} - \mu_0| \leq k\}$$

Por lo tanto un intervalo de confianza simétrico y de nivel α, bajo H_0, se traduce en:

$$P(|\overline{x} - \mu| \leq k) = 1 - \alpha$$

$$P(|\overline{x} - \mu| \leq k) = P(-k \leq \overline{x} - \mu \leq k)$$

Por otro lado sabemos que $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$, luego estandarizando se tiene que $\frac{\sqrt{n}(\overline{x}-\mu)}{\sigma} \sim N(0,1)$.

$$P(-k \leq \overline{x} - \mu \leq k) = P\left(-\frac{k\sqrt{n}}{\sigma} \leq \frac{\sqrt{n}(\overline{x} - \mu)}{\sigma} \leq \frac{k\sqrt{n}}{\sigma}\right)$$

$$= P\left(Z \leq \frac{k\sqrt{n}}{\sigma}\right) - P\left(Z < -\frac{k\sqrt{n}}{\sigma}\right)$$

$$= P\left(Z \leq \frac{k\sqrt{n}}{\sigma}\right) - 1 + P\left(Z < \frac{k\sqrt{n}}{\sigma}\right)$$

$$= P\left(Z \leq \frac{k\sqrt{n}}{\sigma}\right) = 1 - \frac{\alpha}{2}$$

Por lo tanto debe cumplirse que:

$$\frac{k\sqrt{n}}{\sigma} = Z_{1-\frac{\alpha}{2}}$$

$$k = Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

Finalmente, el intervalo de $100(1-\alpha)\,\%$ confianza será de la forma

$$-Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \overline{x}-\mu \leq Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

$$\overline{x}-Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x}+Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

EJERCICIO 26

Sea X una v.a. con densidad $f(x) = e^{-x}, x > 0$
Se obtiene una observación de la v.a. $Y = X^{\theta}$, $\theta \in (1,2)$ y se necesita construir un test para

$$H_0 : \theta = 1 \qquad versus \qquad H_1 : \theta = 2$$

(a) Encuentre el test óptimo al nivel de significancia $\alpha = 0{,}01$

(b) Calcule la potencia del test en la parte (a).

SOLUCIÓN

(a) Lo primero que debemos hacer, es encontrar la función densidad de Y, para poder realizar el TRV.
Utilizando técnicas de Probabilidades (Teorema Cambio de Variable), se tiene que:

$$f_Y(y) = \frac{1}{\theta}e^{-y^{\frac{1}{\theta}}}y^{\frac{1-\theta}{\theta}}, \quad y > 0$$

Recordemos que solo tenemos una observación, por lo que la verosimilitud es la misma densidad.

$$\lambda = \frac{f_Y(y)_{H_0}}{f_Y(y)_{H_1}}$$

$$\rightarrow \quad \frac{e^{-y}}{e^{-y^{\frac{1}{2}}} y^{-\frac{1}{2}}} \leq c$$

$$\rightarrow \quad e^{-y+\sqrt{y}} y^{\frac{1}{2}} \leq c_1 \qquad \backslash \ln$$

$$\rightarrow \quad -y + \sqrt{y} + \ln(\sqrt{y}) \leq c_2$$

Esta inecuación, se refleja en la siguiente gráfica

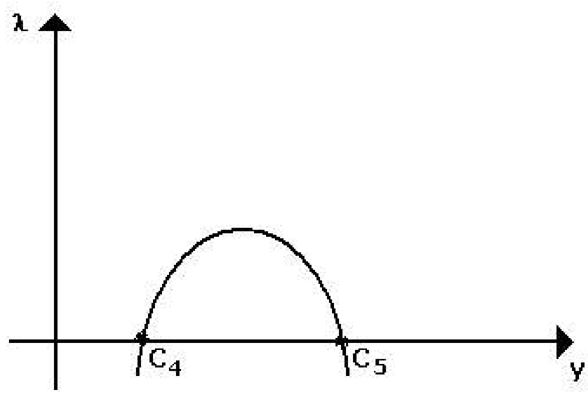

Figura 2.1: TRV

Luego las soluciones cumplen con $y \leq c_4$ o $y \geq c_5$ (Región de Rechazo).
Por lo tanto, bajo H_0, $Y \sim Exp(1)$ y resulta:

$$P(Y \leq c_4 \vee Y \geq c_5) = \int_0^{c_4} e^{-x} dx + \int_{c_5}^{\infty} e^{-x} dx$$

Ahora, hay varias alternativas, por ejemplo que $c_4 = 0$.

$$P(Rechazo) = \int_{c_5}^{\infty} e^{-x} dx = -e^{-x}\Big|_{c_5}^{\infty}$$

$$= e^{-c_5} = \alpha \quad \backslash \ln$$

$$-c_5 = \ln(\alpha)$$

$$c_5 = -\ln(\alpha)$$

Por lo tanto $R = \{y \geq -\ln(\alpha)\}$ es el test óptimo de acuerdo al Lema de Neyman - Pearson. Reemplazando $\alpha = 0{,}10$, se tiene que $R = \{y \geq 2{,}3\}$.

(b) Recordemos que la potencia del test, es $P(Rechazar|H_1)$, por lo que se tiene:

$$P(Y \geq 2{,}3) = P(X^2 \geq 2{,}3)$$

$$= P(X \geq \sqrt{2{,}3})$$

$$= \int_{\sqrt{2{,}3}}^{\infty} e^{-x} dx$$

$$= -e^{-x}\Big|_{\sqrt{2{,}3}}^{\infty}$$

$$= e^{-\sqrt{2{,}3}} = 0{,}219$$

EJERCICIO 27

La legislación impone a los aeropuertos ciertas normas con respecto al ruido emitido por los aviones en el despegue y aterrizaje. En los alrededores, el limite aceptado es de 80 decibels. Mas allá de este limite el aeropuerto debe pagar multa. Los habitantes aseguran que el ruido en dicha zona sobrepasa el valor limite. El aeropuerto, que asegura que no es cierto, decide contratar un grupo de expertos, quienes para concluir asumen que la intensidad del ruido sigue una distribución $\mathrm{N}(\mu, 49)$. Si se considera

$$H_0 : \mu = 80 \qquad versus \qquad H_1 : \mu = 78$$

Encuentre la region critica de Neyman-Pearson, para un nivel de significación del 0.05. ¿Qué decisión tomarán los expertos si se obtiene un promedio de 79.1 decibels en una muestra de 100 aviones ?
Encuentre el error tipo I y el error tipo II.

SOLUCIÓN

Comenzamos por calcular la verosimilitud de los datos, para posteriormente reemplazar los valores de μ, según Neyman-Pearson.

$$L(\boldsymbol{x}) = \prod_{i=1}^{n} \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}\right\}$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right\}$$

Por lo tanto el test queda de la siguiente forma:

$$\lambda = \frac{L_{H_0}(\boldsymbol{x})}{L_{H_1}(\boldsymbol{x})}$$

$$= \frac{(2\pi 49)^{-\frac{n}{2}} \exp\left\{-\frac{\sum_{i=1}^{n}(x_i-80)^2}{2\cdot 49}\right\}}{(2\pi 49)^{-\frac{n}{2}} \exp\left\{-\frac{\sum_{i=1}^{n}(x_i-78)^2}{2\cdot 49}\right\}}$$

$$= \exp\left\{ -\frac{\sum\limits_{i=1}^{n}(x_i-80)^2}{2\cdot 49} + \frac{\sum\limits_{i=1}^{n}(x_i-78)^2}{2\cdot 49} \right\} \quad \backslash \ln$$

$$= -\frac{\sum\limits_{i=1}^{n}(x_i-80)^2}{2\cdot 49} + \frac{\sum\limits_{i=1}^{n}(x_i-78)^2}{2\cdot 49}$$

Empezando a despejar la inecuación, resulta:

$$-\frac{\sum\limits_{i=1}^{n}(x_i-80)^2}{2\cdot 49} + \frac{\sum\limits_{i=1}^{n}(x_i-78)^2}{2\cdot 49} < c \quad \backslash 2\cdot 49$$

$$-\sum_{i=1}^{n}(x_i-80)^2 + \sum_{i=1}^{n}(x_i-78)^2 < c_1$$

$$4\sum_{i=1}^{n} x_i - 316n < c_2$$

$$4\sum_{i=1}^{n} x_i < c_3$$

$$\sum_{i=1}^{n} x_i < k$$

Luego la región de rechazo es

$$R = \left\{ \sum_{i=1}^{n} x_i < k \right\}$$

Ahora para encontrar k, calcularemos el error Tipo I, recordando que $\overline{X} \sim N(\mu, \frac{49}{100})$.

$$P(Rechazar H_0|H_0) = \alpha$$

$$P\left(\sum_{i=1}^{n} x_i < k \Big| \mu = 80\right) = \alpha$$

$$P\left(\frac{\sum_{i=1}^{n} x_i}{n} < \frac{k}{n}\Big| \mu = 80\right) = \alpha$$

$$P(\overline{X} < k_2|\mu = 80) = \alpha$$

$$P\left(\frac{\overline{x} - \mu}{\sqrt{0{,}49}} < \frac{k_2 - \mu}{\sqrt{0{,}49}}\Big| \mu = 80\right) = \alpha$$

$$P\left(Z < \frac{k_2 - 80}{\sqrt{0{,}49}}\right) = 0{,}05$$

$$\frac{k_2 - 80}{\sqrt{0{,}49}} = -1{,}65$$

$$k_2 = 78{,}845$$

Luego la región de rechazo es

$$R = \left\{\frac{\sum_{i=1}^{n} x_i}{n} < 78{,}845\right\}$$

En este caso $\overline{x} = 79{,}1 > 78{,}845$

Por lo tanto no existe evidencia en los datos para rechazar H_0, luego el límite del ruido es aceptado.

$$
\begin{aligned}
P(\text{error Tipo I}) &= P(\textit{rechazar } H_0|H_0) \\
&= P\left(\frac{\sum_{i=1}^{n} x_i}{n} < 78{,}845 \Big| \mu = 80\right) \\
&= P\left(\frac{\overline{x}-\mu}{0{,}7} < \frac{78{,}845-\mu}{0{,}7} \Big| \mu = 80\right) \\
&= P(Z < -1{,}65) \\
&= 0{,}05 = \alpha
\end{aligned}
$$

$$
\begin{aligned}
P(\text{error Tipo II}) &= P(\textit{aceptar } H_0|H_1) \\
&= P\left(\frac{\sum_{i=1}^{n} x_i}{n} > 78{,}845 \Big| \mu = 78\right) \\
&= 1 - P\left(\frac{\overline{x}-\mu}{0{,}7} < \frac{78{,}845-\mu}{0{,}7} \Big| \mu = 78\right) \\
&= 1 - P(Z < 1{,}21) \\
&= 1 - 0{,}8869 \\
&= 0{,}1131 = \beta
\end{aligned}
$$

Luego la potencia del Test es $1 - \beta = 1 - 0{,}1131 = 0{,}8869$.

EJERCICIO 28

Considere una muestra aleatoria $X_1, \ldots, X_n$ proveniente de una distribución $Bernoulli(\theta)$, en donde $\theta \in \{0{,}2, 0{,}4\}$. Encuentre un test de tamaño $\alpha = 0{,}05$ para la hipótesis $H_0 : \theta = 0{,}2$ versus $H_1 : \theta = 0{,}4$.

SOLUCIÓN

Comenzaremos por calcular la verosimilitud asociada a esta distribución, para luego aplicar el Lema de Neyman-Pearson.

$$L(\boldsymbol{x}, \theta) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$$

$$= \left(\frac{\theta}{1-\theta}\right)^{\sum x_i} (1-\theta)^n$$

Por lo tanto, queda lo siguiente:

$$T = \frac{L_{H_1}(\boldsymbol{x})}{L_{H_0}(\boldsymbol{x})}$$

$$= \frac{\left(\frac{0{,}4}{0{,}6}\right)^{\sum x_i} 0{,}6^{10}}{\left(\frac{0{,}2}{0{,}8}\right)^{\sum x_i} 0{,}8^{10}}$$

$$= 2{,}666^{\sum x_i} 0{,}75^{10} \geq k \quad \backslash \ln$$

$$\sum_{i=1}^{n} x_i \ln(2{,}666) + 10\ln(0{,}75) \geq k_1$$

$$\sum_{i=1}^{n} x_i \geq \frac{k_1 - 10\ln(0{,}75)}{0{,}98}$$

$$\sum_{i=1}^{n} x_i \geq \frac{k_1 + 2{,}876}{0{,}98} = k_2$$

Note que $\sum_{i=1}^{n} x_i \sim Bin(n, \theta)$, por lo tanto para encontrar k, se calcula el error tipo I.

$$P(rechazar\ H_0|H_0) = 0{,}05$$

$$P\left(\sum_{i=1}^{n} x_i \geq k_2|\theta = 0{,}2\right) = 0{,}05$$

EJERCICIO 29

Sea X una variable aleatoria proveniente de una distribución logística, esto es,

$$f(x) = \frac{e^{x-\theta}}{[1+e^{x-\theta}]^2}, \quad x \in \mathbb{R}$$

(a) Encuentre el test de Neyman-Pearson para $H_0 : \theta = 0$ versus $H_1 : \theta = 1$.

(b) Determine la región de rechazo

(c) Determine la potencia del test.

SOLUCIÓN

(a) Según Neyman-Pearson, se rechaza H_0 si

$$\frac{L_{H_1}(\boldsymbol{x})}{L_{H_0}(\boldsymbol{x})} \geq k$$

En este caso, como solo se tiene una observación, las verosimilitudes se reducen simplemente a la densidad. Por lo tanto la razón queda de la siguiente forma:

$$\frac{\frac{e^{x-1}}{[1+e^{x-1}]^2}}{\frac{e^{x}}{[1+e^{x}]^2}} \geq k$$

$$(1+e^x)^2 \geq ke(1+e^{x-1})^2 \quad \backslash\sqrt{}$$

$$1+e^x \geq \sqrt{ke}(1+e^{x-1})$$

$$e^x - e^{x-1}\sqrt{ke} \geq \sqrt{ke} - 1$$

$$e^x(1 - \frac{\sqrt{ke}}{e}) \geq \sqrt{ke} - 1$$

$$e^x \geq \frac{\sqrt{ke} - 1}{1 - \frac{\sqrt{k}}{\sqrt{e}}} = \frac{e\sqrt{k} - \sqrt{e}}{\sqrt{e} - \sqrt{k}} \quad \backslash \ln$$

$$x \geq \ln\left(\frac{e\sqrt{k} - \sqrt{e}}{\sqrt{e} - \sqrt{k}}\right)$$

(b) Para encontrar la región de rechazo con $\alpha = 0{,}20$, k debe ser tal que

$$P(R|H_0) = \alpha$$

Por lo tanto queda lo siguiente

$$P\left[X > \ln\left(\frac{e\sqrt{k} - \sqrt{e}}{\sqrt{e} - \sqrt{k}}\right)\right] = 0{,}2$$

$$= \int_{\ln(*)}^{\infty} \frac{e^x}{[1 + e^x]^2} dx$$

$$= -\frac{1}{1 + e^x}\Bigg|_{\ln(*)}^{\infty}$$

$$= \frac{1}{1 + e^{\ln(*)}}$$

$$= \frac{1}{1 + \frac{e\sqrt{k} - \sqrt{e}}{\sqrt{e} - \sqrt{k}}} = 0{,}2$$

$$\longrightarrow \ (5-1)(\sqrt{e}-\sqrt{k}) = e\sqrt{k}-\sqrt{e}$$

$$4\sqrt{e}-4\sqrt{k} = 2{,}7183\sqrt{k}-1{,}6487$$

$$6{,}7183\sqrt{k} = 8{,}2436$$

$$k = 1{,}5056$$

Luego con un nivel de confianza del 80 % se rechaza H_0 cuando $x \geq 1{,}3862$.

(c) La potencia del Test es $P(R|H_1)$, luego queda

$$P(R|H_1) = P\left(x > \ln\left[\frac{e\sqrt{1{,}51}-\sqrt{e}}{\sqrt{e}-\sqrt{1{,}51}}\right] \middle| H_1 : \theta = 1\right)$$

$$= \int_{\ln(1{,}39)}^{\infty} \frac{e^{x-1}}{(1+e^{x-1})^2}dx$$

$$= -\frac{1}{1+e^{x-1}}\Bigg|_{1{,}39}^{\infty}$$

$$= \frac{1}{1+e^{1{,}39-1}} = 0{,}4029$$

Luego la potencia del test es de 0.4029 cuando $\alpha = 0{,}20$.

EJERCICIO 30

Use el Lema de Neyman Pearson para establecer la región crítica al hacer un test para

$$H_0 : \lambda = \lambda_0 \quad versus \quad H_1 : \lambda = \lambda_1 \ \ con\ \lambda_1 < \lambda_0$$

con $\alpha = 0{,}05$ y a partir de una muestra aleatoria $(Y_1, \ldots, Y_n)$ de una población con función de densidad

$$f(y,\lambda) = \lambda^2 y e^{-\lambda y} \quad y > 0, \lambda > 0$$

SOLUCIÓN

Según Neyman Pearson, necesitamos

$$L_{\lambda_1}(\boldsymbol{y}) = \prod_{i=1}^{n} \lambda_1^2 y_i e^{-\lambda_1 y_i}$$

$$= \lambda_1^{2n} \left(\prod_{i=1}^{n} y_i \right) e^{-\lambda_1 \sum y_i}$$

$$L_{\lambda_0}(\boldsymbol{y}) = \prod_{i=1}^{n} \lambda_0^2 y_i e^{-\lambda_0 y_i}$$

$$= \lambda_0^{2n} \left(\prod_{i=1}^{n} y_i \right) e^{-\lambda_0 \sum y_i}$$

Por lo tanto haciendo el cuociente, se obtiene

$$\frac{\lambda_1^{2n} \left(\prod_{i=1}^{n} y_i \right) e^{-\lambda_1 \sum y_i}}{\lambda_0^{2n} \left(\prod_{i=1}^{n} y_i \right) e^{-\lambda_0 \sum y_i}} > k$$

$$\left(\frac{\lambda_1}{\lambda_0} \right)^{2n} e^{\sum y_i (\lambda_0 - \lambda_1)} > k \qquad \backslash \ln$$

$$2n \ln \left(\frac{\lambda_1}{\lambda_0} \right) + \sum_{i=1}^{n} y_i (\lambda_0 - \lambda_1) > k_1$$

$$\sum_{i=1}^{n} y_i (\lambda_0 - \lambda_1) > k_1 - 2n \ln \left(\frac{\lambda_1}{\lambda_0} \right)$$

$$\sum_{i=1}^{n} y_i > \frac{k_1 - 2n \ln \left(\frac{\lambda_1}{\lambda_0} \right)}{\lambda_0 - \lambda_1}$$

Como se quiere conocer la región para $\alpha = 0{,}05$, necesitamos encontrar el valor de $\frac{k_1 - 2n\ln\left(\frac{\lambda_1}{\lambda_0}\right)}{\lambda_0 - \lambda_1}$ correspondiente. Para tal efecto es necesario conocer la distribución de $\sum Y_i$, pero como $Y_i \sim Gamma(2, \lambda)$ y son independientes, se tiene que $\sum Y_i \sim Gamma(2n, \lambda)$. Luego la región crítica está dada por:

$$P(R|H_0) = \alpha$$

$$P\left(\sum_{i=1}^{n} Y_i > \frac{k_1 - 2n\ln\left(\frac{\lambda_1}{\lambda_0}\right)}{\lambda_0 - \lambda_1}\middle| H_0 : \lambda = \lambda_0\right) = 0{,}05$$

$$\int_{\frac{k_1 - 2n\ln\left(\frac{\lambda_1}{\lambda_0}\right)}{\lambda_0 - \lambda_1}}^{\infty} \lambda_0^2 y e^{-\lambda_0 y} dy = 0{,}05$$

Luego la región crítica, resultará de la integral evaluada y despejando el valor de k_1.

EJERCICIO 31

Sea $Y_1, \ldots, Y_n$ una muestra de la distribución Normal con media μ y varianza σ^2 conocida. Considere las hipótesis

$$H_0 : \mu \leq \mu_0 \quad vs \quad H_1 : \mu > \mu_0$$

Determine el valor de k (la región crítica) en función de un nivel de significancia α^* predeterminado si se rechaza H_0 para $\overline{Y} > k$.

<u>SOLUCIÓN</u>

En primer lugar tenemos que si los $Y_i \sim N(\mu, \sigma^2)$ entonces $\overline{Y} \sim N(\mu, \sigma^2/n)$.
Luego para encontrar tal k, recordando que $Z \sim N(0, 1)$, se precede de la siguiente manera:

$$P(R|H_0) = \alpha^*$$

$$P(\overline{Y} > k|H_0) = \alpha^*$$

$$P\left(\frac{\sqrt{n}(\overline{Y} - \mu)}{\sigma} > \frac{\sqrt{n}(k - \mu)}{\sigma}\middle| H_0 : \mu \leq \mu_0\right) = \alpha^*$$

$$P\left(Z > \frac{\sqrt{n}(k-\mu_0)}{\sigma}\right) = \alpha^*$$

$$1 - P\left(Z \leq \frac{\sqrt{n}(k-\mu_0)}{\sigma}\right) = \alpha^*$$

$$P\left(Z \leq \frac{\sqrt{n}(k-\mu_0)}{\sigma}\right) = 1-\alpha^*$$

$$\frac{\sqrt{n}(k-\mu_0)}{\sigma} = Z_{1-\alpha^*}$$

$$\sqrt{n}(k-\mu_0) = \sigma Z_{1-\alpha^*}$$

$$k = \mu_0 + \frac{\sigma}{\sqrt{n}} Z_{1-\alpha^*}$$

Luego la región de rechazo es

$$R = \left\{ \boldsymbol{y} : \overline{y} > \mu_0 + \frac{\sigma}{\sqrt{n}} Z_{1-\alpha^*} \right\}$$

EJERCICIO 32

Sean $X_1, \ldots, X_n$ variables aleatorias iid con función de densidad

$$f(x,\theta) = \frac{2x}{\theta} e^{-\frac{x^2}{\theta}}, \quad x > 0, \theta > 0$$

Construya un TRV para verificar las hipótesis $H_0 : \theta \geq \theta_0$ versus $H_1 : \theta < \theta_0$ y encuentre su región crítica.
Ayuda: ¿Cómo distribuye X_i^2?.

SOLUCIÓN

El supremo de $\theta \in \Theta$, en $L(\boldsymbol{x})$ está dado por el EMV de θ. Luego se tiene:

$$L_\theta(\boldsymbol{x}) = \prod_{i=1}^{n} \frac{2x_i}{\theta} e^{-\frac{x_i^2}{\theta}}$$

$$= \left(\frac{2}{\theta}\right)^n \left(\prod_{i=1}^{n} x_i\right) e^{-\frac{1}{\theta}\sum x_i^2} \qquad \backslash \ln$$

$$\ell(\boldsymbol{x}) = n[\ln(2) - \ln(\theta)] + \sum_{i=1}^{n} \ln(x_i) - \frac{1}{\theta}\sum_{i=1}^{n} x_i^2 \qquad \backslash \partial\theta$$

$$\frac{\partial \ell(\boldsymbol{x})}{\partial\theta} = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i^2 = 0$$

$$\frac{\sum_{i=1}^{n} x_i^2}{\theta^2} = \frac{n}{\theta}$$

$$\frac{\sum_{i=1}^{n} x_i^2}{n} = \frac{\theta^{\cancel{2}}}{\cancel{\theta}}$$

$$\hat{\theta} = \frac{\sum_{i=1}^{n} x_i^2}{n} = \overline{x^2}$$

Luego el TRV, toma la siguiente forma:

$$\frac{\left(\frac{2}{\frac{\sum_{i=1}^{n} x_i^2}{n}}\right)^n \cancel{\prod_{i=1}^{n} x_i} \exp\left\{-\frac{n}{\cancel{\sum_{i=1}^{n} x_i^2}} \cancel{\sum_{i=1}^{n} x_i^2}\right\}}{\left(\frac{2}{\theta_0}\right)^n \cancel{\prod_{i=1}^{n} x_i} \exp\left\{-\frac{n}{\theta_0}\sum_{i=1}^{n} x_i^2\right\}} > k$$

$$\left(\frac{n\theta_0}{\displaystyle\sum_{i=1}^{n} x_i^2}\right)^n \exp\left\{-n + \frac{1}{\theta_0}\sum_{i=1}^{n} x_i^2\right\} > k \qquad \backslash \ln$$

$$n\left[\ln\left(\frac{n\theta_0}{\displaystyle\sum_{i=1}^{n} x_i^2}\right)\right] - n + \frac{1}{\theta_0}\sum_{i=1}^{n} x_i^2 > k_1$$

$$n\left[\ln\left(\frac{n\theta_0}{\displaystyle\sum_{i=1}^{n} x_i^2}\right) - 1\right] + \frac{1}{\theta_0}\sum_{i=1}^{n} x_i^2 > k_1$$

$$\sum_{i=1}^{n} x_i^2 > \underbrace{\frac{\theta_0 k_1}{n\left[\ln\left(\frac{n\theta_0}{\displaystyle\sum_{i=1}^{n} x_i^2}\right)\right]}}_{c}$$

Por lo tanto la región de crítica está dada por:

$$R = \left\{\boldsymbol{x} : \sum_{i=1}^{n} x_i^2 > c\right\}$$

donde c debe cumplir que

$$P\left(\sum_{i=1}^{n} x_i^2 > c \middle| H_0\right) = \alpha$$

Luego para determinar c se necesita conocer la distribución de $\displaystyle\sum_{i=1}^{n} x_i^2$.

Por teoría de Probabilidades (Teorema Cambio de Variable) se tiene que

$$W = \sum_{i=1}^{n} x_i^2 \sim Gamma\left(n, \frac{1}{\theta}\right)$$

Por lo tanto, para conocer c, dado un cierto nivel de significancia α, se deberá resolver la siguiente ecuación:

$$P\left(\sum_{i=1}^{n} x_i^2 > c \middle| H_0\right) = \int_c^\infty \frac{1}{\theta^n \Gamma(n)} w^{n-1} e^{-\frac{w}{\theta}} dw = \alpha$$

EJERCICIO 33

Sea X_i la diferencia entre los tiempos de duración de las i-ésimas piezas de dos equipos similares, $i = 1, \ldots, n$. Si todas las piezas funcionan independientemente y los tiempos de duración de estas piezas siguen una distribución exponencial de parámetro λ, se puede demostrar que $X_1, \ldots, X_n$ es una muestra aleatoria iid con función de densidad dada por:

$$f(x, \lambda) = \frac{1}{2}\lambda e^{-\lambda |x|} \qquad x \in \mathbb{R}, \lambda > 0$$

Construya un TRV para verificar las hipótesis $H_0 : \lambda \geq \lambda_0$ versus $H_1 : \lambda < \lambda_0$ y encuentre su región crítica.
Ayuda: ¿Cómo distribuye $|X_i|$?.

SOLUCIÓN

El TRV dice que rechace H_0 si:

$$\frac{L_{H_1}(\boldsymbol{x})}{L_{H_0}(\boldsymbol{x})} > k$$

Luego para comenzar, es necesario conocer el EMV de λ para su posterior reemplazo en el TRV.

$$L(\boldsymbol{x}; \lambda) = \prod_{i=1}^{n} \frac{1}{2}\lambda e^{-\lambda |x_i|}$$

$$= \left(\frac{\lambda}{2}\right)^n e^{-\lambda \sum |x_i|} \qquad \backslash \ln$$

$$\ell(\boldsymbol{x}; \lambda) = n(\ln(\lambda) - \ln(2)) - \lambda \sum_{i=1}^{n} |x_i| \qquad \backslash \partial\lambda$$

$$\frac{\partial \ell(\boldsymbol{x}; \lambda)}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} |x_i| = 0$$

$$\frac{n}{\lambda} = \sum_{i=1}^{n} |x_i|$$

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} |x_i|}$$

Luego el TRV nos queda como sigue:

$$\frac{\left(\frac{n}{2\sum_{i=1}^{n}|x_i|}\right)^n \exp\left\{-\frac{n}{\cancel{\sum_{i=1}^{n}|x_i|}} \cancel{\sum_{i=1}^{n}|x_i|}\right\}}{\left(\frac{\lambda_0}{2}\right)^n \exp\left\{-\lambda_0 \sum_{i=1}^{n}|x_i|\right\}} > k$$

$$\left(\frac{n}{\lambda_0 \sum_{i=1}^{n}|x_i|}\right)^n \exp\left\{-n + \lambda_0 \sum_{i=1}^{n}|x_i|\right\} > k \qquad \backslash \ln$$

$$n \ln\left(\frac{n}{\lambda_0 \sum_{i=1}^{n}|x_i|}\right) - n + \lambda_0 \sum_{i=1}^{n}|x_i| < k_1$$

$$\lambda_0 \sum_{i=1}^{n}|x_i| < k_1 - n\left[\ln\left(\frac{n}{\lambda_0 \sum_{i=1}^{n}|x_i|}\right) - 1\right]$$

$$\sum_{i=1}^{n}|x_i| < \underbrace{\frac{k_1 - n\left[\ln\left(\dfrac{n}{\lambda_0\displaystyle\sum_{i=1}^{n}|x_i|}\right) - 1\right]}{\lambda_0}}_{c}$$

Luego la región de rechazo está dada por

$$R = \left\{\boldsymbol{x} : \sum_{i=1}^{n}|x_i| < c\right\}$$

donde c se obtiene para algún α dado como:

$$P\left(\sum_{i=1}^{n}|x_i| < c \middle| H_0\right) = \alpha$$

considerando que $\displaystyle\sum_{i=1}^{n}|x_i| \sim Gamma(n, \lambda)$.

EJERCICIO 34[1]

Sea $X_1, \ldots, X_{25}$ una muestra aleatoria de una población normal con varianza igual a 100.

(a) Derive, paso a paso, el test óptimo para determinar la región de rechazo de $H_0 : \mu = 0$ versus $H_1 : \mu = 2$ con un nivel de significancia de $\alpha = 10\,\%$. ¿Cuál es la potencia del test?

(b) Para la misma situación anterior, pero con $H_1 : \mu > 0$. Evalue la potencia en $\mu = 1$ y $\mu = 2$ para $n = 25$ y $n = 100$. A partir de estos puntos, bosqueje las curvas de potencia del test en el mismo eje de coordenadas (indique claramente la leyenda de los ejes).

SOLUCIÓN

(a) Considerando las hipótesis, el test óptimo está dado por Neyman-Pearson, por lo tanto se tiene:

[1]Pregunta I2 II Sem. '03

$$\lambda_{NP} = \frac{L_{H_0}(\boldsymbol{x})}{L_{H_1}(\boldsymbol{x})}$$

$$= \frac{(2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum x_i^2}}{(2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum (x_i-2)^2}}$$

$$= e^{-\frac{1}{2\sigma^2}\left[\sum x_i^2 - \sum (x_i-2)^2\right]}$$

$$= e^{-\frac{1}{2\sigma^2}[4(\sum x_i - n)]}$$

Ahora lo que se necesita es el valor de c tal que $P(\lambda_{NP} < c \big| H_0) = \alpha$, es decir

$$P\left(e^{-\frac{1}{2\sigma^2}[4(\sum x_i - n)]} < c \Big| H_0 : \mu = 0\right) = \alpha$$

$$P\left(\sum_{i=1}^{n} x_i > c^* \Big| H_0 : \mu = 0\right) = \alpha$$

Además tenemos que bajo H_0, $\frac{\sum x_i}{n} \sim N(0, \sigma^2/n)$ con $\sigma^2 = 100$ y $n = 25$, luego resulta:

$$P\left(\sum_{i=1}^{n} x_i > c^* \Big| H_0 : \mu = 0\right) = P\left(\frac{\sum_{i=1}^{n} x_i}{n} > c^{**}\right)$$

Se estandariza: $$P\left(\frac{\overline{x}}{10/\sqrt{25}} > \frac{c^{**}}{2}\right) = \alpha = 0{,}1$$

$$\longrightarrow \frac{c^{**}}{2} = 1{,}28$$

$$c^{**} = 2{,}56 \longrightarrow c^* = 64$$

Luego la región de rechazo para H_0 es

$$R = \left\{ \boldsymbol{x} : \sum_{i=1}^{n} x_i > 64 \right\} \text{ o bien } R = \{\boldsymbol{x} : \overline{x} > 2{,}56\}$$

Ahora la potencia del test, está dada por

$$\begin{aligned} P\left(\sum_{i=1}^{n} x_i > 64 \middle| H_1 \right) &= P\left(\frac{\overline{x} - \mu_1}{\sigma/\sqrt{n}} > \frac{2{,}56 - 2}{2} \right) \\ &= P(Z > 0{,}28) \\ &= 1 - 0{,}6103 \\ &= 0{,}3897 \end{aligned}$$

(b) Cuando $n = 25$ y $\sigma^2 = 100$ con $\alpha = 10\,\%$.
Se rechaza H_0 si $\overline{x} > 2{,}56$, luego tenemos que

$$\begin{aligned} \prod(\mu = 1) &= P(\overline{x} > 2{,}56) \\ &= P(Z > \frac{2{,}56 - 1}{2}) \\ &= P(Z > 0{,}78) \\ &= 0{,}2177 \end{aligned}$$

$$\begin{aligned} \prod(\mu = 2) &= P(\overline{x} > 2{,}56) \\ &= P(Z > \frac{2{,}56 - 2}{2}) \\ &= P(Z > 0{,}28) \\ &= 0{,}3897 \end{aligned}$$

Cuando $n = 100$ y $\sigma^2 = 100$ con $\alpha = 10\,\%$.
Se rechaza H_0 si $\overline{x} > 2{,}56$, luego tenemos que

$$
\begin{aligned}
\prod(\mu = 1) &= P(\overline{x} > 1{,}28) \\
&= P(Z > \frac{1{,}28 - 1}{1}) \\
&= P(Z > 0{,}28) \\
&= 0{,}3897 \\
\prod(\mu = 2) &= P(\overline{x} > 1{,}28) \\
&= P(Z > \frac{1{,}28 - 2}{1}) \\
&= P(Z > -0{,}72) \\
&= 0{,}7642
\end{aligned}
$$

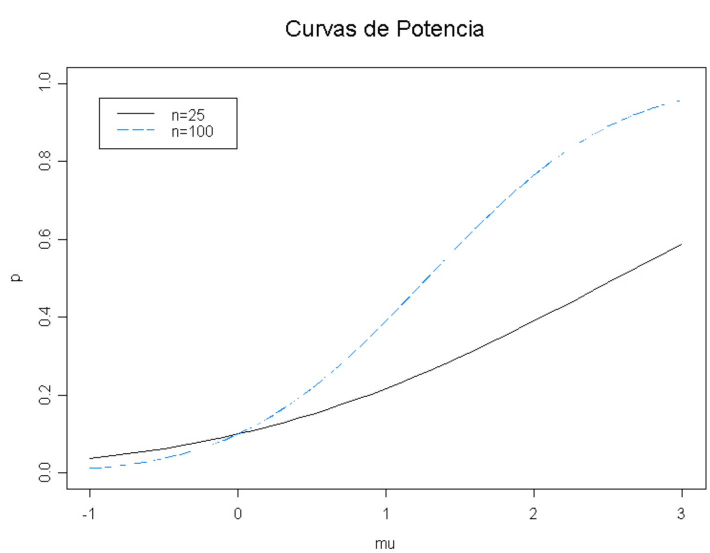

Figura 2.2: Curvas de Potencia

EJERCICIO 35[2]

Sea $X_1, \ldots, X_n$ variables aleatorias independientes distribuidas según la siguiente función de probabilidad discreta, con $0 < \theta < 1$

x	1	2	3
$P(X = x)$	θ^2	$2\theta(1-\theta)$	$(1-\theta)^2$

Sea N_i el número de observaciones igual a i, para $i = 1, 2, 3$, con $\sum N_i = n$. Determine la regla óptima de nivel $\alpha = 0{,}1$, en términos de N_1, N_2, N_3, para la hipótesis $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$.

SOLUCIÓN

En este caso la regla óptima se traduce a utilizar el TRV, para lo cual en esta ocasión necesitamos el EMV de θ.

$$L(\boldsymbol{x}) = \frac{n!}{N_1!N_2!N_3!}(\theta^2)^{N_1}(2\theta(1-\theta))^{N_2}((1-\theta)^2)^{N_3}$$

$$= \frac{n!}{N_1!N_2!N_3!}2^{N_2}\theta^{2N_1+N_2}(1-\theta)^{2N_3+N_2} \qquad \backslash \ln$$

$$\ell(\boldsymbol{x}) = (2N_1 + N_2)\ln(\theta) + (2N_3 + N_2)\ln(1-\theta) \qquad \backslash \partial\theta$$

$$\frac{\partial \ell(\boldsymbol{x})}{\partial \theta} = \frac{2N_1 + N_2}{\theta} - \frac{2N_3 + N_2}{1-\theta} = 0$$

$$\hat{\theta} = \frac{2N_1 + N_2}{2(N_1 + N_2 + N_3)}$$

Por lo tanto, construyendo la razón de verosimilitudes, se obtiene:

$$\Lambda = \frac{\frac{n!}{N_1!N_2!N_3!}2^{N_2}\theta_0^{2N_1+N_2}(1-\theta_0)^{2N_3+N_2}}{\frac{n!}{N_1!N_2!N_3!}2^{N_2}\hat{\theta}^{2N_1+N_2}(1-\hat{\theta})^{2N_3+N_2}}$$

$$= \left(\frac{\theta_0}{\hat{\theta}}\right)^{2N_1+N_2}\left(\frac{1-\theta_0}{1-\hat{\theta}}\right)^{2N_3+N_2}$$

[2] Pregunta I2 II Sem. '03

Recordemos que $-2\ln(\Lambda) \sim \chi^2_{(m)}$.

Por lo tanto se rechaza H_0 si $-2\ln(\Lambda) > \chi^2_{\alpha(m)}$, esto es:

$$-2\left[(2N_1+N_2)\ln\left(\frac{\theta_0}{\hat{\theta}}\right)+(2N_3+N_2)\ln\left(\frac{1-\theta_0}{1-\hat{\theta}}\right)\right] > 2{,}71 = \chi^2_{0,1}(1)$$

EJERCICIO 36

Un fabricante de fibras textiles está investigando una nueva fibra para tapicería, la cual tiene una elongación media por hilo de 12 kg. con una desviación estándar de 0.5 kg. La compañía desea probar la hipótesis $H_0 : \mu_1 \geq 12$, utilizando para ello una muestra aleatoria de tamaño 4.

(a) ¿Cuál es la probabilidad del error tipo I si la región crítica está definida como $\overline{x} < 11{,}5kg$.

(b) Encuentre el error tipo II para el caso donde la verdadera elongación promedio es 11.5 kg.

<u>SOLUCIÓN</u>

(a) Error Tipo I : $P(rechazar\ H_0 \big| H_0) = \alpha$.

Considerando (bajo H_0) que $X \sim N(12, 0{,}5^2) \longrightarrow \overline{X} \sim N(12, 0{,}5^2/4)$

$$P(\overline{x} < 11{,}5 \big| H_0) = P\left(\frac{(\overline{x}-12)\sqrt{4}}{0{,}5} < \frac{(11{,}5-12)\sqrt{4}}{0{,}5}\right)$$

$$= P(Z < -2)$$

$$= 0{,}02275$$

(b) Error Tipo II: $P(aceptar\ H_0 \big| H_1) = \beta$

Considerando (bajo H_1) que $X \sim N(11{,}5, 0{,}5^2) \longrightarrow \overline{X} \sim N(11{,}5, 0{,}5^2/4)$

$$P(\overline{x} \geq 11{,}5 \big| H_1) = P\left(\frac{(\overline{x} - 11{,}5)\sqrt{4}}{0{,}5} \geq \frac{(11{,}5 - 11{,}5)\sqrt{4}}{0{,}5}\right)$$

$$= P(Z \geq 0)$$

$$= 0{,}5$$

2.2. Ejercicios Propuestos

1. Sean $Y_1, \ldots, Y_n$ v.a i.i.d representando la proporciónón de cierta componente química presente en n muestras de un determinado producto, y cuya función de densidad es dada por
$$f(y|\theta) = \theta y^{\theta-1}, \qquad 0 < y < 1$$

   a) Encuentre el EMV de θ.

   b) Utilizando el Lema de Neyman-Pearson, construya un test de nivel α para contrastar
$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta = \theta_1 \quad (\theta_1 > \theta_0)$$

   c) Evalúe el test cuando $\theta_0 = 0{,}3$, $\theta_1 = 0{,}6$, $\alpha = 0{,}05$ y

y_1	0.1
y_2	0.8
y_3	0.5
y_4	0.4
y_5	0.6

   Ayuda: Utilice que $-2\sum_{i=1}^{n} \log Y_i \sim \chi^2_{2n}$ y que $\chi^2_{10}(0{,}95) = 18{,}30$

2. Suponga que el número de tiros al aro de Michael Jordan hasta que éste convierte su primera anotación sigue una distribución geométrica de parámetro θ, donde θ representa la probabilidad de convertir una anotación, es decir
$$P(X = x) = (1-\theta)^{x-1}\theta \quad x = 1, 2, \ldots \quad 0 \leq \theta \leq 1$$

   a) En una m.a de n partidos, en los cuales M. Jordan ha convertido al menos un gol se observan $X_1, \ldots, X_n$ donde X_i = Número de tiros al arco del "aéreo" hasta convertir una anotación en el i-ésimo partido. Determine la región crítica de nivel α para contrastar las siguientes hipótesis:
$$H_0 : \theta = \theta_0 \qquad H_1 : \theta = \theta_1 \quad (\theta_1 > \theta_0)$$
   Utilice aproximación Normal.

   b) Suponga que $\theta_0 = 0{,}5$ y $\theta_1 = 0{,}8$ ¿Cuál sería si conclusión a partir de una muestra de 32 partidos con media muestral de 2.73? Use $\alpha = 0{,}05$.

3. Sean $X_1, \ldots, X_n$ v.a i.i.d con distribución de Pareto, es decir
$$f(x|\theta, \nu) = \frac{\theta \nu^\theta}{x^{\theta+1}} \mathbf{1}_{[\nu,\infty)} \quad \theta > 0 \quad \nu > 0$$

   a) Muestre que el TRV de $H_0 : \theta = 1$ vs $H_1 : \theta \neq 1$ rechaza H_0 si $-T + n \log T < k$ con
$$T = \log\left(\frac{\prod_{i=1}^{n} X_i}{(\text{mín}_i\, X_i)^n}\right)$$

4. Un fabricante de neumáticos afirma que, bajo condiciones normales, la vida promedio de sus neumáticos es mayor que 50,000 kms. En una muestra aleatoria simple de 15 neumáticos sometidos a pruebas de duración, se obtiene un promedio de 54,000 kms. con una desviación estándar de 10,000 kms. Suponiendo que la duración de los neumáticos sigue una distribución normal:

   a) Encuentre un intervalo de 99 % de confianza para la duración promedio de los neumáticos.

   b) Realice un test de hipótesis para contrastar la afirmación realizada por el fabricante de neumáticos. Formule H_0 y H_1 suponiendo que una publicidad engañosa es el error más grave. Use nivel de significancia 5 %. Concluya.

   c) Calcule la potencia del test realizado en (b) para $\mu = 52,000$ kms.

5. Un fabricante sostiene que el diámetro interior de un tubo de aluminio que produce es de 100 mm. Una muestra de 10 tubos fue medida con los siguientes resultados:

100.36	100.31	99.99	100.11	100.64
100.85	99.42	99.91	99.35	100.51

   a) ¿ Se aceptaría lo que dice el fabricante?

   b) ¿Cuál es la probabilidad de aceptar lo que el fabricante afirma cuando el diámetro es realmente 100.55 mm.? (Utilice S^2 como estimador de σ^2).

6. Muchas casas antiguas tienen sistemas eléctricos que utilizan fusibles en lugar de interruptores automáticos de circuito. Un fabricante de fusibles de 40 A quiere estar seguro de que la media de corriente a la que se queman los fusibles es 40. Si la media de corriente es menor de 40, los clientes se quejan porque deben cambiar los fusibles con mucha frecuencia. Si la media de corriente es mayor de 40, el fabricante podría ser el culpable de los daños al sistema eléctrico por el mal funcionamiento de los fusibles. Para verificar la corriente de los fusibles se debe seleccionar e inspeccionar una muestra de éstos. Si fuera a realizarse una prueba de hipótesis sobre los datos resultantes, ¿cuáles hipótesis nula y alternativa serían de interés para el fabricante?

7. Una mezcla de ceniza pulverizada de combustible y cemento Portland para techar debe tener una resistencia a la compresión de más de 1,300KN/m^2. La mezcla no se utilizará a menos que una evidencia experimental indique de manera concluyente que se ha cumplido la especificación de resistencia. Supongamos que la resistencia a la compresión para especímenes de esta mezcla está distribuida normalmente con $\sigma = 60$. Representemos con μ el verdadero promedio de resistencia a la compresión.

   a) ¿Cuáles son las hipótesis nula y alternativa adecuadas?

   b) Representemos con $\overline{X}$ el promedio muestral de resistencia a la compresión para $n = 20$ especímenes seleccionados al azar. Considere el procedimiento de prueba utilizando el estadístico de prueba $\overline{X}$ y la región de rechazo $\overline{x} \geq 1331{,}26$. ¿Cuál es la distribución de probabilidad del estadístico de prueba cuando H_0 es verdadera?. ¿Cuál es la probabilidad de un error tipo I para el procedimiento de prueba?

8. La consejalía de la juventud de un ayuntamiento maneja el dato de que la edad a la que los hijos se independizan de sus padres es una v.a. normal con media 29 años. Aunque la desviación estándar no plantea dudas, se sospecha que la media ha aumentado, sobre todo por el poco apoyo a la política de ayuda al empleo que ha llevado a cabo el ayuntamiento. Así de un estudio reciente sobre 100 jóvenes, que se acaban de independizar, se ha obtenido una media de 30.7 años de edad y una desviación estándar de 3 años.

    a) Con un nivel de significancia del 1 %, ¿es correcta la sospecha que se tiene, acerca de la edad media que se independizan los jóvenes?

    b) Se sabe que el porcentaje de personas, que corresponden al sexo femenino y se independizan antes de los 29 años, no supera el 45 %. Si una muestra de 60 jóvenes son mujeres, y 35 de ellas cumple con las características antes expuestas, que se puede concluir con un nivel de significancia del 5 %.

9. Un consumidor de cierto producto acusa al fabricante diciendo que más del 20 % de las unidades producidas eran defectuosas para confirmar su acusación se utilizó una muestra de tamaño 50 donde el 27 % de los artículos eran defectuosos ¿Qué concluye usted?

10. Una fábrica de hamburguesas inició un proceso de revisión de los estándares de calidad de sus productos. Dichos estándares establecen ciertas dimensiones para el diámetro de sus hamburguesas, el diámetro medio es de 13.9 cm con una desviación estándar estimada de 0.9 cm. Una estándar de calidad establece que el diámetro medio de las hamburguesas debe ser de 14.5 cm ¿Hay alguna evidencia en los datos que las hamburguesas tienen un diámetro incorrecto? ¿Qué supuesto utilizó?

11. Se efectúa una prueba de impacto $Izod$ sobre 20 muestras de tubería PVC. El estándar ASTM para este material requiere que la resistencia al impacto $Izod$ sea mayor que 1.0 ft-lb/in. El promedio y la desviación estándar muestrales son $\overline{x} = 1{,}25$ y $s = 0{,}25$, respectivamente. Pruebe $H_0 : \mu = 1{,}0$ contra $H_1 : \mu > 1{,}0$ utilizando $\alpha = 0{,}01$. Obtenga conclusiones.

12. El sistema de enfriamiento de un submarino nuclear está formado por un ensamble de tuberías soldadas por donde circula un líquido refrigerante. Las especificaciones requieren que la resistencia de la soldadura sea mayor o igual que 150 psi.

    a) Suponga que los ingenieros del diseño deciden probar la hipótesis $H_0 : \mu = 150$ contra $H_1 : \mu > 150$. Explique por qué esta elección de hipótesis alternativa mejor que $H_1 : \mu < 150$.

    b) Al tomar una muestra aleatoria de 20 soldaduras se tiene que $\overline{x} = 153{,}7$ psi. y $s = 11{,}3$ psi. ¿Qué conclusiones pueden obtenerse con respecto a la hipótesis del inciso a)? Utilice $\alpha = 0{,}05$.

Capítulo 3

Ejercicios Resueltos de Interrogaciones

3.1. Interrogaciones I

1. Sea $X_1, X_2, \ldots, X_n$, una muestra aleatoria con función de densidad dada por:

$$f(x|\theta) = \theta x^{\theta-1} I_{[0,1]}(x)$$

   determine el EMV y el estimador de momentos de θ.

2. Sean $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{E}(\lambda)$. Determine la distribución asintótica del EMV de la tasa media μ, la cual se define por $\mu = \frac{1}{\lambda}$.

3. Sea $X_1, X_2, \ldots, X_n$, una muestra aleatoria con función de densidad dada por:

$$f(x|\theta) = \frac{\theta}{(1+x)^{\theta+1}} I_{(0,\infty)}(x),$$

   donde $\theta > 0$. Determine un estadístico suficiente para θ.

4. Sean $X_1, X_2, \ldots, X_n \overset{iid}{\sim} U\left(\theta - \frac{1}{2}, \theta + \frac{1}{2}\right)$. Determine el error cuadrático medio del estimador de momentos de θ.

5. Suponga que $\overline{X}_1$ es la media de una muestra aleatoria de tamaño n de una población con media μ y varianza σ_1^2, $\overline{X}_2$ es la media de una muestra aleatoria (independiente de la muestra aleatoria anterior) de tamaño n de una población con media μ y varianza σ_2^2 y que además $\hat{\mu} = \alpha\overline{X}_1 + (1-\alpha)\overline{X}_2$ es un estimador de μ.

   a) ¿$\hat{\mu}$ es insesgado?

   b) ¿Cuál debe ser el valor de α para que la varianza de $\hat{\mu}$ sea mínima?

6. Sean $X_1, X_2, \ldots, X_n \overset{iid}{\sim} Poisson(\lambda)$.

a) Determine un intervalo de confianza a proximado del 95 % de confianza para λ si $n = 100$ y $\overline{x} = 4$.

b) Obtenga la distribución asintótica del estimador máximo verosímil de $P(X = 0)$.

7. Sean $X_1, X_2, \ldots, X_n$ v.a. *iid* con función de densidad dada por:

$$f(x\,|\beta) = \frac{\beta}{x^{\beta+1}} I_{(1,\infty)}(x),$$

donde $\beta > 0$.

a) Determine el EMV de β.

b) Demuestre que $\sqrt[n]{\prod_{i=1}^{n} X_i}$ es el EMV de $\exp\left(\frac{1}{\beta}\right)$.

c) Demuestre que un intervalo de confianza aproximado de nivel $(1-\alpha)\,100\,\%$ para $\exp\left(\frac{1}{\beta}\right)$ viene dado por

$$\sqrt[n]{\prod_{i=1}^{n} X_i} \pm z_{1-\frac{\alpha}{2}} \frac{\sqrt[n]{\prod_{i=1}^{n} X_i} \times \sum_{i=1}^{n} \ln(X_i)}{n^{\frac{3}{2}}}.$$

8. Sean $Y_1, \ldots, Y_n$ v.a. tales que $Y_i = \beta x_i + \varepsilon_i$, $i = 1, \ldots, n$, con $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ y σ es conocido; y x_i son constantes predeterminadas.

a) Sean

$$\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n} \frac{Y_i}{x_i} \qquad \text{y} \qquad \tilde{\beta} = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} x_i}$$

dos estimadores posibles para β. Calcule el sesgo y el error cuadrático medio para cada estimador.

b) ¿Bajo qué condiciones sobre los x_i los estimadores propuestos anteriormente son consistentes en media cuadrática?

9. En el tiempo $t = 0$ hay un individuo vivo en cierta población. Supongamos ahora que el tiempo que transcurre hasta el primer nacimiento está distribuído exponencialmente con parámetro λ. Después del primer nacimiento, hay dos individuos vivos que se pueden volver a reproducir de forma independiente. El tiempo hasta que el primero vuelve a dar a luz es exponencial con parámetro λ y del mismo modo para el segundo individuo. Por lo tanto, el tiempo hasta el siguiente nacimiento es el mínimo de dos variables aleatorias exponenciales independientes de parámetro λ. Análogamente, una vez que haya ocurrido el segundo nacimiento, hay tres individuos vivos, así que el tiempo hasta el siguiente nacimiento distribuye como el mínimo de tres v.a. exponenciales independientes de parámetro λ, y así sucesivamente. Suponga que se observan n nacimientos y se anotan los tiempos entre los nacimientos sucesivos, es decir, se dispone de $X_1, \ldots, X_n$. Derive el EMV de λ y obtenga la distribución asintótica de éste.

10. Cada una de las variables aleatorias $X_1, \ldots, X_n \overset{iid}{\sim} Poisson\,(\lambda)$ miden el número de clientes que solicitan información a una empresa constructora durante un día. Se quiere saber el número esperado de clientes que solicitan información en un día y para esto se tomó una muestra aleatoria durante 50 días de la cantidad de clientes que llegaron por día, obteniéndose:

Número de clientes por día	0	1	2	3	4
Cantidad de días observados	17	22	7	3	1

    a) En base a los datos observados, encuentre la estimación máximo verosímil de la probabilidad de que no hayan clientes en un día.

    b) En base a los datos observados, calcule un intervalo de confianza aproximado de nivel $(1-\alpha)\,100\,\%$ para la cantidad estimada en el inciso (a).

11. Una urna contiene fichas rojas y negras. Se extrae una ficha de ella. Si la ficha obtenida es roja, se lanza 9 veces la misma moneda de forma independiente, anotándose los resultados de cada lanzamiento. Si la ficha es negra, se lanza de forma independiente la misma moneda hasta obtener 5 caras, anotándose los resultados. Sea p la probabilidad de obtener una ficha roja de la urna y θ la probabilidad de obtener cara en un lanzamiento de esa moneda.

    a) Si se observó ({ficha roja}, {cara, sello, sello, cara, cara, cara, sello, sello, sello}), cuál es la estimación máximo verosímil de p y θ.

    b) Si en otro experimento se definen las v.a. $Y = \begin{cases} 1, \text{ si la ficha es roja} \\ 0, \text{ si la ficha es negra} \end{cases}$, X =número de caras anotadas y Z =número de lanzamientos de la moneda, escriba la función de probabilidad del vector aleatorio (Y, X, Z) en función de los parámetros p y θ.

    c) Suponga ahora que en la función de probabilidad del inciso anterior, p es conocido, ¿pertenecerá esta función de probabilidad a la familia exponencial? y encuentre un estadístico suficiente para θ.

12. Cuando la frecuencia genética está en equilibrio, los genotipos AA, Aa y aa ocurren con probabilidad $(1-\theta)^2$, $2\theta\,(1-\theta)$ y θ^2, respectivamente. Plato *et al.* (1964) publicó los siguientes datos para una muestra de 190 personas:

Tipo de Hemoglobina		
Hp1-1	Hp1-2	Hp2-2
10	68	112

    a) Determine el EMV de θ.

    b) Determine la varianza asintótica del EMV de θ.

    c) Determine un I.C. de nivel α para θ.

13. La edad en la que se adquiere la capacidad de lecto-escritura puede considerarse como una v.a. continua uniforme en (τ, a), con a una constante conocida (por ejemplo, $a = 7$ años). A fin de estimar τ, se toma una muestra aleatoria $X_1, \ldots, X_n$ de la edad exacta en que adquieren la capacidad de lecto-escritura n niños.

    a) Obtenga el estimador de momentos de τ.

    b) ¿es consistente en media cuadrática?. Justifique.

14. La distribución de Pareto ha sido muy utilizada en economía, dado que su función de densidad presenta una cola pesada (es decir decae lentamente). Su función de densidad está dada por:

$$f\left(x|\, \theta, x_0\right) = \theta x_0^{\theta} x^{-(\theta+1)} I_{[x_0,\infty)}\left(x\right),$$

    donde x_0 es una constante conocida y $\theta > 1$. Para una muestra aleatoria $X_1, \ldots, X_n$ con esta distribución responda:

    a) ¿Esta distribución pertenece a la familia exponencial?. Justifique. Identifique un estadístico suficiente para θ.

    b) Demuestre que el EMV de θ es función del estadístico suficiente y obtenga la varianza asintótica de dicho estimador.

15. Suponga que tres componentes electrónicas independientes son puestas a funcionar y después de T horas se presenta una falla. Asuma que la distribución de los tiempos de vida de cada componente sigue una ley $\mathcal{E}\left(\theta^{-1}\right)$.

    a) Determine el EMV de θ.

    b) Obtenga la distribución asintótica del EMV de θ.

3.2. Soluciones

1. El logaritmo de la función de verosimilitud es $l(\theta) = n\ln(\theta) + (\theta - 1)\ln\left(\prod_{i=1}^n x_i\right)$. Luego, al resolver la ecuación $\frac{dl(\theta)}{d\theta} = 0$ para θ, se tiene el punto estacionario $\hat{\theta} = -\frac{n}{\sum_{i=1}^n \ln(x_i)}$. Pero como $\left.\frac{d^2 l(\theta)}{d\theta^2}\right|_{\theta=\hat{\theta}} = -\frac{n}{\hat{\theta}^2} < 0$, entonces el punto estacionario es un máximo, y por tanto el EMV para θ es
$$\hat{\theta} = -\frac{n}{\sum_{i=1}^n \ln(X_i)}.$$
Por otra parte, $\mu = E(X) = \int_0^1 \theta x^\theta dx = \frac{\theta}{\theta+1}$, luego $\theta = -\frac{\mu}{\mu-1}$, y el estimador de momentos para θ es
$$\hat{\theta} = \frac{\overline{X}}{1-\overline{X}}.$$

2. Haciendo la reparametrización $\lambda = \frac{1}{\mu}$, el logaritmo de la función de verosimilitud es $l(\mu) = -n\ln(\mu) - \frac{\sum_{i=1}^n x_i}{\mu}$. Luego, al resolver la ecuación $\frac{dl(\mu)}{d\mu} = 0$ para μ, se tiene el punto estacionario $\hat{\mu} = \overline{x}$. Pero como $\left.\frac{d^2 l(\mu)}{d\mu^2}\right|_{\mu=\hat{\mu}} = -\frac{n}{\overline{x}^2} < 0$, entonces el punto estacionario es un máximo, y por tanto el EMV para μ es
$$\hat{\mu} = \overline{X}.$$
Por otra parte, como la distribución exponencial pertenece a la familia exponencial, entonces la información de Fisher para μ viene dada por
$$I(\mu) = -E\left[\frac{d^2 l(\mu)}{d\mu^2}\right] = -E\left[\frac{n}{\mu^2} - 2\frac{n\overline{X}}{\mu^3}\right] = \frac{n}{\mu^2}.$$
Finalmente, la distribución de $\hat{\mu}$ para n suficientemente grande es aproximadamente $N\left(\mu, \frac{\mu^2}{n}\right)$.

3. Notemos que
$$f(x_1, \ldots, x_n | \theta) = \frac{\theta^n}{\left[\prod_{i=1}^n (1+x_i)\right]^{\theta+1}} \prod_{i=1}^n I_{(0,\infty)}(x_i) = g\left(\prod_{i=1}^n (1+x_i), \theta\right) h(x_1, \ldots, x_n),$$
por tanto, debido al teorema de factorización, un estadístico suficiente para θ podría ser $\prod_{i=1}^n (1+X_i)$ o alguna función biyectiva de éste. Otro estadístico suficiente para θ podría ser toda la muestra, pero éste no aportaría ninguna utilidad extra.

4. La media y varianza de una variable aleatoria con distribución $U(\alpha, \beta)$ son $\frac{\alpha+\beta}{2}$ y $\frac{(\beta-\alpha)^2}{12}$ respectivamente. Por tanto, en nuestro caso tenemos, $\mu = E(X) = \theta$. Luego, el estimador de momentos para θ es $\hat{\theta} = \overline{X}$. Este estimador es insesgado: $E\left(\hat{\theta}\right) = \frac{1}{n}\sum_{i=1}^x E(X_i) = \theta$, lo que implica
$$E\left[\left(\hat{\theta} - \theta\right)^2\right] = Var\left(\hat{\theta}\right) = Var\left(\overline{X}\right) = \frac{1}{n} Var(X) = \frac{1}{n}\frac{\left[\left(\theta + \frac{1}{2}\right) - \left(\theta - \frac{1}{2}\right)\right]^2}{12} = \frac{1}{12n}.$$

5. Notemos que para $i = 1, 2$ se cumple $E\left(\overline{X}_i\right) = \mu$ y $Var\left(\overline{X}_i\right) = \frac{\sigma_i^2}{n}$.

   a) $E(\hat{\mu}) = \alpha E\left(\overline{X}_1\right) + (1-\alpha) E\left(\overline{X}_2\right) = \alpha\mu + (1-\alpha)\mu = \mu$, por tanto $\hat{\mu}$ es insesgado para μ.

   b) $Var(\hat{\mu}) = \alpha^2 Var\left(\overline{X}_1\right) + (1-\alpha)^2 Var\left(\overline{X}_2\right) = \left[Var\left(\overline{X}_1\right) + Var\left(\overline{X}_2\right)\right]\alpha^2 - 2\alpha Var\left(\overline{X}_2\right) + Var\left(\overline{X}_2\right)$, luego $Var(\hat{\mu})$ es una función de α que es una parábola que abre hacia arriba y que por tanto su valor mínimo lo alcanza en

$$\tilde{\alpha} = \frac{Var\left(\overline{X}_2\right)}{Var\left(\overline{X}_1\right) + Var\left(\overline{X}_2\right)} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

6. Es fácil ver que la log-verosimilitud de λ cumple que $l(\lambda) \propto n\overline{x}\ln(\lambda) - n\lambda$, $l'(\lambda) = \frac{n\overline{x}}{\lambda} - n$ y $l''(\lambda) = -\frac{n\overline{x}}{\lambda^2} < 0$.

   a) De $l'(\lambda) = 0$ se tiene que $\hat{\lambda} = \overline{x}$, pero como $l''\left(\hat{\lambda}\right) = -\frac{n}{\overline{x}} < 0$, entonces $\hat{\lambda} = \overline{X}$ es el EMV para λ. Además como la distribución Poisson pertenece a la familia exponencial se cumple que $I(\lambda) = -E\left[l''(\lambda)\right] = \frac{n}{\lambda^2}E\left(\overline{X}\right) = \frac{n}{\lambda}$. Por tanto, la distribución asintótica para $\frac{\left(\hat{\lambda}-\lambda\right)\sqrt{n}}{\sqrt{\lambda}}$ es $N(0,1)$, y de aquí que un intervalo de confianza aproximado de nivel $(1-\alpha)\,100\,\%$ para λ sea $\overline{X} \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{X}}{n}}$, donde $z_{1-\frac{\alpha}{2}}$ es el cuantil $1-\frac{\alpha}{2}$ de una distribución $N(0,1)$. Luego, para los datos observados y $\alpha = 0{,}05$ se tiene

$$\lambda \in \left[4 - 1{,}96 \times \sqrt{\frac{4}{100}}, 4 + 1{,}96 \times \sqrt{\frac{4}{100}}\right] = [3{,}608, 4.392].$$

   b) Notemos que $P(X=0) = e^{-\lambda} =: g(\lambda)$, por tanto $g'(\lambda) = -e^{-\lambda}$. Pero como sabemos que la distribución asintótica de $g\left(\hat{\lambda}\right)$ es $N\left(g(\lambda), \frac{[g'(\lambda)]^2}{I(\lambda)}\right)$, entonces la distribución asintótica del EMV de $P(X=0)$ es

$$N\left(e^{-\lambda}, \frac{\lambda e^{-2\lambda}}{n}\right).$$

7. La log-verosimilitud de β cumple $l(\beta) \propto n\ln(\beta) - \beta\ln\left(\prod_{i=1}^n x_i\right)$, $l'(\beta) = \frac{n}{\beta} - \sum_{i=1}^n \ln(x_i)$ y $l''(\beta) = -\frac{n}{\beta^2} < 0$.

   a) De $l'(\beta) = 0$ se tiene que $\hat{\beta} = \frac{n}{\sum_{i=1}^n \ln(x_i)}$, pero como $l''\left(\hat{\beta}\right) = -\frac{n}{\hat{\beta}^2} < 0$, entonces $\hat{\beta} = \frac{n}{\sum_{i=1}^n \ln(X_i)}$ es el EMV para β.

   b) El EMV de $\exp\left(\frac{1}{\beta}\right)$ es

$$\exp\left(\frac{1}{\hat{\beta}}\right) = \exp\left[\frac{\sum_{i=1}^n \ln(X_i)}{n}\right] = \left\{\exp\left[\ln\left(\prod_{i=1}^n X_i\right)\right]\right\}^{\frac{1}{n}} = \sqrt[n]{\prod_{i=1}^n X_i}.$$

c) Como debemos buscar un I.C para $g(\beta) = \exp\left(\frac{1}{\beta}\right)$, entonces $g'(\beta) = -\frac{\exp\left(\frac{1}{\beta}\right)}{\beta^2}$. Además, como $\frac{\beta}{x^{\beta+1}} = \exp\{-(\beta+1)\ln(x) + \ln(\beta)\}$, entonces dicha distribución pertenece a la familia exponencial, teniendo así,

$$I(\beta) = -E\left[l''(\beta)\right] = \frac{n}{\beta^2}.$$

Por tanto, el error estándar asintótico de $\exp\left(\frac{1}{\hat{\beta}}\right) = \sqrt[n]{\prod_{i=1}^{n} X_i}$ es

$$\sqrt{\frac{\left[g'(\beta)\right]^2}{I(\beta)}} = \frac{\exp\left(\frac{1}{\beta}\right)}{\beta\sqrt{n}}.$$

Luego, un intervalo de confianza aproximado de nivel $(1-\alpha)\,100\,\%$ para $\exp\left(\frac{1}{\beta}\right)$ viene dado por

$$\sqrt[n]{\prod_{i=1}^{n} X_i} \pm z_{1-\frac{\alpha}{2}} \frac{\exp\left(\frac{1}{\hat{\beta}}\right)}{\hat{\beta}\sqrt{n}} = \sqrt[n]{\prod_{i=1}^{n} X_i} \pm z_{1-\frac{\alpha}{2}} \frac{\sqrt[n]{\prod_{i=1}^{n} X_i} \times \sum_{i=1}^{n} \ln(X_i)}{n^{\frac{3}{2}}}.$$

8. Observemos que las v.a. Y_i son **independientes** y con distribución $N(\beta x_i, \sigma^2)$ con σ conocido.

a) $E\left(\hat{\beta}\right) = \frac{1}{n}\sum_{i=1}^{n} \frac{E(Y_i)}{x_i} = \beta$, por tanto el sesgo de $\hat{\beta}$ es cero. De aquí se tiene que

$$E\left[\left(\hat{\beta} - \beta\right)^2\right] = Var\left(\hat{\beta}\right) = \frac{1}{n^2}\sum_{i=1}^{n} \frac{Var(Y_i)}{x_i^2} = \frac{\sigma^2}{n^2}\sum_{i=1}^{n} x_i^{-2}.$$

Por tanto, para que $\hat{\beta}$ sea consistente en media cuadrática debe cumplirse que $\lim\limits_{n\to\infty} n^{-2}\sum_{i=1}^{n} x_i^{-2} = 0$.

Por otra parte, $E\left(\tilde{\beta}\right) = \frac{1}{\sum_{i=1}^{n} x_i}\sum_{i=1}^{n} E(Y_i) = \beta$, por lo que $\tilde{\beta}$ también es insesgado. De aquí se tiene que

$$E\left[\left(\tilde{\beta} - \beta\right)^2\right] = Var\left(\tilde{\beta}\right) = \frac{1}{\left(\sum_{i=1}^{n} x_i\right)^2}\sum_{i=1}^{n} Var(Y_i) = \frac{\sigma^2}{n\overline{x}}.$$

Por tanto, para que $\tilde{\beta}$ sea consistente en media cuadrática debe cumplirse que $\lim\limits_{n\to\infty} n\overline{x} = \infty$.

b) Fue hecho al resolver el inciso (a).

9. Por la descripción del problema se tiene que las v.a. $X_i = \min\limits_{j=1,\ldots,i}\{Y_j\}$ son independientes y donde $Y_1, \ldots, Y_i \overset{iid}{\sim} \mathcal{E}(\lambda)$. La función de densidad de las v.a. X_i viene dada por

$$\begin{aligned} f_{X_i}(x) &= i\left[1 - F_Y(x)\right]^{i-1} f_Y(x) = i\left[1 - \left(1 - e^{-\lambda x}\right)\right]^{i-1} \lambda e^{-\lambda x} \\ &= i\lambda\exp(-i\lambda x). \end{aligned}$$

Por tanto, la función de verosimilitud en dependencia de los datos observados es

$$L(\lambda \,|x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x) = n!\lambda^n \exp\left(-\lambda \sum_{i=1}^{n} ix_i\right),$$

y su logaritmo cumple,

$$l(\lambda) \propto n \ln(\lambda) - \lambda \sum_{i=1}^{n} ix_i.$$

Luego, de $l'(\lambda) = 0$ se tiene $\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} ix_i}$, y como $l''\left(\hat{\lambda}\right) = -\frac{n}{\hat{\lambda}^2} < 0$, entonces el EMV de λ es

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} iX_i}.$$

Por otra parte, como $X_i \sim \mathcal{E}(i\lambda)$, la información de Fisher viene dada por

$$I(\lambda) = -E\left[l''(\lambda)\right] = \frac{n}{\lambda^2}$$

y la distribución asintótica de $\hat{\lambda}$ es $N\left(\lambda, \frac{\lambda^2}{n}\right)$.

10. Es conocido que si $X_1, \ldots, X_n \overset{iid}{\sim} Poisson(\lambda)$ entonces el EMV de λ es $\hat{\lambda} = \overline{X}$ con distribución asintótica $N\left(\lambda, \frac{\lambda}{n}\right)$.

   a) Como $P(X=0) = e^{-\lambda} = g(\lambda)$ entonces el EMV para $P(X=0)$ es $g\left(\hat{\lambda}\right) = e^{-\hat{\lambda}}$. Pero de los datos observados tenemos $\overline{x} = (0 \times 17 + 1 \times 22 + 2 \times 7 + 3 \times 3 + 4 \times 1)/50 = \frac{49}{50} \approx 1$, y por tanto la estimación para $P(X=0)$ es $g(1) = e^{-1}$.

   b) Del inciso (a) se tiene que la varianza asintótica de $e^{-\hat{\lambda}}$ es $\frac{[g'(\lambda)]^2}{I(\lambda)} = \frac{\lambda \exp(-2\lambda)}{n} \hat{=} \frac{1}{50}e^{-2}$. Por tanto, un intervalo de confianza aproximado de nivel $(1-\alpha)100\,\%$ para $P(X=0) = e^{-\lambda}$ es

$$e^{-1} \pm z_{1-\frac{\alpha}{2}} \frac{1}{5\sqrt{2}} e^{-1}.$$

11. Siempre hay que tener en cuenta que la extracción de una ficha es un suceso independiente del lanzamiento de la moneda, sin embargo no es independiente del número de veces que se lanzará la moneda.

   a) La v.a. Y definida en 4.(b) sigue una distribución $Bin(1, p)$, pero es conocido que el EMV de p es $\hat{p} = Y$, por tanto, como se observó $Y = 1$ (se observó una fciha roja), entonces la estimación máximo verosímil de p es $\hat{p} = 1$.

   Como se observó una ficha roja, entonces la v.a. X definida en 4.(b) distribuye $Bin(9, \theta)$, pero es conocido que el EMV de θ es $\hat{\theta} = \frac{X}{n}$, por tanto, como se observaron 4 caras, entonces $X = 4$ y la estimación máximo verosímil de θ es $\hat{\theta} = \frac{4}{9}$.

b) Por la descripción del experimento, si se obtiene una ficha roja (con probabilidad p), luego se realiza un experimento $Bin\,(9,\theta)$; y si la ficha es negra (con probabilidad $1-p$), luego se realiza un experimento $BinNeg\,(5,\theta)$. Luego, por la fórmula de probabilidad total, se tiene

$$\begin{aligned} P_{(Y,X,Z)}\,(y,x,z) &= P\,(X=x, Z=z|\,Y=1)\,P\,(Y=1)+P\,(X=x,Z=z|\,Y=0)\,P\,(Y=0) \\ &= P\,(X=x|\,X\sim Bin\,(9,\theta))\,I_{\{9\}}\,(z)\,P\,(Y=1) \\ &\quad +P\,(Z=z|\,Z\sim BinNeg\,(5,\theta))\,I_{\{5\}}\,(x)\,P\,(Y=0) \\ &= \begin{pmatrix} 9 \\ x \end{pmatrix} \theta^x\,(1-\theta)^{9-x}\,p^y+\begin{pmatrix} z-1 \\ 5-1 \end{pmatrix}\theta^5\,(1-\theta)^{z-5}\,(1-p)^{1-y}, \end{aligned}$$

donde la última igualdad sólo se cumple si $y\in\{0,1\}$, $x\in\{0,1,\ldots,9\}$ y $z=5,6,\ldots$, en otro caso $P_{(Y,X,Z)}\,(y,x,z)=0$.

Pero como 9 es un número de lanzamientos y 5 es un número de caras entonces, para el caso general (Y,X,Z) (no necesariamente con 9 lanzamientos y 5 caras), se tiene

$$P_{(Y,X,Z)}\,(y,x,z)=\begin{pmatrix} z \\ x \end{pmatrix}\theta^x\,(1-\theta)^{z-x}\,p^y+\begin{pmatrix} z-1 \\ x-1 \end{pmatrix}\theta^x\,(1-\theta)^{z-x}\,(1-p)^{1-y}$$

si $y\in\{0,1\}$, $x\in\{0,1,\ldots,z\}$ y $z=x,x+1,\ldots$; y $P_{(Y,X,Z)}\,(y,x,z)=0$ en otro caso.

c) El conjunto $\{(y,x,z): y\in\{0,1\}\,; x\in\{0,1,\ldots,z\}\,; z=x,x+1,\ldots\}$ no depende del parámetro θ, y como en este caso p es conocido, entonces

$$\begin{aligned} &\left[\begin{pmatrix} z \\ x \end{pmatrix}p^y+\begin{pmatrix} z-1 \\ x-1 \end{pmatrix}(1-p)^{1-y}\right]\theta^x\,(1-\theta)^{z-x} \\ =\ &\exp\left\{x\ln\left(\frac{\theta}{1-\theta}\right)+z\ln\,(1-\theta)+\ln\left[\begin{pmatrix} z \\ x \end{pmatrix}p^y+\begin{pmatrix} z-1 \\ x-1 \end{pmatrix}(1-p)^{1-y}\right]\right\}. \end{aligned}$$

De aquí se ve claramente que $P_{(Y,X,Z)}\,(y,x,z)$ pertenece a la familia exponencial, por tanto un estadístico suficiente para θ es (T_1,T_2) donde $T_1\,(Y,X,Z)=X$ y $T_2\,(Y,X,Z)=Z$.

Nota: Aplicando el teorema de factorización también se puede llegar a que (X,Z) es un estadístico suficiente para θ.

12. En este caso las observaciones provienen de una distribución $Multinomial\left(n=190,(1-\theta)^2,2\theta\,(1-\theta)\,,\theta^2\right)$.

a) La verosimilitud de θ cumple, $L\,(\theta)\propto(1-\theta)^{2x_1}\,[2\theta\,(1-\theta)]^{x_2}\,\theta^{2x_3}$ y su logaritmo, $l\,(\theta)\propto(2x_1+x_2)\ln\,(1-\theta)+(x_2+2x_3)\ln\,(\theta)$. Luego, al resolver la ecuación $0=l'\,(\theta)=-\frac{2x_1+x_2}{1-\theta}+\frac{x_2+2x_3}{\theta}$, se tiene $\hat{\theta}=\frac{x_2+2x_3}{2n}$. Como $l''\left(\hat{\theta}\right)=-\frac{2x_1+x_2}{\left(1-\hat{\theta}\right)^2}-\frac{x_2+2x_3}{\hat{\theta}^2}<0$ entonces efectivamente el EMV de θ es $\hat{\theta}=\frac{X_2+2X_3}{2n}$, y la estimación según los datos es $\hat{\theta}=\frac{68+2*112}{2*190}=\frac{73}{95}$.

b) Como

$$P(x_1,\ldots,x_m)=\frac{n!}{\prod_{i=1}^m x_i!}\prod_{i=1}^m p_i^{x_i}=\exp\left\{\sum_{i=1}^m x_i\ln(p_i)+\ln(n!)-\sum_{i=1}^m \ln(x_i!)\right\},$$

entonces la distribución multinomial pertenece a la familia exponencial, por lo que

$$\begin{aligned}I(\theta) &= -E\left[l''(\theta)\right]=\frac{2E(X_1)+E(X_2)}{(1-\theta)^2}+\frac{E(X_2)+2E(X_3)}{\theta^2}\\ &= n\left[\frac{2(1-\theta)^2+2\theta(1-\theta)}{(1-\theta)^2}+\frac{2\theta(1-\theta)+2\theta^2}{\theta^2}\right]=\frac{2n}{(1-\theta)\theta}\end{aligned}$$

y la varianza asintótica de $\hat{\theta}=\frac{X_2+2X_3}{2n}$ es $\frac{(1-\theta)\theta}{2n}$.

c) Un I.C. para θ de nivel α es

$$\begin{aligned}\hat{\theta}\pm z_{1-\frac{\alpha}{2}}\sqrt{I^{-1}\left(\hat{\theta}\right)} &= \frac{X_2+2X_3}{2n}\pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{(2n-X_2-2X_3)(X_2+2X_3)}{8n^3}}\\ &= \frac{1}{2n}\left(X_2+2X_3\pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{n^2-(X_1-X_3)^2}{2n}}\right).\end{aligned}$$

13. Debemos tener en cuenta que si $X\sim U_{(\tau,a)}$ entonces $E(X)=\frac{\tau+a}{2}$ y $Var(X)=\frac{(a-\tau)^2}{12}$.

a) Al resolver la ecuación $\overline{X}=\frac{\hat{\tau}+a}{2}$, nos queda el estimador de momentos para τ es $\hat{\tau}=2\overline{X}-a$.

b) Como $E(\hat{\tau})=2E\left(\overline{X}\right)-a=\tau$, entonces $E\left[(\hat{\tau}-\tau)^2\right]=Var(\hat{\tau})=4Var\left(\overline{X}\right)=\frac{(a-\tau)^2}{3n}\to 0$. Por tanto $\hat{\tau}$ es consistente en media cuadrática.

14. Estamos suponiendo que $X_1,\ldots,X_n\overset{iid}{\sim}f(x|\,\theta)$.

a) Sí pertenece a la familia exponencial ya que

$$\theta x_0^{\theta}x^{-(\theta+1)}I_{[x_0,\infty)}(x)=\exp\left\{-\theta\ln(x)+\ln\left(\theta x_0^{\theta}\right)-\ln(x)+\ln\left(I_{[x_0,\infty)}(x)\right)\right\}.$$

Por tanto un estadístico suficiente para θ es $T(X_1,\ldots,X_n)=\frac{1}{n}\sum_{i=1}^n\ln(X_i)$.

b) La log-verosimilitud de θ cumple, $l(\theta)\propto-\theta\sum_{i=1}^n\ln(x_i)+n\ln(\theta)+n\theta\ln(x_0)$, $l'(\theta)=-\sum_{i=1}^n\ln(x_i)+\frac{n}{\theta}+n\ln(x_0)$ y $l''(\theta)=-\frac{n}{\theta^2}$. El EMV para θ es

$$\hat{\theta}=\frac{n}{\sum_{i=1}^n\ln(X_i)-n\ln(x_0)}=\frac{1}{T(X_1,\ldots,X_n)-\ln(x_0)}$$

ya que cumple $l'\left(\hat{\theta}\right)=0$ y $l''\left(\hat{\theta}\right)<0$.

Por otra parte, la varianza asintótica de $\hat{\theta}$ es

$$I^{-1}(\theta)=-\left[E\left(l''(\theta)\right)\right]^{-1}=\frac{\theta^2}{n}.$$

15. Que después de T horas se presente una falla, significa que el menor tiempo de vida de las tres componentes fue de T horas. Por tanto si X_1, X_2 y X_3 respresentan el tiempo de vida de cada componente, entonces $T = \text{mín}\{X_1, X_2, X_3\}$ donde $X_1, X_2, X_3 \overset{iid}{\sim} \mathcal{E}\left(\theta^{-1}\right)$.

    a) La función de densidad de la v.a. T viene dada por

    $$\begin{aligned} f_T(t) &= 3\left[1 - F_X(t)\right]^{3-1} f_X(t) = 3\left[1 - \left(1 - e^{-\theta^{-1}t}\right)\right]^2 \theta^{-1} e^{-\theta^{-1}t} \\ &= 3\theta^{-1} \exp\left(-3\theta^{-1}t\right). \end{aligned}$$

    Por tanto $l(\theta) \propto -\ln(\theta) - 3t\theta^{-1}$, $l'(\theta) = -\theta^{-1} + 3t\theta^{-2}$ y $l''(\theta) = \theta^{-2} - 6t\theta^{-3}$. De $l'(\theta) = 0$ se tiene $\hat{\theta} = 3t$, pero como $l''\left(\hat{\theta}\right) = (3t)^{-2} - 6t\,(3t)^{-3} = -9t^{-2} < 0$, entonces $\hat{\theta} = 3T$ es el EMV de θ.

    b) Para buscar una distribución asintótica es necesario suponer que se observará una muestra aleatoria $T_1, \ldots, T_n \overset{iid}{\sim} \mathcal{E}\left(3\theta^{-1}\right)$. Pero como la $\mathcal{E}\left(3\theta^{-1}\right)$ pertenece a la familia exponencial, entonces $I(\theta) = -nE\left(l''(\theta)\right) = 6nE(T)\,\theta^{-3} - n\theta^{-2} = n\theta^{-2}$. Luego, la distribución asintótica de $\hat{\theta} = 3\overline{T}$ es

    $$N\left(\theta, \frac{\theta^2}{n}\right).$$

3.3. Interrogaciones II

1. En una muestra aleatoria de 200 residentes de cierta ciudad en Los Estados Unidos que conducen un automóvil de fabricación estadounidense, 165 indicaron que emplean el cinturón de seguridad. Otra muestra aleatoria, independiente de la anterior, de 250 residentes de la misma ciudad que conducen un automóvil de fabricación extranjera, reveló que 198 utilizan el cinturón de seguridad.

   a) ¿Existe alguna evidencia que indique una diferencia en el uso del cinturón de seguridad entre los conductores de automóviles estadounidenses y extrenjeros? (utilice $\alpha = 0{,}05$).

   b) Determine el valor-p de esta prueba y bosqueje el gráfico de la función de potencia respectiva (identifique claramente las leyendas de los ejes).

2. Una empresa de sondeos prueba dos barrenas perforando pozos hasta una profundidad máxima de 125 metros y anotando el número de horas que requiere el proceso. La primera barrena, (1), se utilizó en 12 casos con un tiempo medio de 17.3 horas. Con la segunda, (2), se perforaron 10 pozos, obteniéndose un resultado de 22.3 horas (tiempo medio). Las desviaciones estándar muestrales son $s_1 = 6$ y $s_2 = 6$. ¿Hay evidencia que permita afirmar que la barrena 1 es más eficiente que la barrena 2, en términos de menor tiempo medio?

   a) Plantee las hipótesis y lleve a cabo el test considerando $\alpha = 0{,}01$. ¿Cuál es su conclusión?

   b) ¿Cuál es el menor valor de α para el cual el test es significativo?

   c) Esboce la curva de potencia. Calcule la probabilidad (usando distribución normal en vez de t-student) del error de tipo II si $\mu_1 - \mu_2 = -3$.

3. Sea X una v.a. cuyas funciones de probabilidad bajo H_0 y H_1 están dadas por,

X	1	2	3	4	5	6	7
$P(X = x \mid H_0)$	0.01	0.01	0.01	0.01	0.01	0.01	0.94
$P(X = x \mid H_0)$	0.06	0.05	0.04	0.03	0.02	0.01	0.79

   a) Encuentre el test más potente de nivel $\alpha = 0{,}04$ para contrastar H_0 vs. H_1.

   b) Determine la probabilidad del error de tipo II del test encontrado en (a).

4. Se dice que la aleación reduce la resistencia de un tipo estándar de conductor eléctrico y que las mediciones de la resistencia en conductores estándares y con aleación distribuye normal. Se hicieron diez mediciones en un conductor estándar, obteniéndose una resistencia promedio de 0.19 ohm, y una desviación estándar de 0.03 ohm. Otras diez mediciones independientes se hicieron en un alambre aleado y se obtuvo una resistencia promedio de 0.11 ohm y una desviación estándar de 0.02. ¿La aleación parece reducir la resistencia promedio del alambre? Asumiendo varianzas iguales, probar en el nivel de significación de 5 %.

5. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria con distribución $Geo(\pi_1)$, y sea $Y_1, Y_2, \ldots, Y_m$ otra muestra aleatoria independiente de la anterior con distribución $Geo(\pi_2)$. Derive el TRV (paso a paso) para la hipótesis: $H_0 : \pi_1 = \pi_2$ versus $H_1 : \pi_1 \neq \pi_2$.

6. Un especialista afirma que el tiempo en desarrollar la I2 sigue una distribución $f(x) = \begin{cases} kx \text{ si } x \in (0, 120) & \text{min.} \\ 0 & \text{si no} \end{cases}$. A una muestra de 50 alumnos que rindieron la I2 de ingeniería se les midió el tiempo. La tabla adjunta muestra los resultados obtenidos. ¿Existe evidencia que permita afirmar que los tiempos considerados no se ajustan a la distribución especificada por el especialista? (use un nivel de significancia del 5 %).

Categoría	Frec. Obs.
<1 hora	10
1 a 1.5 horas	20
1.5 a 2 horas	20

7. Sea $X_1, X_2, \ldots, X_n$, una muestra aleatoria con función de densidad dada por $f_X(x) = \frac{1}{\mu} e^{-\frac{x}{\mu}}$.

   a) Construya el test uniformemente más potente para probar las hipótesis $H_0 : \mu \leq \mu_0$ vs. $H_1 : \mu > \mu_0$.

   b) Suponga que para comparar las hipótesis $H_0 : \mu \leq 8$ contra $H_1 : \mu > 8$ se toma una muestra de tamaño $n = 1$ y se considera un nivel de significación $\alpha = 0{,}1$. Encuentre el punto crítico del test.

   c) En base al inciso (b):

      1) Calcule la potencia de este test para $\mu = 12$.

      2) Calcule el valor-p para $X_1 = 15$.

8. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria con distribución $U_{[0,\theta]}$ con $\theta > 0$ y sea $T = \max\limits_{i=1,\ldots,n} \{X_i\}$. Se quiere comparar las hipótesis $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$ y para esto se utiliza el test que rechaza H_0 cuando $T > t_0$.

   a) Obtenga la función de potencia de este test, demuestre que es creciente en θ y evalúela para $n = 3$, $\theta_0 = 2$ y $t_0 = 1{,}9$.

   b) Calcule el punto crítico t_0 para un nivel de significación del 10 % en función de θ_0.

   c) Encuentre el valor-p cuando $n = 3$ y

      1) $\max\limits_{i=1,\ldots,n} \{x_i\} = 1{,}9$ y $\theta_0 = 2$.

      2) $\max\limits_{i=1,\ldots,n} \{x_i\} = 2{,}1$ y $\theta_0 = 2$.

9. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria proveniente de una población Bernoulli de parámetro π. Derive el test óptimo, paso a paso, para las hipótesis: $H_0 : \pi = \pi_0$ versus $H_1 : \pi = \pi_1$, con $\pi_0 > \pi_1$.

   a) Obtenga la forma de la región de rechazo de nivel α.

b) Para $n = 25$, $\pi_0 = 0{,}20$, $\pi_1 = 0{,}10$, $\alpha = 5\,\%$ y usando el TLC, obtenga la región de rechazo.

10. Suponga que se realizan n mediciones bajo las condiciones de un tratamiento (1), y otras n mediciones son tomadas, independientemente, bajo otro tratamiento (2). Asuma que los datos tienen distribución normal y que la desviación estándar poblacional de cualquier observación es 10 bajo cualquier tratamiento.

    a) ¿Cuán grande debe ser n de modo que el intervalo de confianza al 95 % para $\mu_1 - \mu_2$ tenga un ancho igual a 2?

    b) ¿Cuán grande debe ser n de modo que el test $H_0 : \mu_1 = \mu_2$ contra $H_1 : \mu_1 > \mu_2$ tenga potencia igual a 0.5 si $\mu_1 - \mu_2 = 2$ y $\alpha = 0{,}10$?

3.4. Soluciones

1. Sea p_l la proporción de uso de cinturón en autos locales, y p_e la proporción de uso de cinturón en autos extranjeros. Luego, debemos probar las hipótesis $H_0 : p_l = p_e$ vs. $H_1 : p_l \neq p_e$.

   a) El estadístico observado para el test aproximado de comparación de proporciones es
   $$Z = \frac{\hat{p}_l - \hat{p}_e}{\sqrt{\overline{p}\left(1-\overline{p}\right)\left(\frac{1}{n_l}+\frac{1}{n_e}\right)}}$$
   donde
   $$\overline{p} = \frac{n_l\hat{p}_l + n_e\hat{p}_e}{n_l+n_e} = \frac{200\frac{165}{200}+250\frac{198}{250}}{200+250} = 0{,}81.$$
   Luego,
   $$Z = \frac{\frac{165}{200}-\frac{198}{250}}{\sqrt{0{,}81\left(1-0{,}81\right)\left(\frac{1}{200}+\frac{1}{250}\right)}} = 0{,}89.$$
   Como $z_{1-\alpha/2} = z_{0{,}975} = 1{,}96$, entonces de $Z = 0{,}89 < 1{,}96$ se tiene que no se rechaza H_0, y por tanto, no hay suficientes evidencias para decir que existen diferencias significativas en el uso del cinturón de seguridad entre los conductores de autos estadounidenses y extranjeros.

   b) En este caso se tiene
   $$p = P\left(|Z| > 0{,}89\right) = 2P\left(Z > 0{,}89\right) = 2\left[1 - P\left(Z \leq 0{,}89\right)\right] = 2\left[1-0{,}8133\right] = 0{,}3734.$$
   Sean $X_i \overset{iid}{\sim} B\left(1, p_l\right)$ v.a. que toman el valor 1 si el conductor de auto local usa cinturón $Y_j \overset{iid}{\sim} B\left(1,p_e\right)$ otras v.a. que toman el valor 1 si el conductor de auto extrajero usa cinturón, donde $i = 1,\ldots,200$ y $j = 1,\ldots,250$. Luego, por el teorema de límite central se tiene
   $$\begin{aligned}\hat{p}_l &= \frac{1}{200}\sum_{i=1}^{200} X_i \mathrel{\dot\sim} N\left(p_l, \frac{p_l\left(1-p_l\right)}{200}\right),\\ \hat{p}_e &= \frac{1}{250}\sum_{j=1}^{250} Y_j \mathrel{\dot\sim} N\left(p_e, \frac{p_e\left(1-p_e\right)}{250}\right) \text{ y}\\ &\frac{\hat{p}_l-\hat{p}_e-\left(p_l-p_e\right)}{\sqrt{\frac{p_l(1-p_l)}{200}+\frac{p_e(1-p_e)}{250}}} \mathrel{\dot\sim} N\left(0,1\right).\end{aligned}$$
   Por tanto, el gráfico de la función de potencia es similar al del test para comparar $H_0 : \mu_x = \mu_y$ vs. $H_1 : \mu_x \neq \mu_y$ cuando los datos distribuyen normal. Luego, el gráfico de la función de potencia es similar al de la figura 3.1.

2. Este problema se resuelve con un test de comparación de dos medias con datos normales de varianzas desconocidas, pero consideradas iguales.

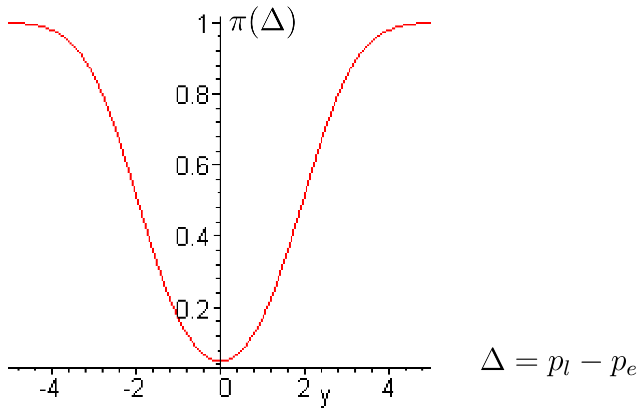

Figura 3.1: Función de potencia

a) Sea μ_i el tiempo medio en que la barrena i $(i = 1, 2)$ se demora en perforar el pozo. Luego, debemos probar las hipótesis $H_0 : \mu_1 \geq \mu_2$ vs. $H_1 : \mu_1 < \mu_2$.

Como
$$S_p^2 = \frac{(n_1 - 1)\, S_1^2 + (n_2 - 1)\, S_2^2}{n_1 + n_2 - 2} = \frac{36\,(11 + 9)}{12 + 10 - 2} = 36,$$
entonces
$$t = \frac{\overline{X} - \overline{Y}}{S_p\sqrt{n_1^{-1} + n_2^{-1}}} = \frac{17{,}3 - 22{,}3}{6\sqrt{12^{-1} + 10^{-1}}} = -1{,}94.$$

De la tabla se tiene que $t_{(n_1+n_2-2),1-\alpha} = t_{(20),0{,}99} = 2{,}528$, pero como $t = -1{,}94 > -2{,}528 = -t_{(20),0{,}99}$, entonces no se rechaza H_0.

Por tanto, los datos no ofrecen suficientes evidencias que permitan afirmar que la barrena 1 es más eficiente que la barrena 2.

b) Debemos buscar o acotar el valor-p. De acuerdo a los percentiles de la distribución $t_{(20)}$ se tiene $t_{(20),0{,}95} = |1{,}725| < |1{,}94| < |2{,}086| = t_{(20),0{,}975}$. Luego, el menor valor de α para el cual el test es significativo, p, cumple $0{,}025 < p < 0{,}05$.

c) Consideremos $\Delta := \mu_1 - \mu_2$, luego

$$\begin{aligned}
\pi(\Delta) &= P\left(\text{rechazar } H_0 \,|\Delta\right) = P\left(\left.\frac{\overline{X} - \overline{Y}}{S_p\sqrt{n_1^{-1} + n_2^{-1}}} < -2{,}528 \right| \Delta\right) \\
&= P\left(\left.\frac{\overline{X} - \overline{Y} - \Delta}{S_p\sqrt{n_1^{-1} + n_2^{-1}}} < -2{,}528 - \frac{\Delta}{S_p\sqrt{n_1^{-1} + n_2^{-1}}} \right| \Delta\right) = P\left(t_{(20)} < -2{,}528 - \frac{\Delta}{2{,}57}\right).
\end{aligned}$$

De aquí se tiene que $\pi(\Delta)$ es decreciente al aumentar Δ, $\lim\limits_{\Delta \to \infty} \pi(\Delta) = 0$, $\lim\limits_{\Delta \to -\infty} \pi(\Delta) = 1$ y $\pi(0) = 0{,}01$. Por lo que el gráfico de la función de potencia se puede ver en la figura 3.2.

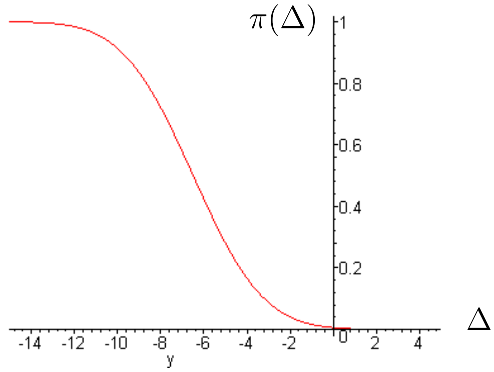

Figura 3.2: Función de potencia

Por otra parte, se tiene

$$\begin{aligned}\beta(-3) &= P(\text{error tipo II}|\,\Delta=-3) = 1-\pi(-3) = 1-P\left(t_{(20)} < -2{,}528-\frac{-3}{2{,}57}\right)\\ &\approx 1-P\left(Z<-2{,}528-\frac{-3}{2{,}57}\right) = P(Z\geq -1{,}36) = P(Z<1{,}36) = 0{,}913.\end{aligned}$$

3. Por el lema de Neyman-Pearson debemos buscar $k \geq 0$ tal que $P\left[\frac{f(X|H_1)}{f(X|H_0)} > k \middle|\, H_0\right] = \alpha$.

   a) De los datos se tiene

X	1	2	3	4	5	6	7
$\frac{f(X\|H_1)}{f(X\|H_0)}$	6	5	4	3	2	1	79/94
$P(X=x\|\,H_0)$	0.01	0.01	0.01	0.01	0.01	0.01	0.94

   Luego, debemos buscar $k \geq 0$ tal que $P\left[\frac{f(X|H_1)}{f(X|H_0)} > k \middle|\, H_0\right] = \alpha = 0{,}04$.

   Para $k=2$ se tiene $P\left[\frac{f(X|H_1)}{f(X|H_0)} > 2 \middle|\, H_0\right] = 0{,}04$. Por tanto, el test que rechaza H_0 si $\frac{f(x|H_1)}{f(x|H_0)} > 2$ es el test más potente de nivel $\alpha = 0{,}04$.

   b) El error de tipo II para el test encontrado en (a) es

$$\beta = P\left(\frac{f(X|\,H_1)}{f(X|\,H_0)} \leq 2 \middle|\, H_1\right) = P(X\in\{5,6,7\}|\,H_1) = 0{,}02+0{,}01+0{,}79 = 0{,}82.$$

4. Sea μ_e la media de la resistencia en un conductor estándar y μ_a, en un conductor aleado. Sea $\overline{X}_e$ la media muestral de la resistencia en un conductor estándar y $\overline{X}_a$, en un conductor aleado. Debemos probar las hipótesis $H_0: \mu_a \geq \mu_e$ vs. $H_1: \mu_a < \mu_e$. Pero como

$$S_p^2 = \frac{(n-1)\,S_e^2+(m-1)\,S_a^2}{m+n-2} = \frac{9\,(0{,}0009+0{,}0004)}{18} = 0{,}00065,$$

entonces

$$t = \frac{\overline{X}_e - \overline{X}_a}{S_p\sqrt{n^{-1} + m^{-1}}} = \frac{0{,}19 - 0{,}11}{0{,}018\sqrt{10^{-1} + 10^{-1}}} = 9{,}94.$$

De la tabla de distribuciones t-student se tiene

$$|t| = 9{,}94 > 1{,}734 = t_{(18),0{,}95} = t_{(m+n-2),1-\alpha}.$$

Por lo que se rechaza H_0 y se concluye que los datos muestran evidencias de que la aleación reduce la resistencia promedio del alambre.

5. Para $\mathbf{X} = (X_1, \ldots, X_n)$ y $\mathbf{Y} = (Y_1, \ldots, Y_m)$, la función de verosimilitud es

$$L(\pi_1, \pi_2 | \mathbf{X}, \mathbf{Y}) = \prod_{i=1}^{n} f(X_i | \pi_1) \prod_{j=1}^{m} f(Y_j | \pi_2) = \pi_1^n \pi_2^m (1 - \pi_1)^{n\overline{X} - n} (1 - \pi_2)^{m\overline{Y} - m},$$

y su logaritmo es

$$l(\pi_1, \pi_2) = n \ln(\pi_1) + m \ln(\pi_2) + \left(n\overline{X} - n\right) \ln(1 - \pi_1) + \left(m\overline{Y} - m\right) \ln(1 - \pi_2).$$

De

$$0 = \frac{\partial l(\pi_1, \pi_2)}{\partial \pi_1} = \frac{n}{\pi_1} - \frac{n\overline{X} - n}{1 - \pi_1}$$

se tiene $\hat{\pi}_1 = \overline{X}^{-1}$. Análogamente, se tiene que $\hat{\pi}_2 = \overline{Y}^{-1}$.

Por otra parte, bajo $H_0 : \pi_1 = \pi_2 = \pi$ se tiene

$$l(\pi) = (n + m) \ln(\pi) + \left(n\overline{X} - n + m\overline{Y} - m\right) \ln(1 - \pi).$$

Por lo que de

$$0 = \frac{dl(\pi)}{d\pi} = \frac{n + m}{\pi} - \frac{n\overline{X} - n + m\overline{Y} - m}{1 - \pi}$$

se tiene

$$\hat{\pi} = \left(\frac{n\overline{X} + m\overline{Y}}{n + m}\right)^{-1}.$$

Por tanto,

$$\Lambda(\mathbf{X}, \mathbf{Y}) = \frac{\sup\limits_{\Theta_0} L(\pi | \mathbf{X}, \mathbf{Y})}{\sup\limits_{\Theta} L(\pi_1, \pi_2 | \mathbf{X}, \mathbf{Y})} = \frac{L(\hat{\pi} | \mathbf{X}, \mathbf{Y})}{L(\hat{\pi}_1, \hat{\pi}_2 | \mathbf{X}, \mathbf{Y})} = \frac{\hat{\pi}^{n+m} (1 - \hat{\pi})^{n\overline{X} - n + m\overline{Y} - m}}{\hat{\pi}_1^n \hat{\pi}_2^m (1 - \hat{\pi}_1)^{n\overline{X} - n} (1 - \hat{\pi}_2)^{m\overline{Y} - m}}.$$

Pero como $-2\ln(\Lambda) \overset{\cdot}{\sim} \chi^2_{(2-1)}$, entonces se rechaza H_0 si

$$\begin{aligned} & 2\left[n \ln(\hat{\pi}_1) + m \ln(\hat{\pi}_2) + \left(n\overline{X} - n\right) \ln(1 - \hat{\pi}_1) + \left(m\overline{Y} - m\right) \ln(1 - \hat{\pi}_2)\right] \\ & -2\left[(n + m) \ln(\hat{\pi}) + \left(n\overline{X} - n + m\overline{Y} - m\right) \ln(1 - \hat{\pi})\right] \\ > \; & \chi^2_{(1),1-\alpha}. \end{aligned}$$

6. Sean $X_i \overset{iid}{\sim} f(x)$ las v.a. que representan los tiempos en desarrollar la I2, y se quiere comparar las hipótesis $H_0 : X_i \sim f(x)$ contra $H_1 : X_i \nsim f(x)$. De acuerdo a los datos obtenidos debemos hacer un test de bondad de ajuste en donde los valores esperados vienen dados por $E_i = np_i$, donde $p_i = \int_i kxdx$.

   Como $1 = \int_0^{120} kxdx = k\frac{120^2}{2}$, entonces $k = \frac{1}{7200}$.

   Por tanto,
$$p_1 = \frac{1}{7200}\int_0^{60} xdx = \frac{60^2}{7200\times 2} = \frac{1}{4},$$
$$p_2 = \frac{1}{7200}\int_{60}^{90} xdx = \frac{90^2-60^2}{7200\times 2} = \frac{5}{16}$$
   y
$$p_3 = \frac{1}{7200}\int_{90}^{120} xdx = \frac{120^2-90^2}{7200\times 2} = \frac{7}{16}.$$

   Finalmente,
$$\begin{aligned}\chi^2_{calc} &= \sum_{i=1}^{3}\frac{(O_i-E_i)^2}{E_i} = \sum_{i=1}^{3}\frac{(O_i-np_i)^2}{np_i}\\ &= \frac{\left(10-\frac{50}{4}\right)^2}{\frac{50}{4}} + \frac{\left(20-50\frac{5}{16}\right)^2}{50\frac{5}{16}} + \frac{\left(20-50\frac{7}{16}\right)^2}{50\frac{7}{16}} = 4{,}74.\end{aligned}$$

   Pero como $\chi^2_{calc} = 4{,}74 < 5{,}99 = \chi^2_{(2),0,95}$, no se rechaza H_0 y asumimos que los tiempos para desarrollar la I2 siguen la distribución dada.

7. La única vía que hemos visto para buscar test potentes ha sido usando el lema de Neyman-Pearson.

   a) Luego, para $\alpha > 0$ dado, debemos buscar $c \geq 0$ tal que $\alpha = P\left[\frac{f_{\mu_0}(\mathbf{X})}{f_{\mu 1}(\mathbf{X})} < c \middle| \mu = \mu_0\right]$ donde $\mu_1 > \mu_0$. De aquí se tiene
$$\begin{aligned}\alpha &= P\left[\frac{\mu_0^{-n}\exp\left(-\frac{1}{\mu_0}\sum_{i=1}^n X_i\right)}{\mu_1^{-n}\exp\left(-\frac{1}{\mu_1}\sum_{i=1}^n X_i\right)} < c \middle| \mu = \mu_0\right]\\ &= P\left[\exp\left\{\left(\frac{1}{\mu_1}-\frac{1}{\mu_0}\right)\sum_{i=1}^n X_i\right\} < c' \middle| \mu = \mu_0\right]\\ &= P\left[\left(\frac{1}{\mu_1}-\frac{1}{\mu_0}\right)\sum_{i=1}^n X_i < c'' \middle| \mu = \mu_0\right] = P\left[\sum_{i=1}^n X_i > k \middle| \mu = \mu_0\right].\end{aligned}$$
   Por tanto se rechaza H_0 si $\sum_{i=1}^n X_i > k$.

   b) Como $n = 1$, se rechaza $H_0 : \mu \leq 8$ si $X_1 > k$. Luego,
$$0{,}1 = P\left[X_1 > k \middle| \mu = 8\right] = \int_k^\infty \frac{1}{8}e^{-\frac{y}{8}}dy = e^{-\frac{k}{8}},$$
   de donde se tiene $k = 8\ln(10) \approx 18.421$.

c) Sabemos que $Potencia\,(\mu) = P\left[X_1 > k|\,\mu\right]$. Luego,

1) $Potencia\,(12) = P\left[X_1 > 8\ln(10)|\,\mu = 12\right] = \int_{8\ln(10)}^{\infty} \frac{1}{12} e^{-\frac{y}{12}} dy = \exp\left[-\frac{8\ln(10)}{12}\right] = \frac{1}{10}\sqrt[3]{10} \approx 0.215\,44.$

2) $p = P\left[X_1 > 15|\,\mu = 8\right] = \int_{15}^{\infty} \frac{1}{8} e^{-\frac{y}{8}} dy = e^{-\frac{15}{8}} \approx 0.153\,35.$

8. Como $X_i \overset{iid}{\sim} U_{[0,\theta]}$ entonces $f_{X_1}(x|\,\theta) = \theta^{-1} I_{[0,\theta]}(x)$ y $F_{X_1}(x|\,\theta) = \begin{cases} 0 \text{ si } x < 0 \\ x/\theta \text{ si } x \in [0,\theta] \\ 1 \text{ si } x > \theta \end{cases}$.

Luego,

a) $f_T(t|\,\theta) = n\left[F_{X_1}(t|\,\theta)\right]^{n-1} f_{X_1}(t|\,\theta) = \frac{n}{\theta^n} t^{n-1} I_{[0,\theta]}(x)$. Por tanto la función de potencia es
$$\pi(\theta) = P(T > t_0) = \int_{t_0}^{\theta} \frac{n}{\theta^n} t^{n-1} dt = 1 - \left(\frac{t_0}{\theta}\right)^n.$$
Pero como $\frac{d\pi(\theta)}{d\theta} = \frac{t_0^n}{\theta^{n+1}} > 0$ para todo $\theta > 0$, entonces $\pi(\theta)$ es creciente. Además,
$$\pi(\theta) = 1 - \left(\frac{t_0}{\theta}\right)^n = 1 - \left(\frac{1{,}9}{2}\right)^3 = 0{,}142625.$$

b) Como la función de potencia es creciente tenemos,
$$\alpha = \sup_{\theta \le \theta_0} \alpha(\theta) = \sup_{\theta \le \theta_0} \pi(\theta) = \pi(\theta_0) = 1 - \left(\frac{t_0}{\theta_0}\right)^n.$$
Luego, si $\alpha = 0{,}1$ se tiene $t_0 = (1-\alpha)^{\frac{1}{n}}\theta_0 = (0{,}9)^{\frac{1}{n}}\theta_0$.

c) En ambos casos consideraremos $n = 3$.

1) $Valor - p = P(T > 1{,}9|\,\theta = \theta_0) = 1 - \left(\frac{1{,}9}{\theta_0}\right)^n = 1 - \left(\frac{1{,}9}{2}\right)^3 = 0{,}142625$. Observen que coincide con el resltado del inciso (a), y esto no es por casualidad.

2) Notemos que si $\theta_0 = 2$, entonces el recorrido de la v.a. T es $[0,2]$, y por tanto
$$Valor - p = P(T > 2{,}1|\,\theta = \theta_0) = P(T > 2{,}1|\,\theta = 2) = 0.$$

9. Notemos que los siguientes resultados servirán también para comparar las hipótesis $H_0 : \pi \ge \pi_0$ vs. $H_1 : \pi < \pi_0$, y que no se utilizará el valor específico de π_1, salvo que $\pi_0 > \pi_1$.

a) Por el lema de Neyman-Pearson tenemos
$$\begin{aligned}
\alpha &= P\left[\left.\frac{\pi_0^{\sum_{i=1}^n X_i}(1-\pi_0)^{n-\sum_{i=1}^n X_i}}{\pi_1^{\sum_{i=1}^n X_i}(1-\pi_1)^{n-\sum_{i=1}^n X_i}} < c\,\right|\,\pi = \pi_0\right] \\
&= P\left\{\left.\left[\frac{\pi_0(1-\pi_1)}{\pi_1(1-\pi_0)}\right]^{\sum_{i=1}^n X_i}\left[\frac{1-\pi_0}{1-\pi_1}\right]^n < c\,\right|\,\pi = \pi_0\right\} \\
&= P\left\{\left.\sum_{i=1}^n X_i \ln\left[\frac{\pi_0(1-\pi_1)}{\pi_1(1-\pi_0)}\right] < c'\,\right|\,\pi = \pi_0\right\},
\end{aligned}$$

pero como $\pi_0 > \pi_1$ implica que $\ln\left[\frac{\pi_0(1-\pi_1)}{\pi_1(1-\pi_0)}\right] > 0$, entonces debemos buscar k tal que

$$\alpha = P\left\{\sum_{i=1}^{n} X_i < k \middle| \pi = \pi_0\right\}.$$

Luego, como $Y = \sum_{i=1}^{n} X_i \sim B(n, \pi_0)$ bajo $H_0 : \pi = \pi_0$, entonces se rechaza H_0 si,

$$\sum_{i=1}^{n} X_i < k_0,$$

donde k_0 es el mayor entero k tal que

$$\sum_{y=0}^{k-1} \binom{n}{y} \pi_0^y (1-\pi_0)^{n-y} \leq \alpha.$$

b) Por el TLC se tiene que

$$\frac{(\hat{p} - \pi_0)\sqrt{n}}{\sqrt{\pi_0(1-\pi_0)}} \sim N(0,1)$$

bajo H_0, donde $\hat{p} = \overline{X}$. Luego, se rechaza H_0 si

$$\frac{(\hat{p} - \pi_0)\sqrt{n}}{\sqrt{\pi_0(1-\pi_0)}} < z_\alpha = -z_{1-\alpha}.$$

10. Estamos suponiendo que $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu_1, \sigma^2)$, $Y_1, \ldots, Y_n \overset{iid}{\sim} N(\mu_2, \sigma^2)$, $\sigma = 10$ y que $X_i \perp\!\!\!\perp Y_j$ para todo $i, j = 1, \ldots, n$.

a) Sabemos que un I.C. para $\mu_1 - \mu_2$ al 95 % es $\overline{X} - \overline{Y} \pm 1{,}96\frac{10}{\sqrt{n}}$, cuya longitud debe ser igual a 2, o sea, $2 \cdot 1{,}96\frac{10}{\sqrt{n}} = 2$, de donde se tiene $n = 384.16$. Luego, para asegurar la confiabilidad deseada y tener un valor entero de n, debe tomarse una muestra de tamaño 385.

b) Sabemos que el test rechaza $H_0 : \mu_1 = \mu_2$ en favor de $H_1 : \mu_1 > \mu_2$ si $\frac{(\overline{X}-\overline{Y})\sqrt{n}}{\sigma\sqrt{2}} > z_{1-\alpha}$. Pero como

$$\begin{aligned}
Potencia\,(\Delta = \mu_1 - \mu_2) &= P\left(\overline{X} - \overline{Y} > z_{1-\alpha}\sigma\sqrt{\frac{2}{n}} \middle| \mu_1 - \mu_2 = \Delta\right) \\
&= P\left(\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{2}{n}}} > z_{1-\alpha} - \frac{(\mu_1 - \mu_2)\sqrt{n}}{\sigma\sqrt{2}} \middle| \mu_1 - \mu_2 = \Delta\right) \\
&= P\left(Z > z_{1-\alpha} - \frac{\Delta\sqrt{n}}{\sigma\sqrt{2}}\right),
\end{aligned}$$

donde $Z = \frac{\left(\overline{X}-\overline{Y}\right)-\Delta}{\sigma\sqrt{\frac{2}{n}}}$ distribuye $N\left(0,1\right)$ bajo $H_1 : \mu_1 - \mu_2 = \Delta > 0$. De lo anterior tenemos que

$$P\left(Z \leq z_{1-\alpha} - \frac{\Delta\sqrt{n}}{\sigma\sqrt{2}}\right) = 1 - Potencia\left(\Delta\right).$$

Por tanto,

$$z_{1-Potencia(\Delta)} = z_{1-\alpha} - \frac{\Delta\sqrt{n}}{\sigma\sqrt{2}}.$$

Del problema tenemos, $\Delta = 2$, $Potencia\left(2\right) = 0{,}5$, $\alpha = 0{,}1$ y $\sigma = 10$. Por tanto,

$$z_{0,5} = z_{0,9} - \frac{2\sqrt{n}}{10\sqrt{2}},$$

y de aquí, $0 = 1{,}28 - \frac{2\sqrt{n}}{10\sqrt{2}}$, lo que implica $n = 81.92$. Pero como debemos tomar un tamaño de muestra entero y que cumpla las exigencias pedidas entonces n debe ser igual a 82.

Apéndice A

Formulario de Distribuciones

	$P(X=x) \quad \mid \quad f_X(x)$	$E(X)$	$V(X)$	$M_X(t)$	$R_X(x)$
$X \sim \text{Bernoulli}(p)$	$p^x(1-p)^{1-x}$	p	$p(1-p)$	$(1-p)+pe^t$	$x=0,1.$
$X \sim \text{Binomial}(n,p)$	$\binom{n}{x}p^x(1-p)^{n-x}$	np	$np(1-p)$	$((1-p)+pe^t)^n$	$x=0,1,\ldots,n$
$X \sim \text{Geom}(p)$	$p(1-p)^{x-1}$	$\frac{1}{p}$	$\frac{(1-p)}{p^2}$	$\frac{pet}{1-(1-p)e^t}$, si $(1-p)e^t<1$	$x=1,2,\ldots$
$X \sim \text{BN}(r,p)$	$\binom{x-1}{r-1}p^r(1-p)^{x-r}$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$	$\left(\frac{e^tp}{1-(1-p)e^t}\right)^r$	$x=r,r+1,r+2,\ldots$
$X \sim \text{Hiper}(M,N,n)$	$\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$	np	$n\,\frac{M}{N}\,(\frac{N-M}{N})(\frac{N-n}{N-1})$		$x=0,1,\ldots,\text{mín}(M,n)$
$X \sim \text{Poisson}(\mu)$	$\frac{\mu^x e^{-\mu}}{x!}$	μ	μ	$e^{\mu(e^t-1)}$	$x=0,1,\ldots$
$X \sim \text{Unif}(a,b)$	$\frac{1}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{tb}-e^{at}}{t(b-a)}$, $t\neq 0$	$a\leq x\leq b$
$X \sim \text{Exp}(\lambda)$	$\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda-t}$,$t<\lambda$	$x>0$
$X \sim \text{Normal}(\mu,\sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2	$e^{\mu t+\frac{\sigma^2t^2}{2}}$	$-\infty<x<\infty$
$X \sim \text{Gamma}(\alpha,\beta)$	$\frac{1}{\Gamma(\alpha)}\beta^\alpha x^{\alpha-1}e^{-\beta x}$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$	$\left(\frac{\beta}{\beta-t}\right)^\alpha$, $t<\beta$	$x>0$
$X \sim \text{Beta}(\alpha,\beta)$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$		$0\leq x\leq 1$

Apéndice B

Formulario de Análisis de Regresión Simple

1. **Modelo de Regresión Estimado**

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

$$\hat{\beta}_1 = \frac{\displaystyle\sum_{i=1}^{n} x_i y_i - \frac{\displaystyle\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}}{\displaystyle\sum_{i=1}^{n} x_i^2 - \frac{\left(\displaystyle\sum_{i=1}^{n} x_i\right)^2}{n}}$$

2. **Suma de cuadrados**

a) $S_{xx} = \displaystyle\sum_{i=1}^{n} (x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{\left(\displaystyle\sum_{i=1}^{n} x_i\right)^2}{n}.$

b) $S_{yy} = \displaystyle\sum_{i=1}^{n} (y_i - \overline{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{\left(\displaystyle\sum_{i=1}^{n} y_i\right)^2}{n}.$

c) $S_{xy} = \displaystyle\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=1}^{n} x_i y_i - \frac{\displaystyle\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}$

d) $SSE = S_{yy} - \hat{\beta}_1 S_{xy}$.

e) $SSR = \hat{\beta}_1 S_{xy}$.

f) $SST = SSR + SSE = S_{yy}$

3. **Varianzas y Desviaciones Estándar**

   a) $\hat{\sigma}^2 = \frac{SSE}{n-2}$

   b) $se(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{\overline{x}^2}{S_{xx}}\right)}$

   c) $se(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}$

4. **Test de Hipótesis para los coeficientes**

   a) $H_0 : \beta_0 = 0 \; H_1 : \beta_0 \neq 0$

$$T_0 = \frac{\hat{\beta}_0}{se(\hat{\beta}_0)}$$

   b) $H_0 : \beta_1 = 0 \; H_1 : \beta_1 \neq 0$

$$T_1 = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$$

   En ambos caso se rechaza la hipótesis nula si $|T_i| > t_{n-2,1-\alpha/2}$

5. **Intervalos de Confianza**

   a) Intervalos de Confianza para los coeficientes

$$IC(\beta_0) = \hat{\beta}_0 \mp t_{n-2,1-\alpha/2} \cdot se(\hat{\beta}_0)$$

$$IC(\beta_1) = \hat{\beta}_1 \mp t_{n-2,1-\alpha/2} \cdot se(\hat{\beta}_1)$$

   b) Intervalo de Confianza para la Predicción y_0 en el valor x_0, donde $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$

$$IC(y_0) = \hat{y}_0 \mp t_{n-2,1-\alpha/2} \sqrt{\hat{\sigma}^2 \left\{1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}\right\}}$$

c) Intervalo de Confianza para la respuesta media, donde $\hat{\mu}_{y|x_0} = \hat{\beta}_0 + \hat{\beta}_1 x_0$

$$IC(\mu_{y|x_0}) = \hat{\mu}_{y|x_0} \mp t_{n-2,1-\alpha/2} \sqrt{\hat{\sigma}^2 \left\{ \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}} \right\}}$$

6. **Coeficiente de Determinación** R^2

$$R^2 = \hat{\beta}_1 \frac{S_{xy}}{S_{yy}} = 1 - \frac{SSE}{S_{yy}}$$

Apéndice C

Tablas de distribución

C.1. Distribución t de Student

	Magnitud de α en una cola							
gl	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.0005**
1	1.38	1.96	3.08	6.31	12.71	31.82	63.66	636.58
2	1.06	1.39	1.89	2.92	4.30	6.96	9.92	31.60
3	0.98	1.25	1.64	2.35	3.18	4.54	5.84	12.92
4	0.94	1.19	1.53	2.13	2.78	3.75	4.60	8.61
5	0.92	1.16	1.48	2.02	2.57	3.36	4.03	6.87
6	0.91	1.13	1.44	1.94	2.45	3.14	3.71	5.96
7	0.90	1.12	1.41	1.89	2.36	3.00	3.50	5.41
8	0.89	1.11	1.40	1.86	2.31	2.90	3.36	5.04
9	0.88	1.10	1.38	1.83	2.26	2.82	3.25	4.78
10	0.88	1.09	1.37	1.81	2.23	2.76	3.17	4.59
11	0.88	1.09	1.36	1.80	2.20	2.72	3.11	4.44
12	0.87	1.08	1.36	1.78	2.18	2.68	3.05	4.32
13	0.87	1.08	1.35	1.77	2.16	2.65	3.01	4.22
14	0.87	1.08	1.35	1.76	2.14	2.62	2.98	4.14
15	0.87	1.07	1.34	1.75	2.13	2.60	2.95	4.07
16	0.86	1.07	1.34	1.75	2.12	2.58	2.92	4.01
17	0.86	1.07	1.33	1.74	2.11	2.57	2.90	3.97
18	0.86	1.07	1.33	1.73	2.10	2.55	2.88	3.92
19	0.86	1.07	1.33	1.73	2.09	2.54	2.86	3.88
20	0.86	1.06	1.33	1.72	2.09	2.53	2.85	3.85
21	0.86	1.06	1.32	1.72	2.08	2.52	2.83	3.82
22	0.86	1.06	1.32	1.72	2.07	2.51	2.82	3.79
23	0.86	1.06	1.32	1.71	2.07	2.50	2.81	3.77
24	0.86	1.06	1.32	1.71	2.06	2.49	2.80	3.75
25	0.86	1.06	1.32	1.71	2.06	2.49	2.79	3.73
26	0.86	1.06	1.31	1.71	2.06	2.48	2.78	3.71
27	0.86	1.06	1.31	1.70	2.05	2.47	2.77	3.69
28	0.85	1.06	1.31	1.70	2.05	2.47	2.76	3.67
29	0.85	1.06	1.31	1.70	2.05	2.46	2.76	3.66
30	0.85	1.05	1.31	1.70	2.04	2.46	2.75	3.65
∞	0.84	1.04	1.28	1.64	1.96	2.33	2.58	3.29

C.2. Distribución χ^2

	Proporción del Area hasta $+\infty$						
gl	**0.995**	**0.99**	**0.975**	**0.95**	**0.90**	**0.75**	**0.50**
1	0.00	0.00	0.00	0.00	0.02	0.10	0.45
2	0.01	0.02	0.05	0.10	0.21	0.58	1.39
3	0.07	0.11	0.22	0.35	0.58	1.21	2.37
4	0.21	0.30	0.48	0.71	1.06	1.92	3.36
5	0.41	0.55	0.83	1.15	1.61	2.67	4.35
6	0.68	0.87	1.24	1.64	2.20	3.45	5.35
7	0.99	1.24	1.69	2.17	2.83	4.25	6.35
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34
13	3.57	4.11	5.01	5.89	7.04	9.30	12.34
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34

	Proporción del Area hasta $+\infty$						
gl	**0.25**	**0.10**	**0.05**	**0.03**	**0.01**	**0.005**	**0.001**
1	1.32	2.71	3.84	5.02	6.63	7.88	10.83
2	2.77	4.61	5.99	7.38	9.21	10.60	13.82
3	4.11	6.25	7.81	9.35	11.34	12.84	16.27
4	5.39	7.78	9.49	11.14	13.28	14.86	18.47
5	6.63	9.24	11.07	12.83	15.09	16.75	20.51
6	7.84	10.64	12.59	14.45	16.81	18.55	22.46
7	9.04	12.02	14.07	16.01	18.48	20.28	24.32
8	10.22	13.36	15.51	17.53	20.09	21.95	26.12
9	11.39	14.68	16.92	19.02	21.67	23.59	27.88
10	12.55	15.99	18.31	20.48	23.21	25.19	29.59
11	13.70	17.28	19.68	21.92	24.73	26.76	31.26
12	14.85	18.55	21.03	23.34	26.22	28.30	32.91
13	15.98	19.81	22.36	24.74	27.69	29.82	34.53
14	17.12	21.06	23.68	26.12	29.14	31.32	36.12
15	18.25	22.31	25.00	27.49	30.58	32.80	37.70
16	19.37	23.54	26.30	28.85	32.00	34.27	39.25
17	20.49	24.77	27.59	30.19	33.41	35.72	40.79
18	21.60	25.99	28.87	31.53	34.81	37.16	42.31
19	22.72	27.20	30.14	32.85	36.19	38.58	43.82

C.3. Distribución F ($\alpha = 0{,}05$)

Grados de libertad denominador	Grados de libertad para el numerador									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	161	199	216	225	230	234	237	239	241	242
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

(Continuación)

Grados de libertad denominador	Grados de libertad para el numerador								
	12	**15**	**20**	**24**	**30**	**40**	**60**	**120**	∞
1	244	246	248	249	250	251	252	253	254
2	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5
3	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
30	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

C.4. Distribución Normal

	Segunda cifra decimal en z									
z	0.0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998