Estimación Puntual

Jaime Vázquez • Lizbeth Naranjo • Ruth Fuentes • Margarita Chávez

Proyecto PAPIME UNAM PE107117
"Estadística para estudiantes de ciencias"

Índice general

Introducción

La estadística inferencial es una disciplina que se basa en gran medida en la probabilidad y que ayuda a resolver problemas mediante inferencias de alguna característica de la población usando datos muestrales de la misma.

La estadística involucra conceptos y resultados que pueden resumirse en grandes temas: análisis exploratorio de datos, distribuciones muestrales, estimación puntual, estimación por intervalo y pruebas de hipótesis, los cuales son fundamentales en el estudio y la aplicación de esta disciplina. En esta parte se abordarán los tópicos relacionados con estimación puntual.

Se inicia con la exposición de los métodos de estimación más importantes, tales como el de máxima verosimilitud, el de momentos y otros como el de medianas y percentiles. Asímismo, se da una introducción al método Bayesiano y más adelante al de mínimos cuadrados.

Posteriormente se revisan las propiedades deseables de un estimador puntual como una forma de analizar su bondad. Se habla del error cuadrático medio, estimadores insesgados y la propiedad de consistencia, para dar lugar al ulterior desarrollo de la teoría para encontrar a los mejores estimadores insesgados.

Para la lectura de este documento, es importante contar con conocimientos de teoría de la probabilidad, así como de cálculo diferencial e integral en una y varias variables. También se recomienda el estudio previo del Capítulo 3 de las notas *Introducción a la Estadística* de los mismos autores o algún texto que cubra los temas relacionados con estadísticas y distribuciones muestrales, así como los de suficiencia y completez.

CAPÍTULO 1

Estimación puntual

Suponga que se dispone de una población en la que se estudia una variable aleatoria X con distribución conocida y parámetro (o parámetros) desconocido(s) y se tiene interés en disminuir el grado de desconocimiento de θ en $f\,(x;\theta)$. De la población se extrae una muestra aleatoria simple de tamaño n: $X_1, X_2, ..., X_n$ y se trata de calcular, a partir de los valores muestrales, una función de los mismos que proporcione una estadística $\widehat{\theta} = T\,(X_1, X_2, ..., X_n)$ que le asigne un valor al parámetro desconocido de la población, de forma que sean lo más cercanos en algún sentido. A $\widehat{\theta}$ se le llama *estimador*.

El objetivo de la *estimación puntual* es entonces encontrar un valor para θ, denotado como $\widehat{\theta}$, que sea función de la muestra aleatoria y que permita modelar o describir de manera adecuada el fenómeno aleatorio.

Definición 1.1 *Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución con función de densidad $f(x;\theta)$. Un estimador es una estadística $T(\underline{X})$ cuyos valores $t(\underline{x})$ sirven para aproximar o estimar los valores de θ.*

La notación $\hat{\theta} = T(\underline{X})$ expresa que el estimador de θ es la estadística $T(\underline{X})$. Los valores del estimador, o sea $t(\underline{x})$, son realizaciones de la variable aleatoria $T(\underline{X})$.

Si por ejemplo, se tiene una población $N(\mu, \sigma^2)$, un posible estimador para μ es $\hat{\mu} = \bar{X}$, es decir, en este caso el estimador de μ sería la estadística $\overline{X}$ (la media muestral). En los siguientes párrafos se presentarán los métodos más conocidos para encontrar estimadores.

En ocasiones, en lugar del parámetro θ, se desea estimar una función de dicho parámetro. En general, se denotará como $\tau\,(\theta)$ a cualquier función de θ.

1.1. Métodos de estimación

El primero de los métodos que se abordará fue aportación de Karl Pearson (1857-1936) y se conoce como el método de momentos para la estimación de parámetros.

Figura 1.1: Karl Pearson con Francis Galton. Ambos fundaron la revista *Biometrika* en 1901. Imagen tomada de commons.wikipedia.org (public domain).

Karl Pearson "fue historiador, escribió sobre folklore, fue socialista convencido, abogado, matemático aplicado, biómetra, estadístico, maestro y biógrafo. Sin duda que su contribución más importante es el nacimiento de la estadística aplicada. Es por lo que se le debe mayor crédito, en frase de él mismo: *Hasta que los fenómenos de cualquier rama del conocimiento no hayan sido sometidos a medida y número, no se puede decir que se trate de una ciencia.* Además del método de momentos para la obtención de estimadores, introdujo el sistema de curvas de frecuencias para disponer de distribuciones que pudieran aplicarse a los distintos fenómenos aleatorios, desarrolló la correlación lineal para aplicarla a la teoría de la herencia y de la evolución. Introdujo el método de la ji cuadrada para dar una medida del ajuste entre datos y distribuciones, para contrastar la homogeneidad entre varias muestras y la independencia entre variables. Fundó los Anales de la Eugenesia y en 1900, junto con Galton y Weldon, fundó la revista Biometrika de la que fue editor hasta su muerte. En una descripción autobiográfica decía: *una explicación para mi vida, se debe a una combinación de dos características que he heredado: capacidad para trabajar mucho y capacidad para relacionar las observaciones de los demás*"[1].

[1] Gómez Villegas, M.A. (2009). Karl Pearson, el Creador de la Estadística Matemática. *Historia de la Probabilidad y la Estadística IV, J. Basulto y J.J. García (eds.).* Congreso Internacional de Historia de la Estadística y la Probabilidad, 351-356.

También se introducirá el método de máxima verosimilitud propuesto por Ronald A. Fisher en 1922, y que intuitivamente pretende obtener el estimador de un parámetro seleccionando el que maximiza la probabilidad de obtener los datos que realmente fueron observados.

Figura 1.2: Placa *English Heritage* dedicada a Fisher en la casa Inverforth. Imagen tomada de commons.wikipedia.org (public domain). By AnemoneProjectors (talk) (Flickr) (Sir Ronald Aylmer Fisher plaque) [CC BY-SA 2.0 (https://creativecommons.org/licenses/by-sa/2.0)], via Wikimedia Commons

Hablar de Fisher, equivale a referirse al desarrollo de la estadística durante el siglo XX. Basta decir que la mayor parte de los términos que se usan en Inferencia Estadística los introdujo él, por ejemplo, parámetro, estadística (función de la muestra aleatoria), verosimilitud, score o puntaje, estadística auxiliar, información, hipótesis nula y errores tipo I y II, sólo por mencionar algunos.

Sin duda que el trabajo de Fisher es la piedra angular sobre la que se sustenta la estadística como ciencia. Egon Pearson (1974), hijo de Karl Pearson, habla de las diferencias conceptuales entre su padre y Fisher[2]: Galton y K. Pearson trabajaron con muestras grandes por su interés en la reproducción libre de las especies en su medio natural, esto ocurre con humanos, animales y plantas. Por su parte, Fisher trabajó con muestras pequeñas relacionadas con datos experimentales, por lo que era necesario analizar con cuidado las bases de la inferencia estadística para una adecuada interpretación. Fisher estudió resultados exactos en muestras pequeñas, pero también obtuvo propiedades asintónticas óptimas de los estimadores máximo verosímiles.

En esta parte se hablará además del enfoque Bayesiano en la teoría de la estimación puntual, el cual se basa en el teorema de Bayes.

[2]Pearson, E. S. (1974). "Memories on the impact of Fisher's work in the 1920's". *Int. Stat. Rev. 42* (1)

1.1.1. Método de momentos

Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución con función de densidad $f(x;\theta)$. A $E(X_i^r)$ se le conoce como el *r-ésimo momento poblacional* y se denota por μ_r, mientras que $\frac{\sum_{i=1}^n X_i^r}{n}$ es el *r-ésimo momento muestral* y se denota por M_r.

El método de estimación por momentos consiste en igualar los momentos muestrales con los momentos poblacionales y resolver para θ (o $\theta_1, ..., \theta_k$, si la distribución tiene k parámetros). Esto es, $\mu_r = M_r$, donde $r = 1, ..., k$ y k representa el número de parámetros a estimar.

De manera general, si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria de una población con función de densidad $f(x;\theta_1, \theta_2, ..., \theta_k)$, en la estimación por momentos se resuelve el siguiente sistema de ecuaciones

$$\underbrace{\mu_1 = M_1, \quad \mu_2 = M_2, \ldots, \mu_k = M_k}_{k \text{ ecuaciones con } k \text{ incógnitas}}$$

La solución a este sistema $\hat{\underline{\theta}} = (\hat{\theta}_1, ..., \hat{\theta}_k)$ se conoce como el *estimador por el método de momentos.*

Ejemplo 1.1 *Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con distribución $Poisson(\theta)$. Como*

$$\mathbb{E}(X) = \theta,$$

entonces

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

Ejemplo 1.2 *Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con distribución $exp(\theta)$. Como*

$$\mathbb{E}(X) = \frac{1}{\theta},$$

entonces

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

Por lo tanto,

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

Ejemplo 1.3 *Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con distribución $N(\mu, \sigma^2)$. Como*

$$\mathbb{E}(X) = \mu \text{ y } Var(X) = \sigma^2,$$

entonces

$$\mathbb{E}(X^2) = \mu^2 + \sigma^2.$$

Por lo tanto,

$$\bar{X} = \hat{\mu} \text{ } y \text{ } \bar{X}^2 + \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

Es decir,

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

Pero note que

$$\begin{aligned}
\sum_{i=1}^{n}(X_i - \bar{X})^2 &= \sum_{i=1}^{n}(X_i^2 - 2\bar{X}X_i + \bar{X}^2) \\
&= \sum_{i=1}^{n} X_i^2 - 2\bar{X}\sum_{i=1}^{n} X_i + n\bar{X}^2 \\
&= \sum_{i=1}^{n} X_i^2 - 2n\bar{X}^2 + n\bar{X}^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2.
\end{aligned}$$

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

Entonces los estimadores por momentos para μ y σ^2 son

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

Ejemplo 1.4 *Suponga que se tiene la siguiente muestra de tamaño* 10 :

$$1, 1, 1, 2, 2, 3, 5, 7, 8, 10.$$

Estimar los parámetros μ y σ^2 usando el método de momentos si la distribución normal se ajusta a través de los datos de la muestra.

En este caso, las estadísticas muestrales están dadas por

$$\sum_{i=1}^{10} x_i = 40 \text{ } y \text{ } \sum_{i=1}^{10} x_i^2 = 258.$$

Usando el método de momentos y el ejemplo anterior:

$$\widehat{\mu} = \frac{40}{10} = 4$$

y

$$\widehat{\sigma^2} + 4^2 = \frac{258}{10}.$$

De donde

$$\widehat{\sigma^2} = 9.8.$$

Ejemplo 1.5 *Hallar los estimadores por el método de momentos de la distribución Gamma y usar los datos del ejemplo anterior para dar valores numéricos de $\widehat{r}$ y $\widehat{\lambda}$. Igualando los primeros momentos muestrales y poblacionales, se obtiene:*

$$\mathbb{E}(X) = \frac{r}{\lambda} = \frac{40}{10} = 4. \tag{1.1}$$

y

$$\mathbb{E}\left(X^2\right) = \frac{r(r+1)}{\lambda^2} = \frac{258}{10} = 25.8.$$

Se tiene un sistema de dos ecuaciones. Para encontrar la solución, se considera el cociente:

$$\frac{\mathbb{E}(X^2)}{[\mathbb{E}(X)]^2} = \frac{\widehat{r}+1}{\widehat{r}} = \frac{25.8}{4^2} = 1.6125.$$

y resolviendo para r :

$$\widehat{r} = 1.6327.$$

Sustituyendo este valor en (1.1) y despejando λ, se obtiene:

$$\widehat{\lambda} = \frac{1.6327}{4} = 0.408\,2.$$

El rango del estimador no necesariamente coincide con el espacio del parámetro a estimar. Considere en el siguiente ejemplo una distribución Binomial con k y p desconocidos.

Ejemplo 1.6 *Suponga que X ~Binomial(k,p). Una posible aplicación con esta distribución es que se busque estimar las tasas de ocurrencia de un crimen, conociendo que existe un número importante de casos que no se denuncian o reportan, es decir, no se sabe cuál es exactamente el valor de k. Utilizando el método de momentos, se consideran los dos primeros momentos poblacionales de la distribución binomial, así como los correspondientes momentos muestrales:*

$$\bar{X}_n = kp \tag{1.2}$$

y

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = kp(1-p) + k^2p^2. \tag{1.3}$$

De (1.2) se obtiene

$$\hat{p} = \frac{\bar{X}_n}{k}, \tag{1.4}$$

sustituyendo este resultado en lugar de p en (1.3):

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = \bar{X}_n(1 - \frac{\bar{X}_n}{k}) + \bar{X}_n{}^2 \tag{1.5}$$

y como

$$\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 + \bar{X}_n{}^2,$$

(1.5) es equivalente a

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 &= \bar{X}_n(1 - \frac{\bar{X}_n}{k}) \\ &= \bar{X}_n - \frac{\bar{X}_n{}^2}{k},\end{aligned}$$

así que despejando el valor de k, se obtiene

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

Por lo tanto, el valor estimado de p es

$$\hat{p} = \frac{\bar{X}_n}{\hat{k}}.$$

Observe que no se puede garantizar que k será un número entero positivo y que $p \in [0, 1]$. Sin embargo, en general el método permite obtener una propuesta para estimar rápidamente los parámetros desconocidos.

1.1.2. Estimadores basados en verosimilitud

Para introducir este método se presenta primero el siguiente ejemplo, tomado de Mood, Graybill y Boes (1973), el cual considera que se tiene una urna en donde se sabe que hay bolas negras y blancas, de las cuales se desconoce el número de cada una de ellas, pero se tiene el dato de que la razón es de 3 a 1, aunque también se ignora si hay más bolas blancas que negras o viceversa. Se desea estimar la proporción de bolas negras en la urna y para ello, se toma una muestra de tamaño 3 con reemplazo de esta población, X_1, X_2, X_3. Note que X_i tiene distribución $Bernoulli(\theta)$, $i = 1, 2, 3$.

Observe que:

- θ es la probabilidad de obtener bola negra, es decir, $\theta = \mathbb{P}(X_i = 1)$.
- θ solo puede tomar los valores $\frac{1}{4}$ y $\frac{3}{4}$, debido a que la razón establecida es de 3 a 1.
- $X := \sum_{i=1}^{3} X_i$ es el número de bolas negras en la muestra. Y por lo tanto, X puede tomar los valores $x = 0, 1, 2, 3$.
- $X \sim Bin(n = 3, \theta)$.

En el cuadro 1.1 se presentan las probabilidades de obtener $0, 1, 2$ y 3 bolas negras, con ambos valores del parámetro. Si en la muestra se obtienen 0 bolas negras, es decir

Valor de θ	$P(X=0)$	$P(X=1)$	$P(X=2)$	$P(X=3)$
$\theta = 1/4$	$27/64$	$27/64$	$9/64$	$1/64$
$\theta = 3/4$	$1/64$	$9/64$	$27/64$	$27/64$

Cuadro 1.1: Probabilidad de obtener $0, 1, 2$ y 3 bolas negras cuando $\theta = \frac{1}{4}$ y $\theta = \frac{3}{4}$.

$$(x_1 = 0, x_2 = 0, x_3 = 0)\,,$$

entonces $\hat{\theta} = 1/4$ porque es más probable obtener (con este valor de θ) 0 bolas negras que con $\hat{\theta} = 3/4$. Ahora, si en la muestra se obtienen 2 bolas negras entonces $\hat{\theta} = 3/4$, ya que es más probable obtener 2 bolas negras con $\hat{\theta} = 3/4$ que con $\hat{\theta} = 1/4$. Es decir, se escoge el valor de θ que maximiza la probabilidad de obtener una muestra específica (x_1, x_2, x_3).

Esta es la idea subyacente en el método de máxima verosimilitud para estimar parámetros. De manera general, es necesario definir una función que represente la "probabilidad"de obtener una cierta muestra de entre todas las posibles en $\mathfrak{X}$ (el espacio muestral o de las muestras). Dicha función, para un valor muestral fijo, depende únicamente de los parámetros de la distribución en cuestión y el problema es encontrar aquél valor del parámetro o de los parámetros que maximicen esta función para una realización fija de la muestra. En el ejemplo anterior, el parámetro sólo puede tomar dos valores, pero en general se estará resolviendo un problema de optimización sobre el espacio paramétrico correspondiente a la distribución con la que se esté trabajando (una vez que se ha observado una muestra).

Para abordar este tema, se iniciará con la definición de la función de verosimilitud.

Definición 1.2 *Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con función de densidad $f(x;\theta)$. Se define la* ***función de verosimilitud*** *como la función de densidad conjunta de la muestra y se denota como $L(\theta)$ o $L(\theta \mid \underline{x})$. Es decir:*

$$L(\theta) = f_{X_1,\ldots,X_n}(x_1,...,x_n;\theta) = \prod_{i=1}^{n} f_{X_i}(x_i;\theta).$$

Definición 1.3 *Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con función de densidad $f(x;\theta)$ y $L(\theta)$ la correspondiente función de verosimilitud. A $\hat{\theta} = T(\underline{X})$ se le llama el estimador máximo verosímil de θ, si satisface que para cualquier $\theta \in \Theta$, se tiene que $L(\hat{\theta}) \geq L(\theta)$.*

Método general

Sea $f(x;\theta_1,\theta_2,\ldots,\theta_k)$ una función de densidad con k parámetros. Si $(\hat{\theta_1},...\hat{\theta_k})$ satisface el sistema

$$\frac{\partial L(\theta_1,\theta_2,...,\theta_k)}{\partial \theta_i} = 0 \quad i = 1,2,...,k;$$

entonces $(\hat{\theta_1},...,\hat{\theta_k})$ es el estimador máximo verosímil de θ.

Note que

$$\frac{\partial}{\partial \theta_i} \ln(L(\theta_1,\theta_2,\ldots,\theta_k)) = \frac{1}{L(\theta_1,\theta_2,\ldots,\theta_k)} \cdot \frac{\partial}{\partial \theta_i} L(\theta_1,\theta_2,\ldots,\theta_k).$$

Por lo tanto,

$$\frac{\partial}{\partial \theta_i} \ln(L(\theta_1,\theta_2,\ldots,\theta_k)) = 0 \Leftrightarrow \frac{\partial}{\partial \theta_i} L(\theta_1,\theta_2,\ldots,\theta_k) = 0.$$

Es decir, $\ln(L(\theta_1,\theta_2,\ldots,\theta_k))$ alcanza su máximo en el mismo punto que $L(\theta_1,\theta_2,\ldots,\theta_k)$.

En virtud de la observación anterior se define la *log-verosimilitud* de $f(x;\theta_1,\theta_2,\ldots,\theta_k)$ como

$$l(\theta_1,\theta_2,\ldots,\theta_k) = \ln(L(\theta_1,\theta_2,\ldots,\theta_k)).$$

Frecuentemente, por practicidad, se calcula el máximo de $l(\theta)$ en vez del de $L(\theta)$.

Ejemplo 1.7 *(Distribución Bernoulli). Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con distribución Bernoulli(θ). Obtener el estimador máximo verosímil de θ.*

$$L(\theta) = \prod_{i=1}^{n} f(x_i,\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}.$$

Entonces,

$$\begin{aligned} l(\theta) &= \ln(L(\theta)) = \ln\left(\theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}\right) \\ &= \ln(\theta)\sum_{i=1}^n x_i + \ln(1-\theta)\left(n-\sum_{i=1}^n x_i\right). \end{aligned}$$

Luego,

$$\frac{\partial}{\partial\theta}l(\theta) = \frac{\sum_{i=1}^n x_i}{\theta} - \frac{n-\sum_{i=1}^n x_i}{1-\theta}.$$

Por lo tanto,

$$\frac{\partial}{\partial\theta}l(\theta) = 0 \Leftrightarrow \frac{\sum_{i=1}^n x_i}{\widehat{\theta}} = \frac{n-\sum_{i=1}^n x_i}{1-\widehat{\theta}} \Leftrightarrow \frac{1}{\widehat{\theta}} - 1 = \frac{n}{\sum_{i=1}^n x_i} - 1 \Leftrightarrow \frac{1}{\widehat{\theta}} = \frac{1}{\bar{x}},$$

de donde se concluye que

$$\widehat{\theta} = \overline{X}.$$

Ahora se verificará que es un máximo

$$\frac{\partial^2 l(\theta)}{\partial\theta^2}\mid_{\widehat{\theta}} = -\frac{\sum_{i=1}^n x_i}{\widehat{\theta}^2} - \frac{n-\sum_{i=1}^n x_i}{(1-\widehat{\theta})^2} = -\left(\frac{\sum_{i=1}^n x_i}{\widehat{\theta}^2} + \frac{n-\sum_{i=1}^n x_i}{(1-\widehat{\theta})^2}\right) < 0.$$

$\therefore$ *El estimador máximo verosímil de θ es $\hat{\theta}_{MV} = \bar{X}$.*

Ejemplo 1.8 *(Distribución exp(θ)). Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con distribución exp(θ). Obtener el estimador máximo verosímil de θ.*

$$L(\theta) = \prod_{i=1}^n f(x_i,\theta) = \prod_{i=1}^n \theta e^{-\theta x_i} I_{(0,\infty)}(x_i) = \theta^n e^{-\theta\sum_{i=1}^n x_i}\prod_{i=1}^n I_{(0,\infty)}(x_i).$$

Entonces,

$$\begin{aligned} l(\theta) &= \ln(L(\theta)) = \ln\left(\theta^n e^{-\theta\sum_{i=1}^n x_i}\prod_{i=1}^n I_{(0,\infty)}(x_i)\right) \\ &= n\ln(\theta) - \theta\sum_{i=1}^n x_i + \sum_{i=1}^n \ln(I_{(0,\infty)}(x_i)). \end{aligned}$$

Luego,

$$\frac{\partial}{\partial\theta}l(\theta) = \frac{n}{\theta} - \sum_{i=1}^n x_i.$$

Por lo tanto,

$$\frac{\partial}{\partial\theta}l(\theta) = 0 \Leftrightarrow \frac{n}{\widehat{\theta}} = \sum_{i=1}^{n} x_i \Leftrightarrow \frac{1}{\widehat{\theta}} = \frac{\sum_{i=1}^{n} x_i}{n} \Leftrightarrow \frac{1}{\widehat{\theta}} = \bar{x},$$

y así

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

Ahora se verificará que es un máximo

$$\frac{\partial^2 l(\theta)}{\partial\theta^2} \mid_{\widehat{\theta}} = -\frac{n}{\widehat{\theta}^2} < 0.$$

$$\therefore \textit{El estimador máximo verosímil de } \theta \textit{ es } \hat{\theta}_{MV} = \frac{1}{\bar{X}}.$$

Ejemplo 1.9 *(Distribución Poisson(θ)). Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con distribución Poisson(θ). Obtener el estimador máximo verosímil de θ.*

$$L(\theta) = \prod_{i=1}^{n} f(x_i, \theta) = \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{x_i}}{x_i!} I_{\{0,1,2,\ldots\}}(x_i) = e^{-n\theta}\theta^{\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} \frac{I_{\{0,1,2,\ldots\}}(x_i)}{x_i!}.$$

Entonces,

$$\begin{aligned} l(\theta) &= \ln(L(\theta)) = \ln\left(e^{-n\theta}\theta^{\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} \frac{I_{\{0,1,2,\ldots\}}(x_i)}{x_i!}\right) \\ &= -n\theta + \ln(\theta)\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \ln\left(\frac{I_{\{0,1,2,\ldots\}}(x_i)}{x_i!}\right). \end{aligned}$$

Luego,

$$\frac{\partial}{\partial\theta}l(\theta) = -n + \frac{\sum_{i=1}^{n} x_i}{\theta}.$$

Por lo tanto,

$$\frac{\partial}{\partial\theta}l(\theta) = 0 \Leftrightarrow n = \frac{\sum_{i=1}^{n} x_i}{\widehat{\theta}} \Leftrightarrow \widehat{\theta} = \frac{\sum_{i=1}^{n} x_i}{n}.$$

De esta manera,

$$\widehat{\theta} = \overline{X}.$$

Ahora se verificará que es un máximo

$$\frac{\partial^2 l(\theta)}{\partial\theta^2} \mid_{\widehat{\theta}} = -\frac{\sum_{i=1}^{n} x_i}{\widehat{\theta}^2} < 0.$$

$$\therefore \textit{El estimador máximo verosímil de } \theta \textit{ es } \hat{\theta}_{MV} = \bar{X}.$$

Ejemplo 1.10 *(Distribución normal). Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con distribución $N(\mu, \sigma^2)$. Obtener los estimadores máximo-verosímiles de μ y σ^2.*

Primero se obtiene la función de verosimilitud:

$$\begin{aligned} L\left(\mu, \sigma^2\right) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \\ &= \left[\frac{1}{2\pi\sigma^2}\right]^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}. \end{aligned}$$

La log-verosimilitud está dada por

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

Obteniendo las derivadas con respecto a los parámetros

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

Igualando a cero se obtiene

$$\sum_{i=1}^{n} x_i - n\widehat{\mu} = 0$$

y

$$-\frac{n}{2\widehat{\sigma}^2} + \frac{1}{2\widehat{\sigma}^4}\sum_{i=1}^{n}(x_i-\widehat{\mu})^2 = 0,$$

de donde

$$\widehat{\mu} = \overline{X} \tag{1.6}$$

y

$$\widehat{\sigma}^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n}. \tag{1.7}$$

Las segundas derivadas están dadas por

$$\begin{aligned} \frac{\partial^2 l}{\partial \mu^2} &= -\frac{n}{\sigma^2}, \\ \frac{\partial^2 l}{\partial \left(\sigma^2\right)^2} &= \frac{n}{2\sigma^4} - \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^6}, \\ \frac{\partial^2 l}{\partial \mu \partial \sigma^2} &= \frac{\partial^2 l}{\partial \sigma^2 \partial \mu} = -\frac{\sum_{i=1}^{n}(x_i-\mu)}{\sigma^4}. \end{aligned}$$

Sea

$$H = \begin{pmatrix} \frac{\partial^2 l}{\partial \mu^2} & \frac{\partial^2 l}{\partial \sigma^2 \partial \mu} \\ \frac{\partial^2 l}{\partial \mu \partial \sigma^2} & \frac{\partial^2 l}{\partial (\sigma^2)^2} \end{pmatrix}$$

la matriz de segundas derivadas. Observe que

$$\left.\frac{\partial^2 l}{\partial \mu^2}\right|_{(\widehat{\mu},\widehat{\sigma}^2)} = \left.-\frac{n}{\sigma^2}\right|_{\widehat{\sigma}^2} < 0, \tag{1.8}$$

mientras que

$$\begin{aligned}
\det H|_{(\widehat{\mu},\widehat{\sigma}^2)} &= \det \begin{pmatrix} -\frac{n}{\widehat{\sigma}^2} & -\frac{\sum_{i=1}^n (x_i-\overline{x})}{\widehat{\sigma}^4} \\ -\frac{\sum_{i=1}^n (x_i-\overline{x})}{\widehat{\sigma}^4} & \frac{n}{2\widehat{\sigma}^4} - \frac{\sum_{i=1}^n (x_i-\overline{x})^2}{\widehat{\sigma}^6} \end{pmatrix} \\
&= \det \begin{pmatrix} -\frac{n}{\widehat{\sigma}^2} & 0 \\ 0 & \frac{n}{2\widehat{\sigma}^4} - \frac{\sum_{i=1}^n (x_i-\overline{x})^2}{\widehat{\sigma}^6} \end{pmatrix} \\
&= -\frac{n^2}{2\widehat{\sigma}^6} + \frac{n^2 \sum_{i=1}^n (x_i - \overline{x})^2}{n\widehat{\sigma}^8} \\
&= -\frac{n^2}{2\widehat{\sigma}^6} + \frac{n^2\widehat{\sigma}^2}{\widehat{\sigma}^8} \\
&= -\frac{n^2}{2\widehat{\sigma}^6} + \frac{n^2}{\widehat{\sigma}^6} = \frac{n^2}{2\widehat{\sigma}^6} > 0.
\end{aligned} \tag{1.9}$$

Por (1.8) y (1.9), se ve que H es definida negativa y por lo tanto, el óptimo $\left(\widehat{\mu}, \widehat{\sigma}^2\right)$ es un máximo, concluyéndose que (1.6) y (1.7) son los estimadores máximo verosímiles de μ y σ^2.

Ejemplo 1.11 *(Distribución uniforme). Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con distribución Uniforme en el intervalo $[\theta - \frac{1}{2}, \theta + \frac{1}{2}]$, es decir,*

$$f(x;\theta) = I_{[\theta-\frac{1}{2},\theta+\frac{1}{2}]}(x).$$

Obtener el estimador máximo verosímil de θ.

La función de verosimilitud está dada por

$$L(\theta) = \prod_{i=1}^n I_{[\theta-\frac{1}{2},\theta+\frac{1}{2}]}(x_i) = \begin{cases} 1 & \textit{si para toda } i,\ x_i \in [\theta - \frac{1}{2}, \theta + \frac{1}{2}] \\ 0 & \textit{si para alguna } i,\ x_i \notin [\theta - \frac{1}{2}, \theta + \frac{1}{2}]. \end{cases}$$

Es decir, el máximo valor de $L(\theta)$ es 1 cuando $x_i \in [\theta - \frac{1}{2}, \theta + \frac{1}{2}]$ para toda i, lo cual ocurre si

$$y_1 \geq \theta - \frac{1}{2} \quad y \quad y_n \leq \theta + \frac{1}{2},$$

es decir, si

$$y_n - \frac{1}{2} \leq \theta \leq y_1 + \frac{1}{2}.$$

Por lo tanto,

$$L(\theta) = \begin{cases} 1 & \text{si } \theta \in \left[y_n - \frac{1}{2}, y_1 + \frac{1}{2}\right] \\ 0 & \text{en otro caso.} \end{cases}$$

Cualquier valor $\widehat{\theta}$ *de* θ *en* $\left[Y_n - \frac{1}{2}, Y_1 + \frac{1}{2}\right]$ *es un estimador máximo verosímil, por ejemplo,* $T(\underline{X}) = \frac{Y_1 + Y_n}{2}$.

Ejemplo 1.12 *(Distribución uniforme). Sea* $X_1, \ldots, X_n$ *una muestra aleatoria de una población con distribución Uniforme en el intervalo* $[0, \theta]$. *Hallar el estimador máximo verosímil de* θ.

La función de densidad está dada por

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

La función de verosimilitud es

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} f(x_i;\theta) \\ &= \begin{cases} \frac{1}{\theta^n} & \text{si } x_i \leq \theta, \text{ para toda } i \\ 0 & \text{si al menos una de las } x_i > \theta \end{cases} \\ &= \begin{cases} \frac{1}{\theta^n} & \text{si } \max\{x_1, x_2, ..., x_n\} \leq \theta \\ 0 & \text{en otro caso} \end{cases} \\ &= \begin{cases} \frac{1}{\theta^n} & \text{si } \theta \geq y_n \\ 0 & \text{si } \theta < y_n, \end{cases} \end{aligned}$$

donde $y_n = \max\{x_1, x_2, ..., x_n\}$. *Note que*

$$\frac{dL(\theta)}{d\theta} = -\frac{n}{\theta^{n+1}} < 0.$$

Así, la función de verosimilitud vale cero si $\theta < y_n$ *y* $\frac{1}{\theta^n}$ *si* $\theta \geq y_n$, *siendo en este caso una función decreciente, como se muestra en la figura 1.3.*

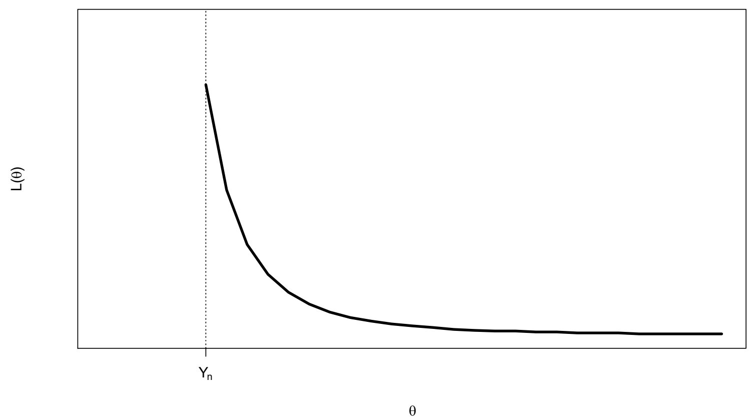

Figura 1.3: Gráfica de la función de verosimilitud para una muestra de tamaño n de la distribución uniforme continua en el intervalo $[0, \theta]$.

Así, el estimador máximo verosímil de θ *es*

$$\widehat{\theta} = Y_n = \text{máx}\{X_1, ..., X_n\}.$$

Propiedad de invarianza de los estimadores máximo-verosimiles

En algunas ocasiones, el objetivo no es estimar un parámetro de la distribución sino una función de éste, $\tau(\theta)$. Por ejemplo, el interés podría ser estimar la desviación estándar de una distribución normal, σ, en lugar de la varianza σ^2; o estimar la trasformación del momio en una distribución Bernoulli, $\theta/(1-\theta)$, en lugar de la probabilidad de éxito θ. Por lo tanto, se busca un estimador de la función $\tau(\theta)$, es decir, $\widehat{\tau(\theta)}$.

Una propiedad de los estimadores máximo verosímiles es la *propiedad de invarianza.* Esto significa que si buscamos un estimador máximo verosímil para una función de θ, denotada por $\tau(\theta)$, y si sabemos que el estimador máximo verosímil de θ es $\hat{\theta}$, entonces el estimador máximo verosímil de $\tau(\theta)$, denotado por $\widehat{\tau(\theta)}$, es $\tau(\hat{\theta})$.

Por lo tanto, la propiedad de invarianza de los estimadores máximo verosímiles enuncia que

$$\widehat{\tau(\theta)} = \tau(\widehat{\theta}),$$

es decir, que para encontrar el estimador máximo verosímil de una función del parámetro basta simplemente evaluar la función en el estimador máximo verosímil.

Para esto es necesario tomar en cuenta las características de la función $\tau(\theta)$, por ejemplo, si la función de $\tau(\theta)$ es uno-a-uno, entonces la propiedad de invarianza se cumple y existe un sólo máximo, dado que podemos invertir la función. Esto se puede ver fácilmente considerando que si $\eta = \tau(\theta)$, entonces la función inversa de $\tau(\theta)$ es $\tau^{-1}(\eta) = \theta$, la cual está bien definida por ser una función uno-a-uno, y la función de verosimilitud de $\tau(\theta)$, escrita como una función de η, está dada por

$$L^*(\eta) = \prod_{i=1}^{n} f(x_i; \tau^{-1}(\eta)) = L(\tau^{-1}(\eta)),$$

y para obtener el estimador máximo verosímil, basta obtener lo siguiente

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

Por tanto, el máximo de $L^*(\eta)$ se alcanza en $\eta = \tau(\theta) = \tau(\hat{\theta})$, mostrando así que el estimador máximo verosímil de $\tau(\theta)$ es $\tau(\hat{\theta})$.

En algunos casos, no es posible usar la propiedad de invarianza de los estimadores máximo verosímiles debido a que muchas funciones de interés no son funciones uno-a-uno. Por ejemplo, para estimar μ^2, donde μ es la media de una distribución normal, la función $\tau(\mu) = \mu^2$ no es una función uno-a-uno. Si $\tau(\theta)$ no es una función uno-a-uno, entonces para algún valor η puede haber más de un valor de θ que satisfaga que $\tau(\theta) = \eta$. En estos casos, la correspondencia entre la maximización sobre η y la correspondiente maximización sobre θ deben analizarse. Por ejemplo, si $\hat{\theta}$ es el estimador máximo verosímil de θ, podría existir otro valor de θ, digamos θ_0, para el cual también se cumple que $\tau(\hat{\theta}) = \tau(\theta_0)$. Así, en el caso de que $\tau(\theta)$ no sea una función uno-a-uno, no necesariamente existirá una única solución. En estos casos será necesario usar una definición más general de la función máximo verosímil de $\tau(\theta)$. Una definición de verosimilitud más general para $\tau(\theta)$ es la siguiente.

Definición 1.4 *La función de verosimilitud inducida por $\tau(\theta)$, denotada por L^*, está dada por*

$$L^*(\eta) = \sup_{\{\theta:\tau(\theta)=\eta\}} L(\theta).$$

En este caso, el valor $\hat{\eta}$ que maximiza a la función $L^*(\eta)$ es el estimador máximo verosímil de $\eta = \tau(\theta)$. Además, puede verse a partir de las igualdades anteriores que el máximo de L^* y el máximo de L coinciden.

Teorema 1.1 *Si $\hat{\theta}$ es el estimador máximo verosímil de θ, entonces para cualquier función $\tau(\theta)$, el estimador máximo verosímil de $\tau(\theta)$ es $\tau(\hat{\theta})$ [3].*

Demostración:

[3] Esta propiedad fue demostrada por Zehna (1966) en el artículo *Invariance of Maximum Likelihood Estimators* en la revista *Annals of Mathematical Statistics.*

Sea $\hat{\eta}$ el valor que maximiza $L^*(\eta)$. Es necesario mostrar que $L^*(\hat{\eta}) = L^*(\tau(\hat{\theta}))$. Además, como se mencionó anteriormente, el máximo de L y el máximo de L^* coinciden, así que se tiene lo siguiente,

$$\begin{aligned} L^*(\hat{\eta}) &= \sup_{\eta} \sup_{\{\theta:\tau(\theta)=\eta\}} L(\theta) \\ &= \sup_{\theta} L(\theta) \\ &= L(\hat{\theta}), \end{aligned}$$

donde la primera igualdad se cumple por definición de $L^*(\eta)$ ya que $L^*(\eta) = \sup_{\{\theta:\tau(\theta)=\eta\}} L(\theta)$, la segunda igualdad se obtiene debido a que la maximización iterada es igual a la maximización no condicional sobre θ, y la última igualdad se cumple por definición de $\hat{\theta}$, ya que $\hat{\theta}$ es el estimador máximo verosímil de $L(\theta)$. Además,

$$\begin{aligned} L(\hat{\theta}) &= \sup_{\{\theta:\tau(\theta)=\tau(\hat{\theta})\}} L(\theta) \\ &= L^*(\tau(\hat{\theta})), \end{aligned}$$

donde la primera igualdad se obtiene debido a que $\hat{\theta}$ es el estimador máximo verosímil de θ, y la segunda igualdad se obtiene por la definición de $L^*(\eta)$.

Por lo tanto, se muestra que $L^*(\hat{\eta}) = L^*(\tau(\hat{\theta}))$ y que $\tau(\hat{\theta})$ es el estimador máximo verosímil de $\tau(\theta)$.

$\blacksquare$

Con este teorema es posible encontrar estimadores máximo verosímiles de funciones de parámetros que no son uno-a-uno, por ejemplo, se puede ver que el estimador máximo verosímil de μ^2, donde μ es la media de una distribución normal, es $\bar{X}^2$.

Observación 1.1 *La propiedad de invarianza de los estimadores máximo verosímiles también se cumple en el caso multivariado. La demostración del teorema anterior es válida aún si θ es un vector de parámetros. Si el estimador máximo verosímil de $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ es $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)$, y si $\tau(\boldsymbol{\theta}) = \tau(\theta_1, \ldots, \theta_k)$ es alguna función de los parámetros, entonces el estimador máximo verosímil de $\tau(\theta_1, \ldots, \theta_k)$ es $\tau(\hat{\theta}_1, \ldots, \hat{\theta}_k)$.*

Ejemplo 1.13 *Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con distribución $N(\mu, 1)$, con μ desconocido. Se busca el estimador máximo verosímil de $\tau(\mu) = \log(\mu)$. Como $\hat{\mu} = \bar{X}$ es el estimador máximo verosímil de μ, entonces por la propiedad de invarianza $\log(\bar{X})$ es el estimador máximo verosímil de $\log(\mu)$.*

Ejemplo 1.14 *Sea $X_1, \ldots, X_n$ una muestra aleatoria de la distribución normal $N(\mu, \sigma^2)$. Se sabe que el estimador máximo verosímil de μ es $\bar{X}$. Para encontrar el estimador máximo verosímil de $\tau(\mu) = sen\ (\mu)$*

$$\widehat{\tau(\mu)} = \tau(\hat{\mu}) = sen(\hat{\mu}) = sen(\bar{X}).$$

Ejemplo 1.15 *Sea $X_1, \ldots, X_n$ una muestra aleatoria de la distribución Bernoulli(θ). Se desea encontrar el estimador máximo verosimil de $\tau(\theta) = \theta(1-\theta)$. Se sabe que el estimador máximo verosímil de θ es $\hat{\theta}_{MV} = \bar{X}$. Entonces*

$$\widehat{\tau(\theta)}_{MV} = \tau(\hat{\theta}_{MV}) = \tau(\bar{X}) = \bar{X}(1-\bar{X}).$$

Ejemplo 1.16 *Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con distribución Bernoulli(θ), con θ desconocido. Se busca el estimador máximo verosímil del momio $\tau(\theta) = \frac{\theta}{(1-\theta)}$. Como $\hat{\theta} = \bar{X}$ es el estimador máximo verosímil de θ, entonces por la propiedad de invarianza $\frac{\bar{X}}{(1-\bar{X})}$ es el estimador máximo verosímil de $\frac{\theta}{(1-\theta)}$.*

1.1.3. Verosimilitud en el enfoque Bayesiano

Considere una muestra aleatoria $X_1, \ldots, X_n$ de una población con distribución gama(μ, s), con media μ. En esta sección se denotará a la función de densidad como $f_X(x|\underline{\theta})$ y a la verosimilitud como $L\left(\underline{\theta}|x_1, \ldots, x_n\right)$, la razón de este cambio se comprenderá más adelante. Así, para el caso de la distribución gama(μ, s), su función de densidad está dada por

$$f_X(x|\mu, s) = \frac{s^s}{\Gamma(s)\mu^s} x^{s-1} \exp\{-sx/\mu\},$$

mientras que la correspondiente función de verosimilitud es

$$\begin{aligned} L(\mu, s|x_1, \ldots, x_n) &= \prod_{i=1}^{n} \frac{s^s}{\Gamma(s)\mu^s} x_i^{s-1} \exp\{-sx_i/\mu\} \\ &= \frac{s^{ns}}{\Gamma^n(s)\mu^{ns}} T_1^{s-1} \exp\{-sT_2/\mu\}, \end{aligned}$$

donde $T_1 = \prod_{i=1}^{n} x_i$ y $T_2 = \sum_{i=1}^{n} x_i$. Observe que la verosimilitud depende de la muestra sólo a través de estas estadísticas suficientes T_1 y T_2. Suponga que se desea estimar el parámetro μ. Al otro parámetro que no es el que se está estimando, en este caso s, se le suele llamar un parámetro de ruido.

Utilizando exclusivamente la verosimilitud para proponer un estimador para el parámetro de interés μ y teniendo en cuenta la posibilidad de un muestreo repetido, el estimador máximo verosímil para μ es el valor del parámetro $\widehat{\mu} \in \Theta$ que maximiza la función de verosimilitud; es decir, el valor del parámetro para el cual la muestra observada es más probable.

Por construcción, el rango del estimador coincide con el espacio paramétrico, aunque hereda las dificultades de cualquier problema de optimización, a saber:

- Encontrar el máximo global y verificar que lo sea.
- Que sea sensible a cambios numéricos.

Ahora, si se consideran dos valores de μ, el cociente de verosimilitudes para el caso de la distribución gama, estaría dado por:

$$\begin{aligned}\frac{L(\mu_1, s|x_1, \ldots, x_n)}{L(\mu_2, s|x_1, \ldots, x_n)} &= \frac{\frac{s^{ns}}{\Gamma^n(s)\mu_1^{ns}} T_1^{s-1} \exp\{-sT_2/\mu_1\}}{\frac{s^{ns}}{\Gamma^n(s)\mu_2^{ns}} T_1^{s-1} \exp\{-sT_2/\mu_2\}}, \\ &= \left(\frac{\mu_1}{\mu_2}\right)^{ns} \exp\{-sT_2[1/\mu_1 - 1/\mu_2]\},\end{aligned}$$

el cual depende del valor del parámetro s. En general no es claro cómo tratar el parámetro de ruido, pero desde el enfoque clásico de maximizar la función de verosimilitud como se vio en la sección 1.1.2, simplemente se toma el valor del estimador $\widehat{s}$.

Existe otro enfoque conocido como *Inferencia Bayesiana*, en el cual se condiciona completamente en los datos observados y se concluye sobre la población en estudio basándose en:

1. La verosimilitud $L(\theta|\underline{\mathbf{x}})$, que representa la información que hay en los datos $\underline{\mathbf{x}} = (x_1, \ldots, x_n)$; y
2. $\pi(\mu)$ una distribución de probabilidad que se conoce como *distribución inicial o a priori* y que describe las ideas subjetivas que se tienen sobre el valor de μ. Estas ideas se conciben como externas a los datos y pueden ser deducidas de experiencias previas o bien de conocimiento experto.

La inferencia se expresa a través de una *distribución posterior, final o a posteriori* de los párametros que se denotará como $\pi(\theta|\underline{\mathbf{x}})$ y que se obtiene a través del teorema de Bayes:

$$\pi(\theta|\underline{\mathbf{x}}) = \frac{L(\theta|\underline{\mathbf{x}})\pi(\theta)}{\int L(\theta|\underline{\mathbf{x}})\pi(\theta)d\theta}. \tag{1.10}$$

La estadística Bayesiana (por Thomas Bayes (1702-1761), matemático del siglo XVIII), representa un enfoque diferente a la estadística inferencial clásica o frecuentista. En el enfoque Bayesiano también se supone que los datos se obtienen de una distribución perteneciente a una familia paramétrica conocida; sin embargo, a diferencia de la estadística clásica, que considera que los parámetros son fijos pero desconocidos, aquí se hace el supuesto de que son variables aleatorias.

En resumen, el enfoque denominado frecuentista no supone que hay conocimiento previo de θ. El enfoque Bayesiano, por el contrario, se basa en el supuesto de que se tiene alguna información previa acerca de θ. Esta información se expresa por medio de una distribución sobre θ, llamada distribución inicial o a priori. Aquí se supondrá que esta distribución a priori tiene una densidad $\pi(\theta)$ y puede tener distintas interpretaciones según el problema que se esté abordando, por ejemplo, que dicha distribución está sustentada en experiencias previas similares o que expresa una creencia subjetiva.

En ambos casos, la verosimilitud provee la información que hay en las observaciones y que permite evaluar y elegir un valor del parámetro sobre otros, pues en el proceso inferencial se busca obtener estimadores que concilien de la mejor manera el modelo con los datos observados. Será entonces de interés examinar la incertidumbre que hay en este proceso para elegir un buen estimador.

En el contexto Bayesiano se debe considerar la evaluación de la dependencia de las conclusiones con respecto a las distribuciones iniciales, las cuales se han dado de manera subjetiva. En muchos casos, la selección de la distribución inicial también contempla la posibilidad de calcular de forma cerrada el denominador en (1.10). Un caso particular de esta selección se da con las familias conjugadas.

Definición 1.5 *Una distribución inicial $\pi(\theta)$ es conjugada si para $\pi(\theta) \in \mathcal{P}$ y $L(\theta|\underline{\mathbf{x}}) \in \mathcal{F}$, se tiene que $\pi(\theta|\underline{\mathbf{x}}) \in \mathcal{P}$, donde $\mathcal{P}$ y $\mathcal{F}$ son familias de distribuciones.*

A continuación se da un primer ejemplo para ilustrar algunas de las funciones que se han mencionado en el enfoque Bayesiano.

Ejemplo 1.17 *Los paquetes de los llamados dulces Smarties vienen con k colores diferentes, los cuales se repiten con la misma frecuencia.*

Figura 1.4: Dulces smarties. Tomada de *pixabay.com* (imágenes gratuitas de alta calidad).

Suponga que no se conoce k y secuencialmente se examinan 3 dulces, resultando un rojo, un verde y un rojo. La densidad para X = el segundo dulce es de diferente color que el *primero, pero el tercero es del mismo color que el primero, está dada por:*

$$\begin{aligned} f(x \mid k) &= \mathbb{P}(\text{el segundo es de diferente color que el primero}) \times \\ &\quad \times \mathbb{P}(\text{el tercero es del mismo color que el primero}) \\ &= \left(\frac{k-1}{k}\right)\left(\frac{1}{k}\right) = \frac{k-1}{k^2}. \end{aligned}$$

A la luz de los datos $x =$ rojo, verde, rojo, se tiene entonces que $f(x|k) = (k-1)/k^2$. Si en lugar de 3 se examinan 4 y resulta que ese cuarto es de color naranja (con los tres primeros rojo, verde, rojo), se tiene que

$$f(x \mid k) = \frac{(k-1)(k-2)}{k^3}.$$

Ahora suponga que se tiene información a priori o se cree que el número de colores es 5,6,7 u 8, con probabilidades iniciales $\frac{1}{10}, \frac{3}{10}, \frac{3}{10}$ y $\frac{3}{10}$, respectivamente.

Para el caso de tres dulces, si $k = 5$, entonces

$$f(x|k) = (5-1)/5^2 = \frac{4}{25} = 0.16,$$

$$f(x|k)\pi(k) = (0.16)\left(\frac{1}{10}\right) = 0.016$$

y

$$\pi(k \mid x) = \frac{(0.16)\left(\frac{1}{10}\right)}{\sum_{k=5}^{8} f(x|k)\pi(k)} = 0.13.$$

A continuación se resumen los valores de estas funciones para los distintos valores de k y para ambos escenarios, es decir, cuando se tiene rojo, verde y rojo (cuadro 1.2) y para cuando

k	$\pi(k)$	$f(x\|k)$	$\pi(k)f(x\|k)$	$\pi(k\|x)$
5	.1	.160	.016	.13
6	.3	.139	.042	.33
7	.3	.122	.037	.29
8	.3	.109	.033	.26

Cuadro 1.2: Cálculo de la distribución a posteriori cuando los dulces examinados son rojo, verde y rojo.

el cuarto dulce es naranja (cuadro 1.3).

Observe que la distribución a posteriori para k es una pequeña modificación de la a priori.

k	$\pi(k)$	$f(x\|k)$	$\pi(k)f(x\|k)$	$\pi(k\|x)$
5	.1	.096	.010	.11
6	.3	.093	.028	.31
7	.3	.087	.026	.30
8	.3	.082	.025	.28

Cuadro 1.3: Cálculo de la distribución a posteriori cuando los dulces examinados son rojo, verde, rojo y naranja.

La estadística Bayesiana se basa en el cálculo de distribuciones condicionales. Los siguientes ejemplos ilustran el uso de la definición de densidades condicionales en términos de las densidades conjuntas y marginales.

Ejemplo 1.18 *Una moneda cargada se lanza n veces. Suponga que x_i vale 1 si se obtiene sol y 0 si no, en el i-ésimo lanzamiento. No se tiene idea de qué tan cargada está la moneda, entonces se considera una distribución uniforme a priori para θ, de tal manera que la densidad a priori está dada por:*

$$\pi(\theta) = 1, \quad 0 \leq \theta \leq 1.$$

Sea t el número de soles. Entonces la distribución a posteriori de θ es:

$$\pi(\theta|x_1, \ldots, x_n) = \frac{\theta^t(1-\theta)^{n-t} \times 1}{\int_0^1 \phi^t(1-\phi)^{n-t} \times 1 d\phi}$$

$$\pi(\theta|\underline{\mathbf{x}}) \propto \theta^t(1-\theta)^{n-t},$$

de donde se puede ver que si se inserta una constante de proporcionalidad apropiada, entonces se tiene una densidad Beta$(t+1, n-t+1)$, que sería la distribución a posteriori de θ dada x.

En el ejemplo anterior se utiliza $\propto$ para denotar que $\pi(\theta|\underline{\mathbf{x}})$ es proporcional a $\theta^t(1-\theta)^{n-t}$. En general, $\propto$ se lee como "es proporcional a".

Ejemplo 1.19 *Suponga que $X_1, \ldots, X_n$ es una muestra aleatoria de una población con distribución $N(\mu, 1)$ y que $\pi(\mu) \sim N(0, \tau^{-2})$ para τ^{-2} conocida. Entonces*

$$\pi(\mu|x_1, \ldots, x_n) \propto \exp\left\{-\frac{1}{2}\left(\sum_{i=1}^n (x_i-\mu)^2 + \mu^2\tau^2\right)\right\}$$

$$\propto \exp\left\{-\frac{1}{2}(n+\tau^2)\left(\mu - \frac{\sum_{i=1}^n x_i}{n+\tau^2}\right)^2\right\}.$$

Así,

$$\mu|x_1, \ldots, x_n \sim N\left(\frac{\sum_{i=1}^n x_i}{n+\tau^2}, \frac{1}{n+\tau^2}\right).$$

Ejemplo 1.20 *Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con distribución exponencial(λ) y la distribución a priori para el parámetro λ es una exponencial(μ), donde μ es fija y conocida. Entonces:*

$$\pi(\lambda|x_1, \ldots, x_n) \propto \mu e^{-\lambda\mu} \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda(\mu + \sum_{i=1}^{n} x_i)},$$

es decir, $\lambda \sim gama(n+1, \mu + \sum_{i=1}^{n} x_i)$.

Ejemplo 1.21 *Suponga que se examina una máquina que hace partes de automóviles y se denota a θ como la proporción de marcas defectuosas. Un día se examinan 10 piezas denotadas por $X_1, \ldots, X_{10}$, donde $X_i = 1$ indica que la pieza i es defectuosa y $X_i = 0$ que no tiene defecto. Esto puede verse como una muestra aleatoria con distribución Bernoulli de parámetro θ, cuya función de densidad es $f_X(x;\theta) = \theta^x(1-\theta)^{1-x} I_{\{0,1\}}(x)$ para $0 \leq \theta \leq 1$, que indica que la probabilidad de que una parte sea defectuosa es θ. Así que la densidad conjunta de las 10 observaciones es*

$$\begin{aligned} f_{\underline{X}}(\underline{x};\theta) &= \theta^{\sum_{i=1}^{10} x_i}(1-\theta)^{10-\sum_{i=1}^{10} x_i} \prod_{i=1}^{10} I_{\{0,1\}}(x_i) \\ &= L\left(\theta \mid \underline{\mathbf{x}}\right). \end{aligned}$$

Con el método de máxima verosimilitud el estimador de θ es $\hat{\theta} = \bar{X}$. Suponga que el experto tiene información adicional acerca de θ y que él ha observado que a lo largo de los días la proporción de partes defectuosas cambia, es decir, el valor de θ cambia y que este cambio puede representarse como una variable aleatoria con función de densidad $\pi(\theta) = 6\theta(1-\theta)I_{[0,1]}(\theta)$, esto es, θ tiene una distribución Beta con parámetros 2 y 2, denotada como $Beta(2,2)$. ¿Cómo se puede usar esta información adicional para estimar θ?.

*Como ya se ha señalado, en el método Bayesiano se considera que θ es una cantidad cuya variación puede describirse por medio de una distribución de probabilidad (llamada **distribución a priori**). La distribución a priori es una distribución subjetiva, basada en las creencias del experto y se formula antes de obtener los datos. Se selecciona una muestra a partir de una población sujeta al parámetro θ, entonces la distribución a priori se actualiza utilizando la información de la muestra y se obtiene la **disribución a posteriori**. Esta actualización se hace usando la regla de Bayes. La distribución a posteriori es una distribución condicional, y es condicional dada la muestra. La distribución a posteriori se usa para hacer inferencia acerca de θ (obtener el estimador puntual, intervalos de credibilidad y pruebas de hipótesis).*

La distribución conjunta de $X_1, \ldots, X_{10}$ y θ es

$$\begin{aligned}
g(\underline{x}, \theta) &= \underbrace{f_{\underline{X}}(\underline{x}; \theta)}_{\textit{distribución conjunta}} \times \underbrace{\pi(\theta)}_{\textit{distribución a priori}} \\
&= \theta^{\sum_{i=1}^{10} x_i}(1-\theta)^{10-\sum_{i=1}^{10} x_i} \times 6\theta(1-\theta) \\
&= \theta^{y}(1-\theta)^{10-y} \times 6\theta(1-\theta) \\
&= 6\theta^{y+1}(1-\theta)^{10-y+1},
\end{aligned}$$

donde $y = \sum_{i=1}^{10} x_i$. Calculando la distribución marginal de la muestra, $m(\underline{x})$,

$$\begin{aligned}
m(\underline{x}) &= \int f_{\underline{X}}(\underline{x}; \theta)\pi(\theta) d\theta = \int g(\underline{x}, \theta) d\theta \\
&= \int 6\theta^{y+1}(1-\theta)^{10-y+1} d\theta \\
&= 6\frac{\Gamma(y+2)\Gamma(10-y+2)}{\Gamma(10+2+2)} \\
&= 6\frac{\Gamma(y+2)\Gamma(12-y)}{\Gamma(14)}.
\end{aligned}$$

Así, la distribución a posteriori de θ dada la muestra $\underline{x}$ es

$$\begin{aligned}
\pi(\theta|\underline{x}) &= \frac{g(\underline{x}, \theta)}{m(\underline{x})} \\
&= \frac{6\theta^{y+1}(1-\theta)^{10-y+1}}{6\frac{\Gamma(y+2)\Gamma(12-y)}{\Gamma(14)}} \\
&= \frac{\Gamma(14)}{\Gamma(y+2)\Gamma(12-y)}\theta^{y+1}(1-\theta)^{11-y},
\end{aligned}$$

que es una distribución $Beta(y+2, 12-y)$.

Un estimador para θ es la media de la distribución a posteriori (ver Sección 1.2.3), la cual daría el estimador de Bayes de θ,

$$\hat{\theta} = \frac{y+2}{14}.$$

En el cuadro 1.4 se resumen los valores de los estimadores máximo verosímil y de Bayes para distintos valores de la muestra.

Las gráficas de la figura 1.5 muestran el comportamiento de la distribución a posteriori ante la evidencia de los datos y el conocimiento previo del parámetro.

Muestra y	0	1	2	3	4	5	6	7	8	9	10
$\hat{\theta}$ EMV	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$\hat{\theta}$ Bayes	0.14	0.21	0.29	0.36	0.43	0.5	0.57	0.64	0.71	0.79	0.86

Cuadro 1.4: Valores de los estimadores máximo verosímil y de Bayes para distintos valores de la muestra.

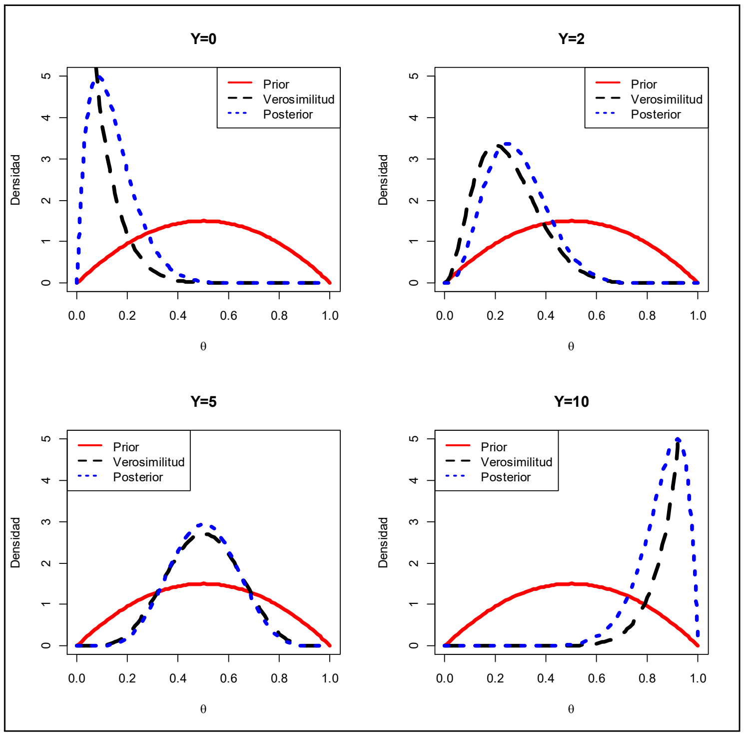

Figura 1.5: Comportamiento de las funciones a priori, a posteriori y de verosimilitud para la proporción de marcas defectuosas y distintos valores de la muestra en el ejemplo 1.21.

En el caso de no utilizar familias conjugadas, la obtención de la constante de normalización para las distribuciones posteriores no se puede hacer de manera analítica y es necesario recurrir a métodos de simulación.

1.1.4. Otros métodos de estimación

Además de los métodos de momentos y los basados en verosimilitud, existen alternativas para encontrar el estimador de un parámetro, por ejemplo, el **método de medianas**, su extensión, el **método de percentiles,** y el **método de mínimos cuadrados**. Los primeros dos se revisan en esta sección, mientras que el último se abordará en la Sección 1.2.1.

El método de medianas únicamente se puede aplicar a modelos dependientes de un sólo parámetro y consiste en lo siguiente: suponga que $\widetilde{x}_{0.5}$ representa a la mediana de la muestra mientras que $x_{0.5}$ denota a la mediana de la distribución. Recuerde que $x_{0.5}$ es el valor de x para el cual $F_X(x_{0.5}) = \frac{1}{2}$, donde $F_X(x)$ es la función de distribución de la variable aleatoria continua X, o alternativamente,

$$\int_{-\infty}^{x_{0.5}} f(x;\theta)\,dx = \int_{x_{0.5}}^{-\infty} f(x;\theta)\,dx = \frac{1}{2}.$$

El método consiste en igualar las medianas y resolver para el parámetro desconocido, el cual está involucrado en la expresión resultante para $x_{0.5}$. La extensión de este método para el caso de dos o más parámetros puede hacerse a través del método de percentiles que se describe a continuación.

Método de percentiles

Si x_p es el valor de x tal que $F_X(x_p) = p$, entonces x_p es el p-ésimo($\times 100$) percentil de la distribución. Para usar este método se calculan los correspondientes percentiles de la muestra y se igualan con los de la distribución (los cuales se encuentran en términos de los parámetros desconocidos) y se resuelve para cada parámetro. Observe que si $p = \frac{1}{2}$, entonces x_p es la mediana, así que el método de las medianas se puede ver como un caso particular.

Ejemplo 1.22 *Suponiendo que se tiene una muestra aleatoria de tamaño n de una población con distribución exponencial, se desea estimar el parámetro θ en $f(x;\theta) = \theta\exp(-\theta x)$. Primero se resuelve $F_X(x_{0.5}) = \frac{1}{2}$ o*

$$\int_0^{x_{0.5}} \theta e^{-\theta x} dx = \frac{1}{2},$$

de donde

$$1 - e^{-\theta x_{0.5}} = \frac{1}{2}$$

o

$$e^{-\theta x_{0.5}} = \frac{1}{2},$$

resultando:

$$x_{0.5} = -\frac{\ln\frac{1}{2}}{\theta}. \tag{1.11}$$

Igualando (1.11) con $\widetilde{x}_{0.5}$, es decir con la mediana muestral, se obtiene que

$$\widehat{\theta} = -\frac{\ln\frac{1}{2}}{\widetilde{x}_{0.5}}.$$

Ejemplo 1.23 *Usando el método de percentiles, estimar los parámetros en una distribución Weibull con función de densidad*

$$f(x;\theta) = \gamma\lambda x^{\gamma-1} \exp\{-\lambda x^{\gamma}\}.$$

Dados los percentiles muestrales $\widetilde{x}_{0.5} = 10000$ *y* $\widetilde{x}_{0.9} = 100000$, *¿cuál es el estimador para el parámetro* γ?

La función de distribución correspondiente es

$$\begin{aligned} F_X(x) &= \int_0^x \gamma\lambda u^{\gamma-1} \exp\{-\lambda u^{\gamma}\}\, du \\ &= 1 - \int_x^{\infty} \gamma\lambda u^{\gamma-1} \exp\{-\lambda u^{\gamma}\}\, du \end{aligned}$$

y

$$\begin{aligned} \int_x^{\infty} \gamma\lambda u^{\gamma-1} \exp\{-\lambda u^{\gamma}\}\, du &= \left.\frac{-\gamma\lambda u^{\gamma-1}}{-\gamma\lambda u^{\gamma-1}} e^{-\lambda u^{\gamma}}\right|_x^{\infty} \\ &= e^{-\lambda x^{\gamma}}. \end{aligned}$$

Por lo tanto:

$$F_X(x) = 1 - \exp\{-\lambda x^{\gamma}\},$$

la cual tiene dos parámetros. Si x_p *es el valor de* x *tal que* $F_X(x_p) = p$, *entonces* x_p *es el* $100p$*-ésimo percentil de la distribución, que para este caso daría las ecuaciones:*

$$F_X(x_{0.5}) = 1 - \exp\{-\lambda x_{0.5}^{\gamma}\} = 0.50$$

y

$$F_X(x_{0.9}) = 1 - \exp\{-\lambda x_{0.9}^{\gamma}\} = 0.90,$$

de donde:

$$\exp\{-\lambda x_{0.5}^{\gamma}\} = 0.50$$

y

$$\exp\{-\lambda x_{0.9}^{\gamma}\} = 0.10,$$

que es equivalente a

$$\lambda x_{0.5}^{\gamma} = -\ln(0.50) = 0.69315$$

y

$$\lambda x_{0.9}^{\gamma} = -\ln(0.10) = 2.30259.$$

Es decir:

$$x_{0.5} = \left(\frac{0.69315}{\lambda}\right)^{\frac{1}{\gamma}}$$

y

$$x_{0.9} = \left(\frac{2.30259}{\lambda}\right)^{\frac{1}{\gamma}}.$$

Igualando con los respectivos percentiles muestrales, se obtiene:

$$\left(\frac{0.69315}{\lambda}\right)^{\frac{1}{\gamma}} = 10000 \tag{1.12}$$

y

$$\left(\frac{2.30259}{\lambda}\right)^{\frac{1}{\gamma}} = 100000. \tag{1.13}$$

Dividiendo (1.13) entre (1.12):

$$\left(\frac{2.30259}{0.69315}\right)^{\frac{1}{\gamma}} = 10$$

o sea,

$$10^{\gamma} = \frac{2.30259}{0.69315} = 3.32192.$$

De donde se obtiene que

$$\widehat{\gamma} = \frac{\ln 3.32192}{\ln 10} = 0.521.$$

Usando (1.12), se puede obtener el estimador de λ :

$$\frac{0.69315}{\lambda} = (10000)^{0.521} = 121.34.$$

Así,

$$\widehat{\lambda} = \frac{0.69315}{121.34} = \frac{457}{80\,000}.$$

Si los percentiles muestrales no están dados explícitamente, se puede usar el siguiente método para calcularlos: para una muestra de tamaño n, sean $x_1, x_2, ..., x_n$ los valores muestrales en orden ascendente. Sea $k = (n+1)\,p$, donde p es el orden del percentil que se busca. Sea l la parte entera de k ($l = 1, 2, ..., n-1$) y sea m la parte fraccional de k, $0 \leq m < 1$. Se define

$$\widetilde{x}_p = (1-m)\,x_l + m x_{l+1} \tag{1.14}$$

como el p-ésimo percentil ($\times 100$) de la muestra. Observe que x_l y x_{l+1} representan los elementos l-ésimo y $(l+1)$-ésimo de la muestra, respectivamente.

Ejemplo 1.24 *En una muestra de ratas de laboratorio se obtienen los tiempos de muerte dados a continuación:* $x = 3, 4, 5, 7, 7, 8, 10, 10$ *y* 12, *donde el tiempo se mide en días. Usando el método de percentiles, estimar los parámetros* B *y* c *del modelo de supervivencia Gompertz, cuya función de distribución está dada por:*

$$F_X(x) = 1 - \exp\left[\frac{B}{\ln c}(1 - c^x)\right],$$

con los percentiles 0.25 y 0.65.

Los percentiles 0.25 y 0.65 son tales que

$$1 - \exp\left[\frac{B}{\ln c}(1 - c^{x_{0.25}})\right] = 0.25$$

y

$$1 - \exp\left[\frac{B}{\ln c}(1 - c^{x_{0.65}})\right] = 0.65$$

$\Leftrightarrow$

$$\frac{B}{\ln c}(1 - c^{x_{0.25}}) = \ln 0.75$$
$$\frac{B}{\ln c}(1 - c^{x_{0.65}}) = \ln 0.35$$

$\Leftrightarrow$

$$c^{x_{0.25}} = 1 - \ln 0.75\frac{\ln c}{B}$$
$$c^{x_{0.65}} = 1 - \ln 0.35\frac{\ln c}{B}.$$

Así,

$$x_{0.25} = \frac{\ln\left[1 - \ln 0.75\frac{\ln c}{B}\right]}{\ln c} \tag{1.15}$$

y

$$x_{0.65} = \frac{\ln\left[1 - \ln 0.35\frac{\ln c}{B}\right]}{\ln c}. \tag{1.16}$$

Usando (1.14), se tiene que para el percentil 0.25, $k = (9 + 1)(0.25) = 2.5$, *de donde se obtiene* $\widetilde{x}_{0.25} = (0.5)(x_2) + (0.5)(x_3) = (0.5)(4) + (0.5)(5) = 4.5$. *Para el cuantil* 0.65, $k = (9 + 1)(0.65) = 6.5$, *por lo tanto* $\widetilde{x}_{0.65} = (0.5)x_6 + (0.5)x_7 = (0.5)8 + (0.5)10 = 9$.

Igualando los percentiles obtenidos en (1.15) y (1.16) con los percentiles muestrales, resultan las siguientes ecuaciones:

$$\frac{\ln\left[1 - \ln 0.75\frac{\ln c}{B}\right]}{\ln c} = 4.5 \tag{1.17}$$

y

$$\frac{\ln\left[1-\ln 0.35\frac{\ln c}{B}\right]}{\ln c}=9.$$

Dividiendo la segunda entre la primera, resulta

$$\frac{\ln\left[1-\ln 0.35\frac{\ln c}{B}\right]}{\ln\left[1-\ln 0.75\frac{\ln c}{B}\right]}=2$$

$\Rightarrow$

$$1-\ln 0.35\frac{\ln c}{B}=\left[1-\ln 0.75\frac{\ln c}{B}\right]^2$$

$\Rightarrow$

$$\ln 0.35\frac{\ln c}{B}=1-\left[1-\ln 0.75\frac{\ln c}{B}\right]^2$$

$\Rightarrow$

$$\ln 0.35z=1-\left(1-2z\ln 0.75+(\ln 0.75)^2 z^2\right),$$

donde $z=\frac{\ln c}{B}$. *Simplificando la última expresión se obtiene:*

$$(\ln 0.75)^2 z^2 - z\left[2\ln 0.75-\ln 0.35\right]=0,$$

o

$$0.83z^2-(0.4744)\,z=0,$$

de donde

$$z=\frac{0.4744}{0.083}=5.7163.$$

Es decir, $\frac{\ln c}{B}=5.7163$, *lo que a su vez implica que* $\ln c=5.7163B$. *Sustituyendo este último valor en (1.17), se tiene:*

$$\frac{\ln\left[1-\ln 0.75\frac{5.7163B}{B}\right]}{5.7163B}=4.5$$

y despejando B, *se llega al resultado* $\widehat{B}=0.03780$. *Finalmente,* $\widehat{c}=1.2412$.

Existen diferentes propuestas para obtener estimadores, entonces es necesario establecer criterios para evaluarlos y compararlos. En las siguiente secciones se abordará este tema.

1.2. Evaluación de estimadores

Dado que hay varios métodos para encontrar estimadores, una pregunta natural es, si se pueden tener estimadores distintos para una parámetro, ¿cuál es mejor o cuál se debe elegir?. Es necesario contar con criterios para responder a esta pregunta y poder decidir cuál estimador es mejor en algún sentido.

1.2.1. Error cuadrático medio y estimadores insesgados

El primer criterio que se analizará es el del error cuadrático medio, concepto que se introduce a continuación.

Definición 1.6 *Sea $T(X_1, \ldots, X_n)$ un estimador de $\tau(\theta)$. Se define el error cuadrático medio (ECM) de T como*

$$ECM_T(\theta) = \mathbb{E}[(T(\underline{X}) - \tau(\theta))^2].$$

Es decir, el error cuadrático medio mide el error cometido al estimar $\tau(\theta)$ con $T(\underline{X})$. Esta medida es un error promedio al considerar los valores que puede tomar la variable aleatoria $T(\underline{X})$ y se calcula como la esperanza de los errores al cuadrado, tomando los errores como la diferencia entre los valores de la variable aleatoria y el valor del parámetro.

La idea es que si se tienen dos estimadores $T_1(\underline{X})$ y $T_2(\underline{X})$ para $\tau(\theta)$ y $ECM_{T_1}(\theta) < ECM_{T_2}(\theta)$, entonces se elige a T_1 como estimador para $\tau(\theta)$.

Si se desarrolla la expresión $\mathbb{E}[(T(\underline{X}) - \tau(\theta))^2]$, se obtiene que

$$\begin{aligned}
ECM_T(\theta) &= \mathbb{E}[T^2 - 2\tau(\theta)T + (\tau(\theta))^2] \\
&= \mathbb{E}(T^2) - 2\tau(\theta)\mathbb{E}(T) + \tau^2(\theta) \\
&= \mathbb{E}(T^2) - \mathbb{E}^2(T) + \mathbb{E}^2(T) - 2\tau(\theta)\mathbb{E}(T) + \tau^2(\theta) \\
&= Var(T) + [\underbrace{\mathbb{E}(T) - \tau(\theta)}_{\text{sesgo de } T}]^2
\end{aligned}$$

A $\mathbb{E}(T) - \tau(\theta)$ se le conoce como *sesgo de T*. Es importante hacer notar que si el sesgo de T es cero, entonces $ECM_T(\theta) = Var(T)$.

Definición 1.7 *Un estimador $T(\underline{X})$ de $\tau(\theta)$ es insesgado si $\mathbb{E}[T(\underline{X})] = \tau(\theta)$ (es decir, en promedio, el estimador es igual al parámetro).*

Nota: Si T es insesgado, entonces $ECM_T(\theta) = Var(T)$.

Observaciones 1.1

1. *En el caso continuo, el error cuadrático medio $\mathbb{E}[(T-\tau(\theta))^2]$ puede calcularse como*

$$\int \ldots \int (t(x_1, x_2, \ldots, x_n) - \tau(\theta))^2 f_{X_1}(x_1;\theta) \ldots f_{X_n}(x_n;\theta) dx_1 \ldots dx_n.$$

2. *El ECM puede pensarse también como una medida de la dispersión de T alrededor de $\tau(\theta)$.*

3. *$ECM_{T_1}(\theta)$ y $ECM_{T_2}(\theta)$ pueden cruzarse, es decir, en general no se cumple que para todo $\theta \in \Theta$, $ECM_{T_1}(\theta) < ECM_{T_2}(\theta)$ (ó $ECM_{T_1}(\theta) > ECM_{T_2}(\theta)$).*

Ejemplo 1.25 *Sea $X_1, \ldots, X_n$ una muestra aleatoria de la distribución $N(\mu, \sigma^2)$. Considere*

$$T_1(\underline{X}) = \bar{X}$$

un estimador para μ. Sean

$$T_2(\underline{X}) = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

y

$$T_3(\underline{X}) = \hat{\sigma}^2_{MV} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{n-1}{n}S^2,$$

estimadores para σ^2.
Para revisar si son insesgados:

- *Note que*
$$\mathbb{E}(T_1(\underline{X})) = \mathbb{E}(\bar{X}) = \mu.$$
Por lo tanto, T_1 sí es insesgado.

- *También note que*
$$\mathbb{E}\left[T_2(\underline{X})\right] = \mathbb{E}(S^2) = \sigma^2.$$
Por lo tanto, T_2 sí es insesgado.

- *Sin embargo,*
$$\mathbb{E}(T_3(\underline{x})) = \mathbb{E}\left[\frac{n-1}{n}S^2\right] = \frac{n-1}{n}\mathbb{E}\left[S^2\right] = \frac{n-1}{n}\sigma^2.$$
Por lo tanto, T_3 no es insesgado.

Para encontrar el error cuadrático medio de T_1, T_2 y T_3:

- *Como T_1 es insesgado, entonces*
$$ECM_{T_1}(\mu, \sigma^2) = Var(\bar{X}) = \frac{\sigma^2}{n}.$$

- *Como T_2 es insesgado, entonces*
$$ECM_{T_2}(\mu, \sigma^2) = Var(S^2) = \frac{2\sigma^4}{n-1}.$$

- *Sin embargo, como T_3 no es insesgado, entonces*

$$ECM_{T_3}(\mu, \sigma^2) = Var(T_3) + (sesgo^2).$$

Pero

$$Var(T_3) = Var\left(\frac{n-1}{n}S^2\right) = \frac{(n-1)^2}{n^2}\frac{2\sigma^4}{n-1} = \frac{2(n-1)}{n^2}\sigma^4.$$

Y

$$\begin{aligned}(sesgo)^2 &= (\mathbb{E}(T_3) - \sigma^2)^2 = \left(\frac{n-1}{n}\sigma^2 - \sigma^2\right)^2 \\ &= \left(\frac{n-1-n}{n}\right)^2 \sigma^4 = \frac{\sigma^4}{n^2}.\end{aligned}$$

Por lo tanto,

$$ECM_{T_3}(\mu, \sigma^2) = \frac{2(n-1)}{n^2}\sigma^4 + \frac{\sigma^4}{n^2} = \frac{2n-1}{n^2}\sigma^4.$$

Pero note que

$$\begin{aligned}\frac{2}{n} &< \frac{2}{n-1} \Rightarrow \frac{2}{n} - \frac{1}{n^2} < \frac{2}{n-1} \Rightarrow \frac{2n-1}{n^2} < \frac{2}{n-1} \\ &\Rightarrow \frac{(2n-1)\sigma^4}{n^2} < \frac{2\sigma^4}{n-1} \Rightarrow ECM_{T_3} < ECM_{T_2}.\end{aligned}$$

Con esto se puede observar que aunque T_2 es insesgado, T_3 tiene un menor ECM, lo cuál exhibe que no siempre un estimador insesgado tiene el menor ECM.

Ilustración mediante simulación

Se simula un conjunto de $M = 1000$ muestras de tamaño $n = 3$ cada una. Los estimadores $T_1(\underline{X})$, $T_2(\underline{X})$ y $T_3(\underline{X})$ del ejemplo 1.25, se grafican en las figuras 1.6 y 1.7.

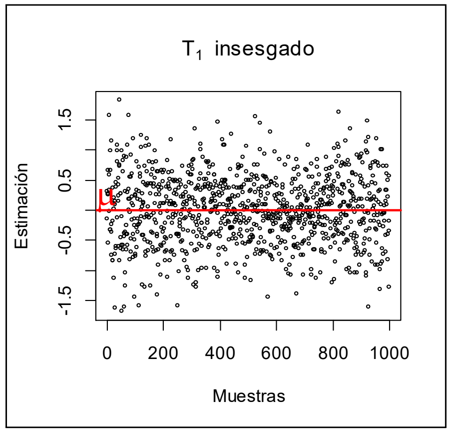

Figura 1.6: El estimador $T_1 = \overline{X}$ es insesgado para μ en el ejemplo 1.25.

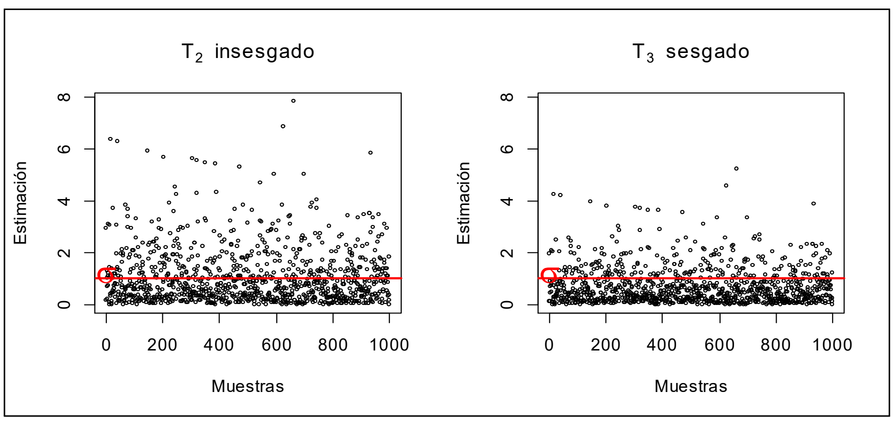

Figura 1.7: El estimador $T_2 = S^2$ es insesgado para σ^2. T_3 no es insesgado para σ^2, pero tiene un error cuadrático medio menor que T_2 (ejemplo 1.25).

Método de mínimos cuadrados para estimación de parámetros

Existe otro procedimiento de estimación conocido como el *método de mínimos cuadrados*, el cual se usa en distintas aplicaciones para encontrar los estimadores de los parámetros relacionados con modelos de diversa índole. Se ilustrará con un ejemplo en el marco del criterio del error cuadrático medio.

Ejemplo 1.26 *Considere un conjunto de n puntos en el plano*

$$(x_1, y_1), \ldots, (x_n, y_n)$$

y el siguiente experimento: se escoge X con $P\{X = x_i\} = 1/n$, para $i = 1, \dots, n$; si $X = x_i$ se asigna $Y = y_i$. Suponga que Y tiene la forma $aX + b$ y se desea encontrar un estimador para Y, de tal manera que se minimice el error cuadrático medio, el cual es:

$$E[(Y - (aX + b))^2] = \frac{1}{n} \sum_{i=1}^{n} [y_i - (ax_i + b)]^2.$$

Para la función $S(a, b) = \sum_{i=1}^{n} [y_i - (ax_i + b)]^2$, los valores que minimizan la expresión satisfacen:

$$\frac{\partial S(a,b)}{\partial a} = -2 \sum_{i=1}^{n} y_i x_i + 2 \sum_{i=1}^{n} a x_i^2 + 2b \sum_{i=1}^{n} x_i = 0,$$

$$\frac{\partial S(a,b)}{\partial b} = -2 \sum_{i=1}^{n} y_i + 2 \sum_{i=1}^{n} a x_i + 2nb = 0,$$

de donde, las soluciones que minimizan el ECM están dadas por

$$\widehat{b} = \frac{1}{n} \sum_{i=1}^{n} y_i - \widehat{a} \frac{1}{n} \sum_{i=1}^{n} x_i$$

y

$$\widehat{a} \left(\sum_{i=1}^{n} x_i^2 - \frac{1}{n} \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i \right) = \sum_{i=1}^{n} y_i x_i - \frac{1}{n} \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i.$$

A estos estimadores se les conoce como el estimador de mínimos cuadrados para a y b. A $\widehat{Y} = \widehat{a}X + \widehat{b}$ se le llama el estimador de mínimos cuadrados de Y.

1.2.2. Consistencia

La consistencia es otra propiedad deseable en un estimador y tiene que ver con tamaños de muestra grandes, es decir, es una propiedad asintótica. Esencialmente, un estimador es consistente, si para n (el tamaño de muestra) grande, el error cometido al estimar $\tau(\theta)$ con $T_n(\underline{X})$, es pequeño (tiende a cero).

Definición 1.8 *(Consistencia en ECM). Sea $T_1, T_2, ..., T_n$ una sucesión de estimadores de $\tau(\theta)$, donde T_n está basado en una muestra de tamaño n. Esta sucesión de estimadores de $\tau(\theta)$ es consistente en error cuadrático medio (ECM) si:*

$$\lim_{n \to \infty} \mathbb{E}[(T_n(\underline{X}) - \tau(\theta))^2] = 0. \tag{1.18}$$

Note que (1.18) es una convergencia en media cuadrática, de la sucesión $\{T_n\}$ a $\tau(\theta)$.

Ejemplo 1.27 *Sea $X_1, \ldots, X_n$ una muestra aleatoria de la distribución $N(\mu, \sigma^2)$. Considere los estimadores $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ para μ y $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ para σ^2. Note que*

$$\mathbb{E}[(\bar{X}_n - \mu)^2] = Var(\bar{X}_n) = \frac{\sigma^2}{n} \stackrel{n\to\infty}{\longrightarrow} 0.$$

Por lo tanto $\bar{X}_n$ es consistente para μ. También note que

$$\mathbb{E}[(S_n^2 - \sigma^2)^2] = Var(S_n^2) = \frac{2\sigma^4}{n-1} \stackrel{n\to\infty}{\longrightarrow} 0.$$

Por lo tanto S_n^2 es consistente para σ^2.

El error cuadrático medio, ECM, es el criterio para medir la bondad de un estimador. Una propiedad desable de un estimador es que proporcione, para muestras grandes, un error (ECM) pequeño en la estimación, es decir, que sea consistente.

Ilustración del concepto de consistencia mediante simulación

Se simula un conjunto de $n = 1000$ muestras de tamaño i, para $i = 2, \ldots, n$. Los estimadores $\overline{X}_n$ y S_n^2 son consistentes, y se pueden observar las gráficas correspondientes en las figuras 1.8 y 1.9.

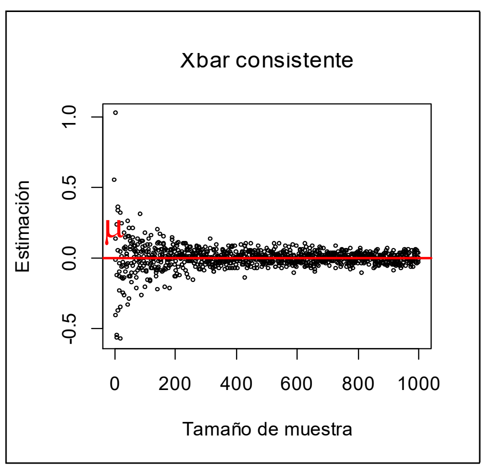

Figura 1.8: Ilustración de la consistencia de $\overline{X}$ en el contexto del ejemplo 1.27.

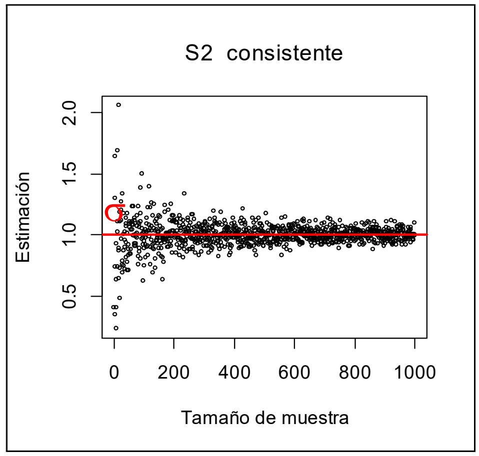

Figura 1.9: Ilustración de la consistencia de S^2 en el contexto del ejemplo 1.27.

Definición 1.9 *Se dice que una sucesión de estimadores* $\{T_n\}_{n\in\mathbb{N}}$ *es consistente simple si y sólo si*

$$\forall \epsilon > 0 \quad \lim_{n\to\infty} \mathbb{P}(|T_n - \tau(\theta)| < \epsilon) = 1. \tag{1.19}$$

La consistencia en ECM implica la consistencia simple. Esto se puede analizar desde dos perspectivas: la primera, notando que (1.19) es una convergencia en probabilidad y usando el hecho de que la convergencia en r-ésima media implica la convergencia en probabilidad; la segunda, utilizando la desigualdad de Chebyshev:

$$\mathbb{P}(|T_n - \tau(\theta)| \geq \epsilon) = \mathbb{P}([T_n - \tau(\theta)]^2 \geq \epsilon^2) \leq \frac{\mathbb{E}[(T_n - \tau(\theta))^2]}{\epsilon^2}$$

y la definición de consistencia en ECM.

1.2.3. Funciones de pérdida y estimación

El enfoque Bayesiano al problema de estimación de parámetros es a través de una función de pérdida $\mathbb{L}(\theta, a)$, la cual mide la pérdida en que se incurre cuando se estima el valor de un parámetro mediante a, siendo que el verdadero valor es θ. Entonces $\hat{\theta}$ se selecciona de tal manera que minimice $\mathbb{E}[\mathbb{L}(\theta, \hat{\theta})]$, donde esta esperanza se toma con respecto a θ usando la distribución a posteriori $\pi(\theta|x)$.

Definición 1.10 *A* $\mathbb{L}(\theta, a) = (a - \theta)^2$ *se le llama la* ***función de pérdida del error cuadrático****.*

Observe que:

$$\mathbb{E}[\mathbb{L}(\theta,a)] = \int \mathbb{L}(\theta,a)\pi(\theta|x_1,\ldots,x_n)d\theta = \int (a-\theta)^2\pi(\theta|x_1,\ldots,x_n)d\theta.$$

Diferenciando esta expresión con respecto a a, se obtiene:

$$2\int (a-\theta)\pi(\theta|x_1,\ldots,x_n)d\theta = 0 \Longrightarrow a = \int \theta\pi(\theta|x_1,\ldots,x_n)d\theta$$

Por lo tanto, la pérdida del error cuadrático se minimiza en $\hat{\theta}$, la **media** o esperanza **a posteriori** de θ.

Definición 1.11 *A $\mathbb{L}(\theta,a) = |a-\theta|$ se le llama la* ***función de pérdida del error absoluto.***

En este caso,

$$\mathbb{E}[\mathbb{L}(\theta,a)] = \int \mathbb{L}(\theta,a)\pi(\theta|x_1,\ldots,x_n)d\theta$$

$$= \int_{-\infty}^{a} (a-\theta)\pi(\theta|x_1,\ldots,x_n)d\theta + \int_{a}^{\infty} (\theta-a)\pi(\theta|x_1,\ldots,x_n)d\theta.$$

Diferenciando con respecto a a, se llega a que el mínimo debe cumplir que:

$$\int_{\theta=-\infty}^{a} \pi(\theta|x_1,\ldots,x_n)d\theta - \int_{a}^{\infty} \pi(\theta|x_1,\ldots,x_n)d\theta = 0$$

Así, ambas integrales deberían ser iguales a $\frac{1}{2}$ y $\hat{\theta}$ es la **mediana a posteriori**.

Ejemplo 1.28 *Sea $X_1,\ldots,X_n$ una muestra aleatoria de una población con distribución $Poisson(\lambda)$, Suponga que $\lambda \sim$exponencial(1), de modo que*

$$\pi(\lambda) = e^{-\lambda}, \lambda > 0.$$

La distribución a posteriori es

$$\pi(\lambda|x_1,\ldots,x_n) = e^{-\lambda}\prod_{i=1}^{n}\frac{e^{-\lambda}\lambda^{x_i}}{x_i!} \propto e^{-\lambda(n+1)}\lambda^{\sum_{i=1}^{n}x_i},$$

es decir, gamma$(\sum_{i=1}^{n}x_i+1, n+1)$. Entonces, usando la función de pérdida del error cuadrático medio:

$$\hat{\theta} = \textit{media a posteriori} = \frac{\sum_{i=1}^{n}x_i+1}{n+1}.$$

Y bajo la función de pérdida del error absoluto, $\hat{\theta}$ es la solución a:

$$\int_{0}^{\hat{\theta}} \frac{e^{-\lambda(n+1)}\lambda^{\sum_{i=1}^{n}x_i}(n+1)^{\sum_{i=1}^{n}x_i+1}}{(\sum_{i=1}^{n}x_i)!}d\lambda = \frac{1}{2}.$$

1.3. Estimación insesgada

En esta sección se hará una restricción considerando únicamente a los estimadores insesgados, es decir a los estimadores $T(\underline{X})$ que pertenecen a la clase:

$$\mathcal{C}_{\tau(\theta)} = \{T(\underline{X}) \mid \mathbb{E}[T(\underline{X})] = \tau(\theta)\},$$

la clase de estimadores insesgados para $\tau(\theta)$.

El siguiente ejemplo muestra la idea subyacente en esta sección en cuanto a la estimación basada en minimizar la varianza de estimadores insesgados.

Ejemplo 1.29 *Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con distribución $Poisson(\lambda)$. Primero note que*

$$\mathbb{E}(X_i) = \lambda, Var(X_i) = \lambda \text{ y } \mathbb{E}(X_i^2) = Var(X_i) + \mathbb{E}^2(X_i) = \lambda + \lambda^2.$$

Considerando ahora la estadística $G(X_1, \ldots, X_n) = \sum_{i=1}^n X_i$, note que $G(\underline{X})$ tiene distribución $Poisson(n\lambda)$; por lo que $\mathbb{E}(G) = n\lambda$ y $Var(G) = n\lambda$. Sean $a \in (0,1)$ una constante y

$$T_a(X_1, \ldots, X_n) = a\bar{X} + (1-a)S^2.$$

Entonces,

$$\begin{aligned}
\mathbb{E}(\bar{X}) &= \frac{1}{n}\mathbb{E}(G) = \lambda; \\
Var(\bar{X}) &= \frac{1}{n^2}Var(G) = \frac{\lambda}{n}; \\
\mathbb{E}(\bar{X}^2) &= Var(\bar{X}) + \mathbb{E}^2(\bar{X}) = \frac{\lambda}{n} + \lambda^2; \\
\mathbb{E}(S^2) &= \frac{1}{n-1}\mathbb{E}(\sum_{i=1}^n X_i^2 - n\bar{X}^2) \\
&= \frac{1}{n-1}(n(\lambda+\lambda^2) - n(\frac{\lambda}{n} + \lambda^2)) \\
&= \frac{1}{n-1}(n\lambda - \lambda) = \lambda, \text{ y} \\
\mathbb{E}(T_a(X_1, \ldots, X_n)) &= \mathbb{E}(a\bar{X} + (1-a)S^2) = a\mathbb{E}(\bar{X}) + (1-a)\mathbb{E}(S^2) = \lambda.
\end{aligned}$$

Así, se tiene una familia infinita de estimadores insesgados para λ; entonces se puede optar por utilizar el estimador que tenga el menor ECM.

Definición 1.12 *Un estimador $T^*(\underline{X})$ insesgado de varianza mínima uniformemente (UMVUE[4]) para $\tau(\theta)$ satisface:*

[4] Por Uniformly Minimum Variance Unbiased Estimator.

(a) $T^*(\underline{X}) \in \mathcal{C}_{\tau(\theta)}$, *es decir,* $\mathbb{E}[T^*(\underline{X})] = \tau(\theta)$.

(b) Para todo $\theta \in \Theta$, $Var(T^*(\underline{X})) \leq Var(T(\underline{X}))$, *donde* $T(\underline{X})$ *es cualquier otro estimador en* $\mathcal{C}_{\tau(\theta)}$.

El UMVUE se refiere entonces al mejor estimador insesgado para $\tau(\theta)$ en el sentido de que tiene el menor error cuadrático medio para toda $\theta \in \Theta$.

El objetivo de esta sección es encontrar el UMVUE para $\tau(\theta)$, para ello se discutirán tres resultados en donde se utilizan los conceptos analizados previamente. En primer lugar se analizará el planteamiento que Cramèr y Rao hicieron con base en el cálculo de una cota inferior para la varianza de un estimador insesgado. Esta propuesta tiene ciertas restricciones, como el hecho de que requiere el cumplimiento de ciertas condiciones de regularidad para la densidad, entre otras. Posteriormente se enunciará el teorema de Rao-Blackwell, el cual utiliza la suficiencia de una estadística para la construcción de UMVUEs bajo la idea de que un estimador que se basa en una estadística suficiente será mejor que otro que no lo hace. Finalmente se enuncia el Teorema de Lehmann-Scheffé, el cual, además de la suficiencia, utiliza el concepto de completez y permite encontrar un UMVUE construyendo un estimador insesgado a partir de una estadística suficiente y completa, la que a su vez puede hallarse usando los resultados antes vistos o, en su caso, identificando a un miembro de la familia exponencial.

1.3.1. La propuesta de Cramèr y Rao

En esta sección se estudia un resultado propuesto por Cramèr y Rao, el cual se basa en el hecho de que, para ver qué tan bueno es un estimador insesgado con respecto a otro, es necesario analizar la varianza de dicho estimador. Así, si la varianza o el error estándar de un estimador es una cantidad de interés para hablar de su bondad, sería deseable contar con una expresión con la cual pueda compararse esta varianza. Esta expresión será una cota inferior para la varianza, de tal manera que si la varianza de un estimador insesgado es igual a esa cota, se puede afirmar que el estimador es el UMVUE. Antes de presentar el teorema de Cramèr y Rao, en donde se da la cota mencionada, es necesario enunciar algunas definiciones y resultados que servirán para la demostración de dicho teorema.

Definición 1.13 *Sea* $X_1, \ldots, X_n$ *una muestra aleatoria de* $f(x;\theta)$ *y sea* $T(\underline{X})$ *un estimador insesgado de* $\tau(\theta)$. *Las siguientes se conocen como* ***condiciones de regularidad:***

- El soporte de $f(x;\theta)$ se define como $sop(f) = \{x : f(x) > 0\}$ y este es el mismo para toda θ.
- Para todo $x \in sop(f)$, $\frac{\partial}{\partial\theta} \ln f(x;\theta)$ existe.
- $\frac{\partial}{\partial\theta} \int \int \ldots \int T(\underline{x}) f(\underline{x};\theta) dx_1 \ldots dx_n = \int \int \ldots \int \frac{\partial}{\partial\theta} T(\underline{x}) f(\underline{x};\theta) dx_1 \ldots dx_n.$

- $\frac{\partial}{\partial\theta}\int\int...\int f(\underline{x};\theta)dx_1...dx_n = \int\int...\int\frac{\partial}{\partial\theta}f(\underline{x};\theta)dx_1...dx_n$.
- $0 < \mathbb{E}\left[\left(\frac{\partial\ \ln f(x;\theta)}{\partial\theta}\right)^2\right] < \infty$.

Observe que esta definición establece principalmente la condición que debe cumplir una función para que se puedan intercambiar derivadas e integrales, lo cual no siempre se cumple. En general, los miembros de la familia exponencial cumplen las condiciones de regularidad, pero densidades como la uniforme contínua no. Para ver este caso específico de la uniforme considere su función de densidad

$$f(x;\theta) = \frac{1}{\theta}\mathbb{I}_{(0,\theta)}(x).$$

A continuación se obtiene la derivada con respecto a θ de la integral, de la siguiente manera:

$$\frac{\partial}{\partial\theta}\int_0^\theta t(x)\, f(x;\theta)dx = \frac{\partial}{\partial\theta}\int_0^\theta t(x)\,\frac{1}{\theta}dx.$$

Utilizando la regla de Leibnitz, la cual es una aplicación del Teorema Fundamental del Cálculo y de la regla de la cadena, y que establece que si $h(x;\theta)$, $a(\theta)$ y $b(\theta)$ son diferenciables con respecto a θ, entonces

$$\begin{aligned}\frac{\partial}{\partial\theta}\int_{a(\theta)}^{b(\theta)} h(x;\theta)dx &= h(b(\theta),\theta)\frac{\partial}{\partial\theta}b(\theta) - h(a(\theta),\theta)\frac{\partial}{\partial\theta}a(\theta) \\ &\quad + \int_{a(\theta)}^{b(\theta)}\frac{\partial}{\partial\theta}h(x;\theta)dx.\end{aligned}$$

En el caso que se está analizando, $a(\theta) = 0$, $b(\theta) = \theta$ y $h(x;\theta) = t(x)\frac{1}{\theta}$.

$$\begin{aligned}\frac{\partial}{\partial\theta}\int_0^\theta t(x)\frac{1}{\theta}dx &= \frac{t(\theta)}{\theta} + \int_0^\theta t(x)\frac{\partial}{\partial\theta}\left(\frac{1}{\theta}\right)dx \\ &\neq \int_0^\theta t(x)\frac{\partial}{\partial\theta}\left(\frac{1}{\theta}\right)dx,\end{aligned}$$

al menos que $\frac{t(\theta)}{\theta} = 0$.

Ahora se definirán algunas funciones que están involucradas en la cota inferior para la varianza propuesta por Cramèr y Rao.

Definición 1.14 *La **función score** o **función de puntaje** se define como:*

$$Sc(\underline{x};\theta) = \frac{\partial}{\partial\theta}\ln f(\underline{x};\theta).$$

Definición 1.15 *La* ***información esperada de Fisher*** *se define como:*

$$I_{\underline{X}}(\theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ln f(\underline{X};\theta)\right)^2\right] = \mathbb{E}\left[(Sc)^2\right].$$

Observación 1.2 *La función score también puede escribirse como:*

$$Sc(\underline{x};\theta) = \frac{\partial}{\partial\theta}\ln f(\underline{x};\theta) = \frac{f'(\underline{x};\theta)}{f(\underline{x};\theta)} = \frac{\partial}{\partial\theta}\ln\prod_{i=1}^{n} f(x_i;\theta) = \sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln f(x_i;\theta). \tag{1.20}$$

Lema 1.2 *Si se satisfacen las condiciones de regularidad, entonces:*

(a) $\mathbb{E}(Sc) = 0$.

(b) $Var(Sc) = I_{\underline{X}}(\theta)$.

Demostración:

(a)

$$\begin{aligned}
\mathbb{E}\left[Sc(\underline{x};\theta)\right] &= \int\int ... \int \left(\frac{\partial}{d\theta}\ln f(\underline{x};\theta)\right) f(\underline{x};\theta)dx_1...dx_n \\
&= \int\int ... \int \frac{f'(\underline{x};\theta)}{f(\underline{x};\theta)} f(\underline{x};\theta)dx_1...dx_n \\
&= \int\int ... \int \frac{\partial}{\partial\theta} f(\underline{x};\theta)dx_1...dx_n \\
&= \frac{\partial}{\partial\theta}\int\int ... \int f(\underline{x};\theta)dx_1...dx_n \\
&= \frac{\partial}{\partial\theta}(1) = 0
\end{aligned}$$

$\therefore\ \mathbb{E}(Sc) = 0.$

(b)

$$Var(Sc) = \mathbb{E}(Sc^2) - \mathbb{E}^2(Sc) = I_{\underline{X}}(\theta) - 0 = I_{\underline{X}}(\theta).$$

$$\therefore\ Var(Sc) = I_{\underline{X}}(\theta).$$

■

Definición 1.16 *Si X es una variable aleatoria, entonces a*

$$I_X(\theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ln f(X;\theta)\right)^2\right]$$

se le conoce como ***información esperada de Fisher por unidad muestral****.*

Es más sencillo calcular la información esperada de Fisher por unidad muestral y el siguiente resultado la relaciona con la información esperada de Fisher para la muestra, así como con otras expresiones.

Lema 1.3 *Si se cumplen las condiciones de regularidad, entonces:*

(a) $I_{\underline{X}}(\theta) = nI_X(\theta)$.

(b) $I_{\underline{X}}(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2} ln\ f(\underline{X};\theta)\right]$.

(c) $I_{\underline{X}}(\theta) = -n\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2} ln\ f(X;\theta)\right]$.

Demostración:

(a) Como $I_{\underline{X}}(\theta) = \mathbb{E}(Sc^2)$ y usando (1.20), así como el hecho de que $\left(\sum a_i\right)^2 = \sum a_i^2 + \sum_{i\neq j} a_i a_j$,

$$\begin{aligned} I_{\underline{X}}(\theta) &= \mathbb{E}\left[\left(\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln f(X_i;\theta)\right)^2\right] \\ &= \sum_{i=1}^{n}\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ln f(X_i;\theta)\right)^2\right] \\ &\quad + \sum_{i\neq j}\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ln f(X_i;\theta)\right)\left(\frac{\partial}{\partial\theta}\ln f(X_j;\theta)\right)\right]. \end{aligned}$$

Como las variables $X_1,...,X_n$ son independientes, se tiene que

$$\left(\frac{\partial}{\partial\theta}\ln f(X_i;\theta)\right)$$

y

$$\left(\frac{\partial}{\partial\theta}\ln f(X_j;\theta)\right),$$

también lo son y

$$\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ln f(X_i;\theta)\right)\left(\frac{\partial}{\partial\theta}\ln f(X_j;\theta)\right)\right]$$

es igual a

$$\mathbb{E}\left(\frac{\partial}{\partial\theta}\ln f(X_i;\theta)\right)\mathbb{E}\left(\frac{\partial}{\partial\theta}\ln f(X_j;\theta)\right),$$

donde, para el caso continuo, y bajo el supuesto de que se cumplen las condiciones de regularidad:

$$\begin{aligned}
\mathbb{E}\left(\frac{\partial}{\partial\theta}\ln f(X_i;\theta)\right) &= \int_{-\infty}^{\infty}\frac{\frac{\partial}{\partial\theta}f(x_i;\theta)}{f(x_i;\theta)}f(x_i;\theta)dx_i \\
&= \int_{-\infty}^{\infty}\frac{\partial}{\partial\theta}f(x_i;\theta)dx \\
&= \frac{\partial}{\partial\theta}\int_{-\infty}^{\infty}f(x_i;\theta)dx = \frac{\partial}{\partial\theta}(1) = 0.
\end{aligned}$$

Así,

$$\begin{aligned}
I_{\underline{X}}(\theta) &= \sum_{i=1}^{n}\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ln f(X_i;\theta)\right)^2\right] \\
&= n\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ln f(X;\theta)\right)^2\right],
\end{aligned}$$

debido a que las X_i's son idénticamente distribuidas.

(b) Observe que

$$\begin{aligned}
\frac{\partial^2}{\partial\theta^2}\ln f(\underline{x};\theta) &= \frac{\partial}{\partial\theta}\frac{f'(\underline{x};\theta)}{f(\underline{x};\theta)} \\
&= \frac{f(\underline{x};\theta)f''(\underline{x};\theta) - f'(\underline{x};\theta)f'(\underline{x};\theta)}{[f(\underline{x};\theta)]^2} \\
&= \frac{f''(\underline{x};\theta)}{f(\underline{x};\theta)} - \left[\frac{f'(\underline{x};\theta)}{f(\underline{x};\theta)}\right]^2.
\end{aligned}$$

Así,

$$-\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}ln\ f(\underline{X};\theta)\right] = -\mathbb{E}\left\{\frac{f''(\underline{X};\theta)}{f(\underline{X};\theta)} - \left[\frac{f'(\underline{X};\theta)}{f(\underline{X};\theta)}\right]^2\right\},$$

y como

$$\begin{aligned}\mathbb{E}\left[\frac{f''(\underline{X};\theta)}{f(\underline{X};\theta)}\right] &= \int\cdots\int \frac{f''(\underline{x};\theta)}{f(\underline{x};\theta)} f(\underline{x};\theta)\,dx_1\cdots dx_n \\ &= \frac{\partial^2}{\partial\theta^2}\int\cdots\int f(\underline{x};\theta)\,dx_1\cdots dx_n = 0,\end{aligned}$$

$$\begin{aligned}-\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2} ln\ f(\underline{X};\theta)\right] &= \mathbb{E}\left[\left(\frac{f'(\underline{X};\theta)}{f(\underline{X};\theta)}\right)^2\right] \\ &= \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ln f(\underline{X};\theta)\right)^2\right] \\ &= I_{\underline{X}}(\theta).\end{aligned}$$

(c) Se deduce de los dos resultados anteriores.

■

Teorema 1.4 *(de **Cramèr y Rao**). Sean $X_1, \ldots, X_n$ una muestra aleatoria de $f(x;\theta)$ y $T(\underline{X})$ un estimador insesgado de $\tau(\theta)$. Si se satisfacen las condiciones de regularidad, entonces*

$$Var(T) \geq \underbrace{\frac{(\tau'(\theta))^2}{I_{\underline{X}}(\theta)}}_{CICR(\tau(\theta))}. \tag{1.21}$$

*Esta desigualdad se conoce como la **Desigualdad de Cramèr-Rao** o desigualdad de la información y a la cantidad $\frac{[\tau'(\theta)]^2}{I_{\underline{X}(\theta)}}$ como la **Cota Inferior de Cramèr y Rao** (CICR).*

En (1.21) la igualdad se da si sólo si:

$$\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln f(x_i;\theta) = k(\theta;n)[T(\underline{x}) - \tau(\theta)],$$

donde k puede depender de θ y de n.

Demostración:

Este resultado se deduce de la conocida desigualdad de Cauchy-Schwarz, la cual establece que si X y Y son variables aleatorias, entonces:

$$\{Cov(X,Y)\}^2 \leq Var(X)\,Var(Y),$$

dándose la igualdad si y sólo si

$$Y - \mathbb{E}(Y) = k[X - \mathbb{E}(X)] \tag{1.22}$$

Aplicando esta desigualdad a las variables $T(\underline{X})$ y $Sc(\underline{X};\theta)$, se obtiene:

$$\{Cov(T, S_C)\}^2 \leq Var(T)\, Var(S_C). \tag{1.23}$$

Usando el Lema 1.2, se tiene que $Var(Sc) = I_{\underline{X}}(\theta)$, por lo que (1.23) se puede escribir como:

$$Var(T) \geq \frac{\{Cov(T, S_C)\}^2}{I_{\underline{X}}(\theta)}.$$

Por otro lado,

$$Cov(T, S_C) = \mathbb{E}(TS_C) - \mathbb{E}(T)\,\mathbb{E}(S_C)$$

y nuevamente por el Lema 1.2, $\mathbb{E}(S_C) = 0$, mientras que:

$$\begin{aligned}
\mathbb{E}(TS_C) &= \int \cdots \int t(\underline{x}) \frac{\frac{\partial}{\partial\theta} f(\underline{x};\theta)}{f(\underline{x};\theta)} f(\underline{x};\theta)\, dx_1 \cdots dx_n \\
&= \frac{\partial}{\partial\theta} \int \cdots \int t(\underline{x})\, f(\underline{x};\theta)\, dx_1 \cdots dx_n \\
&= \frac{\partial}{\partial\theta} \mathbb{E}(T(\underline{X})) = \frac{\partial}{\partial\theta} \tau(\theta) = \tau'(\theta),
\end{aligned}$$

los pasos anteriores se justifican por la definición de S_C, las condiciones de regularidad y el hecho de que T es insesgado para $\tau(\theta)$. Así,

$$Var(T) \geq \frac{\{\tau'(\theta)\}^2}{I_{\underline{X}}(\theta)}.$$

Para ver la condición en la que se alcanza la cota, es decir, en la que se da la igualdad, se usa (1.22), obteniendo:

$$S_C - \mathbb{E}(S_C) = k[T - \mathbb{E}(T)]$$

y como $\mathbb{E}(S_C) = 0$, $S_C = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln f(x_i;\theta)$ y $\mathbb{E}(T) = \tau(\theta)$, se comprueba la segunda parte del teorema.

■

Ejemplo 1.30 *Sea $X_1, \ldots, X_n$ una muestra aleatoria de la distribución $N(0, \sigma^2)$. Para encontrar $I_{\underline{X}}(\sigma^2)$:*

$$\begin{aligned}
\ln f(x;\theta) &= \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}x^2}\right) \\
&= -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}x^2, \\
\frac{\partial}{\partial\sigma^2}\ln f(x;\sigma^2) &= -\frac{1}{2\sigma^2} + \frac{x^2}{2(\sigma^2)^2}, \\
\frac{\partial^2}{\partial(\sigma^2)^2}\ln f(x;\sigma^2) &= \frac{1}{2(\sigma^2)^2} - \frac{x^2}{(\sigma^2)^3}.
\end{aligned}$$

Entonces,

$$\begin{aligned}
I_{\underline{X}}(\sigma^2) &= -n\mathbb{E}\left[\frac{\partial^2}{\partial(\sigma^2)^2}\ln f(X;\sigma^2)\right] = n\left[\frac{\mathbb{E}(X^2)}{\sigma^6} - \frac{1}{2\sigma^4}\right] \\
&= n\left[\frac{\sigma^2}{\sigma^6} - \frac{1}{2\sigma^4}\right] = n\left[\frac{1}{\sigma^4} - \frac{1}{2\sigma^4}\right] = \frac{n}{2\sigma^4}.
\end{aligned}$$

Entonces, la CICR para estimadores insesgados de σ^2 es $\frac{2\sigma^4}{n}$.

Ejemplo 1.31 *Sea $X_1, \ldots, X_n$ una muestra aleatoria de la distribución $exp(\theta)$. Para encontrar $I_{\underline{X}}(\theta)$:*

$$\ln f(x;\theta) = \ln\left(\theta e^{-\theta x}\right) = \ln(\theta) - \theta x,$$

de donde

$$\frac{\partial}{\partial\theta}\ln f(x;\theta) = \frac{1}{\theta} - x.$$

Entonces,

$$I_{\underline{X}}(\theta) = n\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ln f(X;\theta)\right)^2\right] = n\mathbb{E}\left[\left(\frac{1}{\theta} - X\right)^2\right] = n\ Var(X) = \frac{n}{\theta^2}.$$

Para encontrar la CICR para estimadores insesgados de θ:

$$\tau_1(\theta) = \theta \Rightarrow \tau_1'(\theta) = 1.$$

Entonces,

$$CICR(\theta) = \frac{1}{I_{\underline{X}}(\theta)} = \frac{\theta^2}{n}.$$

Para encontrar la CICR para estimadores insesgados de $\tau_2(\theta) = \frac{1}{\theta}$:

$$\tau_2(\theta) = \frac{1}{\theta} \Rightarrow \tau_2'(\theta) = -\frac{1}{\theta^2}.$$

Entonces,

$$CICR(\tau_2(\theta)) = \frac{1/\theta^4}{I_{\underline{X}}(\theta)} = \frac{1/\theta^4}{n/\theta^2} = \frac{1}{n\theta^2}.$$

Observación 1.3 *Para responder a la pregunta: ¿existe alguna función de θ, $\tau(\theta)$, para la cual hay un estimador insesgado cuya varianza coincide con la CICR?, se usa la segunda parte del teorema, es decir, la condición para la alcanzabilidad de la cota.*

Ejemplo 1.32 *Para la distribución exponencial, ¿existe alguna función de θ, $\tau(\theta)$, para la cual hay un estimador cuya varianza coincide con la CICR?.*

Usando la segunda parte del Teorema de Cramèr y Rao, se tiene que

$$\begin{aligned}
\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln f(x_i;\theta) &= \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln\left[\theta e^{-\theta x_i}\right] = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \left[\ln\theta - \theta x_i\right] \\
&= \sum_{i=1}^{n} \left[\frac{1}{\theta} - x_i\right] = \frac{n}{\theta} - \sum_{i=1}^{n} x_i \\
&= -n\left[\frac{\sum_{i=1}^{n} x_i}{n} - \frac{1}{\theta}\right] = -n\left[\overline{x} - \frac{1}{\theta}\right].
\end{aligned}$$

Así, se puede afirmar que $\tau(\theta) = \frac{1}{\theta}$ es una función de θ para la cual existe un estimador insesgado $T(\underline{X}) = \overline{X}$, cuya varianza coincide con la CICR. En otras palabras, $\overline{X}$ es el UMVUE de $\tau(\theta) = \frac{1}{\theta}$. Aunque en general no es necesario probarlo, es claro que en este caso: $Var\left(\overline{X}\right) = \frac{1}{\theta^2 n} = CICR(\tau(\theta))$.

Observaciones 1.2

1. *Si la varianza de un estimador insesgado coincide con la CICR, entonces el estimador es un UMVUE. Pero el UMVUE puede existir sin que su varianza coincida con la CICR.*

2. *Si la muestra aleatoria es de algún miembro de la familia exponencial, siempre existe una función de θ para la cual hay un estimador insesgado cuya varianza coincide con la CICR (basta factorizar*

$$\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln a(\theta)b(x)\exp\{c(\theta)d(x_i)\}$$

   en la forma indicada en la segunda parte del teorema de Cramèr y Rao).

3. *Aún cuando la varianza de un estimador insesgado alcance la CICR, esta situación se da para una función específica de θ, que puede no ser la que se esté analizando. En el caso de la distribución exponencial, en el ejemplo 1.32 se obtuvo que $\overline{X}$ es el UMVUE*

de $\frac{1}{\theta}$ usando la segunda parte del teorema de Cramèr y Rao; sin embargo, si el objetivo es encontrar el UMVUE de θ, este resultado no da información adicional (salvo la expresión correspondiente para la CICR que sirve para compararla con la varianza de algún estimador que se proponga).

4. *La teoría desarrollada por Cramèr y Rao sólo es para densidades que satisfacen las condiciones de regularidad.*

5. *Cuando la varianza de un estimador alcanza la CICR también se dice que es eficiente y la eficiencia de un estimador insesgado se mide como $\frac{CICR}{Var(T)}$, cantidad que es menor o igual a 1. Por lo que un estimador es eficiente si y sólo si el cociente anterior es* 1.

Dadas estas restricciones se analizarán otros resultados que incorporan los conceptos de suficiencia y completez, lo cual se hará en las secciones 1.3.2 y 1.3.3.

Generalización

Aquí se considerarán brevemente la generalización de la teoría de Cramèr y Rao para cuando se tienen distribuciones de dos o más parámetros. En el caso de dos parámetros, la información esperada de Fisher (para una muestra de tamaño n), llamada la *matriz de información de Fisher*, se define como:

$$I_{\underline{X}}(\underline{\theta}) = -\begin{bmatrix} E\left[\frac{\partial^2}{\partial\theta_1}\ln f(\underline{X};\underline{\theta})\right] & E\left[\frac{\partial^2}{\partial\theta_1\partial\theta_2}\ln f(X;\underline{\theta})\right] \\ E\left[\frac{\partial^2}{\partial\theta_2\partial\theta_1}\ln f(\underline{X};\underline{\theta})\right] & E\left[\frac{\partial^2}{\partial\theta_2}\ln f(\underline{X};\underline{\theta})\right] \end{bmatrix},$$

y para el caso de k parámetros $I_{\underline{X}}(\underline{\theta})$ toma la forma:

$$\begin{bmatrix} E\left[\frac{\partial^2}{\partial\theta_1^2}\ln f(\underline{X};\underline{\theta})\right] & E\left[\frac{\partial^2}{\partial\theta_1\partial\theta_2}\ln f(X;\underline{\theta})\right] & \cdots & E\left[\frac{\partial^2}{\partial\theta_1\partial\theta_k}\ln f(X;\underline{\theta})\right] \\ E\left[\frac{\partial^2}{\partial\theta_2\partial\theta_1}\ln f(\underline{X};\underline{\theta})\right] & E\left[\frac{\partial^2}{\partial\theta_2^2}\ln f(\underline{X};\underline{\theta})\right] & \cdots & E\left[\frac{\partial^2}{\partial\theta_2\partial\theta_k}\ln f(\underline{X};\underline{\theta})\right] \\ \vdots & \vdots & \ddots & \vdots \\ E\left[\frac{\partial^2}{\partial\theta_k\partial\theta_1}\ln f(\underline{X};\underline{\theta})\right] & E\left[\frac{\partial^2}{\partial\theta_k\partial\theta_2}\ln f(\underline{X};\underline{\theta})\right] & \cdots & E\left[\frac{\partial^2}{\partial\theta_k^2}\ln f(\underline{X};\underline{\theta})\right] \end{bmatrix}.$$

Y la cota inferior de Cramèr y Rao es la inversa de la matriz de información, es decir, $I_{\underline{X}}^{-1}(\underline{\theta})$.

1.3.2. El teorema de Rao-Blackwell

Como se ha visto, una estadística suficiente conserva toda la información relevante contenida en la muestra acerca del parámetro de interés. Así, los estimadores basados en estadísticas suficientes son mejores (que los que no están basados en estadísticas suficientes) como establece el siguiente resultado.

Teorema 1.5 ***(Rao-Blackwell)***. *Sean $T(\underline{X})$ un estimador insesgado para $\tau(\theta)$ y S una estadística suficiente. Sea $T^*(\underline{X}) := \mathbb{E}(T|S)$. Entonces,*

(a) *T^* es una estadistica función de S.*

(b) *T^* es insesgado para $\tau(\theta)$, es decir, $\mathbb{E}(T^*) = \tau(\theta)$.*

(c) *$Var(T^*) \leq Var(T)$ para toda $\theta \in \Theta$.*

Demostración.

(a) Usando la definición de la esperanza condicional en el caso continuo,

$$T^* = \int_{-\infty}^{\infty} t f_{T/S}(t/s)\, dt$$

es una función de S, además $f_{T/S}$ no depende de θ por ser S una estadística suficiente, por lo que T^* es una estadística.

(b) Por las propiedades de la esperanza condicional,

$$\mathbb{E}(T^*) = \mathbb{E}(\mathbb{E}(T/S)) = \mathbb{E}(T) = \tau(\theta).$$

(c) Usando las propiedades de la varianza condicional,

$$Var(T) = Var(\mathbb{E}(T/S)) + \mathbb{E}(Var(T/S)),$$

lo cual implica que

$$Var(T) = Var(T^*) + \mathbb{E}(Var(T/S)),$$

y como $Var(T/S) \geq 0$, se obtiene el resultado.

■

Ejemplo 1.33 *Sea $X_1, \ldots, X_n$ una muestra aleatoria de la distribución Bernoulli(θ). $S(\underline{X}) = \sum_{i=1}^{n} X_i$ es una estadística suficiente para θ, lo cual se ha verificado (basta ver que la distribución Bernoulli pertenece a la familia exponencial). $T(\underline{X}) = X_1$ es un estimador insesgado de θ (pues $\mathbb{E}(X_1) = \theta$).*

Entonces

$$
\begin{aligned}
T^*\left(\underline{X}\right) = \mathbb{E}(T|S=s) =& \mathbb{E}\left(X_1 \mid \sum_{i=1}^{n} X_i = s\right) \\
=& 0 \cdot \mathbb{P}\left(X_1 = 0 | \sum_{i=1}^{n} X_i = s\right) \\
&+ 1 \cdot \mathbb{P}\left(X_1 = 1 | \sum_{i=1}^{n} X_i = s\right) \\
=& \mathbb{P}\left(X_1 = 1 \mid \sum_{i=1}^{n} X_i = s\right) \\
=& \frac{\mathbb{P}\left(X_1 = 1, \sum_{i=1}^{n} X_i = s\right)}{\mathbb{P}\left(\sum_{i=1}^{n} X_i = s\right)}.
\end{aligned}
$$

Donde $\sum_{i=1}^{n} X_i \sim Bin(n,\theta)$. *Pero*

$$
\frac{\mathbb{P}(X_1 = 1; \sum_{i=1}^{n} X_i = s)}{\mathbb{P}(\sum_{i=1}^{n} X_i = s)} = \frac{\mathbb{P}(X_1 = 1) P(\sum_{i=2}^{n} X_i = s-1)}{\binom{n}{s}\theta^s(1-\theta)^{n-s}}
$$

$$
= \frac{\theta\binom{n-1}{s-1}\theta^{s-1}(1-\theta)^{n-1-s+1}}{\binom{n}{s}\theta^s(1-\theta)^{n-s}} = \frac{\binom{n-1}{s-1}}{\binom{n}{s}} = \frac{\frac{(n-1)!}{(s-1)!(n-s)!}}{\frac{n!}{s!(n-s)!}} = \frac{s}{n}.
$$

Por lo tanto,

$$
T^*(\underline{X}) = \frac{\sum_{i=1}^{n} X_i}{n} = \overline{X}.
$$

El estimador resultante es insesgado, pues $\mathbb{E}(\bar{X}) = \theta$ *y tiene varianza menor que* X_1, *ya que*

$$
Var(\bar{X}) = \frac{\theta(1-\theta)}{n} \leq \theta(1-\theta) = Var(X_1),
$$

las cuales son iguales sólo en el caso $n = 1$.

1.3.3. El teorema de Lehmann-Scheffé

El siguiente resultado muestra que un estimador insesgado función de la estadística suficiente y completa será el UMVUE.

Teorema 1.6 *(**Lehmann-Scheffé**). Sea* $X_1, \ldots, X_n$ *una muestra aleatoria de* $f(x;\theta)$ *y sea* S *una estadística suficiente y completa. Sea* $T^*\left(\underline{X}\right)$ *una función de* S *tal que* $E(T^*) = \tau(\theta)$ *(*T^* *es insesgado para* $\tau(\theta)$*), entonces* T^* *es el UMVUE de* $\tau(\theta)$.

Demostración:

Sea $T'(\underline{X})$, función de S, tal que $\mathbb{E}(T') = \tau(\theta)$. Sea $g(S) = T^* - T'$. Note que

$$\mathbb{E}[g(S)] = \mathbb{E}(T^* - T') = \mathbb{E}[T^*] - \mathbb{E}[T'] = \tau(\theta) - \tau(\theta) = 0$$

$$\therefore\ \mathbb{E}[g(S)] = 0$$

Entonces, por la completez de S se tiene que $\mathbb{P}[g(S) = 0] = 1$, para toda $\theta \in \Theta$. De esta manera, $\mathbb{P}[T^* = T'] = 1$ para toda $\theta \in \Theta$ y, por lo tanto, T^* es único (c.s.). Es decir, T^* es el único estimador insesgado de $\tau(\theta)$ que es función de S.

Por otro lado, sea T tal que $\mathbb{E}[T] = \tau(\theta)$. Por el teorema de Rao-Blackwell, $\mathbb{E}[T|S]$ es estimador insesgado de $\tau(\theta)$ y es función de S, lo que implica que $T^* = \mathbb{E}[T|S]$. Así, por el teorema de Rao-Blackwell, $Var(T^*) \leq Var(T)$, para toda $\theta \in \Theta$.

■

Ejemplo 1.34 *Sea $X_1, \ldots, X_n$ una muestra aleatoria de la distribución $exp(\theta)$, donde $\theta > 0$. Como $f(x;\theta) = \theta e^{-\theta x}$, entonces $f(x;\theta)$ es de la familia exponencial con $d(x) = x$. Entonces, $S(\underline{X}) = \sum_{i=1}^{n} X_i$ es suficiente y completa. Para encontrar el UMVUE de θ y de $\tau(\theta) = \frac{1}{\theta}$, se obtiene*

$$\mathbb{E}(S) = \mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = n\mathbb{E}(X_i) = n\frac{1}{\theta} = \frac{n}{\theta}$$

y como $\mathbb{E}[\bar{X}] = \frac{1}{\theta}$, entonces $\overline{X}$ es el UMVUE de $\frac{1}{\theta}$, pues es función de la estadística suficiente y completa y además es insesgado para θ (note que este resultado coincide con el obtenido mediante la teoría Cramèr y Rao, ejemplo 1.32).

Para encontrar el UMVUE de θ, éste será de la forma $\frac{k}{\sum_{i=1}^{n} X_i}$, donde $S = \sum_{i=1}^{n} X_i$ tiene distribución Gama (n, θ). Entonces, observe que:

$$\begin{aligned}
\mathbb{E}\left[\frac{k}{\sum_{i=1}^{n} X_i}\right] &= \mathbb{E}\left[\frac{k}{S}\right] = k\mathbb{E}\left[\frac{1}{S}\right] \\
&= k\int_0^\infty \frac{1}{s}\frac{\theta^n}{\Gamma(n)} s^{n-1} e^{-\theta s} ds \\
&= k\int_0^\infty \frac{\theta^n}{\Gamma(n)} s^{n-2} e^{-\theta s} ds \\
&= k\frac{\theta^n}{\Gamma(n)}\frac{\Gamma(n-1)}{\theta^{n-1}} \underbrace{\int_0^\infty \frac{\theta^{n-1}}{\Gamma(n-1)} s^{(n-1)-1} e^{-\theta s} ds}_{1} \\
&= k\frac{\theta^n \Gamma(n-1)}{\theta^{n-1}\Gamma(n)} = k\frac{\Gamma(n-1)}{(n-1)\Gamma(n-1)\theta^{-1}} = \frac{k\theta}{n-1}.
\end{aligned}$$

Por lo que, para que k/Y sea insesgado, k debe ser igual a $n-1$. Por lo tanto,

$$T^*(\underline{X}) = \frac{n-1}{\sum_{i=1}^{n} X_i}$$

es el UMVUE de θ.

Para encontrar la CICR para estimadores insesgados de θ (ver ejemplo 1.31):

$$\begin{aligned}
I_{\underline{X}}(\theta) =& n\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ln f(X;\theta)\right)^2\right] \\
=& n\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ln \theta e^{-\theta X}\right)^2\right] \\
=& n\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}(\ln \theta - \theta X)\right)^2\right] \\
=& n\mathbb{E}\left[\left(\frac{1}{\theta} - X\right)^2\right] \\
=& nVar(X) = \frac{n}{\theta^2}.
\end{aligned}$$

Entonces, la CICR para estimadores insesgados de θ es:

$$CICR(\theta) = \frac{1}{I_{\underline{X}}(\theta)} = \frac{\theta^2}{n}.$$

El segundo momento de $T^*(\underline{X}) = \frac{n-1}{\sum_{i=1}^{n} X_i}$ está dado por:

$$\begin{aligned}
\mathbb{E}\left[\frac{(n-1)^2}{S^2}\right] &= (n-1)^2\mathbb{E}\left[\frac{1}{S^2}\right] \\
&= (n-1)^2\int_0^\infty \frac{1}{s^2}\frac{\theta^n}{\Gamma(n)}s^{n-1}e^{-\theta s}ds \\
&= (n-1)^2\frac{\theta^n}{\Gamma(n)}\int_0^\infty s^{n-3}e^{-\theta s}ds \\
&= (n-1)^2\frac{\theta^n}{\Gamma(n)}\frac{\Gamma(n-2)}{\theta^{n-2}}\underbrace{\int_0^\infty \frac{\theta^{n-2}}{\Gamma(n-2)}s^{(n-2)-1}e^{-\theta s}ds}_{1} \\
&= (n-1)^2\frac{\theta^n\Gamma(n-2)}{\theta^{n-2}\Gamma(n)} = (n-1)^2\frac{\Gamma(n-2)}{(n-1)(n-2)\Gamma(n-2)\theta^{-2}} \\
&= \frac{(n-1)\theta^2}{n-2}.
\end{aligned}$$

Entonces,

$$Var(T^*(\underline{X})) = \frac{(n-1)\theta^2}{n-2} - \theta^2 = \frac{\theta^2}{n-2}$$

es la varianza del UMVUE de θ.

Note que

$$Var(T^*(\underline{X})) = \frac{\theta^2}{n-2} > \frac{\theta^2}{n} = CICR(\theta).$$

Ejemplo 1.35 *Sea $X_1, \ldots, X_n$ una muestra aleatoria de la distribución $Poisson(\theta)$. Sea $\tau(\theta) = \mathbb{P}(X = 0) = e^{-\theta}$.*

(a) Encontrar el estimador máximo verosímil de θ y $\tau(\theta)$.

(b) Encontrar el estimador por momentos de θ.

(c) ¿Pertenece $f(x;\theta)$ a la familia exponencial?

(d) Encontrar una estadística suficiente minimal y completa.

(e) Encontrar la CICR para estimadores insesgados de θ y $\tau(\theta)$.

(f) ¿Existirá una función de θ, para la cual hay un estimador insesgado cuya varianza coincide con la CICR? Si es así, encontrarlo.

(g) Encontrar un estimador insesgado de $\tau(\theta)$ y usar el teorema de Rao-Blackwell para hallar un estimador insesgado función de la estadística suficiente.

(h) Decir cuáles son los UMVUEs de θ y $\tau(\theta)$, respectivamente.

Solución:

(a)

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{x_i}}{x_i!}\mathbb{I}_{\{0,1,\ldots\}}^{(x_i)} = e^{-n\theta}\theta^{\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} \frac{\mathbb{I}_{\{0,1,\ldots\}}^{(x_i)}}{x_i!}$$

y el logaritmo de la verosimilitud es

$$l(\theta) = -n\theta + (\sum_{i=1}^{n} x_i)\ln\theta + \ln\left(\prod_{i=1}^{n} \frac{\mathbb{I}_{\{0,1,\ldots\}}^{(x_i)}}{x_i!}\right),$$

de donde,

$$\frac{\partial}{\partial\theta}l(\theta) = -n + \frac{\sum_{i=1}^{n} x_i}{\theta}.$$

Entonces, $\frac{\partial}{\partial\theta}l(\theta) = 0$ si y sólo si

$$-n + \frac{\sum_{i=1}^{n} x_i}{\widehat{\theta}} = 0 \Leftrightarrow n = \frac{\sum_{i=1}^{n} x_i}{\widehat{\theta}} \Leftrightarrow \widehat{\theta} = \frac{\sum_{i=1}^{n} x_i}{n}.$$

Por lo tanto $\hat{\theta}_{M.V.} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$. Para $\tau(\theta)$, aplicando la propiedad de invarianza de los estimadores máximo verosímiles $\tau(\hat{\theta}_{M.V.})$ es estimador máximo versímil de $\tau(\theta)$. Por lo tanto $e^{-\bar{X}}$ es estimador máximo verosímil de $\tau(\theta) = e^{-\theta}$.

(b) Recuerde que $\mathbb{E}(X) = \theta$, entonces el estimador por momentos está dado por

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

(c) Como

$$f(x;\theta) = \frac{e^{-\theta}\theta^{x}}{x!}\mathbb{I}_{\{0,1,\ldots\}}^{(x)}$$

si

$$a(\theta) = e^{-\theta}, \qquad b(x) = \frac{1}{x!}\mathbb{I}_{\{0,1,\ldots\}}^{(x)}, \qquad c(\theta) = \ln(\theta), \qquad d(x) = x.$$

Entonces,

$$f(x;\theta) = a(\theta)b(x)e^{c(\theta)d(x)}.$$

Por lo tanto pertenece a la familia exponencial.

(d) Como $f(x;\theta)$ pertenece a la familia exponencial entonces $T(x) = \sum_{i=1}^{n} d(X_i) = \sum_{i=1}^{n} X_i$ es una estadística suficiente minimal y completa.

(e) La información esperada de Fisher está dada por

$$
\begin{aligned}
I_{\underline{X}}(\theta) =& n\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ln f(X;\theta)\right)^2\right] \\
=& n\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ln\left(\frac{e^{-\theta}\theta^X}{X!}\right)\right)^2\right] \\
=& n\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}(-\theta + X\ln\theta - \ln X!)\right)^2\right] \\
=& n\mathbb{E}\left[\left(-1 + \frac{X}{\theta}\right)^2\right] \\
=& n\mathbb{E}\left[\left(\frac{1}{\theta}(X-\theta)\right)^2\right] \\
=& \frac{n}{\theta^2}\mathbb{E}\left[(X-\theta)^2\right] = \frac{n}{\theta^2}Var(X) = \frac{n\theta}{\theta^2} = \frac{n}{\theta}.
\end{aligned}
$$

Para θ se tiene que

$$CICR(\theta) = \frac{\theta}{n}.$$

Para $\tau(\theta) = e^{-\theta}$ se tiene que

$$CICR(\tau(\theta)) = \frac{(\tau'(\theta))^2}{\frac{n}{\theta}} = \frac{\theta e^{-2\theta}}{n}.$$

(f) Utilizando la segunda parte del teorema de Cramèr-Rao

$$
\begin{aligned}
\sum_{i=1}^n \frac{\partial}{\partial\theta}\ln f(x_i;\theta) =& \sum_{i=1}^n \frac{\partial}{\partial\theta}\ln\frac{e^{-\theta}\theta^{x_i}}{x_i!} \\
=& \sum_{i=1}^n \frac{\partial}{\partial\theta}(-\theta + x_i\ln(\theta) - \ln(x_i!)) \\
=& \sum_{i=1}^n \left(-1 + \frac{x_i}{\theta}\right) \\
=& -n + \frac{1}{\theta}\sum_{i=1}^n x_i = -n + \frac{n}{\theta}\bar{x} = \frac{n}{\theta}(\bar{x}-\theta).
\end{aligned}
$$

Por lo tanto, para θ hay un estimador insesgado, $T^*(\underline{X}) = \bar{X}$, cuya varianza coincide con la Cota de Cramèr-Rao, es decir, $\bar{X}$ es el UMVUE de θ.

(g) Considere $T(\underline{X}) = I_{\{0\}}(X_1)$. Note que $\mathbb{E}(T(\underline{X})) = \mathbb{E}(I_{\{0\}}(X_1)) = \mathbb{P}(X_1 = 0) = e^{-\theta}$. Por tanto $T(\underline{X})$ es un estimador insesgado de $\tau(\theta)$ y ya se vió que $S(X) = \sum_{i=1}^{n} X_i$ es una estadística suficiente minimal y completa. Entonces,

$$\begin{aligned}
\mathbb{E}(T|S &= s) = \mathbb{E}(\mathbb{I}_{\{0\}}(X_1)|S = s) \\
&= \mathbb{P}\left(X_1 = 0|\sum_{i=1}^{n} X_i = s\right) \\
&= \frac{\mathbb{P}(X_1 = 0)\mathbb{P}(\sum_{i=2}^{n} X_i = s)}{\mathbb{P}(\sum_{i=1}^{n} X_i = s)} \\
&= \frac{e^{-\theta}\frac{e^{-(n-1)\theta}((n-1)\theta)^s}{s!}}{\frac{e^{-n\theta}(n\theta)^s}{s!}} \\
&= \frac{e^{-\theta}e^{-(n-1)\theta}((n-1)\theta)^s}{e^{-n\theta}(n\theta)^s} = \frac{e^{-\theta}e^{-n\theta}e^{\theta}(n-1)^s\theta^s}{e^{-n\theta}n^s\theta^s} \\
&= \left(\frac{n-1}{n}\right)^s.
\end{aligned}$$

Entonces, por el teorema de Rao-Blackwell $T^*(\underline{X}) = \left(\frac{n-1}{n}\right)^{\sum_{i=1}^{n} X_i}$.

(h) $\bar{X}$ es el UMVUE de θ, lo cual se justifica utilizando el inciso (f) o bien, notando que $\bar{X}$ es insesgado para θ y función de la estadística suficiente y completa, $\sum_{i=1}^{n} X_i$, por lo que usando el Teorema de Lehmann-Scheffé se llega a la misma conclusión. Además, $\left(\frac{n-1}{n}\right)^{\sum_{i=1}^{n} X_i}$es el UMVUE de $\tau(\theta)$ por el inciso (g) y el Teorema de Lehmann-Scheffé.

1.4. Propiedades asintóticas de los estimadores

Hasta ahora se han estudiado distintas propiedades de los estimadores, pero la mayoría, a excepción de la consistencia vista en el apartado 1.2.2, se refiere a tamaños de muestra pequeños. En esta sección se abordarán propiedades que describen el comportamiento de un estimador cuando el tamaño de muestra es grande, es decir, las propiedades asintóticas de los estimadores.

Como ya se señaló, la consistencia tiene que ver con la precisión asintótica de un estimador, esto es, que el error cometido al estimar $\tau(\theta)$ con $T_n(\underline{X})$ es pequeño cuando el tamaño de muestra es grande. Existe otra propiedad que tiene que ver con la varianza asintóntica de un estimador, la cual se conoce como eficiencia.

Ya se ha reiterado que la varianza (de los estimadores) juega un papel importante en la elección del mejor estimador. De hecho en las observaciones 1.2, se menciona la propiedad de eficiencia. A continuación se da una definición formal de eficiencia asintótica.

Definición 1.17 *Una sucesión de estimadores $\{T_n\}$ es asintóticamente eficiente para un parámetro $\tau(\theta)$ si*

$$\sqrt{n}[T_n - \tau(\theta)] \longrightarrow N[0, CICR(\theta)]$$

en distribución, donde

$$CICR(\theta) = \frac{[\tau\prime(\theta)]^2}{\mathbb{E}\left[(\frac{\partial}{\partial\theta}\ln f(\underline{X};\theta))^2\right]};$$

esto es, la varianza asintótica de T_n alcanza la cota inferior de Cramér-Rao.

Bajo las condiciones de regularidad (Definición 1.13), se puede demostrar que los estimadores máximo verosímiles cumplen las propiedades de consistencia y eficiencia. Es decir, si la muestra aleatoria proviene de una población con función de densidad que satisface las condiciones de regularidad, entonces el estimador máximo verosímil del parámetro θ (o de una función $\tau(\theta)$) tiene estas propiedades asintóticas.

Para el caso de la consistencia simple, puede consultarse Stuart, Ord y Arnold (1999). En cuanto a la eficiencia, se utilizará el siguiente resultado conocido como el método delta:

Lema 1.7 *si $\{X_n\}$ es una sucesión de variables aleatorias que satisfacen que $\sqrt{n}(X_n - \theta) \rightarrow N(0, \sigma^2)$ en distribución, entonces para una función τ y un valor específico de θ, se tiene que*

$$\sqrt{n}[\tau(X_n) - \tau(\theta)] \rightarrow N\left(0, \sigma^2\left(\tau'(\theta)^2\right)\right)$$

en distribución.

Demostración.

El resultado es consecuencia del Teorema de Slutsky, el cual establece que para dos sucesiones de variables aleatorias $\{X_n\}_{n\geq 1}$ y $\{Y_n\}_{n\geq 1}$, tales que $X_n \rightarrow X$ en distribución y $Y_n \rightarrow c$ en probabilidad , donde X es una variable aleatoria y c es una constante, se tiene que:

(i) $X_n + Y_n \rightarrow X + c$, en distribución,

(ii) $X_n Y_n \rightarrow cX$, en distribución,

(iii) Si $c \neq 0$ entonces,

$$\frac{X_n}{Y_n} \rightarrow \frac{X}{c},$$

en distribución.

■

El siguiente resultado se refiere a la eficiencia asintótica de los estimadores máximo verosímiles.

Teorema 1.8 *Sea $X_1, X_2, .., X_n$ una muestra aleatoria de una población con función de densidad $f(x;\theta)$, sea $\widehat{\theta}$ el estimador máximo verosímil de θ, y sea $\tau(\theta)$ una función continua y diferenciable de θ. Bajo las condiciones de regularidad sobre $f(x;\theta)$ y, por lo tanto de la función de verosimilitud $L(\theta)$, se tiene que*

$$\sqrt{n}[\tau(\widehat{\theta}) - \tau(\theta)] \longrightarrow N[0, CICR(\tau(\theta))],$$

donde $CICR(\tau(\theta))$ es la cota inferior de Cramér-Rao para estimadores insesgados de $\tau(\theta)$. Esto es, el estimador máximo verosímil de $\tau(\theta)$, $\tau(\widehat{\theta})$, es un estimador eficiente de $\tau(\theta)$.

Demostración.
Se demostrará el caso $\tau(\theta) = \theta$, es decir, que $\widehat{\theta}$ es asintóticamente eficiente. Para ello, recuerde que

$$l(\theta) = \sum_{i=1}^{n} \ln f(x_i;\theta)$$

es la función de log-verosimilitud. Sean $l', l'', ..$ las derivadas (con respecto a θ) . Expandiendo la primera derivada de la log-verosimilitud alrededor del valor verdadero del parámetro, el cual se denotará por θ_0,

$$l'(\theta) = l'(\theta_0) + (\theta - \theta_0)l''(\theta_0) + ...,$$

donde se ignoran los términos de orden superior.
Sustituyendo el estimador máximo verosímil $\widehat{\theta}$ en lugar de θ, se tiene que

$$l'(\widehat{\theta}) = l'(\theta_0) + (\widehat{\theta} - \theta_0)l''(\theta_0) + ...,$$

pero el estimador máximo verosímil es el cero de la función de verosimilitud, por lo que

$$l'(\theta_0) + (\widehat{\theta} - \theta_0)l''(\theta_0) + ... = 0.$$

Así que reacomodando los términos y multiplicando por $\sqrt{n}$, se obtiene que:

$$\begin{aligned} \sqrt{n}(\widehat{\theta} - \theta_0) &= \sqrt{n}\frac{-l'(\theta_0)}{l''(\theta_0)} \\ &= \frac{-\frac{1}{\sqrt{n}}l'(\theta_0)}{\frac{1}{n}l''(\theta_0)}. \end{aligned}$$

En 1.3.1 , se vio que

$$I_{\underline{X}}(\theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ln f(\underline{X};\theta)\right)^2\right].$$

Como
$$l'(\theta) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln f(x_i;\theta),$$
por (1.20)
$$I_{\underline{X}}(\theta_0) = E\left\{[l'(\theta_0)]^2\right\} = \frac{1}{CICR(\theta)}$$
denota la información esperada de Fisher.

Ahora observe que
$$\frac{1}{\sqrt{n}} l'(\theta_0) = \sqrt{n}\left[\frac{1}{n}\sum_i \frac{\frac{\partial}{\partial\theta} f(x_i;\theta)}{f(x_i;\theta)}\right],$$
donde $Sc(\underline{x};\theta) = \frac{f'(\underline{x};\theta)}{f(\underline{x};\theta)}$ es tal que $E\left[Sc(\underline{X};\theta)\right] = 0$ y $Var\left[Sc(\underline{X};\theta)\right] = I_{\underline{X}}(\theta)$, lo cual se probó en el Lema 1.2. Así, por el Teorema de Límite Central,
$$\frac{1}{\sqrt{n}} l'(\theta_0) \longrightarrow N[0, I(\theta_0)]$$
en distribución y
$$-\frac{1}{\sqrt{n}} l'(\theta_0) \longrightarrow N[0, I(\theta_0)]$$
en distribución. Por otro lado,
$$\frac{1}{n} l''(\theta_0) = \frac{1}{n}\sum_i \left[\frac{\frac{\partial}{\partial\theta} f(x_i;\theta)}{f(x_i;\theta)}\right]^2 - \frac{1}{n}\sum_i \frac{\frac{\partial^2}{\partial\theta^2} f(x_i;\theta)}{f(x_i;\theta)}.$$
Note que la esperanza del primer sumando es $I_{\underline{X}}(\theta_0)$, mientras que la del segundo es cero (ver la demostración del Lema 1.3). Entonces por la Ley Débil de los Grandes Números:
$$\frac{1}{n} l''(\theta_0) \longrightarrow I(\theta_0),$$
en probabilidad.

En consecuencia, si W es una variable aleatoria tal que $W \sim N[0, I(\theta_0)]$, entonces
$$\sqrt{n}(\widehat{\theta} - \theta_0) = \frac{-\frac{1}{\sqrt{n}} l'(\theta_0)}{\frac{1}{n} l''(\theta_0)}$$
converge en distribución a $W/I(\theta_0) \sim N[0, 1/I_{\underline{X}}(\theta_0)]$, es decir, a una variable aleatoria normal con media cero y varianza igual a la Cota Inferior de Cramèr y Rao, lo que demuestra el resultado.

El caso general es consecuencia del método delta. Una forma alternativa para calcular la varianza del estimador máximo verosímil de $\tau(\theta)$ es considerando que, debido a la propiedad de invarianza que tiene el método de máxima verosimilitud,

$$\widehat{\tau(\theta)} = \tau\left(\widehat{\theta}\right).$$

Si se aproxima $\tau\left(\widehat{\theta}\right)$ mediante una expansión en series de Taylor alrededor de θ, considerando solamente la primera derivada, se obtiene:

$$\tau\left(\widehat{\theta}\right) \approx \tau(\theta) + \left(\widehat{\theta} - \theta\right)\tau'(\theta).$$

Tomando la varianza de ambos lados, se llega a

$$Var\left[\tau\left(\widehat{\theta}\right)\right] \approx \left(\tau'(\theta)\right)^2 Var\left(\widehat{\theta}\right),$$

debido a que θ es una constante. Como ya se había visto, $Var\left(\widehat{\theta}\right)$ está dada por $I_{\underline{X}}^{-1}(\theta)$, asi que

$$Var\left[\tau\left(\widehat{\theta}\right)\right] \approx \frac{\left(\tau'(\theta)\right)^2}{I_{\underline{X}}(\theta)},$$

expresión que corresponde a la Cota Inferior de Cramèr-Rao para estimadores insesgados de $\tau(\theta)$, con lo que puede observarse que la varianza del estimador máximo verosímil alcanza dicha cota (al igual que en el caso $\tau(\theta) = \theta$).

En conclusión:

$$\sqrt{n}[\tau(\widehat{\theta}) - \tau(\theta)] \longrightarrow N[0, CICR(\tau(\theta))].$$

■

Ejemplo 1.36 *Considere una muestra aleatoria, $X_1, \ldots, X_n$, de la población con distribución Bernoulli(p); se desea obtener un estimador puntual para el momio, $\tau(p) = \frac{p}{(1-p)}$, así como la varianza de dicho estimador.*

El estimador máximo versoímil para $\hat{p}$ es $\bar{X}$. Por la propiedad de invarianza de los estimadores máximo verosímiles, se tiene que el estimador máximo verosímil para $\tau(p)$ es $\frac{\bar{X}_n}{(1-\bar{X}_n)}$. La varianza de este estimador puede aproximarse de la siguiente manera:

$$\begin{aligned}
\hat{V}\left(\frac{\hat{p}}{(1-\hat{p})}\right) &= \left(\frac{\left[\frac{d}{dp}\left(\frac{p}{(1-p)}\right)\right]^2}{I_X(p)}\right)_{p=\hat{p}} \\
&= \left(\frac{\left[\frac{1}{(1-p)^2})\right]^2}{\frac{n}{p(1-p)}}\right)_{p=\hat{p}} \\
&= \frac{\hat{p}}{n(1-\hat{p})^3} = \frac{\bar{X}_n}{n(1-\bar{X}_n)^3}.
\end{aligned}$$

Ejemplo 1.37 *Considere una sucesión de variables aleatorias, $X_1, \dots, X_n$, independientes e idénticamente distribuidas de una población con distribución $F(\cdot)$ que es diferenciable. Suponga que se satisface que $P(X_i \leq \psi) = 1/2$, es decir, ψ es la mediana poblacional. Sea M_n la mediana muestral y también suponga que n es impar para simplificar el argumento. Se desea obtener la distribución asintótica de la mediana muestral.*

Se calculará

$$\lim_{n\to\infty} P(\sqrt(n)(M_n - \psi) \leq a),$$

para alguna a. Sean las variables aleatorias $Y_i's$ definidas como

$$Y_i = \begin{cases} 1 & si\ X_i \leq \psi + a/\sqrt(n) \\ 0 & en\ otro\ caso, \end{cases}$$

se tiene que las $Y_i's$ son variables aleatorias Bernoulli con probabilidad de éxito

$$p_n = F(\psi + a/\sqrt(n)).$$

Note que el evento $\{M_n \leq \psi + a/\sqrt(n)\}$ es equivalente al evento $\{\sum_i Y_i \geq (n+1)/2\}$. Dado que

$$p_n \to p = F(\psi) = 1/2,$$

se puede utilizar el Teorema de Límite Central, de donde $\frac{\sum_i Y_i - np_n}{\sqrt(np_n(1-p_n))}$ converge a una variable aleatoria Z con distribución normal estándar. Ahora,

$$\begin{aligned} \lim_{n\to\infty} \frac{(n+1)/2 - np_n}{\sqrt(np_n(1-p_n)} &= \lim_{n\to\infty} \frac{(n+1)F(\psi) - nF(\psi + a/\sqrt{n})}{\sqrt{np_n(1-p_n)}} \\ &= \lim_{n\to\infty} \frac{n(F(\psi) - F(\psi + a/\sqrt{n}))}{\sqrt{np_n(1-p_n)}} \\ &= \lim_{n\to\infty} a\frac{n(F(\psi) - F(\psi + a/\sqrt{n})}{a/\sqrt{n}} \frac{1}{p_n(1-p_n)} \\ &= -2af(\psi). \end{aligned}$$

Por lo que

$$P(\sqrt(n)(M_n - \psi) \leq a) \to P(Z \geq -2af\psi).$$

Así$\sqrt(n)(M_n - \psi)$ tiene una distribución normal con media 0 y varianza $1/[2f(\psi)]^2$.

1.5. Ejercicios

1. Sea X una variable aleatoria con distribución $Gamma(\alpha+1, \beta)$, cuya función de densidad es

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

con $x > 0$, $\alpha > -1$ y $\beta > 0$. Obtenga los estimadores de los parámetros α y β por el método de momentos, para una muestra aleatoria de tamaño n.

2. Una urna contiene bolas negras y blancas. Se toma una muestra aleatoria de tamaño n con reemplazo. ¿Cuál es el estimador máximo verosímil de la razón, R, de blancas a negras en la urna? Para esto suponga que la bola se obtiene una por una con reemplazo hasta que la bola negra aparezca. Sea X el número de bolas requeridas no contando la última obtenida; este procedimiento se repite n veces para una muestra aleatoria $X_1, \ldots, X_n$.

3. Se toma una observación de una variable aleatoria discreta X con función de densidad $f(x;\theta)$ dada en la siguiente tabla, donde $\theta \in \{1, 2, 3\}$.

x	$f(x;1)$	$f(x;2)$	$f(x;3)$
0	1/3	1/4	0
1	1/3	1/4	0
2	0	1/4	1/4
3	1/6	1/4	1/2
4	1/6	0	1/4

Encuentre el estimador máximo verosímil de θ.

4. Sea X una variable aleatoria discreta con función de densidad $f(x;\theta)$ dada en la siguiente tabla, donde $\theta \in \{1, 2, 3\}$ y $X \in \{0, 1, 2, 3, 4\}$. Se toma una muestra aleatoria de tamaño dos, (X_1, X_2). Determine el estimador máximo verosímil de θ.

x	$f(x;1)$	$f(x;2)$	$f(x;3)$
0	1/3	1/4	0
1	1/3	1/4	0
2	0	1/4	1/4
3	1/6	1/4	1/2
4	1/6	0	1/4

5. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con función de densidad

$$f(x;\theta) = \theta x^{-2} \mathbb{I}_{[\theta,\infty)}(x).$$

Encuentre el estimador máximo verosímil de θ. También encuentre el estimador por momentos para θ.

6. Sea X_1, X_2, X_3 una muestra aleatoria de la población con distribución $U(\theta, 2\theta)$, con $\theta > 0$.

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

(b) Encuentre el estimador máximo verosímil de θ, $\hat{\theta}_{MV}$, y encuentre una constante k tal que $\mathbb{E}(k\hat{\theta}_{MV}) = \theta$.

7. Sea $X_1, \ldots, X_n$ una muestra aleatoria de la población con función de densidad

$$f(x; \mu, \sigma) = \frac{1}{\sigma} \exp\left(-\frac{(x-\mu)}{\sigma}\right) \mathbb{I}_{(\mu,\infty)}(x),$$

donde $\mu \in \mathbb{R}$ y $\sigma \in \mathbb{R}^+$ son desconocidos.

(a) Demuestre que el estimador máximo verosímil de μ es $X_{(1)}$ (la mínima estadística de orden) y el estimador máximo verosímil de σ es $\frac{1}{n}\sum_{i=1}^{n}(X_i - X_{(1)})$.

(b) ¿Cuáles son los estimadores máximo verosímiles de $\frac{\mu}{\sigma}$, $\frac{\mu}{\sigma^2}$ y de $\mu + \sigma$?.

8. Sea $X_1, \ldots, X_n$ una muestra aleatoria de la población con función de densidad

$$f(x; \theta) = \frac{\theta^2}{\theta + 1}(x+1)e^{-\theta x}\mathbb{I}_{(0,\infty)}(x), \quad \theta > 0.$$

(a) Demuestre que la densidad de X pertenece a la familia exponencial.

(b) Encuentre una estadística suficiente minimal y completa.

(c) Encuentre el estimador por momentos.

(d) Encuentre el estimador máximo verosímil.

9. Sea $X_1, \ldots, X_n$ una muestra aleatoria de la población con función de densidad

$$f(x; \theta) = \theta x^{\theta-1}\mathbb{I}_{[0,1]}(x), \quad \theta > 0.$$

(a) Encuentre el estimador por momentos de θ.

(b) Encuentre el estimador máximo verosímil de θ.

(c) Suponga que el verdadero valor de θ es 2. Utilice simulación en R para comparar el error cuadrático medio (numéricamente) de los estimadores en los apartados anteriores en muestras de tamaño $n = 30$. ¿Qué conclusiones puede extraer?.

10. Sea $X_1, \ldots, X_n$ una muestra aleatoria de la población con función de densidad

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

(a) Analice la suficiencia en esta densidad.

(b) ¿Pertenece $f(x; \theta)$ a la familia exponencial?.

(c) Halle el estimador por el método de momentos para θ.

(d) Halle el estimador máximo verosímil para θ.

11. Sea $X_1, \ldots, X_n$ una muestra aleatoria de la población con distribución $Poisson(\lambda)$, con función de densidad $f(x|\lambda)$. Considere que la función de distribución a priori de λ es una distribución $Gamma(\alpha, \beta)$, con función de densidad $\pi(\lambda)$.

    (a) Encuentre la distribución a posteriori de λ.

    (b) Encuentre el estimador Bayesiano de λ usando la función de pérdida del error cuadrático.

12. Sea $X_1, \ldots, X_n$ una muestra aleatoria de la población con distribución $Geométrica(\theta)$. Considere que la función de distribución a priori de θ es una distribución $Beta(\alpha, \beta)$.

    (a) Encuentre la distribución a posteriori de θ.

    (b) Encuentre el estimador Bayesiano de θ usando la función de pérdida del error cuadrático.

13. Sea $X_1, \ldots, X_n$ una muestra aleatoria de la población con distribución $Normal(\mu, \sigma^2)$, donde σ^2 es conocida. Considere que la función de distribución a priori de μ es una distribución $Normal(\eta, \lambda^2)$.

    (a) Encuentre la distribución a posteriori de μ.

    (b) Encuentre el estimador Bayesiano de μ usando la función de pérdida del error cuadrático.

14. Suponga que ciertas pérdidas siguen una distribución $Weibull$ con parámetros θ y τ. Se tiene la siguiente muestra de 16 pérdidas: 54, 70, 75, 81, 84, 88, 97, 105, 109, 114, 122, 125, 128, 139, 146, 153. Estime los parámetros utilizando el método de percentiles, usando los percentiles 20^{th} y 70^{th}.

15. Se practican n mediciones del radio de un círculo. Si las mediciones son independientes entre sí y los errores se distribuyen $N(0, \sigma^2)$ con σ^2 desconocida, proponer un estimador insesgado para el perímetro del círculo y otro para el área.

16. Sea X_1, X_2, X_3, X_4 una muestra aleatoria de tamaño cuatro de una población con distribución $N(0, \sigma^2)$, donde σ es desconocida. Considere los siguientes estimadores $T_1 = X_1^2 - X_2 + X_4$, $T_2 = \frac{1}{3}(X_1^2 + X_2^2 + X_4^2)$, $T_3 = \frac{1}{4}\sum_{i=1}^{4} X_i^2$, $T_4 = \frac{1}{3}\sum_{i=1}^{4}(X_i - \bar{X})^2$ y $T_5 = \frac{1}{2}|X_1 - X_2|$.

    (a) ¿T_1, T_2, T_3, T_4 son insesgados?

(b) De entre T_1, T_2, T_3, T_4, ¿cuál tiene el menor error cuadrático medio?

(c) ¿T_5 es un estimador insesgado para σ?. Si no lo es, encuentre un múltiplo de T_5 que lo sea. Calcule el error cuadrático medio de T_5.

17. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con función de densidad con media μ y varianza σ^2.

(a) Pruebe que $\sum_{i=1}^{n} a_i X_i$ es un estimador insesgado para μ para cualquier valor de las constantes $a_1, a_2, \ldots, a_n$ que satisfagan que $\sum_{i=1}^{n} a_i = 1$.

(b) Si $\sum_{i=1}^{n} a_i = 1$, prueba que $Var\left[\sum_{i=1}^{n} a_i X_i\right]$ se minimiza cuando $a_i = 1/n$, para $i = 1, \ldots, n$.

18. Sea $X_1, \ldots, X_n$ una muestra aleatoria de la población con función de densidad

$$f(x;\theta) = \frac{2x}{\theta^2}\mathbb{I}_{(0,\theta)}(x), \quad \theta > 0.$$

(a) Estime θ por el método de momentos. Llame a este estimador T_1. Encuentre su media y su ECM.

(b) Encuentre el estimador máximo verosímil de θ. Llame a este estimador T_2. Encuentre su media y su ECM.

(c) De entre todos los estimadores de la forma aY_n, donde $Y_n = \text{máx}\{X_1, \ldots, X_n\}$ y a es un valor constante que depende de n, encuentre un estimador para θ con error cuadrático medio uniformemente más pequeño. Llame a este estimador T_3. Encuentre su media y su ECM de T_3.

(d) Encuentre un UMVUE de θ. Nombre a este estimador T_4. Encuentre su media y su ECM.

(e) Define $T_5 = (Y_1 + Y_n)/2$, donde $Y_1 = \text{mín}\{X_1, \ldots, X_n\}$ y $Y_n = \text{máx}\{X_1, \ldots, X_n\}$. Encuentre su media y su ECM. ¿Qué estimador de θ preferiría y por qué?.

19. Sea $X_1, \ldots, X_n$ una muestra aleatoria de la población con distribución $Weibull(\alpha, \beta)$, cuya función de densidad es

$$f_X(x;\alpha) = \frac{1}{\alpha}\beta x^{\beta-1}\exp\left(\frac{-x^{\beta}}{\alpha}\right)\mathbb{I}_{(0,\infty)}(x),$$

donde $\alpha > 0$ es un parámetro desconocido, pero $\beta > 0$ se supone conocido. Encuentre los estimadores máximo verosímiles de α, α^2 y $\frac{1}{\alpha}$ y demuestre que son consistentes en ECM.

20. Sea $X_1, \ldots, X_n$ una muestra aleatoria de la población con distribución tal que

$$\mathbb{E}(X_i) = \theta + b, \quad Var(X_i) = \sigma^2,$$

donde $b \neq 0$, es una constante conocida. Pruebe que $\overline{X}$ no es un estimador consistente en error cuadrático medio para θ. Construya un estimador insesgado para θ que sea consistente.

21. Sea $X_1, \ldots, X_n$ una muestra aleatoria de la población con distribución $Pareto(\alpha, \theta)$, cuya función de densidad es

$$f(x;\theta) = \frac{\alpha\theta^{\alpha}}{x^{\alpha+1}}, \quad x > \theta, \ \theta > 0,$$

con α conocida.

(a) Encuentre el estimador máximo verosímil para θ. ¿Es este un estimador insesgado para θ?. Si la respuesta es negativa, encontrar el estimador insesgado.

(b) Encuentre el estimador por el método de momentos para θ. Nuevamente verifique si el estimador es insesgado; en caso contrario, obtenga el estimador insesgado.

(c) ¿Son consistentes los estimadores obtenidos en los incisos anteriores?.

22. Sea $X_1, \ldots, X_n$ una muestra aleatoria de la población con distribución $Bernoulli(\theta)$, donde $x \in \{0, 1\}$ y $0 \leq \theta \leq \frac{1}{2}$. Note que el espacio paramétrico es $\Theta = \{\theta : 0 \leq \theta \leq \frac{1}{2}\}$.

(a) Encuentre el estimador de θ por medio del método de momentos. Calcule su media y su ECM.

(b) Encuentre el estimador máximo verosímil de θ. Calcule su media y su ECM.

(c) ¿Los estimadores son consistentes en ECM?.

(d) ¿Qué estimador es más eficiente en ECM?.

23. Sea $X_1, \ldots, X_n$ una muestra aleatoria discreta de la población con distribución $Poisson(\lambda)$, donde $x \in \{0, 1, 2, \ldots\}$ y $0 < \lambda \leq 2$. Note que el espacio paramétrico es $\Theta = \{\lambda : 0 < \lambda \leq 2\}$.

(a) Encuentre el estimador de λ por el método de momentos. Calcule su media y su ECM.

(b) Encuentre el estimador máximo verosímil de λ. Calcule su media y su ECM.

(c) ¿Los estimadores son consistentes en ECM?.

24. Considere las siguientes funciones de densidad:

$$f_1(x;p) = p^x(1-p)^{1-x}\mathbb{I}_{\{0,1\}}(x) \;\; \text{donde} \;\; 0 < p < 1,$$

$$f_2(x;\theta) = \theta^x \frac{\log(\theta)}{\theta - 1}\mathbb{I}_{(0,1)}(x) \;\; \text{donde } \theta > 1.$$

En cada caso, para una muestra aleatoria de tamaño n, ¿existirán estadísticas $T_1(\underline{X})$ y $T_2(\underline{X})$ para ciertas funciones $\tau_1(p)$ y $\tau_2(\theta)$, para las cuales la varianza de $T_i(\underline{X})$, $i = 1, 2$, coincidan con la CICR?.

25. Sea $X_1, \ldots, X_n$ una muestra aleatoria de la población con distribución $N(\theta, 1)$.

   (a) Encuentre la CICR para la varianza de los estimadores insesgados de $\tau_1(\theta) = \theta$, $\tau_2(\theta) = \theta^2$ y $\tau_3(\theta) = \mathbb{P}(X > 0)$.

   (b) ¿Existe un estimador insesgado para $\tau_2(\theta) = \theta^2$?. Si es así, encuéntrelo.

   (c) ¿Existe un estimador insesgado para $\tau_3(\theta) = \mathbb{P}(X > 0)$?. Si es así, encuéntrelo.

   (d) Encuentre el UMVUE para $\tau_2(\theta) = \theta^2$.

26. Sea $X_1, \ldots, X_n$ una muestra aleatoria de la población con distribución $Beta(\theta, 1)$, donde $\theta > 0$, es decir, con función de densidad

$$f(x;\theta) = \theta x^{\theta - 1}\mathbb{I}_{(0,1)}(x).$$

   (a) Encuentre el estimador máximo verosímil de $\tau(\theta) = \theta/(1+\theta)$.

   (b) Encuentre una estadística suficiente, y compruebe si es completa.

   (c) ¿Es $S = \sum_{i=1}^n X_i$ una estadística suficiente?. ¿Es $S = \sum_{i=1}^n X_i$ una estadística completa?.

   (d) ¿Existe una función de θ, $\tau(\theta)$, para el cual exista una estimador insesgado cuya varianza coincida con la CICR?. Justifique.

   (e) Encuentre un UMVUE para las siguientes funciones de θ:

      (i) $\tau(\theta) = \theta$

      (ii) $\tau(\theta) = 1/\theta$

      (iii) $\tau(\theta) = \theta/(1+\theta)$

27. Sea $X_1, \ldots, X_n$ una muestra aleatoria de la población con distribución $Bernoulli(p)$, con $p \in (0, 1)$ con $n \geq 3$.

   (a) Sea $U = \sum_{i=1}^n X_i$. Calcule $\mathbb{E}(X_1|U = u)$ y obtenga $\mathbb{E}(X_1|U)$.

(b) Use el teorema de Rao-Blackwell para mejorar el estimador $T_1(\underline{X})$ de $\tau_1(p) = p^2$, dado por
$$T_1(\underline{X}) = X_1X_2.$$

(c) Use el teorema de Rao-Blackwell para mejorar el estimador $T_2(\underline{X})$ de $\tau_2(p) = p^2(1-p)$, dado por
$$T_2(\underline{X}) = X_1X_2(1-X_3).$$

28. Sea $X_1, \ldots, X_n$ una muestra aleatoria de distribución $Poisson(\lambda)$, con $\lambda > 0$ y $n \geq 2$.

(a) Use el teorema de Rao-Blackwell para mejorar el estimador $T_1(\underline{X})$ de $\tau_1(\lambda) = \lambda$, dado por
$$T_1(\underline{X}) = \frac{1}{2}(X_1 + X_2).$$

(b) Use el teorema de Rao-Blackwell para mejorar el estimador $T_2(\underline{X})$ de $\tau_2(\lambda) = e^{-\lambda}$, dado por
$$T_2(\underline{X}) = \mathbb{I}_{\{0\}}(X_1),$$
(llegará a $\left(1-\frac{1}{n}\right)^{\sum_{i=1}^n X_i}$).

(c) Use el teorema de Rao-Blackwell para mejorar el estimador $T_3(\underline{X})$ de $\tau_3(\lambda) = \lambda e^{-\lambda}$, dado por
$$T_3(\underline{X}) = \mathbb{I}_{\{1\}}(X_1).$$

29. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con función de densidad
$$f(x;\theta) = \frac{1}{2\theta}\mathbb{I}_{(-\theta,\theta)}(x), \quad \theta > 0.$$
Encuentre, si existe, el UMVUE para θ.

30. Sea $X_1, \ldots, X_n$ una muestra aleatoria de la población con distribución $U(0,\theta)$. Sean Y_1 y Y_n la mínima y máxima estadísticas de orden, respectivamente.

(a) Encuentre el estimador por momentos para θ. Llame T_1 a dicho estimador y encuentre su media y error cuadrático medio.

(b) Encuentre el estimador máximo verosímil de θ. Llame T_2 a dicho estimador y encuentre su media y error cuadrático medio.

(c) De entre todos los estimadores de la forma aY_n, donde a es una constante que podría depender de n. Encuentre un estimador para θ que tenga el error cuadrático medio uniformemente más pequeño. Llame T_3 a dicho estimador y encuentre su media y error cuadrático medio.

(d) Encuentre el UMVUE de θ. Llame T_4 a dicho estimador y encuentre su media y error cuadrático medio.

(e) Sea $T_5 = Y_1 + Y_n$. Encuentre su media y error cuadrático medio.

(f) Diga ventajas y desventajas de los estimadores $T_1, \ldots, T_5$.

31. Sea $X_1, \ldots, X_n$ una muestra aleatoria de la población con función de densidad

$$f(x;\theta) = \frac{\theta}{(1+x)^{1+\theta}} \mathbb{I}_{(0,\infty)}(x), \quad \theta > 0$$

(a) Encuentre el estimador por momentos para θ suponiendo que $\theta > 1$.

(b) Encuentre el estimador máximo verosímil de $\tau(\theta) = 1/\theta$.

(c) Encuentre una estadística suficiente y completa (si es que existe).

(d) Encuentre la CICR para los estimadores insesgado de $\tau(\theta) = 1/\theta$.

(e) ¿Existe el UMVUE de $\tau(\theta)$?. Si es así, encuéntrelo.

(f) ¿Existe el UMVUE de θ?. Si es así, encuéntrelo.

32. Sea $X_1, \ldots, X_n$ una muestra aleatoria de la población con función de densidad

$$f(x;\mu) = e^{-(x-\mu)} \mathbb{I}_{(\mu,\infty)}(x), \quad \mu \in \mathbb{R}.$$

(a) Demuestre que $T(\underline{X}) = X_{(1)} = \text{mín}\{X_1, \ldots, X_n\}$ es una estadística suficiente y completa.

(b) Encuentre la única función de $X_{(1)}$ que sea el UMVUE de μ.

33. Sea $X_1, \ldots, X_n$ una muestra aleatoria de la distribución $Exp(\lambda)$. Demuestre que

$$T(\underline{X}) = \frac{n-1}{\sum_{i=1}^{n} X_i}$$

es el UMVUE de λ.

34. Sea $X_1, \ldots, X_n$ una muestra aleatoria de la población con distribución $Poisson(\lambda)$. Sea $\tau(\lambda) = (1+\lambda)e^{-\lambda}$.

(a) Obtenga el estimador máximo verosímil para $\tau(\lambda)$.

(b) Obtenga un estimador insesgado para $\tau(\lambda)$.

(b) Obtenga un UMVUE para $\tau(\lambda)$. Sugerencia: Encuentre un estimador insesgado de $\tau(\lambda)$ y utilice el teorema de Rao-Blackwell para mejorarlo.

35. Sea $X_1, \ldots, X_n$ una muestra aleatoria de la población con distribución *Geométrica*(θ) con función de densidad

$$\mathbb{P}(X = x) = \theta(1-\theta)^x, \quad x = 0, 1, 2, \ldots, \ 0 < \theta < 1.$$

(a) Obtenga el estimador por el método de momentos para θ .

(b) Obtenga el estimador máximo verosímil para θ.

(c) Calcule la CICR para la varianza de los estimadores insesgados de θ.

(d) Encuentre un UMVUE para θ.

36. Sea $X_1, \ldots, X_n$ una muestra aleatoria de la población con distribución $N(\theta, \theta^2)$, $\theta \in \mathbb{R}$.

(a) ¿Existe una estadística suficiente unidimensional para θ?

(b) Encuentra una estadística suficiente bidimensional para θ.

(c) ¿Es $\bar{X}$ un UMVUE para θ?.

(d) ¿θ es un parámetro de localización o escala?.

Bibliografía

[1] Casella, G. y Berger, R. L. (2002). *Statistical Inference*. Duxbury Advanced Series. 2nd. ed.

[2] Gómez V., M.A. (2009). "Karl Pearson, el Creador de la Estadística Matemática". *Historia de la Probabilidad y la Estadística IV, J. Basulto y J.J. García (eds.).* Congreso Internacional de Historia de la Estadística y la Probabilidad, 351-356.

[3] Hogg, R.V., McKean, J. W., Craig, A. T. (2014). *Introduction to Mathematical Statistics*. Pearson Education International. 7th. ed.

[4] Kapadia, A.S., Chan, W. y Moyé, L. (2005). *Mathematical Statistics with Applications*. Chapman & Hall, Inc./CRC Press.

[5] Kellison, S.G. y London, R.L. (2011). *Risk Models and Their Estimation*. Actex Publications, Inc.

[6] Lindgren, B.W. (1993). Statistical Theory. Chapman & Hall, Inc. 4th ed.

[7] Mood, A. M., Graybill, F. A. y Boes, D. C. (1974). *Introduction to the theory of stastistics*. Mc Graw-Hill, Inc. 3rd. ed.

[8] Pearson, E. S. (1974). "Memories on the impact of Fisher's work in the 1920's". *Int. Stat. Rev. 42* (1).

[9] Stuart, A., Ord, J. K. y Arnold, S. (1999). *Advanced Theory of Statistics, 2A: Clasical Inference and the Linear Model.* London: Oxford University Press, 6th ed.

[10] Zehna, P.W. (1966). *Invariance of Maximum Likelihood Estimators. Annals of Mathematical Statistics 37* 744.

[11] http://www.dm.uba.ar/materias/estadistica_M/

[12] http://www.statslab.cam.ac.uk/~rrw1/