INFERENCIA ESTADISTICA

FCFM
FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

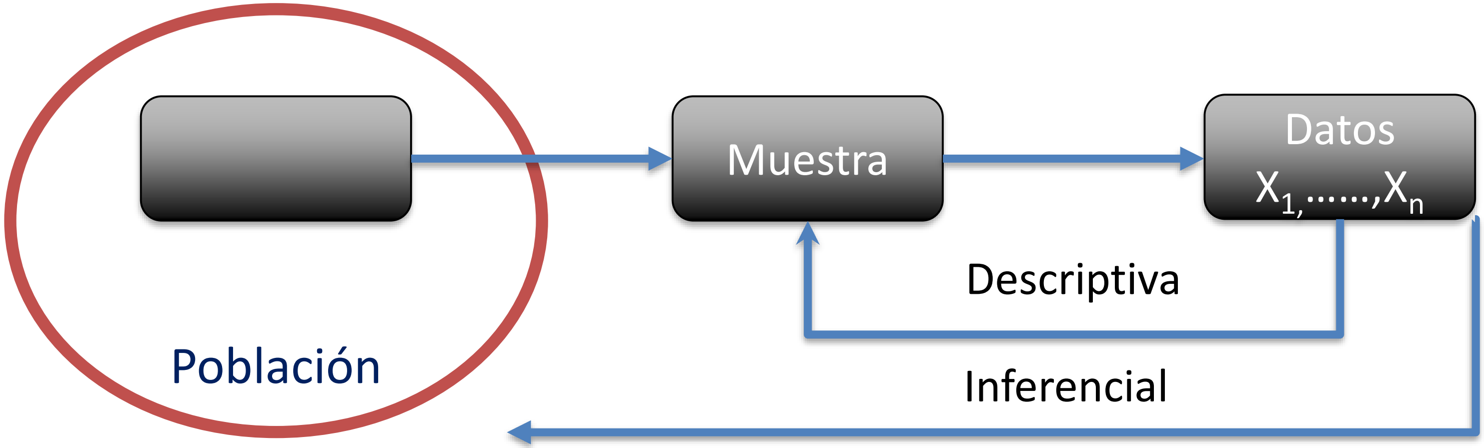
Tema 3

- ✓ Estimación puntual

3.1 Estadísticas y estimadores.

¿Qué es la estadística inferencial?

En la Estadística inferencial se estudian técnicos y procedimientos con el objetivo de extender o generalizar la información de una muestra a la población.



3.1 Estadísticas y estimadores.

¿Qué es la estimación puntual?

Sea x una v.a. con función de densidad o de probabilidad conocida f(x, θ), pero dependiente de un parámetro desconocido θ.

3.1 Estadísticas y estimadores.

Problema: Cómo podemos estimar a partir de, uno serie de observaciones $X_1,\ldots, X_n$ de la v.a. X?

Ejemplo
X = "Tiempo de, traslado de una persona del trabajo a la Casa”
Supongamos X~exp(θ)
Datos (en minutos)
$x_1=100$, $x_2=35$, $x_3=90$, $x_4=35$, $x_5=70$, $x_6=10$

¿Cómo estimamos θ?

3.1 Estadísticas y estimadores.

Definición: Al Conjunto de todos los posibles valores de un parámetro de una distribución de probabilidad se le llama **espacio parametral** y se le denota por letra Θ (θ mayúscula).

Ejemplos:

1) Es espacio parametral de la dist. Exp(θ):

$\Theta = (0, \infty)$

2) Para la dist. Bin(n, p)

$\Theta = \{1, 2, \ldots\} * (0, 1)$

3) Para la dist. N($\mu, \sigma 2$)

$\Theta = \{-\infty, 0\} * (0, \infty)$

3.1 Estadísticas y estimadores.

Muestras aleatorias, estadísticas y estimadores

Definición: Una **muestra aleatoria** (m.a.) es una colección de v.a.s. $X_1,\ldots,X_n$ que son:

a) Independientes
b) Tienen la misma distribución que X.

Definición: Una **estadística** es una función de una muestra aleatoria $X_1,\ldots,X_n$ que no depende de parámetros desconocidos, A tales funciones se les denotara por $T=T(X_1,\ldots,X_n)$.

3.1 Estadísticas y estimadores.

Muestras aleatorias, estadísticas y estimadores

Ejemplos de Estadísticas:

1. $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i =$ Media Muestral
2. $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = Varianza\ Muestral$
3. $T = \frac{1}{n}\sum_{i=1}^{n} x_i^k = k - esimo\ momuento\ muestral$
4. $X_{(k)} = k - esimo\ min\{X_1, \ldots, X_n\} = k - esimo\ estadistica\ de\ orden$
5. $T = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$

$No\ es\ una\ estadistica\ cuando\ X_1, \ldots, X_n \sim N(\mu, \sigma^2)\ \ con\ \mu\ desconocida.$

Ya que por definición, una estadística no depende de parámetros desconocidos.

3.1 Estadísticas y estimadores.

Definición: Un **estimador puntual** para un parámetro desconocido θ es una estadística $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ que se propone para estimar θ.

Por ejemplo, la estadística $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n) = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ puede ponerse como estimador del parámetro desconocido θ en la distribución de Poisson (θ).

Nota: Si $X_1, \dots, X_n$ son valores de la **estimación** para θ es $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$.

3.2 Métodos de construcción de estimadores.

Método de Momentos

Definición: Sea X una v.a. y sea k≥1 un entero. El **k-esimo momento** de X, si existe, es el numero $\boldsymbol{E\left(X^k\right)}$.

A los números $\boldsymbol{E\left(X^1\right), E\left(X^2\right)}$, ... se les llama también **momentos poblacionales.**

Observación: Sea $X_1, \ldots, X_n$ una m.a. de la distribución f(x; θ).

Definición: Sea $X_1, \ldots, X_n$ una m.a. y sea k≥1 un entero. El **k-esimo momento** es la v.a. $\frac{1}{n}\sum_{i=1}^{n} X_i^k$

3.2 Métodos de construcción de estimadores.

Método de Momentos

Consiste en igualar los momentos poblacionales con los correspondientes momentos muestrales y resolver esta ecuación (o sistema de ecuaciones) para el parámetro o vector de parámetros.

$$1er.m.poblacional\ E(X) = \frac{1}{n}\sum_{i=1}^{n} X_i \ \ 1er.m.muestral$$

$$2do.m.poblacional\ E(X^2) = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \ \ 2do.m.muestral$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

3.2 Métodos de construcción de estimadores.

Ejemplo 1: Sea X una v.a. con función de densidad $f(x;\theta) = \begin{cases} \theta x^{\theta-1} & si\ 0 < x < 1 \\ 0 & otro\ caso \end{cases}$ en donde θ > 0, estimar el parámetro por el método de momentos.

Entonces $E(X) = \frac{\theta}{1+\theta}$, se calcula el primer momento, respecto a la función de densidad.

Si $X_1, \dots, X_n$ es una m.a. de esta distribución, $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \;\; 1er.m.muestral$

Entonces por el método de momentos,

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

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

$$\therefore Este\ es\ el\ estimador\ para\ \theta.$$

3.2 Métodos de construcción de estimadores.

Ejemplo 2: Sea X una v.a. con distribución de Poisson(θ) en donde θ > 0, estimar el parámetro por el método de momentos.
Entonces $E(x) = \theta$

Si $X_1, \ldots, X_n$ es una m.a. de esta distribución, $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \;\; 1er.m.muestral$
Entonces por el método de momentos,

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

$$\therefore Este\ es\ el\ estimador\ para\ \theta.$$

3.2 Métodos de construcción de estimadores.

Ejemplo 3: Sea X una v.a. con distribución N(μ, σ^2), estimar los parámetros por el método de momentos.

Entonces $E(X) = \mu$, $E(X^2) = \sigma^2 + \mu^2$

Igualamos con los momentos muestrales, tenemos el sistema de ecuaciones:

$\hat{\mu}=\bar{X}$ … … (1)

$$\hat{\sigma}^2 + \hat{\mu}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \text{ … … (2)}$$

Desarrollamos: $\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n}(X_i^2 - 2X_i\bar{X} + \bar{X}^2) = \sum_{i=1}^{n} X_i^2 - 2\bar{X}\sum_{i=1}^{n} X_i + n\bar{X}^2 = \sum_{i=1}^{n} X_i^2 - 2n\bar{X}^2 + n\bar{X}^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2$

$Sustituimos\ (1) en\ (2)$

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

$$\hat{\sigma}^2 = \frac{n-1}{n}s^2$$

$\therefore Estos\ son\ los\ estimadores.$

3.2 Métodos de construcción de estimadores.

Ejercicio: Suponga que tiene la siguiente muestra de tamaño 10:

$$\{1,1,1,2,2,3,5,7,8,10\}$$

Estimar los parámetros $\mu\ y\ \sigma^2$ usando el Método de Momentos si la distribución normal se ajusta a través de los datos de la muestra.

3.2 Métodos de construcción de estimadores.

Observaciones de Método de Momentos:

<u>Ley débil de los grandes números:</u>

Sean $(X_1, X_2, \ldots, X_n)$ n variables aleatorias independientes todas las cuales tienen la misma esperanza $\mu = E(X)$ y varianza $\sigma^2 = V(X)$. Sea $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Entonces

$$\lim_{n\to\infty} P\left(\left|\overline{X} - \mu\right| \geq \varepsilon\right) = 0$$

Decimos que $\overline{X}$ converge a μ en probabilidad y lo indicamos: $\overline{X} \xrightarrow{p} \mu$.

3.2 Métodos de construcción de estimadores.

Observaciones de Método de Momentos:

1. El método es razonable ya que por la ley de loa grandes números cuando $n \to \infty$, los momentos muestras se aproximan a los momentos poblacionales.
2. El método puede aplicarse para distribuciones discretas como continuas.
3. El método presupone la existencia de los primeros momentos poblacionales, y que estos dependen de los parámetros.
4. El método presupone que el sistema de ecuaciones tiene una única solución u esta es sencilla de encontrar.
5. El método no garantiza que el estimador tome valores en el espacio parametral.

3.2 Métodos de construcción de estimadores.

Método de Máxima Verosimilitud

Supongamos que X es una v.a. discreta con función de distribución de probabilidad $p(x,\theta)$, donde θ es un parámetro desconocido. Sean $x_1, x_2, \ldots, x_n$ los valores observados de una muestra aleatoria de tamaño n.

Se define la ***función de verosimilitud*** como la función de distribución conjunta de las observaciones:

$$L(x_1, x_2, \ldots, x_n, \theta) = P(X_1 = x_1)P(X_2 = x_2)\ldots P(X_n = x_n) = p(x_1,\theta).p(x_2,\theta).\ldots.p(x_n,\theta)$$

Notar que la función de verosimilitud es una función de θ.

El estimador de máxima verosimilitud de θ es aquel valor de θ que maximiza la función de verosimilitud

3.2 Métodos de construcción de estimadores.

Método de máxima verosimilitud

La interpretación del método sería: el estimador de máxima verosimilitud es aquel valor del parámetro que maximiza la probabilidad de ocurrencia de los valores muestrales

La adaptación para el caso en que X es una v.a. continua sería la siguiente:

Supongamos que X es una v.a. continua con función de densidad de probabilidad $f(x,\theta)$, donde θ es un parámetro desconocido. Sean $x_1, x_2, \ldots, x_n$ los valores observados de una muestra aleatoria de tamaño n.

Se define la ***función de verosimilitud*** como la función de distribución conjunta de las observaciones:

$$L(x_1, x_2, \ldots, x_n, \theta) = f(x_1,\theta).f(x_2,\theta).....f(x_n,\theta)$$

La función de verosimilitud es una función de θ.

El estimador de máxima verosimilitud de θ es aquel valor de θ que maximiza la función de verosimilitud

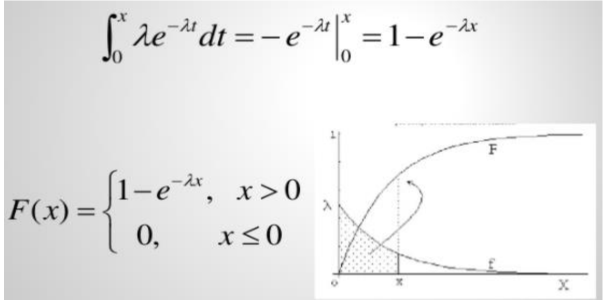
3.2 Métodos de construcción de estimadores.

Método de máxima verosimilitud

Consiste en obtener el valor de θ que maximiza a la función de verosimilitud $L(x_1, \dots, x_n\,; \theta) = f(x_1; \theta) \dots f(x_n; \theta)$. Al valor de θ en donde $L(x_1, \dots, x_n\,; \theta)$ alcanza su máximo si este existe, se le llama "estimación de máxima verosimilitud" o "estimación máximo verosimilitud".

Esta es la Idea: Se debe de encontrar el valor θ, tal que el valor numérico observado $(x_1, \dots, x_n)$ de la m.a. tenga probabilidad máxima.

3.2 Métodos de construcción de estimadores.



Método de máxima verosimilitud

Ejemplo 1: Sea $X_1, \dots, X_n$ una m.a. de la dist. Exp(θ).

La función de verosimilitud es

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

Observación: $L(\theta)$ es máxima en el mismo punto en donde $\ln L(\theta)$ lo es, ya que esto se debe a las propiedades de continuidad y monotonía de la función ln.

Entonces, aplicando ln:

$$\ln L(x_1, \dots, x_n\,; \theta) = n\ln\theta - \theta n \bar{x}$$

$$\therefore \frac{d}{d\theta} \ln L(x_1, \dots, x_n\,; \theta) = \frac{n}{\theta} - n\bar{x}$$

Esta derivada es cero $\leftrightarrow \theta = \frac{1}{\bar{x}}$ Además $\frac{d^2}{d\theta^2} \ln L(x_1, \dots, x_n\,; \theta) = -\frac{n}{\theta^2} < 0$

$\therefore \hat{\theta} = \frac{1}{\bar{x}}$ Es **la estimación** para θ

Estimación es un número

$\therefore \hat{\theta} = \frac{1}{\bar{X}}$ Es la **estimador** para θ

Estimador es una estadística

3.2 Métodos de construcción de estimadores.

Método de máxima verosimilitud

Ejemplo 2: Sea $X_1, \dots, X_n$ una m.a. de la distribución $f(x;\theta) = \begin{cases} \theta x^{\theta-1} & si\ 0 < x < 1 \\ 0 & otro\ caso \end{cases}$

La función de verosimilitud es

$$L(x_1, \dots, x_n\,; \theta) = f(x_1;\theta) \dots f(x_n;\theta)$$
$$= \theta^n (x_1 \dots x_n)^{\theta-1}$$
$$\ln L(x_1, \dots, x_n\,; \theta) = n ln\theta + (\theta - 1) \sum_{i=1}^{n} ln x_i$$
$$\therefore \frac{d}{d\theta} lnL(x_1, \dots, x_n\,; \theta) = \frac{n}{\theta} + \sum_{i=1}^{n} ln x_i$$

Esta derivada es cero $\leftrightarrow$ $\theta = {}^{-n}/_{\sum_{i=1}^{n} ln x_i}$

Además $\frac{d^2}{d\theta^2} lnL(x_1, \dots, x_n\,; \theta) = -\frac{n}{\theta^2} < 0$

$\therefore \hat{\theta} = {}^{-n}/_{\sum_{i=1}^{n} ln x_i}$ Es **la estimación** para θ

$\therefore \hat{\theta} = {}^{-n}/_{\sum_{i=1}^{n} ln X_i}$ Es la **estimador** para θ

3.2 Métodos de construcción de estimadores.

Método de máxima verosimilitud

Ejemplo 3: Sea $X_1, \dots, X_n$ una m.a. de la distribución $\mathrm{N}(\mu, \sigma^2)$.

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

La función de verosimilitud es

$$L(x_1, \dots, x_n; \mu, \sigma^2) = f(x_1; \mu, \sigma^2) \dots f(x_n; \mu, \sigma^2)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_1-\mu)^2/2\sigma^2} \dots \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_n-\mu)^2/2\sigma^2}$$

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

$$\ln L(x_1, \dots, x_n; \mu, \sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

Derivamos cada variable, lo que tenemos:

$$\frac{\partial}{\partial \mu} L(x_1, \dots, x_n; \mu, \sigma^2) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 \quad y$$

$$\frac{\partial}{\partial \sigma^2} L(x_1, \dots, x_n; \mu, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

3.2 Métodos de construcción de estimadores.

Método de máxima verosimilitud

Sigue el Ejemplo 3, Igualamos a cero ambas derivadas:

$$\sum_{i=1}^{n}(x_{\mathrm{i}}-\mu)=0\ldots\ldots\ldots(1)$$

$$-n\sigma^{2}+\sum_{i=1}^{n}(x_{\mathrm{i}}-\mu)^{2}=0\ldots\ldots\ldots(2)$$

De (1), $\hat{\mu}=\bar{x}$ y substituyendo en (2), $\hat{\sigma}^{2}=\frac{1}{n}\sum_{i=1}^{n}(x_{\mathrm{i}}-\bar{x})^{2}$, punto critico.

Para verificar que $(\hat{\mu},\hat{\sigma}^{2})$ es un máximo que se calcula por la matriz hessiana,

$$\mathrm{H}(\mu,\sigma^{2})=\begin{pmatrix}\frac{\partial^{2}}{\partial\mu\partial\mu}L(x_{1},\ldots,x_{n};\mu,\sigma^{2}) & \frac{\partial^{2}}{\partial\mu\partial\sigma^{2}}L(x_{1},\ldots,x_{n};\mu,\sigma^{2})\\ \frac{\partial^{2}}{\partial\sigma^{2}\partial\mu}L(x_{1},\ldots,x_{n};\mu,\sigma^{2}) & \frac{\partial^{2}}{\partial\sigma^{2}\partial\sigma^{2}}L(x_{1},\ldots,x_{n};\mu,\sigma^{2})\end{pmatrix}$$

3.2 Métodos de construcción de estimadores.

Método de máxima verosimilitud
Sigue el Ejemplo 3:

$$\mathrm{H}(\mu,\sigma^2)=\begin{pmatrix} -n\sigma^2 & -(\sigma^2)^{-2}\sum_{i=1}^{n}(x_\mathrm{i}-\mu) \\ -(\sigma^2)^{-2}\sum_{i=1}^{n}(x_\mathrm{i}-\mu) & \frac{n}{2}(\sigma^2)^{-2}-(\sigma^2)^{-3}\sum_{i=1}^{n}(x_\mathrm{i}-\mu)^2 \end{pmatrix}$$

Al evaluar en el punto critico y obtenemos la siguiente Matriz de la forma reducida:

$$\mathrm{H}(\hat{\mu},\hat{\sigma}^2)=\begin{pmatrix} -n(\hat{\sigma}^2)^{-1} & 0 \\ 0 & -\frac{n}{2}(\hat{\sigma}^2)^2 \end{pmatrix}=\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

El que esta matriz se definida negativa significa que

$1.\ Primer\ entreda\ a_{11}<0\ y\ 2.\ El\ determinante\ |\mathrm{H}(\hat{\mu},\hat{\sigma}^2)|>0$

$\therefore\ \hat{\mu}=\bar{x}\quad y\ \ \hat{\sigma}^2=\sum_{i=1}^{n}(x_\mathrm{i}-\bar{x})^2$ son las estimaciones por máxima verosimilitud.

$\therefore\ \hat{\mu}=\bar{X}\quad y\ \ \hat{\sigma}^2=\sum_{i=1}^{n}(X_\mathrm{i}-\bar{X})^2$ son los estimadores por máxima verosimilitud.

3.2 Métodos de construcción de estimadores.

Observaciones Método de máxima verosimilitud

1. El método de máxima verosimilitud puedes aplicarse tanto para distribuciones discretas como continuas.
2. El método de máxima verosimilitud no genera necesariamente los mismo estimadores que el método de momentos.
3. Cuando existe solución, el método en el máxima verosimilitud siempre produce un valor del parámetro dentro del espacio parametral.
4. El estimador máximo verosimil, puede no existir, un ejemplo puede es para la distribución unif$(0, \theta)$.
5. Si el espacio parametral se reduce, el punto máximo puede cambiar.

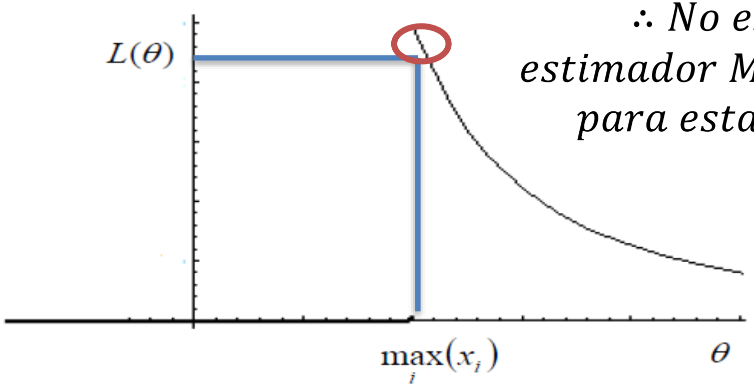
3.2 Métodos de construcción de estimadores.

Método de máxima verosimilitud

Ejemplo 4: Sea $X_1, \ldots, X_n$ una m.a. de la distribución Unif(0, θ).

La función de verosimilitud es

$$L(x_1, \ldots, x_n; \theta) = f(x_1; \theta) \ldots f(x_n; \theta)$$
$$= \frac{1}{\theta} \ldots \frac{1}{\theta} * 1_{(0,\theta)}(x_1) \ldots 1_{(0,\theta)}(x_n)$$
$$= \left(\frac{1}{\theta}\right)^n * 1_{(0,\theta)}(x_1) \ldots 1_{(0,\theta)}(x_n)$$



$\therefore$ *No es existe el estimador Maximo Verosimil para esta distribución*

El valores de θ, debe de ser estrictamente mayor a todos los valores observados "$x_1, \ldots, x_n$", en otras θ debe ser mayor al máximo de los valores observados "x_n".
No existe un valor Máximo para esta función.

3.2 Métodos de construcción de estimadores.

Funciones Parametrales y Máxima Verosimilitud

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

Ejemplos:

1) $\tau(\theta) = \theta^2$
2) $\tau(n, p) = np$ "$funcion\ parametral\ de\ 2\ parametros$"
3) $\tau(\theta) = E(x)$
4) Cuantiles "Los cuantiles suelen usarse por grupos que dividen la distribución en partes iguales"
5) $\tau(\theta) = P(X\epsilon A)$ "$La\ prob. de\ que\ la\ v. a., tomr\ algún\ valor\ en\ algún\ conjunto$"

3.2 Métodos de construcción de estimadores.

Funciones Parametrales y Máxima Verosimilitud

Definición: La función de verosimilitud asociada a una función parametral $\tau(\theta)$ es la siguiente: para cada posible valor η de $\tau(\theta)$,

$$L^*(\eta) = sup\{L(\theta): \theta \epsilon \tau^{-1}(\eta)\}$$

En otras palabras, tomamos un valor η y se calcula su imagen inversa bajo τ, eso es un conjunto y se evalúa L en cada punto de ese conjunto, donde el supremo de estos valores es la función de verosimilitud de $\tau(\theta)$.

Ejemplo: Para la distribución $\exp(\theta)$

$$L(\theta) = \theta^n e^{-\theta n\bar{x}}$$

$Sea\ \tau(\theta) = \theta^2, es\ 1 - 1\ en\ el\ intervalo\ (0, \infty).$

$Entonces\ L^*(\eta) = L(\sqrt{\eta}) = (\sqrt{\eta})^n e^{-\sqrt{\eta} n\bar{x}}$

3.2 Métodos de construcción de estimadores.

Principio de Invarianza

Teorema: Sea $\hat{\theta}$ el estimador máximo verosímil para un parámetro θ. El estimador máximo verosímil para una función parametral $\tau(\theta)$ es $\tau\left(\hat{\theta}\right)$.

Demostración:

Sabiendo que $L^{*}(\eta) = sup\{L(\theta): \theta \epsilon \tau^{-1}(\eta)\}$.

Sea $\theta \rightarrow \tau(\theta)$ uno a uno y sea $\eta = \tau(\theta)$.

Entonces $L^{*}(\eta) = L\left(\tau^{-1}(\eta)\right) = L(\theta)$

$\therefore L^{*}(\eta)\ es\ máximo\ en\ el\ valor\ de\ \hat{\eta}\ definido\ en\ \tau(\hat{\theta})$

$Es\ decir\ \tau\left(\hat{\theta}\right)\ es\ el\ estimador\ de\ maximo\ verosimil\ para\ \tau(\theta).$

3.2 Métodos de construcción de estimadores.

Principio de Invarianza
Sigue Demostración de Teorema:

Sea $\theta \rightarrow \tau(\theta)$ no necesariamente uno a uno.

Entonces $\underset{\eta}{máx} L^*(\eta) = máx\{L(\theta) : \theta\epsilon\Theta\} = L(\hat{\theta})$

Sea $\hat{\eta}$ $definido\ en\ \tau(\hat{\theta})$. Entonces

$L^*(\eta) = L^*\left(\tau(\hat{\theta})\right) = sup\{L(\theta) : \theta\epsilon\tau^{-1}(\tau(\hat{\theta}))\} = L(\hat{\theta})$

$\therefore L^*(\hat{\eta})\ es\ el\ valor\ máximo\ de\ la\ función\ L^*$

$Es\ decir\ \hat{\eta} = \tau(\hat{\theta})\ es\ el\ estimador\ de\ maximo\ verosimil\ para\ \tau(\theta).$

3.2 Métodos de construcción de estimadores.

Ejemplo 1: Para la distribución $\exp(\theta)$ y definida en $\tau(\theta)=\theta^2$

Tenemos que $\widehat{\tau(\theta)}=\tau(\hat{\theta})\ es\ decir\ \hat{\theta}^2=(\hat{\theta})^2$.

Ejemplo 2: Sea $X_1,\ldots,X_n$ una m.a. de una población con distribución $N(\theta,1)$, con θ desconocido. Se busca el estimador máximo verosímil de $\tau(\theta)=\log(\theta)$.

Como $\hat{\theta}=\bar{X}$ es el estimador máximo verosímil θ, entonces por la propiedad de varianza $\log(\bar{X})$ es el estimador máximo verosímil de $\log(\theta)$.

Ejemplo 3: Sea $X_1,\ldots,X_n$ una m.a. de una población con distribución $N(\theta,\sigma^2)$.

Se sabe que el estimador máximo verosímil $\hat{\theta}\ es\ \bar{X}$. Para encontrar el estimador máximo verosímil $\tau(\theta)=sen(\theta)$.

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

3.3 Criterios de evaluación de estimadores.

Insesgamiento

Definición: Un estimador $\hat{\theta}$ es insesgado para un parámetro θ si $\mathrm{E}(\hat{\theta}) = \theta$.

Ejemplo 1: El estimador $\hat{\theta} = \bar{x}$ (media muestral) es insesgado para el parámetro θ de la distribución Poisson(θ) pues

$$\mathrm{E}(\hat{\theta}) = E(\bar{x})$$

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

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

$Al\ ser = de\ \theta\ es\ un\ estimador\ insesgado.$

3.3 Criterios de evaluación de estimadores.

Insesgamiento

Ejemplo 2: El estimador $\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$es insesgado para la varianza de la distribución $\mathrm{N}(\mu, \sigma^2)$, en donde

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

$$= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu + \mu - \bar{x})^2 = \frac{1}{n-1}\sum_{i=1}^{n}\big((x_i - \mu) - (\bar{x} - \mu)\big)^2$$

Se desarrolla al cuadrado

$$= \frac{1}{n-1}\sum_{i=1}^{n}[(x_i - \mu)^2 - (2 * (x_i - \mu) * (\bar{x} - \mu)) + (\bar{x} - \mu)^2]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}(x_i - \mu)^2 + n(\bar{x} - \mu)^2 - 2(\bar{x} - \mu)\sum_{i=1}^{n}(x_i - \mu)\right]$$

3.3 Criterios de evaluación de estimadores.

Insesgamiento

Sigue el Ejemplo 2:

Separamos en $\sum_{i=1}^{n}$ y tomamos en cuenta si:

$\bar{x} = \frac{\sum_i^n x_i}{n} \Rightarrow n\bar{x} = \sum_i^n x_i$ y Recíprocamente con $\mu = \frac{\sum_i^n x_i}{n} \Rightarrow n\mu = \sum_i^n x_i$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}(x_i - \mu)^2 + n(\bar{x}-\mu)^2 - 2(\bar{x}-\mu)\left(\sum_{i=1}^{n} x_i - n\mu\right)\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}(x_i - \mu)^2 + n(\bar{x}-\mu)^2 - 2(\bar{x}-\mu)(n\bar{x} - n\mu)\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}(x_i - \mu)^2 + n(\bar{x}-\mu)^2 - 2n(\bar{x}-\mu)^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}(x_i - \mu)^2 - n(\bar{x}-\mu)^2\right]$$

3.3 Criterios de evaluación de estimadores.

Insesgamiento

Sigue el Ejemplo 2:

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}(x_i - \mu)^2 - \sum_{i=1}^{n}(\bar{x} - \mu)^2\right] = \frac{1}{n-1}\sum_{i=1}^{n}[(x_i - \mu)^2 - (\bar{x} - \mu)^2]$$

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

$$= \frac{1}{n-1}\sum_{i=1}^{n}[E(x_i - \mu)^2 - E(\bar{x} - \mu)^2]$$

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

$$Al\ ser = de\ \theta\ es\ un\ estimador\ insesgado.$$

3.3 Criterios de evaluación de estimadores.

Insesgamiento

Ejemplo 3: Suponga que los valores $\mathbf{x_1}$, $\mathbf{x_2}$, $\mathbf{x_3}$, representa una muestra simple de tamaño 3 seleccionada en forma aleatoria, donde sus valores siempre son positivos y provienen de una población con media μ y desviación típica σ. Si se tiene como posibles estimadores de μ a los siguientes estadísticos:

$$\hat{\mu}_1 = \frac{1}{5}(x_1 + 2x_2 + x_3)$$

$$\hat{\mu}_2 = \frac{1}{6}(2x_1 + x_2 + 3x_3)$$

¿Cuál de los dos estimadores utilizaría para realizar la estimación?

Solución: Al tener 2 estimadores, se selecciona al que tenga menos sesgo con respecto al parámetro a estimar., en otras palabras, aquel cuyo valor esperado sea igual al valor del parámetro en cuestión.

3.3 Criterios de evaluación de estimadores.

Insesgamiento

Sigue Ejemplo 3: Se procede a calcular los valores esperados de los estimadores

$$\hat{\mu}_1 = \frac{1}{5}(x_1 + 2x_2 + x_3)$$

$$E(\hat{\mu}_1) = E\left(\frac{1}{5}(x_1 + 2x_2 + x_3)\right) = \frac{1}{5}(E(x_1) + 2E(x_2) + E(x_3))$$

$$Sabiebdo\ queE(x) = \mu$$

$$E(\hat{\mu}_1) = \frac{1}{5}(\mu + 2\mu + \mu)$$

$$\therefore E(\hat{\mu}_1) = \frac{4}{5}\mu = \frac{4}{5}\theta$$

$$Al\ ser\ \ \neq de\ \theta\ es\ un\ estimador\ sesgado.$$

3.3 Criterios de evaluación de estimadores.

Insesgamiento

Sigue Ejemplo 3: Se procede a calcular los valores esperados de los estimadores

$$\hat{\mu}_2 = \frac{1}{6}(2x_1 + x_2 + 3x_3)$$

$$E(\hat{\mu}_2) = E\left(\frac{1}{6}(2x_1 + x_2 + 3x_3)\right) = \frac{1}{6}(2E(x_1) + E(x_2) + 3E(x_3))$$

$$Sabiebdo\ queE(x) = \mu$$

$$E(\hat{\mu}_2) = \frac{1}{6}(2\mu + \mu + 3\mu)$$

$$\therefore E(\hat{\mu}_2) = \frac{6}{6}\mu = \mu = \theta$$

$$Al\ ser = de\ \theta\ es\ un\ estimador\ insesgado.$$

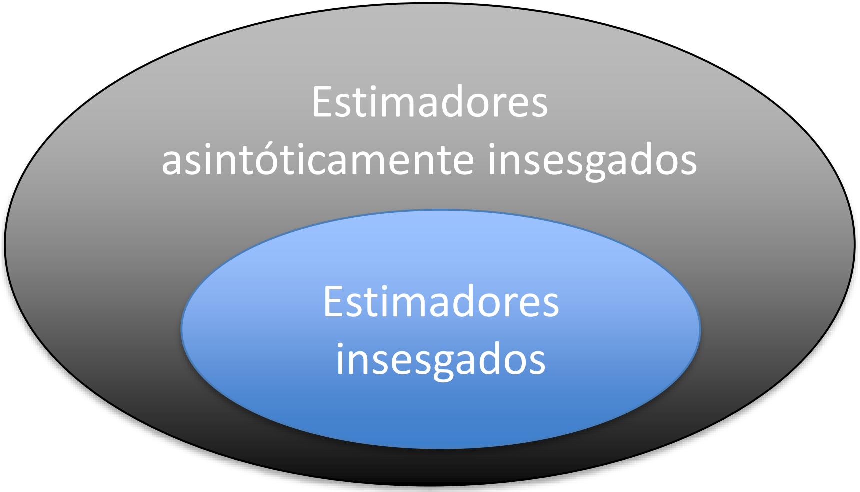
3.3 Criterios de evaluación de estimadores.

Insesgamiento asintótico

Definición: Una estadística $\widehat{\theta_n}$, basada en una muestra aleatoria de tamaño n, es un estimador asintóticamente insesgado para un parámetro θ si

$$\lim_{n\to\infty} E\left(\widehat{\theta_n}\right) = \theta$$

Podemos ver que, todo estimador insesgado es asintóticamente insesgado, pero hay casos donde el estimador es asintóticamente insesgado pero que no es insesgado.



3.3 Criterios de evaluación de estimadores.

Insesgamiento asintótico

Ejemplo 4: Sea $X_1, \dots X_n$ una muestra aleatoria de la dist. $\mathrm{N}(\mu, \theta)$, donde el estimador para θ

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

$$\Rightarrow \mathrm{E}\left(\widehat{\theta_n}\right) = \mathrm{E}\left(\frac{n-1}{n}s^2\right) = \frac{n-1}{n}E(s^2)$$

$$= \frac{n-1}{n}\theta \rightarrow \theta \; cuando \; n \rightarrow \infty$$

Ya demostramos que $E(s^2)$, que es un estimador insesgado para la varianza.

Podemos concluir $\widehat{\theta_n}$ no es insesgado, pero es asintóticamente insesgado.

3.3 Criterios de evaluación de estimadores.

Consistencia

Sea $\widehat{\theta_n}$ un estimador para un parámetro θ, basado en una m.a. $X_1, \ldots X_n$

Definición: $\widehat{\theta_n}$ es consistente para θ si $\widehat{\theta_n} \underset{n \to \infty}{\overset{p}{\rightarrow}} \theta$, esto es para cualquier $\in > 0$,

$$\lim_{n\to\infty} P\left(\left|\widehat{\theta_n} - \theta\right| > \epsilon\right) = 0$$

Ejemplo: Sea X una variable aleatoria con función de probabilidad o densidad f$(x;$θ$)$ tal que $E(x) =$ θ. Por ejemplo, las distribuciones $Poisson(\theta)$ ó $N(\theta, \sigma^2)$, satisfacen esta condición. El estimador $\widehat{\theta_n}$ $definido\ como\ \bar{x}$, es consistente para θ pues por la "ley débil de los grandes números" $\bar{x} \underset{n \to \infty}{\overset{p}{\rightarrow}} E(x)$

3.3 Criterios de evaluación de estimadores.

Consistencia

Proposición: $\widehat{\theta_n}$ es consistente θ si:

a) $\lim_{n\to\infty} \mathrm{E}(\widehat{\theta_n}) = \theta$

b) $\lim_{n\to\infty} \mathrm{Var}(\widehat{\theta_n}) = 0$

Tenemos:

$$E(\widehat{\theta_n} - E(\widehat{\theta_n}))^2 = Var\widehat{\theta_n}$$

$$\left(E(\widehat{\theta_n}) - \theta\right) \textit{ es una contante}$$

$$(\widehat{\theta_n} - E(\widehat{\theta_n})) = 0$$

Demostración: Chebyshev, para c ∈ ℝ, ∈> 0, $P(|x - c| > \epsilon) \leq \frac{1}{\epsilon^2} E(x - c)^2$

Entonces

$$P\left(\left|\widehat{\theta_n} - \theta\right| > \epsilon\right) \leq \frac{1}{\epsilon^2} E\left(\widehat{\theta_n} - \theta\right)^2$$

$$= \frac{1}{\epsilon^2} E\left[(\widehat{\theta_n} - E(\widehat{\theta_n})) + \left(E(\widehat{\theta_n}) - \theta\right)\right]^2 = \frac{1}{\epsilon^2}\left[Var(\widehat{\theta_n}) + 0 + \left(E(\widehat{\theta_n}) - \theta\right)^2\right]$$

$$\textit{Podemos ver por} \text{ con } a)\ y\ b)\ \to 0 \textit{ cuando } n \to \infty$$

∴ Podemos concluir que todo estimador, a) asintóticamente insesgado y b) con varianza convergente a cero es consistente.

3.3 Criterios de evaluación de estimadores.

Consistencia

Ejemplo 1: Sea $X_1, \ldots X_n$ una muestra aleatoria de la distribución Bernulli(θ). Sea z una distribución Bernulli(1/n) y definida en: $\hat{\theta}_n = \begin{cases} \bar{x} \; si \; z = 0 \\ n \; si \; z = 1 \end{cases}$

$$E(\hat{\theta}_n) = E(\hat{\theta}_n/z = 0)\left(1 - \frac{1}{n}\right) + E(\hat{\theta}_n/z = 1)\left(\frac{1}{n}\right)$$

$$= E(\bar{x})\left(\frac{n-1}{n}\right) + n\frac{1}{n} = \theta\left(\frac{n-1}{n}\right) + 1$$

Como es diferente de θ entonces $\hat{\theta}_n$ no es insesgado y cuando n tiende a infinito es $\theta + 1$ por lo tanto tampoco es asintóticamente insesgado.

3.3 Criterios de evaluación de estimadores.

Consistencia

Continua Ejemplo 1: Pero es consistente pues para cualquier $\epsilon > 0$,

$$\mathrm{P}\left(\left|\hat{\theta}_n - \theta\right| > \epsilon\right)$$

$$= \mathrm{P}\left(\left|\hat{\theta}_n - \theta\right| > \epsilon | z = 0\right)\left(\frac{n-1}{n}\right) + \mathrm{P}\left(\left|\hat{\theta}_n - \theta\right| > \epsilon | z = 1\right)\left(\frac{1}{n}\right)$$

$$= \mathrm{P}(|\bar{x} - \theta| > \epsilon)\left(\frac{n-1}{n}\right) + \mathrm{P}(|n - \theta| > \epsilon)\left(\frac{1}{n}\right)$$

Nota: Cuando n tiende ∞ podemos ver que $\mathrm{P}(|\bar{x} - \theta| > \epsilon)$ es cero, ya que el estimador $\bar{x}$ es consistente y por otro lado $\left(\frac{1}{n}\right)$ también tiende a cero.

$\therefore$ *Tenemos que esta probabilidad tiende cero cuando n tiende a infinito.*

$\therefore$ *El estimador no es insegado ni asintotimante insesgado, pero si es consistente*

3.3 Criterios de evaluación de estimadores.

Consistencia

Ejemplo 2: Sea $\widehat{\theta_n} = \bar{X}$ un estimador consistente para el parámetro μ.

$$E\left(\widehat{\theta_n}\right) = E(\bar{X}) = \mu = \theta$$

$$Al\ ser = de\ \theta\ es\ un\ estimador\ insesgado$$

Ahora calcular la varianza del estimador

$$V\left(\widehat{\theta_n}\right) = V(\bar{X}) = \frac{\sigma^2}{n};$$

$$\lim_{\mathrm{n}\to\infty} \mathrm{Var}\left(\widehat{\mathrm{\theta_n}}\right) = \lim_{\mathrm{n}\to\infty} \left(\frac{\sigma^2}{n}\right) = 0$$

Como el estimador fue insesgado y la varianza convergente a cero, confirmaos que es $\bar{X}$ es un estimador consistente.

3.3 Criterios de evaluación de estimadores.

Consistencia

Ejemplo 3: Sea $\widehat{\theta_n} = 3\bar{X}$ un estimador del parámetro θ en una población con función de densidad dada por:

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

Determinar si $\widehat{\theta_n}$ es un estimador consistente de θ.

$$E\left(\widehat{\theta_n}\right) = E(3\bar{X}) = 3E(\bar{X}) = 3\mu = 3\left({}^{\theta}/_{3}\right) = \theta$$

$$ya\ que\ \mu = \int_0^{\theta} x \frac{2}{\theta^2}(\theta - x)\,dx = \frac{\theta}{3}$$

$$Al\ ser = de\ \theta\ es\ un\ estimador\ insesgado$$

3.3 Criterios de evaluación de estimadores.

Consistencia

Sigue Ejemplo 3: Calcular la varianza del estimador

$$\mathrm{V}(\widehat{\theta_n}) = V(3\bar{X}) = 9V(\bar{X}) = 9\left(\frac{\sigma^2}{n}\right)$$

$$donde\ \sigma^2 = {\theta^2}/{18}$$

$$\lim_{n\to\infty} \mathrm{Var}(\widehat{\theta_n}) = \lim_{n\to\infty} 9\left(\frac{\sigma^2}{n}\right) = \lim_{n\to\infty} 9\left(\frac{\theta^2/18}{n}\right) = \lim_{n\to\infty}\left(\frac{\theta^2}{2n}\right) = 0$$

Como el estimador fue insesgado y la varianza convergente a cero, confirmaos que es $3\bar{X}$ es un estimador consistente.

3.3 Criterios de evaluación de estimadores.

Error Cuadrático Medio y Sesgo

Definición: El sesgo de un estimador $\hat{\theta}$ para un parámetro θ es

$$\mathrm{B}(\hat{\theta}) = \mathrm{E}(\hat{\theta}) - \theta$$

Definición: El error cuadrático medio $\hat{\theta}$ es

$$ECM(\hat{\theta}) = \mathrm{E}(\hat{\theta} - \theta)^2$$

Observación: Si $\hat{\theta}$ es insesgado para θ, entonces $ECM(\hat{\theta}) = Var(\hat{\theta})$

Donde el Error Cuadrático Medio se puede escribir de la siguiente manera:

$$ECM(\hat{\theta}) = Var(\hat{\theta}) + B^2(\hat{\theta})$$

3.3 Criterios de evaluación de estimadores.

Error Cuadrático Medio y Sesgo

$$ECM(\hat{\theta}) = Var(\hat{\theta}) + B^2(\hat{\theta})$$

Demostración:

$$ECM(\hat{\theta}) = \mathrm{E}(\hat{\theta} - \theta)^2$$
$$= E(\hat{\theta}^2 - 2\theta\hat{\theta} + \theta^2)$$
$$= E(\hat{\theta}^2) - 2\theta E(\hat{\theta}) + \theta^2$$
$$= E(\hat{\theta}^2) - E^2(\hat{\theta}) + E^2(\hat{\theta}) - 2\theta E(\hat{\theta}) + \theta^2$$
$$= Var(\hat{\theta}) + (\mathrm{E}(\hat{\theta}) - \theta)^2$$
$$= Var(\hat{\theta}) + B^2(\hat{\theta})\ LQD$$

3.3 Criterios de evaluación de estimadores.

Error Cuadrático Medio y Sesgo

Ejemplo 1: Sea $X_1, \dots, X_n$ una m.a. de la dist. Exp(θ) en donde $\theta > 0$ es desconocido. Sea $\hat{\theta} = {}^{1}/_{\bar{x}}$.

Solución:
$$B(\hat{\theta}) = E(\hat{\theta}) - \theta = \frac{n}{n-1}\theta - \theta = \frac{\theta}{n-1}$$
$$Var(\hat{\theta}) = E(\hat{\theta}^2) - E^2(\hat{\theta})$$
$$= \frac{n^2}{(n-1)(n-2)}\theta^2 - \frac{n^2}{(n-1)^2}\theta^2$$

Observación:
$$\frac{n^2+1}{(n-1)^2} = \frac{(n-1)(n+1)}{(n-1)^2}$$

$$ECM(\hat{\theta}) = Var(\hat{\theta}) + B^2(\hat{\theta})$$
$$= \frac{n^2}{(n-1)(n-2)}\theta^2 - \frac{n^2}{(n-1)^2}\theta^2 + \frac{\theta^2}{(n-1)^2} = \left[\frac{n^2}{(n-1)(n-2)} - \frac{(n-1)(n+1)}{(n-1)^2}\right]\theta^2$$
$$= \frac{n^2 - (n+1)(n-2)}{(n-1)(n-2)}\theta^2 = \frac{n+2}{(n-1)(n-2)}\theta^2$$

3.3 Criterios de evaluación de estimadores.

Error Cuadrático Medio y Sesgo

Ejemplo 2: Las observaciones $X_1, X_2\ y\ X_3$ componen una muestra simple, de valores positivos, procedentes de una población con media θ y desviación típica 6. A partir de los siguientes estimadores de θ:

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

Obtener: El Sesgo y el Error Cuadrático Medio.

$$E(\hat{\theta}) = E\left[\frac{1}{2}(X_1 + X_2) + \frac{1}{3}X_3\right] = E\left(\frac{1}{2}(X_1 + X_2)\right) + E\left(\frac{1}{3}X_3\right)$$

$$= \frac{1}{2}(E(X_1) + E(X_2)) + \frac{1}{3}E(X_3) = \frac{1}{2}(\theta + \theta) + \frac{1}{3}\theta = \frac{4}{3}\theta$$

$$\mathrm{B}(\hat{\theta}) = \mathrm{E}(\hat{\theta}) - \theta = \frac{4}{3}\theta - \theta = \frac{\theta}{3}$$

3.3 Criterios de evaluación de estimadores.

Error Cuadrático Medio y Sesgo

Sigue Ejemplo 2: Se Calcula la Var($\hat{\theta}$)

$$Var(\hat{\theta}) = Var\left[\frac{1}{2}(X_1 + X_2) + \frac{1}{3}X_3\right] = Var\left(\frac{1}{2}(X_1 + X_2)\right) + Var\left(\frac{1}{3}X_3\right)$$

$$= \left(\frac{1}{2}\right)^2 (Var(X_1) + Var(X_2)) + \left(\frac{1}{3}\right)^2 Var(X_3) = \frac{1}{4}(\sigma^2 + \sigma^2) + \frac{1}{9}\sigma^2 = \frac{11}{18}\sigma^2$$

$$Sabiendo\ que\ \sigma = 6, tenemos:$$

$$Var(\hat{\theta}) = \frac{11}{18}\sigma^2 = \frac{11}{18}(6)^2 = 22$$

$$ECM(\hat{\theta}) = Var(\hat{\theta}) + B^2(\hat{\theta}) = 22 + \left(\frac{\theta}{3}\right)^2$$

$$ECM(\hat{\theta}) = 22 + \frac{\theta^2}{9}$$

3.4 Suficiencia.

Estadísticas Suficientes.

Definición: Una estadististica $T = T(x_1, \dots x_n)$ es suficiente para un parámetro θ si la distribución conjunta de $x_1, \dots, x_n$ condicionada al evento $T = t$, no depende de θ, para cualquier valor t de la estadística.

Definición: Un estadístico es suficiente respecto al parámetro θ si la distribución de probabilidad de la muestra aleatoria $X_1, \dots, X_n$ condicionada al estadístico $\mathrm{t} = T(X_1, \dots, X_n)$ no depende del parámetro θ, es decir

$$P((X_1, \dots, X_n)/T(X_1, \dots, X_n) = t\ ) = \frac{P\big((X_1, \dots, X_n) \cap T(X_1, \dots, X_n) = t\big)}{P(T(X_1, \dots, X_n) = t)}$$

$$No\ depende\ de\ \theta$$

3.4 Suficiencia.

Estadísticas Suficientes.

Ejemplo 1: Sea $x_1, \ldots, x_n$ una m.a. de la distribución Ber(θ), comprobar que la estadística T definida como $x_1 + \cdots + x_n$ es suficientes para θ.

$$P(X_1 = x_1, \ldots, X_n = x_n / T = t\ ) = \frac{P(X_1 = x_1, \ldots, X_n = x_n \cap T = t)}{P(T = t)}$$

$$= \frac{P(X_1 = x_1, \ldots, X_n = x_n)}{P(T = t)} 1_{\{t\}}(x_1 + \cdots + x_n)$$

$$= \frac{P(X_1 = x_1) \ldots P(X_n = x_n)}{P(T = t)} 1_{\{t\}}(n\bar{x})$$

$$= \frac{\theta^{x_1}(1-\theta)^{1-x_1} \ldots \theta^{x_n}(1-\theta)^{1-x_n}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} 1_{\{t\}}(n\bar{x}) = \frac{\theta^{n\bar{x}}(1-\theta)^{n-n\bar{x}}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} 1_{\{t\}}(n\bar{x})$$

$$= \frac{1}{\binom{n}{t}} 1_{\{t\}}(n\bar{x})$$

Nota: Función indicadora $= 1_{\{t\}} = 1_{\{0,1\}}$ para el caso de la función Bernoulli y tomando en cuenta que T es una v.a. Binomial con parámetros (n, θ)

$\therefore$ *Como NO depende de θ, concluimos que la estadistica T SI es suficiente para θ*

3.4 Suficiencia.

Estadísticas Suficientes.

Ejemplo 2: Sea $x_1, \dots, x_n$ una m.a. de la distribución $\text{Exp}(\theta)$, comprobar que la estadística T definida como $x_1 + \cdots + x_n$ es suficientes para θ. Donde T tiene dist. Gama de parámetros (n, θ)

$$f(x_1, \dots, x_n / T = t\,) = \frac{f(x_1, \dots, x_n \cap T = t)}{f_T(t)}$$

$$= \frac{f(x_1) \dots f(x_n)}{f_T(t)} 1_{\{t\}}(x_1 + \cdots + x_n) = \frac{\theta e^{-\theta x_1} \dots \theta e^{-\theta x_n}}{\frac{(\theta t)^{n-1}}{(n-1)!} \theta e^{-\theta t}} 1_{\{t\}}(n\bar{x})$$

$$= \frac{\theta^n e^{-\theta n \bar{x}}}{\frac{\theta^n t^{n-1}}{(n-1)!} e^{-\theta t}} 1_{\{t\}}(n\bar{x}) = \frac{(n-1)!}{t^{n-1}} 1_{\{t\}}(n\bar{x})$$

$\therefore$ *Como NO depende de θ, concluimos que la estadistica T SI es suficiente para θ*

3.4 Suficiencia.

Estadísticas Suficientes.

Ejemplo 3: Sea x_1, x_2, x_3 una m.a. de tamaño n=3 de la distribución Ber(θ), comprobar que la estadística T definida como $x_1 + x_2 + x_3$ no es suficientes para θ. Donde $(x_1, x_2, x_3) = (1,1,0)\ y\ t = 3$.

$$P(X_1 = 1, X_2 = 1, X_3 = 0/T = 3\ ) = \frac{P(X_1 = 1, X_2 = 1, X_3 = 0 \cap T = 3)}{P(T = 3)}$$

$$= \frac{P(X_1 = 1, X_2 = 1, X_3 = 0)}{P(x_1, x_2, x_3)\epsilon\{(1,1,0), (0,0,1)\}}$$

Si se cumple que $X_1 = 1, X_2 = 1, X_3 = 0$, entonces el evento $T = 3$ se satisface

$$= \frac{\theta\theta(1-\theta)}{\theta^2(1-\theta) + (1-\theta)^2\theta}$$

$$= \frac{\theta}{\theta + (1-\theta)} = \theta$$

$$\therefore\ Como\ SI\ depende\ de\ \theta, concluimos\ que\ la\ estadistica\ T\ NO\ es\ suficiente\ para\ \theta$$

3.4 Suficiencia.

El Teorema de Factorización.

Teorema de Factorización de Fisher-Neyman:

Una estadística $\mathrm{T} = \mathrm{T}(x_1, \dots, x_n)$ es suficiente para θ

$$\Leftrightarrow f(x_1, \dots, x_n; \theta) = \mathrm{g}\big(\mathrm{T}(x_1, \dots, x_n; \theta)\big) h(x_1, \dots, x_n) \qquad \text{(a)}$$

Demostración: Para el Caso Discreto

⇒Supongamos que T es suficiente para θ. Por la definición de Suficiencia podemos escribir:

$$\begin{aligned} f(x_1, \dots, x_n; \theta) &= f(x_1, \dots, x_n \cap T(x_1, \dots, x_n)) \\ &= f(x_1, \dots, x_n / T(x_1, \dots, x_n))\, f_T(T(x_1, \dots, x_n)) \qquad \text{(b)} \\ &= h(x_1, \dots, x_n) \mathrm{g}\big(\mathrm{T}(x_1, \dots, x_n; \theta)\big) \end{aligned}$$

Como (a) con (b) tenemos que $h(x_1, \dots, x_n) = f(x_1, \dots, x_n / T(x_1, \dots, x_n))$ y $\mathrm{g}\big(\mathrm{T}(x_1, \dots, x_n; \theta)\big) = f_T(T(x_1, \dots, x_n))$. Pero como hemos asumido que T es suficiente para θ, entonces por la definición de suficiencia, $f(x_1, \dots, x_n / T(x_1, \dots, x_n))$ no puede depender de θ, luego la función $h(x_1, \dots, x_n)$ solo depende de $x_1, \dots, x_n$.

3.4 Suficiencia.

El Teorema de Factorización.

Demostración: Para el Caso Discreto

$\Leftarrow$Por hipótesis $f(x_1, \dots, x_n; \theta) = \mathrm{g}\big(\mathrm{T}(x_1, \dots, x_n; \theta)\big)h(x_1, \dots, x_n)$

$$P(X_1 = x_1, \dots, X_n = x_n / T = t\ ) = \frac{P(X_1 = x_1, \dots, X_n = x_n \cap T = t)}{P(T = t)}$$

$$= \frac{P(X_1 = x_1, \dots, X_n = x_n)}{P(T = t)} 1_{\{t\}}(T(x_1 + \cdots + x_n))$$

Si $T(x_1 + \cdots + x_n) \neq t$, la probabilidad condicionada será cero y si $T(x_1 + \cdots + x_n) = t$ tenemos que:

$$= \frac{\mathrm{g}\big(\mathrm{T}(x_1, \dots, x_n; \theta)\big)h(x_1, \dots, x_n)}{\sum_{(y_1, \dots, y_n): T(y_1, \dots, y_n) = t} P(X_1 = y_1, \dots, X_n = y_n)} 1_{\{t\}}(T(x_1 + \cdots + x_n))$$

Utilizamos la hipótesis para la probabilidad condicional, $P(X_1 = x_1, \dots, X_n = x_n)$ admite la factorización de g por h.

3.4 Suficiencia.

El Teorema de Factorización.

Sigue Demostración: Para el Caso Discreto

$$=\frac{\mathrm{g}\big(\mathrm{T}(x_1,\ldots,x_n;\theta)\big)h(x_1,\ldots,x_n)}{\sum_{(y_1,\ldots,y_n):T(y_1,\ldots,y_n)=t}\mathrm{g}\big(\mathrm{T}(y_1,\ldots,y_n;\theta)\big)h(y_1,\ldots,y_n)}1_{\{t\}}(T(x_1+\cdots+x_n))$$

Sabiendo que $T(x_1,\ldots,x_n)=t$, nos queda:

$$=\frac{\mathrm{g}(\mathrm{t};\theta)h(x_1,\ldots,x_n)}{\mathrm{g}(\mathrm{t};\theta)\sum_{(y_1,\ldots,y_n):T(y_1,\ldots,y_n)=t}h(y_1,\ldots,y_n)}1_{\{t\}}(T(x_1+\cdots+x_n))$$

$$=\frac{h(x_1,\ldots,x_n)}{\sum_{(y_1,\ldots,y_n):T(y_1,\ldots,y_n)=t}h(y_1,\ldots,y_n)}1_{\{t\}}(T(x_1+\cdots+x_n))$$

Ya que la expresión no depende de θ, podemos concluir que T es suficiente.

3.4 Suficiencia.

El Teorema de Factorización.

Ejemplo 1: La estadística $\mathrm{T} = x_1 + \cdots + x_n$ es suficientes para θ de la distribución $\mathrm{Ber}(\theta)$.

$$\begin{aligned}
&P(X_1 = x_1, \ldots, X_n = x_n) \\
&= P(X_1 = x_1) \ldots P(X_n = x_n) \\
&= \theta^{x_1}(1-\theta)^{1-x_1} 1_{\{0,1\}}(x_1) \ldots \theta^{x_n}(1-\theta)^{1-x_n} 1_{\{0,1\}}(x_n) \\
&= \underbrace{\theta^{x_1+\cdots+x_n}(1-\theta)^{n-(x_1+\cdots+x_n)}}_{\mathrm{g}(\mathrm{T}(x_1,\ldots,x_n;\theta))} \underbrace{1_{\{0,1\}}(x_1) \ldots 1_{\{0,1\}}(x_n)}_{h(x_1,\ldots,x_n)}
\end{aligned}$$

Como esta factorización se cumple, podemos concluir que T es suficiente para θ.

3.4 Suficiencia.

El Teorema de Factorización.

Ejemplo 2: La estadística $\mathrm{T} = x_1 + \cdots + x_n$ es suficientes para θ de la distribución $\exp(\theta)$.

$$f(x_1, \dots, x_n; \theta) = f(x_1; \theta) \dots f(x_n; \theta)$$
$$= \theta e^{-\theta x_1} 1_{\{0,\infty\}}(x_1) \dots \theta e^{-\theta x_n} 1_{\{0,\infty\}}(x_n)$$
$$= \underbrace{\theta^n e^{-\theta(x_1 + \cdots + x_n)}}_{\mathrm{g}(\mathrm{T}(x_1,\dots,x_n;\theta))} \underbrace{1_{\{0,\infty\}^n}(x_1, \dots, x_n)}_{h(x_1,\dots,x_n)}$$

Como esta factorización se cumple, podemos concluir que T es suficiente para θ.

3.4 Suficiencia.

El Teorema de Factorización.

Ejemplo 3: La estadística $\mathrm{T} = máx\{x_1 + \cdots + x_n\} = X_{(n)}$ es suficientes para θ de la distribución unif$(0, \theta)$.

$$f(x_1, \dots, x_n; \theta) = f(x_1; \theta) \dots f(x_n; \theta)$$

$$= \frac{1}{\theta} 1_{\{0,\theta\}}(x_1) \dots \frac{1}{\theta} 1_{\{0,\theta\}}(x_n) = \frac{1}{\theta^n} 1_{\{0,\theta\}}(x_1) \dots 1_{\{0,\theta\}}(x_n)$$

$$= \frac{1}{\theta^n} 1_{\{0,x_n\}}(\theta) = \frac{1}{\theta^n} 1_{\{máx\{x_1+\cdots+x_n\},\infty\}}(\theta)$$

$$= \underbrace{\frac{1}{\theta^n} 1_{\{0,x_n\}}(\theta)}_{\mathrm{g}(\mathrm{T}(x_1,\dots,x_n;\theta))} \underbrace{1}_{h(x_1,\dots,x_n)}$$

Como esta factorización se cumple, podemos concluir que T es suficiente para θ.

3.4 Suficiencia.

El Estadísticas Suficientes Minimales.
Sea $x_1, \dots, x_n$ una m.a. de una distribución $f(x;\theta)$ y sean S y T dos estadísticas.

Definición: T es función de S si para cualesquiera valores $(x_1, \dots, x_n)$ y $(y_1, \dots, y_n)$ de la m.a. se cumple:

$$\mathrm{S}(x_1, \dots, x_n) = \mathrm{S}(y_1, \dots, y_n)$$
$$\Rightarrow T(x_1, \dots, x_n) = T(y_1, \dots, y_n)$$

Definición: Un estadística T es suficiente minimal para θ si

a) Es suficiente
b) Es minimal, es decir, T es función de cualquier otra estadística suficiente para θ.

3.4 Suficiencia.

El Estadísticas Suficientes Minimales.

Teorema: Sea T una estadística y sean $(x_1, \dots, x_n)$ y $(y_1, \dots, y_n)$ dos valores cualesquiera de la m.a. si se cumple las dos implicaciones

$$\frac{f(x_1, \dots, x_n; \theta)}{f(y_1, \dots, y_n; \theta)} \; no \; depende \; de \; \theta \; si \; y \; solo \; si \; T(x_1, \dots, x_n) = T(y_1, \dots, y_n)$$

Entones T es suficiente minimal para θ.

3.4 Suficiencia.

El Estadísticas Suficientes Minimales.

Demostración: "*T Suficiente*"

Sea $(x_1, \dots, x_n)$ un valor cualquiera de la m.a., Sea t $= T(x_1, \dots, x_n)$ y sea $(y_1, \dots, y_n)$ otro valor de la m.a. tal que $T(y_1, \dots, y_n) = t$.

Entonces $T(x_1, \dots, x_n) = T(y_1, \dots, y_n)$ y, por la implicación de derecha a izquierda,

$$\frac{f(x_1,\dots,x_n;\theta)}{f(y_1,\dots,y_n;\theta)}\ no\ depende\ de\ \theta$$

$$\Rightarrow \frac{f(x_1,\dots,x_n;\theta)}{f(y_1,\dots,y_n;\theta)} = h_0(x_1, \dots, x_n, y_1, \dots, y_n; \theta)$$

Podemos afirmar h_0 No depende θ

$$\Rightarrow f(x_1, \dots, x_n; \theta) = f(y_1, \dots, y_n; \theta) h_0(x_1, \dots, x_n, y_1, \dots, y_n)$$

$$\Rightarrow f(x_1, \dots, x_n; \theta) = g(T(x_1, \dots, x_n); \theta) h_0(x_1, \dots, x_n, y_1, \dots, y_n)$$

$$Luego\ de\ utilizar\ el\ Teorema\ de\ Fisher{-}Neyman\ el\ estadistico$$

$$T\ es\ suficiente\ para\ \theta$$

3.4 Suficiencia.

El Estadísticas Suficientes Minimales.

Demostración: "*T Minimal*"

Sea S otro estadística suficiente θ. Demostraremos que T es función de S. Sean $(x_1, \ldots, x_n)$ y sea $(y_1, \ldots, y_n)$ dos valor de la m.a. tales que $\mathrm{S}(x_1, \ldots, x_n) = \mathrm{S}(y_1, \ldots, y_n)$. Entonces tenemos que:

$$\frac{f(x_1, \ldots, x_n; \theta)}{f(y_1, \ldots, y_n; \theta)} = \frac{g(S(x_1, \ldots, x_n); \theta)h(x_1, \ldots, x_n)}{g(S(y_1, \ldots, y_n); \theta)h(y_1, \ldots, y_n)} \, no\ depende\ de\ \theta$$

$$= \frac{h(x_1, \ldots, x_n)}{h(y_1, \ldots, y_n)} \, no\ depende\ de\ \theta$$

$$\Rightarrow T(x_1, \ldots, x_n) = T(y_1, \ldots, y_n)$$

$$esto\ sifnifica\ que\ T\ es\ funcion\ de\ S \ \blacksquare$$

3.4 Suficiencia.

El Estadísticas Suficientes Minimales.

Ejemplo 1: Sea $x_1, \dots, x_n$ una m.a. de la distribución $\text{Ber}(\theta)$. Demostrar que la estadística $T = x_1 + \cdots + x_n$ es suficiente minimal.

Sean $(x_1, \dots, x_n)$ y sea $(y_1, \dots, y_n)$ dos valores de la m.a. Entonces:

$$\frac{f(x_1, \dots, x_n; \theta)}{f(y_1, \dots, y_n; \theta)} = \frac{\theta^{x_1}(1-\theta)^{1-x_1} \dots \theta^{x_n}(1-\theta)^{1-x_n}}{\theta^{y_1}(1-\theta)^{1-y_1} \dots \theta^{y_n}(1-\theta)^{1-y_n}}$$

$$= \frac{\theta^{n\bar{x}}(1-\theta)^{n-n\bar{x}}}{\theta^{n\bar{y}}(1-\theta)^{n-ny}} = \theta^{n(\bar{x}-\bar{y})}(1-\theta)^{n(\bar{y}-\bar{x})}$$

$$= \left(\frac{\theta}{1-\theta}\right)^{n(\bar{x}-\bar{y})}$$

$$\frac{f(x_1, \dots, x_n; \theta)}{f(y_1, \dots, y_n; \theta)} \textit{ no depende de } \theta \Leftrightarrow n(\bar{x}-\bar{y}) = 0$$

$$\Leftrightarrow T(x_1, \dots, x_n) = T(y_1, \dots, y_n)$$

$\therefore$ el estadístico T es suficiente minimal para θ ■

3.4 Suficiencia.

El Estadísticas Suficientes Minimales.

Ejemplo 2: Sea $x_1, \dots, x_n$ una m.a. de la distribución $N(\mu, \sigma^2)$. Demostrar que la estadística $T = (T_1, T_2) = (x_1 + \cdots + x_n, x_1^2 + \cdots + x_n^2)$ es suficiente minimal para θ=(μ, σ^2). Sean $(x_1, \dots, x_n)$ y sea $(y_1, \dots, y_n)$ dos valores de la m.a. Entonces:

$$\frac{f(x_1, \dots, x_n; \theta)}{f(y_1, \dots, y_n; \theta)} = \frac{(2\pi\sigma^2)^{-n/2} e^{\left(-\sum_{i=1}^{n} (x_i - \mu)^2 / 2\sigma^2\right)}}{(2\pi\sigma^2)^{-n/2} e^{\left(-\sum_{i=1}^{n} (y - \mu)^2 / 2\sigma^2\right)}}$$

$$= \frac{e^{\left(-\sum_{i=1}^{n} (x_i^2 - 2\mu n\bar{x} + n\mu) / 2\sigma^2\right)}}{e^{\left(-\sum_{i=1}^{n} (y_i^2 - 2\mu n\bar{y} + n\mu) / 2\sigma^2\right)}}$$

Desarrollamos los cuadrados, se cancela $(2\pi\sigma^2)^{-n/2}$ y también $n\mu$

$$= e^{\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} y_i^2\right) + \frac{n\mu}{\sigma^2}(\bar{x} - \bar{y})\right)}$$

3.4 Suficiencia.

El Estadísticas Suficientes Minimales.

Continúa Ejemplo 2: Sea $x_1, \dots, x_n$ una m.a. de la distribución $\mathrm{N}(\mu, \sigma^2)$. Demostrar que la estadística $T = (T_1, T_2) = (x_1 + \cdots + x_n, x_1^2 + \cdots + x_n^2)$ es suficiente minimal para θ=(μ, σ^2). Sean $(x_1, \dots, x_n)$ y sea $(y_1, \dots, y_n)$ dos valores de la m.a. Entonces:

$$\frac{f(x_1, \dots, x_n; \theta)}{f(y_1, \dots, y_n; \theta)} \; no\ depende\ de\ \theta \Leftrightarrow \begin{pmatrix} Exponete\ es\ cero \\ \forall\ \mu\ y\ \sigma^2 \end{pmatrix}$$

$$\Leftrightarrow \left(\bar{x} = \bar{y} \;\; y \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i^2 \right)$$

$$\Leftrightarrow \begin{pmatrix} T_1(x_1, \dots, x_n) = T_1(y_1, \dots, y_n) \\ T_2(x_1, \dots, x_n) = T_2(y_1, \dots, y_n) \end{pmatrix}$$

$$\therefore (T_1, T_2)\ es\ suciciente\ minimal\ para\ (\mu, \sigma^2) \;\; \blacksquare$$

3.5 Estimación insesgada.

Cota inferior de Crámer-Rao.

Sea $x_1, \dots, x_n$ una m.a. de una distribución con función de probilidad o función de densidad $f(x;\theta)$, dependiente de un parámetro desconocido θ. Sea T un estimador insesgado para una función parametral $\tau(\theta)$.

Teorema de Cota inferior de Crámer-Rao (CICR).

Si $\mathrm{T} = \mathrm{T}(X_1, \dots, X_n)$ es un estimador insesgado para la función parametral $\tau(\theta)$. Entonces:

$$Var(T) \geq \frac{\left(\tau'(\theta)\right)^2}{nE\left[\left(\frac{\partial}{\partial\theta} lnf(X;\theta)\right)^2\right]} \qquad \theta \in \Theta$$

Lema: Si se cumplen las condiciones de Regularidad

$$\Rightarrow nE\left[\left(\frac{\partial}{\partial\theta} lnf(X;\theta)\right)^2\right] = -nE\left(\frac{\partial^2}{\partial\theta^2} lnf(X;\theta)\right)$$

3.5 Estimación insesgada.

Cota inferior de Crámer-Rao.

Teorema de Cota inferior de Crámer-Rao (CICR).

Suponiendo que $f(x;\theta)$ cumple las condiciones de regularidad:

1) El soporte de $f(x;\theta)$ dado el conjunto $\{x: f(x;\theta) > 0\}$ no depende de θ.
2) Para toda x en el soporte de $f(x;\theta)$, la función $\ln f(x;\theta)$ es diferenciable respecto de θ.
3) Es valido el siguiente intercambio de derivada e integral:

$$0 = \frac{d}{d\theta}\int_{\mathbb{R}} f(x;\theta)\,dx = \int_{\mathbb{R}} \frac{\partial}{\partial\theta} f(x;\theta)\,dx$$

4) $$0 < nE\left[\left(\frac{\partial}{\partial\theta} ln f(x;\theta)\right)^2\right] < \infty$$

5) Es valido el siguiente intercambio de derivada e integral:

$$\frac{d}{d\theta}\int_{\mathbb{R}^n} T(\underline{x}) f(x;\theta)\,d\underline{x} = \int_{\mathbb{R}^n} \frac{\partial}{\partial\theta} T(\underline{x}) f(x;\theta)\,d\underline{x}$$

Observación:

$$\mathrm{T}(x_1, \ldots, x_n) = T(\underline{x})$$

$$x_1, \ldots, x_n = \underline{x}$$

3.5 Estimación insesgada.

Demostración Teorema de Cota inferior de Crámer-Rao (CICR).

Supongamos el Caso Continuo:

$$Como\ \frac{d}{d\theta}\int_{\mathbb{R}} f(x;\theta)\,dx = 1,$$

$$0 = \frac{d}{d\theta}\int_{\mathbb{R}} f(x;\theta)\,dx = \int_{\mathbb{R}} \frac{\partial}{\partial\theta} f(x;\theta)\,dx$$

$$= \int_{\mathbb{R}} \frac{\partial}{\partial\theta} e^{\ln f(x;\theta)}dx$$

$$= \int_{\mathbb{R}} f(x;\theta)\frac{\partial}{\partial\theta}\ln f(x;\theta)\,dx$$

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

3.5 Estimación insesgada.

Sigue Demostración Teorema de Cota inferior de Crámer-Rao (CICR).

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

Utilizando lo siguiente:

$$\mathrm{T}(x_1, \dots, x_n) = T(\underline{x})$$
$$x_1, \dots, x_n = \underline{x}$$

$$\tau'(\theta) = \frac{d}{d\theta} E(T)$$

$$= \frac{d}{d\theta} \int_{\mathbb{R}^n} T(x_1, \dots, x_n) f(x_1, \dots, x_n; \theta)\, dx_1 \dots dx_n = \frac{d}{d\theta} \int_{\mathbb{R}^n} T(\underline{x}) f(\underline{x}; \theta)\, d\underline{x}$$

$$= \int_{\mathbb{R}^n} T(\underline{x}) \frac{\partial}{\partial\theta} f(\underline{x}; \theta)\, d\underline{x} = \int_{\mathbb{R}^n} T(\underline{x}) \frac{\partial}{\partial\theta} e^{\ln f(\underline{x};\theta)}\, d\underline{x}$$

$$= \int_{\mathbb{R}^n} T(\underline{x}) f(\underline{x}; \theta) \frac{\partial}{\partial\theta} \ln f(\underline{x}; \theta)\, d\underline{x} = \int_{\mathbb{R}^n} T(\underline{x}) f(\underline{x}; \theta) \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln f(X_i; \theta)\, d\underline{x}$$

$$= E\left[T \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln f(X_i; \theta) \right]$$

3.5 Estimación insesgada.

Sigue Demostración Teorema de Cota inferior de Crámer-Rao (CICR).

Utilizando la $\mathrm{COV}(X,Y) = E(XY) - E(X)E(Y) \leq \sqrt{Var(X)}\sqrt{Var(Y)}$

$$\left(\tau'(\theta)\right)^2 = \left(\mathrm{COV}\left(T, \sum_{i=1}^{n} \frac{\partial}{\partial\theta} lnf(X_i;\theta)\right)\right)^2$$

$$= Var(T)Var\left(\sum_{i=1}^{n} \frac{\partial}{\partial\theta} lnf(X_i;\theta)\right) = Var(T)\sum_{i=1}^{n} Var\left(\frac{\partial}{\partial\theta} lnf(X_i;\theta)\right)$$

$$= Var(T)nVar\left(\frac{\partial}{\partial\theta} lnf(X_i;\theta)\right) = Var(T)nE\left[\left(\frac{\partial}{\partial\theta} lnf(X;\theta)\right)^2\right]$$

$$\therefore Var(T) \geq \frac{\left(\tau'(\theta)\right)^2}{nE\left[\left(\frac{\partial}{\partial\theta} lnf(X;\theta)\right)^2\right]} \quad \blacksquare$$

Usando la hipótesis de muestra aleatoria tenemos que $f(x_1, \dots, x_n;\theta) = (f(x;\theta))^n$

Sabiendo que: $E\left[\frac{\partial}{\partial\theta} lnf(X_i;\theta)\right] = 0$

3.5 Estimación insesgada.

Teorema de Cota inferior de Crámer-Rao (CICR).

Observaciones:

1. El resultado es valido bajo las "condiciones de regularidad", que se consideraron.
2. $\boldsymbol{f}(\boldsymbol{X};\boldsymbol{\theta})$ es la función de densidad o de probabilidad $f(x;\theta)$ evaluada en la v.a.x. la expresión $\boldsymbol{f}(\boldsymbol{X};\boldsymbol{\theta})$ es una v.a.
3. Cuando la función parametral $\tau(\theta)=\theta$, su derivada es 1, entonces:

$$Var(T) \geq \frac{1}{nE\left[\left(\frac{\partial}{\partial\theta} lnf(X;\theta)\right)^2\right]}$$

4. Si T es insegado para $\tau(\theta)$ y $\mathrm{Var}(T)=\mathrm{CICR}(\theta), \theta \ \epsilon\ \Theta$, entonces T es un UMVUE (Estimador Insesgado de Mínima Varianza), en ingles (Uniformly Minimum Variance Unbiased Estimator)

3.5 Estimación insesgada.

Cota inferior de Crámer-Rao (CICR).

Ejemplo 1: Sea X una v.a. con distribución $Ber(\theta)$, con $\theta \epsilon (0,1)$ desconocido. Sea $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ cualquier estimador insesgado para $\tau(\theta) = \theta$, Calcular CICR(θ):

Recordando: $f(x;\theta) = \theta^x (1-\theta)^{1-x}, con\ x = 0, 1.$

$$\Rightarrow \frac{\partial}{\partial\theta} lnf(X;\theta) = \frac{\partial}{\partial\theta} ln\theta^x (1-\theta)^{1-x}$$

$$= [xln\theta + (1-x)\ln(1-\theta)] = \frac{x}{\theta} - \frac{1-x}{1-\theta}$$

Ahora calculamos,

$$E\left[\left(\frac{\partial}{\partial\theta} lnf(X;\theta)\right)^2\right] = E\left[\left(\frac{x}{\theta} - \frac{1-x}{1-\theta}\right)^2\right] = E\left(\left(\frac{x}{\theta}\right)^2 - 2\left(\frac{x}{\theta}\right)\left(\frac{1-x}{1-\theta}\right) + \left(\frac{1-x}{1-\theta}\right)^2\right)$$

$$= \frac{\theta}{\theta^2} - \frac{2E(x(1-x))}{\theta(1-\theta)} + \frac{1-\theta}{(1-\theta)^2} = \frac{1}{\theta} + \frac{1}{1-\theta} = \frac{1}{\theta(1-\theta)}$$

Tenemos que CIRC:

$$\boldsymbol{CIRC} = \frac{\left(\boldsymbol{\tau}'(\boldsymbol{\theta})\right)^2}{\boldsymbol{nE}\left[\left(\frac{\boldsymbol{\partial}}{\boldsymbol{\partial\theta}} \boldsymbol{lnf}(\boldsymbol{X};\boldsymbol{\theta})\right)^2\right]}$$

$$= \frac{\boldsymbol{\theta}(\boldsymbol{1}-\boldsymbol{\theta})}{\boldsymbol{n}} \quad \boldsymbol{0} < \boldsymbol{\theta} < \boldsymbol{1}$$

3.5 Estimación insesgada.

Cota inferior de Crámer-Rao (CICR).

Ejemplo 2: Sea X una v.a. con distribución $\exp(\theta)$, con $\theta \epsilon (0, \infty)$ desconocido. Sea $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ cualquier estimador insesgado para $\tau(\theta) = \theta$, Calcular CICR(θ):

Recordando: $f(x;\theta) = \theta e^{-\theta x}\ para\ x > 0$

$$\Rightarrow \frac{\partial}{\partial\theta} lnf(X;\theta) = \frac{\partial}{\partial\theta} ln\theta e^{-\theta x} = \frac{\partial}{\partial\theta}(ln\theta - \theta x) = \frac{1}{\theta} - x$$

Ahora calculamos,

$$E\left[\left(\frac{\partial}{\partial\theta} lnf(X;\theta)\right)^2\right] = E\left(\frac{1}{\theta} - x\right)^2 = Var(x) = \frac{1}{\theta^2}$$

$$\Rightarrow CIRC = \frac{\left(\tau'(\theta)\right)^2}{nE\left[\left(\frac{\partial}{\partial\theta} lnf(X;\theta)\right)^2\right]} = \frac{\theta^2}{n}\ donde\ \theta > 0$$

3.5 Estimación insesgada.

Cota inferior de Crámer-Rao (CICR).

Ejemplo 3: Sea $X_1, \dots, X_n$ una muestra aleatoria con distribución $\mathrm{N}(0, \sigma^2)$. Calcular $\mathrm{CICR}(\sigma^2)$:

$$\Rightarrow \frac{\partial}{\partial \sigma^2} lnf(X; \sigma^2) = \frac{\partial}{\partial \sigma^2} \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}x^2}\right) = \frac{\partial}{\partial \sigma^2}\left(-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}x^2\right)$$

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

Calculamos: $\frac{\partial^2}{\partial (\sigma^2)^2} lnf(X; \sigma^2) = \frac{1}{2(\sigma^2)^2} - \frac{x^2}{(\sigma^2)^3}$

Ahora calculamos,

$$E\left[\frac{\partial^2}{\partial (\sigma^2)^2} lnf(X; \sigma^2)\right] = E\left(\frac{1}{2(\sigma^2)^2} - \frac{x^2}{(\sigma^2)^3}\right) = \frac{1}{2\sigma^4} - \frac{E(X^2)}{\sigma^6} = \frac{1}{2\sigma^4} - \frac{\sigma^2}{\sigma^6} = -\frac{1}{2\sigma^4}$$

$$\Rightarrow CIRC = \frac{\left(\tau'(\theta)\right)^2}{-nE\left[\frac{\partial^2}{\partial (\sigma^2)^2} lnf(X; \sigma^2)\right]} = \frac{n}{2\sigma^4}$$

3.5 Estimación insesgada.

Eficiencia.

Definición: Un estimador insesgado $\hat{\theta}_1$ es relativamente más eficiente que otro estimador insesgado $\hat{\theta}_2$ si $Var(\hat{\theta}_1) \leq Var(\hat{\theta}_2)$

Definición: Un estimador insesgado es eficiente si su varianza alcanza la CICR.

Definición: La eficiencia estimador insesgado $\hat{\theta}$ es

$$Efi(\hat{\theta}) = \frac{CIRC(\theta)}{Var(\hat{\theta})} \leq 1$$

Definición: Un estimador insesgado $\hat{\theta}_n = \hat{\theta}(X_1, \dots, X_n)$ es asintóticamente eficiente si

$$\lim_{n \to \infty} Efi(\hat{\theta}_n) = 1$$

3.5 Estimación insesgada.

Eficiencia.

Ejemplo 1: Sea $X_1, \ldots, X_n$ una m.a. de la distribución $Ber(\theta)$, con $0 < \theta < 1$ desconocido. Recordemos que:

$$CIRC(\theta) = \frac{\theta(1-\theta)}{n}$$

a) El estimador insesgado $\hat{\theta}_n$ definido como $(X_1 + \cdots + X_{n-1})/(n-1)$ tiene varianza

$$Var(\hat{\theta}_n) = {\theta(1-\theta)}\big/_{n-1}$$

Podemos ver que NO es eficiente porque:

$$Efi(\hat{\theta}) = \frac{CIRC(\theta)}{Var(\hat{\theta})} = \frac{n-1}{n} < 1$$

Pero SI es asintóticamente eficiente, ya que:

$$\lim_{n\to\infty} Efi(\hat{\theta}_n) = 1$$

3.5 Estimación insesgada.

Eficiencia.

b) El estimador insesgado $\hat{\theta}_n$ definido como $2(X_1 + 2X_2 + \cdots + nX_{n-1})/n(n-1)$ tiene varianza

$$Var(\hat{\theta}_n) = \frac{2(2n+1)}{3(n+1)} \frac{\theta(1+\theta)}{n}$$

Podemos ver que NO es eficiente porque:

$$Efi(\hat{\theta}) = \frac{CIRC(\theta)}{Var(\hat{\theta})} = \frac{3(n+1)}{2(2n+1)} < 1$$

Y NO es asintóticamente eficiente, ya que:

$$\lim_{n\to\infty} Efi(\hat{\theta}_n) = \frac{3}{4} < 1$$

3.5 Estimación insesgada.

Eficiencia.

Ejemplo 2: Las observaciones X_1 y X_2 componen una muestra aleatoria de tamaño 2 extraída de una población exponencial con paremtero θ, Determinar cual de los siguientes estimadores insesgados es más eficiente para θ.

$$\hat{\theta}_1 = \frac{X_1+X_2}{2} \quad \text{Y} \quad \hat{\theta}_2 = \frac{X_1+2X_2}{3}$$

Primero hay que demostrar que son estimadores insesgados:

$$E(\hat{\theta}_1) = E\left(\frac{X_1+X_2}{2}\right) = E(\bar{x}) = \theta \; por \; lo \; tanto \; es \; insegado$$

Podemos ver que $\frac{X_1+X_2}{2}$ es la media muestral

$$E(\hat{\theta}_2) = E\left(\frac{X_1+2X_2}{3}\right) = \frac{1}{3}E(x_1) + \frac{2}{3}E(x_2) = \frac{1}{3}\theta + \frac{2}{3}\theta = \theta \; por \; lo \; tanto \; es \; insegado$$

3.5 Estimación insesgada.

Eficiencia.

Sigue Ejemplo 2: Ahora hay que calcular sus Varianzas:

$$Var(\hat{\theta}_1) = Var\left(\frac{X_1 + X_2}{2}\right) = Var\left(\frac{1}{2}(X_1 + X_2)\right) = \left(\frac{1}{2}\right)^2 (Var(X_1) + Var(X_2))$$

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

$$Var(\hat{\theta}_2) = Var\left(\frac{X_1 + 2X_2}{3}\right) = \left(\frac{1}{3}\right)^2 Var(X_1) + \left(\frac{2}{3}\right)^2 Var(X_2) = \frac{1}{9}\sigma^2 + \frac{4}{9}\sigma^2 = \frac{5\sigma^2}{9}$$

Recordando la Definición: Un estimador insesgado $\hat{\theta}_1$es relativamente más eficiente que otro estimador insesgado $\hat{\theta}_2$ si $Var(\hat{\theta}_1) \leq Var(\hat{\theta}_2)$

$$\Rightarrow Var(\hat{\theta}_1) \leq Var(\hat{\theta}_2) = \frac{\sigma^2}{2} \leq \frac{5\sigma^2}{9} \quad \therefore \hat{\theta}_1 \; es \; más \; eficiente$$

Bibliografía

Ipiña, S. L. (2008). Inferencia Estadística y Análisis de Datos. Madrid: PEARSON Prentice-Hall.

Vazquez, J. (2017). Estimación Puntual. México: Proyecto PAIME UNAM.