PROGRAMACIÓN TEMA 2

Definición de estimador de un parámetro

Metodos de obtención de estimadores:

a) Método de los momentos

b) Método de máxima verosimilitud

Propiedades deseables de un estimador

- Insesgadez:

Proposición 1: Si $T_1$ y $T_2$ son dos estimadores insesgados de $\theta$ entonces toda combinación lineal convexa de $T_1$ y $T_2$ es un estimador de $\theta$.

Proposición 2: Si $T$ es estimador insesgado de $\theta \Rightarrow$ $\Rightarrow g(T)$ es estimador insesgado de $g(\theta)$ siendo $g(x) = ax + b$ (función lineal)

- Consistencia

Error cuadratico medio ECM

Proposición: $ECM(T) = Var[T] + (E(T) - \theta)^2$

Consistencia en Media cuadrática: $\lim_{n\to\infty} ECM(T) = 0$

Proposición: T consistente en media cuadr. $\Leftrightarrow \begin{cases} \lim_{n\to\infty} Var[T] = 0 \\ T \text{ es asintoticamente insesgado} \\ \lim_{n\to\infty} (T - \theta) = 0 \end{cases}$

Consistencia simple $\lim_{n\to\infty} P[\,|T_n - \theta| \leq \varepsilon\,] = 1 \quad \forall \varepsilon > 0$

- EFICIENCIA

Estimador insesgado de mínima varianza uniforme

T es e.i.m.v.u si:

T es estimador insesgado de $\theta$

Si T' es estimador insesgado de $\theta \Rightarrow Var[T] < Var[T']$

Estimador Eficiente: T estimador de $\theta$. Se dice que T es eficiente si: T es insesgado, $Var(T) =$ Cota $FCR(\theta)$

Condiciones de regularidad de Frechet-Cramer-Rao

1ª El campo de variación de X no puede depender de $\theta$

2ª El espacio paramétrico debe ser abierto, acotado o no.

3ª Existe $E\left[\dfrac{\partial \ln L_\theta(x_1, x_2, \dots x_n)}{\partial \theta}\right]^2$

Cota de Frechet-Cramer-Rao

$$\text{Cota } FCR(\theta) = \frac{\left(\frac{\partial h(\theta)}{\partial \theta}\right)^2}{-E\left[\frac{\partial^2 \ln L_\theta(x)}{\partial \theta^2}\right]}$$

Teorema de Frêchet-Cramer-Rao

Supongamos que se cumplen las condiciones de regularidad de FCR y que T es un estimador insesgado de $h(\theta)$. Siempre se cumple Cota $FCR(\theta) \leq Var[T]$. Entonces:

T es e.i.m.v.u $\Leftrightarrow \dfrac{\partial \ln L_\theta(x_1, x_2, \dots x_n)}{\partial \theta} = [T(x_1, x_2 \dots x_n) - h(\theta)] \cdot g(\theta)$

donde $g(\theta)$ depende tan sólo de $\theta$. En este caso:

$Var[T] =$ Cota $FCR(\theta)$

SUFICIENCIA: Un estadístico T es suficiente para $\theta$ si:

$P[x_1, x_2, \dots x_n / T = t]$ no depende de $\theta$ (V. d)

$f(x_1, x_2, \dots x_n / T = t]$ no depende de $\theta$ (V. c)

Teorema de factorización de Fisher-Neymann.

T es suficiente para $\theta$ $\Rightarrow \begin{cases} f_\theta(x_1, x_2, \dots x_n) = h(x_1, x_2, \dots x_n) \cdot g[T(x_1, x_2 \dots x_n), \theta] & V.c \\ P_\theta[x_1, x_2, \dots x_n] = h(x_1, x_2, \dots x_n) \cdot g[T(x_1, x_2, \dots x_n), \theta] & V.d \end{cases}$

donde h depende tan solo de la m.a.s. y g depende de $\theta$ y del valor que toma T en la muestra.

Familia exponencial

Si la funcion de densidad o de probabilidad son de la forma:

$$f_\theta(x) = a(\theta)\cdot b(x)\, e^{c(\theta)\cdot d(x)} \;;\; P_\theta(x) = a(\theta)\cdot b(x)\cdot e^{c(\theta)\cdot d(x)}$$

entonces $T(X_1, X_2, \dots X_n) = \sum_{i=1}^{n} d(X_i)$ es estadistico suficiente para $\theta$.

Si se trata de un parámetro k-dimensional $\theta = (\theta_1, \theta_2, \dots, \theta_k)$. Si la función de densidad o de probabilidad tienen la forma:

$$f_\theta(x) = a(\theta)\cdot b(x)\,.\, e^{\sum_{i=1}^{k} c_i(\theta)\cdot d_i(x)}$$

entonces el estadístico $T(X_1, X_2, \dots, X_n) = \left(\sum_{i=1}^{n} d_1(X_i), \sum_{i=1}^{n} d_2(X_i), \dots, \sum_{i=1}^{n} d_k(X_i)\right)$ es suficiente para $\theta = (\theta_1, \theta_2, \dots, \theta_k)$

# TEMA 2 - ESTIMACION PUNTUAL

Consideremos una población con una variable aleatoria X cuya función de probabilidad o densidad depende de un parámetro desconocido $\theta$ que toma valores en un conjunto determinado llamado espacio paramétrico. El problema de la estimación puntual consiste en asignar un valor al parámetro desconocido $\theta$.

## Estimador puntual

Sea X una v.a. distribuida como $X \rightsquigarrow F(x, \theta)$, $\theta$ desconocido. Sea $X_1, X_2, \ldots X_n$ una m.a.s. Diremos que $T(X_1, X_2, \ldots X_n)$ es un estimador de $\theta$ si:

i) T es un estadístico

ii) $T(x_1, x_2, \ldots x_n) = \hat{\theta}$ está en el espacio paramétrico de $\theta$, para toda realización muestral $x_1, x_2, \ldots x_n$.

Ejemplos

1º Sea $X \rightsquigarrow P(\lambda)$ $\lambda \in (0, +\infty)$. Sea $X_1, X_2, \ldots, X_n$ una m.a.s y consideremos

$$T(X_1, X_2, \ldots X_n) = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}$$

Entonces T es un estimador de $\lambda$ ya que:

i) T es un estadístico pues no contiene parámetros desconocidos

ii) $\bar{x} = T(x_1, x_2, \ldots x_n)$ está en el espacio paramétrico de $\lambda$ que es $(0, +\infty)$.

2º Sea $X \rightsquigarrow N(\mu, \sigma)$. Sea $X_1, X_2, \ldots, X_n$ una m.a.s. y $T(X_1, X_2, \ldots X_n) = \sum_{i=1}^{n} X_i$. ¿T es un estimador de $\mu$?

i) T es un estadístico evidentemente

ii) $T(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} x_i \in (-\infty, +\infty)$ que es el espacio paramétrico de $\mu$.

3º Sea $X \rightsquigarrow N(\mu, \sigma)$. Sea $X_1, X_2, \ldots, X_n$ una m.a.s. y consideremos $T(X_1, X_2, \ldots X_n) = \sum_{i=1}^{n} X_i$

¿ T es estimador de $\sigma$ ?

i) T es estadístico ya que no contiene parámetros desconocidos.

ii) $T(X_1, X_2, \ldots, X_n) = \bar{X} \in (-\infty, \infty)$ no tiene porqué estar en el espacio paramétrico de $\sigma$ que es $(0, +\infty)$ luego $\bar{X}$ no es estimador de $\sigma$.

Podemos concluir diciendo que para un parámetro existen infinidad de estimadores. Nos interesan aquellos que cumplan el mayor número de propiedades deseables en un estimador: insesgadez, suficiencia, consistencia y eficiencia.

## MÉTODOS DE OBTENCIÓN DE ESTIMADORES

### 1 Método de los momentos.

- Recordemos los momentos poblacionales

El momento poblacional de orden k es

$$\alpha_k = E[X^k] \quad k = 1, 2, \ldots$$

Si $k = 1 \Rightarrow \alpha_1 = E[X]$ media poblacional

Si $k = 2 \Rightarrow \alpha_2 = E[X^2]$ y esto, junto con el anterior valor nos sirve para calcular la varianza poblacional

$Var[X] = E[X^2] - (E[X])^2$

- Recordemos los momentos muestrales

El momento muestral de orden k, de la m.a.s. $X_1, X_2, \ldots X_n$ es:

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

Si $k = 1$ obtenemos la media muestral $O_1 = \bar{X} = \dfrac{\sum_{i=1}^{n} x_i}{n}$

Si $k = 2$ $\quad O_2 = \dfrac{\sum_{i=1}^{n} x_i^2}{n}$ y con estos dos valores podemos obtener la varianza muestral $S^2 = O_2 - O_1^2 = \dfrac{\sum_{i=1}^{n} x_i^2}{n} - \bar{X}^2$

Sea $X$ una variable aleatoria distribuida como:
$X \rightsquigarrow F_X(\theta_1, \theta_2, \dots \theta_K)$ es decir que la función de distribución o densidad de $X$ está afectado por $K$-parámetros desconocidos $\theta_1, \theta_2, \dots \theta_K$.
El método de los momentos consiste en obtener los estimadores puntuales de $\theta_1, \theta_2, \dots \theta_K$ igualando los $K$-primeros momentos poblacionales a los $K$-primeros momentos muestrales, es decir:

$$\left.\begin{aligned} \alpha_1(\theta_1, \theta_2, \dots \theta_K) &= a_1 \\ \alpha_2(\theta_1, \theta_2, \dots \theta_K) &= a_2 \\ &\vdots \\ \alpha_K(\theta_1, \theta_2, \dots \theta_K) &= a_K \end{aligned}\right\}$$

De este sistema despejamos $\theta_1, \theta_2, \dots \theta_K$.

Ejemplos:

1º Sea $X$ una v.a distribuida según una distribución de Poisson de parámetro $\lambda$ desconocido $X \rightsquigarrow P(\lambda)$
Se trata de estimar $\lambda$ por el método de los momentos.
Tomamos una m.a.s $X_1, X_2, \dots X_n$. Como tenemos que determinar un parámetro igualamos el primer momento poblacional al primer momento muestral es decir:

$$\alpha_1(\lambda) = a_1$$

$\alpha_1(\lambda) = E[X] = a_1 = \bar{x} \Rightarrow \lambda = \bar{x}$. En realidad esta expresión no es del todo correcta porque $\lambda$ no se puede saber con total exactitud. Mejor es poner que la estimación de $\lambda$, que representaremos por $\hat{\lambda}$ es $\bar{x}$, es decir $\hat{\lambda} = \bar{x}$.

2º Sea la v.a. $X \rightsquigarrow Exp(\lambda)$, con $\lambda$ desconocido
Tratamos de obtener un estimador de $\lambda$ por el método de los momentos.
De nuevo tomamos una m.a.s. $X_1, X_2, \dots X_n$

y como tan sólo queremos determinar un parámetro:

$$\alpha_1(\lambda) = O_1$$

$$\alpha_1(\lambda) = E[X] = \frac{1}{\lambda} \qquad O_1 = \bar{x} = \frac{\sum_1^n x_i}{n}$$ . Entonces

el estimador de $\lambda$, $\hat{\lambda}$ verifica:

$$\frac{1}{\hat{\lambda}} = \bar{x} \Rightarrow \hat{\lambda} = \frac{1}{\bar{x}}$$

3º Sea la v.a. $X \rightsquigarrow B(n,p)$ con $n$ conocido y $p$ desconocido. Obtener el estimador de los momentos para $p$.

Como tan sólo tenemos que determinar un parámetro, planteamos la ecuación:

$$\alpha_1(p) = O_1$$

Si tomamos una m.a.s $x_1, x_2, \dots, x_n$

$O_1 = \dfrac{\sum_1^n x_i}{n} = \bar{x}$, $\alpha_1(p) = n \cdot p = E[X]$ (ver distribución binomial en las tablas)

luego $n\hat{p} = \bar{x} \Rightarrow \hat{p} = \dfrac{\bar{x}}{n}$

4º Sea la v.a. $X \rightsquigarrow G(a,\lambda)$ donde son desconocidos tanto $a$ como $\lambda$. Se trata de estimar los parámetros $a$ y $\lambda$ por el método de los momentos. Sea la m.a.s. $x_1, x_2, \dots, x_n$. Como tenemos que determinar dos parametros montamos un sistema de dos ecuaciones con dos incógnitas, a saber:

$$\alpha_1(a,\lambda) = O_1$$

$$\alpha_2(a,\lambda) = O_2$$

De la distribucion gamma obtenemos:

$\alpha_1(a,\lambda) = \dfrac{\lambda}{a} = E[X]$ $\qquad \alpha_2(a,\lambda) = E[X^2] = Var[X] + (E[X])^2$

$$\alpha_2(a,\lambda) = \frac{\lambda}{a^2} + \left(\frac{\lambda}{a}\right)^2 = \frac{\lambda(1+\lambda)}{a^2}$$

Para la muestra tenemos

$$O_1 = \bar{x} = \frac{\sum x_i}{n}$$

$$O_2 = S^2 + \bar{x}^2 \quad \text{ya que} \quad S^2 = \frac{\sum x_i^2}{n} - \bar{x}^2 = O_2 - \bar{x}^2$$

Entonces:

$$\left.\begin{array}{l} \alpha_1(a,d) = O_1 \Rightarrow \dfrac{d}{a} = \bar{x} \\ \alpha_2(a,d) = \dfrac{d}{a^2} + \dfrac{d^2}{a^2} = S^2 + \bar{x}^2 \end{array}\right\}$$

Si sustituimos en la segunda ecuación $\bar{x}$:

$$\frac{d}{a^2} + \cancel{\frac{d^2}{a^2}} = S^2 + \cancel{\frac{d^2}{a^2}}$$

Nos queda

$$\left.\begin{array}{l} d = \bar{x} \cdot a \\ d = S^2 \cdot a^2 \end{array}\right\} \bar{x} \cdot \cancel{a} = S^2 \cdot a^{\cancel{2}}$$

$$\Rightarrow \boxed{a = \frac{\bar{x}}{S^2}} \Rightarrow d = \bar{x} \cdot \frac{\bar{x}}{S^2} \Rightarrow \boxed{d = \frac{\bar{x}^2}{S^2}}$$

## 2 Método de máxima verosimilitud

### Función de verosimilitud de una v.a. discreta

Supongamos que la v.a. $X$ se distribuye con función de probabilidad $P[X = x_i, \theta]$ siendo $\theta$ un parámetro desconocido. Si tomamos una m.a.s. $X_1, X_2, \dots X_n$, entonces la función de probabilidad de la v.a $n$-dimensional $(X_1, X_2, \dots X_n)$ es:

$$L(X_1, X_2, \dots X_n) = P[X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$$

$$= P[X_1 = x_1, \theta] \cdot P[X_2 = x_2, \theta] \cdots P[X_n = x_n, \theta]$$ ya que las variables $X_1, X_2, \dots X_n$ son independientes, es decir:

$$L(X_1, X_2, \dots X_n) = \prod_{i=1}^{n} P[X_i = x_i, \theta]$$. Al valor que toma la función $L$ en la muestra $X_1, X_2, \dots X_n$ se le

llama verosimilitud de la muestra $x_1, x_2, \dots x_n$

**Función de verosimilitud de una v.a. continua**

Supongamos que la v.a. $X$ se distribuye con función de densidad $f(x, \theta)$, siendo $\theta$ un parámetro desconocido. Si tomamos una m.a.s. $x_1, x_2, \dots, x_n$ entonces la función de densidad de la v.a. n-dimensional $(X_1, X_2, \dots X_n)$ es:

$L(X_1, X_2, \dots, X_n) = f(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$. Como $X_1, X_2, \dots, X_n$ son independientes entonces la función de verosimilitud de la muestra es:

$L(X_1, X_2, \dots X_n) = f(X_1 = x_1) \cdot f(X_2 = x_2) \dots f(X_n = x_n)$

y como las $X_i$ están idénticamente distribuidas; la función de verosimilitud (credibilidad) de la muestra es:

$$L_\theta(X_1, X_2, \dots, X_n) = \prod_{i=1}^{n} f(X = x_i; \theta)$$

Nos debe de quedar claro que la función de verosimilitud de una muestra tan sólo depende del parámetro desconocido $\theta$.

**Método de máxima verosimilitud**

Consiste en hallar un valor del parámetro $\hat{\theta}$ tal que

$$\max_{\theta} L_\theta(x_1, x_2, \dots x_n) = L_{\hat{\theta}}(x_1, x_2, \dots x_n)$$

Para hallar el parámetro de máxima verosimilitud seguiremos los siguientes pasos:

1º Hallaremos la función de verosimilitud $L_\theta(x_1, x_2 \dots x_n)$

2º Calcularemos el logaritmo neperiano de esta función $z = \ln L_\theta(x_1, x_2, \dots x_n)$

3º Hallamos el máximo $z$ al variar los valores de $\theta$:

a) C.N. $\frac{\partial z}{\partial \theta} = 0$ y obtenemos un valor $\hat{\theta}$

b) C.S. Hemos de comprobar que el anterior valor $\hat{\theta}$

es efectivamente un máximo:

$$\frac{\partial^2 Z(\hat{\theta})}{\partial \theta} < 0$$

Veamos un ejemplo.

Ejemplo 1:

Sea la v.a. $X \rightsquigarrow P(\lambda)$ con $\lambda$ desconocido. Se trata de obtener el estimador de máxima verosimilitud (EMV) de $\lambda$.

La función de probabilidad de la distribución de Poisson es:

$$P[X = x] = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$$

1º Tomemos una m.a.s. $X_1, X_2, \dots X_n$. Entonces la función de verosimilitud es:

$$L_\lambda(x_1, x_2, \dots x_n) = \frac{e^{-\lambda} \cdot \lambda^{x_1}}{x_1!} \cdot \frac{e^{-\lambda} \cdot \lambda^{x_2}}{x_2!} \cdots \frac{e^{-\lambda} \cdot \lambda^{x_n}}{x_n!} =$$

$$= \prod_{i=1}^{n} \frac{e^{-\lambda} \cdot \lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda} \cdot \lambda^{\Sigma x_i}}{\prod_{i=1}^{n} x_i!}$$

2º Hallamos $\ln L_\lambda$:

$$Z = \ln L_\lambda = \ln \frac{e^{-n\lambda} \cdot \lambda^{\Sigma x_i}}{\prod_{i=1}^{n} x_i!} = \ln e^{-n\lambda} + \ln \lambda^{\Sigma x_i} - \ln \Pi x_i!$$

$$= -n\lambda + \sum_{i=1}^{n} x_i \cdot \ln \lambda - \ln\left(\prod_{i=1}^{n} x_i!\right)$$

3º Hallamos el máximo de $Z$:

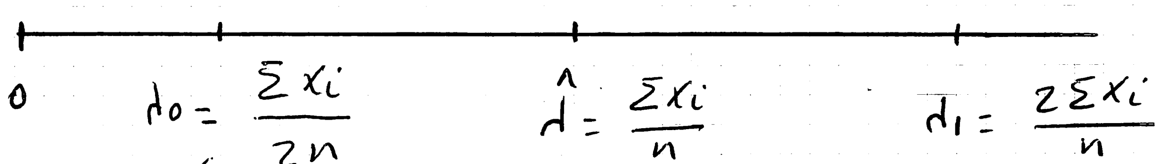
$$\text{C.N. } \frac{\partial Z}{\partial \lambda} = 0 \Rightarrow -n + \sum_{i=1}^{n} x_i \cdot \frac{1}{\lambda} = 0 \Rightarrow$$

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

C.S. $\frac{\partial^2 z}{\partial \lambda^2} = \frac{-\sum_{i=1}^{n} x_i}{\lambda^2} \Rightarrow \frac{\partial^2 z}{\partial \lambda^2}(\hat{\lambda}) < 0 \Rightarrow$

$\hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n}$ es máximo.

Con el método de la 2ª derivada obtenemos un máximo local (válido para nuestro propósito de calcular el máximo de la función $L_\theta$). Podríamos haber procedido de otra forma, asegurándonos que el máximo es global. Para comprobar que el valor obtenido con la C.N. $\hat{\lambda} = \frac{\sum x_i}{n}$ maximiza la función $z = L_\lambda$ podemos estudiar el signo de la derivada primera (si $\frac{\partial z}{\partial \lambda} > 0$ $z$ es creciente y si $\frac{\partial z}{\partial \lambda} < 0$ $z$ es decreciente) del siguiente modo:




Un valor cualquiera a la izquierda de $\hat{\lambda}$

Un valor cualquiera a la derecha de $\hat{\lambda}$

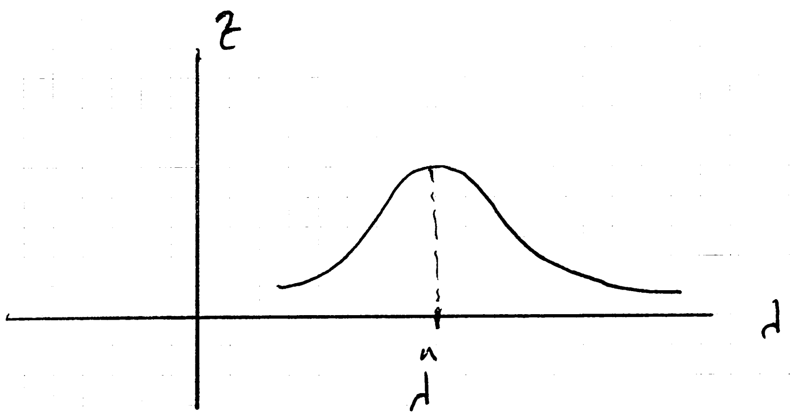
Para $\lambda_0 = \frac{\sum x_i}{2n} \Rightarrow \frac{\partial z}{\partial \lambda}(\lambda_0) = -n + \cancel{\sum x_i} \cdot \frac{1}{\frac{\cancel{\sum x_i}}{2n}} = n > 0 \Rightarrow$

$\Rightarrow z$ es creciente a la izquierda de $\hat{\lambda}$

Para $\lambda_1 = \frac{2\sum x_i}{n} \Rightarrow \frac{\partial z}{\partial \lambda}(\lambda_1) = -n + \cancel{\sum x_i} \cdot \frac{1}{\frac{2\cancel{\sum x_i}}{n}} = -\frac{n}{2} < 0$

$\Rightarrow z$ es decreciente a la derecha de $\hat{\lambda}$

Por lo tanto la grafica de $z$ tendría esta forma:




Lo cual nos indica que $\hat{\lambda}$ es el valor donde $z$ alcanza un máximo.

Ejemplo 2: Sea la variable aleatoria $X \rightsquigarrow N(\mu, \sigma)$ donde $\mu$ es conocida. Se trata de obtener el EMV para $\sigma$ y para $\sigma^2$

Solución:

La función de densidad de la distribución normal es:

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

Tomamos una m.a.s. $x_1, x_2, \ldots x_n$. Entonces la función de verosimilitud es:

$$L_\sigma(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma}\cdot e^{-\frac{1}{2}\cdot\frac{(x_i-\mu)^2}{\sigma^2}}$$ , es decir:

$$L_\sigma(x_1, x_2, \ldots x_n) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \cdot e^{-\frac{1}{2\sigma^2}\cdot\sum_{i=1}^{n}(x_i-\mu)^2}$$

Si tomamos logaritmos neperianos:

$$z = \ln L_\sigma(x_1, x_2, \ldots, x_n) = n \ln \frac{1}{\sqrt{2\pi}\,\sigma} - \frac{1}{2\sigma^2}\cdot\sum_{i=1}^{n}(x_i-\mu)^2$$

$$z = -n \ln \sqrt{2\pi}\,\sigma - \frac{\sum_1^n (x_i-\mu)^2}{2}\cdot\frac{1}{\sigma^2}$$ . Hemos de tener en cuenta que la variable en la función de verosimili-

tud es la desviación típica $\sigma$.
La condición de primer orden (condición necesaria) es en este caso:

$$\frac{\partial z}{\partial \sigma} = 0 \Rightarrow -n \cdot \frac{1}{\sqrt{2\pi}\cdot\sigma} \cdot \sqrt{2\pi} + \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2} \cdot \frac{2\sigma}{\sigma^4} = 0$$

$$\frac{\partial z}{\partial \sigma} = -\frac{n}{\sigma} + \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^3} = 0 \Rightarrow -n\sigma^2 + \sum_{i=1}^{n}(x_i-\mu)^2 = 0$$

$$\Rightarrow \sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n} \Rightarrow \sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}}$$

Veamos que estos estimadores son E.M.V.

$$\frac{\partial z}{\partial \sigma} = \frac{-n}{\sigma} + \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^3}$$ si derivamos nuevamente:

$$\frac{\partial^2 z}{\partial \sigma^2} = \frac{n}{\sigma^2} - \frac{3\sum(x_i-\mu)^2}{\sigma^4} = \frac{n\sigma^2 - 3\sum_{1}^{n}(x_i-\mu)^2}{\sigma^4}$$

$$\frac{\partial^2 z}{\partial \sigma^2} = \frac{\sum(x_i-\mu)^2 - 3\sum(x_i-\mu)^2}{\sigma^4} = \frac{-2\sum(x_i-\mu)^2}{\sigma^4} < 0$$

lo que nos indica que hay un máximo en

$$\sigma = \sqrt{\frac{\sum_{1}^{n}(x_i-\mu)^2}{n}}$$

## PROPIEDADES DESEABLES DE UN ESTIMADOR

### Insesgadez.

Sea $X$ una v.a. con distribución $F(x,\theta)$. Sea $T$ un estimador puntual de $\theta$. Se dice que $T$ es insesgado si su valor esperado coincide con $\theta$, es decir que $\forall\, X_1, X_2, \dots X_n$ m.a.s $E[T(X_1, X_2, \dots X_n)] = \theta$. Si $T$ no es insesgado se dice que es sesgado. En este caso se llama sesgo de $T$ al valor $E[T] - \theta$

Se dice que $T$ es asintoticamente insesgado si $\lim_{n\to\infty} E[T(X_1, X_2, \dots X_n)] = \theta$

Ejemplo 1: Sea $X \rightsquigarrow P(\lambda)$, $\lambda \in (0,+\infty)$. Entonces la media muestral es un estimador insesgado de $\lambda$.

Demostración:

Tomemos una m.a.s. $X_1, X_2, \dots X_n$ y consideremos el estadístico: $T(X_1, X_2, \dots X_n) = \bar{X} = \frac{\sum_1^n X_i}{n}$. En primer lugar podemos afirmar que $T$ es un estimador de $\lambda$ ya que:

i) $T$ no contiene parámetros desconocidos

ii) $\bar{X}$ está en el espacio paramétrico de $\lambda$ que es $(0,+\infty)$ ya que $\bar{X} = T(X_1, X_2, \dots X_n) > 0$ al ser positivos todos los $X_i$ ($X_i = 0, 1, 2, \dots$ en la distribución de Poisson)

Veamos que el estimador $T(X_1, X_2, \dots X_n) = \bar{X} = \frac{\sum_{i=1}^n X_i}{n}$ es insesgado

$$E[T(X_1, X_2, \dots X_n)] = E[\bar{X}] = \frac{1}{n} \cdot E\left[\sum_{i=1}^n X_i\right] = \frac{1}{n} \cdot \sum_{i=1}^n E[X_i]$$

Como en la distribución de Poisson $E[X] = \lambda \Rightarrow$

$\Rightarrow E[T(X_1, X_2, \dots X_n)] = \frac{1}{n} \sum_{i=1}^n \lambda = \frac{1}{n} \cdot n\lambda = \lambda$. Luego el estimador es insesgado.

Ejemplo 2: Sea la v.a. $X \rightsquigarrow Exp(\lambda)$. Entonces la media muestral no es un estimador insesgado para $\lambda$.

Demostración:

Tomemos una m.a.s. $X_1, X_2, \dots, X_n$ y consideremos el estadístico $T(X_1, X_2, \ldots X_n) = \bar{X} = \dfrac{\sum_{i=1}^{n} x_i}{n}$

Entonces:

$$E[T(X_1, X_2, \ldots X_n)] = E[\bar{X}] = \frac{1}{n} \cdot \sum_{i=1}^{n} E(x_i) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\lambda}$$

$= \frac{1}{n} \cdot \frac{n}{\lambda} = \frac{1}{\lambda} \neq \lambda$ no es estimador insesgado

Ejemplo 3: ¿Es la varianza muestral un estimador insesgado de la varianza poblacional? NO, la razón es la siguiente:

Sabemos que la relación entre la varianza muestral y la varianza poblacional es $S^2 = \frac{n-1}{n} \sigma^2$. Por lo tanto:

$$E[S^2] = E\left[\frac{n-1}{n} \sigma^2\right] \Rightarrow E[S^2] = \frac{n-1}{n} \sigma^2 \neq \sigma^2$$

donde $n$ es el tamaño de la muestra

Ejemplo 4: ¿Es la varianza muestral un estimador asintóticamente insesgado de la varianza poblacional? SI ya que $E[S^2] = \frac{n-1}{n} \sigma^2$. Por tanto:

$$\lim_{n \to \infty} E[S^2] = \lim_{n \to \infty} \frac{n-1}{n} \sigma^2 = \sigma^2$$

**Proposición**: Si $T_1$ y $T_2$ son estimadores insesgados para el parámetro $\theta$, entonces toda combinación lineal convexa de estos estimadores es también un estimador insesgado para $\theta$, es decir $\forall \alpha \in [0,1] \Rightarrow T = \alpha T_1 + (1-\alpha) T_2$ es estimador insesgado para $\theta$

Demostración:

En primer lugar veamos que si $T_1$ y $T_2$ son estimadores de $\theta$ también lo es $\alpha T_1 + (1-\alpha) T_2$:

* Para toda m.a.s. $X_1, X_2, \ldots X_n$ se verifica que $T(X_1, X_2, \ldots X_n) = \alpha\, T_1(X_1, X_2, \ldots X_n) + (1-\alpha) \cdot T_2(X_1, X_2, \ldots X_n)$ no contiene parametros de la poblacion que sean desconocidos ya que $\alpha$ no es parámetro poblacional.

* $\alpha T_1(x_1, x_2, \dots, x_n) + (1-\alpha) T_2(x_1, x_2, \dots x_n)$ están en el espacio paramétrico de $\theta$ ya que lo están $T_1$ y $T_2$, también lo está cualquier combinación lineal convexa.
* Como $E[T_1] = \theta$, $E[T_2] = \theta$ $\forall \alpha \in (0,1)$:

$$E[\alpha \cdot T_1 + (1-\alpha) T_2] = \alpha \cdot E[T_1] + (1-\alpha) E[T_2] = \alpha\theta + (1-\alpha)\theta$$

$= \alpha\theta + \theta - \alpha\theta = \theta \Rightarrow \alpha T_1 + (1-\alpha) T_2$ es insesgado.

**Proposición**: Sea $T$ un estimador insesgado para $\theta$ y $g(x) = ax + b$ una función lineal, siendo $a, b \in \mathbb{R}$. Entonces $g(T)$ es un estimador insesgado para $g(\theta)$

Demostración:

Si $T$ es un estimador de $\theta \Rightarrow g(T)$ es estimador de $g(\theta)$ ya que

- $T$ no contiene parámetros desconocidos $\Rightarrow g(T) = aT + b$ tampoco contiene parámetros desconocidos ya que $a$ y $b$ se suponen conocidos
- Si el espacio paramétrico de $\theta$ es $(K, L)$ entonces al ser $T$ estimador de $\theta$, $\forall x_1, x_2, \dots x_n$ m.a.s se verifica que $K < T(x_1, x_2, \dots x_n) < L \Rightarrow$

$$aK + b < aT(x_1, x_2, \dots x_n) + b < aL + b \Rightarrow$$

$\Rightarrow g(K) < g(T) < g(L)$ es decir que $g(T)$ está en el espacio paramétrico de $g(\theta)$.
- Si $T$ es insesgado $\Rightarrow E[T] = \theta$. Por tanto:

$$E[g(T)] = E[aT+b] = a \cdot E[T] + b = a \cdot \theta + b = g(\theta)$$

$\Rightarrow g(T)$ es estimador insesgado de $g(\theta)$

Ejemplo:

Si $\bar{x}$ es insesgado para estimar $\mu$ entonces $5\bar{x} + 3$ es insesgado para estimar $5\mu + 3$.

## CONSISTENCIA

En adelante nos colocamos en la siguiente situación: Sea X una v.a. con función de distribución $X \rightsquigarrow F(x, \theta)$ siendo $\theta$ un parámetro desconocido. Consideremos una función $h(\theta)$ del parámetro $\theta$, llamando $\hat{h}(\theta)$ a la estimación de la función $h(\theta)$. Con el estudio de la consistencia pretendemos controlar la proximidad entre $\hat{h}(\theta)$ y $h(\theta)$

### Error cuadrático medio

Sea T un estimador del parámetro $\theta$. Se define el error cuadrático medio de T como:

$$ECM(T) = E[(T-\theta)^2]$$

**Proposición**: $ECM(T) = Var[T] + [E[T]-\theta]^2$

Demostración

$$ECM(T) = E[(T-\theta)^2] = E[((T-E(T)+(E(T)-\theta))^2]$$

$$ECM(T) = E[(T-E(T))^2 + 2\cdot(T-E(T))(E(T)-\theta) + (E(T)-\theta)^2]$$

Hemos de tener en cuenta que $E(T)$ es constante y $\theta$ es también constante. Por lo tanto:

$$ECM(T) = E[(T-E(T))^2] + 2[E(T)-\theta]\cdot E[T-E[T]] + E[(E(T)-\theta)^2]$$

Como $E[T-E[T]] = E[T]-E[T] = 0$ y $E[(E[T)-\theta)^2] = (E(T)-\theta)^2$ se tiene que:

$$ECM(T) = E[(T-E(T))^2] + (E[T]-\theta)^2$$

Además $var[T] = E[(T-E[T])^2]$ con lo que:

$$ECM(T) = Var[T] + (E[T]-\theta)^2$$

Lo que nos interesa es que $ECM(T)$ sea lo más pequeño posible.

**Definición: consistencia en media cuadrática.**

Sea $\{T_n\}_{n\in\mathbb{N}}$ una sucesión de estimadores del parámetro $\theta$. Se dice que $\{T_n\}_{n\in\mathbb{N}}$ es consistente en media cuadrática si converge en media cuadrática a $\theta$, es decir:

$$\lim_{n\to\infty} E[(T_n-\theta)^2] = \lim_{n\to\infty} ECM(T_n) = 0$$ donde

EMC es el error cuadrático medio.

Recordemos la propiedad del error cuadrático medio:

$$ECM(T_n) = Var[T_n] + [E(T_n)-\theta]^2$$

**Proposición**: Una sucesión de estimadores de $\theta$ $\{T_n\}_{n\in\mathbb{N}}$ es consistente en media cuadrática si y sólo si $\{T_n\}_{n\in\mathbb{N}}$ es asintóticamente insesgado ($\lim_{n\to\infty} E[T_n] = \theta$) y tambien verifica $\lim_{n\to\infty} Var[T_n] = 0$

Demostración:

$\{T_n\}_{n\in\mathbb{N}}$ es consistente en media cuadrática $\Leftrightarrow$

$$\lim_{n\to\infty} ECM(T_n) = 0 \Leftrightarrow \lim_{n\to\infty} [Var[T_n] + (E[T_n]-\theta)^2] = 0$$

$$\Leftrightarrow \lim_{n\to\infty} Var[T_n] + \lim_{n\to\infty} (E(T_n)-\theta)^2 = 0$$ y como

estos dos sumandos no pueden ser negativos $\Leftrightarrow$

$$\Leftrightarrow \lim_{n\to\infty} Var[T_n] = 0, \lim_{n\to\infty} (E(T_n)-\theta)^2 = 0 \Leftrightarrow$$

$$\Leftrightarrow \lim_{n\to\infty} Var[T_n] = 0, \lim_{n\to\infty} E[T_n] = \theta.$$

**Definición: consistencia simple**

Sea $\{T_n\}_{n\in\mathbb{N}}$ una sucesión de estimadores de $\theta$. Diremos que esta sucesión es simplemente consistente si $\{T_n\}_{n\in\mathbb{N}} \xrightarrow{P} \theta$ es decir que $\forall \varepsilon > 0$ se cumple:

$$\lim_{n\to\infty} P[|T_n-\theta| > \varepsilon] = 0$$, o lo que es lo mismo:

$$\lim_{n\to\infty} P[|T_n-\theta| \le \varepsilon] = 1 \quad \forall \varepsilon > 0$$

La relación entre consistencia en media cuadrática

y consistencia simple es:
$\{T_n\}_{n\in\mathbb{N}}$ es consistente en media cuadrática $\Rightarrow$ $\{T_n\}_{n\in\mathbb{N}}$ es simplemente consistente. El recíproco no es cierto.

Comparación de estimadores
Sean $T_1$ y $T_2$ dos estimadores de un mismo parámetro $\theta$. Se dice que $T_1$ es mejor que $T_2$ si
$ECM(T_1) < ECM(T_2)$

## EFICIENCIA Y MINIMA VARIANZA

Definición de estimador insesgado de mínima varianza uniforme (eimvu)
Un estimador $T$ de $\theta$ es insesgado de mínima varianza uniforme si verifica:

$$ECM(T) < ECM(T') \quad \forall\, T' \text{ estimador insesgado de } \theta$$

Al ser $T$ insesgado $E[T]=\theta$ y lo mismo para $T'$ que también es insesgado $E[T']=\theta$.
Sabemos que:
$ECM(T) = Var[T] + (E(T)-\theta)^2 = Var[T]$ y lo mismo para $T'$, $ECM(T') = Var[T']$

Entonces $T$ es estimador insesgado de mínima varianza para $\theta$ si:

$$Var[T] < Var[T'] \quad \forall\, T' \text{ estimador insesgado de } \theta$$

El eimvu es de todos los estimadores insesgados aquel que tiene menor varianza.
Para que exista e.i.m.v.u debe haber un estimador insesgado de $\theta$ en el que la varianza de los estimadores insesgados de $\theta$ alcanza un mínimo.
Este mínimo es una cota inferior alcanzable:
$Cota(\theta) = ECM(T) = Var[T]$, donde $T$ es eimvu de $\theta$

## CONDICIONES DE REGULARIDAD DE FRECHET-CRAMER-RAO

Sea X una v.a. distribuida como $X \rightsquigarrow F(x,\theta)$

Se dice que X cumple las condiciones de regularidad de Frechet-Cramer-Rao si verifica:

1º El campo de variación de X no puede depender de $\theta$

Por ejemplo si se definiera la función de densidad de X como:

$$f(x) = \begin{cases} \dfrac{1}{\theta} & \text{si } 0 < x < \theta \\ 0 & \text{en el resto} \end{cases}$$

Entonces X no cumpliría esta condición. Sería correcto el poner $f(x) = \frac{1}{\theta}$, lo que no es admisible es que $0 < x < \theta$ ya que el campo de variación de X dependería de $\theta$.

2º El espacio paramétrico (campo de variación de $\theta$) debe ser un conjunto abierto, acotado o no.

Por ejemplo si $a < \theta < b \Rightarrow \theta \in (a,b)$ se cumple la propiedad, pero si $a \leq \theta < b \Rightarrow \theta \in [a,b)$ o bien $a < \theta \leq b \Rightarrow \theta \in (a,b]$ o $a \leq \theta \leq b \Rightarrow \theta \in [a,b]$ se infringe la propiedad. Además a o b, o ambos pueden ser $\infty$.

3º Si $X_1, X_2, \dots X_n$ es m.a.s. y $L_\theta(x_1, x_2, \dots x_n)$ es la función de verosimilitud, ha de existir:

$$E\left[\frac{\partial \ln L_\theta(x_1,x_2,\dots x_n)}{\partial \theta}\right]^2$$

**Cota de Frechet-Cramer-Rao**

Sea la v.a. $X \rightsquigarrow F(x,\theta)$. Sea $X_1, X_2, \dots, X_n$ una m.a.s. y $L_\theta(x_1, x_2, \dots x_n)$ la función de verosimilitud. Entonces la cota de FCR es:

$$\text{Cota FCR}(\theta) = \frac{\left(\dfrac{\partial h(\theta)}{\partial \theta}\right)^2}{-E\left[\dfrac{\partial^2 \ln L_\theta(x_1,x_2,\dots x_n)}{\partial \theta^2}\right]}$$

Teorema de Frechet- Cramer - Rao
Sea X una v. a. que verifica las condiciones de regularidad de FCR. Sea T un estimador insesgado de $h(\theta)$ (una función del parámetro $\theta$).
Entonces:
i) cota $FCR(\theta) \leq Var(T)$
ii) T es eimvu $\iff$ $\dfrac{\partial \ln L_\theta(x_1, x_2, \dots x_n)}{\partial \theta} = (T(x_1,x_2,\dots x_n) - h(\theta))g(\theta)$

donde $g(\theta)$ es una función que no depende de la muestra, tan solo depende de $\theta$.

**Pasos a seguir para estudiar la existencia y el alcance del e.i.m.v.u.**

1º Si no se cumplen las condiciones de regularidad de FCR, la cota de FCR no es relevante, el teorema no se puede aplicar.

2º Si se cumplen las condiciones de regularidad de FCR, tenemos una primera condición de existencia del eimvu. La condición ii) del teorema nos dice si la cota de FCR se alcanza en $T(x_1, x_2, \dots x_n)$. En caso de que no podamos afirmar que se cumple esta condición no se puede concluir que no exista el eimvu.

**Definición: estimador eficiente.**
Decimos que el estimador T de $h(\theta)$ es un estimador eficiente si cumple:
i) T es insesgado, es decir $E[T] = h(\theta)$

ii) $Var[T]$ coincide con la cota $FCR(\theta)$ $\forall \theta$, es decir:
$$Var[T] = \text{Cota } FCR(\theta) \quad \forall \theta.$$

Observación:
El denominador de la cota FCR tiene varias expresiones equivalentes:

$$-E\left[\frac{\partial^2 \ln L_\theta(x_1,x_2,\dots x_n)}{\partial\theta^2}\right]=E\left[\left(\frac{\partial \ln L_\theta(x_1,x_2,\dots x_n)}{\partial\theta}\right)^2\right]$$

$$-n\cdot E\left[\frac{\partial^2 \ln L_\theta(x)}{\partial\theta^2}\right]=n\cdot E\left[\left(\frac{\partial \ln L_\theta(x)}{\partial\theta}\right)^2\right]$$

La que más se utiliza es:

$$-n\cdot E\left[\frac{\partial^2 \ln L_\theta(x)}{\partial\theta^2}\right]$$

Veamos un ejemplo, el ejercicio 18 de la relación 2.

Ejercicio 18 relación 2

Dada la función de densidad $f_\theta(x)=\begin{cases}\dfrac{x^{-(1+\frac{1}{\theta})}}{\theta} & \text{si } x>1,\ \theta>1\\ 0 & \text{en el resto}\end{cases}$

halle la distribución de $Y=\ln x$ y demuestre que $E[\ln X]=\theta$ y que $Var[\ln X]=\theta^2$. Obtenga el estimador máximo verosímil de $\theta$ y demuestre sus propiedades de insesgadez, consistencia, suficiencia y eficiencia.

Según la regla del cambio de variable tenemos que

$f_\theta(y)=f_\theta(x)\cdot\left|\dfrac{dx}{dy}\right|$; si $y=\ln x \Rightarrow x=e^y \Rightarrow$

$\Rightarrow \dfrac{dx}{dy}=e^y$. Por tanto:

$f_\theta(y)=f_\theta(x)\cdot e^y=f_\theta(e^y)\cdot e^y$. Entonces:

$$f_\theta(y)=\begin{cases}\dfrac{(e^y)^{-(1+\frac{1}{\theta})}}{\theta}\cdot e^y & \text{si } e^y>1 \quad \theta>1\\ 0 & \text{en el resto}\end{cases}$$

Si $e^y>1 \Rightarrow \ln e^y>\ln 1 \Rightarrow y>0$. Además:

$$\frac{(e^y)^{-(1+\frac{1}{\theta})}}{\theta}\cdot e^y=\frac{\cancel{e^{-y}}\cdot e^{-\frac{1}{\theta}y}\cdot \cancel{e^y}}{\theta}=\frac{1}{\theta}\cdot e^{-\frac{1}{\theta}y}$$

Podemos poner entonces que la función de densidad de $y = \ln x$ es:

$$f_\theta(y) = \begin{cases} \frac{1}{\theta} \cdot e^{-\frac{1}{\theta} y} & \text{si } y > 0 \\ 0 & \text{en el resto} \end{cases}$$

Si recordamos la función de densidad de $y \rightsquigarrow Exp(\lambda)$

$$f_\lambda(y) = \begin{cases} \lambda \cdot e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{en el resto} \end{cases}$$

si $\lambda = \frac{1}{\theta}$ $\quad f_\lambda(y) = f_\theta(y)$

Por tanto $y = \ln x \rightsquigarrow Exp\left(\frac{1}{\theta}\right)$

Si utilizamos la tabla de las distribuciones de probabilidad y miramos la exponencial, rápidamente podemos observar que:

$$E[\ln x] = \frac{1}{\frac{1}{\theta}} = \theta \qquad Var[\ln x] = \frac{1}{\left(\frac{1}{\theta}\right)^2} = \theta^2$$

Como nos piden que lo demostremos no tenemos más remedio que hacer lo siguiente:

$$E[\ln x] = E[y] = \int_{-\infty}^{+\infty} y \cdot f_\theta(y)\, dy = \int_0^\infty y \cdot \frac{1}{\theta} \cdot e^{-\frac{1}{\theta} y} dy$$

$$= \frac{1}{\theta} \cdot \int_0^\infty y \cdot e^{-\frac{1}{\theta} y} dy = \frac{1}{\theta} \cdot \lim_{t \to \infty} \int_0^t y \cdot e^{-\frac{1}{\theta} y} dy$$

Hacemos por partes la siguiente integral:

$$\int y \cdot e^{-\frac{1}{\theta} y} dy \begin{cases} u = y \Rightarrow du = dy \\ dv = e^{-\frac{1}{\theta} y} dy \begin{cases} r = -\frac{1}{\theta} y \\ dr = -\frac{1}{\theta} dy \\ dy = -\theta dr \end{cases} \Rightarrow v = -\theta \int e^r dr = -\theta \cdot e^{-\frac{1}{\theta} y} \end{cases}$$

$$\int y \cdot e^{-\frac{1}{\theta} y} dy = -\theta y \cdot e^{-\frac{1}{\theta} y} + \theta \int e^{-\frac{1}{\theta} y} dy = -\theta y e^{-\frac{1}{\theta} y} - \theta^2 \cdot e^{-\frac{1}{\theta} y}$$

Luego:

$$\int y\cdot e^{-\frac{1}{\theta}y}\,dy = -\theta\cdot e^{-\frac{1}{\theta}y}(y+\theta) \quad \text{con lo cual:}$$

$$\int_0^t y\cdot e^{-\frac{1}{\theta}y}\,dy = -\theta\cdot e^{-\frac{1}{\theta}t}(t+\theta)+\theta^2 = \frac{-\theta(t+\theta)}{e^{\frac{1}{\theta}t}}+\theta^2$$

$$\lim_{t\to\infty}\int_0^t y\cdot e^{-\frac{1}{\theta}y}\,dy = \lim_{t\to\infty}\frac{-\theta(t+\theta)}{e^{\frac{1}{\theta}t}}+\theta^2 = \theta^2 \quad \text{ya}$$

que $\displaystyle\lim_{t\to\infty}\frac{-\theta(t+\theta)}{e^{\frac{1}{\theta}t}} = \frac{\infty}{\infty} \overset{\text{L'Hop}}{=} \lim_{t\to\infty}\frac{-\theta}{\frac{1}{\theta}\cdot e^{\frac{1}{\theta}t}} = \frac{-\theta}{\infty} = 0$

Como:

$$E[\ln x] = \frac{1}{\theta}\cdot\lim_{t\to\infty}\int_0^t y\cdot e^{-\frac{1}{\theta}y}\,dy = \frac{1}{\theta}\cdot\theta^2 = \theta.$$

De forma análoga se demostraría que $Var[\ln x] = \theta^2$

$$Var[\ln x] = Var[y] = \int_{-\infty}^{\infty} y^2\cdot f(y)\,dy =$$

$$= \int_0^{\infty} y^2\cdot\frac{1}{\theta}\cdot e^{-\frac{1}{\theta}y}\,dy$$ (se integra aplicando dos veces integración por partes).

Vamos a obtener ahora el estimador de máxima verosimilitud de $\theta$.

1º Hallamos la función de verosimilitud de $\theta$:

Sea $X_1, X_2, \dots X_n$ una m.a.s.

$$L_\theta(x_1,x_2,\dots x_n) = \prod_{i=1}^n f_\theta(x_i) = \prod_{i=1}^n \frac{x_i^{-(1+\frac{1}{\theta})}}{\theta} \Rightarrow$$

$$L_\theta(x_1,x_2,\dots x_n) = \frac{\left[\prod_{i=1}^n x_i\right]^{-(1+\frac{1}{\theta})}}{\theta^n}$$

2º Tomamos $\ln L_\theta(x)$:

$$\ln L_\theta(x_1, x_2, \ldots, x_n) = \ln \frac{\left[\prod_{i=1}^{n} x_i\right]^{-(1+\frac{1}{\theta})}}{\theta^n}$$

3º $$\ln L_\theta(x_1, x_2, \ldots x_n) = -\left(1+\frac{1}{\theta}\right) \ln\left[\prod_{i=1}^{n} x_i\right] - n \ln\theta$$

$$\frac{\partial \ln L_\theta(x_1, x_2, \ldots x_n)}{\partial\theta} = \frac{1}{\theta^2} \cdot \ln\left[\prod_{i=1}^{n} x_i\right] - \frac{n}{\theta} = 0$$

$$\frac{1}{\theta}\left[\frac{1}{\theta} \ln\left[\prod_{i=1}^{n} x_i\right] - n\right] = 0 \Rightarrow \frac{1}{\theta} \ln\left[\prod_{i=1}^{n} x_i\right] = n$$

$$\Rightarrow \theta = \frac{\ln\left[\prod_{i=1}^{n} x_i\right]}{n} \Rightarrow \theta = \frac{\sum_{i=1}^{n} \ln x_i}{n}$$

Luego el estimador de máxima verosimilitud es:

$$\theta_{EMV} = \frac{\sum_{i=1}^{n} \ln x_i}{n}$$

Vamos a comprobar las propiedades de $\theta_{EMV}$:

Insesgadez: $E[\theta_{EMV}] = \theta$

Recordemos que $y = \ln x \rightsquigarrow Exp\left(\frac{1}{\theta}\right)$ y que por tanto $E[\ln x] = \frac{1}{\frac{1}{\theta}} = \theta$.

$$E[\theta_{EMV}] = E\left[\frac{\sum_{i=1}^{n} \ln x_i}{n}\right] = \frac{1}{n} \cdot E\left[\sum_{i=1}^{n} \ln x_i\right] =$$

$$= \frac{1}{n} \cdot \sum_{i=1}^{n} E[\ln x_i] = \frac{1}{n} \sum_{i=1}^{n} \theta = \frac{1}{n} \cdot n\theta = \theta$$

Consistencia:

Tenemos que demostrar que $\lim_{n\to\infty} ECM(\theta_{EMV}) = 0$

Como $ECM(\theta_{EMV}) = Var[\theta_{EMV}] + (E(\theta_{EMV}) - \theta)^2$ y al ser insesgado $E(\theta_{EMV}) - \theta = 0$ tenemos que probar que se verifica: $\lim_{n\to\infty} Var[\theta_{EMV}] = 0$.

$$Var[\theta_{EMV}] = Var\left[\frac{\sum_1^n \ln x_i}{n}\right] = \frac{1}{n^2} Var\left[\sum_{i=1}^n \ln x_i\right]$$

Al ser los $x_i$ independientes, también son independientes $\ln x_i$ y por tanto:

$$Var\left[\sum \ln x_i\right] = \sum_{i=1}^n Var[\ln x_i]$$

Como $\ln x_i \rightsquigarrow Exp\left(\frac{1}{\theta}\right) \Rightarrow Var[\ln x_i] = \theta^2$

Luego:

$$Var[\theta_{EMV}] = \frac{1}{n^2} \cdot n\theta^2 = \frac{\theta^2}{n}$$ y entonces:

$$\lim_{n\to\infty} Var[\theta_{EMV}] = \lim_{n\to\infty} \frac{\theta^2}{n} = 0 \Rightarrow \theta_{EMV} = \frac{\sum_{i=1}^n \ln x_i}{n}$$

es consistente.

Eficiencia.

Para probar que $\theta_{EMV}$ es eficiente hay que demostrar que:

i) $\theta_{EMV}$ es insesgado (ya lo hemos probado)

ii) $Var[\theta_{EMV}] =$ Cota $FCR(\theta)$

Como ya sabemos $Var[\theta_{EMV}] = \frac{\theta^2}{n}$ veamos cuanto vale Cota $FCR(\theta)$

$$\text{Cota } FCR(\theta) = \frac{\left(\frac{\partial h(\theta)}{\partial\theta}\right)^2}{-nE\left[\frac{\partial^2 \ln L_\theta(x)}{\partial\theta^2}\right]}$$

En este caso $h(\theta) = \theta \Rightarrow \frac{\partial h(\theta)}{\partial\theta} = 1$

$$f_\theta(x) = \frac{x^{-(1+\frac{1}{\theta})}}{\theta} \Rightarrow \ln f_\theta(x) = -\left(1+\frac{1}{\theta}\right)\ln x - \ln\theta$$

Luego: $\ln L_\theta(x) = -\left(1+\frac{1}{\theta}\right)\ln x - \ln\theta \Rightarrow$

$$\Rightarrow \frac{\partial \ln L_\theta(x)}{\partial\theta} = \frac{1}{\theta^2}\ln x - \frac{1}{\theta} = \frac{\ln x - \theta}{\theta^2}$$

$$\frac{\partial^2 \ln L_\theta(x)}{\partial\theta^2} = \frac{-\theta^2 - (\ln x - \theta)\cdot 2\theta}{\theta^4} = \frac{\theta^2 - \ln x \cdot 2\theta}{\theta^4}$$

es decir:

$$\frac{\partial^2 \ln L_\theta(x)}{\partial \theta^2} = \frac{\theta - 2\ln x}{\theta^3}$$

Entonces la cota de Frêchet-Cramer-Rao es:

$$\text{Cota FCR} = \frac{1}{-n \cdot E\left[\frac{\theta - 2\ln x}{\theta^3}\right]}$$

$$E\left[\frac{\theta - 2\ln x}{\theta^3}\right] = \frac{1}{\theta^3}\cdot\left(E[\theta] - 2\cdot E[\ln x]\right) =$$

$$= \frac{1}{\theta^3}(\theta - 2\cdot\theta) = -\frac{1}{\theta^2}$$

Por tanto Cota FCR $= \frac{\theta^2}{n}$ y se verifica entonces que:

$$Var[\theta_{EMV}] = \text{Cota FCR} \Rightarrow \theta_{EMV} \text{ es eficiente}$$

Ejemplo:

Sea la v.a. $X \rightsquigarrow U(a, \theta)$. Sea $x_1, x_2, \dots, x_n$ una m.a.s. y consideremos el estadístico $T(x_1, x_2, \dots, x_n) = \max\{x_1, x_2, \dots x_n\}$. Entonces el estadístico $T$ es asintoticamente insesgado para $\theta$

Solución:

Queremos demostrar que $\lim_{n\to\infty} E[T(x_1, x_2, \cdots x_n)] = \theta$ y para ello necesitamos la función de densidad de $T$. Como conocemos la función de densidad de $X$, si pudiéramos aplicar la fórmula del cambio de variable:

$f_T(t) = f_X(X(t)) \cdot \left|\frac{d\,X(t)}{dT}\right|$, sería fácil obtener la función de densidad de $T$. Pero ocurre que

$T(X_1, X_2, \dots X_n) = \max\{X_1, X_2, \dots X_n\}$ no es ni continua ni derivable. No nos queda otro remedio que obtener la función de densidad de T a partir de la función de distribución de T, derivándolo. La función de distribución de T es:

$$F_T(t) = P[T \leq t] = P[\max\{X_1, X_2, \dots, X_n\} \leq t] =$$

$$= P[X_1 \leq t, X_2 \leq t, \dots, X_n \leq t] = \prod_{i=1}^{n} P(X_i \leq t) = \prod_{i=1}^{n} F_{X_i}(t)$$ ya

que las $X_i$ son independientes. Veamos cual es la función de distribución de $F_{X_i}(t)$. La función de densidad de cada $X_i$ es:

$$f_{X_i}(x_i) = \begin{cases} \frac{1}{\theta - a} & \text{si } a \leq t < \theta \\ 0 & \text{en el resto} \end{cases}$$

Entonces:

$$F_{X_i}(t) = \begin{cases} 0 & \text{si } t < a \\ \int_a^t \frac{1}{\theta - a} \, dt & \text{si } a \leq t < \theta \\ 0 & \text{si } t > \theta \end{cases}$$

$$F_{X_i}(t) = \begin{cases} 0 & \text{si } t \leq a \\ \frac{t-a}{\theta - a} & \text{si } a \leq t < \theta \\ 0 & \text{si } t \geq \theta \end{cases}$$

La función de distribución de T es:

$$F_T(t) = \begin{cases} 0 & \text{si } t < a \\ \prod_{i=1}^{n} \left(\frac{t-a}{\theta - a}\right) & \text{si } a \leq t < \theta \\ 0 & \text{si } t > \theta \end{cases}$$

es decir:

$$F_T(t) = \begin{cases} 0 & \text{si } t < a \\ \left(\dfrac{t-a}{\theta-a}\right)^n & \text{si } a \leq t < \theta \\ 1 & \text{si } t > \theta \end{cases}$$

la función de densidad de $T$ la obtenemos derivando la función de distribución, es decir:

$$f_T(t) = F_T'(t) = \begin{cases} 0 & \text{si } t < a \\ n \cdot \left(\dfrac{t-a}{\theta-a}\right)^{n-1} \cdot \dfrac{1}{\theta-a} & \text{si } a \leq t < \theta \\ 0 & \text{si } t > 0 \end{cases}$$

con lo que:

$$f_T(t) = \begin{cases} \dfrac{n}{(\theta-a)^n} \cdot (t-a)^{n-1} & \text{si } a \leq t < \theta \\ 0 & \text{en el resto} \end{cases}$$

Ahora podemos calcular $E[T(X_1, X_2, \cdots X_n)] = E[T]$
Por definición se tiene que:

$$E[T] = \int_{-\infty}^{+\infty} t \cdot f_T(t)\, dt = \int_a^\theta t \cdot \frac{n}{(\theta-a)^n} \cdot (t-a)^{n-1} dt$$

$$E[T] = \frac{n}{(\theta-a)^n} \int_a^\theta t \cdot (t-a)^{n-1} dt$$

Vamos a hallar por partes la integral indefinida:

$$\int t\,(t-a)^{n-1} dt \begin{cases} u = t \Rightarrow du = dt \\ dv = (t-a)^{n-1} dt \Rightarrow v = \dfrac{(t-a)^n}{n} \end{cases} =$$

$$= t \cdot \frac{(t-a)^n}{n} - \frac{1}{n} \int (t-a)^n dt = t \cdot \frac{(t-a)^n}{n} - \frac{1}{n(n+1)} (t-a)^{n+1}$$

Luego:

$$\int_a^\theta t\cdot(t-a)^{n-1}dt = t\cdot\frac{(t-a)^n}{n} - \frac{1}{n(n+1)}(t-a)^{n+1}\Big|_a^\theta =$$

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

Por tanto:

$$E[T] = \frac{n}{(\theta-a)^n}\cdot\frac{(\theta-a)^n}{n}\left(\theta-\frac{\theta-a}{n+1}\right) = \theta - \frac{\theta-a}{n+1}$$

y se verifica que:

$\lim\limits_{n\to\infty} E[T] = \lim\limits_{n\to\infty}\left(\theta - \frac{\theta-a}{n+1}\right) = \theta \Rightarrow T$ es asintóticamente insesgado para $\theta$.

Ejemplo:

Sea la v.a. $X \rightsquigarrow U(\theta, b)$. Tomemos una m.a.s $X_1, X_2, \dots X_n$ y definamos el estadístico:

$T(X_1, X_2, \dots, X_n) = \min\{X_1, X_2, \dots X_n\}$. Entonces $T$ es asintóticamente insesgado para el límite inferior del intervalo, $\theta$.

Solución:

Este ejercicio es muy parecido al anterior. Tenemos que probar que $\lim\limits_{n\to\infty} E[T] = \theta$. Necesitamos la función de densidad de $T$ que, de nuevo, hemos de obtener derivando la función de distribución.

La función de distribución de $T$ es:

$$F_T(t) = P[T \le t] = P[\min\{X_1, X_2, \dots X_n\} \le t] =$$

$$= 1 - P[\min\{X_1, X_2, \dots, X_n\} > t] =$$

$= 1 - P[X_1 > t, X_2 > t \dots, X_n > t]$. Al ser los $X_i$ independientes:

$F_T(t) = 1 - \prod_{i=1}^{n} P[x_i > t]$. Como:

$$P(x_i > t) = \int_t^b \frac{1}{b-\theta}\, dt = \frac{b-t}{b-\theta}$$

$$F_T(t) = 1 - \prod_{i=1}^{n} \left(\frac{b-t}{b-\theta}\right) = 1 - \left(\frac{b-t}{b-\theta}\right)^n$$

Luego

$$F_T(t) = \begin{cases} 0 & \text{si } t < \theta \\ 1 - \left(\frac{b-t}{b-\theta}\right)^n & \text{si } \theta \leq t < b \\ 0 & \text{si } t \geq b \end{cases}$$

La función de densidad de $T$ es: $f_T(t) = F_T'(t)$

$$f_T(t) = \begin{cases} n\left(\frac{b-t}{b-\theta}\right)^{n-1} \cdot \frac{1}{b-\theta} & \text{si } \theta \leq t \leq b \\ 0 & \text{en el resto} \end{cases}$$

Es decir que:

$$f_T(t) = \begin{cases} \frac{n}{(b-\theta)^n} \cdot (b-t)^{n-1} & \text{si } \theta \leq t \leq b \\ 0 & \text{en el resto.} \end{cases}$$

Vamos a calcular ahora $E[T]$

$$E[T] = \int_{-\infty}^{+\infty} t \cdot f_T(t)\, dt = \int_\theta^b t \cdot \frac{n}{(b-\theta)^n} (b-t)^{n-1} dt$$

Calculemos en primer lugar la integral indefinida:

$$\int t \frac{n}{(b-\theta)^n} \cdot (b-t)^{n-1} dt = \frac{n}{(b-\theta)^n} \int t \cdot (b-t)^{n-1} dt$$

Tenemos que:

$$\int t\cdot(b-t)^{n-1}dt \left\{\begin{array}{l} u=t \Rightarrow du=dt \\ dv=(b-t)^{n-1}dt \\ v=-\dfrac{(b-t)^n}{n}\end{array}\right\} = -t\cdot\frac{(b-t)^n}{n}+\frac{1}{n}\int(b-t)^n dt$$

$$= -t\cdot\frac{(b-t)^n}{n} - \frac{1}{n}\cdot\frac{(b-t)^{n+1}}{n+1} = \frac{(b-t)^n}{n}\left(-t-\frac{b-t}{n+1}\right)$$

Entonces:

$$\int_\theta^b t\cdot\frac{n}{(b-\theta)^n}\cdot(b-t)^{n-1}dt = \frac{n}{(b-\theta)^n}\cdot\left[\frac{(b-t)^n}{n}\left(-t-\frac{b-t}{n+1}\right)\right]_\theta^t =$$

$$\frac{n}{(b-\theta)^n}\cdot\left[\frac{(b-\theta)^n}{n}\cdot\left(\theta+\frac{b-\theta}{n+1}\right)\right] = \theta+\frac{b-\theta}{n+1}$$

Como $E[T]=\theta+\dfrac{b-\theta}{n+1} \Rightarrow \lim\limits_{n\to\infty} E[T]=\theta$

Ejercicio 12 Relación 2

Considere una población $U(0,\theta)$ definida para $\theta>0$

a) Demuéstrese que el estadístico $T(X_1,X_2\dots X_n) = \dfrac{n+1}{n}\cdot\max\{x_1,x_2,\dots x_n\}$ definido a partir de una m.a.s. $X_1,X_2,\dots,X_n$, es insesgado para $\theta$ y tiene varianza igual a $\dfrac{\theta^2}{n(n+2)}$

b) Calcule la cota FCR. ¿Se viola en este caso la tesis del Teorema correspondiente? En caso afirmativo, ¿cuál es la razón de que esto ocurra?

Solución

a) La función de densidad de $X \rightsquigarrow U(0,\theta)$ es:

$$f_X(x)=\begin{cases}\dfrac{1}{\theta} & \text{si } 0\le x\le\theta \\ 0 & \text{en el resto}\end{cases}$$

Nuestro propósito es demostrar que $E[T] = \theta$.
Como:

$$E[T] = E\left[\frac{n+1}{n} \cdot \max\{x_1, x_2, \dots, x_n\}\right] =$$

$$= \frac{n+1}{n} \cdot E[\max\{x_1, x_2, \dots x_n\}]$$, hemos de hallar la función de densidad de $Y = \max\{x_1, x_2, \dots x_n\}$
Esto ya lo hemos realizado en un ejemplo anterior y la función de densidad es:

$$f_Y(y) = \begin{cases} \dfrac{n}{\theta^n} \cdot y^{n-1} & \text{si } 0 < y < \theta \\ 0 \text{ en el resto} \end{cases}$$

Por tanto $E[Y] = \displaystyle\int_0^\theta y \cdot \frac{n}{\theta^n} \cdot y^{n-1} dy =$

$$= \frac{n}{\theta^n} \cdot \int_0^\theta y^n dy = \frac{n}{\theta^n} \cdot \left[\frac{y^{n+1}}{n+1}\right]_0^\theta = \frac{n \cdot \theta^{n+1}}{(n+1) \cdot \theta^n} = \frac{n}{n+1} \cdot \theta$$

Al ser:

$$E[T] = \frac{n+1}{n} \cdot E[Y] \Rightarrow E[T] = \theta$$

La varianza de $T$ es:

$$Var[T] = Var\left[\frac{n+1}{n} \cdot Y\right] = \frac{(n+1)^2}{n^2} \cdot Var[Y]$$

$$Var[Y] = \int_{-\infty}^{\infty} y^2 \cdot f(y)\, dy = \int_0^\theta y^2 \cdot \frac{n}{\theta^n} \cdot y^{n-1} dy$$

$$Var[Y] = \frac{n}{\theta^n} \cdot \int_0^\theta y^{n+1} dy = \frac{n}{\theta^n} \cdot \left.\frac{y^{n+2}}{n+2}\right|_0^\theta = \frac{n \cdot \theta^{n+2}}{\theta^n (n+2)}$$

$Var[Y] = \dfrac{n\theta^2}{n+2}$. Luego:

$$Var[T] = \frac{(n+1)^2}{n^2} \cdot \frac{n \cdot \theta^2}{n+2} = \frac{(n+1)^2}{n(n+2)} \cdot \theta^2$$

b) Cálculo de la cota de FCR

Como $X \rightsquigarrow U(0,\theta)$ la función de densidad de $X$ es:

$$f_X(x) = \begin{cases} \frac{1}{\theta} & \text{si } 0 \leq x < \theta \\ 0 & \text{en el resto} \end{cases}$$

Como no se cumplen las condiciones de regularidad de FCR ya que la función de densidad $f(x)$ depende de $\theta$, tampoco podemos asegurar que se vayan a cumplir las tesis del teorema, es decir, que no podemos ni afirmar ni desmentir que el estimador de $\theta$, $T$, sea el eimvu. La cota de FCR no es relevante:

$$f_\theta(x) = \frac{1}{\theta} \Rightarrow \ln f_\theta(x) = \ln\left(\frac{1}{\theta}\right) = -\ln\theta$$

$$\frac{\partial \ln f_\theta(x)}{\partial\theta} = -\frac{1}{\theta} \Rightarrow \frac{\partial^2 \ln f_\theta(x)}{\partial\theta^2} = \frac{1}{\theta^2}$$

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

Luego la Cota de FCR $= \dfrac{1}{-n\cdot\frac{1}{\theta^2}} = -\dfrac{\theta^2}{n}$

**Observación: conjunto de nivel de una función real de variable vectorial.**

Sea la función $f: \mathbb{R}^n \longrightarrow \mathbb{R}$. Se define la curva o conjunto de nivel $k$ de $f$ como:

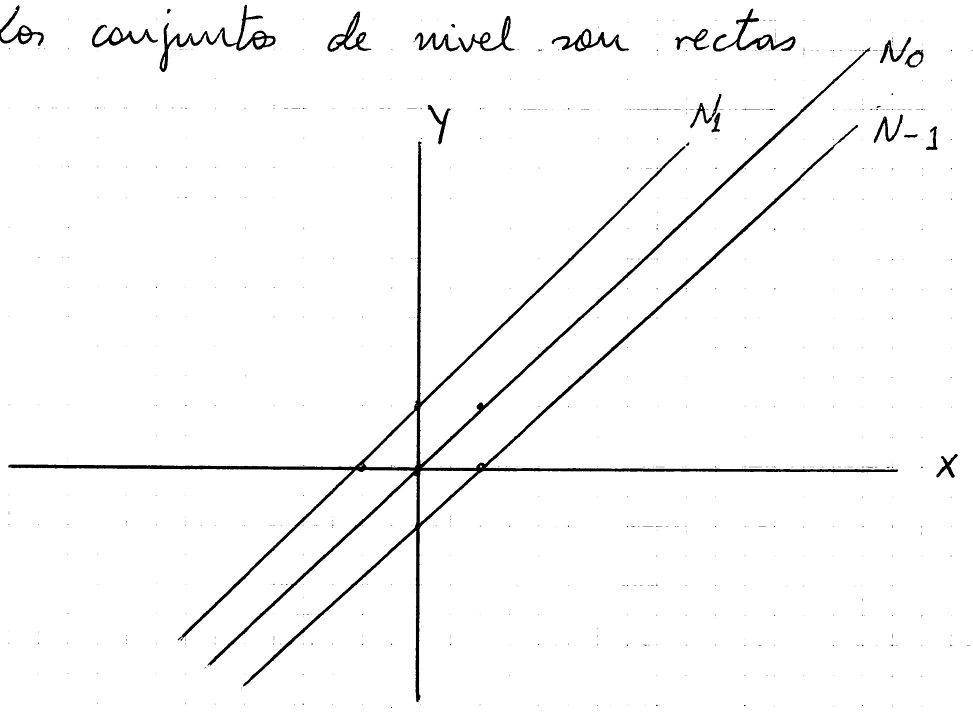
$$N_k = \{(x_1, x_2, \dots x_n) \in \mathbb{R}^n \,/\, f(x_1, x_2, \dots x_n) = k\}$$

Por ejemplo $f: \mathbb{R}^2 \longrightarrow \mathbb{R}$ definida por $f(x,y) = y - x$

El conjunto de nivel $k$ es:

$$N_k = \{(x,y) \in \mathbb{R}^2 \,/\, y - x = k\}$$

Los conjuntos de nivel son rectas




Si $k=0$ tenemos el conjunto de nivel $N_0$
Si $k=1$ " " " $N_1$
Si $k=-1$ " " " $N_{-1}$

Los conjuntos de nivel son rectas paralelas.

PROPIEDAD DE SUFICIENCIA

Intuitivamente, el que un estimador $T$ de $\theta$ sea suficiente para el parámetro $\theta$ quiere decir que $T$ recoge toda la información que sobre $\theta$ contiene una muestra seleccionada, es decir no se pierde información si se destruye la muestra tras calcular el valor de $T$ en ella. Esto es maravilloso: imagina que tomas muestras de tamaño $n = 1.000.000$ y que tienes encima de la mesa este millón de datos. Si calculamos el valor $T(X_1, X_2, \ldots X_{1000000}) = t$, podemos tirar a la basura ese millón de datos, es decir condensamos un millon de datos en un único número $t$ sin que haya pérdida de información respecto a $\theta$. Vamos a dar ahora una definición más rigurosa de suficiencia

**Definición (suficiencia)**

Sea $X$ una variable aleatoria con función de densidad $f_X(x)$ si se trata de variable continua o con función de probabilidad $P_X(x)$ si se trata de variable discreta. Sea $X_1, X_2, \ldots X_n$

una m.a.s. y consideremos el estadístico $T(X_1, X_2, \ldots X_n)$. Se dice que T es suficiente para estimar $\theta$ si la distribución de la muestra condicionada a que $T=t$ no depende de $\theta$, es decir:

$f_{X_1,X_2,\ldots X_n / T=t}$ no depende de $\theta$ (variable continua)

$P_{X_1,X_2,\ldots X_n / T=t}$ no depende de $\theta$ (variable discreta)

¿Qué quiere decir todo esto?

$N_t$ es la curva de nivel $t$ para la función

$$T: \mathbb{R}^n \longrightarrow \mathbb{R}$$

$$N_t = \{(x_1, x_2, \ldots x_n) / T(x_1, x_2, \ldots x_n) = t\}$$

Pues bien, si restringimos la función de densidad conjunta a $N_t$, esta función no depende de $\theta$, de manera más clara:

$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots x_n)$ no depende de $\theta$ si $(x_1, x_2, \ldots, x_n) \in N_t$ esto si se trata de variable continua. Si la variable fuera discreta

$P_{X_1,X_2,\ldots X_n}(x_1, x_2, \ldots x_n)$ no depende de $\theta$ si $(x_1, x_2, \ldots x_n) \in N_t$.

## Criterio de factorización de Fisher-Neyman

Sea X una v.a. con función de densidad $f_\theta(x)$ (variable continua) o con función de probabilidad $P_\theta(x)$ (variable discreta). Sea $X_1, X_2, \ldots X_n$ una m.a.s. y consideremos el estadístico $T(X_1, X_2, \ldots X_n)$. Entonces:

$$T \text{ es suficiente para } \theta \iff \begin{cases} f_\theta(x_1,x_2,\ldots x_n) = h(x_1,x_2,\ldots x_n) \cdot g_\theta(T(x_1,x_2\ldots x_n)) \ (V.C) \\ \text{o bien} \\ P_\theta(x_1,x_2,\ldots x_n) = h(x_1,x_2,\ldots x_n) \cdot g_\theta(T(x_1,x_2\ldots x_n)) \ (V.D) \end{cases}$$

donde:

- h tan sólo depende de la muestra y es no negativa (en la definición de h no puede entrar $\theta$)
- g depende de $\theta$ pero no depende de la muestra si no

a través del valor que toma $T$ en esa muestra. Es no negativa.

Observación: función inyectiva.

Una función $f: D \subset \mathbb{R} \longrightarrow \mathbb{R}$ se dice inyectiva si elementos distintos tienen imágenes distintas, es decir:

Si $x_1 \neq x_2 \Rightarrow f(x_1) \neq f(x_2)$ o lo que es lo mismo

Si $f(x_1) = f(x_2) \Rightarrow x_1 = x_2$

**Proposición:**
Sea $T$ un estadístico suficiente para $\theta$ y $m$ una función inyectiva. Entonces $m(T)$ es suficiente para $\theta$.

Ejemplo:
Sea $X \rightsquigarrow P(\lambda)$. Obtener un estadístico suficiente para $\lambda$

Solución:
La función de probabilidad de la distribución de Poisson es:

$$P_\lambda(X = x) = \frac{\lambda^x \cdot e^{-\lambda}}{x!}$$

. Por tanto la función de probabilidad conjunta es:

$$P_\lambda(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P_\lambda[X_1 = x_1, X_2 = x_2, \ldots X_n = x_n] =$$

$$= \prod_{i=1}^{n} \frac{\lambda^{x_i} \cdot e^{-\lambda}}{x_i!} = \frac{1}{\prod_{i=1}^{n} x_i!} \cdot e^{\sum x_i} \cdot e^{-n\lambda}$$

Si tomamos $T(x_1, x_2, \ldots x_n) = \frac{\sum_{i=1}^{n} x_i}{n} \Rightarrow \sum_{i=1}^{n} x_i = T(x_1 \ldots x_n) \cdot n$

y entonces podemos poner:

$$P_\lambda(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \frac{1}{\prod_{i=1}^{n} x_i!} \cdot e^{nT(x_1 \ldots x_n) - n\lambda}$$

Si llamamamos

$h(x_1, x_2, \dots x_n) = \dfrac{1}{\prod_{i=1}^{n} x_i!}$ que tan solo depende de

la muestra y $g_\lambda(T(x_1, x_2, \dots x_n)) = e^{nT(x_1, x_2 \dots x_n) - n\lambda}$

que depende de la muestra a través del valor que en esta toma T y de $\lambda$, entonces:

$f_\lambda(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = h(x_1, x_2, \dots x_n) \cdot g_\lambda(T(x_1, x_2, \dots x_n))$

Aplicando el teorema de Fisher-Neyman T es suficiente para $\lambda$ siendo

$$T(X_1, X_2, \dots X_n) = \frac{\sum_{i=1}^{n} X_i}{n}$$

¿Es suficiente el estadístico $T = \sum_{i=1}^{n} X_i$ si $X \rightsquigarrow P(\lambda)$ y $X_1, X_2, \dots X_n$ es una m.a.s?

Sabemos que $T = \dfrac{\sum X_i}{n}$ es suficiente para $\lambda$.

La función $m(x) = nx$ es inyectiva ya que si $m(x_1) = m(x_2) \Rightarrow \cancel{n} x_1 = \cancel{n} x_2 \Rightarrow x_1 = x_2$. Por tanto $m(T) = m\left(\dfrac{\sum X_i}{n}\right) = n \cdot \dfrac{\sum_1^n X_i}{n} = \sum X_i$ es suficiente para $\lambda$.

## Familia de distribución exponencial

La familia de distribución exponencial es un conjunto de distribuciones estadísticas cuya función de probabilidad (v.d) o de densidad (v.c) pueden expresarse como:

$f_\theta(x) = a(\theta) \cdot b(x) \cdot e^{c(\theta) \cdot d(x)}$

o bien:

$P_\theta(x) = a(\theta) \cdot b(x) \cdot e^{c(\theta) \cdot d(x)}$

siendo $a$ y $c$ dos funciones de $\theta$ y $b$, $d$ dos funciones de $x$.

Ejemplo:

Averiguar si la distribución binomial pertenece a la familia exponencial

Solución:

Si $X \rightsquigarrow B(n,p)$ el parámetro es $p$

$$P_p(X=x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}$$

$$P_p(X=x) = \binom{n}{x} \cdot p^x \cdot \frac{(1-p)^n}{(1-p)^x}$$

$$P_p(X=x) = \binom{n}{x} \cdot \left(\frac{p}{1-p}\right)^x \cdot (1-p)^n$$

teniendo en cuenta que $e^{\ln\left(\frac{p}{1-p}\right)^x} = \left(\frac{p}{1-p}\right)^x$

podemos poner:

$P_p(X=x) = \binom{n}{x}(1-p)^n \cdot e^{\ln\left(\frac{p}{1-p}\right)^x}$ y finalmente:

$$P_p(X=x) = \binom{n}{x}(1-p)^n \cdot e^{x \cdot \ln\frac{p}{1-p}}$$

La función de probabilidad de la binomial tiene la forma:

$$P_p(X=x) = a(p) \cdot b(x) \cdot e^{c(p) \cdot d(x)}$$

$a(p) = (1-p)^n$ ; $c(p) = \ln \frac{p}{1-p}$

$b(x) = \binom{n}{x}$ $\quad d(x) = x$

Luego la binomial pertenece a la familia exponencial.

Teorema de Koopman-Darmois (o teorema de la familia exponencial)

Si la variable $X$ se distribuye según una distribución perteneciente a la familia exponencial con función de densidad (o de probabilidad):

$$f_\theta(x) = a(\theta)\cdot b(x)\cdot e^{c(\theta)\cdot d(x)} \quad \left(P_\theta(x) = a(\theta)\cdot b(x)\cdot e^{c(\theta)\cdot d(x)}\right)$$

Si tomamos la m.a.s. $x_1, x_2, \dots x_n$ entonces el estadístico:

$T(x_1, x_2, \dots x_n) = d(x_1) + d(x_2) + \dots + d(x_n) = \sum_{i=1}^{n} d(x_i)$ es suficiente para $\theta$.

Ejemplo:

Obtener un estadístico suficiente para $p$, en la distribución $X \rightsquigarrow B(n, p)$ siendo $n$ conocido.

Solución:

Hemos visto que la binomial pertenece a la familia exponencial y que su función de probabilidad vale:

$P_p(X=x) = \binom{n}{x}(1-p)^x \cdot e^{x \ln \frac{p}{1-p}}$. Aplicando el teorema de la familia exponencial

$T(x_1, x_2, \dots x_n) = \sum_{i=1}^{n} x_i$ es un estadístico suficiente para $p$.

Observación:

Si el estadístico $\theta$ es $K$-dimensional, es decir $\theta = (\theta_1, \theta_2, \dots \theta_K)$ y la función de densidad de la v.a. $X$ es:

$$f_\theta(x) = a(\theta)\cdot b(x)\cdot e^{\sum_{i=1}^{k} c_i(\theta)\cdot d_i(x)}$$

Entonces el estadístico suficiente para $\theta$ es:

$$T(x_1, x_2, \dots, x_n) = \left(\sum_{i=1}^{n} d_1(x_i), \sum_{i=1}^{n} d_2(x_i), \dots, \sum_{i=1}^{n} d_k(x_i)\right)$$

$\forall\, x_1, x_2, \dots x_n$ m.a.s, así sería el teorema de la familia exponencial en este caso.