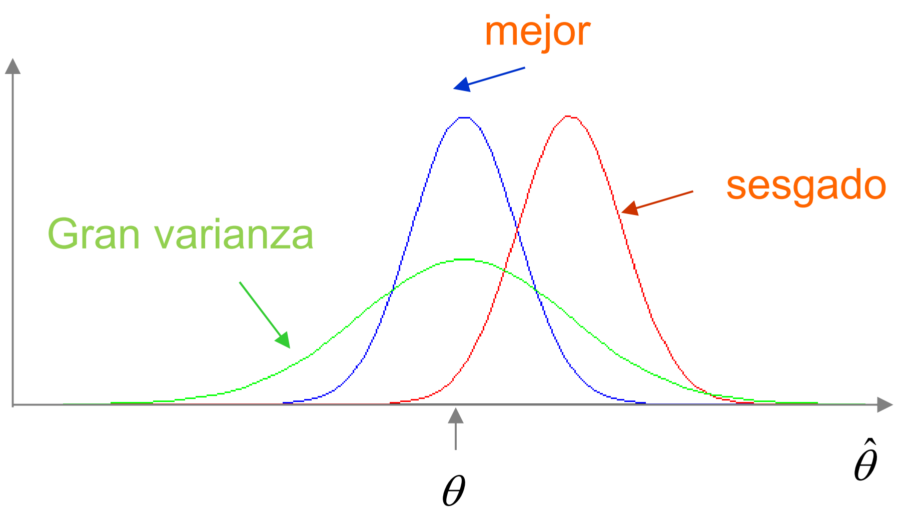
Tema 7 Estimación de parámetros

1. Introducción.
2. Propiedades de los estimadores
3. Información y "*likelihood*"
4. Límite de mínima varianza. Desigualdad de Raó-Cramér
5. Eficiencia de un estimador
6. Ejemplo



1.- Introducción

¿Cuál es la diferencia entre Probabilidad y Estadística?

$T(\vec{x};\vec{p})$ **Teoría** $\xrightarrow{\textbf{Probabilidad / Calculus}}$ **Datos** $D(\vec{x})$

Dada una distribución de probabilidad con sus parámetros conocidos ¿qué podemos decir sobre los datos?

$T(\vec{x};\vec{p})$ **Teoría** $\xleftarrow{\textbf{Estadística / Inferencia estadística}}$ **Datos** $D(\vec{x})$

Realizado un experimento y obtenidos unos datos ¿qué podemos decir sobre la teoría? ¿Qué podemos decir sobre la pdf y sus parámetros a partir de los datos?

Estimación de parámetros.- Procedimiento para obtener $\vec{p}$ a partir de $D(\vec{x})$

1. Introducción. Definiciones

Supongamos que tenemos un conjunto de n medidas x_i como resultado de un experimento y queremos medir un parámetro cuyo valor verdadero pero desconocido es θ.

Estadístico

Un ***estadístico*** es cualquier función de las observaciones de una muestra, $\phi(x_i)$, y que no depende de ninguna de las características desconocidas de la población de origen (***parent pdf***).

Estimador

Un ***estimador*** es un procedimiento, método, o estadístico que proporciona un valor del parámetro o propiedad de la distribución, en función de las observaciones de la muestra.

A la estimación obtenida en general se le llama con un acento circunflejo, $\hat{\theta}$, es decir, se dice entonces que $\hat{\theta}$ es el estimador del parámetro θ

Estimación de parámetros

Dado un conjunto de datos, ***estimar un parámetro*** consiste en determinar un valor tan cercano como sea posible al valor verdadero

Estimación de intervalos

Dado un conjunto de datos, ***estimar un intervalo*** consiste en determinar el rango de valores que con mayor probabilidad contiene al valor verdadero

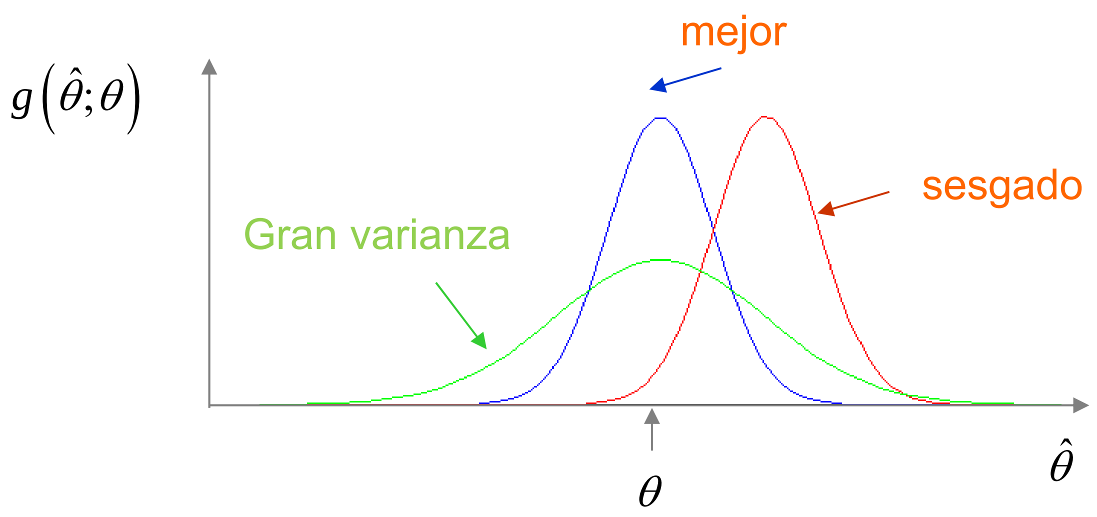
2.- Propiedades de los estimadores

Puesto que un estadístico se calcula a partir de variables aleatorias, cualquier estimador basado en un estadístico es también una variable aleatoria y se distribuirá según una distribución de probabilidad i.e., si repetimos el experimento, las estimaciones obtenidas seguirán una pdf:

$$g\left(\hat{\theta};\theta\right)$$

La calidad o mérito de un estimador se mide por las propiedades de las distribuciones a que dan lugar.

En general, un buen estimador debe producir estimaciones que no se desvíen sistemáticamente (***unbiased***) del valor verdadero, su exactitud aumente con el número de medidas (eficiente) y tenga la mínima varianza.



2.- Propiedades de los estimadores

Sesgo (*Bias*)

Para un estimador que utiliza un conjunto de n observaciones, se define el sesgo o *bias* b_n como la diferencia entre su valor esperado y el valor verdadero del parámetro

$$b_n\left(\hat{\theta}\right) = E\left[\hat{\theta}\right] - \theta = E\left[\hat{\theta} - \theta\right]$$

Se dice que un estimador es no sesgado (unbiased) si para todo n y todo θ :

$$b_n\left(\hat{\theta}\right) = 0 \quad \Rightarrow \quad E\left[\hat{\theta}\right] = \theta$$

Ejemplo

Media.- En general la media muestral es un estimador no sesgado de la media verdadera de la distribución

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \Rightarrow \quad E[\hat{\mu}] = E[\overline{x}] = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[x_i] = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu$$

2.- Propiedades de los estimadores

Varianza.- Si conocemos la media verdadera μ

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

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

Se dice que S^2 es un estimador no sesgado de la varianza σ^2

Varianza.- Si no conocemos la media verdadera y la estimamos a partir del valor medio $\hat{\mu} = \overline{x}$

$$S_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2 = \overline{x^2} - \overline{x}^2$$

$$E\left[\sum_{i=1}^{n} x_i^2\right] = nE\left[x^2\right] = n\left(\sigma^2+\mu^2\right)$$

$$E\left[\left(\sum_{i=1}^{n} x_i\right)^2\right] = V\left[\left(\sum_{i=1}^{n} x_i\right)\right] + \left(E\left[\sum_{i=1}^{n} x_i\right]\right)^2 = n\sigma^2 + n^2\mu^2$$

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

Se dice que S_x^2 es un estimador sesgado de la varianza σ^2

2.- Propiedades de los estimadores

Varianza: La anchura de los datos alrededor de la media muestral siempre es menor que la anchura alrededor de la media verdadera ➡ S_x^2 siempre subestima la varianza

Un estimador no sesgado de la varianza vendrá dado por tanto por:

$$\boxed{s^2 = \frac{n}{n-1} S_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(x_i - \overline{x} \right)^2}$$

Además, si la pdf es gaussiana o $n \to \infty$ (CLT) la variable:

$$\frac{(n-1)\, s^2}{\sigma^2} \quad \text{se distribuye como} \quad \chi^2(n-1)$$

(Tema 6 Distribuciones de Probabilidad (II), pag. 12)

Si las variables son gaussianas otra forma de verlo es la siguiente:

$$z_i = \frac{x_i - \overline{x}}{\sigma} \quad \Rightarrow \quad \sum_{i=1}^{n} z_i^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} \left(x_i - \overline{x} \right)^2$$

Se distribuye como una variable $\chi^2(n-1)$

$$E\left[s^2\right] = E\left[\frac{\sigma^2}{n-1} \sum_{i=1}^{n} z_i^2\right] = E\left[\frac{\sigma^2}{n-1} \chi^2(n-1)\right] = \frac{\sigma^2}{n-1} E\left[\chi^2(n-1)\right] = \sigma^2$$

2.- Propiedades de los estimadores

Consistencia

Un estimador se dice que es consistente cuando converge al valor verdadero a medida que aumenta el número de observaciones.

Definición

Un estimador $\hat{\theta}$ de un parámetro θ es consistente si para todo $\varepsilon > 0$ (no importa lo pequeño que sea)

$$\lim_{n\to\infty} P\left(\left|\hat{\theta} - \theta\right| \geq \varepsilon\right) = 0$$

Si el estimador $\hat{\theta}$ es el promedio de datos que se distribuyen de acuerdo con una pdf para la cual se aplica el teorema del límite central (CLT), entonces $\hat{\theta}$ es un estimador consistente, puesto que la anchura de la pdf tiende a:

$$N\left(x;\mu,\frac{\sigma^2}{n}\right) \Rightarrow \frac{\sigma^2}{n} \to 0 \qquad n \to \infty$$

2.- Propiedades de los estimadores

Varianza

Experimentos repetidos dan lugar a valores diferentes del estimador distribuidos con una determinada varianza bien alrededor del valor verdadero, o de un valor erróneo si el estimador es sesgado.

Un estimador se dice que es eficiente si tiene la varianza más pequeña posible.

Ejemplo Varianza de la media muestral

$$V[\overline{x}] = V\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} V[x_i] = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n} \quad \Rightarrow \quad \boxed{V[\overline{x}] = \frac{\sigma^2}{n}}$$

Ejemplo Varianza de la varianza muestral

$$V\left[\sum_{i=1}^{n} z_i^2\right] = V\left[\chi^2(n-1)\right] = 2(n-1)$$

$$V\left[s^2\right] = V\left[\frac{\sigma^2}{n-1}\sum_{i=1}^{n}\frac{(x_i-\overline{x})^2}{\sigma^2}\right] = \left[\frac{\sigma^2}{n-1}\right]^2 V\left[\sum_{i=1}^{n}\frac{(x_i-\overline{x})^2}{\sigma^2}\right] = \left[\frac{\sigma^2}{n-1}\right]^2 V\left[\sum_{i=1}^{n} z_i^2\right] = \left[\frac{\sigma^2}{n-1}\right]^2 2(n-1) = \frac{2\left(\sigma^2\right)^2}{n-1} \quad \Rightarrow \quad \boxed{V\left[s^2\right] = \frac{2\left(\sigma^2\right)^2}{n-1}}$$

Si no conocemos σ^2

$$\boxed{\hat{V}[\overline{x}] = \frac{s^2}{n}} \quad \Rightarrow \quad \boxed{\hat{\mu} = \overline{x} \pm \sqrt{V[\overline{x}]} = \overline{x} \pm \frac{s}{\sqrt{n}}}$$

$$\boxed{\hat{V}\left[s^2\right] = \frac{2\left(s^2\right)^2}{n-1}} \quad \Rightarrow \quad \boxed{\widehat{\sigma^2} = s^2 \pm \sqrt{V\left[s^2\right]} = s^2 \pm \sqrt{\frac{2}{n-1}}\, s^2}$$

3.- Información y "likelihood"

Likelihood

Supongamos una variable aleatoria X con pdf $f\left(x;\theta\right)$.

Sea Ω_θ el espacio muestral o conjunto de posibles valores de X (puede depender de θ)

Tanto X como θ pueden ser un conjunto de valores $\overline{X}$, $\overline{\theta}$ no necesariamente de la misma dimensión

Consideremos un conjunto de n observaciones independientes x_i . Puesto que las variables son independientes, la pdf conjunta viene dada por:

$$L\left(\overline{x};\overline{\theta}\right)=L\left(x_i,x_2,\ldots,x_n;\overline{\theta}\right)=\prod_{i=1}^{n} f\left(x_i;\overline{\theta}\right)$$

Función de verosimilitud o *likelihood*

Función que depende tanto de las medidas x_i como de los parámetros $\overline{\theta}$

Una vez realizadas las medidas, estas son fijas y la función solo depende de los parámetros.

También se define su logaritmo como:

$$\ell\equiv\ln L\left(x_i,x_2,\ldots,x_n;\overline{\theta}\right)=\ln\left(\prod_{i=1}^{n} f\left(x_i;\overline{\theta}\right)\right)=\sum_{i=1}^{n}\ln f\left(x_i;\overline{\theta}\right)$$

3.- Información y "*likelihood*"

Información

Se define la información (de R. A. Fischer) dada sobre un parámetro θ por la observación de la variable aleatoria $\overline{x}$:

$$I_{\overline{x}}(\theta) = E\left[\left(\frac{\partial \ln L(\overline{x};\theta)}{\partial\theta}\right)^2\right] = E\left[\left(\frac{\partial \ell}{\partial\theta}\right)^2\right] = \int_{\Omega_\theta}\left(\frac{\partial \ln L(\overline{x};\theta)}{\partial\theta}\right)^2 L(\overline{x};\theta)\,d\overline{x}$$

$$I_{\overline{x}}(\theta) = E\left[\left(\frac{\partial \ln L(\overline{x};\theta)}{\partial\theta}\right)^2\right] = E\left[\left(\frac{\partial \ln \prod f(x_i;\theta)}{\partial\theta}\right)^2\right] = E\left[\left(\frac{\partial}{\partial\theta}\sum \ln f(x_i;\theta)\right)^2\right] = E\left[\left(\sum \frac{\partial}{\partial\theta}\ln f(x_i;\theta)\right)^2\right]$$

Supongamos que:

1.- Ω_θ es independiente de $\overline{\theta}$

2.- $L(\overline{x};\overline{\theta})$ es suficientemente regular para poder intercambiar el orden entre derivadas e integrales

$$E\left[\left(\frac{\partial \ln f(x_i;\theta)}{\partial\theta}\right)\right] = \int\left[\frac{\partial \ln f(x_i;\theta)}{\partial\theta}\right] f(x_i;\theta)\,dx_i = \int \frac{1}{f(x_i;\theta)}\left[\frac{\partial f(x_i;\theta)}{\partial\theta}\right] f(x_i;\theta)\,dx_i = \int \frac{\partial f(x_i;\theta)}{\partial\theta}dx_i = \frac{\partial}{\partial\theta}\int f(x_i;\theta)\,dx_i = \frac{\partial}{\partial\theta}1 = 0$$

$$E\left[\left(\frac{\partial \ln L(\overline{x};\theta)}{\partial\theta}\right)\right] = E\left[\left(\frac{\partial \ln \prod f(x_i;\theta)}{\partial\theta}\right)\right] = E\left[\left(\frac{\partial}{\partial\theta}\sum \ln f(x_i;\theta)\right)\right] = E\left[\left(\sum \frac{\partial \ln f(x_i;\theta)}{\partial\theta}\right)\right] = \sum E\left[\frac{\partial \ln f(x_i;\theta)}{\partial\theta}\right] = 0$$

El valor esperado de la derivada del logaritmo de la likelihood es nulo

$$E\left[\left(\frac{\partial \ln L(\overline{x};\theta)}{\partial\theta}\right)\right] = 0$$

4.- Límite de mínima varianza

Se puede demostrar que existe un límite inferior para la varianza de un estimador bajo ciertas condiciones generales:

Desigualdad de Rao-Cramér

Supongamos que tenemos un estimador $\hat{\theta}$ del parámetro θ con un sesgo dado por, $b_n\left(\hat{\theta}\right)=E\left[\hat{\theta}\right]-\theta$ varianza finita $V\left[\hat{\theta}\right]$ y que el rango de X no depende de θ. Entonces:

$$E\left[\hat{\theta}\frac{\partial}{\partial\theta}\ln L\left(\overline{x};\theta\right)\right]=\int\cdots\int\hat{\theta}\left[\frac{\partial}{\partial\theta}\ln L\left(\overline{x};\theta\right)\right]L\left(\overline{x};\theta\right)dx_1dx_2\ldots dx_n=$$

$$=\int\cdots\int\hat{\theta}\left[\frac{1}{L\left(\overline{x};\theta\right)}\frac{\partial}{\partial\theta}L\left(\overline{x};\theta\right)\right]L\left(\overline{x};\theta\right)dx_1dx_2\ldots dx_n=$$

$$=\int\cdots\int\hat{\theta}\left[\frac{\partial}{\partial\theta}L\left(\overline{x};\theta\right)\right]dx_1dx_2\ldots dx_n=\int\cdots\int\hat{\theta}\frac{\partial}{\partial\theta}L\left(\overline{x};\theta\right)dx_1dx_2\ldots dx_n=$$

$$=\int\cdots\int\frac{\partial}{\partial\theta}\left[\hat{\theta}L\left(\overline{x};\theta\right)\right]dx_1dx_2\ldots dx_n=\frac{\partial}{\partial\theta}\int\cdots\int\hat{\theta}L\left(\overline{x};\theta\right)dx_1dx_2\ldots dx_n=$$

$$=\frac{\partial}{\partial\theta}E\left[\hat{\theta}\right]=\frac{\partial}{\partial\theta}\left(\theta+b_n\left(\hat{\theta}\right)\right)=1+\frac{\partial}{\partial\theta}b_n\left(\hat{\theta}\right)$$

$b_n\left(\hat{\theta}\right)=E\left[\hat{\theta}\right]-\theta$

4.- Límite de mínima varianza

El resultado anterior se puede expresar como: $E\left[\hat{\theta}\frac{\partial}{\partial\theta}\ln L\left(\overline{x};\theta\right)\right]=1+\frac{\partial}{\partial\theta}b_n\left(\hat{\theta}\right)$

Recordemos además que: $I_{\overline{x}}\left(\theta\right)=E\left[\left(\frac{\partial\ln L\left(\overline{x};\theta\right)}{\partial\theta}\right)^2\right]$ y que: $E\left[\left(\frac{\partial\ln L\left(\overline{x};\theta\right)}{\partial\theta}\right)\right]=0$

Y por tanto: $V\left[\frac{\partial}{\partial\theta}\ln L\left(\overline{x};\theta\right)\right]=E\left[\left(\frac{\partial}{\partial\theta}\ln L\left(\overline{x};\theta\right)\right)^2\right]-\left(E\left[\frac{\partial}{\partial\theta}\ln L\left(\overline{x};\theta\right)\right]\right)^2=E\left[\left(\frac{\partial}{\partial\theta}\ln L\left(\overline{x};\theta\right)\right)^2\right]=I_{\overline{x}}\left(\theta\right)$

Tanto $\hat{\theta}$ como $\frac{\partial}{\partial\theta}\ln L\left(\overline{x};\theta\right)$ son variables aleatorias y podemos hablar de su covarianza

$$\operatorname{cov}\left[\hat{\theta}\left(\overline{x}\right),\frac{\partial}{\partial\theta}\ln L\left(\overline{x};\theta\right)\right]=E\left[\hat{\theta}\left(\overline{x}\right)\frac{\partial}{\partial\theta}\ln L\left(\overline{x};\theta\right)\right]-E\left[\frac{\partial}{\partial\theta}\ln L\left(\overline{x};\theta\right)\right]E\left[\hat{\theta}\left(\overline{x}\right)\right]=1+\frac{\partial}{\partial\theta}b_n\left(\hat{\theta}\right)$$

El coeficiente de correlación vendrá dado por:

$$\rho^2=\frac{\operatorname{cov}\left[\hat{\theta}\left(\overline{x}\right),\frac{\partial}{\partial\theta}\ln L\left(\overline{x};\theta\right)\right]^2}{V\left[\frac{\partial}{\partial\theta}\ln L\left(\overline{x};\theta\right)\right]V\left[\hat{\theta}\right]}=\frac{\left(1+\frac{\partial}{\partial\theta}b_n\left(\hat{\theta}\right)\right)^2}{I_{\overline{x}}\left(\theta\right)V\left[\hat{\theta}\right]}\leq 1$$

Desigualdad de Rao-Cramér

$$\sigma^2\left(\hat{\theta}\right)=V\left[\hat{\theta}\right]\geq\frac{\left(1+\frac{\partial}{\partial\theta}b_n\left(\hat{\theta}\right)\right)^2}{I_{\overline{x}}\left(\theta\right)}$$

Para un conjunto de datos dado y una cierta cantidad de información $I\left(\theta\right)$ sobre un parámetro θ nunca podemos encontrar un estimador con una varianza más baja que el límite de Rao-Cramér

4.- Límite de mínima varianza

La información se puede expresar también como: $I_{\overline{x}}(\theta)=E\left[\left(\frac{\partial \ln L(\overline{x};\theta)}{\partial\theta}\right)^2\right]=-E\left[\frac{\partial^2 \ln L(\overline{x};\theta)}{\partial\theta^2}\right]$

Demostración

$$E\left[\left(\frac{\partial \ln L(\overline{x};\theta)}{\partial\theta}\right)^2\right]=E\left[\left(\frac{\partial \ln \prod_{i=1}^{n} f(x_i;\theta)}{\partial\theta}\right)^2\right]=E\left[\left(\frac{\partial}{\partial\theta}\sum_{i=1}^{n}\ln f(x_i;\theta)\right)^2\right]=E\left[\left(\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln f(x_i;\theta)\right)^2\right]=E\left[\sum_{i=1}^{n}\left(\frac{\partial}{\partial\theta}\ln f(x_i;\theta)\right)^2\right]+E\left[\left(\sum_{i\neq j}^{n}\frac{\partial}{\partial\theta}\ln f(x_i;\theta)\frac{\partial}{\partial\theta}\ln f(x_j;\theta)\right)\right]$$

Donde el segundo término se anula por: $E\left[\left(\frac{\partial \ln f}{\partial\theta}\right)\right]=\int\left[\frac{\partial \ln f}{\partial\theta}\right]fdx_i=\int\frac{1}{f}\left[\frac{\partial f}{\partial\theta}\right]fdx_i=\int\frac{\partial f}{\partial\theta}dx_i=\frac{\partial}{\partial\theta}\int fdx_i=\frac{\partial}{\partial\theta}1=0$

$$E\left[\left(\frac{\partial \ln L(\overline{x};\theta)}{\partial\theta}\right)^2\right]=E\left[\sum_{i=1}^{n}\left(\frac{\partial}{\partial\theta}\ln f(x_i;\theta)\right)^2\right]=\sum_{i=1}^{n}E\left[\left(\frac{\partial}{\partial\theta}\ln f(x_i;\theta)\right)^2\right]=\sum_{i=1}^{n}E\left[\left(\frac{f'(x_i;\theta)}{f(x_i;\theta)}\right)^2\right]= \Rightarrow$$

$$E\left[\left(\frac{f'}{f}\right)\right]=\int_\Omega\frac{f'}{f}fdx=\int_\Omega f'dx=\frac{\partial}{\partial\theta}\int_\Omega fdx=\frac{\partial 1}{\partial\theta}=0$$

Derivando de nuevo:

$$\frac{\partial}{\partial\theta}\int_\Omega\left(\frac{f'}{f}\right)fdx=\int_\Omega\left[\left(\frac{f'}{f}\right)'f+\left(\frac{f'}{f}\right)f'\right]dx=\int_\Omega\left[\left(\frac{f'}{f}\right)'+\left(\frac{f'}{f}\right)^2\right]fdx=E\left[\left(\frac{f'}{f}\right)'\right]+E\left[\left(\frac{f'}{f}\right)^2\right]=0 \quad\Rightarrow\quad E\left[\left(\frac{f'}{f}\right)^2\right]=-E\left[\left(\frac{f'}{f}\right)'\right]$$

$$=\sum_{i=1}^{n}E\left[\left(\frac{f'(x_i;\theta)}{f(x_i;\theta)}\right)^2\right]=-\sum_{i=1}^{n}E\left[\left(\frac{f'(x_i;\theta)}{f(x_i;\theta)}\right)'\right]=-E\left[\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\left(\frac{f'(x_i;\theta)}{f(x_i;\theta)}\right)\right]=-E\left[\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\frac{\partial}{\partial\theta}\ln f(x_i;\theta)\right]=$$

$$\Rightarrow\quad =-E\left[\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\ln f(x_i;\theta)\right]=-E\left[\frac{\partial^2}{\partial\theta^2}\sum_{i=1}^{n}\ln f(x_i;\theta)\right]=-E\left[\frac{\partial^2}{\partial\theta^2}\sum_{i=1}^{n}\ln f(x_i;\theta)\right]=-E\left[\frac{\partial^2 \ln L(\overline{x};\theta)}{\partial\theta^2}\right]$$

5.- Eficiencia de un estimador

Desigualdad de Rao-Cramér

$$\sigma^2\left(\hat{\theta}\right)=V\left[\hat{\theta}\right]\geq\frac{\left(1+\frac{\partial}{\partial\theta}b_n\left(\hat{\theta}\right)\right)^2}{I_{\bar{x}}\left(\theta\right)}$$

$$I_{\bar{x}}\left(\theta\right)=E\left[\left(\frac{\partial\ln L\left(\bar{x};\theta\right)}{\partial\theta}\right)^2\right]=-E\left[\frac{\partial^2\ln L\left(\bar{x};\theta\right)}{\partial\theta^2}\right]$$

El resultado anterior se puede expresar como:

$$\sigma^2\left(\hat{\theta}\right)\geq\frac{\left[1+\frac{\partial}{\partial\theta}b_n\left(\hat{\theta}\right)\right]^2}{E\left[\left(\frac{\partial\ln L}{\partial\theta}\right)^2\right]}=\frac{\left(1+\frac{\partial}{\partial\theta}b_n\left(\hat{\theta}\right)\right)^2}{-E\left[\frac{\partial^2\ln L}{\partial\theta^2}\right]}$$

Existe por tanto un límite inferior en la varianza ***MVB (minimun variance bound)*** del estimador que:

- Disminuye con la información $I_{\bar{x}}\left(\theta\right)$
- Aumenta con el sesgo si depende de θ y si es *unbiased*

$$\sigma^2\left(\hat{\theta}\right)\geq\frac{1}{-E\left[\frac{\partial^2\ln L}{\partial\theta^2}\right]}$$

Eficiencia

La eficiencia de un estimador se define por comparación con su ***MVB*** :

$$\varepsilon\left(\hat{\theta}\right)=\frac{\sigma^2_{\min}\left(\hat{\theta}\right)}{\sigma^2\left(\hat{\theta}\right)}\leq 1$$

Se dice que un estimador es eficiente si su varianza es igual a la mínima, es decir si su eficiencia es: $\varepsilon\left(\hat{\theta}\right)=1$

6.- Ejemplo estimador eficiente

Ejemplo **Distribución exponencial**

$$f(x,\mu)=\frac{1}{\mu}e^{-x/\mu}$$

Consideremos n medidas independientes de una distribución exponencial, queremos estimar el valor de μ.

$$\ln L(\overline{x};\mu)=\ln\prod_{i=1}^{n} f(x_i;\mu)=\sum_{i=1}^{n}\ln f(x_i;\mu)=\sum_{i=1}^{n}-\ln\mu-\frac{x_i}{\mu}$$

$$\frac{\partial \ln L(\overline{x};\mu)}{\partial\mu}=\frac{\partial}{\partial\mu}\sum_{i=1}^{n}\left(-\ln\mu-\frac{x_i}{\mu}\right)=\sum_{i=1}^{n}\left(-\frac{1}{\mu}+\frac{x_i}{\mu^2}\right)$$

$$\frac{\partial^2 \ln L(\overline{x};\theta)}{\partial\mu^2}=\frac{\partial}{\partial\mu}\sum_{i=1}^{n}\left(-\frac{1}{\mu}+\frac{x_i}{\mu^2}\right)=\sum_{i=1}^{n}\left(\frac{1}{\mu^2}-\frac{2x_i}{\mu^3}\right)$$

$$I(\mu)=-E\left[\frac{\partial^2 \ln L(\overline{x};\theta)}{\partial\theta^2}\right]=-E\left[\sum_{i=1}^{n}\left(\frac{1}{\mu^2}-\frac{2x_i}{\mu^3}\right)\right]=-\sum_{i=1}^{n}\left(\frac{1}{\mu^2}-\frac{2}{\mu^2}\right)=\frac{n}{\mu^2}$$

Si $\hat{\mu}$ es un estimador sin sesgo su varianza mínima es:

$$\boxed{\sigma^2_{\min}=\frac{\mu^2}{n}}$$

Si utilizamos como estimador la media de la muestra: $\hat{\mu}=\overline{x}$ su varianza coincide con la varianza mínima, luego la media de la muestra es un estimador eficiente

$$V(\overline{x})=\frac{1}{n}V(x)=\frac{1}{n}\left(E\left[x^2\right]-\mu^2\right)=\frac{2\mu^2}{n}-\frac{\mu^2}{n}=\frac{\mu^2}{n}$$