Definición 39: Un estimador W^* es el mejor estimador insesgado de $Z(\theta)$ si satisface

a) $E_\theta(W^*) = Z(\theta)$ para todo θ

b) Para otro estimador W con $E_\theta(W) = Z(\theta)$ se cumple que $Var_\theta(W^*) \leq Var_\theta(W)$ para todo θ

obs Δ W^* también se le llama estimador insesgado de mínima varianza uniformemente de $\tau(\theta)$ (UMVUE)

Teorema 40 (Cramér - Rao) Sea $X_1, X_2, \ldots, X_n$ una muestra con densidad $f(x|\theta)$ y sea $W(x)$ un estimador que es diferenciable como función de θ.

Supongamos que la función de densidad conjunta $f(x|\theta) = f(x_1, \ldots, x_n|\theta)$ satisface

$$(**) \quad \frac{d}{d\theta}\int \cdots \int h(x) f(x|\theta)\, dx_1 \ldots dx_n = \int \cdots \int h(x) \frac{\partial}{\partial\theta} f(x_1, \ldots, x_n|\theta)\, dx_1 \ldots dx_n$$

para toda función $h(x)$ con $E_\theta |h(X)| < \infty$ entonces

$$\mathrm{Var}_\theta (W(X)) \geq \frac{\left(\frac{d}{d\theta} E_\theta W(X)\right)^2}{E_\theta\left(\left(\frac{\partial}{\partial\theta} \log f(X|\theta)\right)^2\right)}$$

Demostración: Se usará la desigualdad de Cauchy-Schwarz

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

o sea $$\mathrm{Var}(X) \geq \frac{(\mathrm{Cov}(X,Y))^2}{\mathrm{Var}(Y)}$$

Consideremos $X = W(X)$ y $Y = \frac{\partial}{\partial\theta} \log f(X|\theta)$

$$E_\theta\left(\frac{\partial}{\partial\theta} \log f(X|\theta)\right) = E_\theta\left[\frac{\frac{\partial}{\partial\theta} f(X|\theta)}{\ldots}\right] =$$

$$\underset{\theta}{E}\left(\frac{\partial}{\partial\theta}\log f(x|\theta)\right)=E_\theta\left[\frac{\frac{\partial}{\partial\theta}f(x|\theta)}{f(x|\theta)}\right]=$$

$$=\int\cdots\int\frac{\frac{\partial}{\partial\theta}f(x|\theta)}{f(x|\theta)}f(x|\theta)\,dx_1\cdots dx_n=$$

$$=\int\cdots\int\frac{\partial}{\partial\theta}f(x|\theta)\,dx_1\cdots dx_n=\frac{d}{d\theta}\int\cdots\int f(x|\theta)\,dx_1\cdots dx_n$$

$$=\frac{d}{d\theta}E_\theta(1)=0$$

Entonces

$$Cov\left(W(x),\frac{\partial}{\partial\theta}\log f(x|\theta)\right)=\underset{\theta}{E}\left(W(x)\frac{\partial}{\partial\theta}\log f(x|\theta)\right)-\underset{\theta}{E}(W(x))\underbrace{\underset{\theta}{E}\left(\frac{\partial}{\partial\theta}\log f(x|\theta)\right)}_{0}=$$

$$=\underset{\theta}{E}\left(W(x)\frac{\partial}{\partial\theta}\log f(x|\theta)\right)=$$

$$=E_\theta\left(W(x)\frac{\frac{\partial}{\partial\theta}f(x|\theta)}{f(x|\theta)}\right)=\int\cdots\int W(x)\frac{\frac{\partial}{\partial\theta}f(x|\theta)}{f(x|\theta)}f(x|\theta)\,dx_1\cdots dx_n$$

$$=\int\cdots\int W(x)\frac{\partial}{\partial\theta}f(x|\theta)\,dx_1\cdots dx_n=$$

$$=\frac{d}{d\theta}\int\cdots\int W(x)f(x|\theta)\,dx_1,\ldots,dx_n=\frac{d}{d\theta}E_\theta(W(x))$$

Además

$$\mathrm{Var}_\theta\left(\frac{\partial}{\partial\theta}\log f(x/\theta)\right)=E_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(x/\theta)\right)^2\right)-\underbrace{E^2\left(\frac{\partial}{\partial\theta}\log f(x/\theta)\right)}_{\overset{\shortparallel}{0}}$$

$$=E_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(x/\theta)\right)^2\right)$$

luego usando Cauchy-Schwarz

$$\mathrm{Var}_\theta(W(X))\geq\frac{\left(\frac{d}{d\theta}E_\theta W(x)\right)^2}{E_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(x/\theta)\right)^2\right)}$$

□

Corolario 41: Sean $X_1,\ldots,X_n$ iid con densidad $f(x|\theta)$ y sea $W(X)$ un estimador con $E_\theta(W(X))$ es una función diferenciable de θ. Si la función de densidad conjunta $f(x|\theta)=\prod_{i=1}^{n}f(x_i|\theta)$ satisface (**)

entonces $$\mathrm{Var}_\theta(W(x))\geq\frac{\left(\frac{d}{d\theta}E_\theta(W(x))\right)^2}{n\,E_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(x/\theta)\right)^2\right)}$$

↳ la función de densidad de X_i

Demostración: Veamos que

$$E_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(X/\theta)\right)^2\right] = E_\theta\left(\left(\frac{\partial}{\partial\theta}\log\prod_{i=1}^n f(X_i|\theta)\right)^2\right) =$$

$$= E_\theta\left(\left(\sum_{i=1}^n \frac{\partial}{\partial\theta}\log f(X_i|\theta)\right)^2\right) =$$

$$= \sum_{i=1}^n E_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(X_i|\theta)\right)^2\right) +$$

$$+ \sum_{i\neq j} E_\theta\left(\frac{\partial}{\partial\theta}\log f(X_i|\theta)\,\frac{\partial}{\partial\theta}\log f(X_j|\theta)\right)$$

Para $i\neq j$ tenemos

$$E_\theta\left(\frac{\partial}{\partial\theta}\log f(X_i|\theta)\,\frac{\partial}{\partial\theta}\log f(X_j|\theta)\right) =$$

independencia

$$= E_\theta\left(\frac{\partial}{\partial\theta}\log f(X_i|\theta)\right) E_\theta\left(\frac{\partial}{\partial\theta}\log f(X_j|\theta)\right) = 0$$

de donde

$$\sum_{i=1}^n E_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(X_i|\theta)\right)^2\right) = n\,E_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2\right)$$

con lo que se demuestra el corolario $\square$

Es claro a partir del teorema de Cramér-Rao que se encontramos un estimador insesgado que cumple la cota de Cramer-Rao este es el mejor estimador insesgado.

El teorema de Cramer-Rao lo demostramos para variables aleatorias continuas pero también es cierto para variables aleatorias discretas, realizando los cambios obvios, es decir si $f(x|\theta)$ es una función de masa de probabilidad se debe poder intercambiar diferenciación con sumación finita estamos considerando derivabilidad respecto a θ

Definición 42: Llamamos información de Fisher de la muestra a la cantidad

$$E_\theta\left(\left(\frac{\partial}{\partial\theta}\log\prod_{i=1}^{n} f(x_i|\theta)\right)^2\right)$$

Si W es el mejor estimador insesgado de θ tenemos que $\left(\frac{\partial}{\partial\theta}E_\theta(W(X))\right)^2=\left(\frac{\partial\theta}{\partial\theta}\right)^2=1$

y se alcanza la cota de Cramér-Rao

luego $\operatorname{Var}(W(X))=\dfrac{1}{E_\theta\left[\left(\frac{\partial}{\partial\theta}\log\prod_{i=1}^{n} f(x_i|\theta)\right)^2\right]}$

o sea la información de Fisher es el inverso de la varianza de $W(X)$

Si tenemos una función $\tau(\theta)$ y un estimador tal que $E_\theta(W(X)) = \tau(\theta)$ vemos que la cota inferior de la varianza del estimador W sólo depende de $\tau(\theta)$, de $f(x/\theta)$ y es independiente del estimador W que consideramos, de aquí que esta es una cota inferior uniforme de la varianza y cualquier estimador W que cumple $E_\theta(W) = \tau(\theta)$, que alcance la cota inferior de la varianza es el mejor estimador insesgado de $\tau(\theta)$

Propiedad 43. Si $f(x/\theta)$ satisface

$$\frac{d}{d\theta}\left(E_\theta\left(\frac{\partial \log f(X/\theta)}{\partial\theta}\right)\right) = \int \frac{\partial}{\partial\theta}\left[\left(\frac{\partial}{\partial\theta}\log f(x/\theta)\right) f(x/\theta)\right]dx$$

entonces $$E_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(X/\theta)\right)^2\right) = -E_\theta\left(\frac{\partial^2}{\partial\theta^2}\log f(X/\theta)\right)$$

Dem: Se verá en el Práctico

Corolario 44: Sean $X_1, \dots, X_n$ variables aleatorias independientes idénticamente distribuidas de $f(x/\theta)$. La función $f(x/\theta)$ satisface la condición del teorema de Cramér-Rao.
Sea $L(\theta|x) = \prod_{i=1}^{n} f(x_i|\theta)$ la función de verosimilitud. Si $W(X)$ es un estimador insesgado de $\tau(\theta)$, entonces $W(X)$ alcanza la

cota de Cramer-Rao si y solo si

$$a(\theta)\left[W(x)-\tau(\theta)\right]=\frac{\partial}{\partial\theta}\log(L(\theta|x))$$

para alguna función $a(\theta)$

Demostración.

A partir de $[\mathrm{Cov}(X,Y)]^2 \leq \mathrm{Var}(X)\mathrm{Var}(Y)$ podemos escribir

$$\left[\mathrm{Cov}_\theta\left(W(X),\frac{\partial}{\partial\theta}\log\prod_{i=1}^{n} f(x_i|\theta)\right)\right]^2 \leq \mathrm{Var}_\theta(W(X))\,\mathrm{Var}_\theta\left(\frac{\partial}{\partial\theta}\log\prod_{i=1}^{n} f(x_i|\theta)\right)$$

Si se da la igualdad entonces

$$\left[\frac{\mathrm{Cov}_\theta\left(W(X),\frac{\partial}{\partial\theta}\log\prod_{i=1}^{n} f(x_i|\theta)\right)}{\sqrt{\mathrm{Var}_\theta(W(X))}\sqrt{\mathrm{Var}_\theta\left(\frac{\partial}{\partial\theta}\log\prod_{i=1}^{n} f(x_i|\theta)\right)}}\right]^2=1$$

o sea $\left|\rho_{W(X),\frac{\partial}{\partial\theta}\log\prod_{i=1}^{n} f(x_i|\theta)}\right|=1$ o sea

$$\frac{\partial}{\partial\theta}\log\prod_{i=1}^{n} f(x_i|\theta)=a(\theta)W(X)+b(\theta)$$

Tomando esperanzas

$$\underbrace{E_\theta\left[\frac{\partial}{\partial\theta}\log\prod_{i=1}^{n} f(x_i|\theta)\right]}_{0} = a(\theta)\underbrace{E_\theta(W(X))}_{Z(\theta)} + b(\theta)$$

o sea $\quad b(\theta) = -a(\theta)\,Z(\theta)$

de donde

$$\frac{\partial}{\partial\theta}\left(\log\prod_{i=1}^{n} f(x_i|\theta)\right) = \underbrace{a(\theta)\,W(X) - a(\theta)\,Z(\theta)}_{a(\theta)\left(W(X)-Z(\theta)\right)}$$

$\square$