La desigualdad de Cramer-Rao

Sea $X \sim f(x,\theta)$ con $x \in \mathcal{X}$ (espacio muestral que podemos pensar incluido en $\mathbb{R}^k$)

y $\theta \in \Theta$ (espacio de parámetros que podemos pensar incluido en $\mathbb{R}^m$)

f es una densidad de probabilidad (para fijar ideas podemos pensar que X es absolutamente continua, aunque no es imprescindible)

tendremos $\int_{\mathcal{X}} f(x,\theta)\, dx = 1$

Sea $X_1, \dots, X_n$ las variables aleatorias muestrales correspondientes a una muestra aleatoria simple de tamaño "n" i.e. $X_1 \dots X_n$ i.i.d X

La densidad conjunta de $X_1 \dots X_n$ será pues:

$$\tilde{f}(x_1,\dots,x_n,\theta) = f(x_1,\theta)\cdot\ldots\cdot f(x_n,\theta)$$

Supongamos además que el soporte de la densidad no depende de θ.

es la adherencia del subconjunto del dominio donde la densidad es positiva

Sea $\mathcal{U} = \mathcal{U}(X_1 \dots X_n)$ un estimador de θ. Tendremos en notación matricial

$$\mathcal{U} = \begin{pmatrix} \mathcal{U}_1 \\ \vdots \\ \mathcal{U}_m \end{pmatrix} \qquad \theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_m \end{pmatrix}$$

El sesgo será:

$$B(\theta) = E_\theta(\mathcal{U}) - \theta = \begin{pmatrix} E_\theta(\mathcal{U}_1) \\ \vdots \\ E_\theta(\mathcal{U}_m) \end{pmatrix} - \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_m \end{pmatrix}$$

Calculemos:

$$\frac{\partial}{\partial\theta_\beta} B_\alpha(\theta) = \frac{\partial}{\partial\theta_\beta}\left(E_\theta(\mathcal{U}_\alpha) - \theta_\alpha\right) = \frac{\partial}{\partial\theta_\beta} E_\theta(\mathcal{U}_\alpha) - \delta_{\alpha\beta}$$

donde $\delta_{\alpha\beta} = \begin{cases} 1 & \text{si } \alpha = \beta \\ 0 & \text{si } \alpha \neq \beta \end{cases}$ (deltas de Kronecker)

y
$$\frac{\partial}{\partial\theta_\beta} E_\theta(\mathcal{U}_\alpha) = \frac{\partial}{\partial\theta_\beta} \int_{\mathcal{X}^n} \mathcal{U}_\alpha(x_1 \dots x_n)\, \tilde{f}(x_1,\dots,x_n,\theta)\, dx_1 \dots dx_n$$

$$= \int_{\mathcal{X}^n} \mathcal{U}_\alpha(x_1 \dots x_n) \frac{\partial \ln \tilde{f}(x_1,\dots,x_n,\theta)}{\partial\theta_\beta} \tilde{f}(x_1,\dots,x_n)\, dx_1 \dots dx_n =$$

$$= E_\theta\left(\mathcal{U}_\alpha(X_1,\dots,X_n) \frac{\partial \ln \tilde{f}(X_1,\dots,X_n,\theta)}{\partial\theta_\beta}\right)$$

asumimos las condiciones de regularidad necesarias para garantizar

Por tanto tendremos:

$$\frac{\partial}{\partial \theta_\beta} B_\alpha(\theta) + \delta_{\alpha\beta} = E_\theta\left(u_\alpha(x_1, \dots, x_n) \frac{\partial \ln \tilde{f}(x_1, \dots, x_n, \theta)}{\partial \theta_\beta} \right)$$

pero como

$$E_\theta\left(\frac{\partial \ln \tilde{f}(x_1, \dots, x_n, \theta)}{\partial \theta_\beta} \right) = \int_{\chi^n} \frac{\partial \ln \tilde{f}(x_1, \dots, x_n, \theta)}{\partial \theta_\beta} \tilde{f}(x_1 \dots x_n, \theta)\, dx_1 \cdots dx_n =$$

$$= \int_{\chi^n} \frac{\partial \tilde{f}(x_1 \dots x_n, \theta)}{\partial \theta_\beta}\, dx_1 \cdots dx_n =$$

$$= \frac{\partial}{\partial \theta_\beta} \Big(\underbrace{\int_{\chi^n} \tilde{f}(x_1 \dots x_n, \theta)\, dx_1 \dots dx_n}_{= 1} \Big) = 0$$

tendremos que

$$E_\theta\left(u_\alpha(x_1 \dots x_n) \frac{\partial \ln \tilde{f}(x_1 \dots x_n, \theta)}{\partial \theta_\beta} \right) = E_\theta\left(\left(u_\alpha(x_1 - x_n) - \theta_\alpha - B_\alpha(\theta)\right) \frac{\partial \ln \tilde{f}(x_1 \dots x_n, \theta)}{\partial \theta_\beta} \right)$$

ya que θ_α y $B_\alpha(\theta)$ no dependen de $x_1 \dots x_n$ y pueden "salir fuera" de la esperanza al desarrollar el término derecho.

por tanto, de momento, podemos escribir:

$$\frac{\partial}{\partial \theta_\beta} B_\alpha(\theta) + \delta_{\alpha\beta} = E_\theta\left(\left(u_\alpha(x_1 \dots x_n) - \theta_\alpha - B_\alpha(\theta)\right) \frac{\partial \ln \tilde{f}(x_1 \dots x_n, \theta)}{\partial \theta_\beta} \right) \qquad (1)$$

A partir de esta expresión tendremos:

$$\sum_{\alpha=1}^{m} \left\{ \frac{\partial}{\partial \theta_\alpha} B_\alpha(\theta) + \underbrace{\delta_{\alpha\alpha}}_{=1} \right\} = \sum_{\alpha=1}^{m} E_\theta\left(\left(u_\alpha - \theta_\alpha - B_\alpha(\theta)\right) \frac{\partial \ln \tilde{f}(x_1 \dots x_n, \theta)}{\partial \theta_\alpha} \right)$$

$$\left(\sum_{\alpha=1}^{m} \frac{\partial}{\partial \theta_\alpha} B_\alpha(\theta) \right) + m = E_\theta\left(\sum_{\alpha=1}^{m} \left(u_\alpha - \theta_\alpha - B_\alpha(\theta)\right) \frac{\partial \ln \tilde{f}(x_1 - x_n, \theta)}{\partial \theta_\alpha} \right)$$

Si introducimos el vector columna:

$$\frac{\partial \ln \tilde{f}(x_1 \cdots x_m, \theta)}{\partial \theta} = \begin{pmatrix} \dfrac{\partial \ln \tilde{f}(x_1 \cdots x_m, \theta)}{\partial \theta_1} \\ \vdots \\ \dfrac{\partial \ln \tilde{f}(x_1 \cdots x_m, \theta)}{\partial \theta_m} \end{pmatrix}$$

podemos observar que:

$$\sum_{\alpha=1}^{m} \left(U_\alpha - \theta_\alpha - B_\alpha(\theta) \right) \frac{\partial \ln \tilde{f}(x_1 \cdots x_m, \theta)}{\partial \theta_\alpha} = \underbrace{\left(U - \theta - B(\theta) \right) \cdot}_{\uparrow\ \text{producto escalar}} \frac{\partial \ln \tilde{f}(x_1 - x_m, \theta)}{\partial \theta}$$

por lo que tendremos:

$$m + \sum_{\alpha=1}^{m} \frac{\partial}{\partial \theta_\alpha} B_\alpha(\theta) = E_\theta\left(\left(U - \theta - B(\theta) \right) \cdot \frac{\partial \ln \tilde{f}(x_1 \cdots x_m, \theta)}{\partial \theta} \right) \qquad (2)$$

Recordemos que dados dos vectores $|v \cdot w| \leq \|v\| \, \|w\|$ (Desigualdad de Cauchy-Schwarz)

por tanto:

$$\left| m + \sum_{\alpha=1}^{m} \frac{\partial}{\partial \theta_\alpha} B_\alpha(\theta) \right| = \left| E_\theta\left(\left(U - \theta - B(\theta) \right) \cdot \frac{\partial \ln \tilde{f}(x_1 \cdots x_m, \theta)}{\partial \theta} \right) \right|$$

$$\leq E_\theta\left(\left| \left(U - \theta - B(\theta) \right) \cdot \frac{\partial \ln \tilde{f}(x_1 - x_m, \theta)}{\partial \theta} \right| \right)$$

ya que $|E(X)| \leq E(|X|)$ (siempre que $E(|X|)$ exista)

$$\leq E_\theta\left(\| U - \theta - B(\theta) \| \, \left\| \frac{\partial \ln \tilde{f}(x_1 - x_m, \theta)}{\partial \theta} \right\| \right)$$

Cauchy-Schwarz

$$\leq \sqrt{E_\theta\left(\| U - \theta - B(\theta) \|^2 \right)} \sqrt{E_\theta\left(\left\| \frac{\partial \ln \tilde{f}(x_1 \cdots x_m, \theta)}{\partial \theta} \right\|^2 \right)}$$

Cauchy-Schwarz de nuevo: $|E(XY)| \leq \sqrt{E(X^2)} \sqrt{E(Y^2)}$

la "esperanza del producto" es un producto escalar en un espacio vectorial conveniente de variables aleatorias

Elevando al cuadrado:

$$\left(m + \sum_{\alpha=1}^{m} \frac{\partial}{\partial \theta_\alpha} B_\alpha(\theta) \right)^2 \leq E_\theta\left(\| u - \theta - B(\theta) \|^2 \right) E_\theta\left(\left\| \frac{\partial \ln \tilde{f}(x_1 - x_n, \theta)}{\partial \theta} \right\|^2 \right) \quad (3)$$

pero:

$$\| u - \theta - B(\theta) \|^2 = \| u - \theta \|^2 + \| B(\theta) \|^2 - 2 (u-\theta) \cdot B(\theta)$$

y al tomar esperanza, teniendo en cuenta que $E_\theta(u - \theta) = B(\theta)$

$\| B(\theta) \|^2 = B(\theta) \cdot B(\theta)$ resultará:

$$E_\theta\left(\| u - \theta - B(\theta) \|^2 \right) = E_\theta\left(\| u - \theta \|^2 \right) - \| B(\theta) \|^2 \quad (4)$$

pero $E_\theta\left(\| u - \theta \|^2 \right)$

es el error cuadrático medio del estimador

Introduzcamos ahora la denominada matriz de información de Fisher correspondiente a X. Es una matriz $m \times m$

$$I(\theta) = \left(I_{ij}(\theta) \right)_{m \times m} \text{ definida por } I_{ij}(\theta) = E_\theta\left(\frac{\partial \ln f(x,\theta)}{\partial \theta_i} \frac{\partial \ln f(x,\theta)}{\partial \theta_j} \right)$$

y la matriz de información de Fisher correspondiente a toda la muestra:

$$\tilde{I}(\theta) = \left(\tilde{I}_{ij}(\theta) \right)_{m \times m} \text{ tal que } \tilde{I}_{ij}(\theta) = E_\theta\left(\frac{\partial \ln \tilde{f}(x_1 - x_n, \theta)}{\partial \theta_i} \frac{\partial \ln \tilde{f}(x_1 - x_n, \theta)}{\partial \theta_j} \right)$$

observar que $I(\theta)$ y $\tilde{I}(\theta)$ son simétricas.

Observar que al ser $\tilde{f}(x_1 - x_n, \theta) = f(x_1, \theta) \cdots f(x_n, \theta)$

tendremos que

$$\frac{\partial \ln \tilde{f}(x_1 - x_n, \theta)}{\partial \theta_i} = \sum_{\lambda=1}^{n} \frac{\partial \ln f(x_\lambda, \theta)}{\partial \theta_i}$$

Observación:
No confundir la matriz de información de Fisher $I(\theta)$, $\tilde{I}(\theta)$, I_θ, $\tilde{I}_\theta$ con la matriz identidad I ó I_n

y por tanto:

$$\tilde{I}_{ij}(\theta) = E_\theta\left(\left\{ \sum_{\lambda=1}^{n} \frac{\partial \ln f(x_\lambda, \theta)}{\partial \theta_i} \right\} \left\{ \sum_{\gamma=1}^{n} \frac{\partial \ln f(x_\gamma, \theta)}{\partial \theta_j} \right\} \right) = \sum_{\lambda=1}^{n} \sum_{\gamma=1}^{n} E_\theta\left(\frac{\partial \ln f(x_\lambda, \theta)}{\partial \theta_i} \frac{\partial \ln f(x_\gamma, \theta)}{\partial \theta_j} \right)$$

pero $\dfrac{\partial \ln f(X_\lambda,\theta)}{\partial\theta_i}$ y $\dfrac{\partial \ln f(X_\gamma,\theta)}{\partial\theta_j}$ son independientes si $\lambda\neq\gamma$

por tanto, teniendo en cuenta que

$$E_\theta\left(\frac{\partial \ln f(X,\theta)}{\partial\theta_i}\right)=\int_{\mathcal{X}}\frac{\partial \ln f(x,\theta)}{\partial\theta_i}f(x,\theta)\,dx=\int_{\mathcal{X}}\frac{\partial f(x,\theta)}{\partial\theta_i}dx=$$

$$=\frac{\partial}{\partial\theta_i}\underbrace{\int_{\mathcal{X}} f(x,\theta)\,dx}_{1}=0$$

resultará:

$$\boxed{\tilde{I}_{ij}(\theta)=}\sum_{\lambda=1}^{n}E_\theta\left(\frac{\partial \ln f(X_\lambda,\theta)}{\partial\theta_i}\frac{\partial \ln f(X_\lambda,\theta)}{\partial\theta_j}\right)=$$

$$=\sum_{\lambda=1}^{n}I_{ij}(\theta)\ \boxed{=n\,I_{ij}(\theta)\qquad(5)}$$

Por tanto:

$$E_\theta\left(\left\|\frac{\partial \ln\tilde{f}(X_1\ldots X_n,\theta)}{\partial\theta}\right\|^2\right)=E_\theta\left(\sum_{\alpha=1}^{m}\left(\frac{\partial \ln\tilde{f}(X_1\ldots X_n,\theta)}{\partial\theta_\alpha}\right)^2\right)=$$

$$=\sum_{\alpha=1}^{m}E_\theta\left(\left(\frac{\partial \ln\tilde{f}(X_1-X_n,\theta)}{\partial\theta_\alpha}\right)^2\right)$$

$$=\sum_{\alpha=1}^{m}\tilde{I}_{\alpha\alpha}(\theta)=\sum_{\alpha=1}^{m}n\,I_{\alpha\alpha}(\theta)$$

$$=n\,\mathrm{tr}(I(\theta))\qquad(6)$$

Si combinamos ahora (3), (4) y (6) obtendremos

$$\left(m+\sum_{\alpha=1}^{m}\frac{\partial}{\partial\theta_\alpha}B_\alpha(\theta)\right)^2\leq\left(E_\theta(\|U-\theta\|^2)-\|B(\theta)\|^2\right)n\,\mathrm{tr}(I(\theta))$$

igual a:

$$\boxed{E_\theta\left(\|\mathcal{U}-\theta\|^2\right) \geq \|B(\theta)\|^2 + \frac{\left(m + \sum_{\alpha=1}^{m} \frac{\partial}{\partial\theta_\alpha} B_\alpha(\theta)\right)^2}{n\, tr\,(I(\theta))} \qquad (7)}$$

Como

$$\underbrace{E_\theta\left(\|\mathcal{U}-\theta\|^2\right)}_{\text{error cuadrático medio}} = tr\left(cov_\theta(\mathcal{U})\right) + \|B(\theta)\|^2$$

Observar que: $E_\theta(\|\mathcal{U}-\theta\|^2) = E_\theta(\|\mathcal{U}-E_\theta(\mathcal{U}) + E_\theta(\mathcal{U})-\theta\|^2) =$
$= E_\theta(\|\mathcal{U}-E_\theta(\mathcal{U}) + B(\theta)\|^2) = E_\theta(\|\mathcal{U}-E_\theta(\mathcal{U})\|^2) + \|B(\theta)\|$
pero $E_\theta(\|\mathcal{U}-E_\theta(\mathcal{U})\|^2) = E_\theta(tr\,(\mathcal{U}-E_\theta(\mathcal{U}))(\mathcal{U}-E_\theta(\mathcal{U}))^t)$
$= tr\, E_\theta((\mathcal{U}-E_\theta(\mathcal{U}))(\mathcal{U}-E_\theta(\mathcal{U}))^t) =$
$= tr\,(cov_\theta(\mathcal{U}))$

podemos escribir también:

$$\boxed{tr\left(cov_\theta(\mathcal{U})\right) \geq \frac{\left(m + \sum_{\alpha=1}^{m} \frac{\partial}{\partial\theta_\alpha} B_\alpha(\theta)\right)^2}{n\, tr\,(I(\theta))} \qquad (8)}$$

Consideremos el caso particular $\mathcal{U}$ insesgado. Entonces $B(\theta)=0$ $\quad \frac{\partial B_\alpha(\theta)}{\partial\theta_\alpha} = 0$ y tendremos:

$$\boxed{E_\theta\left(\|\mathcal{U}-\theta\|^2\right) \geq \frac{m^2}{n\, tr\,(I(\theta))} \qquad (9)}$$

o bien:

$$\boxed{tr\left(cov_\theta(\mathcal{U})\right) \geq \frac{m^2}{n\, tr\,(I(\theta))} \qquad (10)}$$

Otro caso particular: si θ es unidimensional, $m=1$, entonces:

$cov_\theta(\mathcal{U}) = var_\theta(\mathcal{U})$ (la matriz de covarianzas es 1×1 y su único elemento es la varianza de $\mathcal{U}=(\mathcal{U}_1)$ (no pondremos subíndice a $\mathcal{U}$ ya que sólo tiene una componente)

$I(\theta) = (I_{11}(\theta))$

$$I(\theta) = E_\theta\left(\left(\frac{\partial \ln f(X,\theta)}{\partial\theta}\right)^2\right)$$

(no pondremos índices a θ ni a $I(\theta)$ ni $B(\theta)$)

$B'(\theta) = \partial B_1(\theta)/\partial\theta_1$

Por tanto para $m=1$ tendremos:

$$\boxed{E_\theta\left((\mathcal{U}-\theta)^2\right) \geq B(\theta)^2 + \frac{(1+B'(\theta))^2}{n\, I(\theta)} \qquad (11)}$$

o bien:

$$\boxed{\operatorname{var}_\theta(\mathcal{U}) \geq \frac{(1+B'(\theta))^2}{n\, I(\theta)} \qquad (12)}$$

Si además $\mathcal{U}$ es insesgado:

$$\boxed{E_\theta\left((\mathcal{U}-\theta)^2\right) = \operatorname{var}_\theta(\mathcal{U}) \geq \frac{1}{n\, I(\theta)} \qquad (13)}$$

Las expresiones (12) y (13) son las conocidas desigualdades (para $m=1$) de Cramér-Rao, y las cotas inferiores correspondientes se conocen como cotas de Cramér-Rao. Para el caso multidimensional tendremos las expresiones (7), (8) y (9) ó (10).

Para llegar a estas expresiones hemos requerido ciertas "condiciones de regularidad" para el modelo: soporte independiente del parámetro θ, derivadas parciales de f respecto θ_α existentes, información de Fisher existente, etc..

NOTA adicional I:

En cálculo vectorial, dado un vector $(Y_1(x_1 \dots x_m), \dots, Y_m(x_1, \dots, x_m))$ se define el operador "divergencia del vector" (respecto un sistema de coordenadas cartesiano como:

$$\operatorname{div} Y = \sum_{\alpha=1}^{m} \frac{\partial Y_\alpha}{\partial x_\alpha}$$

desde esta óptica, la expresión $\displaystyle\sum_{\alpha=1}^{m} \frac{\partial}{\partial \theta_\alpha} B_\alpha(\theta)$ puede ser entendida como la "divergencia del sesgo"

$$\operatorname{div} B = \sum_{\alpha=1}^{m} \frac{\partial}{\partial \theta_\alpha} B_\alpha(\theta)$$

pudiendo escribir (7), (8) de forma más compacta como:

$$\boxed{E_\theta\left(\|U-\theta\|^2\right) \geq \|B_\theta\|^2 + \frac{\left(m + \operatorname{div}(B_\theta)\right)^2}{n \operatorname{tr}(I_\theta)} \qquad (14)}$$

donde hemos escrito B_θ y I_θ en vez de $B(\theta)$ y $I(\theta)$ por simplificar la notación. También :

$$\boxed{\operatorname{tr}\left(\operatorname{cov}_\theta(U)\right) \geq \frac{\left(m + \operatorname{div}(B_\theta)\right)^2}{n \operatorname{tr}(I_\theta)} \qquad (15)}$$

Nota adicional II :

En el caso multidimensional, $m>1$, las desigualdades de Cramer-Rao pueden expresarse de forma matricial.

Si introducimos la matriz Δ $m\times m$ definida por

$$\Delta_\theta = \left(\frac{\partial}{\partial\theta_\beta} B_\alpha(\theta) + \delta_{\alpha\beta} \right) \qquad (\text{ver } (1)) \qquad \begin{array}{l} \alpha \text{ índice de fila} \\ \beta \text{ índice de columna} \end{array}$$

podemos reescribir (1) en forma matricial como

$$\Delta_\theta = E_\theta\left((U-\theta-B(\theta)) \left(\frac{\partial \ln \tilde{f}(x_1 \dots x_m,\theta)}{\partial\theta} \right)^t \right) =$$

$$= E_\theta\left(U \left(\frac{\partial \ln \tilde{f}(x_1 \dots x_m,\theta)}{\partial\theta} \right)^t \right)$$

Si definimos el vector Y, $m\times 1$, igual a :

$$Y = U - \Delta_\theta \tilde{I}^{-1}(\theta) \frac{\partial \ln \tilde{f}(x_1 \dots x_m,\theta)}{\partial\theta}$$

donde suponemos implícitamente que $\tilde{I}(\theta)$ posee inversa.

Por otra parte sabemos que dados dos vectores aleatorios V y W resulta:

$$\operatorname{cov}(V+W) = \operatorname{cov}(V) + \operatorname{cov}(W) + E\left((V-E(V)) (W-E(W))^t \right) + E\left((W-E(W)) (V-E(V))^t \right)$$

siempre que estas existan.

Por tanto tendremos :

$$\operatorname{cov}_\theta(Y) = \operatorname{cov}_\theta(u) + \operatorname{cov}_\theta\left(\Delta_\theta \tilde{I}^{-1}(\theta) \frac{\partial \ln \tilde{f}(x_1 - x_n, \theta)}{\partial \theta}\right)$$

$$- E_\theta\left(u\left(\frac{\partial \ln \tilde{f}(x_1 - x_n, \theta)}{\partial \theta}\right)^t \tilde{I}^{-1}(\theta)\, \Delta_\theta^t\right)$$

$$- E_\theta\left(\Delta_\theta \tilde{I}^{-1}(\theta) \frac{\partial \ln \tilde{f}(x_1 - x_n, \theta)}{\partial \theta}\, u^t\right)$$

pero

$$\operatorname{cov}_\theta\left(\Delta_\theta \tilde{I}^{-1}(\theta) \frac{\partial \ln \tilde{f}(x_1 - x_n, \theta)}{\partial \theta}\right) = E_\theta\left(\Delta_\theta \tilde{I}^{-1}(\theta) \frac{\partial \ln \tilde{f}(x_1 - x_n, \theta)}{\partial \theta} \cdot \left(\frac{\partial \ln \tilde{f}(x_1 - x_n, \theta)}{\partial \theta}\right)^t \tilde{I}^{-1}(\theta)\, \Delta_\theta^t\right)$$

$$= \Delta_\theta \tilde{I}^{-1}(\theta) \underbrace{E_\theta\left(\frac{\partial \ln \tilde{f}(x_1 - x_n, \theta)}{\partial \theta}\left(\frac{\partial \ln \tilde{f}(x_1 - x_n, \theta)}{\partial \theta}\right)^t\right)}_{\tilde{I}(\theta)} \tilde{I}^{-1}(\theta)\, \Delta_\theta^t$$

$$= \Delta_\theta \tilde{I}_\theta^{-1} \Delta_\theta^t$$

mientras que

$$E_\theta\left(u\left(\frac{\partial \ln \tilde{f}(x_1 - x_n, \theta)}{\partial \theta}\right)^t \tilde{I}^{-1}(\theta)\, \Delta_\theta^t\right) =$$

$$= E_\theta\left(u\left(\frac{\partial \ln \tilde{f}(x_1 - x_n, \theta)}{\partial \theta}\right)^t\right) \tilde{I}^{-1}(\theta)\, \Delta_\theta^t =$$

$$= \Delta_\theta \tilde{I}^{-1}(\theta)\, \Delta_\theta^t$$

y su transpuesta

$$E_\theta\left(\Delta_\theta \tilde{I}^{-1}(\theta) \frac{\partial \ln \tilde{f}(x_1 - x_n, \theta)}{\partial \theta}\, u^t\right) = \Delta_\theta \tilde{I}^{-1}(\theta)\, \Delta_\theta^t$$

Por tanto:

$$\operatorname{cov}_\theta(Y) = \operatorname{cov}_\theta(u) - \Delta_\theta \tilde{I}^{-1}(\theta)\, \Delta_\theta^t$$

Además, como $\tilde{I}(\theta) = n\, I(\theta)$ resulta que

$$\tilde{I}(\theta)^{-1} = \frac{1}{n}\, I(\theta)^{-1}$$

Al ser $cov_\theta(Y)$, como cualquier matriz de covarianzas, definida o semidefinida positiva, tendremos:

$$0 \leq cov_\theta(Y) = cov_\theta(\mathcal{U}) - \frac{1}{n}\, \Delta_\theta\, I(\theta)^{-1} \Delta_\theta^t$$

o de forma equivalente

$$\boxed{cov_\theta(\mathcal{U}) \geq \frac{1}{n}\, \Delta_\theta\, I_\theta^{-1} \Delta_\theta^t \qquad (16)}$$

donde hemos escrito I_θ en vez de $I(\theta)$ por simplicidad.

En el caso que $\mathcal{U}$ sea insesgado, $\Delta_\theta = I_m$ (la identidad $m \times m$) por tanto tendremos:

$$\boxed{cov_\theta(\mathcal{U}) \geq \frac{1}{n}\, I_\theta^{-1} \qquad (17)}$$

OBSERVACIÓN:

Si A y B son matrices simétricas $A \leq B \iff B-A$ es definida o semidefinida positiva (valores propios no negativos $x^t(B-A)x \geq 0 \quad \forall x$ vector $m \times 1$)

OBSERVACIÓN FINAL:

Puede comprobarse que si $\mathcal{U}$ es tal que puede escribirse como:

$$\boxed{\sum_{i=1}^{n} \frac{\partial \ln f(x_i, \theta)}{\partial \theta} = K(\theta, n)\, \big(\mathcal{U}(x_1 \dots x_n) - \theta \big) \qquad (18)}$$

donde K es una matriz $m \times m$ no singular, entonces:

$\mathcal{U}$ es insesgado para θ y además alcanza la cota de Cramér-Rao

Observar que si se cumple (18), como

$$E_\theta\left(\frac{\partial \ln f(x_i,\theta)}{\partial\theta}\right)=0$$ resulta

$$0 = E_\theta\big(K(\theta,m)\,(\mathcal{U}(x_1-x_n)-\theta)\big) = K(\theta,m)\big(E_\theta(\mathcal{U}(x_1-x_n))-\theta\big)$$

por tanto al ser $K(\theta,m)$ regular,

$$0 = E_\theta(\mathcal{U}(x_1\cdots x_n))-\theta = B(\theta)$$

$\mathcal{U}$ es insesgado: $B(\theta)=0$ y Δ_θ, la matriz introducida en la página 8, es $\Delta_\theta = I_m$ (identidad $m\times m$)

no confundir con la matriz de información de Fisher I_θ, etc.

Además, observar que

$$\sum_{i=1}^{n}\frac{\partial \ln f(x_i\,\theta)}{\partial\theta} = \frac{\partial \ln \tilde f(x_1\cdots x_n,\theta)}{\partial\theta}$$

por tanto, tomando covarianzas tendremos:

$$\operatorname{cov}_\theta\left(\frac{\partial \ln\tilde f(x_1\cdots x_n,\theta)}{\partial\theta}\right) = E_\theta\left(\frac{\partial \ln\tilde f(x_1-x_n\,\theta)}{\partial\theta}\left(\frac{\partial \ln\tilde f(x_1-x_n\,\theta)}{\partial\theta}\right)^t\right)$$

$$= \tilde I(\theta)$$

(ya que $E_\theta\left(\frac{\partial \ln\tilde f(x_1\,x_n\,\theta)}{\partial\theta}\right)=0$)

por tanto

$$\tilde I(\theta) = \operatorname{cov}_\theta\big(K(\theta,m)\,(\mathcal{U}(x_1-x_n)-\theta)\big) =$$

$$= \operatorname{cov}_\theta\big(K(\theta,m)\,\mathcal{U}(x_1-x_n)\big) =$$

$$= K(\theta,m)\operatorname{cov}_\theta(\mathcal{U})\,K(\theta,m)^t \qquad (19)$$

(recordar $\operatorname{cov}(Ax+c) = A\operatorname{cov}(x)A^t$) si A y c son constantes: no dependen de la v.a. x

Por otra parte, de (18) obtenemos:

$$\frac{\partial \ln\tilde f(x_1-x_n\,\theta)}{\partial\theta}\left(\frac{\partial \ln\tilde f(x_1-x_n|\theta)}{\partial\theta}\right)^t = K(\theta,m)(\mathcal{U}(x_1-x_n)-\theta)\left(\frac{\partial \ln f(x_1-x_n\,\theta)}{\partial\theta}\right)^t$$

tomando esperanzas:

$$\tilde I(\theta) = K(\theta,m)\,E_\theta\left((\mathcal{U}(x_1-x_n)-\theta)\left(\frac{\partial \ln f(x_1-x_n,\theta)}{\partial\theta}\right)^t\right) \qquad (20)$$

Escribiendo matricialmente (1), teniendo en cuenta que $B(\theta)=0$ y $\Delta_m = I$, resulta:

$$I_m = E_\theta\left((\mathcal{U}-\theta)\left(\frac{\partial \ln \tilde{f}(x_1 - x_n, \theta)}{\partial \theta}\right)^t \right) \quad (21)$$

por tanto combinando (20) y (21) resulta:

matriz identidad

$$\tilde{I}(\theta) = K(\theta, n)\, I_m = K(\theta, n) \quad (22)$$

En otras palabras, si $\mathcal{U}$ satisface la condición (18) concluimos que $\mathcal{U}$ es insesgado y $cov_\theta(\mathcal{U}) = \tilde{I}_\theta^{-1} = \frac{1}{n} I_\theta^{-1}$

NOTA: observar que si se cumple (18), entonces

$$K(\theta, n) = \tilde{I}(\theta) = n\, I(\theta) \quad (23)$$
