Desigualdad de Cràmer-Rao

Yoselin Arvelaiz

29 de noviembre de 2018

Introducción

¿Cómo escoger el mejor estimador?

Podemos escoger siempre estimadores, algunos muy buenos y otros no tanto, por supuesto que las bondades de los estimadores deben estar establecidas siguiendo algún criterio. En este caso las daremos en términos de los errores cuadráticos. Entonces una pregunta nos viene de inmediato, que tan bueno puede ser un estimador en términos de su error cuadrático. La respuesta la otorga la desigualdad de Cràmer-Rao, esta expresa el mínimo valor que puede alcanzar el error cuadrático, cuando se estima un parámetro insesgado.

Sí w_1 y w_2 son estimadores insesgados del parámetro θ. Es decir, $\mathbb{E}[w_1] = \mathbb{E}[w_2] = \theta$, entonces el error cuadrático medio (ECM) son iguales a la varianza. Deberiamos escoger el estimador que tiene la menor varianza.

Sí $\mathbb{B}(\theta)$ es una cota inferior de cualquier estimador insesgado θ podemos encontrar un estimador w que satisface que $Var(w) = \mathbb{B}(\theta)$ entonces w es el mejor estimador insesgado. Esto se hace con la cota inferior de Cràmer-Rao.

La desigualdad de Cràmer-Rao, llamada así en honor a **Harald Cràmer** y **Calyampudi Radhakrishna Rao**, expresa una cota inferior para la varianza de un estimador insesgado, basado en la información de Fisher.

Antes de enunciar el teorema de desigualdad de Cràmer-Rao, veremos el siguiente resultado,

Regla de Leibnitz

Sí $f(x,\theta), a(\theta)$ y $b(\theta)$ son diferenciables con respecto a θ entonces,

$$\frac{d}{d\theta}\int_{a(\theta)}^{b(\theta)} f(x,\theta)dx = f(b(\theta),\theta)\frac{db(\theta)}{d\theta} - f(a(\theta),\theta)\frac{da(\theta)}{d\theta} + \int_{a(\theta)}^{b(\theta)} \frac{\partial}{\partial\theta} f(x,\theta)dx$$

Sí $a(\theta)$ y $b(\theta)$ son constantes

$$\frac{d}{d\theta}\int_{a(\theta)}^{b(\theta)} f(x,\theta)dx = \int_{a(\theta)}^{b(\theta)} \frac{\partial}{\partial\theta} f(x,\theta)dx.$$

Para enunciar el teorema de Cràmer-Rao supondremos que tanto la función de probabilidad en el caso discreto como la función de densidad en el caso continuo tienen dos derivadas continuas con respecto a la variable θ.

Teorema (Desigualdad de Cràmer-Rao)

Sean $X_1, ..., X_n$ una muestra aleatoria con función de probabilidad $f(X|\theta)$ y sea $w(X) = w(X_1, ..., X_n)$ un estimador de θ que satisface

$$\frac{d}{d\theta}\mathbb{E}[w(X)] = \int \frac{\partial}{\partial\theta}[w(X)f(X|\theta)]dX$$

y $Var(X) < \infty$. Entonces,

$$Var_\theta[w(X)] \geq \frac{\left(\frac{d}{d\theta}\mathbb{E}_\theta[w(X)]\right)^2}{\mathbb{E}_\theta[(\frac{\partial}{\partial\theta}\log f(X|\theta))^2]}$$

Demostración:

Ésta demostración se basa en la desigualdad de Cauchy-Schwarz para la covarianza,

$$|Cov(X,Y)|^2 \leq Var(X)Var(Y),$$
$$Var(X) \geq \frac{|Cov(X,Y)|^2}{Var(Y)}$$

<u>Observación 1:</u>

$$\begin{aligned}
\frac{d}{d\theta}\mathbb{E}_\theta[w(X)] &= \frac{d}{d\theta}\int w(X)f(X|\theta)dX \\
&= \int w(X)\frac{\partial}{\partial\theta}f(X|\theta)dX \\
&= \int w(X)\frac{\frac{\partial}{\partial\theta}f(X|\theta)}{f(X|\theta)}f(X|\theta)dX \\
&= \int w(X)\frac{\partial}{\partial\theta}\log(f(X|\theta))f(X|\theta)dX \\
&= \mathbb{E}\left[w(X)\tfrac{\partial}{\partial\theta}\log(f(X|\theta))\right]
\end{aligned}$$

Consideremos $w(X) = 1$,

$$\begin{aligned}
\mathbb{E}\left[w(X)\tfrac{\partial}{\partial\theta}\log(f(X|\theta))\right] &= \mathbb{E}\left[\tfrac{\partial}{\partial\theta}\log(f(X|\theta))\right] \\
&= \int \frac{\partial}{\partial\theta}\log(f(X|\theta))f(X|\theta)dX \\
&= \int \frac{\partial}{\partial\theta}f(X|\theta)dX \\
&= \frac{d}{d\theta}\int f(X|\theta)dX \\
&= \frac{d}{d\theta}(1) \\
&= 0
\end{aligned}$$

Entonces,

$$\begin{aligned}
Cov_\theta[w(X), \tfrac{\partial}{\partial\theta}f(X|\theta)] &= \mathbb{E}_\theta\left[w(X)\tfrac{\partial}{\partial\theta}\log(f(X|\theta))\right] - \mathbb{E}_\theta[w(X)]\,\mathbb{E}_\theta\left[\tfrac{\partial}{\partial\theta}\log(f(X|\theta))\right] \\
&= \mathbb{E}_\theta\left[w(X)\tfrac{\partial}{\partial\theta}\log(f(X|\theta))\right] \\
&= \frac{d}{d\theta}\mathbb{E}_\theta[w(X)], \qquad \text{Por observación 1}
\end{aligned}$$

Cómo $\mathbb{E}_\theta\left[\frac{\partial}{\partial\theta}\log(f(X|\theta))\right]=0$ entonces,

$$Var_\theta\left[\frac{\partial}{\partial\theta}\log(f(X|\theta))\right]=\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log(f(X|\theta))\right)^2\right]$$

Por la desigualdad de C.Sh se tiene:

$$Var_\theta[w(X)]\geq\frac{\left(\frac{d}{d\theta}\mathbb{E}_\theta[w(X)]\right)^2}{\mathbb{E}_\theta[(\frac{\partial}{\partial\theta}\log f(X|\theta))^2]}.$$

Sí asumimos que la muestra es i.i.d con función de densidad $f(X|\theta)$ y las mismas condiciones del teorema anterior. Entonces,

$$Var_\theta[w(X)]\geq\frac{\left(\frac{d}{d\theta}\mathbb{E}_\theta[w(X)]\right)^2}{n\mathbb{E}_\theta[(\frac{\partial}{\partial\theta}\log f(X|\theta))^2]}.$$

Probemos que $\mathbb{E}_\theta[(\frac{\partial}{\partial\theta}\log f(X|\theta))^2]=n\mathbb{E}_\theta[(\frac{\partial}{\partial\theta}\log f(X|\theta))^2]$

Demostración:

Cómo $X_1,...,X_n$ son independientes,

$$\begin{aligned}\mathbb{E}_\theta[(\tfrac{\partial}{\partial\theta}\log f(X|\theta))^2] &= \mathbb{E}_\theta[(\tfrac{\partial}{\partial\theta}\log(\prod_{i=1}^n f(X_i|\theta)))^2]\\ &= \mathbb{E}_\theta[(\sum_{i=1}^n\frac{\partial}{\partial\theta}\log f(X_i|\theta))^2]\\ &= \sum_{i=1}^n\mathbb{E}_\theta[(\frac{\partial}{\partial\theta}\log f(X_i|\theta))^2]+\sum_{i\neq j}^n\mathbb{E}_\theta[(\frac{\partial}{\partial\theta}\log f(X_i|\theta))(\frac{\partial}{\partial\theta}\log f(X_j|\theta))]\end{aligned}$$

Para $i\neq j$,

$$\begin{aligned}\mathbb{E}_\theta\left[(\tfrac{\partial}{\partial\theta}\log f(X_i|\theta))(\tfrac{\partial}{\partial\theta}\log f(X_j|\theta))\right] &= \mathbb{E}_\theta\left[(\tfrac{\partial}{\partial\theta}\log f(X_i|\theta))\right]\mathbb{E}_\theta\left[(\tfrac{\partial}{\partial\theta}\log f(X_j|\theta))\right]\\ &=0\end{aligned}$$

Cómo, $\sum_{i=1}^n\mathbb{E}_\theta[(\frac{\partial}{\partial\theta}\log f(X_i|\theta))^2]=n\mathbb{E}_\theta[(\frac{\partial}{\partial\theta}\log f(X|\theta))^2]$ por ser i.i.d, entonces,

$$\mathbb{E}_\theta[(\frac{\partial}{\partial\theta}\log f(X|\theta))^2]=n\mathbb{E}_\theta[(\frac{\partial}{\partial\theta}\log f(X|\theta))^2].$$

Nota: La desigualdad de Cràmer-Rao también se puede aplicar a variables discretas (se intercambia diferenciación con $\sum$).

La cantidad $\mathbb{E}_\theta[(\frac{\partial}{\partial\theta}\log f(X|\theta))^2]$ es llamada la información de Fisher.

Ejemplo

Información de Fisher para la densidad Gaussiana con respecto a μ. Tenemos que,

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

Calculemos,

$$\begin{aligned}
\mathbb{E}_\mu\left[(\tfrac{\partial}{\partial\mu}\log f(X|\mu,\sigma^2))^2\right] &= \mathbb{E}_\mu\left[(\tfrac{\partial}{\partial\mu}[\log(\tfrac{1}{\sqrt{2\pi\sigma^2}}) - \tfrac{(X-\mu)^2}{2\sigma^2}])^2\right] \\
&= \mathbb{E}_\mu\left[\left(\tfrac{X-\mu}{\sigma^2}\right)^2\right] \\
&= \frac{1}{\sigma^4}\mathbb{E}_\mu\left[(X-\mu)^2\right] \\
&= \frac{1}{\sigma^4}Var_\mu\left[X\right] \\
&= \frac{1}{\sigma^4}n\sigma^2 \\
&= \frac{n}{\sigma^2}
\end{aligned}$$

Luego,

$$\begin{aligned}
\left(\tfrac{d}{d\mu}\mathbb{E}_\mu[w(X)]\right)^2 &= (\tfrac{d}{d\mu}(\mu))^2 \\
&= 1
\end{aligned}$$

La cota de Cràmer-Rao para un estimador insesgado $w(X)$ de μ, es

$$Var_\theta[w(X)] \geq \frac{1}{\frac{n}{\sigma^2}} = \frac{\sigma^2}{n}$$

De esta forma la cota de Cramer-Rao para estimar μ resulta ser: $\dfrac{\sigma^2}{n}$, número que coincide con el riesgo cuadrático del estimador $\overline{X_n}$. Este estimador resulta ser entonces el mejor estimador insesgado. En efecto, para el estimador $\overline{X_n} = \frac{\sum_{i=1}^n X_i}{n}$ de μ, se tiene que:

$$
\begin{aligned}
Var_\mu\left[\overline{X_n}\right] &= Var_\mu\left[\frac{\sum_{i=1}^n X_i}{n}\right] \\
&= \frac{1}{n^2} Var_\mu\left[\sum_{i=1}^n X_i\right] \\
&= \frac{1}{n^2}\sum_{i=1}^n Var_\mu[X_i] \\
&= \frac{1}{n^2} n\sigma^2 \\
&= \frac{\sigma^2}{n}
\end{aligned}
$$