Ya tenemos

- algunos estimadores “naturales”, junto con diversas propiedades de interés (sesgo, varianza, ECM);
- además de métodos generales de construcción de estimadores (momentos, máxima verosimilitud).

Nos disponemos ahora a analizar

- la cota de Cramér-Rao;
- y el comportamiento asintótico (para n, tamaño de la muestra, grande) de estimadores.

Derivadas de funciones de densidad/masa

Variable X con función de densidad/masa $f(x;\theta)$, con $x \in \mathbf{sop}_\theta$ y $\theta \in \Theta$. Suponemos que Θ es un intervalo.

Muestra aleatoria $(X_1, \ldots, X_n)$ con función de densidad conjunta $f(\mathbf{x};\theta)$ dada por

$$f(\mathbf{x};\theta) = \prod_{j=1}^{n} f(x_j;\theta),$$

con $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbf{sop}_\theta^n$ y $\theta \in \Theta$.

Interesa disponer de una expresión explícita de la derivada de $f(\mathbf{x};\theta)$ con respecto a θ.

Usaremos la notación ∂_θ, en lugar de $\frac{\partial}{\partial\theta}$ o $\frac{d}{d\theta}$.

Derivamos el producto:

$$\partial_\theta\big(f(\mathbf{x};\theta)\big) = \partial_\theta\Big(\prod_{j=1}^{n} f(x_j;\theta)\Big) = \sum_{j=1}^{n} \partial_\theta f(x_j;\theta)\frac{f(\mathbf{x};\theta)}{f(x_j;\theta)}\,,$$

es decir,

$$\frac{\partial_\theta f(\mathbf{x};\theta)}{f(\mathbf{x};\theta)} = \sum_{j=1}^{n} \frac{\partial_\theta f(x_j;\theta)}{f(x_j;\theta)}$$

o bien

$$\partial_\theta \ln(f(\mathbf{x};\theta)) = \sum_{j=1}^{n} \partial_\theta \ln(f(x_j;\theta))$$

Información/número de información de una variable aleatoria

Llamamos **información de la variable** X a la variable aleatoria Y dada por

$$Y = \frac{\partial_\theta f(X;\theta)}{f(X;\theta)} = \partial_\theta \ln(f(X;\theta))\,.$$

Para cada $\theta \in \Theta$, la variable Y está definida por la expresión anterior en $\mathbf{sop}_\theta$; fuera de $\mathbf{sop}_\theta$ entendemos que $Y \equiv 0$.

El **número/cantidad de información de** $f(x;\theta)$ se define como la varianza de Y:

$$I_X(\theta) = \mathbf{V}_\theta(Y) = \mathbf{V}_\theta(\partial_\theta \ln(f(X;\theta))\,.$$

I_X es una función definida para $\theta \in \Theta$.

Ejemplo 1. Digamos que X es una variable $\mathrm{EXP}(\lambda)$. El parámetro θ que se desea estimar es la media: $\theta = 1/\lambda$. La función de densidad es

$$f(x;\theta) = \frac{1}{\theta} e^{-x/\theta}, \quad \text{para } x > 0 \,.$$

Obsérvese que $\mathbf{E}_\theta(X) = (1/\lambda) = \theta$ y que $\mathbf{V}_\theta(X) = (1/\lambda^2) = \theta^2$.

Como

$$\ln f(x;\theta) = -\ln(\theta) - \frac{x}{\theta} \quad \Longrightarrow \quad \partial_\theta\big(\ln f(x;\theta)\big) = \frac{(x-\theta)}{\theta^2}\,,$$

y, por tanto,

$$Y = \frac{X-\theta}{\theta^2},$$

de donde

$$I_X(\theta) = \mathbf{V}_\theta(Y) = \mathbf{E}_\theta\big(Y^2\big) = \frac{1}{\theta^4}\mathbf{V}_\theta(X) = \frac{1}{\theta^2}\,.$$

Ejemplo 2. Digamos que X es una variable $\text{POISS}(\lambda)$. El parámetro es $\theta = \lambda$. La función de masa es

$$f(k;\lambda) = e^{-\lambda}\frac{\lambda^k}{k!}\,, \quad \text{para cada entero } k \geq 0\,.$$

Obsérvese que $\mathbf{E}_\lambda(X) = \lambda$ y que $\mathbf{V}_\lambda(X) = \lambda$.

Como

$$\ln f(k;\lambda) = -\lambda + k\ln(\lambda) - \ln(k!) \quad \Longrightarrow \quad \partial_\lambda\big(\ln f(x;\lambda)\big) = -1 + \frac{k}{\lambda}\,,$$

y, por tanto,

$$Y = \frac{X-\lambda}{\lambda},$$

de donde

$$I_X(\lambda) = \mathbf{V}_\lambda(Y) = \mathbf{E}_\lambda\big(Y^2\big) = \frac{1}{\lambda^2}\,\mathbf{V}_\lambda(X) = \frac{1}{\lambda}\,.$$

(No)-ejemplo 3. Digamos que X es una variable $\textsc{unif}[0,b]$. El parámetro es $\theta = b$. La función de densidad es

$$f(x;b) = \frac{1}{b} \quad \text{para } x \in [0,b]\,.$$

Obsérvese que $\mathbf{E}_b(X) = b/2$ y que $\mathbf{V}_b(X) = b^2/12$.

Como

$$\ln f(x;b) = -\ln(b) \quad \Longrightarrow \quad \partial_b\big(\ln f(x;b)\big) = -\frac{1}{b}\,,$$

resulta que

$$Y = -\frac{1}{b},$$

Se trata de un caso "degenerado": obsérvese que $\mathbf{E}_b(Y) = -1/b$, que $\mathbf{E}_b(Y^2) = 1/b^2$, pero

$$I_X(b) = \mathbf{V}_b(Y) = 0\,.$$

Cota de Cramér-Rao

Es deseable disponer de estimadores

- insesgados;
- y con la menor varianza posible.

Vamos a ver que existe una cota, que depende sólo de la "forma" de la función de densidad/masa, que limita cuán pequeña puede llegar ser la varianza de estimadores insesgados.

Ilustraremos el argumento con el caso en el que

- que el conjunto de parámetros Θ es un intervalo de $\mathbb{R}$;
- la variable X es finita;
- y su soporte no depende de $\theta \in \Theta$.

Digamos que, para cada $\theta \in \Theta$, la variable X toma un número finito de valores A con probabilidad positiva. Además $\mathbf{sop}_\theta = A$ para cada $\theta \in \Theta$.

Lema 1

En la situación descrita,

$$\mathbf{E}_\theta(Y) = 0 \quad \text{para cada } \theta \in \Theta\,.$$

Y por tanto

$$I_X(\theta) = \mathbf{V}_\theta(Y) = \mathbf{E}_\theta(Y^2) = \mathbf{E}_\theta\Big(\Big(\frac{\partial_\theta f(X;\theta)}{f(X;\theta)}\Big)^2\Big)\,.$$

Demostración. Como para cualquier $\theta \in \Theta$,

$$(\star) \qquad 1 = \sum_{x \in A} f(x;\theta)$$

(una suma finita), derivando respecto de θ en el intervalo Θ obtenemos

$$0 = \sum_{x \in A} \partial_\theta f(x;\theta) = \sum_{x \in A} \left(\frac{\partial_\theta f(x;\theta)}{f(x;\theta)} \right) f(x;\theta) = \mathbf{E}_\theta(Y)\,.$$

Sea ahora $(X_1, \ldots, X_n)$ una muestra aleatoria de X. Cada X_j tiene su variable de información (clónicas e independientes)

$$Y_j = \frac{\partial_\theta f(X_j;\theta)}{f(X_j;\theta)}\,, \quad \text{para } j = 1, 2, \ldots, n\,,$$

Consideremos

$$Z_n = \sum_{j=1}^{n} Y_j$$

(un estadístico, pues es función de $X_1, \ldots, X_n$).

Lema 2

Para cada $\theta \in \Theta$,

$$\mathbf{E}_\theta(Z_n) = 0 \qquad \text{y} \qquad \mathbf{V}_\theta(Z_n) = n\, I_X(\theta)\,.$$

Sea $T = h(X_1, X_2, \ldots, X_n)$ un estimador **insesgado** de θ, es decir,

$$\theta = \mathbf{E}_\theta(T)\,, \quad \text{para todo } \theta \in \Theta\,,$$

Que reescribimos

$$(\star\star) \qquad \theta = \mathbf{E}_\theta\big(h(X_1, \ldots, X_n)\big) = \sum_{\mathbf{x} \in A^n} h(\mathbf{x})\ f(\mathbf{x}; \theta)$$

Derivamos:

$$1 = \sum_{\mathbf{x} \in A^n} h(\mathbf{x})\ \Big(\sum_{j=1}^{n} \frac{\partial_\theta f(x_j; \theta)}{f(x_j; \theta)}\Big)\ f(\mathbf{x}; \theta).$$

Es decir,

$$1 = \mathbf{E}_\theta(T \cdot Z_n)\,.$$

Como $\mathbf{E}_\theta(Z_n) = 0$,

$$\mathrm{cov}_\theta(T, Z_n) = \mathbf{E}_\theta(T \cdot Z_n) - \mathbf{E}_\theta(T) \cdot \mathbf{E}_\theta(Z_n) = \mathbf{E}_\theta(T \cdot Z_n) = 1\,.$$

De la desigualdad (general)

$$\mathrm{cov}_\theta(T, Z_n) \leq \sqrt{\mathbf{V}_\theta(T)}\sqrt{\mathbf{V}_\theta(Z_n)}\,,$$

se deduce que

$$1 \leq \mathbf{V}_\theta(T)\mathbf{V}_\theta(Z_n) = \mathbf{V}_\theta(T) \cdot n\, I_X(\theta)\,.$$

Teorema 3 (Cramér–Rao, caso finito, soporte independiente de θ)

Para todo estadístico insesgado T de θ,

$$\mathrm{ECM}_\theta(T) = \mathbf{V}_\theta(T) \geq \frac{1}{n\ I_X(\theta)}\,, \quad \text{para todo } \theta \in \Theta.$$

El caso general

Hemos obtenido esta cota bajo las hipótesis

1) variable finita;
2) soporte independiente de θ.

Mantengamos 2). Denotemos por **sop** al soporte común, para cada $\theta \in \Theta$.

En el argumento anterior hemos derivado dos ecuaciones, $(\star)$ y $(\star\star)$ y hemos usado que la derivada de la suma es la suma de las derivadas.

Si la variable X fuera discreta con soporte A numerable, pero no finito, tendríamos series, en lugar de sumas, y para poder intercambiar serie y derivación hace falta que $f(x;\theta)$ y sus derivadas respecto de θ tiendan a 0 (no demasiado lentamente) cuando $|x| \to \infty$.

Cuando X es continua, las esperanzas son integrales. Para poder derivar bajo el signo integral en una expresión como

$$(\star') \quad 1 = \int_{\textbf{sop}} f(x;\theta)dx$$

para obtener que

$$0 = \int_{\textbf{sop}} \partial_\theta f(x;\theta)dx = \int_{\textbf{sop}} \frac{\partial_\theta f(x;\theta)}{f(x;\theta)} f(x;\theta)dx \quad \Longrightarrow \quad \mathbf{E}_\theta(Y) = 0\,,$$

hace falta que $f(x;\theta)$ y sus derivadas respecto de θ decaigan a cero no demasiado lentamente cuando $|x| \to \infty$ (por ejemplo, para garantizar que la integral $\int_{\textbf{sop}} \partial_\theta f(x;\theta)dx$ converge).

Estas condiciones se cumplen para los modelos para X que se usan en la práctica.

Teorema 4 (Cota de Cramér–Rao)

En condiciones muy generales, si $\mathbf{sop}_\theta$ *no depende* de θ, entonces para todo estadístico insesgado T de θ se cumple que

$$\mathbf{V}_\theta(T) \geq \frac{1}{n\ I_X(\theta)}\,, \quad \text{para todo } \theta \in \Theta$$

A un estimador insesgado T del parámetro θ cuya varianza es justamente la cota de Cramér–Rao, es decir, tal que

$$\mathbf{V}_\theta(T) \cdot I_X(\theta) = \frac{1}{n}\,, \quad \text{para todo } \theta \in \Theta$$

se le dice **estimador eficiente** o **insesgado de mínima varianza**.

El caso del soporte dependiente de θ

En este caso, la cota no tiene por qué ser válida.

Ejemplo. $X \sim \text{UNIF}[0, b]$. Como ya vimos (ejemplo 3),

$$Y \equiv -1/b.$$

Para empezar, no se cumpliría que $\mathbf{E}_b(Y) = 0$, como debería. Además, $\mathbf{V}_b(Y) = 0$, y por tanto $I_X(b) = 0$.

Aún así, como $\mathbf{E}_b(Y^2) = 1/b^2$, la cota de Cramér–Rao "debería" ser b^2/n.

Pero ya vimos que el estimador insesgado $T = \frac{n+1}{n} \max(X_1, \ldots, X_n)$ tiene varianza

$$\mathbf{V}_b(T) = \frac{1}{n(n+2)} b^2 \, .$$

Observaciones:

- El cálculo del número de información puede ser complicado.
- No es seguro que exista un estimador eficiente (o que al menos sepamos decidir cuál es).
- En el caso en que exista, ¿será único?
- Desde el punto de vista de la eficiencia, ¿qué tal son, por ejemplo, los estimadores que se obtienen por máxima verosilimitud?

Ejemplo 1. Cota de Cramér–Rao para $X \sim \text{EXP}(\lambda)$, cuando queremos estimar $\theta = 1/\lambda$.

Ya sabemos que

$$I_X(\theta) = \frac{1}{\theta^2}.$$

La cota de Cramér-Rao es

$$\mathbf{V}_\theta(T) \geq \frac{\theta^2}{n}.$$

Recordemos que $\overline{X}$ es estimador insesgado de θ y que

$$\mathbf{V}_\theta(\overline{X}) = \frac{\mathbf{V}_\theta(X)}{n} = \frac{\theta^2}{n}.$$

Así que $\overline{X}$ es estimador insesgado de mínima varianza, es decir, estimador eficiente.

Ejemplo 2. Cota de Cramér–Rao para $X \sim \text{POISSON}(\lambda)$, cuando queremos estimar $\theta = \lambda$.

Ya sabemos que

$$I_X(\lambda) = \frac{1}{\lambda}.$$

La cota de Cramér-Rao es

$$\mathbf{V}_\lambda(T) \geq \frac{\lambda}{n}.$$

De nuevo, $\overline{X}$ es estimador insesgado de λ y su varianza es

$$\mathbf{V}_\lambda(\overline{X}) = \frac{\mathbf{V}_\lambda(X)}{n} = \frac{\lambda}{n}.$$

Así que $\overline{X}$ es estimador insesgado de mínima varianza, es decir, estimador eficiente.

Ejemplo 3. Cota de Cramér–Rao para $X \sim \mathcal{N}(0, \sigma^2)$, cuando queremos estimar $\theta = \sigma^2$. Tenemos

$$f(x;\theta) = \frac{1}{\sqrt{2\pi}\sqrt{\theta}} e^{-\frac{1}{2}x^2/\theta} \quad \text{para todo } x \in \mathbb{R}\,.$$

De manera que

$$\partial_\theta \ln(f(x;\sigma)) = \frac{x^2 - \theta}{2\theta^2}\,.$$

y, por tanto,

$$Y = \frac{X^2 - \theta}{2\theta^2}.$$

La media $\mathbf{E}_\theta(Y) = 0$, pues $\mathbf{E}_\theta(X^2) = \theta$.

Queda calcular

$$I_X(\theta) = \mathbf{V}_\theta(Y) = \mathbf{E}_\theta(Y^2) = \frac{1}{4\theta^4}\mathbf{E}_\theta((X^2 - \theta)^2)$$

Obsérvese que $X = \sqrt{\theta} Z$, donde $Z \sim \mathcal{N}(0,1)$.

Así que, como $\mathbf{E}(Z^4) = 3$,

$$\mathbf{E}_\theta((X^2-\theta)^2) = \theta^2\, \mathbf{E}_\theta((Z^2-1)^2) = \theta^2\, \mathbf{E}(Z^4+1-2Z^2) = \theta^2\,(3+1-2) = 2\,\theta^2$$

En total

$$I_X(\theta) = \frac{1}{2\,\theta^2}$$

La cota es $\dfrac{2\theta^2}{n}$.

S^2 es estimador insesgado de θ y tiene

$$\mathbf{V}(S^2) = \frac{2\theta^2}{n-1}.$$

No se alcanza (por poco) la cota de Cramér–Rao.