Suficiencia y Completitud.
Familias exponenciales. Estimadores IMVU
Cota de Rao–Cramer.

Suficiencia

- $\mathbf{X} \sim F_{\boldsymbol{\theta}}$ con $\boldsymbol{\theta} \in \Theta$.

- Se dice que $\mathbf{T} = t(\mathbf{X})$ es suficiente para $\boldsymbol{\theta}$ si:

  la distribución de $\mathbf{X} \,|\, \mathbf{T} = \mathbf{t}$ es independiente de $\boldsymbol{\theta}$ para todo $\mathbf{t} \in \mathcal{T}$ tal que

  $\mathbb{P}_{\boldsymbol{\theta}}(\mathbf{T} \notin \mathcal{T}) = 0$ para todo $\boldsymbol{\theta} \in \Theta$

Teorema de Factorización

Sea $\mathbf{X}$ un vector aleatorio con función de densidad o función de probabilidad puntual $p(\mathbf{x}, \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$.

El estadístico $\mathbf{T} = t(\mathbf{X})$ es suficiente para $\boldsymbol{\theta}$ si y sólo si existen dos funciones g y h tales que

$$p(\mathbf{x}, \boldsymbol{\theta}) = g(t(\mathbf{x}), \boldsymbol{\theta})h(\mathbf{x})$$

Rao Blackwell

Sea $\mathbf{X}$ un vector de una distribución perteneciente a la familia $\mathcal{F} = \{F_{\boldsymbol{\theta}} : \quad \boldsymbol{\theta} \in \Theta\}$. Sea $\mathbf{T}$ un estadístico suficiente para $\boldsymbol{\theta}$ y $\delta(\mathbf{X})$ un estimador de $q(\boldsymbol{\theta})$.

Supongamos $\mathbb{E}_{\boldsymbol{\theta}}|\delta(\mathbf{X})| < \infty$. Definamos un nuevo estimador

$$\eta(\mathbf{T}) = \mathbb{E}(\delta(\mathbf{X})|\mathbf{T}).$$

- Luego se tiene
  - (i) $ECM_{\boldsymbol{\theta}}(\eta) \leq ECM_{\boldsymbol{\theta}}(\delta), \ \forall \, \theta \in \Theta$
  - (ii) La igualdad en (i) se satisface si y sólo si

$$P_{\boldsymbol{\theta}}(\eta(\mathbf{T}) = \delta(\mathbf{X})) = 1 \quad \forall \, \boldsymbol{\theta} \in \Theta$$

  - (iii) Si $\delta(\mathbf{X})$ es insesgado, entonces $\eta(\mathbf{T})$ también lo es.

Otros resultados

Corolario.

Sea $\mathbf{X}$ un vector aleatorio con función de densidad o función de probabilidad puntual $p(\mathbf{x}, \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$. Supongamos que la familia $\mathcal{F} = \{p(\mathbf{x}, \boldsymbol{\theta})\}$ tiene soporte común, independiente de $\boldsymbol{\theta}$.

Entonces, una condición necesaria y suficiente para que $\mathbf{T}$ sea suficiente para $\boldsymbol{\theta}$ es que fijados $\boldsymbol{\theta}_1$ y $\boldsymbol{\theta}_2$ el cociente

$$\frac{p(\mathbf{x}, \boldsymbol{\theta}_1)}{p(\mathbf{x}, \boldsymbol{\theta}_2)}$$

sea función de $\mathbf{T}$.

Otros resultados

Corolario.

Sea $\mathbf{X}$ un vector aleatorio con función de densidad o función de probabilidad puntual $p(\mathbf{x}, \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$. Supongamos que la familia $\mathcal{F} = \{p(\mathbf{x}, \boldsymbol{\theta})\}$ tiene soporte común, independiente de $\boldsymbol{\theta}$.

Entonces, una condición necesaria y suficiente para que $\mathbf{T}$ sea suficiente para $\boldsymbol{\theta}$ es que fijados $\boldsymbol{\theta}_1$ y $\boldsymbol{\theta}_2$ el cociente

$$\frac{p(\mathbf{x}, \boldsymbol{\theta}_1)}{p(\mathbf{x}, \boldsymbol{\theta}_2)}$$

sea función de $\mathbf{T}$.

Lema. Si $\mathbf{X}$ es un vector aleatorio con una distribución $F(\mathbf{x}, \boldsymbol{\theta})$, con $\boldsymbol{\theta} \in \Theta$ si $\mathbf{T} = t(\mathbf{X})$ es un estadístico suficiente para $\boldsymbol{\theta}$ y si m es una función biunívoca de $\mathbf{T}$ entonces el estadístico $\mathbf{V} = m(\mathbf{T})$ también es suficiente para $\boldsymbol{\theta}$.

Minimal suficiente

Definición: Sea $\mathbf{X}$ un vector aleatorio con distribución $P_{\boldsymbol{\theta}}$ con $\boldsymbol{\theta} \in \Theta$.

Se dice que un *estadístico* $\mathbf{T} = t(\mathbf{X})$ *es minimal suficiente* para $\boldsymbol{\theta}$ si:

- $\mathbf{T}$ es suficiente para $\boldsymbol{\theta}$.
- Dado cualquier otro estadístico $\mathbf{U} = g(\mathbf{X})$ suficiente para $\boldsymbol{\theta}$ existe una función H tal que $\mathbf{T} = H(\mathbf{U})$.

Completitud

Definición. Sea $\mathbf{X}$ un vector aleatorio cuya distribución pertenece a una familia $\mathcal{P} = \{P_{\boldsymbol{\theta}} : \quad \boldsymbol{\theta} \in \Theta\}$.

Un estadístico $\mathbf{T} = t(\mathbf{X})$ se dice completo si

$$\mathbb{E}_{\boldsymbol{\theta}}(g(\mathbf{T})) = 0 \qquad \forall\, \boldsymbol{\theta} \in \Theta$$

implica que

$$\mathbb{P}_{\boldsymbol{\theta}}(g(\mathbf{T}) = 0) = 1 \qquad \forall\, \boldsymbol{\theta} \in \Theta$$

Resultados

Teorema de Basu. Supongamos que $\mathbf{X}$ tiene una distribución perteneciente a una familia de distribuciones $\mathcal{F} = \{F_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta\}$.

Supongamos que

- $\mathbf{T} = t(\mathbf{X})$ es suficiente y completo para $\mathcal{P}$.

Entonces,

Dado cualquier estadístico $\mathbf{U} = g(\mathbf{X})$ anciliario, o sea, tal que su distribución no depende de $\boldsymbol{\theta}$, se tiene que

$\mathbf{U}$ es independiente de $\mathbf{T}$.

Definición.

Se dice que una familia de distribuciones continuas o discretas en $\mathbb{R}^q$, $F_{\boldsymbol{\theta}}$, donde $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$ es una *familia exponencial a* k *parámetros* si su densidad o probabilidad puntual se puede escribir como

$$p(\mathbf{x}, \boldsymbol{\theta}) = A(\boldsymbol{\theta}) e^{\sum_{j=1}^{k} \eta_j(\boldsymbol{\theta})\, t_j(\mathbf{x})} h(\mathbf{x}) \qquad (1)$$

donde

- $\eta_j : \Theta \to \mathbb{R}$,
- $A : \Theta \to \mathbb{R}^+$
- $t_j : \mathbb{R}^q \to \mathbb{R}$
- $h : \mathbb{R}^q \to \mathbb{R}^+$.

Resultados

Teorema. Una familia exponencial a k parámetros cuya función de densidad viene dada por (1) tiene como estadístico suficiente para $\boldsymbol{\theta}$ el vector $\mathbf{T} = \mathbf{t}(\mathbf{X}) = (t_1(\mathbf{X}), \ldots, t_k(\mathbf{X}))$.

Teorema. Sea $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ una muestra aleatoria de una distribución que pertenece a una familia exponencial a k parámetros, cuya función de densidad viene dada por (1).

Luego,

- la distribución conjunta de $\mathbf{X}_1, \ldots, \mathbf{X}_n$ también pertenece a una familia exponencial a k parámetros y
- el estadístico suficiente para $\boldsymbol{\theta}$ es el vector

$$\mathbf{T}^* = (T_1^*, \ldots, T_k^*), \text{ donde } T_j^* = \sum_{i=1}^{n} t_j(\mathbf{X}_i), \quad 1 \leq j \leq k$$

Resultados

Teorema. Sea $\mathbf{X}$ un vector cuya distribución pertenece a una familia exponencial a k parámetros cuya función de densidad satisface (1).

Luego, la función de densidad de los estadísticos suficientes $\mathbf{T} = (t_1(\mathbf{X}), \ldots, t_k(\mathbf{X}))$ es de la forma

$$p_{\mathbf{T}}(t_1, t_2, \ldots, t_k, \boldsymbol{\theta}) = A(\boldsymbol{\theta}) e^{\sum_{j=1}^{k} \eta_j(\boldsymbol{\theta}) t_j} h^*(t_1, \ldots, t_k)$$

Por lo tanto, la familia de distribuciones de $\mathbf{T}$ también forma una familia exponencial a k parámetros.

Lema

Sea $\mathbf{X} = (X_1, \ldots, X_q)$ un vector aleatorio cuya distribución pertenece a una familia exponencial a un parámetro continua

$$p(\mathbf{x}, \boldsymbol{\theta}) = A(\theta) e^{\eta(\theta) t(\mathbf{x})} h(\mathbf{x})$$

- $\theta \in \Theta \subset \mathbb{R}$, Θ abierto
- $\eta(\theta)$ infinitamente derivable.

Luego, si $m(\mathbf{x})$ es un estadístico tal que

$$\int |m(\mathbf{x})| p(\mathbf{x}, \theta) d\mathbf{x} < \infty \quad \forall\, \theta \in \Theta$$

entonces, la expresión

$$\int \ldots \int m(\mathbf{x}) e^{\eta(\theta) t(\mathbf{x})} h(\mathbf{x}) dx_1 \ldots dx_q$$

es infinitamente derivable respecto de θ y se puede derivar dentro del signo integral.

En el caso discreto, se reemplazan las integrales por sumatorias.

Teorema

Sea $\mathbf{X} = (X_1, \ldots, X_q)$ un vector aleatorio cuya distribución pertenece a una familia exponencial a un parámetro con densidad dada por

$$p(\mathbf{x}, \theta) = A(\theta) e^{\eta(\theta) t(\mathbf{x})} h(\mathbf{x}) \qquad \theta \in \Theta \subset \mathbb{R}$$

con Θ abierto y $\eta(\theta)$ infinitamente derivable. Luego se tiene:

(i) $A(\theta)$ es infinitamente derivable.

(ii)

$$\mathbb{E}_\theta(t(\mathbf{X})) = -\frac{A'(\theta)}{A(\theta)\eta'(\theta)}$$

(iii)

$$\mathbb{V}_\theta(t(\mathbf{X})) = \frac{\frac{\partial \mathbb{E}_\theta(t(\mathbf{X}))}{\partial \theta}}{\eta'(\theta)}$$

Teorema

Sea una familia exponencial a k parámetros, discreta o continua con función de densidad dada por

$$p(\mathbf{x}, \boldsymbol{\theta}) = A(\boldsymbol{\theta}) e^{\sum_{j=1}^{q} \eta_j 1(\boldsymbol{\theta})\, t_j(\mathbf{x})} h(\mathbf{x})$$

y sea

$$\Lambda = \{\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_k)^{\mathrm{T}} \in \mathbb{R}^k : \lambda_j = \eta_j(\boldsymbol{\theta});\ \boldsymbol{\theta} \in \Theta\}$$

a) Si Λ contiene $k+1$ puntos $\boldsymbol{\lambda}^{(1)}, \ldots, \boldsymbol{\lambda}^{(k+1)}$ tales que $\{\boldsymbol{\lambda}^{(j)} - \boldsymbol{\lambda}^{(1)}\ , 2 \leq j \leq k+1\}$ son linealmente independientes, entonces el estadístico suficiente $\mathbf{T} = (t_1(\mathbf{X}), \ldots, t_k(\mathbf{X}))$ es minimal suficiente.

b) Si Λ contiene una esfera en $\mathbb{R}^k$, entonces el estadístico suficiente $\mathbf{T} = (t_1(\mathbf{X}), \ldots, t_k(\mathbf{X}))$ es completo.

Teorema de Lehmann-Scheffé

- Sea $\mathbf{X}$ un vector aleatorio de cuya distribución pertenece a la familia de distribuciones $\mathcal{F} = \{F_{\boldsymbol{\theta}} \quad \boldsymbol{\theta} \in \Theta\}$.

- Sea $\mathbf{T}$ un estadístico suficiente y completo para $\boldsymbol{\theta}$.

Luego, dada una función $q : \Theta \to \mathbb{R}$, se tiene que

(i) Existe a lo sumo un estimador insesgado de $q(\boldsymbol{\theta})$, basado en $\mathbf{T}$.

(ii) Si $\delta(\mathbf{T})$ es un estimador insesgado de $q(\boldsymbol{\theta})$, entonces $\delta(\mathbf{T})$ es IMVU.

(iii) Si $\eta(\mathbf{X})$ es un estimador insesgado para $q(\boldsymbol{\theta})$, luego $\delta(\mathbf{T}) = \mathbb{E}(\eta(\mathbf{X})|\mathbf{T})$ es un estimador IMVU para $q(\boldsymbol{\theta})$.

La pregunta que intentamos responder ahora es:

¿Un estimador insesgado puede tener una varianza arbitrariamente pequeña?

Hipotesis

Supongamos que $\mathbf{X} \sim f(\mathbf{x}, \theta)$, con $\theta \in \Theta$; Θ un conjunto abierto de $\mathbb{R}$.

(A) El conjunto $\mathcal{S} = \{\mathbf{x} : f(\mathbf{x}, \theta) > 0\}$ es independiente de θ.

(B) Para todo $\mathbf{x}$, $f(\mathbf{x}, \theta)$ es derivable respecto de θ.

(C) Si $h(\mathbf{X})$ es un estadístico tal que $\mathbb{E}_\theta[|h(\mathbf{X})|] < \infty$ para todo $\theta \in \Theta$ entonces se tiene

$$\frac{\partial}{\partial \theta} \int h(\mathbf{x}) p(\mathbf{x}, \theta) d\mathbf{x} = \int h(\mathbf{x}) \frac{\partial f(\mathbf{x}, \theta)}{\partial \theta} d\mathbf{x}$$

(D)

$$0 < I(\theta) = \mathbb{E}_\theta \left[\left(\frac{\partial \log f(\mathbf{X}, \theta)}{\partial \theta} \right)^2 \right] < \infty$$

$I(\theta)$ se denomina *número de información de Fisher*.

Lema

Supongamos que se cumplan las condiciones A, B, C y D. Sea

$$\psi(\mathbf{x},\theta) = \frac{\partial \log(f(\mathbf{x},\theta))}{\partial\theta}$$

Entonces,

(i) $\mathbb{E}_\theta \psi(\mathbf{X},\theta) = 0$ y $\mathbb{V}_\theta \psi(\mathbf{X},\theta) = I(\theta)$.

(ii) Si además existe la derivada segunda de $f(\mathbf{x},\theta)$ respecto de θ y si para todo estadístico $h(\mathbf{X})$ tal que, $\mathbb{E}_\theta[|h(\mathbf{X})|] < \infty$ para todo $\theta \in \Theta$, se cumple que

$$\frac{\partial^2}{\partial\theta^2}\int h(\mathbf{x})f(\mathbf{x},\theta)d\mathbf{x} = \int h(\mathbf{x})\frac{\partial^2 f(\mathbf{x},\theta)}{\partial\theta^2}d\mathbf{x} \tag{2}$$

entonces

$$I(\theta) = -\mathbb{E}_\theta\frac{\partial^2 \log f(\mathbf{X},\theta)}{\partial\theta^2} = -\mathbb{E}_\theta\frac{\partial\psi(\mathbf{X},\theta)}{\partial\theta}$$

Teorema de Rao–Cramer

Bajo las condiciones A, B, C y D, si $\delta(\mathbf{X})$ es un estimador insesgado de $q(\theta)$ tal que $\mathbb{E}_\theta \delta^2(\mathbf{X}) < \infty$ se tiene

(a)
$$\mathbb{V}_\theta(\delta(\mathbf{X})) \geq \frac{|q'(\theta)|^2}{I(\theta)}$$

(b) La desigualdad en (a) es una igualdad si y sólo si $\delta(\mathbf{X})$ es un estadístico suficiente de una familia exponencial, es decir, si y sólo si
$$f(\mathbf{x}, \theta) = A(\theta) e^{\eta(\theta)\delta(\mathbf{x})} h(\mathbf{x})$$

Teorema

Sea $\mathbf{X}_1, \ldots, \mathbf{X}_n$ una muestra aleatoria de una distribución con densidad $f(\mathbf{x}, \theta)$ con $\theta \in \Theta \subset \mathbb{R}$.

Luego, si se denomina

- $I_n(\theta)$ al número de información de $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ y
- $I_1(\theta)$ al número de información de $\mathbf{X}_1$,

entonces se tiene

$$I_n(\theta) = nI_1(\theta)$$

Teorema

Sean

- $\widehat{\theta}_n$ un estimador de máxima verosimilitud de θ consistente y
- $q(\theta)$ derivable con $q'(\theta) \neq 0$ para todo θ.

Entonces, $q(\widehat{\theta}_n)$ es A.N.E. para estimar $q(\theta)$, o sea,

$$\sqrt{n}\left(q(\widehat{\theta}_n) - q(\theta)\right) \xrightarrow{D} N\left(0, \frac{[q'(\theta)]^2}{I_1(\theta)}\right)$$

Teorema (Le Cam, 1953)

(D) $0 < I_1(\theta) = \mathbb{E}_\theta \left(\dfrac{\partial \log f(\mathbf{X}, \theta)}{\partial \theta} \right)^2 < \infty$

(E) Para cada θ_0 fijo, existen $c > 0$ y una función $M(\mathbf{x}) > 0$ tales que

$$\begin{aligned} \left| \frac{\partial^2 \log f(\mathbf{x}, \theta)}{\partial \theta^2} \right| = \left| \frac{\partial \psi(\mathbf{x}, \theta)}{\partial \theta} \right| &\leq M(\mathbf{x}) \\ \mathbb{E}_{\theta_0} M(\mathbf{X}_1) &< \infty \end{aligned}$$

para todo $\mathbf{x} \in \mathcal{S}$ y para todo $\theta \in \{\theta \in \Theta : |\theta_0 - \theta| < c\}$

Teorema (Le Cam)

Sea $\delta_n(X_1, \ldots, X_n)$ una sucesión de estimadores asintóticamente normales de $q(\theta)$, con q derivable y $q'(\theta) \neq 0$ para todo θ. Es decir,

$$\sqrt{n}\left(\delta_n(X_1, \ldots, X_n) - q(\theta)\right) \xrightarrow{D} N\left(0, \sigma^2(\theta)\right)$$

entonces

$$\sigma^2(\theta) \geq \frac{[q'(\theta)]^2}{I_1(\theta)}$$

excepto en un conjunto $\Theta_0 \subset \Theta$ tal que Θ_0 tiene medida 0.