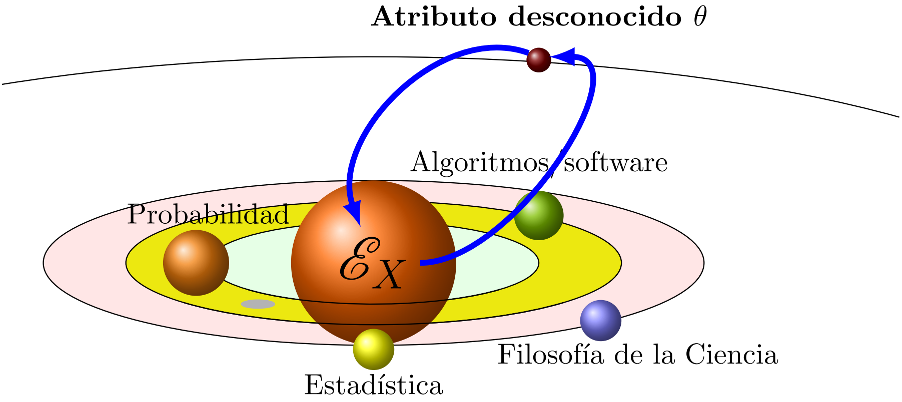
Teoría Estadística Avanzada (DES124). Unidad I

Dr. Jaime Lincovil

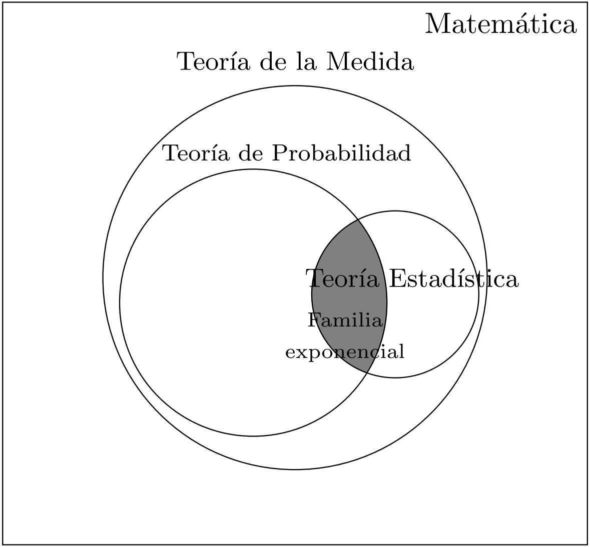
Universidad Nacional de Ingeniería

2025

Métodos de Inferencia Estadística



Teoría de Inferencia Estadística



Notación

Letras

- Letras caligraficas $\mathcal{G}$, $\mathcal{F}$ y $\mathcal{F}_\Theta$.
- Letras griegas: θ, Θ, π, Π, λ (lambda), Λ, η (eta), ξ (xi), ω (omega), Ω, ζ (Zeta) y ν (nu).
- $(\Omega, \mathcal{F}, P_0)$ espacio de probabilidad original.
- $X^{-1}(A)$es la imagen inversa de la variable X, $\sigma(X)$: sigma álgebra generada por X.
- $\mathcal{R}$, y $\mathcal{R}^k$: conjunto de los números reales y el conjunto euclideano de dimensión k
- $\frac{dP}{d\nu}$: derivada de Radon-Nykodym de P con respecto a ν.

Espacio de Probabilidad

- Un **espacio de probabilidad** es una tripla $(\Omega, \mathcal{F}, P_0)$ en que Ω es el espacio muestral, $\mathcal{F}$ es una σ-algebra de Ω y P_0 una medida de probabilidad sobre $\mathcal{F}$ conocida.
- Sea Ω y Λ conjuntos nno vacíos y X una función

$$X : \Omega \to \Lambda.$$

La función X es llamada **variable aleatoria** de $(\Omega, \mathcal{F})$ a $(\Lambda, \mathcal{G})$ si y solo si

$$X^{-1}(\mathcal{G}) = \left\{X^{-1}(G) : G \in \mathcal{G}\right\} \subset \mathcal{F}.$$

$X^{-1}(\mathcal{G})$ es llamada la **imagen inversa de** $\mathcal{G}$ **sobre** X.
- El modelo de probabilidad inducido por X es la tripla

$$(\Lambda, \mathcal{F}, P_X),$$

en que P_X es conocida. En el modelo de probabilidad, la incertidumbre esta en evento $X = x$.

Medidas y densidades

- Si $\nu(A) = 0$ implica $P(A) = 0$, entonces decimos que P es absolutamente continua con respecto a ν denotado por $P \ll \nu$
- Si P es una medida de probabilidad sobre $\mathcal{B}$ (o $\mathcal{B}^k$), entonces

$$f(x) = \frac{dP}{d\nu}: \ \textbf{densidad} \text{ de } P \text{ con respecto a la medida de Lebesgue } \nu.$$

- Si P es una medida de probabilidad sobre $\mathcal{A}$ (discreto), entonces

$$p(x) = \frac{dP}{d\nu}: \ \textbf{densidad} \text{ de } P \text{ con respecto a la medida de Conteo } \nu.$$

- Una función $g : \mathcal{R} \to \mathcal{R}$ es llamada de Boreleana si sus imagenes inversas estan contenidas en $\mathcal{B}$., i.e. $g^{-1}(A) \in \mathcal{B}$ para todo $A \subset \mathcal{R}$.

Inferencia $\rightarrow$ Teoría de Estadística

- Consideremos una variable aleatoria "$X \sim \mathrm{N}(\mu, \sigma^2)$" con espacio muestral $\Lambda = \bar{\mathcal{R}}$ y espacio paramétrico $\Theta = (-\infty, +\infty) \times (0, +\infty)$.
- Por "$X \sim \mathrm{N}(\mu, \sigma^2)$", entendemos que X tiene una función de densidad f_{μ,σ^2}.
- En **Inferencia Estadística**, dada una muestra $x = (x_1, x_2, \ldots, x_n)$ de X, una estimación putual consiste en encontrar una función tal que:

$$(x, f) \rightarrow (\mu_0, \sigma_0^2) \in \Theta.$$

- En **Teoría Estadística**, por "$X \sim \mathrm{N}(\mu, \sigma^2)$" entendemos que X tiene una una medida de probabilidad P_{μ,σ^2}. En estimación puntual buscaremos una función tal que

$$(x, P_{\mu,\sigma^2}) \rightarrow P_{\mu_0,\sigma_0^2} \in \mathcal{P}.$$

Población, muestra y modelos estadísticos. Familias exponenciales

Medidas de probabilidad Normal y de Poisson

- Sea X una variable con densidad normal, media μ y varianza σ^2, luego la medida de probabilidad definida por $\mathrm{P}_{\mu,\sigma^2} : \mathcal{B} \to [0,1]$ es dada por

$$\mathrm{P}_{\mu,\sigma^2}(A) = \int_A f_{\mu,\sigma^2}(x)dx.$$

- Consideremos un vector $(X_1, \ldots, X_n)$ de variables i.i.d con densidad normal con media μ y varianza σ^2, luego la medida de probabilidad conjunta $\mathrm{P}'_{\mu,\sigma^2} : \mathcal{B}^n \to [0,1]$ es dada por

$$\mathrm{P}'_{\mu,\sigma^2}(A_1 \times \ldots \times A_n) = \mathrm{P}_{\mu,\sigma^2}(A_1) \times \ldots \times \mathrm{P}_{\mu,\sigma^2}(A_n).$$

- Consideremos un vector $(X_1, \ldots, X_n)$ de variables i.i.d con función de probabilidad Poisson de parametro λ, luego la medida de probabilidad conjunta es definida por $\mathrm{P}'_\lambda : \mathcal{A}' \to [0,1]$ es dado por

$$\mathrm{P}'_\lambda(A_1 \times \ldots \times A_n) = \mathrm{P}_\lambda(A_1) \times \ldots \times \mathrm{P}_\lambda(A_n) = \prod_{i=1}^{n} \Big[\sum_{x \in A_i} \frac{e^{-\lambda}\lambda^x}{x!} \Big].$$

El problema Estadístico

En Teoría Estadística los *datos* se consideran una observación de una variable aleatoria $X : \Omega \to \Lambda$ modelado por $(\Lambda, \mathcal{G}, P_X)$ definido acorde el experimento aleatorio.

Definición 1

- Los datos o evento observado $\mathbf{x}$ será llamado de una **muestra de una medida de probabilidad** P.
- La *verdadera* medida de probabilidad que genero los datos x será llamada de **población** de x.
- **El problema estadístico: inferir** propiedades de la población P basado una muestra $\mathbf{x} = (x_1, \ldots, x_n)$ observada en un experimento.

Modelo estadístico paramétrico

Definición 2. Modelo Estadístico Clásico

Una clase $\{P_\theta\}$ en $(\Lambda, \mathcal{G})$ indexada por un **parámetro** $\theta \in \Theta$ es una **familia paramétrica** si y solo si $\Theta \subset \mathcal{R}^d$ para $d \in \mathcal{N}_+$ y cada P_θ es una medida de probabilidad totalmente especificada. Θ es llamado **espacio paramétrico** de dimensión d. Un **modelo estadístico paramétrico** para los eventos en $(\Lambda, \mathcal{G})$ es la tripla

$$\Big(\Lambda, \mathcal{G}, \{P_\theta : \theta \in \Theta\}\Big). \tag{1}$$

- Un modelo paramétrico asume que la población P pertenece a una familia de medidas de probabilidad **paramétrizada** o **indexada** por $\theta \in \Theta$.
- Una familia paramétrica $\{P_\theta : \theta \in \Theta\}$ es **identificable** si y solo si $\theta_1 \neq \theta_2$ y $\theta_i \in \Theta$ implican $P_{\theta_1} \neq P_{\theta_2}$.
- Si la familia está dominada por ν, entonces esta puede ser identificada por la familia de densidades $\{f_\theta\} = \left\{\frac{dP}{d\nu} : P_\theta \in \mathcal{P}\right\}$ de P con respecto a ν.

El Problema Estadístico y notación

Sea $(X_1, \ldots, X_n)$ un vector aleatorio con una familia de probabilidad $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$. Adicionalmente, sea Λ^n el espacio muestral del vector, T una funcion sobre Λ^n y I_Θ la clase de todos los intervalos contenidos en Θ.

- (Estimación puntual) Encontrar una función $T : \Lambda^n \to \Theta$ con propiedades óptimas para asi determinar la población $P_{T(x)}$.
- (Estimación intervalar) Encontrar una función $T : \Lambda^n \to I_\Theta$ con propiedade óptimas.
- (Prueba de hipótesis) Sea $\mathcal{P}_0$ y $\mathcal{P}_1$ dos subfamilias disjuntas de $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$. y las hipótesis $H_0 : P \in \mathcal{P}_0$ contra $H_1 : P \in \mathcal{P}_1$. Encontrar una función $T : \Lambda^n \to \{0, 1\}$ tal que $T(x) = 1$ implica rechazar H_0 y $T(x) = 0$ implica no rechazar H_0.

Familia Exponencial

Definición 3 A

Una familia paramétrica $\{P_\theta : \theta \in \Theta\}$ dominada por una medida σ-finita ν en $(\Lambda, \mathcal{G})$ se llama **familia exponencial** si y solo si

$$\frac{dP_\theta}{d\nu}(x) = \exp\left\{\eta^\top(\theta)T(x) - \xi(\theta)\right\} h(x), \quad x \in \Lambda, \tag{2}$$

donde $\exp\{x\} = e^x$, T y $\eta(\theta)$ son vectores de dimensión p, η es una función de θ y T función de X, h es una función Boreliana no negativa y $\xi(\theta)$ es llamada de constante de normalización de la densidad[a].

[a] T y h son funciones de x. η y ξ son funciones de θ.

- La función T tiene la forma $T(x) = (T_1(x), T_2(x), \ldots, T_p(x))$ y η tiene la forma $\eta(\theta) = (\eta_1(\theta), \eta_2(\theta), \ldots, \eta_p(\theta))$.
- Notemos que $\eta^\top(\theta)T(x) = \sum_{k=1}^{p} \eta_k(\theta)T_k(x)$.

Parametrización natural

Definición 3 B

La reparametrización $\eta = \eta(\theta)$ es llamada de **parametrización natural** con $\eta \in \Xi$ y

$$f_\eta(x) = \exp\left\{\eta^\top T(x) - \zeta(\eta)\right\} h(x), \quad \textbf{densidad canonica},$$

$x \in \Lambda$, en que $\zeta(\eta)$ es una función de η obtenida al invertir $\xi(\theta)$. El conjunto $\Xi = \{\eta(\theta) : \theta \in \Theta\}$ es llamado de **espacio parametrico natural**.

(a) Si $\dim(\theta) < p$, la medida de probabilidad pertenece a una **familia exponencial curvada**.
(b) Si $\dim(\theta) = p$, la medida de probabilidad pertenece a una **familia exponencial de rango completo**.

Modelo de Poisson

Sea P_θ una familia de medidas de probabilidad Poisson(θ) con parámetro $\theta > 0$. La función de probabilidad de P_θ puede ser rescrita por:

$$f_\theta(x) = \exp\left\{\log(\theta)x + \theta\right\} \frac{1}{x!} I_{\mathcal{N}}(x)$$

Aqui $p = 1$, $\eta(\theta) = \log(\theta)$, $T(x) = x$, $\xi(\theta) = \theta$ and $h(x) = \frac{1}{x!} I_{\mathcal{N}}(x)$.

Consideremos la reparametrización natural $\eta = \log(\theta)$ y $\zeta(\eta) = \exp(\eta)$. El espacio paramétrico natural es $\Xi = \mathcal{R}_+$. Luego:

$$f_\eta(x) = \exp\{\eta x - \exp(\eta)\} h(x).$$

Dado que $\dim(\theta) = \dim(\eta) = 1$, el modelo de Poisson es de rango completo.

Modelo Normal

Consideremos la familia de medidas de probabilidad Normal $\left(\mu, \sigma^2\right)$ pertenece a la familia exponencial, desde que su función de densidad puede ser escrita

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2 - [\frac{\mu^2}{2\sigma^2} + \log\sigma]\right\}.$$

Tenemos que $p = 2$, $T(x) = \left(x, -x^2\right), \eta(\theta) = \left(\frac{\mu}{\sigma^2}, \frac{1}{2\sigma^2}\right), \theta = \left(\mu, \sigma^2\right), \xi(\theta) = \frac{\mu^2}{2\sigma^2} + \log\sigma$, and $h(x) = 1/\sqrt{2\pi}$.

Sea $\eta = (\eta_1, \eta_2) = \left(\frac{\mu}{\sigma^2}, \frac{1}{2\sigma^2}\right)$. Entonces, $\Xi = \mathcal{R} \times (0, \infty)$ y podemos obtener una familia exponencial natural f_η con $\zeta(\eta) = \eta_1^2/\left(4\eta_2\right) + \log\left(1/\sqrt{2\eta_2}\right)$. Dado que θ y η tienen dimensión 2, la familia natural es de rango completo.

Nota: La familia $N(\mu, \mu^2)$ no tiene rango completo (ejercicio).

Modelo Binomial

Consideremos una familia de medidas de probabilidad Binomial(θ, n) con parametro θ, en que n conocido y fijo. La función de probabilidad para P_θ es

$$f_\theta(x) = \exp\left\{x \log \frac{\theta}{1-\theta} + n\log(1-\theta)\right\} \left(\begin{array}{c} n \\ x \end{array} \right) I_{\{0,1,\ldots,n\}}(x)$$

en que $p = 1$, $T(x) = x$, $\eta(\theta) = \log \frac{\theta}{1-\theta}$, $\xi(\theta) = -n\log(1-\theta)$, y $h(x) = \binom{n}{x} I_{\{0,1,\ldots,n\}}(x)$. Considerando $\eta = \log \frac{\theta}{1-\theta}$, then $\Xi = \mathcal{R}$ con función de densidad

$$f_\eta(x) = \exp\left\{x\eta - n\log\left(1+e^\eta\right)\right\} \left(\begin{array}{c} n \\ x \end{array} \right) I_{\{0,1,\ldots,n\}}(x)$$

es una familia exponencial de rango completo.

Modelo Estadístico Bayesiano

Definición 4

Sea $(\Lambda, \mathcal{G})$ y $(\Theta, \mathcal{F}_\Theta)$ dos espacios medibles, $X : \Omega \to \Lambda$ y $\theta : \Omega \to \Theta$ variables aleatorias. El modelo **estadístico Bayesiano** es una familia de medidas de probabilidad conjunta para el vector aleatorio $(X, \theta) \sim P_{X\theta}$. La **familia paramétrica** para X es

$$\mathcal{P} = \{P_{\theta_0} : P_{\theta_0}(A) = P(X \in A | \theta = \theta_0), \quad \theta_0 \in \Theta\}. \tag{3}$$

Si cada $P_{\theta_0} << \nu$, luego la densidad es $f(x|\theta = \theta_0) := f(x|\theta_0) = \frac{dP_{\theta_0}}{d\nu}$. Por tanto, podemos escribir:

$$P_{\theta_0}(A) = \int_A f(x|\theta_0)dx. \tag{4}$$

La densidad conjunta de (X, θ) es dada por:

$$f(x, \theta) = \pi(\theta) f(x|\theta) \text{ o } f(\mathbf{x}, \theta) = \pi(\theta) f(\mathbf{x}|\theta).$$

Modelo Bayesiano a posteriori

Definición 5

Un modelo de probabilidad **a priori** para θ es $(\Theta, \mathcal{F}_\Theta, \Pi_\theta)$, en que $\Pi_\theta(B) = \int_B \pi(\theta)d\theta$ y $\pi(\theta)$ es la densidad a priori de θ (o probabilidad cambiando la integral por una sumatoria).

Luego de observar $\mathbf{X} = \mathbf{x}$, el **modelo Bayesiano a posteriori** de $\theta|\mathbf{X} = \mathbf{x}$ es la tripla $(\Theta, \mathcal{F}_\Theta, P_{\theta|\mathbf{X}=\mathbf{x}})$ en que

$$P_{\theta|\mathbf{X}=\mathbf{x}}(B) = \int_B \pi(\theta|\mathbf{X} = \mathbf{x})d\theta \text{ para } B \in \mathcal{F}_\Theta. \tag{5}$$

y $\pi(\theta|\mathbf{X} = \mathbf{x})$ es la densidad a posteriori de θ dado la muestra observada $\mathbf{X} = \mathbf{x}$.

Por el teorema de Bayes, la densidad a posteriori se rescribe:

$$\pi(\theta|\mathbf{X} = \mathbf{x}) = \frac{\pi(\theta)f(\mathbf{x}|\theta)}{\int_\Theta \pi(\theta)f(\mathbf{x}|\theta)d\theta}, \quad g(\mathbf{x}) = \int_\Theta \pi(\theta)f(\mathbf{x}|\theta)d\theta \text{ marginal de } \mathbf{x}.$$

Ejemplo de modelo Bernoulli y Poisson

- Consideremos que el parametro de interes es la proporción $\theta \in (0,1)$ y que a priori (antes de observar los datos) le provemos de un modelo Beta(a, b). Supongamos, que $X_i|\theta = \theta_0 \sim$ Bernoulli(θ_0). El modelo de probabilidad a posteriori de $\theta|\mathbf{X} = \mathbf{x}$ es la medida de probabilidad Beta$(\sum_{i=1}^{n} x_i + a; n - \sum_{i=1}^{n} + b)$.
- Supongamos que el parametro de interes es la taza de ocurrencia $\theta > 0$ y que a priori (antes de observar los datos) le provemos de un modelo Gama(a, b). Supongamos, que $X_i|\theta = \theta_0 \sim$ Poisson(θ_0). El modelo de probabilidad a posteriori de $\theta|\mathbf{X} = \mathbf{x}$ es la medida de probabilidad Gama $(\sum_{i=1}^{n} x_i + a; n + b)$.
- Un modelo a priori para θ, representado por π, es **conjugado** con $f(\mathbf{x}|\theta)$ cuando $\pi(\theta)$ y $\pi(\theta|\mathbf{X} = \mathbf{x})$

Ejemplo de modelo Normal

- Si el parametro de interes es la media $\theta \in \mathcal{R}$ y que a priori (antes de observar los datos) le provemos de un modelo Normal(a, b). Supongamos, que $X_i|\theta = \theta_0 \sim$ Normal(θ_0, σ_0^2) (σ_0^2 es conocido). El modelo de probabilidad a posteriori de $\theta|\mathbf{X} = \mathbf{x}$ es la medida de probabilidad

$$\text{Normal}\left\{\frac{\sum_{i=1}^{n} \frac{x_i}{\sigma_o^2} + \frac{a}{b^2}}{\frac{n}{\sigma_0^2} + \frac{1}{b^2}}; \frac{1}{\frac{n}{\sigma_0^2} + \frac{1}{b^2}}\right\}.$$

Modelo no parametrico

Un **modelo estadístico no paramétrico** es una tripla

$$\Big(\Lambda, \mathcal{G}, \{P_s : s \in S\}\Big),$$

en que s es una **función** que pertenece a una clase de funciones S bien definida. Por ejemplo, el espacio L^2 en los reales o en el intervalo unitario.

- Sea $\epsilon \sim N(0, \sigma^2)$, una variable X de modelo desconocido e $Y = s(X) + \epsilon$, en que s es una función Boreleana en L^2. El modelo para $Y|X \sim N(s(X), \sigma^2)$ es no paramétrico. El problema estadístico consiste en estimar la función $s(X)$.
- Sea $H : \mathcal{R} \to [0, 1]$ conocida (por ejemplo Φ o la Logística) y s antes definida. Un modelo no paramétrico para una variable Bernoulli $Y|X$ considera una función de probabilidad

$$p_s(y) = H(s(x))^y [1 - H(s(x))]^{1-y}.$$

Ejemplos de modelo estadístico no paramétrico 1

- Consideremos una familia de funciones de probabilidad acumulada (f.d.a) $\mathcal{L} = \left\{ F : \int_{-\infty}^{\infty} x dF(x) < \infty \right\}$. El modelo de probabilidad indexado por $\mathcal{L}$ es la tripla $(\mathcal{R}, \mathcal{B}, \{P_F : F \in \mathcal{L}\})$, en que

$$\left\{ P_F : P_F(A) = \int_A dF(x) \quad A \subset \mathcal{R} \text{ y } F \in \mathcal{L} \right\}.$$

  Podemos estimar $\theta = \int_{-\infty}^{\infty} x dF(x)$ por $\widehat{\theta} = n^{-1} \sum_{i=1}^{n} X_i$.

- Consideremos una familia de f.d.a $\mathcal{L} = \left\{ F : \int |x|^2 dF(x) < \infty \right\}$. El modelo de probabilidad indexado por $\mathcal{L}$ es la tripla $(\mathcal{R}, \mathcal{B}, \{P_F : F \in \mathcal{L}\})$. Podemos estimar el funcional

$$\theta = \int \int 2^{-1} (x_1 - x_2)^2 dF(x_1) dF(x_2) \quad \text{(varianza de } F\text{)}.$$

  por $\widehat{\theta} = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n}^{n} 2^{-1} (X_i - X_j)^2$.

Ejemplos de modelo estadístico no paramétrico 2

- Consideremos una familia de f.d.a bivariada

$$\mathcal{L} = \left\{ F : \int |xy|^2 dF(x, y) < \infty \right\}.$$

El modelo de probabilidad indexado por $\mathcal{L}$ es la tripla $(\mathcal{R}, \mathcal{B}, \{P_F : F \in \mathcal{L}\})$. Podemos estimar el funcional

$$\theta = \int_{\mathcal{R}} \int_{\mathcal{R}} 2^{-1}(x_1 - x_2)(y_1 - y_2) dF(x_1, y_2) dF(x_2, y_2) \quad \text{(cov. de F)}$$

por $\widehat{\theta} = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n}^{n} 2^{-1}(X_i - X_j)(Y_i - Y_j)$.

Estadísticos, Suficiencia, Ancilaridad y Completitud

Estadístico y disminución de dimensión

Sea $(\Omega, \mathcal{F})$ el espacio medible original del experimento y $(\Lambda, \mathcal{G})$ el espacio medible del vector aleatorio $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ definido sobre $(\Omega, \mathcal{F})$. Una operación de reducción es una función

$$\mathbf{X} \to T, \quad \dim(T) < \dim(\mathbf{X}).$$

Ejemplos:

- Reducción de una muestra: $(x_1, x_2, \ldots, x_n) \in \mathcal{R}^n$ al escalar $x \in \mathcal{R}$.
- Reducción de un matriz: $X_{n\times p}$ al vector $(\bar{x}_1, \ldots, \bar{x}_p)$, en que $\bar{x}_i$ es el promedio de la i-ésima columna de $X_{n\times p}$.
- Reducción de un matriz por condicionamiento: $X_{n\times p}|A = (1, p)$ reduce la matriz a $Y_{2\times(p-1)}$. El estadístico A será conecido como estadístico **ancilar**.

Estadístico y sub σ-álgebra generada

Sea $(\Omega, \mathcal{F})$ el espacio medible original del experimento y $(\Lambda, \mathcal{G})$ el espacio medible del vector X.

Definición 6

Consideremos el espacio medible $(\Phi, \mathcal{T})$, una función mensurable $T : \Lambda \to \Phi$ es llamado de **estadístico**, esto es, siempre que $T(\mathcal{T})^{-1} \subset \mathcal{F}$.

Resultado de la Teoría de la medida de Probabilidad.

$$\sigma(T(x)) \subset \sigma(X) \subset \mathcal{F}.$$

- La media $\bar{X}_n$ y varianza muestral S^2.
- El estadístico conjunto $T = (\bar{X}, S^2)$.
- El vector de estadísticos de orden $T = (X_{(1)}, X_{(2)}, \ldots, X_{(n)})$.
- Una clase de equivalencia puede ser un estadístico. Por ejemplo, sea $T(x) = [x] = \{y \in \mathcal{X} : \exists k, f_\theta(x) = kf_\theta(y)\ \forall \theta \in \Theta\}$. Es decir, $T(x)$ es el conjunto de todos puntos del espacio muestral con función de densidad proporcional a $f_\theta(x)$.

Sufficiencia Clásica

Definición 7. Suficiencia clásica

Sea X un vector aleatorio de una población desconocida $P_\theta \in \mathcal{P}$, donde $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$. Un estadístico $T(X)$ es **suficiente para** $\theta \in \Theta$ en el sentido **clásico** si y solo si la densidad de P_θ para X condicional a $T = t$ no depende de θ, o equivalentemente, es invariante con respecto a θ.

Ejemplos:

- Modelo Bernoulli: $T(X) = \sum_{i=1}^{n} X_i$.
- Modelo Poisson: $T(X) = \sum_{i=1}^{n} X_i$.
- Modelo Normal: $T(X) = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$.
- Modelo Gamma: $T(X) = \left(\prod_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i\right)$.

Criterio de Factorización

Teorema 1 A. Criterio de Factorización

Supongamos que x es una muestra de $P_\theta \in \mathcal{P}$ y $\mathcal{P}$ es una familia de medidas en $(\mathcal{R}^n, \mathcal{B}^n)$ dominada por medida σ-finita ν. Entonces, $T(X)$ es suficiente para $P_\theta \in \mathcal{P}$ si y solo si existe una funcion medible de Borel nonegativa h (que no depende de θ) sobre $\mathcal{R}^n$ and g_θ (la cual depende de θ) sobre el rango de T tal que

$$\frac{dP_\theta}{d\nu}(x) = g_\theta(T(x))h(x).$$

Si $\mathcal{P}$ es una familia exponencial y $X(\omega) = x$, por el teorema, $g_\theta(t) = \exp\left\{[\eta(\theta)]^\top T(x) - \xi(\theta)\right\}$ implica que T es suficiente para $\theta \in \Theta$.

Vector de estadísticos de orden

Sea $X_1, \ldots, X_n$ sean v.a i.i.d. con medida $P \in \mathcal{P}$, en que $\mathcal{P}$ es la familia de medidas en $\mathcal{R}$ que posee f.d.d de Lebesgue. El estadístico $T(X) = \left(X_{(1)}, \ldots, X_{(n)}\right)$ es suficiente para $\theta \in \Theta$.

Definición 8. Suficiencia minimal

Sea T un estadístico suficiente para $\theta \in \Theta$. Luego, T es llamado de **estadístico suficiente minimal** si y solamente si
(a) T es suficiente para $\theta \in \Theta$;
(b) Para cualquier otro estadístico suficiente S para $\theta \in \Theta$, existe una función medible ψ tal que $T = \psi(S)$ casi ciertamente en $\mathcal{P}$.

Si ambos T y S son estadísticos suficiente minimal, entonces, por definición existe una función medible 1 a 1 ψ tal que $T = \psi(S)$ c.c $\mathcal{P}$. es decir, los estadísticos mínimos suficientes son **únicos** en el sentido de que dos de ellas funciones uno a uno del otro con probabiliodad 1.

Ejemplo de estadísticos suficientes minimal para modelos.

- Modelo Bernoulli: $T(X) = \sum_{i=1}^{n} X_i$.
- Modelo Uniforme$(0, \theta)$: $T(X) = \max\{X_1, \ldots, X_n\} = X_{(n)}$.
- Modelo Poisson: $T(X) = \sum_{i=1}^{n} X_i$.
- Modelo Normal: $T(X) = (\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2)$.

Criterio de Factorización

Teorema 1 B. Criterio para un estadístico suficiente minimal

Sea $f_\theta(x)$ la función de densidad o de probabilidad de un vector X y una estadística T tal que:
$T(x) = T(y)$ si y solamente si $y \in D(x)$, en que

$$D(x) = \{y \in \Lambda : f_\theta(y) = f_\theta(x)h(x,y), \quad \theta \in \Theta; h(x,y) > 0\}.$$

Entonces, tal T es un estadístico suficiente minimal.

Ejemplos

- Modelo Bernoulli: $h(x,y) = 1$.
- Modelo Normal:
- Modelo Exponencial:

Interpretaciones de la Suficiencia

- Modelo Binomial: $T(x) = \sum_{i=1}^{n} x_i = t$. La suma total t de caras observadas al lanzar n veces una moneda contiene toda la información necesaria para estimar la probabilidad $\Pr(X_i = 1) = \theta$.
- Modelo Uniforme: $T(X) = X_{(n)}$. El máximo observado $x_{(n)}$ contiene contiene toda la información necesaria para estimar θ.

  En inferencia estadística, un estadístico suficiente es utilizado para la construcción de estimadores que son funciones de este.

Principio de Suficiencia*

Dada una muestra $x = (x_1, \ldots, x_n)$ de una variable X con modelo estadístico que contempla un estadistico suficiente (mínimal) T para $\theta \in \Theta$. Entonces, los estimadores puntuales, intervalares y estadístico de prueba de un contraste hipótesis inferencia acerca de θ deben ser funciones de T. [Reid, N., & Cox, D. (2013). Principles of Statistical Inference].

Estadísticos ancilares

Definición 9. Estadísticos ancilares de primer y segundo orden.

Un estadístico $V(X)$ es llamado de **ancilar de primer orden** si su medida de probabilidad no depende de θ y será llamada de **ancilar de segundo orden** si $E[V(X)]$ no depende de θ.

Consideremos una secuencia de variables aleatorias $X_1, \ldots, X_n$ en que X_i tiene medida Normal(μ, σ^2).

- $\frac{(n-1)S^2}{\sigma^2}$ con medida Chi$(n-1)$ que no depende de $\theta = (\mu, \sigma^2)$.
- $\frac{\sqrt{n}(\bar{X}-\mu)}{S}$ con medida Student$(n-1)$ no depende de $\theta = (\mu, \sigma^2)$.
- $\sum_{i=1}^{n} \frac{(X_i-\mu)^2}{\sigma^2}$ con medida Chi(n) no depende de $\theta = (\mu, \sigma^2)$.
- Estadístico F.

Interpretaciones de la Ancilaridad

- Modelo Normal. La construcción de un intervalo de confianza para μ debe ser basado condicionando en el valor observado de $\frac{\sqrt{n}(\bar{X}-\mu)}{S}$ sin agregar sesgo a la estimación.
- La construcción de un intervalo de confianza para σ debe ser basado condicionando en el valor observado de $\frac{(n-1)S^2}{\sigma^2}$ sin agregar sesgo a la estimación.
- En inferencia estadística, un estadistico ancilar es empleado para condicionar en el y trabajar con las distribución condicional.

Principio de Condicionalidad*

Dada una muestra $x = (x_1, \ldots, x_n)$ de una variable X con modelo estadístico que contempla un estadistico suficiente (mínimal) T y un estadístico ancilar A para $\theta \in \Theta$. Entonces, los estimadores puntuales, intervalares y estadístico de prueba de hipótesis de inferencia acerca de θ deben ser funciones de $T|A$ (condicional).

[Reid, N., & Cox, D. (2013). Principles of Statistical Inference]

Estadístico Completo

Definición 10. Completitud

Un estadístico $T(X)$ es llamado de **completo** para $\theta \in \Theta$ en relación a una familia $\mathcal{P}$ si y solo si para cualquier función Boreleana f, $E_\theta[f(T)] = 0$ para todo $\theta \in \Theta$ implica $f(T) = 0$ casi ciertamente en $\mathcal{P}$. Adicionalmente, T es llamado de **limitadamente completo** si el requerimiento anterior es verdadero solo para funciones Boreleanas limitadas f.

Nota: Un estadístico completo y suficiente podría ser minimal, lo cual es provado por Lehmann and Scheffé (1950) and Bahadur (1957) (ver Exercise 48). Sin embargo, un estadístico suficiente minimal no es necesariamente completo.

Intuitivamente, **un estadístico completo no posee información ancilar dentro de si**, es decir, contiene alta información necesaria para estimar θ.

Goal: Estadístico suficiente, minimo y completo***.

Completitud en familias exponenciales

Proposición 1.

(a) Si P_θ pertenece a la familia exponencial con rango completo con la densidad canonica f_η, entonces $T(X)$ es completo y suficiente para $\eta \in \Xi$.
(b) Si existe una correspondencia 1-1 entre η y θ ($\eta(\theta)$ es inyectora), entonces T es también completo para θ.

Ejemplo de estadísticos completos para modelos.

- Modelo Bernoulli: $T(X) = \sum_{i=1}^{n} X_i$ es completo para θ.
- Modelo Poisson: $T(X) = \sum_{i=1}^{n} X_i$ es completo para θ.
- Modelo Normal: $T(X) = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ es completo para $\eta = \left(\frac{\mu}{\sigma^2}, \frac{1}{2\sigma^2}\right)$ la cual es una función inyectora, por lo tanto, también es completa para $\theta = (\mu, \sigma^2)$.
- (Demostración directa) Modelo Uniforme$(0, \theta)$: $T(X) = X_{(n)}$ es completa para θ.

Teorema 2. Teorema de Basu.

Sea V y T dos estadísticos de X con población $P_\theta \in \mathcal{P}$. Si V es ancilar y T es limitadamente completa y suficiente para $\theta \in \Theta$, entonces V y T son independientes con respecto cualquier $P \in \mathcal{P}$.

El Teorema de Basu es útil para probar la independencia, si fue el caso, de dos estadísticos.

- Para una muestra $X_1, X_2, \ldots, X_n$ del modelo Normal(μ, σ^2), las estadísticas $\bar{X}$ y S^2 son independientes.

Suficiencia Bayesiana

Recordemos que $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad y $(\Lambda, \mathcal{G})$ y (Θ, ν) dos espacios medibles y sean $X : \Omega \to \Lambda$ y $\theta : \Omega \to \Theta$ funciones medibles. Tambien recordemos que $P_{\theta_0}(A) = P(X \in A | \theta = \theta_0)$.

Definición 11. Suficiencia Bayesiana.

El estadístico $T(X)$ es suficiente (en el sentido **Bayesiano**) para θ si y solamente si, para toda medida de probabilidad a priori Π_θ,

$$P_{\theta|\mathbf{X}=\mathbf{x}}(B) = P_{\theta|T(\mathbf{x})=t}(B), \text{ para todo } B \text{ tal que } P_{\mathbf{X}}(B) \neq 0,$$

en que $P_{\mathbf{X}}$ es la medida marginal de $\mathbf{X}$.

Un criterio operacional derivado del anterior de este Teorema es el siguiente: T cumple la Definición 11 si y solo si

$$\pi(\theta|\mathbf{x}) = \pi(\theta|T(\mathbf{x}) = t), \text{ casi ciertamente.}$$

Ver Lema 2.6 en Schervish (1995), página 85.

Equivalencia entre Suficiencia Clásica y Bayesiana

Teorema 2. Equivalencia entre suficiencias

Sea $(\Phi, \mathcal{T})$ un espacio medible y T un estadístico. Supongamos que existe una medida σ-finita ν tal que $P_{X|\theta=\theta_0}(\cdot \mid \theta_0) \ll \nu$ para todo $\theta_0 \in \Theta$. Entonces, T es suficiente en el sentido clásico si y solo si T es suficiente en el sentido Bayesiano.

Ejemplo de estadísticos suficientes para el modelo clásico y Bayesiano.

- Modelo Bernoulli conjugado: $T(X) = \sum_{i=1}^{n} X_i$.
- Modelo Poisson conjugado: $T(X) = \sum_{i=1}^{n} X_i$.
- Modelo Normal conjugado: $T(X) = (\sum_{i=1}^{n} X_i, \prod_{i=1}^{n} X_i^2)$.

Función de Verosimilitud

- Sea $X = (X_1, \dots, X_n)$ un vector aleatorio con función de densidad conjunta $f_\theta(x_1, \dots, x_n)$. La función de verosimilitud θ de la muestra $x = (x_1, \dots, x_n)$ es la función

$$\mathcal{L} = (\theta) = f_\theta(x_1, \dots, x_n) : \Theta \to \mathcal{R}.$$

- Caso Bernoulli.
- Caso Poisson.
- Si cada $X_i \sim N(\mu, \sigma^2)$ para $i = 1, \dots, n$, entonces

$$\mathcal{L}(\theta) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \times \exp\left\{-\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2}\right\},$$

en que $\theta = (\mu, \sigma^2)$

Interpretaciones de la "Verosimilitud"

- La función de verosimilitud L induce una relación de orden (preferencia) sobre Θ dada por $\theta_1 \leq \theta_2$ siempre que $\mathcal{L}(\theta_1) \leq \mathcal{L}(\theta_2)$.
- Criterio del más preferente: $\theta^\star$ tal que $\theta \leq \theta^\star$ para todo $\theta \in \Theta$.

Principio de Verosimilitud*

Dada una muestra $x = (x_1, \ldots, x_n)$ de una variable X con modelo estadístico que contempla una función de verosimilitud $\mathcal{L}(\theta)$. Entonces, las inferencias acerca de θ deben ser basadas en $\mathcal{L}(\theta)$.
[Reid, N., & Cox, D. (2013). Principles of Statistical Inference].

Principio frecuentista clásico

- Sea T un estimador de θ y $L(L, \theta)$ una medida de error en la estimación.
- Consideremos una secuencia infinita de variables de X: $X_1, X_2, \ldots,$.

Principio frecuentista clásico*

Las propiedades de inferencia de T sobre θ medidas por una funcion $L(T, \theta)$ deben ser evaluadas según las propiedades de L el repetir el experimento un número infinito de veces.
Evans(2014).

Perdida, función de perdida, regla de decisión y riesgo medio

Añadir

La **Teoria de la Decisión** se preocupa del proceso de toma de decisión basado en modelo probabilisticos. La **Teoria de la Decisión Estadística** empleo modelo estadísticos y datos.

Abordajes

- Normativa: busca maximizar ganancias y/o minimizar perdidas.
- Descriptiva: empleo técnicas heuristicos en contextos prácticos

El fundamento teorico en el caso normativo esta el la **Teoria de la Utilidad** que toma como base los axiomas de compartamiento racional de los jugadores.

Tipos de metodologías:

- Clasica: basada en modelos de probabilidad.
- Bayesiana: incluye un modelo de probabilidad para el estado de la naturaleza.
- Estadística clasica; empleao modelos estadísticos estimados desde datos.
- Estadística Bayesiana: emplea modelos Bayesianos a posteriores.

Elementos de la Teoria de Decisión

- El objetivo es estimar el **estado de naturaleza** desconocido dentro del **conjunto de los posibles estados**.
- La estimación se realiza por medio de una función de decisión que asume valores en el **espacio de las posibles decisiones**.
- Decimos que ocurre una perdida cuando nuestra decisión no acierta a estimar el verdadero estado de la naturaleza.
- Medimos la magnitud de la perdida con una función de perdida.
- La Teoria de Decisiones Estadística utiliza datos y una función de decisión T para estimar el "verdadero estado de la naturaleza" especifico dentro de un "conjunto de estados posible" **minimizando** una función de perdida (A. conservador) o **maximizando** una función de utilidad (A. riesgoso).

T. Decisiones Estadística $\subset$ T. Decisión

- El parametro θ es el **estado de naturaleza** desconocido y Θ el **conjunto de los posibles estados**.
- El estadístico $T : \Lambda \to \mathbb{A}$ es llamado de función de decisión y $\mathbb{A}$ el **espacio de las posibles decisiones**. Además, $\mathcal{F}_{\mathbb{A}}$ es la σ-álgebra sobre $\mathbb{A}$. El espacio medible $(\mathbb{A}, \mathcal{F}_{\mathbb{A}})$ es llamado de **espacio de decisiones**.
- La Teoria de Decisiones Estadística (TDE) busca utilizar una muestra X de $P_\theta \in \mathcal{P}$ atraves de una función de decisión T para estimar el "verdadero estado de la naturaleza" especifico dentro de un "conjunto de estados posible" **minimizando** una función de perdida L (A. conservador) o **maximizando** una función de utilidad U (A. riesgoso).
- En el contexto de la TDE, existe tres opciones para $\mathbb{A}$
  - Si $\mathbb{A} = \Theta$, entonces T es un estimador puntual de θ.
  - Si $\mathbb{A} = I_\Theta$ (clase de intervalos abiertos), entonces, T es un estimador intervalar de θ.
  - Si $\mathbb{A} = \{0, 1\}$, en donde 1 significa rechazar H_0, entonces, T es una función del tipo *test* .

Perdida en estimación puntual

Definición 12. Función de perdida.

Decimos que **ocurre una perdida** en la estimación de θ por $T(x)$ siempre que $T(x) \neq \theta$.
La magnitud de la perdida es medida por una función. Una **función de perdida** L, es una función $L : \Theta \times \mathbb{A} \to [0, \infty)$ para una medida fija $P_\theta \in \mathcal{P}$. Cuando $X = x$ es observada y tomamos la decisión $T(x)$, entonces nuestra "magnitud de la perdida" es $L(\theta, T(x))$.

Ejemplos

- $L(\theta, T) = \dfrac{\theta}{2}(\theta - T(x))^2$.
- $L(\theta, T) = (\theta - T(X))^2$ llamada de perdida cuadratica.
- $L(\theta, T) = |\theta - T(x)|$ llamada de perdida absoluta.

Riesgo medio

Definición 13. Función de Riesgo medio.

Para una medida P_θ fija, la función $L(\theta, T(X))$ es una función aleatoria de X con medida P_θ. El **riesgo** de T, denotado por R_T es el valor esperado:

$$R_T(\theta) = E_X[L(\theta, T(X))].$$

Ejemplo

- Sea $X_1, \ldots, X_n$ una muestra tal que $X_i \sim \text{Modelo}(\theta)$. Entonces, para $T(X) = \frac{1}{n}\sum_{i=1}^n X_i = \bar{X}$, tenemos que $R_T(\theta) = (\theta - E\bar{X})^2 + \dfrac{\sigma^2}{n}$.
- Sea $X_1, \ldots, X_n$ una muestra tal que $X_i \sim \text{Normal}(\mu, \sigma^2)$. Entonces, para $T(X) = \bar{X}$, tenemos que $R_T(\theta) = (\mu - E\bar{X})^2 + \dfrac{\sigma^2}{n}$.

Admisibilidad e insesgamiento

Definición 14. Admisibilidad

Sea $\mathcal{D}$ una clase de funciones de decisión. Una función de decisión $T \in \mathcal{D}$ es llamada de **admisible** en $\mathcal{D}$ si y solo si no existe alguna otra función $S \in \mathcal{D}$ con un riesgo medio menor que T.

En palabras simples, una función de división es inadmisible si existe otra función con un riesgo menor a ella.

Definición 15. Insesgamiento.

En un problema de estimación, el **sesgo** de un estimador puntual $T(X)$ de un parámetro real $g(\theta)$ de una población desconocida es definida como la función $b_T(g(\theta)) = E[T(X)] - g(\theta)$. Un estimador $T(X)$ es llamado de **insesgado** para $g(\theta)$ si y solo si $b_T(g(\theta)) = 0$ para cualquier medida $P_\theta \in \mathcal{P}$.

Resultado de optimalidad

Teorema 3. Rao-Blackwell.

Sea T un estadistico suficiente para $\theta \in \Theta$, T_0 una función de decisión real tal que $E\,\|T_0\| < \infty$ y definamos $T_1 = E\,[T_0(X) \mid T]$. Entonces, $R_{T_1}(\theta) \leq R_{T_0}(\theta)$ para cualquier $P_\theta \in \mathcal{P}$.

- Sea $X_1, \ldots, X_n$ i.i.d tal que $X_i \sim \text{Poisson}(\theta)$. Queremos estimar el parametro $P(X = 0) = g(\theta) = e^{-\theta}$. Temos que a estatística $T(X) = \sum_{i=1}^n X_i$ es suficiente para θ. Consideremos el estimador insesgado

$$S(X) = \begin{cases} 1, & \text{si } X_1 = 0 \\ 0, & \text{caso contrario.} \end{cases}$$

Luego, el estimador

$$\widehat{g(\theta)} = \left(\frac{n-1}{n}\right)^{\sum_{i=1}^n X_i}.$$

es insesgado com menor varianza que S.

Decisión Minimax

Definición 16

Una función de decisión T_0 es llamada de **decisión minimax** de uma clase $\mathcal{D}$ de funciones siempre que:

$$\sup_{\theta \in \Theta} R_{T_0}(\theta) = \inf_{T \in \mathcal{D}} \sup_{\theta \in \Theta} R_T(\theta),$$

en que $\sup_{\theta \in \Theta} R_{T_0}(\theta)$: es el riesgo máximo al decidir T_0.
$\inf_{T \in \mathcal{D}} \sup_{\theta \in \Theta} R_T(\theta)$: el menor riesgo máximo en Θ de todas las posibles decisiones en $\mathcal{D}$.

- Interpretaciones: La función de decisión minimax es la decisión que nos protege del riesgo máximo.
- Nota: el supremo es la menor de las cotas inferiores.

Decisión de Bayes

Definición 17

El **riesgo de Bayes** de la decisión T con respecto a la función de perdida de $L(\theta, T)$ es la esperanza:

$$r(\pi, T) = E_\pi[R_T(\theta)] = \sum_{\theta \in \Theta} R_T(\theta)\pi(\theta) \quad \text{o} \quad \int_\Theta R_T(\theta)\pi(\theta)d\theta$$

Una función de decisión T_B es llamada de **decisión de Bayes** para una función de perdida de $L(\theta, T)$ siempre que $r(\pi, T_B) = \min_{T \in \mathcal{D}} r(\pi, T)$.

Resultado (Teorema)

Sea $\mathcal{D}$ la clase de todas las funciones decisión. Cuando consideramos la función cuadratica, entonces la decisión de Bayes que minimiza el riesgo medio dada una priori $\pi(\theta)$ es el valor esperado $T_B(\mathbf{X}) = E(\theta|\mathbf{X})$. **Nota:** $R_T(\theta)$ es una variable aleatoria función de θ.

Ejemplo de decisión de Bayes

- Para el modelo Bayesiano Conjugado de $\theta \sim \text{Beta}(a, b)$ y $X_i|\theta = \theta_0 \sim \text{Bernoulli}(\theta_0)$ para $i = 1, 2, \ldots, n$. La decisión de Bayes es:
$$d_B(\mathbf{X}) = E(\theta|\mathbf{X}) = \frac{\sum_{i=1}^{n} X_i + a}{n + a + b}.$$
- Consideremos un modelo conjugado Poisson en que $X_i|\theta = \theta_0 \sim \text{Poisson}(\theta_0)$ y $\theta \sim \text{Gamma}(a, b)$.
$$d_B(\mathbf{X}) = E(\theta|\mathbf{X}) = \frac{\sum_{i=1}^{n} X_i + a}{n + b}.$$

Estimación puntual en Teoría de la Desición

- Si $\mathbb{A} = \Theta$, entonces $T : \Lambda \rightarrow \Theta$ es un función de decisión puntual o estimador puntual de θ.
- La función de perdida L puede ser escrita de forma generalizada de la forma
$$L(\theta, T) = c(\theta)|\theta - T(x)|^r, \quad c(\theta) > 0.$$
- La función de riesgo es el valor esperado $E_T[L(\theta, T(X))]$.
- Podemos considerar la función cuadratica:
$$L(\theta, T) = (\theta - T(X))^2$$
- La función de perdida de diferencia de valor absoluto.
$$L(\theta, T) = |\theta - T(x)|.$$

Estimación intervalar y Teoría de la Desición

- Si $\mathbb{A}$ es la clase de los intervalos abiertos $IC \subset \Theta$, entonces, T es una función de decision intervalar o estimador intervalar de θ.
- Consideremos la perdida $L : \theta \times \mathbb{A}$ dada por [1]:

$$L(\theta, T) = a \times \text{length}(T(x)) - b \times I_{T(x)}(\theta), \quad a, b > 0.$$

- En este caso, la función de riesgo es el valor esperado

$$\begin{aligned} R(\theta, T(X)) &= E_T\left[a \times \text{length(T(X))} - b \times I_{T(X)}(\theta)\right] \\ &= E_T\left[a \times \text{length}(T(X))\right] - b \times P(I_{T(X)}(\theta) = 1) \\ &= aE_T\left[\text{length}(T(X))\right] - b \times P(\theta \in T(X)). \\ &= \text{El valor medio de la longuitud del intervalo por a} \\ &\quad\ \text{menos la probabilidad de } \theta \in T(X) \text{ por b.} \end{aligned}$$

[1] Denotamos $\text{length}(T(X)) = \text{length}([a, b]) = b - a$.

Prueba de hipótesis y Teoría de la Desición

- Si $\mathbb{A} = \{0, 1\}$, en donde 1 significa rechazar H_0, entonces, T es una función de decisiondel tipo *test*.
- Consideremos la partición $\mathcal{P}_0 \cup \mathcal{P}_1 = \mathcal{P}$.

Consideremos las hipótesis

$$H_0 : P_\theta \in \mathcal{P}_0 \ (\theta \in \Theta_0) \quad \text{versus} \quad H_1 : P_\theta \in \mathcal{P}_1 \ (\theta \in \Theta_1).$$

Con

$$L(\theta, T) = \begin{cases} a & \text{si } \theta \in \Theta_0 \text{ y } T(x) = 1 \\ b & \text{si } \theta \in \Theta_1 \text{ y } T(x) = 0. \end{cases}$$

Con riesgo medio $R_T(\theta) = aP_{\theta_0}(T(x) = 1) + bP_{\theta_1}(T(x) = 0)$

$$R_T(\theta_0) = \begin{cases} aP(T(X) = 1) & \text{si } \theta \in \Theta_0 \\ bP(T(X) = 0) & \text{si } \theta \in \Theta_1. \end{cases}$$

Estimación puntual, insesgamento y eficiencia. Estadísticos U

Ejemplo de estimadores puntuales

Ejemplo de estimadores puntuales.

- **Book:** Casella, G., & Berger, R. (2024). Statistical inference. CRC Press.
- Estimadores de Máxima Verosimilitud EMV
- Estimadores de los momentos Generalizados (E MG).
- Estadísticos U. **Book:** Lee, A. J. (2019). U-statistics: Theory and Practice. Routledge.

Definición 19

Sea X una muestra de la población $P_\theta \in \mathcal{P}$ sea $g(\theta)$ un parámetro real $g(\theta)$. Diremos que un estimador $T(X)$ de $g(\theta)$ es dicho **insesgado** si y solo si $E[T(X)] = g(\theta)$ para cualquier $\theta \in \Theta$. Si existe un estimador insesgado de $g(\theta)$, entonces $g(\theta)$ es llamado de **parametro estimable**.

Definición 20

Un estimador $T(X)$ insesgado de $g(\theta)$ es llamado de **Estimador Insesgado de Varianza Minima Uniforme (EIVMU)** si y solamente si $\text{Var}(T(X)) \leq \text{Var}(U(X))$ para todo $\theta \in \Theta$ y cualquier otro estimador insesgado $U(X)$ de $g(\theta)$.

Teorema 4 (Lehmann-Scheffé).

Supongamos que existe un estadístico suficiente y completo $T(X)$ para $\theta \in \Theta$. Si θ es un parametro estimable, entonces:

(i) Existe un unico estimador insesgado de θ que tiene la forma $h(T)$ para una función Boreliana h;

(ii) Adicionalmente, $h(T)$ es el único EIVMU de θ.

Nota: Si dos estimadores son iguales casi ciertamente en toda $\mathcal{P}$, entonces ambos son tratados como un mismo estimador.

- Sea $X_i \sim \text{Uniforme}(0, \theta)$, $\theta > 0$, $g(\theta)$ diferenciable en $(0, \infty)$. Luego, un EIVMU para $g(\theta)$ es

$$h(X_{(n)}) = g(X_{(n)}) + n^{-1} X_{(n)} g(X_{(n)}).$$

- Sea $X_i \sim \text{Poisson}(\theta)$, $\theta > 0$. Luego, un EIVMU para $g(\theta) = \theta^r$ es

$$h(T) = \frac{T!}{n^r (T - r)!}, \quad t \geq r, \quad h(T) = 0, \quad t < r.$$

Teorema 5

Sea $\mathcal{U}$ el conjunto de todos los estimadores insesgados de 0 con varianza finita y sea T un estimador insesgado de θ con $E\left(T^2\right) < \infty$.

(i) Una condición necesaria y suficiente para que $T(X)$ sea un EIVMU de θ es: $E[T(X)U(X)] = 0$ para cualquier $U \in \mathcal{U}$ y cualquier $\theta \in \Theta$.

(ii) Supongamos que $T = h(\tilde{T})$, donde $\tilde{T}$ es un estadístico suficiente para $\theta \in \Theta$ y h es una función Boreliana. Sea $\mathcal{U}_{\tilde{T}}$ un subconjunto de $\mathcal{U}$ que contiene funciones Borelianas de $\tilde{T}$. Entonces una condición necesaria y suficiente para que T sea un EIVMU de θ es que $E[T(X)U(X)] = 0$ para cualquier $U \in \mathcal{U}_{\tilde{T}}$ y cualquiera $\theta \in \Theta$.

Información de Fisher

Definición 21. Información de Fisher y función Score

La cantidad

$$\frac{\partial \log f_\theta(x)}{\partial \theta}$$

es llamada de **función Score** de θ contenida en x.

La cantidad

$$\mathrm{IF}_n(\theta) = E\left[\left(\frac{\partial \log f_\theta(X)}{\partial \theta}\right)^2\right].$$

es llamada de **Información de Fisher** de θ contenida en $X = (X_1, \ldots, X_n)$.

Denotamos por $\mathrm{IF}_i(\theta)$ la información de Fisher contenida en una observación particular x_i.

Ejemplos: escores e información de Fisher

- En el caso de la familia Normal con σ^2 conocida, tenemos la función escore dada por
$$\frac{\partial^2}{\partial\mu^2}\log f_\mu(x) = -\frac{n}{\sigma^2}.$$
Luego, $\mathrm{IF}_n(\mu) = -E\left[\frac{\partial^2}{\partial\mu^2}\log f_\theta(X)\right] = n/\sigma^2$.
- En el caso de la familia Poisson(θ), tenemos la función escore dada por
$$\frac{\partial^2}{\partial\theta^2}\log p_\mu(x) = -\frac{x_i}{\theta^2}.$$
Luego, $\mathrm{IF}_n(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\log p_\theta(X)\right] = n/\theta$.

Teorema 6 (Límite inferior de Cramér-Rao)

Sea $X = (X_1, \ldots, X_n)$ una muestra de P_θ, donde Θ es un conjunto abierto de $\mathcal{R}^k$. Supongamos que $T(X)$ es un estimador tal que $E[T(X)] = g(\theta)$ es una función diferenciable de θ y P_θ tiene p.d.f. f_θ con respecto a una medida ν para todo $\theta \in \Theta$, f_θ es diferenciable como función de θ y satisface

$$\textbf{[Condición T6]} \quad \frac{\partial}{\partial\theta}\int_\Lambda h(x) f_\theta(x) d\nu = \int_\Lambda h(x) \frac{\partial}{\partial\theta} f_\theta(x) d\nu, \ \forall\theta \in \Theta,$$

para $h(x) \equiv 1$ y $h(x) = T(x)$. Entonces,

$$\mathrm{Var}(T(X)) \geq \mathrm{LI}(\theta) = \left[\frac{\partial}{\partial\theta} g(\theta)\right]^2 [\mathrm{IF}_n(\theta)]^{-1}.$$

Definición 22. Estimador eficiente

Un estimador $T(X)$ es llamado de **eficiente** para estimar θ si y solo si

$$\operatorname{Var}(T(X)) = \left[\frac{\partial}{\partial\theta} g(\theta)\right]^2 [\mathrm{IF}_n(\theta)]^{-1}.$$

Ejemplo de estimadoes eficientes:

- En el caso de la familia Normal con σ^2 conocida, $g(\mu) = \mu$, $\mathrm{CI}(\mu) = [\mathrm{IF}_n(\theta)]^{-1} = \sigma^2/n$ que es igual a $\operatorname{Var}(\bar{X}) = \sigma^2/n$. Por ende, $\bar{X}$ es eficiente para estimar μ.
- En el caso de la familia Poisson con media θ, $g(\theta) = \theta$, $\mathrm{CI}(\theta) = [\mathrm{IF}_n(\theta)]^{-1} = \theta/n$ que es igual a $\operatorname{Var}(\bar{X}) = \theta/n$. Por ende, $\bar{X}$ es eficiente para estimar θ.

Proposición 2

(i) Sea X e Y variables independientes con matrices de información de Fisher $\mathrm{IF}_X(\theta)$ y $\mathrm{IF}_Y(\theta)$, respectivamente. Entonces, la información de Fisher acerca de θ contenida en (X, Y) es $\mathrm{IF}_X(\theta) + \mathrm{IF}_Y(\theta)$. En particular, si $X_1, \ldots, X_n$ son i.i.d. y $\mathrm{IF}_1(\theta)$ es la información de Fisher de θ contenida en un único X_i, entonces la información de Fisher acerca de θ contenida en $X_1, \ldots, X_n$ es $n\mathrm{IF}_1(\theta)$.

(ii) Supongamos que X tiene densidad f_θ la cual es dos veces diferenciable en θ y que (3.3) es válido con $h(x) \equiv 1$ y f_θ es reemplazada por $\partial f_\theta / \partial \theta$. Entonces

$$\mathrm{IF}_n(\theta) = -E\left[\frac{\partial^2}{\partial \theta^2} \log f_\theta(X)\right], \quad n \geq 1.$$

Proposición 3

Supongamos que el modelo de probabilidad de X pertenece a la familia exponencial.

$$\frac{dP_\theta}{d\nu}(x) = \exp\left\{\eta^\top(\theta)T(x) - \xi(\theta)\right\} h(x), \quad x \in \Lambda, \tag{6}$$

Entonces:
(i) La Condición 1 es satisfecha para cualquier función h tal que $E(h(X)) < \infty$ y la información de Fisher puede ser escrita como

$$\mathrm{IF}_n(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\log f_\theta(X)\right].$$

(ii) Si $\mathrm{IF}_n(\theta)$ es la información de Fisher de $\theta = E[T(X)]$, entonces $\mathrm{Var}[T(X)] = \mathrm{IF}_n(\theta)^{-1}$.

Teoría de Estadisticos U (No paramétrico)

Book: Lee, A. J. (2019). U-statistics: Theory and Practice. Routledge.

- Consideramos un modelo estadístico no paramétrico $\left(\Lambda, \mathcal{G}, \{P_F : F \in \mathcal{L}\}\right)$, en que $\mathcal{L}$ es un espacio funcional de funciones de probabilidad acumulada $F(x) = P(X \leq x)$.
- Este abordaje incluye posibles variables del tipo **continua, absolutamente continua, discreta y mixta** desde que todas ellas poseen una función F.
- Algunas suposiciones sobre F:
- $\int xdF(x) < \infty$, es decir, la media es finita.
- $\int |x|^2 dF(x) < \infty$, el segundo momento es finito.
- $\int |xy|^2 dF(x, y) < \infty$, en que $F(x, y) = P(X \leq x, Y \leq y)$, la covarianza es finita.
- $\int h(x)dF(x)$ es una integral de Riemann-Stieltjes.

Suposiciones de La Teoría de Estadisticos U

- El parametro θ objetivo es una función $\theta = h(F)$ finita y estimable, es decir, existe un estadístico $T(X)$ tal que $E(T(X)) = \theta$.
- **Teorema:**Si existe $E(T(X)) = \theta$ ssi existe una funcion $\psi(\bullet)$ tal que $E(\psi(X_1, \ldots, X_k)) = E(T(X)) = \theta$ para $k \leq n$.
- θ es llamado de funcional regular de grado $k \leq n$.
- $\psi(\bullet)$ es llamado de kernel de θ.
- **Teorema:** $\psi(\bullet)$ es único kernel para θ casi ciertamente.

Ejemplos de Estadisticos U

- **[El promedio]:** El estadístico $T(X) = n^{-1}\sum_{i=1}^{n} X_i$ es un estimador insesgado para $\theta = \int_{-\infty}^{\infty} x dF(x)$.
- **[La varianza muestral]:** El estadístico $T(X) = \binom{n}{2}^{-1}\sum_{1\leq i<j\leq n}^{n} 2^{-1}(X_i - X_j)^2$ es un estimador insesgado para

$$\theta = \int\int 2^{-1}(x_1 - x_2)^2 dF(x_1)dF(x_2).$$

- **[La covarianza muestral]:** El estadístico $T(X) = \binom{n}{2}^{-1}\sum_{1\leq i<j\leq n}^{n} 2^{-1}(X_i - X_j)(Y_i - Y_j)$ es un estimador insesgado para el funcional

$$\theta = \int_{\mathcal{R}}\int_{\mathcal{R}} 2^{-1}(x_1 - x_2)(y_1 - y_2)dF(x_1, y_2)dF(x_2, y_2).$$

- Los estadísticos U tiene optimas propiedades asintóticas ($n \to \infty$).

Región e intervalos de confianza. Región de credibilidad

Región, intervalos y limites de confianza

Definición 23

Sea $X = (X_1, \ldots, X_n)$ un vector aleatorio en que cada X_i es una muestra de la misma población P_θ y sea $T(x) \subset \Theta$. Si

$$\inf_{\theta \in \Theta} P_\theta \left(\theta \in T(X)\right) \geq 1 - \alpha,$$

entonces $T(x)$ es llamado de **región de confianza** del θ de coeficiente de confianza $(1-\alpha) \times 100\%$ para $\alpha \in (0, 1)$.

Si $T(x)$ tiene la forma de un intervalo $[a, b]$ para $a < b$, entonces $T(x)$ es llamado de **intervalo de confianza** de θ. Si $T(x)$ tiene la forma $(-\infty, b)$ o $(a, +\infty)$, entonces $T(x)$ es llamado de **limite de confianza** de θ.

Nota: En el caso continuo tenemos que

$$P_\theta \left(\theta \in T(X)\right) = 1 - \alpha$$

para los intervalos y limites de confianza.

Muestras de una población Normal

Resultado

Sea $X_1, \ldots, X_n$ una muestra aleatoria de tamaño n de un modelo $N\left(\mu, \sigma^2\right)$. Entonces:

(i) $\bar{X}$ e S^2 son independentes;

(ii) $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$;

(iii) $\dfrac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t_{n-1}$; donde χ^2_{n-1} denota una variable aleatoria con distribución chi-cuadrado con $n-1$ grados de libertad y t_{n-1} denota una variable aleatoria con distribución de Student t con $n-1$ grados de libertad y $\bar{X} = \sum_{i=1}^{n} X_i/n$ e $S^2 = \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2/(n-1)$.

- Los estadísticos (ii) y (iii) son ejemplos de estadísticos ancilares de primera orden.
- $\bar{X}$ e S^2 son independentes por el **Teorema de Basu**.

Métodos del estadístico pivotal

Definición 24

Decimos que una variable aleatoria $Q(X_1, \ldots, X_n; \theta) = Q(\mathbf{X}; \theta)$ es una **cantidad pivotal** para el parámetro θ si su medida de probabilidad es independiente de θ.

Entonces, en el caso continuo, para $1 - \alpha$ fijo, podemos encontrar los cuantiles λ_1 y λ_2 en la densidad de $Q(\mathbf{X}; \theta)$ de modo que

$$P_\theta \left[\lambda_1 \leq Q(\mathbf{X}; \theta) \leq \lambda_2\right] = 1 - \alpha.$$

Luego, determinar t_1 y t_2 tales que:

$$P\left[t_1(\mathbf{X}) \leq \theta \leq t_2(\mathbf{X})\right] = 1 - \alpha.$$

Ejemplo de cantidades pivotales

- Sea $X_1, \ldots, X_n$ una muestra aleatoria de $\operatorname{Exp}(\theta)$. Luego, $Q(\mathbf{X};\theta) = 2\theta\sum_{i=1}^{n} X_i \sim \chi^2_{2n}$.
- Sea $X_1, \ldots, X_n$ una muestra aleatoria $\text{Unifome}(0,\theta)$. La cantidad $Q(\mathbf{X};\theta)$ $X_{(n)}/\theta$ tiene densidad de probabilida $f_Q(q) = nq^{n-1} I_{[0,1]}(q)$.
- Sea $X_1, \ldots, X_n$ una muestra aleatoria de tamaño n de un modelo $N\left(\mu, \sigma^2\right)$.
  - Suponiendo que se conoce σ^2, luego $Q(\mathbf{X};\mu) = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ tiene modelo $N(0,1)$.
  - Si σ^2 es desconocido, tenemos que $Q(\mathbf{X},\mu) = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim \text{t-Student}(n-1)$.
  - Considerando μ desconocido, tenemos que $Q\left(\mathbf{X},\sigma^2\right) = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ es una cantidad pivotal para σ^2.

IC para una media de poblaciones Normales

- Suponiendo que se conoce σ^2, luego $Q(\mathbf{X};\mu) = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ tiene modelo $N(0,1)$.
- Para un coeficiente de confianza $1-\alpha$, determinamos λ_1 y λ_2 tal que

$$P\left[\lambda_1 \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq \lambda_2\right] = 1-\alpha$$

- Entonces sea $\lambda_1 = -z_{\alpha/2}$ y $\lambda_2 = z_{\alpha/2}$, donde $P\left(Z \leq z_{\alpha/2}\right) = 1-\alpha/2$, $Z \sim N(0,1)$ por lo que el intervalo de longitud más pequeño viene dado por

$$\left[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}; \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right].$$

IC para una media de poblaciones NormalesI

- Si σ^2 es desconocido, tenemos que $Q(\mathbf{X}, \mu) = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim$ t-Student$(n-1)$.
- Entonces, dado $1-\alpha$, hay λ_1 y λ_2 en el modelo t_{n-1} de modo que

$$P\left[\lambda_1 \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq \lambda_2\right] = 1-\alpha.$$

- Como Q es simétrica, debemos elegir λ_1 y λ_2 para que el área a la derecha de λ_2 sea igual al área a la izquierda de λ_1, es decir $\lambda_1 = -t_{\alpha/2}$ y $\lambda_2 = t_{\alpha/2}$, donde $P\left(T \leq t_{\alpha/2}\right) = 1-\alpha/2$, $T \sim t_{n-1}$ de modo que el intervalo de longitud más pequeño está dado por

$$\left[\bar{X} - t_{\alpha/2}\frac{S}{\sqrt{n}}; \bar{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right].$$

IC para dos poblaciones Normales

- Considerando μ desconocido, tenemos que $Q\left(\mathbf{X},\sigma^2\right) = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ es una cantidad pivotal para σ^2.
- Por lo tanto, dado $1-\alpha$, podemos determinar λ_1 y λ_2 tal que

$$P\left[\lambda_1 \leq \frac{(n-1)S^2}{\sigma^2} \leq \lambda_2\right] = 1-\alpha.$$

- Considerando o intervalo simétrico, ou seja, $\lambda_1 = q_1$ e $\lambda_2 = q_2$, onde $P\left[\chi^2_{n-1} \geq \; q_2\right] = P\left[\chi^2_{n-1} \leq q_1\right] = \alpha/2$, temos de (5.3.3), o intervalo

$$\left[\frac{(n-1)S^2}{q_2};\frac{(n-1)S^2}{q_1}\right].$$

IC para dos poblaciones Normales

- Consideremos que $X_1, \ldots, X_n$ una muestra $X \sim N\left(\mu_1, \sigma^2\right)$ y $Y_1, \ldots, Y_m$, una muestra $Y \sim N\left(\mu_2, \sigma^2\right)$, donde X y Y son independientes.
- Sabemos que:

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \sigma^2\left(\frac{1}{n} + \frac{1}{m}\right)\right)$$

de modo que, siendo $\theta = \mu_1 - \mu_2$ y σ es conocido.

- Consideramos la cantidad pivotal

$$Q(\mathbf{X}, \mathbf{Y}, \theta) = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim N(0, 1).$$

- Siendo σ^2 conocido, tenemos el intervalo

$$\left[\bar{X} - \bar{Y} - z_{\alpha/2}\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}; \bar{X} - \bar{Y} + z_{\alpha/2}\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}\right].$$

IC para dos poblaciones Normales.

- En el caso que se desconosca σ^2, tenemos que una cantidad pivotal viene dada por

$$Q(\mathbf{X},\mathbf{Y},\theta)=\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}}\sim t_{n+m-2}$$

en donde

$$S_p^2=\frac{(n-1)S_x^2+(m-1)S_y^2}{(n+m-2)},\quad S_x^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$$

$$\text{e}\quad S_y^2=\frac{1}{m-1}\sum_{i=1}^{m}\left(Y_i-\bar{Y}\right)^2.$$

- Un intervalo de confianza para $\theta=\mu_1-\mu_2$, con coeficiente de confianza $1-\alpha$ viene dado por

$$\left[\bar{X}-\bar{Y}-t_{\alpha/2}S_p\sqrt{\frac{1}{n}+\frac{1}{m}};\quad \bar{X}-\bar{Y}+t_{\alpha/2}S_p\sqrt{\frac{1}{n}+\frac{1}{m}}\right].$$

IC para razón de varianzas

- En el caso donde $X \sim N\left(\mu_1, \sigma_1^2\right)$ y $Y \sim N\left(\mu_2, \sigma_2^2\right)$ y el interés es la construcción de un IC para σ_1^2/σ_2^2, teniendo en cuenta que

$$\frac{(n-1)S_x^2}{\sigma_1^2} \sim \chi^2_{n-1} \quad \text{e} \quad \frac{(m-1)S_y^2}{\sigma_2^2} \sim \chi^2_{m-1}$$

  tenemos que

$$Q(\mathbf{X}, \mathbf{Y}, \theta) = \frac{(m-1)S_y^2/\sigma_2^2(m-1)}{(n-1)S_x^2/\sigma_1^2(n-1)} \sim F_{m-1,n-1}$$

  donde $F_{m-1,n-1}$ denota la distribución F con $m-1$ y $n-1$ grados de libertad, respectivamente.

- El IC esta dado por:

$$\left[F_1 \frac{S_x^2}{S_y^2}; F_2 \frac{S_x^2}{S_y^2}\right].$$

Densidades muestrales

- $\dfrac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t_{n-1}$ tiene densidad t-Student:

$$f(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

- $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ tiene desidad Chi-cuadrado

$$f(x) = \frac{1}{2^{(n-1)/2}\Gamma((n-1)/2)}x^{(n-3)/2}e^{-x/2}$$

- $F_{m-1,n-1} = \frac{(m-1)S_y^2/\sigma_2^2(m-1)}{(n-1)S_x^2/\sigma_1^2(n-1)} \sim$ con densidad F de Snedector dada por

$$f(x) = \frac{\sqrt{(\frac{n1x}{n2})^{n1}\cdot(\frac{n2}{n1x+n2})^{n2}}}{x\cdot B(\frac{n1}{2},\frac{n2}{2})}$$

Intervalos de Confianza Aproximados.

- Consideramos IC aproximados para θ basados en la distribución asintótica del EMV $\hat{\theta}$ de θ. Recordemos que

$$\frac{\hat{\theta}-\theta}{\sqrt{(\mathsf{IF}_n(\theta))^{-1}}} \to_d N(0,1)$$

  Como $\mathsf{IF}_n(\theta)$ puede depender de θ, es substituindo por $\mathsf{IF}_n(\hat{\theta})$. Luego,

- $$Q(\mathbf{X},\theta) = \frac{\hat{\theta}-\theta}{\sqrt{\left(\mathsf{IF}_n(\hat{\theta})\right)^{-1}}} \to_d N(0,1)$$

  entonces $Q(\mathbf{X},\theta)$ es una cantidad pivotal con un modelo asintótico $N(0,1)$.

-

Intervalos de Confianza Aproximados

- Sea $X_1, \ldots, X_n$ una muestra $X \sim \text{Bernoulli}(\theta)$. El EMV de θ es $\hat{\theta} = \bar{X}$ y $\text{IF}_n(\theta) = 1/\theta(1-\theta)$, tenemos una cantidad pivote para θ es

$$Q(\mathbf{X}, \theta) = \frac{\bar{X} - \theta}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \to_d N(0,1).$$

- Un IC para θ con un coeficiente de confianza de aproximadamente $1-\alpha$ viene dado por

$$\left[\bar{X} - z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}; \bar{X} + z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right].$$

Intervalos de Confianza Aproximados

- Sea $X_1, \ldots, X_n$ una muestra $X \sim \text{Exp}(\theta)$, con función de densidad $f(x \mid \theta) = \theta e^{-\theta x}; \quad x > 0, \quad \theta > 0.$
- Como $\text{IF}_n^{-1}(\theta) = \theta^2$ y $\hat{\theta} = 1/\bar{X}$, se sigue de (5.4.1) que una cantidad fundamental para θ viene dada por

$$Q(\mathbf{X}, \theta) = \frac{1/\bar{X} - \theta}{\sqrt{\hat{\theta}^2/n}} \rightarrow_d N(0,1).$$

- Un IC con coeficiente de confianza aproximado $1 - \alpha$ viene dado por

$$\left[\frac{1}{\bar{X}} - z_{\alpha/2}\sqrt{\frac{1}{n\bar{X}^2}}; \frac{1}{\bar{X}} + z_{\alpha/2}\sqrt{\frac{1}{n\bar{X}^2}}\right].$$

Límites de Confianza

Para una muestra de una población Normal considerando σ^2 desconocida.

- El limite inferior de confianza para estimar μ es dado por:

$$\left[\bar{X}_n - t_{n-1,1-\alpha}\frac{S}{\sqrt{n}}, \quad \infty\right].$$

- El limite superior de confianza para estimar μ es dado por:

$$\left[\infty, \quad \bar{X}_n + t_{n-1,1-\alpha}\frac{S}{\sqrt{n}}\right].$$

Intervalos de Confianza Bayesianos

Definición 25. Intervalo de Credibilidad (Caso continuo)

Sea $\pi(\theta \mid \mathbf{x})$ una densidad a posteriore de θ dado $\mathbf{X} = \mathbf{x}$. Decimos que $[t_1; t_2]$ es un **Intervalo de Credibilidad** Bayesiano (ICB) para θ, con coeficiente de confianza $1 - \alpha$ si

$$\int_{t1}^{t2} \pi(\theta \mid \mathbf{x}) d\theta = 1 - \alpha.$$

Siempre que sea posible, la longitud del $[t_1; t_2]$ debe ser el menor posible. El ICB más corto ubicados en los valores más altos de $\pi(\theta \mid \mathbf{x})$ conoce como el **Intervalo de Densidad Máxima a Posteriori** (IDMP) satisfaciendo las siguientes condiciones

$$\{\theta : \pi(\theta|x) > k_\alpha\} \subset \text{ICB} \subset \{\theta : \pi(\theta|x) \geq k_\theta\}.$$

Intervalos de Confianza Bayesianos

Sea $X_1, \ldots, X_n$ i.i.d Normal$(\mu, 1)$. Consideremos para μ la modelo a priori $N(\mu_0, 1)$. Tenemos que el modelo a posteriori de μ dada $\mathbf{X}$ que denotamos por $\mu \mid \mathbf{X}$, está dada por $\mu \mid \mathbf{X} \sim N\left(\frac{\sum_{i=1}^{n} X_i + \mu_0}{n+1}, \frac{1}{n+1}\right)$
Siendo $1-\alpha = 0.95$, entonces tenemos de la tabla de probabilidades $N(0,1)$ que $[t_1; t_2]$ debe elegirse como sigue:

$$\frac{t_1 - \frac{\sum_{i=1}^{n} X_i + \mu_0}{n+1}}{\sqrt{\frac{1}{n+1}}} = -1,96 \quad \text{e} \quad \frac{t_2 - \frac{\sum_{i=1}^{n} X_i + \mu_0}{n+1}}{\sqrt{\frac{1}{n+1}}} = 1,96,$$

es decir $t_1 = \frac{\sum_{i=1}^{n} X_i + \mu_0}{n+1} - 1,96\sqrt{\frac{1}{n+1}}$ y $t_2 = \frac{\sum_{i=1}^{n} X_i + \mu_0}{n+1} + 1,96\sqrt{\frac{1}{n+1}}$, por lo que el intervalo bayesiano de menor longitud (HPD) para μ con coeficiente de confianza $\gamma = 0.95$ está dado por

$$\left[\frac{\sum_{i=1}^{n} X_i + \mu_0}{n+1} - 1,96\sqrt{\frac{1}{n+1}}; \frac{\sum_{i=1}^{n} X_i + \mu_0}{n+1} + 1,96\sqrt{\frac{1}{n+1}}\right].$$

Intervalos de Confianza Bayesianos

Sea $X_1, X_2, \ldots, X_n$ uma muestra aleatória de tamaño n da variable aleatória $X \sim U(0, \theta)$. Consideremos para θ una priori com densidad de Pareto

$$\pi(\theta) = \frac{b}{\theta^{b+1}} I_{(a,\infty)}(\theta).$$

Es posible probar que la densidad a posterior de θ dado $X = (X_1 \ldots, X_n)$ es

$$h(\theta|X) = \frac{(n+b)(\max(a, X_{(n)})^{n+b}}{\theta^{n+b+1}} I_{(\max(a,X_{(n)}),\infty)}(\theta).$$

El Intervalo Bayesiano simétrico para θ, con coeficiente de confianza $1-\alpha$, siendo $a < \theta < a'$:

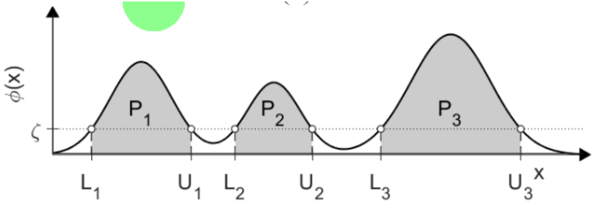
$$\left[\frac{\max\{a, X_{(n)}\}}{(1-\alpha)^{1/n+b}}; \frac{\max\{a, X_{(n)}\}}{\alpha^{1/n+b}}\right] \tag{7}$$

IC para modelo multimodal

Definición 26. IC caso multimodal (Caso continuo)

Consideremos una secuencia de puntos $L_1 < U_1 < L_2 < U_2 < \ldots < L_k < U_k$ del espacio paramétrico Θ de un modelo estadístico indexado por θ con k modas $\theta_1, \ldots, \theta_k$. Definimos un IC multimodal paraθ, a la secuencia de intervalos $[L_1, U_1], \ldots, [L_k, U_k]$ tal que

$$P\Big(\theta_k \in [L_1, U_1] \cup \ldots \cup \theta_k \in [L_k, U_k]\Big) = 1 - \alpha. \tag{8}$$



Regiones generales de confianzal

Una **región de confianza** basada en la verosimilitud con nivel α es

$$\Theta_\alpha = \{\theta \in \Theta : T_\theta \leq F_\alpha^{-1}\},$$

donde $T_\theta = 2(l(\widehat{\theta}) - l(\theta))$, $\widehat{\theta} = \arg\sup_{\theta \in \Theta} l(\theta)$ es el estimador de máxima verosimilitud, l es la función de log-verosimilitud y $F_\alpha^{-1} = (1-\alpha)$-cuantil calculado a partir de una función de distribución acumulativa F, es decir, $F(F_\alpha^{-1}) = 1 - \alpha$.
Aquí, F_θ es (una aproximación para) la función de distribución acumulativa de la variable aleatoria T_θ, que no depende de θ_0, donde θ_0 es el valor verdadero.
Patriota, A. G. (2013). A classical measure of evidence for general null hypotheses. Fuzzy Sets and Systems, 233, 74-88.

Observaciones finales

- Un IC clásico es interpretado en término de la probabilidad de que un estadístico T varie en un intervalo, mientras que θ se interpreta como una constante (vector) fijo.

$$P\left[\lambda_1 \frac{S}{\sqrt{n}} + \mu \leq \bar{X} \leq \lambda_2 \frac{S}{\sqrt{n}} + \mu\right] = 1 - \alpha.$$

- Cuando $X = x$ la muestra es observada, el IC se interpreta de la siguiente manera: "De 100 nuevas muestras de la misma población P_θ, aproximadamente el $(1-\alpha) \times 100\%$ de estas generaran un IC $T(x_i)$ que contenga θ, para $i = 1, \ldots, 100$."
- Cuando $X = x$ la muestra es observada, un ICB es una región de Θ al cual θ puede pertenecer con probabilidad $1-\alpha$

$$\left[\frac{\sum_{i=1}^{n} X_i + \mu_0}{n+1} - 1,96\sqrt{\frac{1}{n+1}} \leq \theta \leq \frac{\sum_{i=1}^{n} X_i + \mu_0}{n+1} + 1,96\sqrt{\frac{1}{n+1}}\right]$$

Convergencia de variables aleatorias, Orden de convergencias

Teoría Asintótica Estadística

- Cuando el modelo estadístico del estimador puntual o del estadístico tiene forma analitica conocida, no es necesario usar resultados asintóticos.
- Cuando el modelo estadístico de los anteriores no tiene una forma de encontrar por medio de métodos analiticos, este se puede aproximar por el modelo que tendría cuando el tamaño de la muestra es grande ($n \to \infty$).
- Es decir

$$P_{\text{asintótico}}(X \in A) \approx P_\theta(X \in A).$$

Repaso: convergencias de variables aleatorias

Sea $\{X_n\}$ una secuencia de vectores aleatorios definidos sobre el mismo espacio de probabilidad.

Definición 27 A. Tipos de convergencia de V.A

(i) Escribimos $X_n \rightarrow_{c.c.} X$ (casi ciertamente) ssi

$$P\left[\{\omega : \lim_{n\rightarrow\infty} X_n(\omega) = X(\omega)\}\right] = 1.$$

(ii) Escribimos $X_n \rightarrow_p X$ (convergencia en probabilidad) ssi

$$\lim_{n\rightarrow\infty} P\left(|X_n - X| > \epsilon\right) = 0. \quad \forall\epsilon > 0.$$

(iii) Escribimos $X_n \rightarrow {}_{L_r} X$ (converge a X en L_r) sii

$$\lim_{n\rightarrow\infty} E|X_n - X|^r = 0, \text{en que } r \in \mathcal{N}_+.$$

Repaso: convergencias de variables aleatorias

Definición 27 B. Tipos de convergencia de V.A

iv) Sea $F, F_n, n = 1, 2, \ldots$, una secuencia de funciones de distribución acumulada definidas sobre $\mathcal{R}^k$ y $P, P_n, n = 1, \ldots$, son las respectivas medidas de probabilidad. Decimos que $\{F_n\}$ convege a F debilmente (o $\{P_n\}$ converge a P debilmente) y escribimos $F_n \rightarrow_w F$ (o $P_n \rightarrow_w P$) si y solo si, para cada punto de continuidad x de F,

$$\lim_{n\rightarrow\infty} F_n(x) = F(x).$$

Nosotros decimos que $\{X_n\}$ converge a X en distribución y escribimos $X_n \rightarrow_d X$ si y solo si $F_{X_n} \rightarrow_w F_X$.

Interpretación de las convergencias

(i) El conjunto de puntos $\omega \in \Omega$ para los cuales X_n converge puntalmente a X tiene probabilidad 1.

(ii) Para todo nivel de proximidad ϵ, la probabilidad de que la diferencia entre X_n y X sea menor que ϵ converge a 1.

(iii) El valor esperado de orden r entre la diferencia entre X_n y X converge a 0.

(iv) La función de probabilidad acumulada F_n converge puntualmente a F en cada punto de continuidad de F.

Repaso: convergencias de variables aleatorias

Teorema 7. Implicancias entre convergencias

Sea $X, X_1, X_2, \ldots$ una secuencia de variables aleatorias definidos sobre el mismo espacio de probabilidad.

(i) Si $X_n \rightarrow_{c.c.} X$, entonces $X_n \rightarrow_p X$.

(ii) Si $X_n \rightarrow_{L_r} X$ para $r > 0$, entonces $X_n \rightarrow_p X$.

(iii) Si $X_n \rightarrow_p X$, entonces $X_n \rightarrow_d X$.

(iv) (Teorema de Skorohod). Si $X_n \rightarrow_d X$, entonces existen vectores aleatorios $Y, Y_1, Y_2, \ldots$ definimos sobre un mismo espacio de probabilidades tal que $P_Y = P_X$, $P_{Y_n} = P_{X_n}, n = 1, 2, \ldots$, and $Y_n \rightarrow_{c.c.} Y$.

Ordenes $O(\cdot), o(\cdot)$, $Op(\cdot)$ y $op(\cdot)$

Las secuencias reales $\{a_n\}$ y $\{b_n\}$,
- Satisface $a_n = O(b_n)$ ssi $|a_n| \leq c\,|b_n|$ para todo n y una constante c.
- Denotamos por $a_n = o(b_n)$ ssi $a_n/b_n \to 0$ cuando $n \to \infty$.

Definición 28. Orden de convergencia

Sea $X_1, X_2, \ldots$ y $Y_1, Y_2, \ldots$ dos secuencia de variables aleatorias definidas sobre un mismo espacio de probabilidad.
(i) $X_n = O(Y_n)$ c.c. si y solo si $P(|X_n| = O(|Y_n|)) = 1$.
(ii) $X_n = o(Y_n)$ c.c. si y solo si $X_n/Y_n \to_{c.c.} 0$.
(iii) $X_n = Op(Y_n)$ si y solo si, para cualquier $\epsilon > 0$, existe una constante $C_\epsilon > 0$ tal que $\sup_n P(|X_n| \geq C_\epsilon\,|Y_n|) < \epsilon$.
(iv) $X_n = op(Y_n)$ si y solo si $X_n/Y_n \to_p 0$.

Interpretación de las ordenes

(i) El evento $|X_n| \leq |Y_n|$ tiene probabilidad 1.

(ii) El conjunto de puntos $\omega \in \Omega$ para los cuales X_n/Y_n converge puntalmente a 0 tiene probabilidad 1.

(iii) Para todo nivel de proximidad ϵ, la probabilidad máxima para todo n de que la diferencia entre X_n y Y_n sea menor que ϵ esta acotada y depende de ϵ.

(iv) X_n/Y_n converge en probabilidad a 0.

Convergencia de funciones aleatorias

Para variables aleatorias X_n que convergen a X en algun sentido, nosotros a menudo queremos conocer si $g(X_n)$ converge a $g(X)$ en algun sentido. El siguiente resultado provee una respuestaa este tipo de preguntas.

Teorema 8. Convergencia de funciones de variables

Sea $X, X_1, X_2, \ldots$ una secuencia de variables aleatorias definidas sobre un mismo espacio de probabilidad y sea g una función medible de $\left(\mathcal{R}^k, \mathcal{B}^k\right)$ to $\left(\mathcal{R}^l, \mathcal{B}^l\right)$. Supongamos que g es continua a.s. P_X. Entonces,
(i) $X_n \rightarrow_{c.c.} X$ implica $g(X_n) \rightarrow_{c.c.} g(X)$;
(ii) $X_n \rightarrow_p X$ implica $g(X_n) \rightarrow_p g(X)$;
(iii) $X_n \rightarrow_d X$ implica $g(X_n) \rightarrow_d g(X)$.

Teorema del Límite Central

El Teorema del Límite Central (CLT), que desempeña un papel fundamental en la teoría asintótica estadística para aproximar la medida de probabilidad de sumas de variables aleatorias.

Teorema 9. T. del Límite Central (TLC) de Lindeberg

Sea $\{X_n\}$ variables aleatorias independientes con $0 < s_n^2 = \text{Var}\left(\sum_{j=1}^n X_j\right) < \infty$ y $E(X_i) < \infty$. Si

$$\textbf{[Condición de Lindebeg]} \sum_{j=1}^{n} E\left[(X_j - EX_j)^2 I_{\{|X_j - EX_j| > \epsilon s_n\}}\right] = o\left(s_n^2\right)$$

para cualquier $\epsilon > 0$, entonces

$$\frac{1}{s_n}\sum_{j=1}^{n}(X_j - EX_j) \to_d N(0,1).$$

II. Propiedades asintóticas de EIVMU, Estadístico U y EMV

Eficiencia asintótica de un EIVMU

Definición 29

Diremos que un estimador T_n de θ tal que $\dfrac{T_n - \theta}{\sqrt{Var(T_n)}} \rightarrow_d N(0, \mathrm{IF}_n(\theta))$ es **asintóticamente eficiente** cuando $\lim_{n\to\infty} Var(T_n) = [\mathrm{IF}_n(\theta)]^{-1}$.

Definición 30

Diremos que un estimador T_n de θ es **asintóticamente consistennte** cuando $T_n \rightarrow_p \theta$.

Propiedad asintótica de un EIVMU

Teorema 11. Eficiencia asintótica de EIVMU

Sobre la condición **Condición T6**, si $T_n(X)$ es insesgado para $g(\theta)$ y si, para cualquier $\theta \in \Theta$, vale

$$T_n(X) - g(\theta) = \left[\frac{\partial}{\partial\theta} g(\theta)\right]^{\tau} [\mathrm{IF}_n(\theta)]^{-1} \frac{\partial}{\partial\theta} \log f_\theta(X) \, [1 + o_p(1)]$$

casi ciertamente en P_θ, entonces

$$\lim_{n\to\infty} \mathrm{Var}_{T_n}(\theta) = \text{ La cota inferior de Cramér-Rao.}$$

siempre que la cota inferior de Cramér-Rao no sea nula.

Ejemplo de estadísticos U

Un estadístico U función de una muestra $X = (X_1, X_2, \ldots, X_n)$ es definida por:

$$U = U(X) = \binom{n}{m}^{-1} \sum_{c} h(X_1, X_2, \ldots, X_n),$$

donde $\sum_c$ es la suma sobre $\binom{n}{m}$ combinaciones de m distintos variables. Recordemos algunos Estadísticos U y h es llamado de kernel de orden $k \leq n$.

- $U_1 = n^{-1} \sum_{i=1}^{n} X_i$, en que $m = 1$ y $k = 1$.
- $U_2 = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n}^{n} 2^{-1}(X_i - X_j)^2$, $m = 2$ y $k = 2$.
- $U_3 = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n}^{n} 2^{-1}(X_i - X_j)(Y_i - Y_j)$, en que $m = 2$ y $k = 4$.

Modelo asintótico de un estadístico U

Teorema 12. Convergencia de Estadísticos U

Sea U_n un estadístico U definida en que cumple $E\left[h(X_1,\ldots,X_m)\right]^2 < \infty$. Sea $h_k(x_1,\ldots,x_k) = E(x_1,x_2,\ldots,x_k,X_{k+1},\ldots,X_m)$ y $\zeta_k = \mathrm{Var}(h_k(X_1,\ldots,X_k))$.

(i) Si $\zeta_1 > 0$, entonces

$$\sqrt{n}\left[U_n - E\left(U_n\right)\right] \rightarrow_d N\left(0, m^2\zeta_1\right)$$

(ii) Si $\zeta_1 = 0$ pero $\zeta_2 > 0$, entonces

$$n\left[U_n - E\left(U_n\right)\right] \rightarrow_d \frac{m(m-1)}{2}\sum_{j=1}^{\infty}\lambda_j\left(\chi_{1j}^2 - 1\right),$$

en que χ_{1j}^2 son variables i.i.d. chi-cuadrado con un grado de libertad y λ_j son algunas constantes (las cuales no dependen de P) que satisfacen$\sum_{j=1}^{\infty}\lambda_j^2 = \zeta_2$.

Consistencia asintótica

Teorema 13

(i) Si T_n es una secuencia de estimadores de un parámetro θ que satisface (a) $\lim_{n\to\infty} \text{Var}_\theta\, T_n = 0$ y (b) $\lim_{n\to\infty} \text{Bias}_\theta\, T_n = 0$, para cada $\theta \in \Theta$, entonces T_n converge en probabilidad a θ (es una secuencia consistente de estimadores).

(ii) Sean $a_1, a_2, \ldots$ y $b_1, b_2, \ldots$ secuencias de constantes que satisfacen (c) $\lim_{n\to\infty} a_n = 1$ y (d) $\lim_{n\to\infty} b_n = 0$. Entonces la secuencia $U_n = a_n T_n + b_n$ es una secuencia consistente de estimadores de θ.

Condiciones de regularidad CI

C1 La muestra $X_1, \ldots, X_n$ es identicamente distribuida e independiente.

C2 $\theta_1 \neq \theta_2$ implican $f_{\theta_1}(x) \neq f_{\theta_2}(x)$ (El modelo es **identificable**).

C3 Cada $f_\theta(x)$ es diferenciable en θ.

C4 El espacio paramétrico Θ contiene un conjunto abierto Θ_0 el cual contiene un punto interior θ_0 que es el verdadero parámetro θ tal que P_{θ_0} es la medida genero los datos.

Consistencia asintótica

Teorema 14

(Consistencia de MLE) Sea $X_1, X_2, \ldots,$ iid con misma medida P_θ y sea $L(\theta \mid \mathbf{x}) = \prod_{i=1}^{n} f\left(x_i \mid \theta\right)$ sea la función de verosimilitud. Sea $\widehat{\theta}_n$ el EMV de θ. Sea $g(\theta)$ una función continua de θ. Bajo las condiciones de regularidad en Miscellanea 10.6.2 en $f(x|\theta)$ y, por lo tanto, $L(\theta \mid \mathbf{x})$, para cada $\epsilon > 0$ y cada $\theta \in \Theta$,

$$\lim_{n\to\infty} P_\theta(g(\widehat{\theta}_n) - g(\theta)| \geq \epsilon) = 0$$

Es decir, $g(\widehat{\theta}_n)$ es un estimador consistente de $g(\theta)$.

Condiciones de regularidad CII

Asumimos C1-C4

C5 Para todo $x \in \Lambda$, la densidad o función de probabilidad $f_\theta(x)$ es tres veces diferenciable con respecto a θ, la tercera derivada es continua en θ, y $\int f_\theta(x)dx$ es tres veces diferenciable sobre la aplicación de la integral.

C6 Para cualquier $\theta_0 \in \Theta$, existe un número positivo c y una función $M(x)$ (ambos podrían depender de θ) tal que

$$\left|\frac{\partial^3}{\partial\theta^3}\log f_\theta(x)\right| \leq M(x) \quad \text{para todo } x \in \Lambda \quad \theta_0 - c < \theta < \theta_0 + c,$$

y además que $E[M(X)] < \infty$.

Eficiencia asintótica de EMV

Ahora mostramos que bajo algunas condiciones de regularidad, una Raíz de la Ecuación de Verosimilitud (REV), que es candidata para un M.L.E, es asintóticamente eficiente.

Teorema 15

(Eficiencia asintótica de los MLE) Sea $X_1, X_2, \ldots$, i.i.d y sea $\widehat{\theta}_n$ el EMV de θ, y sea $g(\theta)$ una función continua de θ. Bajo las condiciones de regularidad $C2$ en $f(x \mid \theta)$ y, por lo tanto, $L(\theta \mid \mathbf{x})$,

$$\sqrt{n}[\widehat{g(\theta)_n} - g(\theta)] \rightarrow_d \mathrm{N}[0, \mathrm{LI}(\theta)]$$

donde $\mathrm{LI}(\theta) = \left[\frac{\partial}{\partial\theta} g(\theta)\right]^2 [\mathrm{IF}_n(\theta)]^{-1}$ es el límite inferior de Cramér-Rao. Es decir, $g(\widehat{\theta}_n)$ es un estimador consistente y asintóticamente eficiente para estimar $g(\theta)$.

Método Delta

En el caso que las condiciones de regularidad N sean satisfecha, entonces

(i) $\sqrt{n}(\widehat{\theta} - \theta) \rightarrow_d N\left(0, \frac{1}{\mathrm{IF}_n(\theta)}\right)$ y

(ii) $\sqrt{n}(g(\widehat{\theta}) - g(\theta)) \rightarrow_d N\left(0, \frac{g'(\theta)^2}{\mathrm{IF}_n(\theta)}\right)$.

Referencias

- Bolfarine, H., & Sandoval, M. C. (2001). Introdução à inferência estatística (Vol. 2). SBM.
- Casella, G., & Berger, R. L. (2021). Statistical inference. Cengage Learning.
- Pereira C. (1993). Teoría Estadística. Monografias do IMPA, Rio de Janeiro. Disponible online: .
- Rincón, L. (2019). Una introducción a la estadística inferencial. México: Las prensas de Ciencia.
- Sen, P. K., & Singer, J. M. (2017). Large sample methods in statistics (1994): An introduction with applications. CRC press.
- Shao, J. (2003). Mathematical Statistics. Springer Science & Business Media.