Estadística

Probabilidades y Estadística (C) - 2019 - Parte 2

Estadística

POBLACION $\leftrightarrow F$	MUESTRA $X_1, \ldots X_n$ i.i.d. $X_i \sim F$
Parámetro: Valor asociado de F	Estimador:estadístico para estimar
$\theta = \theta(F)$	$\widehat{\theta}_n = \widehat{\theta}_n(X_1, \ldots, X_n)$
θ: valor poblacional	$\widehat{\theta}_n$ NUEVA VARIABLE ALEATORIA

Máxima verosimilitud

Caso discreto

- Modelo: $\mathcal{M} = \{p(\cdot, \theta)\ , \theta \in \Theta\}$.
- $\mathbf{x}= x_1, \cdots, x_n$ realización de $X_1, \cdots, X_n$ i.i.d.
- Función de verosimilitud asociada a $\mathbf{x}= x_1, \cdots, x_n$:

$$L(\cdot\ ;\ \mathbf{x}\ )\ : \Theta \to \mathbb{R}$$

$$L(\theta\ ;\ \mathbf{x}\ )\ =\ \mathbb{P}(X_1 = x_1, \cdots, X_n = x_n)\ , X_i \sim p(\cdot, \theta)\ .$$

$$L(\theta\ ;\ \mathbf{x}\ )\ =\ \prod_{i=1}^{n} p(x_i, \theta)\ ,$$

Nos interesa ver cuán **verosímil** es que un determinado parámetro haya generado los datos.

Máxima verosimilitud

Caso discreto

- Modelo: $\mathcal{M} = \{p(\cdot, \theta)\ , \theta \in \Theta\}$.

- $\mathbf{x}= x_1, \cdots, x_n$ realización de $X_1, \cdots, X_n$ i.i.d.

- Función de verosimilitud asociada a $\mathbf{x}= x_1, \cdots, x_n$:

$$L(\cdot\ ;\ \mathbf{x}\ )\ : \Theta \to \mathbb{R}$$

$$L(\theta\ ;\ \mathbf{x}\ )\ =\ \mathbb{P}(X_1 = x_1, \cdots, X_n = x_n)\ , X_i \sim p(\cdot, \theta)\ .$$

$$L(\theta\ ;\ \mathbf{x}\ )\ =\ \prod_{i=1}^{n} p(x_i, \theta)\ ,$$

Nos interesa ver cuán **verosímil** es que un determinado parámetro haya generado los datos.

- Propuesta de máxima verosimilitud:

$$h_n(\mathbf{x}) = \underset{\theta \in \Theta}{\operatorname{argmax}}\ L(\theta\ ;\ \mathbf{x}\ )\ .$$

o sea

$$L(h_n(\mathbf{x}), \mathbf{x}) \geq L(\theta, \mathbf{x})$$

Estimador de Máxima verosimilitud

Caso discreto

Bajo el modelo $\mathcal{M} = \{p(\cdot, \theta)\ , \theta \in \Theta\}$, el estimador de máxima verosimilitud para θ es la variable aleatoria

$$\widehat{\theta}_n = h_n(X_1, \ldots, X_n)$$

Máxima Verosimilitud

Caso continuo

- Modelo: $\mathcal{M} = \{f(\cdot,\theta)\ ,\theta \in \Theta\}$, con $f(\cdot,\theta)$ función de densidad.
- $\mathbf{x} = x_1, \cdots, x_n$ realización correspondiente a $X_1, \ldots, X_n$
- Función de verosimilitud: $L(\cdot\ ;\ \mathbf{x}\ )\ : \Theta \to \mathbb{R}$

$$L(\theta\ ;\ \mathbf{x}\ )\ =\ f_{X_1,\cdots,X_n}(x_1, \ldots, x_n)\ , X_i \sim f(\cdot,\theta)\ .$$

$$L(\theta\ ;\ \mathbf{x}\ )\ =\ \prod_{i=1}^{n} f(x_i,\theta)\ ,$$

- Propuesta de Máxima Verosimilitud:

$$h_n(\mathbf{x}) = \underset{\theta \in \Theta}{\operatorname{argmax}}\ L(\theta\ ;\ \mathbf{x}\ )\ .$$

o sea

$$L(h_n(\mathbf{x}), \mathbf{x}) \geq L(\theta, \mathbf{x})$$

- Definimos el EMV siendo $\widehat{\theta}_n = h_n(X_1, \cdots, X_n)$.

Ejemplo $\mathcal{E}(\lambda)$: $f(x,\lambda) = \lambda e^{-x\lambda}\, \mathcal{I}_{(0,\infty)}(x)$.

$X_1, \cdots, X_n$ v.a. i.i.d. $X_i \sim \mathcal{E}(\lambda), \lambda > 0$

$$L(\lambda; \mathbf{x}) = \prod_{i=1}^{n} f(x_i, \lambda) = \prod_{i=1}^{n} \lambda e^{-x_i\lambda}\, \mathcal{I}_{(0,\infty)}(x_i)$$

- si $x_i \geq 0 \ \forall i$

$$L(\lambda; \mathbf{x}) = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$$

- Si consideramos el $\log L$, resulta

$$\ell(\lambda; \mathbf{x}) = n \log(\lambda) - \lambda \sum_{i=1}^{n} x_i$$

- Derivando e igualando a 0 queda

$$\frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0 \Rightarrow \text{punto crítico es } 1/\bar{x}_n \quad \text{ver que maximiza!}$$

- $\Rightarrow \widehat{\lambda} = 1/\bar{X}_n$

$X_1, \cdots, X_n$ v.a. i.i.d. $X_i \sim \mathcal{N}(\mu, 9)$, $f(x, \mu, 9) = \frac{1}{\sqrt{2\pi}} \, \frac{1}{3} \, e^{-\frac{1}{2}\frac{(x-\mu)^2}{9}}$

$$L(\mu, 9; \mathbf{x}) \quad = \quad \prod_{i=1}^{n} f(x_i, \mu, 9)$$

$X_1, \cdots, X_n$ v.a. i.i.d. $X_i \sim \mathcal{N}(\mu, 9)$, $f(x, \mu, 9) = \frac{1}{\sqrt{2\pi}} \, \frac{1}{3} \, e^{-\frac{1}{2} \frac{(x-\mu)^2}{9}}$

$$L(\mu, 9; \mathbf{x}) \quad = \quad \prod_{i=1}^{n} f(x_i, \mu, 9) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \, \frac{1}{3} \, e^{-\frac{1}{2} \frac{(x_i-\mu)^2}{9}}$$

$X_1, \cdots, X_n$ v.a. i.i.d. $X_i \sim \mathcal{N}(\mu, 9)$, $f(x, \mu, 9) = \frac{1}{\sqrt{2\pi}} \, \frac{1}{3} \, e^{-\frac{1}{2}\frac{(x-\mu)^2}{9}}$

$$L(\mu, 9; \mathbf{x}) = \prod_{i=1}^{n} f(x_i, \mu, 9) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \, \frac{1}{3} \, e^{-\dfrac{1}{2}\dfrac{(x_i-\mu)^2}{9}}$$

$$L(\mu, 9; \mathbf{x}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{3}\right)^n e^{-\dfrac{1}{2}\dfrac{\sum_{i=1}^{n}(x_i-\mu)^2}{9}}$$

$X_1, \cdots, X_n$ v.a. i.i.d. $X_i \sim \mathcal{N}(\mu, 9)$, $f(x, \mu, 9) = \frac{1}{\sqrt{2\pi}} \, \frac{1}{3} \, e^{-\frac{1}{2}\frac{(x-\mu)^2}{9}}$

$$L(\mu, 9; \mathbf{x}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{3}\right)^n e^{-\frac{1}{2}\frac{\sum_{i=1}^n (x_i-\mu)^2}{9}}$$

- Tomemos logaritmo

$$\ell(\mu, 9; \mathbf{x}) = cte - \frac{1}{2}\frac{\sum_{i=1}^n (x_i-\mu)^2}{9}$$

Maximizar a $\ell(\mu, 9; \mathbf{x})$ como función de μ equivale a minimizar

$$h(\mu) = \sum_{i=1}^n (x_i - \mu)^2$$

Un par de clases atrás, vimos que $h(\mu)$ se minimiza en $\bar{x}_n$

EMV de μ : $\widehat{\mu} = \bar{X}_n$

$X_1, \cdots, X_n$ v.a. i.i.d. $X_i \sim \mathcal{N}(\mu, \sigma^2)$, $f\left(x, \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$

$$L(\mu, \sigma^2; \mathbf{x}) = \prod_{i=1}^{n} f(x_i, \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$X_1, \cdots, X_n$ v.a. i.i.d. $X_i \sim \mathcal{N}(\mu, \sigma^2)$, $f\left(x, \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$

$$L(\mu, \sigma^2; \mathbf{x}) = \prod_{i=1}^{n} f(x_i, \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\dfrac{(x_i-\mu)^2}{2\sigma^2}}$$

$$L(\mu, \sigma^2; \mathbf{x}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\sum_{i=1}^{n} \dfrac{(x_i-\mu)^2}{2\sigma^2}}$$

$X_1, \cdots, X_n$ v.a. i.i.d. $X_i \sim \mathcal{N}(\mu, \sigma^2)$, $f\left(x, \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$

$$L(\mu, \sigma^2; \mathbf{x}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}}$$

- Tomando logaritmo y resolviendo las ecuaciones

$$\frac{\partial \ell(\mu, \sigma^2; \mathbf{x})}{\partial \mu} = 0 \quad \text{y} \quad \frac{\partial \ell(\mu, \sigma^2; \mathbf{x})}{\partial \sigma^2} = 0$$

se obtiene que los EMV de μ y σ^2 son

$$\widehat{\mu} = \bar{X}_n \qquad \widehat{\sigma}^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}{n}$$

Notemos que el estimador ...

$$\widehat{\theta}_n \equiv \widehat{\theta}_n(X_1, \ldots, X_n)$$

- $\widehat{\theta}_n$ es una variable aleatoria.
- $\widehat{\theta}_n$ tiene distribución (siempre).

Sampling distribution of $\widehat{\theta}_n$: $f_{\widehat{\theta}_n}$

- $\widehat{\theta}_n$ tiene (en general) esperanza: $\mathbb{E}(\widehat{\theta}_n) = \int u f_{\widehat{\theta}_n}(u) du$
- $\widehat{\theta}_n$ tiene (en general) varianza: $\mathbb{V}(\widehat{\theta}_n)$

Sesgo y Varianza: ejemplo

- $X_i \sim \mathcal{U}(0, \theta)$
- $\widehat{\theta}_n = 2\overline{X}_n$.
- $\widetilde{\theta}_n = \max\{X_1, \ldots, X_n\}$.
- Calcule la esperanza y varianza de cada estimador.

Sesgo - Definiciones (pensando en modelos paramétricos...)

Sesgo (Bias) del estimador $\widehat{\theta}_n$:

$$\text{sesgo}(\widehat{\theta}_n) = \mathbb{E}(\widehat{\theta}_n) - \theta \quad \left(\text{sesgo}(\widehat{\theta}_n, \theta) = \mathbb{E}_\theta(\widehat{\theta}_n) - \theta\right)$$

Sesgo - Definiciones (pensando en modelos paramétricos...)

Sesgo (Bias) del estimador $\widehat{\theta}_n$:

$$\text{sesgo}(\widehat{\theta}_n) = \mathbb{E}(\widehat{\theta}_n) - \theta \quad \left(\text{sesgo}(\widehat{\theta}_n, \theta) = \mathbb{E}_\theta(\widehat{\theta}_n) - \theta\right)$$

- $\widehat{\theta}_n$ se dice INSESGADO si $\text{sesgo}(\widehat{\theta}_n) = 0$ (para todo θ)

$$\text{sesgo}(\widehat{\theta}_n, \theta) = 0 \;, \forall\theta \quad \equiv \quad \mathbb{E}_\theta(\widehat{\theta}_n) = \theta \;, \quad \forall\theta.$$

Dicho en palabras, el estimador $\widehat{\theta}_n$ se dice insesgado si su esperanza coincide con el valor de interés que queremos estimar:

- $(\widehat{\theta}_n)_{n\geq 1}$ se dice ASINTOTICAMENTE INSESGADO si

$$\lim_{n\to\infty} \text{sesgo}(\widehat{\theta}_n) = 0 \quad \text{(para todo } \theta)$$

$$\lim_{n\to\infty} \text{sesgo}(\widehat{\theta}_n, \theta) = 0 \;, \forall\theta \quad \equiv \quad \lim_{n\to\infty} \mathbb{E}_\theta(\widehat{\theta}_n) = 0 \;, \forall\theta$$

Consistencia (de la suceción de ESTIMADORES)

A medida que aumenta el tamaño n de la muestra, el estimador se acerca a lo que queremos conocer.

$$\widehat{\theta}_n \longrightarrow^{\mathcal{P}} \theta \,, \quad \text{cuando } n \to \infty$$

Consistencia (de la suceción de ESTIMADORES)

A medida que aumenta el tamaño n de la muestra, el estimador se acerca a lo que queremos conocer.

$$\widehat{\theta}_n \longrightarrow^{\mathcal{P}} \theta \ , \quad \text{cuando } n \to \infty$$

Definición: $(\widehat{\theta}_n)_{n\geq 1}$ se dice consistente sii

$$\widehat{\theta}_n \longrightarrow^{\mathcal{P}} \theta \ , \quad \text{cuando } n \to \infty, \text{ para todo } \theta.$$

$\widehat{\theta}_n(X_1, \ldots, X_n) \longrightarrow^{\mathcal{P}} \theta(F)$, cuando $X_i \sim F$, $\forall F$ en mi modelo.

Error cuadrático medio (ECM)

$$\text{ECM} : \mathbb{E}\left\{(\widehat{\theta}_n - \theta)^2\right\} \quad \text{ECM}(\widehat{\theta}_n, \theta) : \mathbb{E}\left\{(\widehat{\theta}_n - \theta)^2\right\}.$$

Lema: Si el ECM de un estimador converge a cero entonces vale la consistencia:

$$\mathbb{E}\left\{(\widehat{\theta}_n - \theta)^2\right\} \longrightarrow 0 \quad \text{implica que} \quad \widehat{\theta}_n \longrightarrow^{\mathcal{P}} \theta \ .$$

Propidades

Lema: El error cuadrático medio de un estimador se descompone de la siguiente manera:

$$\text{ECM}(\widehat{\theta}_n, \theta) = \mathbb{V}(\widehat{\theta}_n) + \left\{\mathbb{E}(\widehat{\theta}_n) - \theta\right\}^2$$

En particular... Si

$$\mathbb{V}(\widehat{\theta}_n) \to 0 \quad \text{y} \quad \mathbb{E}(\widehat{\theta}_n) \to \theta$$

tenemos que ECM converge a cero, y por lo tanto el estimador es consistente:

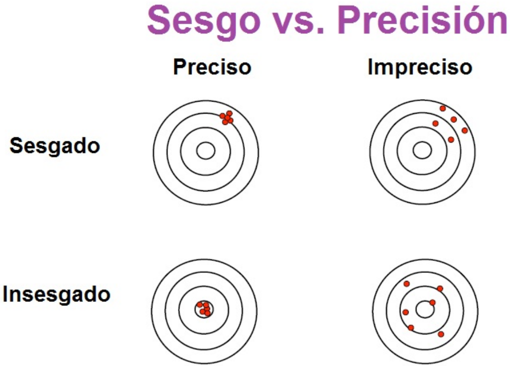
$$\widehat{\theta}_n \longrightarrow^{\mathcal{P}} \theta$$

Precisión y exactitud

Precisión (varianza): se refiere a la dispersión del conjunto de valores obtenidos de mediciones repetidas de una magnitud.

Exactitud (sesgo): se refiere a cuán cerca del valor real se encuentra el valor medido. En términos estadísticos, la exactitud está relacionada con el sesgo de una estimación.

Sesgo - Varianza



Propiedades

- Consistencia

$$\widehat{\theta}_n(X_1, \ldots, X_n) \to \theta(F) \text{ en probabilidad, cuando } X_i \sim F$$

$$\text{abreviado:} \quad \widehat{\theta} \to \theta$$

- Error cuadratico medio: ECM=$\mathbb{E}\{(\widehat{\theta}_n - \theta)^2\}$
- Lema: Si $\mathbb{E}\{(\widehat{\theta}_n - \theta)^2\} \to 0$, entonces $\widehat{\theta}_n \to^{\mathcal{P}} \theta$
- Sesgo: $\mathbb{E}(\widehat{\theta}_n) - \theta$.
- Estimador insesgado: Sesgo=0: $\mathbb{E}(\widehat{\theta}_n) - \theta$
- Lema: $\mathbb{E}\left\{(\widehat{\theta}_n - \theta)^2\right\} = \mathbb{V}(\widehat{\theta}_n) + \left\{\mathbb{E}(\widehat{\theta}_n) - \theta\right\}^2$
- Si $\mathbb{V}(\widehat{\theta}_n) \to 0$ y $\mathbb{E}(\widehat{\theta}_n) \to \theta$, entonces

$$\mathbb{E}\{(\widehat{\theta}_n - \theta)^2\} \to 0$$