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

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

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

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

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:

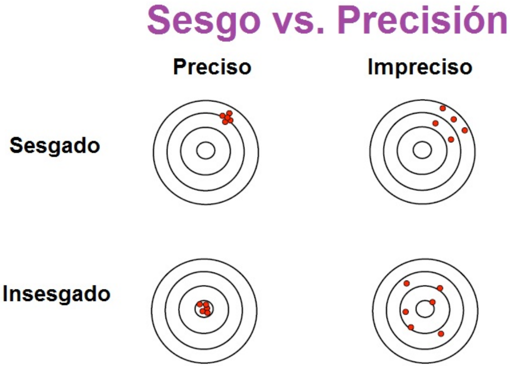
$$\widehat{\theta}_n \longrightarrow^{\mathcal{P}} \theta$$

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

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

Estimación: ejemplo

- $X_1, \ldots, X_n$ i.i.d. $X_i \sim X$.
- Parámetro de interés; $\sigma^2 = \mathbb{E}\{(X-\mu)^2\} = \mathbb{V}(X)$
- ¿Estimador?

Estimación: ejemplo

- $X_1, \ldots, X_n$ i.i.d. $X_i \sim X$.
- Parámetro de interés; $\sigma^2 = \mathbb{E}\{(X-\mu)^2\} = \mathbb{V}(X)$
- ¿Estimador?

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

Estimación: ejemplo

- $X_1, \ldots, X_n$ i.i.d. $X_i \sim X$.
- Parámetro de interés; $\sigma^2 = \mathbb{E}\{(X-\mu)^2\} = \mathbb{V}(X)$
- ¿Estimador?

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

- $\mathbb{E}(\widehat{\sigma}_n) = (n-1)n^{-1}\sigma^2$
- Estimador (insesgado) de la varianza: $S^2 = S_n^2$

$$S^2 = S_n^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n-1} \; .$$

- $\mathbb{E}[S^2] = \sigma^2$ (insesgado)
- $S^2 \to \sigma^2$ en probabilidad (consistencia)