4/02/2025

A grandes rasgos el proceso de la inferencia estadística...

1) FA $\rightarrow$ datos $x_1, x_2, \ldots, x_n$

2) Asumimos que existe una v.a. $X$ que puede describir el comportamiento agregado del FA

$$\underbrace{X \sim f(x|\theta)}_{\text{paramétrica}} \quad \text{ó} \quad \underbrace{X \sim F(x)}_{\text{no-paramétrico}}$$

3) Necesitamos asumir un modelo de recolección para $x_1, x_2, \ldots, x_n$, asumimos que son una realización de una

m.a. $\equiv$ v.a.i.i.d $X_1, X_2, \ldots, X_n$ de $X \sim$

4) Bajo estos supuestos podemos hacer

1) Estimación puntual
2) Estimación por intervalo
3) Prueba de hipótesis

4.1) Estimación puntual. El objetivo es obtener una única mejor estimación del parámetro ó cantidad de interés

Ej si $X_1, \ldots, X_n$ es una m.a. de $X \sim f(x|\theta)$

de alguna manera $\Rightarrow$ $\hat{\theta}_n = h(X_1, X_2, \ldots, X_n)$ estimador

$\hat{\Theta}_n = h(X_1, X_2, \ldots, X_n)$ estimación

$\hat{\Theta}_n$ es "mejor" con respecto a algún o varios criterios

Estimación – función de los valores observados de la m.a $(x_1, \ldots, x_n)$
valor — que usaremos para estimar $\Theta$

Estimador – función de la m.a $X_1, \ldots, X_n$ que podemos analizar
v.a. — usando teoría de probabilidad para - evaluar/conocer - sus propiedades al estimar $\Theta$.

Propiedades de los estimadores

1) Sesgo. El sesgo es una desviación sistemática de lo que se busca estimar.

$$\text{Sesgo}_\theta(\hat{\Theta}_n) = \mathbb{E}(\hat{\Theta}_n) - \Theta,$$

diremos que el estimador $\hat{\Theta}_n$ es insesgado para estimar $\Theta$ si

$$\text{Sesgo}_\theta(\hat{\Theta}_n) = 0 \iff \mathbb{E}(\hat{\Theta}_n) = \Theta$$

2) $\text{Var}(\hat{\Theta}_n) \longleftrightarrow \text{se}(\hat{\Theta}_n) = \sqrt{\text{Var}(\hat{\Theta}_n)}$

La varianza mide la desviación "promedio" de $\hat{\Theta}_n$ con respecto a $\mathbb{E}(\hat{\Theta}_n)$, en general se prefieren estimadores con menos varianza ó al menos que decrezca conforme $n$ aumente

$$\mathrm{Var}(\hat{\theta}_n) = \mathbb{E}\left(\left(\hat{\theta}_n - \mathbb{E}(\hat{\theta}_n)\right)^2\right)$$
$$= \mathbb{E}(\hat{\theta}_n^2) - \mathbb{E}(\hat{\theta}_n)^2$$

3) Error cuadrático medio

$$ECM_\theta(\hat{\theta}_n) = \mathbb{E}\left((\hat{\theta}_n - \theta)^2\right)$$

mide la desviación promedio, al cuadrado, entre la v.a. $\hat{\theta}_n$ y $\underline{\theta}$.
Si tenemos dos estimadores $\hat{\theta}_n^1$ y $\hat{\theta}_n^2$ de $\theta$

$\Rightarrow$ un criterio de decisión podría ser el de elegir el que tenga el menor error cuadrático medio.

Teorema

$$ECM_\theta(\hat{\theta}_n) = \mathrm{Sesgo}^2(\hat{\theta}_n) + \mathrm{Var}(\hat{\theta}_n)$$
$$= \left(\mathbb{E}(\hat{\theta}_n) - \theta\right)^2 + \mathbb{E}\left(\left(\hat{\theta}_n - \mathbb{E}(\hat{\theta}_n)\right)^2\right)$$

Dem

$$ECM_\theta(\hat{\theta}_n) = \mathbb{E}\left((\hat{\theta}_n - \theta)^2\right)$$
$$= \mathbb{E}\left(\left(\hat{\theta}_n - \mathbb{E}(\hat{\theta}_n) + \mathbb{E}(\hat{\theta}_n) - \theta\right)^2\right)$$
$$= \mathbb{E}\Big[\left(\hat{\theta}_n - \mathbb{E}(\hat{\theta}_n)\right)^2 + 2\left(\hat{\theta}_n - \mathbb{E}(\hat{\theta}_n)\right)\left(\mathbb{E}(\hat{\theta}_n) - \theta\right) + \left(\mathbb{E}(\hat{\theta}_n) - \theta\right)^2\Big]$$

Explicar sesgo, varianza y ECM con dibujos

$$= Var(\hat{\theta}_n) + 2\,\cancel{\mathbb{E}\left((\hat{\theta}_n - \mathbb{E}(\hat{\theta}_n))(\mathbb{E}(\hat{\theta}_n)-\theta)\right)}^{=0}$$

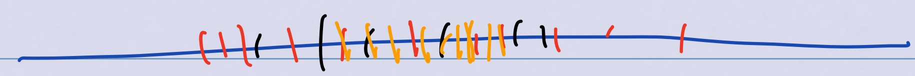
$$+ \mathbb{E}\left[(\mathbb{E}(\hat{\theta}_n)-\theta)^2\right]$$

$$= Var(\hat{\theta}_n) + Sesgo_\theta^2(\hat{\theta}_n)$$

4) Quizá la propiedad más importante actualmente es

$\hat{\theta}_n \xrightarrow{P} \theta$, i.e. $\hat{\theta}_n$ es un estimador consistente para estimar $\theta$

$n=10$ $n=50$ $n=100$



$\theta$

a medida que aumenta $n$ las estimaciones van cayendo más cerca de $\theta$, son realizaciones de una v.a. por eso se pide convergencia en probabilidad

$$\hat{\theta}_n \xrightarrow{P} \theta \iff \lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \varepsilon) = 0,$$

$$\forall \varepsilon > 0.$$

5: $\hat{\theta}_n \xrightarrow{D} Y \sim F_Y(y)$

convergencia en distribución

$\lim_{n\to\infty} F_{\hat{\theta}_n}(y) = F_Y(y)$ ← distribución asintótica

Antes de hacer un ejemplo, recordemos dos resultados fundamentales de su curso de probabilidad

Ley débil de los grandes números (LDGN)

Sea $X_1, X_2, \ldots, X_n$ una m.a. de $X \sim F(x)$, en donde $\mu = \mathbb{E}(X)$ y $\sigma^2 = Var(X)$. Recordemos que

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\Rightarrow \quad \overline{X}_n \xrightarrow{P} \mu$$

en términos estadísticos, sin importar la distribución de la m.a. si $\mu = \mathbb{E}(X)$ existe

$\Rightarrow$ $\overline{X}_n$ es un estimador consistente para $\mu$

Además sabemos que

$$\mathbb{E}(\overline{X}_n) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(X_i)$$

$$= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(X) = \mu$$

$\Rightarrow$ Sin importar la distribución que generó la m.a.

$\overline{X}_n$ es un estimador insesgado para $\mu$.

$$Var(\bar{X}_n) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) =$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} Var(X) = \frac{\sigma^2}{n}$$

$$\Rightarrow ECM_\mu(\bar{X}_n) = \underbrace{\cancel{Sesgo^2_\mu(\bar{X}_n)}}_{\to 0} + Var(\bar{X}_n)$$

$$= \sigma^2/n$$

Teorema Central de Límite (TCL)

Sea $X_1, X_2, \ldots, X_n$ una m.a. de $X \sim F(x)$, en donde $\mu = \mathbb{E}(X)$ y $\sigma^2 = Var(X) < \infty$

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} Z \sim N(0,1)$$

En términos menos formales pero que ayudan a entender el punto.

Para $\underline{n}$ suficientemente grande, pero fija,

$$\bar{X}_n \sim N(\mu, \sigma^2/n)$$

Acabamos de analizar a $\overline{X}_n$ como estimador de $\mu$ en el caso NO paramétrico. Lo que encontramos es que es un excelente estimador

## Ejemplo 2

Sea $X_1, X_2, \ldots, X_n$ una m.a. de $X \sim \text{Bernoulli}(\theta)$

$\Rightarrow$ digamos que queremos evaluar qué tan buen estimador es $\overline{X}_n$ de $\theta$, ¿por qué estimaríamos $\theta$ con $\overline{X}_n$?

1) Sesgo $\mathbb{E}(\overline{X}_n) = \mathbb{E}(X) = \theta$

2) Varianza $Var(\overline{X}_n) = \frac{1}{n} Var(\theta) = \frac{\theta(1-\theta)}{n}$

3) $ECM_\theta(\overline{X}_n) = \frac{\theta(1-\theta)}{n}$

4) $\overline{X}_n \xrightarrow{P} \theta$, es un estimador consistente de $\theta$

LDGN

5) $$\frac{\overline{X}_n - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \xrightarrow{D} Z \sim N(0,1)$$

$\Rightarrow$ pour $\underline{n}$ suf grande $\quad \bar{X}_n \sim N\left(\theta, \frac{\theta(1-\theta)}{n}\right)$