Conceptos fundamentales de inferencia estadística

1. Estimación puntual
2. Estimación por intervalos
3. Pruebas de hipótesis

2 Estimación por intervalo

El objetivo es construir $L_n = L(X_1, X_2, \dots, X_n)$
$U_n = U(X_1, X_2, \dots, X_n)$

tal que

$$P(L_n \leq \theta \leq U_n) = 1-\alpha$$

Desde la perspectiva clásica (L_n, U_n) es un intervalo que cubre a θ con $(1-\alpha)\times 100\%$ de confianza

¿Por qué usar una estimación por intervalo en lugar de una puntual?

a) $P(\hat{\theta}_n = \theta) = 0$ y $P(L_n \leq \theta \leq U_n) = 1-\alpha$

b) La amplitud del intervalo nos da una idea de que tan buena es la estimación.

c) La confianza es una manera de cuantificar la incertidumbre

Teorema (intervalo normal asintótico)

Sea $X_1, X_2, \ldots, X_n$ una m.a. de $f(x|\theta)$ y $\hat{\theta}_n$ un estimador de θ, si se cumple que

$$\frac{\hat{\theta}_n - \theta}{se(\hat{\theta}_n)} \xrightarrow{D} Z \sim N(0,1)$$

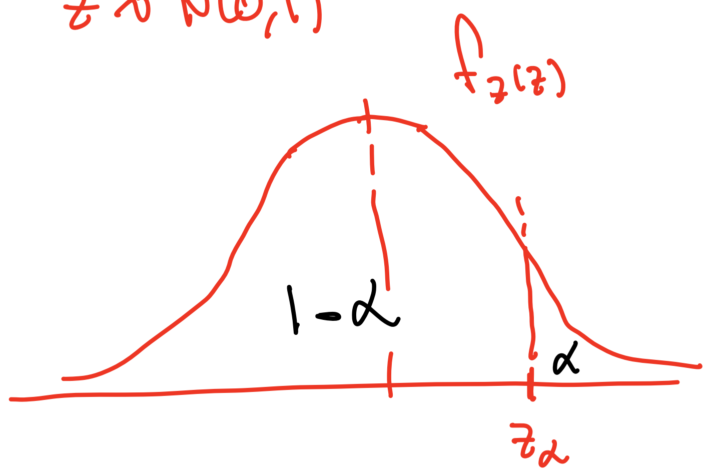
$\Longrightarrow$ Para $\underline{n}$ suf. grande el intervalo (L_n, U_n) es un intervalo del $(1-\alpha)\times 100\%$ de confianza para θ.

En donde

$$L_n = \hat{\theta}_n - z_{\alpha/2}\, se(\hat{\theta}_n)$$
$$U_n = \hat{\theta}_n + z_{\alpha/2}\, se(\hat{\theta}_n)$$

Con $z_\alpha = F_Z^{-1}(1-\alpha) \iff F_Z(z_\alpha) = 1-\alpha$

A z_α se le llama el cuantil $\underline{\underline{\alpha}}$ de la distribución $Z \sim N(0,1)$



Dem

Como $Z_n = \dfrac{\hat{\theta}_n - \theta}{se(\hat{\theta}_n)} \xrightarrow{P} Z \sim N(0,1)$

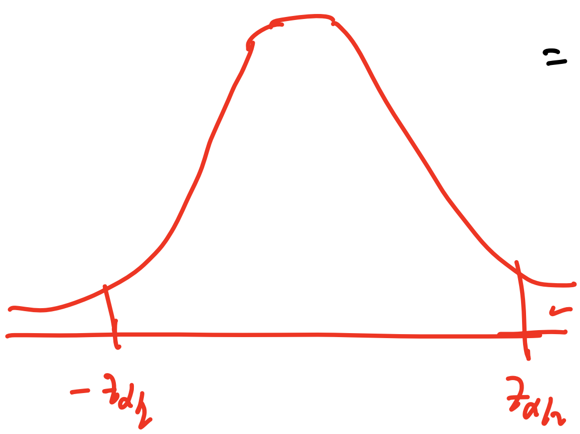
$\Rightarrow$ para n sut grande la distribución de Z_n y de Z son las mismas, i.e. $N(0,1)$

$\Rightarrow P(-z_{\alpha/2} \leq Z_n \leq z_{\alpha/2}) \approx P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2})$

$$P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = P(Z \leq z_{\alpha/2}) - P(Z \leq -z_{\alpha/2})$$

$$= P(Z \leq z_{\alpha/2}) - (1 - P(Z \leq z_{\alpha/2}))$$

$$= 2P(Z \leq z_{\alpha/2}) - 1$$



$$= 2F_Z(z_{\alpha/2}) - 1 = 2(1 - \alpha/2) - 1$$
$$= 2 - \alpha - 1 = 1 - \alpha$$

$\Rightarrow$ acabamos de demostrar que si $Z \sim N(0,1)$

$$\Rightarrow \boxed{\begin{array}{l} P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1-\alpha \\ \text{en donde} \quad z_{\alpha/2} = F_Z^{-1}(1-\alpha/2) \end{array}}$$

Para n suf. grande

$$P\left(-z_{\alpha/2} \leq \frac{\hat{\theta}_n - \theta}{se(\hat{\theta}_n)} \leq z_{\alpha/2}\right) \approx 1-\alpha$$

$$\Leftrightarrow \quad P\left(\hat{\theta}_n - z_{\alpha/2}\, se(\hat{\theta}_n) \leq \theta \leq \hat{\theta}_n + z_{\alpha/2}\, se(\hat{\theta}_n)\right) \approx 1-\alpha$$

$\Rightarrow$ para n suf grande y considerando

$$L_n = \hat{\theta}_n - z_{\alpha/2}\, se(\hat{\theta}_n)$$
$$U_n = \hat{\theta}_n + z_{\alpha/2}\, se(\hat{\theta}_n)$$

(L_n, U_n) es un intervalo del $(1-\alpha) \times 100\%$ de confianza para θ.

Ejemplo

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

$\Rightarrow$ vimos que $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

es un estimador insesgado, i.e. $\mathbb{E}(\hat{\theta}_n) = \theta$

$$ECM_\theta(\hat{\theta}_n) = Var(\hat{\theta}_n)$$

$$= \frac{\theta(1-\theta)}{n}$$

$$\hat{\theta}_n \xrightarrow{P} \theta \quad \text{consistente}$$

y que $$\frac{\hat{\theta}_n - \theta}{se(\hat{\theta}_n)} \xrightarrow{D} Z \sim N(0,1)$$

$\Rightarrow$ un intervalo asintótico (para n suf. grande) para θ está dado por (L_n, U_n)

$$L_n = \hat{\theta}_n - z_{\alpha/2}\, se(\hat{\theta}_n)$$
$$U_n = \hat{\theta}_n + z_{\alpha/2}\, se(\hat{\theta}_n)$$

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

¿?

$$se(\hat{\theta}_n) = \frac{\theta(1-\theta)}{n}$$

$$\hat{se}(\hat{\theta}_n) = \frac{\hat{\theta}_n(1-\hat{\theta}_n)}{n}$$

¡IMPORTANTE! si $\theta \approx 1$ o $\theta \approx 0 \Rightarrow$ la aproximación no funciona muy bien y el n necesario suele ser demasiado grande.

Recordatorio TCL

Sea $X_1, X_2, \ldots, X_n$ una m.a. de $X \sim F(x)$
con $\mu = \mathbb{E}(X)$ y $0 < \sigma^2 = Var(X) < \infty$

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

$$\frac{\hat{\theta}_n - \theta}{se(\hat{\theta}_n)}$$

obviamente si $se(\hat{\theta}_n) \approx 0$
el cociente no está bien definida

$se(\hat{\theta}_n) = \frac{\theta(1-\theta)}{n}$

$\theta \approx 0$ ó 1 hay problemas

③ Pruebas de hipótesis

En pruebas de hipótesis se empieza con una hipótesis acerca de un parámetro poblacional y usamos la información de la m.a. para "intentar rechazarla", si no es posible rechazarla $\Rightarrow$ la tomamos como verdadera

Ejemplo

Sea $X_1, X_2, \dots, X_n$ una m.a. de $X \sim$ Bernoulli(θ)

$$\underbrace{H_0 : \theta = 1/2}_{\text{hipótesis nula}} \quad Vs \quad \underbrace{H_1 : \theta \neq 1/2}_{\text{hipótesis alternativa}}$$

Necesitamos ① Una estadística con la que podamos medir desviaciones con respecto a lo que dice H_0
$\equiv$ estadística de prueba

$$T_0 = \left|\hat{\theta}_n - \frac{1}{2}\right|$$

② Necesitamos obtener la distribución de T_0

③ Encontrar un c, punto de corte - valor crítico - de manera que si

$$T_0 > c \Rightarrow \text{rechazar } H_0.$$

3 = Estadística no paramétrica

① Tenemos una característica de una población (o fenómeno aleatorio de interés), asumimos que existe una v.a. X

que es capaz de modelar su comportamiento agregado

② $X \sim F(x)$ para una F general

③ Sea $X_1, X_2, \ldots, X_n$ una m.a. de $X \sim F(x)$

¿Cómo estimamos $F(x) \to (x, F(x))$ con la m.a. $X_1, X_2, \ldots, X_n$?

Vía la función de distribución empírica

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(X_i \leq x)$$

para cada $x \in \mathbb{R}$

en donde $\mathbb{1}(X_i \leq x) = \begin{cases} 1 & \text{si } X_i \leq x \\ 0 & \text{c.c.} \end{cases}$

Teorema

Para cada $x \in \mathbb{R}$, se tiene que

1) $\mathbb{E}(\hat{F}_n(x)) = F(x)$

2) $\text{Var}(\hat{F}_n(x)) = \dfrac{F(x)(1-F(x))}{n}$

3) $ECM_{F(x)}\left(\hat{F}_n(x)\right) = Var\left(\hat{F}_n(x)\right)$

4) $\hat{F}_n(x) \xrightarrow{P} F(x)$

Dem

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(X_i \leq x)$$

$$Y_i = \mathbb{1}(X_i \leq x)$$

$\Rightarrow$ $Y_i = 1$ ó 0 $\Rightarrow$ $Y_1, Y_2, \dots, Y_n$ es una m.a. de $Y \sim$ Bernoulli$(F(x))$

$$P(Y=1) = P(X \leq x)$$
$$= F(x)$$

$$\Rightarrow \hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(X_i \leq x)$$

$$= \frac{1}{n}\sum_{i=1}^{n} Y_i$$

$\Rightarrow$ $\mathbb{E}\left(\hat{F}_n(x)\right) = F(x)$ ✓

$$Var\left(\hat{F}_n(x)\right) = \frac{1}{n^2}\sum Var(Y_i)$$

$$= \frac{F(x)\left(1-F(x)\right)}{n}$$

③ $$ECM_{F(x)}\left(\hat{F}_n(x)\right) = \frac{F(x)\left(1-F(x)\right)}{n}$$

$$\lim_{n\to\infty} ECM_{F(x)}\left(\hat{F}_n(x)\right) = 0$$

$$\Rightarrow \hat{F}_n(x) \xrightarrow{P} F(x)$$

④ ✓