11/02/2025

Proceso de la inferencia estadística

- FA $\rightarrow x_1, \ldots, x_n$
- Asumimos que $\exists$ una v.a. que puede describir el comportamiento agregado del FA

$$\underline{X} \sim \begin{cases} f(x|\theta) \\ F(x) \end{cases}$$

- Modelo de recolección, $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_n$ una m.a. de $\underline{X}$
- ¿Qué podemos hacer con todo esto?

Si el objetivo es estimar θ

i) Estimación puntual ✓
ii) Estimación por intervalo ←
iii) Prueba de hipótesis

Cambiamos la estimación puntual $\underline{\hat{\theta}_n}$ por un intervalo

$$(L_n, U_n), \quad \begin{aligned} L_n &= L(\underline{X}_1, \ldots, \underline{X}_n) \\ U_n &= U(\underline{X}_1, \ldots, \underline{X}_n) \end{aligned}$$

- Ventajas

- Interpretación

Teorema (intervalo normal asintótico)

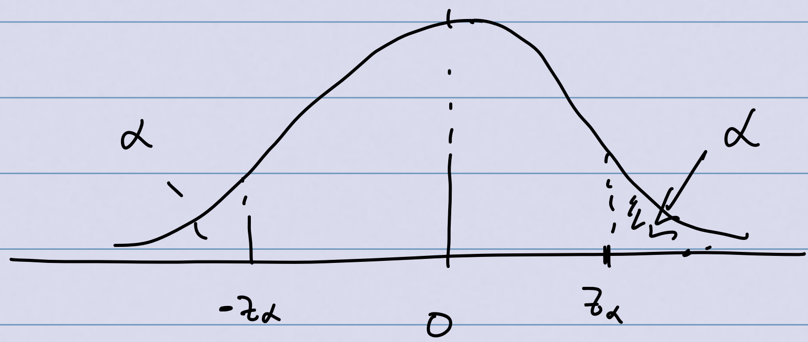
Sea $X_1, X_2, \dots, X_n$ una m.a. de $X \sim f(x|\theta)$ y $\hat{\theta}_n$ un estimador de θ, si se cumple que

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

$\Rightarrow$ para $\underline{n}$ suf grande el intervalo

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

En donde $P(Z \geq z_\alpha) = \alpha$



Entonces

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

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

Es necesario hablar un poco sobre los cuantiles de la distribución normal

De la definición, se tiene que

$$P(Z \geq z_\alpha) = \alpha \iff 1 - P(Z \leq z_\alpha) = \alpha$$

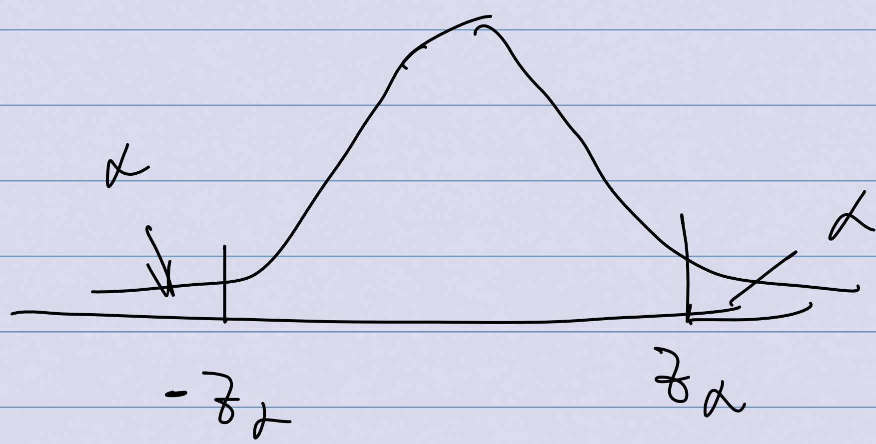
$$\iff P(Z \leq z_\alpha) = 1 - \alpha$$

$$F_Z(z_\alpha) = 1 - \alpha$$

$$\Rightarrow \quad z_\alpha = F_Z^{-1}(1-\alpha)$$

$$z_\alpha = qnorm(1-\alpha, 0, 1)$$

Además por simetría de $f_Z(z)$



$$P(Z \leq -z_\alpha) = P(Z \geq z_\alpha) = \alpha$$

<u>Dem.</u>

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

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

$$= 1 - \alpha/2 - \alpha/2 = \underline{1-\alpha}$$

Ahora,

$$Z_n = \frac{\hat{\theta}_n - \theta}{se(\hat{\theta}_n)} \xrightarrow{D} Z$$

$\Rightarrow$ para n suf grande

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

$$\Rightarrow P\left(\hat{\theta}_n - z_{\alpha/2}\, se(\hat{\theta}_n) \leq \theta \leq \hat{\theta}_n + z_{\alpha/2}\, se(\hat{\theta}_n)\right)$$
$$= 1-\alpha.$$

Ejemplo

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

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$ es mi estimador puntual,

pero además

$$Z_n = \frac{\hat{\theta}_n - \theta}{se(\hat{\theta}_n)} \longrightarrow Z \sim N(0,1)$$

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

$\Rightarrow$ El intervalo asintótico del $(1-\alpha) \times 100\%$ de confianza para θ está dado por

$$\left(\hat{\theta}_n - z_{\alpha/2}\sqrt{\frac{\theta(1-\theta)}{n}},\ \hat{\theta}_n + z_{\alpha/2}\sqrt{\frac{\theta(1-\theta)}{n}}\right)$$

Problema 1, el obvio

Problema 2, ¿Qué pasa si $\underline{\theta \approx 0 \text{ ó } \theta \approx 1}$?

Problema 3, el que siempre sucede en la práctica y mucha gente lo pone por escrito en sus reportes.

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

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

Necesitamos ① Una estadística con la que podamos medir desviaciones con respecto a lo que dice H_0

$\equiv$ estadística de prueba

$$T_0 = |\hat{\theta}_n - \tfrac{1}{2}|$$

(2) Necesitamos obtener la distribución de T_0

(3) Encontrar un $\underline{c}$, punto de corte — valor crítico — de manera que si

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

Con esto terminamos el tema 2. Modelos estadísticos del temario