Examen de la parte de estadística no paramétrica

Estadística no paramétrica

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

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

② Funcionales estadísticos

¿Qué son?

Cualquier función de $F(x) \longrightarrow T(F)$

y...¿hay alguno que sea medio conocido?

pues....

$$\mu = \mathbb{E}(X) = \int dF(x)$$

$$m = F^{-1}(1/2)$$

pensaría que todos los que conocen son funcionales estadísticos, i.e.

media, asimetría, correlación etc.

varianza, kurtosis

Existe una clase especial de funcionales estadísticos $T(F)$, los llamados funcionales estadísticos lineales. Cualquier funcional estadístico de la forma

$$T(F) = \int g(x)\, dF(x) = \int g(x) f(x)\, dx$$

es un funcional estadístico lineal.

¿Qué tienen de especial estos funcionales estadísticos lineales?

Teorema (estimador plug-in)

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

$$\theta = T(F) = \int g(x)\, dF(x)$$

un funcional estadístico lineal

$\Rightarrow$ para obtener un estimador de θ sólo metemos (plug-in) $\hat{F}_n(x)$ en donde veamos $F(x)$, i.e.

$$\hat{\theta}_n = T(\hat{F}_n) = \int g(x)\, d\hat{F}_n(x)$$

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

resultados de teoría de la medida

¡Sí, un promedio!

Ejemplo

1) Media

$$\mu = \mathbb{E}(X) = \int x \, dF(x)$$

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

2) Varianza

$$\sigma^2 = Var(X) = \int (x-\mu) \, dF(x)$$

$$= \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

$$= \int x^2 \, dF(x) - \mu^2$$

$$\Rightarrow \hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \hat{\mu}_n^2$$

Observen que

$$= \frac{1}{n} \sum (X_i - \hat{\mu}_n)^2$$

$$\frac{1}{n} \sum (X_i - \bar{X}_n)^2 = \frac{1}{n} \sum (X_i^2 - 2\bar{X}_n X_i + \bar{X}_n^2)$$

$$= \frac{1}{n} \sum X_i^2 - \frac{2}{n} \bar{X}_n \sum X_i + \bar{X}_n^2$$

$$= \frac{1}{n}\sum x_i^2 - 2\bar{X}_n^2 + \bar{X}_n^2$$

$$= \frac{1}{n}\sum x_i^2 - \bar{X}_n^2$$

3) Asimetría $K = \frac{\mathbb{E}((X-\mu)^3)}{\sigma^3} = \frac{\int (x-\mu)^3 \, dF(x)}{\sigma^3}$

$$\Rightarrow \hat{K}_n = \frac{\frac{1}{n}\sum (x_i - \hat{\mu}_n)^3}{\left(\frac{1}{n}\sum (x_i - \hat{\mu}_n)^2\right)^{3/2}}$$

4) Correlación

Sean $(X_1, Y_1), (X_2, Y_2), .., (X_n, Y_n)$ una m.a. de $(X, Y) \sim F(x,y)$

$$\rho = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = \frac{\mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))}{\sigma_X \; \sigma_Y}$$

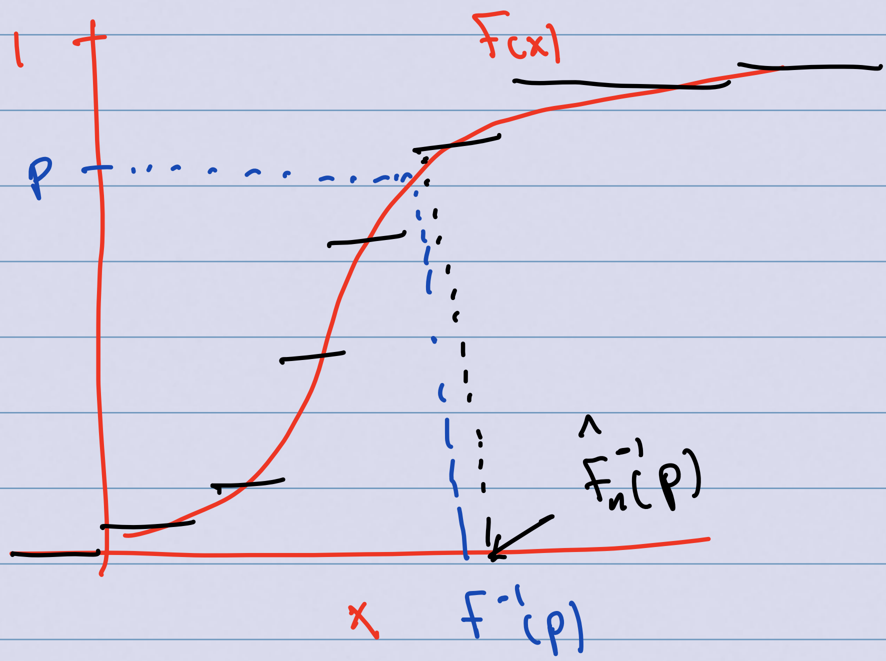
$$\hat{\rho}_n = \frac{\frac{1}{n}\sum (x_i - \hat{\mu}_X)(y_i - \hat{\mu}_Y)}{\hat{\sigma}_{nX} \; \hat{\sigma}_{nY}}$$

5) Cuantiles (no es un funcional estadístico lineal)

Sea $F(x)$ una función de distribución estrictamente creciente. Para $0<p<1$ el cuantil p-ésimo se define como

$$T(F) = F^{-1}(p) \Rightarrow T(\hat{F}_n) = \hat{F}_n^{-1}(p)$$

en donde $\hat{F}_n^{-1}(p) = \inf\{x : \hat{F}_n(x) \geq p\}$



Ya tenemos una forma de obtener estimadores puntuales

de muchos parámetros de una cierta población de interés

$X \sim F(x)$

$\theta = T(F)$

Pero, ¿cómo obtengo estimadores por intervalo?

si $\mu = \mathbb{E}(X)$ y sabemos que $\sigma^2 = Var(X) < \infty$

podemos utilizar directamente el TCL

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

$$se(\hat{\mu}_n) = \sigma / \sqrt{n}$$

$$\Rightarrow \begin{array}{l} L_n = \hat{\mu}_n - z_{\alpha/2}\, \sigma/\sqrt{n} \\ U_n = \hat{\mu}_n + z_{\alpha/2}\, \sigma/\sqrt{n} \end{array} \qquad \text{con } \sigma = \sqrt{Var(X)}$$

$$\hat{\sigma}_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2}$$

$\Rightarrow$ (L_n, U_n) sería un intervalo asintótico del $(1-\alpha)\times 100\%$ de confianza para μ

— Observen que NO hice ningún supuesto acerca de $F(x)$ —

Ok y si por alguna razón quiero el intervalo de confianza para

$$\sigma = \sqrt{Var(X)}$$

$$K = \frac{\mathbb{E}((X-\mu)^3)}{\sigma^3}$$

Tenemos $\hat{\sigma}_n$ y $\hat{K}_n$ fácilmente, pero "asumiendo" normalidad (no sabemos si eso es cierto)

$\Rightarrow$ aun es necesario

$$se(\hat{\sigma}_n) = \sqrt{Var(\hat{\sigma}_n)}$$
$$se(\hat{K}_n) = \sqrt{Var(K_n)}$$

$$Var\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu}_n)^2}\right)$$

El mayor reto (en estadística en general) suele ser estimar la varianza del estimador, i.e.

$$\theta = T(F)$$

$$\hat{\theta}_n = T(\hat{F}_n)$$

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

En estos casos utilizamos el Bootstrap.