- La página del curso ya está lista.

3: Modelos e inferencia estadística

Como vimos la clase pasada, en inferencia estadística siempre asumiremos que

① el fenómeno aleatorio de interés (la variable que nos interesa de nuestra población) es una v.a. X

Salario mensual de los egresados de la LCD del IIMAS $= X$

② $X \sim f(x|\theta)$

ó

$X \sim F(x)$

$X \sim LN(\mu, \sigma^2)$

③ Obtenemos datos de nuestra población

$x_{1:n} = x_1, x_2, \ldots, x_n$

④ Usamos esos datos para estimar la característica de X que nos interesa

- $\underline{\theta}$
- $\mu = \mathbb{E}(X)$
- $f(x|\theta)$
- $F(x)$
- $P(X > c)$

Es necesario describir con más detalle los pasos 2 y 3

En 2 estamos suponiendo que un modelo estadístico describe la variable de interés de nuestra población. Tendremos 2 clases de modelos

- Paramétricos
- No paramétricos

Un modelo estadístico M es simplemente un conjunto de distribuciones que en teoría pueden describir la variable de interés.

$$M = \left\{ f(x|\mu, \sigma^2) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-(\log(x)-\mu)^2/2\sigma^2}, \ \mu \in \mathbb{R} \text{ y } \sigma > 0 \right\}$$

modelo paramétrico pues lo podemos describir mediante 2 parámetros. μ y σ^2

Un modelo paramétrico puede describirse mediante un número finito de parámetros. En general los modelos paramétricos los describiremos como

$$M = \{ f(x|\theta) \mid \theta \in \Omega \}$$

en donde θ es el parámetro desconocido o vector de parámetros

$$\theta = (\theta_1, \theta_2, \ldots, \theta_k)$$

Un modelo estadístico no paramétrico es un conjunto $\underline{\underline{M}}$ que no se puede describir mediante un número finito de parámetros.

$\underline{\underline{M}}$ = { todas las funciones de distribución }

$\underline{\underline{M}}$ = { todas las funciones de distribución para v.a. continuas }

$\underline{\underline{M}}$ = { todas las funciones de distribución simétricas }

En 3 tenemos que considerar que los datos $x_1, x_2, \ldots, x_n$ están sujetos a la variabilidad de observaciones individuales y que si se obtiene otro conjunto de datos este no será el mismo

$\Rightarrow$ se asume que el modelo de recolección de la información es una muestra aleatoria = v.a.i.i.d. de $\underline{\underline{M}}$

$X_1, X_2, \ldots, X_n$ y que $x_1, x_2, \ldots, x_n$ es simplemente una realización de estas v.a.

Sea $X_1, X_2, \ldots, X_n$ una m.a. de $f(x|\theta)$

$\Rightarrow$ 1) $X_i \sim f(x|\theta)$ para $i = 1, \ldots, n$

2) $X_i \perp X_j \quad \forall\, i \neq j$

en este caso $\underbrace{f(x_1, x_2, \dots, x_n|\theta)}_{\text{la distribución conjunta de la m.a}} = \underbrace{\prod_{i=1}^{n} f(x_i|\theta)}_{\text{producto de las marginales}}$

y cada $f(x_i|\theta)$ es simplemente $f(x|\theta)$ evaluada en x_i.

Ahora describiremos con más detalle el paso 4

Existen 3 problemas principales que ataca la inferencia estadística:

i) Estimación puntual
ii) Estimación por intervalo
iii) Prueba de hipótesis.

i) Estimación puntual. Aquí el objetivo es obtener una única mejor estimación para un parámetro de interés.

- El parámetro que gobierna el comportamiento del modelo paramétrico $\longrightarrow \hat{\theta}_n$

- La distribución acumulada que aproxima a $F(x)$ en el modelo NO paramétrico $\hat{F}_n(x)$.

- La distribución que aproxima $f(x)$ en el modelo NO paramétrico. $\hat{f}_n(x)$

Vamos a concentrarnos en la estimación de θ usando $\underline{\underline{\hat{\theta}_n}}$

$$\hat{\theta}_n = \begin{cases} g(x_1, x_2, \ldots, x_n) & \leftarrow \text{estimación} \\ g(X_1, X_2, \ldots, X_n) & \leftarrow \text{estimador} \end{cases}$$

Tenemos el supuesto de que $X_1, X_2, \ldots, X_n$ es una m.a. de $f(x|\theta)$ y $x_1, x_2, \ldots, x_n$ son los valores observados de la m.a.

Primero construimos un <u>estimador</u> de θ que es una función únicamente de la m.a. y que usaremos para estimar θ.

$$\hat{\theta}_n = g(X_1, X_2, \ldots, X_n) \therefore \text{v.a.}$$

Utilizando la teoría de probabilidad podemos evaluar las cualidades del estimador $\hat{\theta}_n$ para estimar θ.

1) Sesgo. El sesgo es una desviación sistemática de lo que buscamos estimar y se mide como

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

Diremos que el estimador $\hat{\theta}_n$ es insesgado para estimar θ si

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

② Error cuadrático medio

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

si tenemos $\underline{2}$ estimadores de $\underline{\theta}$, $\hat{\theta}_n^1$ y $\hat{\theta}_n^2$

$\Rightarrow$ se prefiere el que tenga menor error cuadrático medio.

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

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

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

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

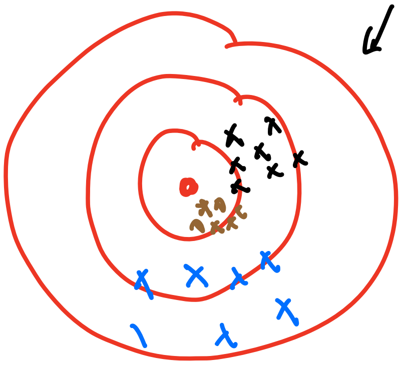
$$+ \mathbb{E}\left(\left(\mathbb{E}(\hat{\theta}_n) - \theta\right)^2\right)$$

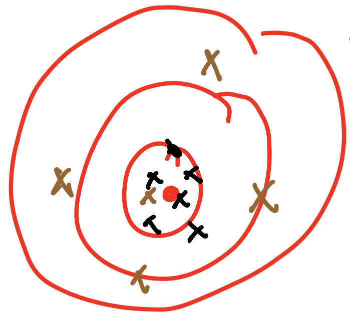
$$= \text{Var}(\hat{\theta}_n) + \left(\mathbb{E}(\hat{\theta}_n) - \theta\right)^2$$

$$= \text{Var}(\hat{\theta}_n) + \text{Sesgo}_\theta^2(\hat{\theta}_n)$$

$$ECM_\theta(\hat{\theta}_n) = Var(\hat{\theta}_n) + Sesgo^2_\theta(\hat{\theta}_n)$$

Sesgo, varianza y error cuadrático medio





③ Consistencia, una propiedad esencial de cualquier estimador es que al aumentar el tamaño de muestra la estimación vaya mejorando

$\hat{\theta}_n \xrightarrow[n\to\infty]{} \theta$, pero $\hat{\theta}_n$ es una v.a.
esto es muy restrictivo

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

$$\forall\, \varepsilon > 0$$

Ley débil de los grandes números (LDGN)

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

y $\mu = \mathbb{E}(X)$

$$\Rightarrow \bar{X}_n \xrightarrow{P} \mu$$

$\Rightarrow \bar{X}_n$ es un estimador consistente para μ.

④ Distribución de muestreo de $\hat{\theta}_n$ (asintótica)

¿Cómo se distribuye $\hat{\theta}_n$?

$\hat{\theta}_n \sim$ ¿?

¿Cómo se distribuye $\hat{\theta}_n$ para n suf grande?

Ejemplo

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

$$\hat{\theta}_n = \bar{X}_n$$

- Sesgo
- ECM
- Consistencia
- Distribución asintótica