Temario  30/01/2025

0 = Introducción y motivación

1 = Estadística descriptiva → Erick

2 = Modelos estadísticos

La clase pasada vimos "el proceso de la inferencia estadística"

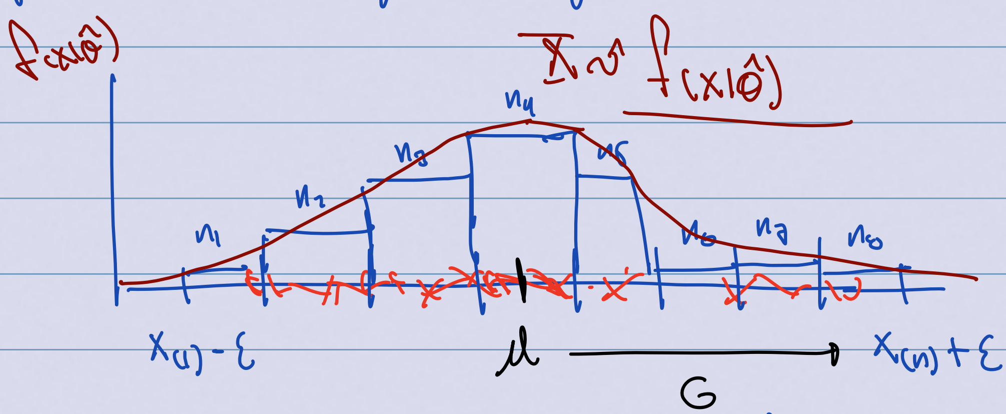
(1) FA que nos interesa conocer/describir/predecir
y podemos recolectar información generada por FA, i.e.
datos $x_1, x_2, x_3, \dots, x_n$

(2) Asumimos que el FA puede describirse mediante una v.a. $X$

¿Qué significa esto?

el comportamiento agregado de $x_1, x_2, \dots, x_n$
puede describirse mediante una v.a. $X$

pensemos en un histograma o gráfica de barras




pero aquí necesitamos decidir si $X \sim f(x|\theta)$ ó $X \sim F(x)$

(3) Para poder echar andar la maquinaria de la inferencia estadística, necesitamos asumir un modelo de recolección de datos, i.e.

los datos $x_1, x_2, \ldots, x_n$ fueron el resultado de algún mecanismo aleatorio → probabilidades

(4) Con estos supuestos podremos estimar

$$\mu = \mathbb{E}(X)$$
$$\sigma = \sqrt{Var(X)}$$
$$P(X > c)$$
$$P(a < X < b)$$ ← nosotros fijamos a y b

ó $P(a < X < b) = 0.95$

fijamos

y encontramos a y b

Quizá nos interese $f(x|\theta) \to f(x|\hat{\theta}_n)$

$F(x) \to \hat{F}_n(x)$

Vamos a describir con mayor detalle los pasos (2) y (3)

En el paso (2), necesitamos elegir/asumir un modelo estadístico que describa razonablemente bien a FA, de manera general tenemos 2 casos

- Modelo estadístico paramétrico
- Modelo estadístico no paramétrico

Por ejemplo

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

en este caso $M$ es un modelo paramétrico pues lo podemos describir mediante un número finito de parámetros $\longrightarrow (\mu, \sigma^2)$

En general a los modelos paramétricos los describiremos como

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

en donde $\theta = (\theta_1, \theta_2, .., \theta_K)$ es el parámetro que gobierna el comportamiento de $f(x|\theta)$ y $K < \infty$

Un modelo estadístico NO paramétrico, es un conjunto $M$ que no se puede describir mediante un número finito de parámetros

$M = \{$ todas las funciones de distribución acumulada $F(x) \}$

$M = \{$ todas las funciones " " $F(x)$ continuas $\}$

$M = \{ \quad \ldots \quad$ , simétricas $\}$

En los viejos días se decía que un modelo paramétrico tenía parámetros y un no paramétrico, no tenía parámetros.

Entonces llegaron las estadísticas bayesianas NO paramétricas y de alguna manera definieron

$$f(x) = \sum_{j=1}^{\infty} \omega_j N(x \mid \mu_j, \sigma_j^2)$$

utilizando el proceso Dirichlet

Ferguson T. S (1973) A Bayesian analysis of some non-parametric problems

Ahora vamos a describir con mayor detalle el punto ③, en este caso necesitamos definir un modelo de recolección de información que considere que los datos que tenemos $x_1, x_2, \ldots, x_n$

1) Están sujetos a variaciones individuales
2) Si obtenemos más información de FA, $x_1^*, \ldots, x_n^*$ no será la misma información que $x_1, \ldots, x_n$
3) La v.a. $X$ describe el comportamiento agregado de los datos generados por FA

Un modelo de recolección que cumple con lo anterior y que es sencillo de trabajar es el de

muestra aleatoria $=$ m.a. $\equiv$ v.a.i.i.d

Entonces asumimos que

1) $\underbrace{X_1, X_2, \ldots, X_n}_{\text{v.a.}}$ es una $\underline{\underline{\text{m.a}}}$ de $X \sim \begin{cases} f(x|\theta) \\ F(x) \end{cases}$

2) $\underbrace{x_1, x_2, \ldots, x_n}_{\text{valores observados al realizar la medición}}$ es una realización de $X_1, X_2, \ldots, X_n$

Del curso de probabilidad

$X_1, X_2, \ldots, X_n$ es una m.a de $X \sim f(x|\theta)$ si

1) $X_i \sim f(x|\theta), \; i=1,\ldots,n$ ← idénticamente distribuidas

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

Bajo nuestro modelo de recolección de información, la densidad conjunta de los datos $x_1, x_2, \ldots, x_n$ está dada por

independencia

$$f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n|\theta) = \prod_{i=1}^{n} f_{X_i}(x_i|\theta)$$

idénticamente distribuidas

$$= \prod_{i=1}^{n} f_{X}(x_i|\theta)$$

por $\underline{\text{Fisher}}$

$$= \prod_{i=1}^{n} f(x_i|\theta)$$

Ahora vamos a describir con mayor detalle el punto (4)

De manera muy general, podemos decir que la inferencia estadística nos da herramientas para resolver 3 problemas

i) Estimación puntual
ii) Estimación por intervalo
iii) Pruebas de hipótesis

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

Ej. $X \sim f(x|\theta) \rightarrow$ la mejor estimación $\hat{\theta}_n = h(x_1, ..., x_n)$ de $\theta$.

$X \sim F(x) \rightarrow$ la mejor aproximación a $F(x)$ para $x$ en $\mathbb{R}$,

$\rightarrow \hat{F}_n(x)$

$X \sim f(x) \rightarrow$ la mejor aproximación a $f(x)$, para $x$ en $\mathbb{R}$

$\hat{f}_n(x)$

Vamos a concentrarnos en el caso en el que $X_1, X_2, \dots, X_n$ es una m.a. de $X \sim f(x|\theta)$ y queremos encontrar el "mejor" estimador de $\underline{\theta}$

De alguna manera vamos a obtener un estimador

$$\hat{\theta}_n = h(X_1, X_2, \dots, X_n)$$ ← es una función de la m.a. que usamos para estimar $\theta$, es una v.a.

Como el estimador es una v.a. podemos analizar algunas propiedades usando teoría de v.a.s. Pero antes definimos la estimación

$$\hat{\theta}_n = h(x_1, x_2, \dots, x_n)$$ es el valor que usamos en la práctica para aproximar $\underline{\theta}$ y se obtiene insertando los valores observados $x_1, x_2, \dots, x_n$ en $\underline{h}$

⇒ Estimador ≡ v.a. → probabilidades

Estimación ≡ valor que usamos para aproximar $\theta$.