RESUMEN ESTADISTICA

CAPITULO 9 – Inferencia estadística

Def: Inferencia

Construir un modelo sobre una base de datos experimentales y extraer conclusiones.

Def: Muestra

Variable aleatoria X, definida sobre $(\Omega, \mathcal{A}, \mathbb{P})$ con distribución $\mathcal{F}_X(x) = \mathbb{P}_{(X \leq x)}$ que se desconoce (al menos parcialmente).

X → "Observable" del experimento aleatorio

Muestra aleatoria:

$X_1, X_2, \dots X_n$ todas iid a X.

Muestra observada:

$x_1, x_2, \dots x_n$

Con:

$$\mathcal{F}_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = \mathbb{P}_{(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)} = \prod_{i=1}^{n} \mathcal{F}_{X_i}(x_i)$$

Def: MODELOS PARAMETRICOS

Familia paramétrica de distribuciones:

$\mathcal{F} = \{F_\theta(x) : \theta \in \Theta\}$ sera una familia de distribuciones de probabilidad parametrizadas por un espacio paramétrico Θ.

Def: FUNCION DE VEROSIMILITUD

$\mathcal{L}_{(\theta)} = \prod_{i=1}^{n} f_\theta(x_i)$ si $\underline{X}$ es continuo

$\mathcal{L}_{(\theta)} = \prod_{i=1}^{n} p_\theta(x_i)$ si $\underline{X}$ es discreto

RESUMEN ESTADISTICA

CAPITULO 9 – Inferencia estadística

<u>Def:</u> FAMILIA PARAMETRICA REGULAR

Una familia paramétrica es **Regular** si:

1. El soporte de $f_\theta(x)$ no depende de θ
2. $f_\theta(x)$ es derivable con respecto a θ $\forall x$
3. El conjunto paramétrico $\Theta \in \mathbb{R}^p$ es abierto

---

<u>Def:</u> FAMILIAS EXPONENCIALES

Una familia de distribuciones es una familia exponencial si su función de densidad/probabilidad se puede escribir como:

$$f_\theta(x) = A_{(\theta)} \cdot e^{\sum_{i=1}^{k} C_i(\theta) \cdot r_i(x)} \cdot h_{(x)}$$

Donde:

$$C_i(\theta): \Theta \in \mathbb{R}$$

$$A(\theta): \Theta \in \mathbb{R}^+$$

$$r_i(x): \mathbb{R}^q \in \mathbb{R}$$

$$h(x): \mathbb{R}^q \in \mathbb{R}^+$$

<u>Def:</u> ESTADISTICOS

Un estadístico es cualquier función medible $T_n = T(\underline{X})$ con valores en un espacio euclídeo de dimensión finita.

Fácil: Dada una m.a. $\underline{X}$, un estadístico es una función de la m.a. que, evaluada en los valores observados, debe poder resultar en un valor numérico.

Obs: esta función NO puede depender de parámetros desconocidos.

RESUMEN ESTADISTICA

CAPITULO 9 – Inferencia estadística

<u>Def:</u> ESTADISTICOS SUFICIENTES

Un estadístico $T = r_{(\underline{X})}$ es suficiente para θ si:

$p_{\underline{X}|T=t}(\underline{x})$ o $f_{\underline{X}|T=t}(\underline{x})$ es independiente de θ, para todo t.

<u>Def:</u> TEOREMA DE FACTORIZACION

Sean dos funciones h y g tales que:

$$f_\theta(x) = g_{(r(\underline{x}),\theta)} \cdot h_{(\underline{x})}$$

$T = r(\underline{X})$ es un estadístico suficiente de <u>X</u>

<u>Def:</u> TEOREMA DE ESTADISTICOS PARA FLIAS EXPONENCIALES

Sean $X_1, X_2, \dots, X_n$una m.a. de una distribución perteneciente a una flia exponencial a k parámetros., entonces el estadístico suficiente para θ basado en la m.a. será:

$$T = \left(\sum_{i=1}^{n} r_1(X_i), \sum_{i=1}^{n} r_2(X_i), \dots, \sum_{i=1}^{n} r_k(X_i)\right)$$

<u>Def:</u> ESTIMADOR DE MAXIMA VEROSIMILITUD

A partir de $\mathcal{L}_{(\theta)}$busco el valor de θ que maximiza dicha función.

<u>Def:</u> PRINCIPIO DE INVARIANZA

Suponiendo $\lambda = q(\theta)$ una función biunívoca de θ. Si $\hat{\theta}$ es el EMV de θ, entonces $\hat{\lambda} = q(\hat{\theta})$

RESUMEN ESTADISTICA

CAPITULO 9 – Inferencia estadística

<u>Def:</u> BONDAD DE LOS ESTIMADORES

Se mide con:

$$\mathcal{R}\left(\theta,\hat{\theta}\right)=ECM\left(\hat{\theta}\right)=\mathbb{E}_{\left[\left(\theta-\hat{\theta}\right)^{2}\right]}=Var_{\theta}\left(\hat{\theta}\right)+\mathfrak{B}\left(\hat{\theta}\right)^{2}$$

Un **<u>ESTIMADOR</u> <u>ÓPTIMO</u>** para θ será $\hat{\theta}$ tal que

$$ECM\left(\hat{\theta}^{*}\right)\leq ECM\left(\hat{\theta}\right),\forall\theta$$

Un **<u>ESTIMADOR</u> <u>INSESGADO</u>** para θ será $\hat{\theta}$ tal que

$$\mathbb{E}_{\theta}\left[\hat{\theta}\right]=\theta,\forall\theta$$

Un **<u>ESTIMADOR</u> <u>ASINTOTICAMENTE</u> <u>INSESGADO</u>** para θ será $\hat{\theta}$ tal que

$$\lim_{n\to\infty}\mathbb{E}_{\theta}\left[\hat{\theta}\right]=\theta,\forall\theta$$

Con un **<u>ESTIMADOR</u> <u>SESGADO</u>** el sesgo se define como:

$$\mathfrak{B}\left(\hat{\theta}\right)=\mathbb{E}_{\left[\hat{\theta}\right]}-\theta$$

<u>Def:</u> CONSISTENCIA

Dada una sucesión de estimadores $\widehat{\theta}_{n}$de θ, decimos que $T=\widehat{\theta}$ es (débilmente) **consistente** si $\forall\mathcal{E}>0$

$$\mathbb{P}(|T-\theta|>\mathcal{E})\xrightarrow{n\to\infty}0$$

RESUMEN ESTADISTICA

CAPITULO 9 – Inferencia estadística

<u>Def:</u> TEOREMA DE CONSISTENCIA

Dada una sucesión de estimadores $\hat{\theta}_n$de θ. Si $Var_\theta(\hat{\theta}) \longrightarrow 0$ y $\mathbb{E}_\theta(\hat{\theta}) \longrightarrow \theta$, entonces $\hat{\theta}_n$ es débilmente consistente

---

<u>Def:</u> ESTIMADORES ASINOTITCAMENTE NORMALES

Se dice que $\hat{\theta}_n$ es una sucesión de estimadores asintóticamente normales si $\sqrt{n} \cdot \left(\hat{\theta}_n - q(\theta)\right)$ converge en distribución a una normal con media cero y varianza $q'(\theta)/I(\theta)$.

$I(\theta)$ se llama al **Numero de información de Fischer** y se calcula como:

$$I(\theta) = \mathbb{E}\left[\left(\frac{d}{d\theta}\ln\left(f_\theta(X)\right)\right)^2\right]$$

$$I(\theta) = -\mathbb{E}\left[\frac{d^2}{d\theta^2}\ln\left(f_\theta(X)\right)\right]$$

(solo para familias regulares)

<u>Def:</u> ESTIMADORES ASINOTITCAMENTE NORMALES

Bajo ciertas condiciones muy generales, sea $\hat{\theta}_n(\underline{X})$ un EMV de θ consistente y sea $q(\theta)$ derivable con $q'(\theta) \neq 0 \;\; \forall\theta$, entonces $q'(\hat{\theta}_n)$ es asintóticamente normal para estimar $q(\theta)$.

Si $\sqrt{n} \cdot \sqrt{I(\theta)} \cdot (\hat{\theta}_n - \theta) \sim^a \mathcal{N}(0,1)$ y $\hat{\theta}_n$ es un estimador consistente para $\hat{\theta}$,

Entonces vale que:

$$\sqrt{n} \cdot \sqrt{I(\hat{\theta}_n)} \cdot (\hat{\theta}_n - \theta) \sim^a \mathcal{N}(0,1)$$

RESUMEN ESTADISTICA

CAPITULO 9 – Inferencia estadística

<u>Def:</u> TEOREMA CHI CUADRADO

Si $X_1, X_2, \dots, X_n$ son iid con $X_i \sim \mathcal{X}^2_{v_i}$, entonces $Y = \sum_{i=1}^{n} X_i$ tendrá distribución $\mathcal{X}^2_{v}$, con:

$$v = \sum_{i=1}^{n} v_i$$

<u>Def:</u> COROLARIO CHI CUADRADO

Se llama distribución $\mathcal{X}^2$ con n grados de libertad a la distribución de

$$U = \sum_{i=1}^{n} {Z_i}^2 \;\; donde\; Z_1, Z_2, \dots, Z_n \sim \mathcal{N}(0,1)$$

<u>Def:</u> DISTRIBUCION T DE STUDENT

Sean $Z \sim \mathcal{N}(0,1)\; y\; U \sim \mathcal{X}^2_n$, entonces si Z y U son independientes:

$$T = \frac{Z}{\sqrt{U/n}} \sim t_n$$

<u>Def:</u> DISTRIBUCION F DE FISHER – SNEDECOR

Sean U y V v.a. indep con distribución $\mathcal{X}^2_{n_1}$ y $\mathcal{X}^2_{n_2}$ respectivamente. Entonces:

$$F = \frac{U/n_1}{V/n_2} \sim \mathfrak{F}_{n_1,n_1}$$

RESUMEN ESTADISTICA

CAPITULO 9 – Inferencia estadística

<u>Def:</u> TEOREMAS

Sean $X_1, X_2, \dots, X_n \sim^{iid} \mathcal{N}(\mu, \sigma^2)$

1- $Z = \sqrt{n}\frac{(\bar{X}-\mu)}{\sigma} \sim \mathcal{N}(0,1)$

2- $W = \sum_{i=1}^{n} \frac{(X_i-\bar{X})^2}{\sigma^2} \sim \mathcal{X}_{n-1}^2$

3- Z y W son independientes.

4- Si $S^2 = \sum_{i=1}^{n} \frac{(X_i-\bar{X})^2}{n-1}$, entonces:

$$T = \sqrt{n}\frac{(\bar{X}-\mu)}{S} \sim t_{n-1}$$