PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
Facultad de Matemáticas
Departamento de Estadística
Segundo semestre 2022

EYP2127 Inferencia Estadística
Ayudantía 2: Estadístico suficiente, minimal, ancilar y completo.

Profesora: Inés M. Varas
Ayudante: Borja Márquez de la Plata

Ejercicio 1

Se esta realizando un estudio sobre la acumulación de agua debido a lluvias en reservas a lo largo del país. Para este año se observa, el mismo día del invierno, el nivel de agua acumulada en n reservas. Se define la variable aleatoria X_i como el nivel del agua en la reserva i. Se asume que el nivel del agua distribuye Gamma(α, β), y que el agua acumulada entre las diferentes reservas es independiente a lo largo del país:

a) Encuentre un estadístico suficiente bidimensional para $\theta = (\alpha, \beta)$

b) Verifique que el estadístico encontrado en a) sea suficiente minimal.

Ejercicio 2

Considere los tiempos de falla de los componentes eléctricos encontrados en las diferentes estaciones del metro de Santiago. Estos tiempos se modelan según una función de densidad dada por:

$$f(x|\theta) = \begin{cases} e^{-(x-\theta)} & x > \theta, \\ 0 & x \leq \theta \end{cases}$$

Donde $\theta \in \mathbb{R}$

a) Verifique que $T = \sum_{i=1}^{n} X_i$ no es un estadístico suficiente.

b) Encuentre un estadístico suficiente y verifique si es también minimal.

c) ¿Es la familia del estadístico suficiente encontrado una familia completa?

Ejercicio 3

Se tiene una muestra aleatoria donde X_1, X_2, ..., X_n distribuyen iid Normal(θ, θ^2):

a) Encuentre un estadístico suficiente minimal para θ.

b) Justifique por que ni $T_1 = \sum_{i=1}^{n} X_i$ ni $T_2 = \sum_{i=1}^{n} X_i^2$ pueden ser suficientes para estimar θ.

c) ¿Es la familia $T = (T_1, T_2)$ una familia completa para estimar θ?

Ejercicio 4

Sea N una variable aleatoria tomando los valores $1, 2, ..., n$ con probabilidades conocidas $p_1, p_2, ..., p_n$, las cuales suman 1. Si se observa $N = n$, se realizan n intentos Bernoulli(θ), donde θ es la probabilidad de éxito θ y se obtienen X éxitos.
Demuestre que el par (X, N) es suficiente minimal y que N es ancilar para θ.

Resumen:

Estadístico: Función de la muestra

Tipos:

→ **Suficiente:**

Suficiente para θ la distribución condicional de la muestra dado el valor del estadístico no depende de θ

Es decir:

$T(x)$ es suf. si $x|T$ distribuye ind. de θ

basta con demostrar

$f(x|\theta)/q(T(x)|\theta)$ no depende de θ

Teorema de Factorización

$$f(x|\theta) = g(T(x)|\theta)\cdot h(x)$$

→ **Minimal:**

Un estadístico $T(x)$ es suficiente minimal si, para cualquier otro estadístico $T'(x)$, $T(x)$ es función de $T'(x)$

si la razón $\frac{f(x|\theta)}{f(y|\theta)}$ es constante como función de θ

Si y solo si $T(x) = T(y)$, $\Rightarrow T(x)$ es S.M. para θ

→ **Ancilar:**

Un estadístico $S(x)$ es ancilar si su distribución no depende del parametro

→ Completitud:

Sea $\{g(T|\theta): \theta \in \Theta\}$ una familia de fdp para T. Esta se denomina completa si $\mathbb{E}[h(T)]=0 \ \forall \theta$ implica que $h(t)=0 \ \forall t$. En este caso se dice que T es un estadístico completo.

$$\int_{t \in \mathcal{T}} h(t) \cdot g(t|\theta)\, dt = 0 \ \forall \theta \implies h(t)=0$$

Ejercicio 1

Se esta realizando un estudio sobre la acumulación de agua debido a lluvias en reservas a lo largo del país. Para este año se observa, el mismo día del invierno, el nivel de agua acumulada en n reservas. Se define la variable aleatoria X_i como el nivel del agua en la reserva i. Se asume que el nivel del agua distribuye Gamma(α, β), y que el agua acumulada entre las diferentes reservas es independiente a lo largo del país:

a) Encuentre un estadístico suficiente bidimensional para $\theta = (\alpha, \beta)$

b) Verifique que el estadístico encontrado en a) sea suficiente minimal.

a) $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Gamma}(\alpha, \beta)$

Sea $\theta = (\alpha, \beta)$

$$f(x|\theta) = \prod_{i=1}^{n} \frac{\beta^{\alpha}}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-\beta x_i}$$

$$= \left(\frac{\beta^{\alpha}}{\Gamma(\alpha)}\right)^{n} \left(\prod_{i=1}^{n} x_i\right)^{\alpha-1} e^{-\beta \sum_{i=1}^{n} x_i}$$

$$g(x|\alpha,\beta) = \left(\frac{\beta^{\alpha}}{\Gamma(\alpha)}\right)^{n} \left(\prod_{i=1}^{n} x_i\right)^{\alpha-1} e^{-\beta \sum_{i=1}^{n} x_i}$$

$$h(x) = 1$$

$T(x) = \left(\prod_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i\right)$ es suf. para θ

b)
$$\frac{f(x|\theta)}{f(y|\theta)} = \frac{\prod_{i=1}^{n} \cancel{\frac{\beta^{\alpha}}{\Gamma(\alpha)}} x_i^{\alpha-1} e^{-\beta x_i}}{\prod_{i=1}^{n} \cancel{\frac{\beta^{\alpha}}{\Gamma(\alpha)}} y_i^{\alpha-1} e^{-\beta y_i}} = \frac{\left(\prod_{i=1}^{n} x_i\right)^{\alpha-1} e^{-\beta \sum x_i}}{\left(\prod_{i=1}^{n} y_i\right)^{\alpha-1} e^{-\beta \sum y_i}}$$

$$= \left(\prod_{i=1}^{n} x_i \Big/ \prod_{i=1}^{n} y_i\right)^{\alpha-1} \cdot e^{-\beta\left(\sum x_i - \sum y_i\right)}$$

$\Rightarrow$ Si $\prod_{i=1}^{n} x_i = \prod_{i=1}^{n} y_i \quad \wedge \quad \sum x_i = \sum y_i$

$\dfrac{f(x|\theta)}{f(y|\theta)}$ es constante

Ejercicio 2

Considere los tiempos de falla de los componentes eléctricos encontrados en las diferentes estaciones del metro de Santiago. Estos tiempos se modelan según una función de densidad dada por:

$$f(x|\theta) = \begin{cases} e^{-(x-\theta)} & x > \theta, \\ 0 & x \leq \theta \end{cases}$$

Donde $\theta \in \mathbb{R}$

a) Verifique que $T = \sum_{i=1}^{n} X_i$ no es un estadístico suficiente.

b) Encuentre un estadístico suficiente y verifique si es también minimal.

c) ¿Es la familia del estadístico suficiente encontrado una familia completa?

a) $f(x|\theta) = e^{-(x-\theta)} \mathbb{1}_{(\theta,\infty)}(x)$

para la muestra

$$f(\underline{x}|\theta) = e^{-\sum(x_i-\theta)} \prod_{i=1}^{n} \mathbb{1}_{(\theta,\infty)}(x_i)$$

$$= \underbrace{e^{n\theta - \sum x_i}}_{g(T|\theta)} \underbrace{\mathbb{1}_{(-\infty, x_{(1)})}(\theta)}_{h(x),\ \text{depende de } \theta}$$

No es suficiente.

b) $f(\underline{x}|\theta) = e^{n\theta - \sum x_i} \mathbb{1}_{(-\infty, x_{(1)})}(\theta)$

$$= \underbrace{e^{-\sum x_i}}_{h(x)} \cdot \underbrace{e^{n\theta} \mathbb{1}_{(-\infty, x_{(1)})}(\theta)}_{g(T|\theta)}$$

$T(x) = x_{(1)}$ es suficiente

minimal:

$$\frac{f(x|\theta)}{f(y|\theta)} = \frac{e^{-\sum x_i} \cdot \cancel{e^{n\theta}} \mathbb{1}_{(-\infty, x_{(1)})}(\theta)}{e^{-\sum y_i} \cdot \cancel{e^{n\theta}} \mathbb{1}_{(-\infty, y_{(1)})}(\theta)}$$

es constante para θ si $x_{(1)} = y_{(1)}$ $\therefore$ es minimal

$U = x_{(1)}$

c) $U = \min\{\underline{X}\}$

$$f_U(u) = n\, f_X(u)\left[1 - F_X(u)\right]^{n-1}$$

$$f_X(u) = e^{-(u-\theta)} \mathbb{1}_{(\theta,\infty)}(u)$$

$$F_X(u) = \int_\theta^u e^{-(x-\theta)}\,dx = 1 - e^{-(u-\theta)} \quad \text{si } u \geq \theta$$

$$\therefore f_U(u) = n e^{-n(u-\theta)} \mathbb{1}_{(\theta,\infty)}(u)$$

Sea $g(U)$ una func. tq $\mathbb{E}[g(U)] = 0 \Rightarrow$

$$\int_\theta^\infty g(u)\, n e^{-n(u-\theta)}\, du = 0$$

$$\int_\theta^\infty g(u) e^{-nu}\, du = 0$$

$$\frac{d}{d\theta}\int_\theta^\infty g(u) e^{-nu}\, du = \frac{d}{d\theta} 0$$

$$-g(\theta) e^{-n\theta} = 0$$

$$g(\theta) = 0$$

$\therefore$ la fam de probabilidades de $U = X_{(1)}$ es completa.

Ejercicio 3

Se tiene una muestra aleatoria donde $X_1, X_2, ..., X_n$ distribuyen iid Normal(θ, θ^2):

a) Encuentre un estadístico suficiente minimal para θ.

b) Justifique por que ni $T_1 = \sum_{i=1}^{n} X_i$ ni $T_2 = \sum_{i=1}^{n} X_i^2$ pueden ser suficientes para estimar θ.

c) ¿Es la familia $T = (T_1, T_2)$ una familia completa para estimar θ?

a) $$f(x|\theta) = \frac{1}{\sqrt{2\pi\theta^2}} \exp\left\{-\frac{1}{2\theta^2}(x-\theta)^2\right\}$$

$$\frac{f(\underline{x}|\theta)}{f(\underline{y}|\theta)} = \frac{(2e\pi\theta^2)^{-n/2} \exp\left\{\frac{1}{\theta}\sum_{i=1}^{n} x_i - \frac{1}{2\theta^2}\sum_{i=1}^{n} x_i^2\right\}}{(2e\pi\theta^2)^{-n/2} \exp\left\{\frac{1}{\theta}\sum_{i=1}^{n} y_i - \frac{1}{2\theta^2}\sum_{i=1}^{n} y_i^2\right\}}$$

$$= \exp\left\{\frac{1}{\theta}\left(\sum x_i - \sum y_i\right) - \frac{1}{2\theta^2}\left(\sum x_i^2 - y_i^2\right)\right\}$$

$\therefore$ si $\sum x_i = \sum y_i \wedge \sum x_i^2 = \sum y_i^2 \Rightarrow$

$T = (\underbrace{\sum x_i}_{T_1}, \underbrace{\sum x_i^2}_{T_2})$ es ESM.

b) ya que T es Suf. minimal, todo otro estadístico suf. T' se puede escribir como $T = g(T')$ con g función 1-1. ya que no existe g que lleve a T_1 a T como T_2 a T, no pueden ser suf.

c) Sea $h(T)$ una función tal que $\mathbb{E}[h(T)] = 0$

$$\mathbb{E}[T_1] = \mathbb{E}[\sum x_i] = \sum \mathbb{E}[x_i] = n\theta$$

$$\mathbb{V}[T_1] = \mathbb{V}[\sum x_i] = n\theta^2$$

$$\mathbb{E}[T_1^2] = \mathbb{V}[T_1] + \mathbb{E}[T_1]^2 = n(n+1)\theta^2$$

$$\mathbb{E}[T_2] = \mathbb{E}\left[\sum x_i^2\right] = \sum \mathbb{E}[x_i^2] = 2n\theta^2$$

$$Z = \frac{x_i - \theta}{\theta} \quad \overset{(\;)^2}{\Rightarrow} \quad Z^2 \cdot \theta^2 = x_i^2 - 2x_i\theta + \theta^2$$

$$Z \sim N(0,1) \qquad X_i^2 = Z^2 \cdot \theta^2 + 2x_i\theta - \theta^2$$

$$Z^2 \sim \chi^2_{(1)} \qquad \mathbb{E}[x_i^2] = \theta^2\, \mathbb{E}[Z^2] + 2\theta \cdot n\theta - \theta^2$$

$$\mathbb{E}[x_i^2] = \cancel{\theta^2} \cdot 1 + 2n\theta^2 - \cancel{\theta^2}$$

$$= 2n\theta^2$$

definimos

$$h(T) = (n+1)\, T_2 - 2\, T_1^2 \neq 0$$

$$\mathbb{E}[h(T)] = (n+1)\, \mathbb{E}[T_2] - 2\, \mathbb{E}[T_1^2]$$

$$= (n+1) \cdot 2n\theta^2 - 2\, n(n+1)\theta^2 = 0$$

$\therefore$ la familia no puede ser completa. ya que encontramos una función $h(T) \neq 0$ tq $\mathbb{E}[h(T)] = 0$

Ejercicio 4

Sea N una variable aleatoria tomando los valores $1, 2, ..., n$ con probabilidades conocidas $p_1, p_2, ..., p_n$, las cuales suman 1. Si se observa $N = n$, se realizan n intentos Bernoulli(θ), donde θ es la probabilidad de éxito θ y se obtienen X éxitos.
Demuestre que el par (X, N) es suficiente minimal y que N es ancilar para θ.

$$\frac{f(x,n|\theta)}{f(y,n'|\theta)} = \frac{f(x|\theta, N=n)\cdot P(N=n)}{f(y|\theta, N=n')\cdot P(N=n')}$$

$$= \frac{\binom{n}{x}\theta^{x}(1-\theta)^{n-x}p_n}{\binom{n'}{y}\theta^{y}(1-\theta)^{n'-y}p_{n'}}$$

$$= \theta^{x-y}(1-\theta)^{n-n'-x+y}\frac{\binom{n}{x}p_n}{\binom{n'}{y}p_{n'}}$$

$\therefore$ si $x=y \wedge n=n'$ (X, N) es suf. min

Como $P(N=n)=p_n$ no depende de θ, N es ancilar

Notar que, aunque N no depende de θ, el estadístico suf. minimal contiene a N.