Suficiencia

Sea $\mathbf{X} = (X_1, \dots, X_n) \sim F(\mathbf{x}; \theta) \in \mathcal{F} = \{F(\cdot; \theta) : \theta \in \Theta\}$
con $\Theta \subset \mathbb{R}^p$.

Def Un estadístico $T(\mathbf{X})$ es suficiente para θ ssi la distribución de $\mathbf{X}$ condicional a $T(\mathbf{X}) = t$ es independiente de θ para cualquier valor de t.

Ej $X_1 \dots X_n$ iid $f(x; \theta) = \theta^{x}(1-\theta)^{1-x}$ $\quad x \in \{0,1\}$, $\theta \in [0,1]$

$$f_{X_1 \dots X_n}(x_1, x_2 \dots x_n; \theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} =$$

$$= \theta^{\sum x_i}(1-\theta)^{n - \sum x_i}$$

Proponemos $T(X_1 \dots X_n) = \sum_{i=1}^{n} X_i$ $\qquad f_T(t) = \binom{n}{t}\theta^{t}(1-\theta)^{n-t}$

$$f_{X_1 \dots X_n, T}(x_1 \dots x_n, t; \theta) = P(X_1 = x_1; \dots; X_n = x_n; \textstyle\sum X_i = t)$$

$$= \begin{cases} \theta^{\sum x_i}(1-\theta)^{n-\sum x_i} & t = \sum_{i=1}^{n} x_i \\ 0 & t \neq \sum x_i \end{cases}$$

$$f_{X_1 \dots X_n | T}(x_1 \dots x_n, t; \theta) = \begin{cases} 1/\binom{n}{t} & \\ 0 & t \neq \sum x_i \end{cases}$$

$X_1 \dots X_n \mid T = t$

no depende de θ.

$\therefore$ $T(X_1 \dots X_n) = \sum_{i=1}^{n} X_i$ es suficiente.

Notar 1) Hay muchos estadísticos suficientes

En el ej. $\tilde{T}(X_1 \dots X_n) = \left(\sum_{i=1}^{m} X_i ; \sum_{m+1}^{n} X_i\right)$

ii) La muestra entera $T(X_1 \dots X_n) = (X_1, \dots, X_n)$ es siempre trivialmente suficiente.

iii) Si $X_1 \dots X_n$ son i.i.d entonces los estadísticos de orden $T = (X_{(1)} \dots X_{(n)})$ son también suficientes.

iv) Toda transformación 1-1 de un estadístico suficiente es también suficiente.

En el ej $\hat{p}_n = \frac{\sum X_i}{n}$

¿Cómo encontrar estadísticos suficientes?

Teorema de Factorización (Fisher - Neyman)

Sea $X = (X_1 \dots X_n) \sim f(x, \theta)$ para $\theta \in \Theta$. $T(X)$ es suficiente para θ ssi. existen funciones g y h tales que

$$f(x;\theta) = \underbrace{g(T(x);\theta)}_{\text{depende del parametro y de la muestra solo a través de } T} \overbrace{h(x)}^{\text{independiente de } \theta} \qquad (*)$$

(caso discreto)

Dem i) ($\Leftarrow$) Supongamos que existen g y h tales que $(*)$ se cumple. Entonces la densidad conjunta

$$f_{X,T}(x,t;\theta) = \begin{cases} g(t;\theta)\, h(x) & T(x)=t \\ 0 & c.c. \end{cases}$$

y la marginal

$$f_T(t;\theta) = \sum_{\{x : T(x)=t\}} g(t;\theta)\, h(x)$$

$$= g(t;\theta) \sum_{\{\dots\}} h(x) \qquad h^*(x)$$

$$f_{X|T}(x,t;\theta) = \frac{f_{X,T}(x,t;\theta)}{f_T(t;\theta)} = \begin{cases} \dfrac{h(x)}{h^*(x)} & T(x)=t \\ 0 & c.c. \end{cases}$$

independiente de θ $\therefore$ $T(X)$ es suficiente.

ii) $(\Rightarrow)$ $P_\theta(X=x) = P_\theta(X=x; T(X)=T(x))$

$$= f_{X,T}(x; T(x); \theta)$$

$$= \underbrace{f_{X|T}(x, T(x), \theta)}_{h(x)} \ \underbrace{f_T(t;\theta)}_{g(t,\theta)} \qquad \diamond$$

Notar: tb valido

caso continuo

mixto

modelos no parametricos. dificil porque preciso medida dominante. Dist condicionales deben ser coherentes.

Ej $f(x_1 \ldots x_n; \theta) = \theta^{\sum x_i} (1-\theta)^{n-\sum x_i}$ $\qquad T = \sum x_i$

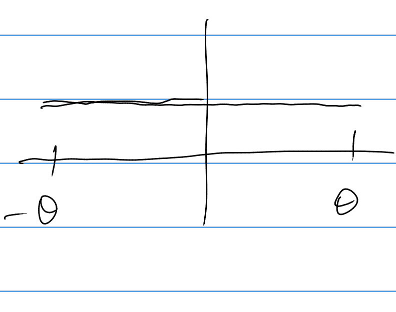
$$= \underbrace{\theta^T (1-\theta)^{n-T}}_{g(T,\theta)} \cdot \underbrace{1}_{h(x_1 \ldots x_n)}$$

$T = (x_1, \ldots, x_n)$

$T = (x_{(1)}, \ldots, x_{(n)})$ $\qquad \sum x_i = \sum x_{(i)}$

Ej $x_1 \ldots x_n \overset{iid}{\sim} U(-\theta, \theta)$

$$f_X(x;\theta) = \begin{cases} \dfrac{1}{2\theta} & -\theta \le x \le \theta \\ 0 & c.c. \end{cases}$$



$$U = \sin\left(\frac{|X_1|}{|X_{(n)}|}\right) \qquad T = \max\{|X_i|\}$$

$$f(x;\theta) = \frac{1}{2\theta}\, 1(|x| \leq \theta)$$

$$f(x_1 \ldots x_n;\theta) = \prod_{i=1}^{n} \frac{1}{2\theta}\, 1(|X_i| \leq \theta) = \underbrace{\left(\frac{1}{2\theta}\right)^n 1\left(\max\{|X_i|\} \leq \theta\right)}_{g(T,\theta)} \qquad 1 = h(x)$$

$$\underset{3}{\max\{|X_i|\}} \neq \underset{2}{|X_{(n)}|} \qquad -3,\ 1,\ 0,\ 2$$

$T_1 = \max\{|X_i|\}$ es suficiente.

Notar Otros estadísticos suficientes son

$$T_2 = (|X_1|; \ldots, |X_n|) \qquad T_1 = \max\{T_{21}, \ldots T_{2n}\}$$

$$T_3 = (X_{(1)}; X_{(n)}) \qquad T_1 = \max\{|T_{31}|; |T_{32}|\}$$

Cor Sea $X \sim f(x;\theta) \in \{f(\cdot;\theta)\}$ con soporte común e independiente de θ. Entonces $T(X)$ es suficiente sii $\forall\, \theta_1, \theta_2 \in \Theta$

$$\frac{f(x;\theta_1)}{f(x;\theta_2)} = H(T(x))$$

función del estadístico suficiente pero no de θ.

Teorema Sea $T(X)$ suficiente y $\tilde{T}(X) = H(T(X))$ con H biunívoca. Entonces $\tilde{T}(X)$ es suficiente.

Dem

$$\begin{aligned} f_X(x;\theta) &= g(T(x);\theta)\, h(x) \\ &= \underbrace{g(H^{-1}(\tilde{T}(x));\theta)}\, h(x) \\ &= g^{*}(\tilde{T}(x);\theta)\, h(x) \end{aligned}$$

$\tilde{T}(X)$ es suficiente $\Diamond$

$\underline{Eg}$ $X_1 \dots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ $f_X(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$

$$f_{X_1\dots X_n}(x_1 \dots x_n;\mu,\sigma^2) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2} \underbrace{\sum_{i=1}^n (x_i-\mu)^2}\right\}$$

$T = \cancel{\sum (x_i - \mu)^2}$ $n\bar{x}$

$$= (2\pi)^{-n/2}\, \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum x_i^2 - 2\mu \sum x_i + n\mu^2\right]\right\}$$

$$T(X_1 \dots X_n) = (\underbrace{\sum X_i}_{t_1};\ \underbrace{\sum X_i^2}_{t_2})$$

$$= \underbrace{(2\pi)^{-n/2}\, \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\left[t_2 - 2\mu t_1 + n\mu^2\right]\right\}}_{g(t_1, t_2;\mu,\sigma^2)}$$

$\tilde{T} = (\tilde{T}_1;\tilde{T}_2)$ $\tilde{T}_1 = T_1/n = \sum X_i/n = \bar{X}$

$\tilde{T}_2 = \sum (X_i - \bar{X})^2/n-1$

$$= \frac{1}{n-1}\left[\sum X_i^2 - n\bar{X}^2\right] = \frac{1}{n-1}\left[T_2 - \frac{T_1^2}{n}\right]$$

$$T = \begin{pmatrix} \sum X \\ \sum X^2 \end{pmatrix} \overset{1-1}{\longleftrightarrow} \tilde{T} = \begin{pmatrix} \bar{X} \\ S^2 \end{pmatrix}$$

$\tilde{t}_1 = t_1/n$ $\frac{\partial \tilde{t}_1}{\partial t_1} = \frac{1}{n}$ $\frac{\partial \tilde{t}_1}{\partial t_2} = 0$

$\tilde{t}_2 = \frac{1}{n-1}\left(t_2 - \frac{1}{n} t_1^2\right)$

$\frac{\partial \tilde{t}_2}{\partial t_1} = \frac{1}{n-1}\frac{1}{n}(-2t_1)$ $\frac{\partial \tilde{t}_2}{\partial t_2} = \frac{1}{n-1}$

$$|J| = \begin{vmatrix} \frac{1}{n} & 0 \\ \frac{1}{n-1}\frac{1}{n}(-2t_1) & \frac{1}{n-1} \end{vmatrix} = \frac{1}{n}\frac{1}{n-1} \neq 0$$

Notar En los ejemplos vistos logramos encontrar un estadístico suficiente de la misma dimensión que el espacio paramétrico

i) $X_1, \ldots, X_n \sim B(\theta) \longrightarrow \hat{p}_n$

ii) $\sim N(\mu, \sigma^2) \longrightarrow (\bar{X}, S^2)$

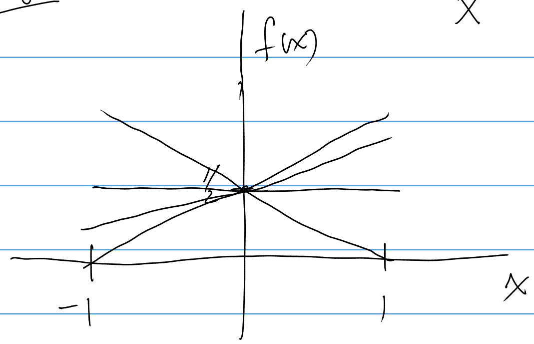
(Estos son estadísticos suficientes minimales)

¿Es esto siempre posible?. **NO**

Si lo es para familias exponenciales de rango completo (próximas atracciones).

Ej $X_1, \ldots, X_n \overset{iid}{\sim} f_X(x;\theta) = \begin{cases} \frac{1}{2}(1+\theta x) & -1 \leq x \leq 1 \\ 0 & c.c. \end{cases}$

$\theta \in [-1, 1]$



$$f_{X_1 \ldots X_n}(x_1 \ldots x_n; \theta) = \prod_{i=1}^{n} \frac{1}{2}(1+\theta x_i) = \frac{1}{2^n} \prod_{i=1}^{n} (1+\theta x_i)$$

$$= \frac{1}{2^n} \prod_{i=1}^{n} (1+\theta x_{(i)})$$

$$T(X_1 \ldots X_n) = (X_{(1)}, \ldots, X_{(n)})$$

¿Se podrá reducir más? No.

Def Un estadístico $T(X)$ es suficiente minimal si para cualquier estadístico suficiente $U(X)$ se tiene que

$$T(X) = H(U(X))$$

Ej $X_1 \dots X_n \overset{iid}{\sim} \mathcal{P}(\lambda)$

$T_1 = (X_1, \dots, X_n)$

$T_2 = (X_{(1)}, \dots, X_{(n)})$

$T_4 = n \sum X_i \overset{1-1}{\longleftrightarrow} \sum X_i$

$T_3 = \left(\sum_{i=1}^{m} X_i ; \sum_{i=m+1}^{n} X_i\right)$

$\underline{T = \sum_{i=1}^{n} X_i}$ suficiente minimal

$$T = T_{11} + \dots + T_{1n}$$
$$= T_{21} + \dots + T_{2n}$$
$$= T_{31} + T_{32}$$

<u>Teorema</u> Sea $X \sim F(x;\theta) \in \mathcal{F} = \{F(\cdot, \theta_j) : j = 1 \dots k\}$
$\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad ? \; f(\cdot, \theta_j)$

El estadístico
$$T(x) = \left(\frac{f(x,\theta_2)}{f(x,\theta_1)} ; \dots ; \frac{f(x,\theta_k)}{f(x,\theta_1)} \right)$$
es minimal suficiente.

<u>Dem</u> $\forall i,j$ $\quad \dfrac{f(x,\theta_i)}{f(x,\theta_j)}$ es función de $T(x)$ por construcción

y por corolario es suficiente
(suficiencia está, falta minimal).

Sea $U(X)$ un estadístico suficiente
por corolario
$$U(X) = \frac{f(x,\theta_i)}{f(x,\theta_1)} \quad i = 2 \dots k$$

entonces T es función de U
$\therefore$ T es suficiente minimal por definición.

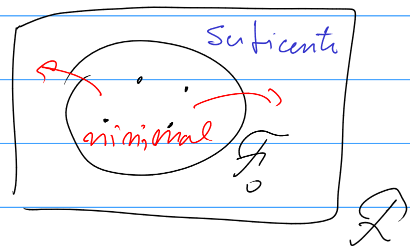
<u>Teorema</u> Sea $X \sim F(x;\theta) \in \mathcal{F} = \{F(\cdot,\theta) : \theta \in \Theta\}$
con soporte común. Y sea $\mathcal{F}_0 \subset \mathcal{F}$ con $\# \mathcal{F}_0 < \infty$.

Además sea $T(X)$ un estadístico que es:

i) suficiente para $\mathcal{F}$

ii) suficiente minimal para $\mathcal{F}_0$

Entonces $T(X)$ es suficiente minimal para $\mathcal{F}$.



Dem Sea $U(X)$ un estadístico suficiente en $\mathcal{F}$

$\therefore$ $U(X)$ es ✓ en $\mathcal{F}_0$

$\therefore$ $T(X) = H(U(X))$

Uso $X_1 ... X_n \overset{iid}{\sim} P(\lambda)$ $\quad \mathcal{F} = \{P(\lambda) : \lambda > 0\}$

$T(X) = \sum X_i$ es suf. en $\mathcal{F}$

Construyo $\mathcal{F}_0 = \{P(\lambda) : \lambda = 1, 2\}$

$$f(x,\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$

$$T(X) = \frac{f(x;2)}{f(x;1)} = \frac{e^{-n2}\, 2^{\sum x_i} / \cancel{\prod x_i!}}{e^{-n}\, 1^{\sum x_i} / \cancel{\prod x_i!}} =$$

$$= e^{-n}\, 2^{\sum X_i}$$

$\tilde{T}(X) = \sum X_i$ suf minimal para $\mathcal{F}_0$ (Teo 1)

✓ ✓ ✓ $\mathcal{F}$. (Teo 2)

Ej reconsiderado $\quad f_X(x;\theta) = \frac{1}{2}(1+\theta x) \qquad -1 \le x \le 1$

$\theta \in [-1;1]$

$$f_{X_1 \dots X_n}(x_1 \dots x_n;\theta) = \frac{1}{2^n}\prod_{i=1}^{n}(1+\theta x_{(i)})$$

$T(X) = (X_{(1)}; \dots ; X_{(n)})$ suficiente en

$\mathcal{F} = \{ F_X(\cdot,\theta) : \theta \in [-1,1] \}$

$\mathcal{F}_0 = \{ F_X(\cdot,\theta) : \theta \in \{0, \underbrace{\theta_1}_{0.5}, \dots, \underbrace{\theta_n}_{0.3} \} \}$

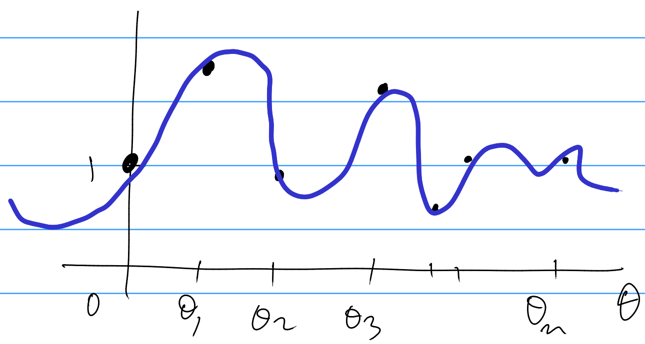
$$T(X) = \left(\frac{f(x,\theta_1)}{f(x,0)}; \dots , \frac{f(x,\theta_n)}{f(x,0)} \right) \qquad f(x,0) = 1/2^n$$

$$= \left(\prod_{i=1}^{n}(1+\theta_1 x_i); \dots , \prod_{i=1}^{n}(1+\theta_n x_i) \right)$$

$$\tilde{T}(X) = \left(\underbrace{\prod_{i=1}^{n}(1+\theta_1 X_{(i)})}; \dots ; \underbrace{\prod_{i=1}^{n}(1+\theta_n X_{(i)})} \right)$$

notar $\displaystyle\prod_{i=1}^{n}(1+\theta X_{(i)})$ = Polinomio en θ = $p_n(\theta)$

con $p_n(0) = 1$



Correspondencia 1-1 con raíces

$$\tilde{T}(X) \overset{1-1}{\longleftrightarrow} (X_{(1)}, \dots , X_{(n)}) = T^*(X)$$

$(X_{(1)}, \dots, X_{(n)})$ es suficiente minimal para $\mathcal{F}_0$

Entonces por teorema es suficiente minimal para θ.

Observar no logré disminuir dimensión!

$$\prod (1+\theta X_{(i)}) = 0$$

$1+\theta X_{(1)} = 0$ o $1+\theta X_{(2)} = 0$ ... $1+\theta X_{(n)} = 0$

$1 = -\theta X_{(1)}$

$\theta = -\dfrac{1}{X_{(1)}}$ - - - - - - $\theta = -\dfrac{1}{X_{(n)}}$