Suficiencia o exhaustividad

Carlos Carleos, Norberto Corral

11 de octubre de 2022

Índice

1. suficiencia o exhaustividad de un estadígrafo

- X población, $X \hookrightarrow F_\theta$, $\theta \in \Theta$
- $\vec{X} = (X_1, \ldots, X_n)$ muestra
- $T = T(\vec{X})$ estadígrafo
- en la muestra $\vec{X}$ puede haber información irrelevante para estimar θ
- por ejemplo, si $X \hookrightarrow B(p)$, para estimar p es irrelevante el orden de éxitos y fracasos; sólo son relevantes sus frecuencias
- se pretende que un estadígrafo conserve toda la información relevante de $\vec{X}$ para estimar θ

1.1. ejemplo

- considerar el lanzamiento de una moneda tres veces, en el que se obtiene la muestra $(\text{cara}, \text{cruz}, \text{cruz})$
- se pretende estimar $p = \Pr(\text{cara})$
- ¿es relevante el orden una vez que se sabe el recuento $(\text{cara}\colon 1, \text{cruz}\colon 2)$?

- no, porque entonces la probabilidad de la muestra ya no depende de p:

$$\Pr\big[\text{cara}, \text{cruz}, \text{cruz} \mid (\text{cara}\colon 1, \text{cruz}\colon 2)\big] = \frac{\Pr(\text{cara}, \text{cruz}, \text{cruz})}{\Pr(\text{cara}\colon 1, \text{cruz}\colon 2)} = \frac{p(1-p)^2}{3p(1-p)^2} = \frac{1}{3}$$

- el recuento es suficiente: conserva por completo la información relevante

1.2. definición

- Un estadígrafo T es **exhaustivo** o **suficiente** para estimar θ si la distribución de $\vec{X}$ condicionada a $T = t$ no depende de θ.

1.3. teorema de factorización

1.3.1. enunciado

- T es exhaustivo si y sólo si existen g y h tales que

$$f(\vec{x}, \theta) = g(t, \theta) \cdot h(\vec{x})$$

con $t = T(\vec{x})$.

1.3.2. demostración (caso discreto)

(⇒) Por un lado, $f(\vec{x}, \theta) = f(\vec{x}, \theta \mid T = t) \cdot f_T(t, \theta)$. Por otro, si T es suficiente entonces $f(\vec{x}, \theta \mid T = t)$ no depende de θ y basta tomar $g(t, \theta) = f_T(t, \theta) = \sum_{T(\vec{x})=t} f(\vec{x}, \theta)$ y $h(\vec{x}) = f\big(\vec{x}, \theta \mid T = T(\vec{x})\big)$

(⇐) Sea $f(\vec{x}, \theta) = g(t, \theta) \cdot h(\vec{x})$ con $t = T(\vec{x})$. Sea $A_t = \{\vec{x} \mid T(\vec{x}) = t\}$. Entonces

$$f(\vec{x}, \theta \mid T = t) = \begin{cases} 0 & T(\vec{x}) \neq t \\ \frac{f(\vec{x},\theta)}{f(t,\theta)} = \frac{f(\vec{x},\theta)}{\sum_{\vec{y}\in A_t} f(\vec{y},\theta)} = \frac{g(T(\vec{x}),\theta)\cdot h(\vec{x})}{\sum_{\vec{y}\in A_t} g(T(\vec{y}),\theta)\cdot h(\vec{y})} = \frac{g(t,\theta)\cdot h(\vec{x})}{\sum_{\vec{y}\in A_t} g(t,\theta)\cdot h(\vec{y})} = \frac{h(\vec{x})}{\sum_{\vec{y}\in A_t} h(\vec{y})} & T(\vec{x}) = t \end{cases}$$

que no depende de θ, luego T es suficiente para θ.

1.3.3. EXTRA caso general

- véanse
  - ejemplos primero y sexto de condicionamiento como desintegración
  - capítulo 2.6 *Characterization of sufficiency* de *Testing statistical hypotheses, second edition* de E.L. Lehmann
  - suplemento IV, pág. 550 de *Estadística matemática* de A.A. Borovkov, KP 519B 3 en la biblioteca
- si el soporte de X depende de θ,
  - Sufficient statistics and intrinsic accuracy, sección 4
  - sea X con soporte $(0, \theta)$; si T no involucra θ, es obvio que al fijar $T = t$ la distribución del máximo $X_{(n)}$ tiene que depender de θ, a menos que T sea $X_{(n)}$ o una función suya; así, si existe un estadígrafo suficiente tiene que ser función sólo de $X_{(n)}$

1.4. demostración de no exhaustividad/suficiencia

- Sean dos muestras $\vec{x} = (x_1, \ldots, x_n)$ y $\vec{y} = (y_1, \ldots, y_n)$ tales que $T(\vec{x}) = T(\vec{y})$.
- Si $\frac{f(\vec{x},\theta)}{f(\vec{y},\theta)}$ depende de θ, T no es exhaustivo.
- Si lo fuera,

$$\frac{f(\vec{x}, \theta)}{f(\vec{y}, \theta)} = \frac{g(T(\vec{x}), \theta)h(\vec{x})}{g(T(\vec{y}, \theta)h(\vec{y})} = \frac{g(t, \theta)h(\vec{x})}{g(t, \theta)h(\vec{y})} = \frac{h(\vec{x})}{h(\vec{y})}$$

no dependería de θ.

1.5. ejemplos

1.5.1. suficiencia normal

- $\vec{X}$, la propia muestra, es trivialmente un estadígrafo exhaustivo
- $\vec{X}_{(\cdot)} = (X_{(1)}, \ldots, X_{(n)})$, la muestra ordenada, es exhaustivo en muestreo aleatorio simple
- $X \hookrightarrow U(0, \theta] \implies X_{(n)}$ es suficiente para θ
  - (sean [] los corchetes de Iverson)
  - $f(\vec{x}, \theta) = \prod_{i=1}^{n} \frac{1}{\theta}[x_i < \theta] = \frac{1}{\theta^n}[x_{(n)} < \theta] = g(t, \theta)h(\vec{x})$ con $t = x_{(n)}$ y $h(\cdot) = 1$
  - hay dificultades al condicionar a sucesos con probabilidad nula
- $X \hookrightarrow N(\mu, \sigma)$, σ conocida $\implies T = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ es exhaustivo para μ
- $X \hookrightarrow N(\mu, \sigma)$, μ conocida $\implies T = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ es exhaustivo para σ
- $X \hookrightarrow N(\mu, \sigma)$, σ conocida $\implies \bar{X}^2$ no es suficiente para μ
- $X \hookrightarrow N(\mu, \sigma)$, μ y σ desconocidas $\implies \bar{X}, S^2$ es exhaustivo para (μ, σ)

1.5.2. EXTRA suficiencia parcial

- $X \hookrightarrow N(\mu, \sigma)$, μ y σ desconocidas $\implies S^2$ no es suficiente para σ
- $X \hookrightarrow N(\mu, \sigma)$, μ y σ desconocidas $\implies \bar{X}$ ¿es suficiente para μ?
  - según la definición de suficiencia:

```
f_gaus(x,mu,sigma) := 1/sigma/sqrt(2*%pi)*%e^(-(x-mu)^2/2/sigma^2) $
f_gaus_muestra(X,mu,sigma) := apply("*", f_gaus(X,mu,sigma)) $
f_gaus_media(xmedia,mu,sigma,n) := f_gaus(xmedia,mu,sigma/sqrt(n)) $
media(X) := apply ("+", X) / length(X) $
f_gaus_cond(X,mu,sigma,t) :=
  f_gaus_muestra(X,mu,sigma) / f_gaus_media(t,mu,sigma,length(X)) $
n : 10 $
X : makelist (x[i], i, 1, n) $
freeof (mu, ratsimp(f_gaus_cond (X, mu, sigma, media(X)))) ; /* true */
```

  - pero $\bar{X} \hookrightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ luego la muestra incluye más información sobre μ a través de la dispersión S^2
  - ejercicio 36 del capítulo 3, pág. 122, de *Testing statistical hypotheses, second edition* de E.L. Lehmann
    - T es **parcialmente suficiente** para μ si
      - $f(\vec{x}, \mu, \sigma \mid T = t)$ no depende de μ
      - la distribución de T no depende de σ
    - $\bar{X}$ no es parcialmente suficiente para μ
  - ejemplo 2.1 de `https://www.jstor.org/stable/1403095`: el estadígrafo (parcialmente) suficiente para μ es $(\bar{X}, S^2)$

1.6. estadígrafo mínimamente exhaustivo / estadístico suficiente minimal

- un estadígrafo T induce una partición ϕ del espacio $X(\Omega)^n$, de forma que $\vec{x} \sim \vec{y} \iff T(\vec{x}) = T(\vec{y})$
- una partición es suficiente si induce un estadígrafo suficiente
- una partición es **suficiente minimal** si es suficiente y cualquier otra partición suficiente es un refinamiento suyo
- un estadígrafo es **suficiente minimal** si induce una partición suficiente minimal

- la partición dada por la relación

$$\vec{x} \sim \vec{y} \iff \frac{f(\vec{x},\theta)}{f(\vec{y},\theta)} \text{ no depende de } \theta$$

  es suficiente minimal;

  demostración para el caso discreto:

  - Sea T cualquier estadígrafo asociado a dicha partición. Sea $\vec{x}'$ tal que $T(\vec{x}') = t$. Entonces

$$f(\vec{x}',\theta \mid T = t) = \frac{f(\vec{x}',\theta)}{f_T(t,\theta)} = \frac{f(\vec{x}',\theta)}{\sum_{T\vec{x}=t} f(\vec{x},\theta)} = \frac{1}{\sum_{T\vec{x}=t} \frac{f(\vec{x},\theta)}{f(\vec{x}',\theta)}}$$

    no depende de θ, luego T es suficiente.
  - Sea T' otro estadígrafo suficiente y $T'(\vec{x}) = T'(\vec{x}') = t'$, es decir, $\vec{x}$ y $\vec{x}'$ pertenecen al mismo elemento de la partición asociada a T'. Entonces

$$f(\vec{x},\theta \mid T' = t') = \frac{f(\vec{x},\theta)}{f_{T'}(t',\theta)}$$

    y

$$f(\vec{x}',\theta \mid T' = t') = \frac{f(\vec{x}',\theta)}{f_{T'}(t',\theta)}$$

    son independientes de θ y su cociente

$$\frac{f(\vec{x},\theta \mid T' = t')}{f(\vec{x}',\theta \mid T' = t')} = \frac{f(\vec{x},\theta)}{f(\vec{x}',\theta)}$$

    también, luego $T(\vec{x}) = T(\vec{x}')$ y $\vec{x}$ y $\vec{x}'$ pertenecen al mismo elemento de la partición asociada a T, luego la de T' es un refinamiento de la de T. Por tanto, T es mínimamente suficiente.

- ejemplo

  - $X \hookrightarrow \text{Exp}(\lambda)$, $T = \sum_{i=1}^n X_i$
  - $f(\vec{x},\lambda) = \lambda^n e^{-\lambda \sum x_i} = \lambda^n e^{-\lambda t} = g(t,\lambda)h(\vec{x})$ con $h(\vec{x}) = 1$, luego T es suficiente
  - T es minimal suficiente porque

$$\frac{f(\vec{x},\lambda)}{f(\vec{y},\lambda)} = \frac{\lambda^n e^{-\lambda \sum x_i}}{\lambda^n e^{-\lambda \sum y_i}} = e^{-\lambda(\sum x_i - \sum y_i)} = e^{-\lambda[T(\vec{x}) - T(\vec{y}_i)]}$$

    no depende de λ sii $T(\vec{x}) = T(\vec{y}_i)$

2. la familia exponencial k-paramétrica

- incluye a la mayoría de distribuciones habituales

2.1. definición

- una familia de distribuciones $\{F_\theta \mid \theta \in \Theta \subset \mathbb{R}^k\}$ pertenece a la familia exponencial k-paramétrica si

  - el soporte $\{\vec{x} \mid f(\vec{x},\theta)\}$ no depende de θ
  - existen D, $Q_1,\ldots,Q_k$, S, $T_1,\ldots,T_k$ tales que

$$f(\vec{x},\vec{\theta}) = \exp\left[S(\vec{x}) + D(\vec{\theta}) + \sum_{j=1}^k Q_j(\vec{\theta})T_j(\vec{x})\right]$$

$$f(\vec{x},\vec{\theta}) = c(\vec{\theta})h(\vec{x})e^{\sum Q_j(\theta)T_j(\vec{x})}$$

2.2. parametrización natural

- parametrizando $\eta_j = Q_j(\theta)$ se tiene la parametrización natural

$$f(\vec{x},\vec{\eta}) = c^*(\vec{\eta})h(\vec{x})e^{\sum \eta_j T_j(\vec{x})}$$

- el espacio paramétrico natural es

$$\mathrm{H} = \left\{\vec{\eta}\ :\ \int_{\mathbb{R}^n} h(\vec{x})e^{\sum \eta_j T_j(\vec{x})}d\vec{x} < \infty\right\}$$

2.3. ejemplos

2.3.1. $B(n,p)$ con n conocido, $0 < p < 1$

- $f(x,p) = \binom{n}{x}p^x(1-p)^{n-x} = \binom{n}{x}(1-p)^n\left(\frac{p}{1-p}\right)^x = \binom{n}{x}(1-p)^n e^{x\ln\frac{p}{1-p}}$
- parámetro natural $\eta = \ln\frac{p}{1-p}$
- $f(x,\eta) = \binom{n}{x}\left(\frac{1}{1+e^{-\eta}}\right)^n e^{x\eta}$

2.3.2. $\gamma(p,a)$, $p,a>0$

- $f(x,p,a) = \frac{a^p}{\Gamma(p)}e^{-ax}x^{p-1} = \frac{a^p}{\Gamma(p)}\frac{1}{x}e^{-ax+p\ln x}$
- p y a son parámetros naturales, con $T_1 = -x$ y $T_2 = \ln x$

2.3.3. Poisson(λ), $\lambda > 0$

- $f(x,\lambda) = e^{-\lambda}\frac{\lambda^x}{x!} = \frac{1}{x!}e^{-\lambda}e^{x\ln\lambda}$
- $\eta = \ln\lambda$ es parámetro natural

2.3.4. $N(\mu,\sigma)$

- $f(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2+\mu^2-2\mu x}{2\sigma^2}} = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{\mu^2}{2\sigma^2}}e^{-\frac{x^2-2\mu x}{2\sigma^2}} = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{\mu^2}{2\sigma^2}}e^{-\frac{x^2}{2}\frac{1}{\sigma^2}+2x\frac{\mu}{\sigma^2}}$
- parámetros naturales $\eta_1 = \frac{1}{\sigma^2}$, $\eta_2 = \frac{\mu}{\sigma^2}$
- $T_1 = -\frac{x^2}{2}$, $T_2 = 2x$

2.4. teorema

- sea $\vec{\vec{X}} = (\vec{X}_1,\ldots,\vec{X}_n)$ una muestra obtenida del vector aleatorio $\vec{X}$, perteneciente a la familia exponencial k-paramétrica con densidad

  $$f(\vec{x},\vec{\eta}) = c(\vec{\eta})h(\vec{x})e^{\sum_{j=1}^k \eta_j T_j(\vec{x})}$$

  para los $\vec{x}$ con $f(\vec{x},\vec{\eta}) > 0$;
- supóngase que el espacio paramétrico natural H contiene un abierto de $\mathbb{R}^k$
- entonces $\vec{W} = (W_1,\ldots,W_k)$ con $W_i = \sum_{j=1}^n T_i(\vec{x}_j)$ es mínimamente exhaustivo
- demostración
  - $\vec{W}$ es exhaustivo por el teorema de factorización: $f(\vec{x}_1,\ldots,\vec{x}_n,\vec{\eta}) = c(\vec{\eta})^n\prod_{i=1}^n h(\vec{x}_i)e^{\sum_{j=1}^k\eta_j\sum_{i=1}^n T_j(\vec{x}_i)} = c(\vec{\eta})^n\prod_{i=1}^n h$ $= g(\vec{w},\vec{\eta})H(\vec{\vec{x}})$ con $H(\vec{\vec{x}}) = \prod_{i=1}^n h(\vec{x}_i)$, $w_j = \sum_{i=1}^n T_j(\vec{x}_i)$, $g(\vec{w},\vec{\eta}) = c(\vec{\eta})^n e^{\sum_{j=1}^k \eta_j w_j}$
  - $\vec{W}$ es minimal suficiente pues $\vec{W}\vec{\vec{x}} = \vec{W}\vec{\vec{y}} \iff \frac{f(\vec{\vec{x}},\vec{\eta})}{f(\vec{\vec{y}},\vec{\eta})}$ es independiente de $\vec{\eta}$:

($\Leftarrow$) sean $\vec{\vec{x}}$ y $\vec{\vec{y}}$ tales que $\frac{f(\vec{\vec{x}},\vec{\eta})}{f(\vec{\vec{y}},\vec{\eta})}$ es independiente de $\vec{\eta}$; entonces

$$\frac{f(\vec{\vec{x}},\vec{\eta})}{f(\vec{\vec{y}},\vec{\eta})} = \frac{H(\vec{\vec{x}})}{H(\vec{\vec{y}})} e^{\sum \eta_j [W_j(\vec{\vec{x}}) - W_j(\vec{\vec{y}})]}$$

que es independiente de $\vec{\eta}$ sii $\vec{W}\vec{\vec{x}} = \vec{W}\vec{\vec{y}}$; si no fuera así, supóngase sin pérdida de generalidad que $W_1\vec{\vec{x}} \neq W_1\vec{\vec{y}}$ y

$$\forall\, \vec{\eta} \in \mathrm{H}, \quad \sum_{j=1}^{k} \eta_j [W_j(\vec{\vec{x}}) - W_j(\vec{\vec{y}})] = 0$$

sean ahora $\vec{\eta} = (\eta_1, \eta_2, \ldots, \eta_k)$ y $\vec{\eta}^* = (\eta_1 + \epsilon, \eta_2, \ldots, \eta_k)$ dentro de un abierto de H; entonces $0 = \sum_{j=1}^{k} \eta_j^* [W_j(\vec{\vec{x}}) -$

$= \epsilon[W_1(\vec{\vec{x}}) - W_1(\vec{\vec{y}})] + \underbrace{\sum_{j=1}^{k} \eta_j [W_j(\vec{\vec{x}}) - W_j(\vec{\vec{y}})]}_{\vec{\eta} \in \mathrm{H} \implies = 0} = \epsilon[W_1(\vec{\vec{x}}) - W_1(\vec{\vec{y}})] \neq 0$ y llegaríamos a una contradicción;

($\Rightarrow$) sean $\vec{\vec{x}}$ y $\vec{\vec{y}}$ tales que $\vec{W}\vec{\vec{x}} = \vec{W}\vec{\vec{y}}$; entonces $\frac{f(\vec{\vec{x}},\vec{\eta})}{f(\vec{\vec{y}},\vec{\eta})} = \frac{H(\vec{\vec{x}})}{H(\vec{\vec{y}})} e^{\sum \eta_j [W_j(\vec{\vec{x}}) - W_j(\vec{\vec{y}})]} = \frac{H(\vec{\vec{x}})}{H(\vec{\vec{y}})}$ que es independiente de $\vec{\eta}$