Lista 3

1) Seja $X_1, \dots, X_n$ uma amostra aleatória para uma população com pdf $f(x-\theta)$. Mostre que a estatística de ordem $T(X_1,\dots,X_n) = (X_{(1)},\dots,X_{(n)})$, é uma estatística suficiente para θ e não pode ser reduzida.

Primeiro, temos que,

$$L(\theta|X_1,\dots,X_n) = f(x_1-\theta)\dots f(x_n-\theta)$$

$$(*) = f(X_{(1)}-\theta)\dots f(X_{(n)}-\theta)$$

(*) pois o produto é invariável por ordenação. Daí temos que T é suficiente.

Como cada entrada de f possui θ com operação de "$-$" não é possível agrupar X_i's em uma estatística de dimensão menor que n.

/ /

(2) Para cada uma das distribuições seja $X_1, \dots, X_n$ uma amostra aleatória. Encontre uma estatística suficiente mínima para θ.

(a) $f(x|\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}}, -\infty < x < \infty, -\infty < \theta < \infty$

(b) $f(x|\theta) = e^{-(x-\theta)}, \quad \theta < x < \infty, -\infty < \theta < \infty$

(c) $f(x|\theta) = \frac{e^{-(x-\theta)}}{(1+e^{-(x-\theta)})^2}, -\infty < x < \infty, -\infty < \theta < \infty$

(d) $f(x|\theta) = \frac{1}{\pi(1+(x-\theta)^2)}, -\infty < x, \theta < \infty$

(e) $f(x|\theta) = \frac{1}{2} e^{-|x-\theta|}, -\infty < x, \theta < \infty.$

(a) $$\cancel{\frac{f(x|\theta)}{f(y|\theta)}} = \frac{L(\theta|x_1, \dots, x_n)}{L(\theta|y_1, \dots, y_n)} = \frac{e^{-\Sigma(x_i-\theta)^2/2}}{e^{-\Sigma(y_i-\theta)^2/2}}$$

$$= \exp\left\{-\frac{1}{2}\left[\left(\Sigma x_i^2 - \Sigma y_i^2\right) + 2\theta n(\bar{y}-\bar{x})\right]\right\}$$

é constante em θ se e somente se $\bar{x} = \bar{x}$, então $\bar{X}$ é uma estatística suficiente mínima para θ.

(b) $$\frac{L(\theta|x_1, \dots, x_n)}{L(\theta|y_1, \dots, y_n)} = \frac{\prod\left(e^{-(x_i-\theta)} \mathbb{I}(\theta < x_i)\right)}{\prod\left(e^{-(y_i-\theta)} \mathbb{I}(\theta < y_i)\right)}$$

$$= \frac{\cancel{e^{n\theta}} e^{-\Sigma x_i} \mathbb{I}(\min x_i > \theta)}{\cancel{e^{n\theta}} e^{-\Sigma y_i} \mathbb{I}(\min y_i > \theta)}$$

Para fazer θ constante, temos $\min x_i = \min y_n$, ou seja $\min x_i$

/ /

é uma estatística suficiente mínima para θ.

(c) $$\frac{L(\theta|X_1,\ldots,X_n)}{L(\theta|Y_1,\ldots,Y_n)} = \frac{e^{-\sum X_i+\theta}}{\prod(1+e^{-(X_i-\theta)})^2} \cdot \frac{\prod(1+e^{-(Y_i-\theta)})^2}{e^{-\sum Y_i+\theta}}$$

$$= e^{-\sum(Y_i-X_i)}\left(\frac{\prod(1+e^{-(Y_i-\theta)})^2}{\prod(1+e^{-(X_i-\theta)})^2}\right)$$

Pelo exercício anterior, ~~a~~ uma estatística suficiente mínima é uma ~~o~~ permutação dos dados $T(X_1,\ldots,X_n)=(X_{(1)},\ldots,X_{(n)})$

(d) A estatística de ordem é uma estatística suficiente mínima, pelo exercício anterior também

(e) Fixe as amostras $\{X_i\}$ e $\{Y_i\}$. Defina $A(\theta)=\{i: X_i\leq\theta\}$, $B(\theta)=\{i: Y_i\leq\theta\}$, $a(\theta)=$ nº de elementos em $A(\theta)$ e $b(\theta)=$ nº de elementos em $B(\theta)$. Então a função ~~f~~ $L(\theta|X_1,\ldots,X_n)/L(\theta|Y_1,\ldots,Y_n)$ depende em θ apenas por

$$\sum|X_i-\theta|-\sum|Y_i-\theta| = \sum_{i\in A(\theta)}(\theta-X_i)+\sum_{i\in A(\theta)^c}(X_i-\theta)-\sum_{i\in B(\theta)}(\theta-Y_i)-\sum_{i\in B(\theta)^c}(Y_i-\theta)$$

$$=(a(\theta)-[n-a(\theta)]-b(\theta)+[n-b(\theta)])\theta$$

$$+\left(-\sum_{i\in A(\theta)}X_i+\sum_{i\in A(\theta)^c}X_i+\sum_{i\in B(\theta)}Y_i-\sum_{i\in B(\theta)^c}Y_i\right)$$

$$=2(a(\theta)-b(\theta))\theta+\left(\sum_{i\in A(\theta)}X_i+\sum_{i\in A(\theta)^c}X_i+\sum_{i\in B(\theta)}Y_i-\sum_{i\in B(\theta)^c}Y_i\right)$$

Considerando um intervalo de θs que não contem nenhum X_i's e Y_i's. O primeiro termo é constante se e só se $a(\theta)=b(\theta)$, ou seja quando X e Y tem a mesma ordem. Então é uma estatística suficiente.

/ /

③ ⓐ Seja $X_1, \ldots, X_n$ iid $N(\mu, \sigma^2)$, onde σ^2 é conhecido.

(i) Mostre que $\bar{X}$ é completa suficiente para μ, e S^2 é auxiliar. Portanto pelo Teorema de Basu, $\bar{X}$ e S^2 são independentes.

(ii) Mostre que essa independencia se estende até quando σ^2 é desconhecido, pois o conhecimento de σ^2 não tem relação com as distribuições.

ⓑ Uma Monte Carlo swindle é uma tecnica para melhorar a variancia do estimador. Suponha que $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ iid e que queremos computar a variancia da media, M.

(i) Aplique o teorema de Basu para mostrar que $Var(M) = Var(M - \bar{X}) + Var(\bar{X})$; assim ~~nos~~ só temos que simular o ~~pequeno~~ pedaço $V(M - \bar{X})$ de $Var\, M$ (desde que $Var(\bar{X}) = \sigma^2/n$)

(ii) Mostre que o estimador é mais preciso mostrando que a variancia de M é aproximadamente $2[Var\, M]^2/(N-1)$ e que $M - \bar{X}$ é aproximadamente $2[Var(M - \bar{X})]^2/(N-1)$, onde N é o número de amostras de Monte Carlo.

ⓒ (i) Se X/Y e Y são independentes, mostre que

$$E\left(\frac{X}{Y}\right)^k = E(X^k)/E(Y^k)$$

(ii) Use este resultado e o Teorema de Basu para mostrar que se $X_1, \ldots, X_n \overset{iid}{\sim} Gamma(\alpha, \beta)$, com α conhecido, então para $T = \sum X_i$

$$E(X_{(i)} | T) = E\left(\frac{X_{(i)}}{T} T \,\middle|\, T\right) = T \frac{E X_{(i)}}{E T}$$

/ /

(a) (i) Pelo teorema 6.2.25 [CB], ΣX_i é uma estatística completa e suficiente. $\bar{X}$ é uma função de ΣX_i, então $\bar{X}$ é também uma estatística suficiente completa. Pelo teorema 5.3.1 [CB] sabemos que $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$ = Gama Gamma$((n-1)/2, 2)$. $S^2 = [\sigma^2/(n-1)][(n-1)S^2/\sigma^2]$, uma transformação simples, então, $S^2 \sim$ Gamma$((n-1)/2, 2\sigma^2/(n-1))$, que não depende de μ, logo S^2 é auxiliar e pelo teorema de Basu, $\bar{X}$, S^2 são independentes.

(ii) A independencia de $\bar{X}$ e S^2 é determinada pela conjunta de $(\bar{X}, S^2)$ para cada valor de (μ, σ^2). Pela parte (i), para ~~dada~~ cada valor de (μ, σ^2), $\bar{X}$ e S^2 são independentes

(b) (i) μ é um parametro de locação. Vamos primeiro que se $X_1, \dots, X_n$ é uma amostra aleatoria de uma familia de locação, então $M - \bar{X}$ é uma estatística auxiliar.

Como $X_1, \dots, X_n$ é uma familia de ~~ob~~ locação, podemos escrever $X_i = Z_i + \mu$, onde $Z_1, \dots, Z_n$ é uma amostra aleatoria para a pdf, $f(z)$. Seja ~~$M(Z)$~~ $M(X)$ denotando a média calculada para $X_1, \dots, X_n$. Então $M(X) = M(Z) + \mu$ e $\bar{X} = \bar{Z} + \mu$. Portanto, $M(X) - \bar{X} = (M(Z) + \mu) - (\bar{Z} + \mu) = M(Z) - \bar{Z}$. Porque $M(X) - \bar{X}$ é uma função de $Z_1, \dots, Z_n$, a distribuição de $M(X) - \bar{X}$ não depende de μ, ou seja, $M(X) - \bar{X}$ é uma estatística auxiliar.

Pelo item (a) $\bar{X}$ é uma estatística suficiente completa. Pelo Teorema de Basu $\bar{X}$ e $M - \bar{X}$ são independentes. Como são independentes, pelo teorema 4.5.6 [CB], $\text{Var}\, M = \text{Var}(M - \bar{X} + \bar{X}) = \text{Var}(M - \bar{X}) + \text{Var}\, \bar{X}$.

/ /

(ii) Se S^2 é uma amostra de variancia calculada para uma normal de tamanho N, $(N-1)S^2/\sigma^2 \sim \chi^2_{N-1}$. Portanto, $(N-1)^2 \text{Var}\, S^2/(\sigma^2)^2 = 2(N-1)$ e $\text{Var}\, S^2 = 2(\sigma^2)^2/(N-1)$. Ambos, M e $M-\bar{X}$ são assintoticamente normais, então, $M_1, \dots, M_N$ e $M_1-\bar{X}_1, \dots, M_N-\bar{X}_N$ são cada um aproximadamente amostras normais se N é grande o bastante. Logo, usando as expressões acima temos 2 expressões onde no caso simples σ^2 se refere a Var M, e no caso de swindle σ^2 se refere a $\text{Var}(M-\bar{X})$

(c) (i) Por

$$EX^k = E\left(\frac{X}{Y}Y\right)^k = E\left[\left(\frac{X}{Y}\right)^k (Y)^k\right] = E\left(\frac{X}{Y}\right)^k \cdot E(Y^k)$$

agora basta dividir os 2 lados por $E(Y^k)$.

(ii) Fixando α, $T = \sum X_i$ é uma estatística suficiente completa pelo teorema 6.2.25 [CB] Como β é um parametro, se $Z_1, \dots, Z_n$ é uma amostra aleatoria para uma gamma$(\alpha, 1)$, então $X_{(i)}/T$ tem a mesma distribuição que $(\beta Z_{(i)})/(\beta \sum Z_i) = Z_{(i)}/(\sum Z_i)$, e esta distribuição não depende de β. Portanto, $X_{(i)}/T$ é auxiliar, e pelo teorema de Basu, é independente de T. Temos

$$E(X_{(i)}|T) = E\left(\frac{X_{(i)}}{T}T \,\middle|\, T\right) = TE\left(\frac{X_{(i)}}{T}\,\middle|\, T\right)$$

$$= T\, E\left(\frac{X_{(i)}}{T}\right) = \frac{TE[X_{(i)}]}{ET}.$$

④ A distribuição beta com parametros $\alpha, \beta > 0$ tem densidade

$$f_{\alpha,\beta}(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, & x \in (0,1) \\ 0, & c.c \end{cases}$$

Suponha $X_1, \ldots, X_n$ iid beta.

ⓐ Determine uma estatística suficiente mínima se α e β variam livremente

ⓑ Determine uma estatística mínima suficiente se $\alpha = 2\beta$

ⓒ Determine " " " " se $\alpha = \beta^2$.

ⓐ $$\frac{L(\alpha,\beta|\{X_i\})}{L(\alpha,\beta|\{Y_i\})} = \frac{\prod X_i^{\alpha-1}(1-X_i)^{\beta-1}}{\prod Y_i^{\alpha-1}(1-Y_i)^{\beta-1}}$$

$$\Rightarrow \frac{\ln L(\alpha,\beta|\{X_i\})}{\ln L(\alpha,\beta|\{Y_i\})} = (\alpha-1)(\sum \ln X_i - \sum \ln Y_i) + (\beta-1)(\sum \ln(1-X_i) - \sum \ln(1-Y_i))$$

α é constante se e só se $\sum \ln X_i = \sum \ln Y_i$

β é constante se e só se $\sum \ln(1-X_i) = \sum \ln(1-Y_i)$.

ⓑ e ⓒ ambos os casos seguem análogo ao ⓐ com a mesma estatística:

$$T(X_1, \ldots, X_n) = \left(\sum \ln X_i, \sum \ln(1-X_i)\right).$$

/ /

(5) (Regressão logística). Seja $X_1, \ldots, X_n$ variáveis independentes de Bernoulli, com $p_i = P(X_i = 1)$, $i = 1, \ldots, n$. Seja $t_1, \ldots, t_n$ uma sequência de constantes conhecidas relacionadas com p_i via

$$\log \frac{p_i}{1-p_i} = \alpha + \beta t_i,$$

onde α e β são parâmetros desconhecidos. Determine uma estatística suficiente mínima para a família de distribuições conjuntas.

$$P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} p_i^{x_i} (1-p_i)^{1-x_i}$$

Podemos escrever $p_i = \dfrac{e^{\alpha+\beta t_i}}{1+e^{\alpha+\beta t_i}}$, daí

$$P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} \left(\frac{e^{\alpha+\beta t_i}}{1+e^{\alpha+\beta t_i}}\right)^{x_i} \left(\frac{1}{1+e^{\alpha+\beta t_i}}\right)^{1-x_i}$$

$$= \prod_{i=1}^{n} \frac{e^{x_i(\alpha+\beta t_i)}}{1+e^{\alpha+\beta t_i}}$$

$\Rightarrow$ ~~$\ln P(X_1 = x_1, \ldots, X_n = x_n)$~~

~~$\ln P\alpha$~~

~~$\Rightarrow \ln L(\alpha, \beta) = \sum_{i=1}^{n} \left[x_i(\alpha + \beta t_i) - \log(1 + e^{\alpha+\beta t_i}) \right]$~~

$$\ln\left(\frac{P(X_1 = x_1, \ldots, X_n = x_n)}{P(Y_1 = y_1, \ldots, Y_n = y_n)}\right) = \ln \prod_i \frac{e^{x_i(\alpha+\beta t_i)}}{e^{y_i(\alpha+\beta t_i)}}$$

$$= \ln \prod_i \exp\{\alpha(x_i - y_i) + \beta(t_i x_i - t_i y_i)\}$$

/ /

$= \alpha\left(\sum X_i - \sum Y_i\right) + \beta\left(\sum t_i X_i - \sum t_i Y_i\right)$

Logo, uma estatística suficiente mínima para α e β é
$T(X_1, \ldots, X_n) = \left(\sum X_i, \sum X_i t_i\right)$

/ /

6) Seja f uma função positiva integrável em $(0,\infty)$. Definida por

$$c(\theta) = 1 \Big/ \int_{\theta}^{\infty} f(x)\,dx,$$

e tome $p_\theta(x) = c(\theta) f(x)$, $x > \theta$, e $p_\theta(x) = 0$, $x \leq \theta$. Seja $X_1, \dots, X_n$ iid com densidade p_θ. Sendo $M = \min X_i$ uma estatística suficiente. Mostre que M é suficiente mínima.

$$\frac{L(\theta \mid X_1, \dots, X_n)}{L(\theta \mid Y_1, \dots, Y_n)} = \frac{\prod c(\theta) f(X_i)\, \mathbb{1}(X_i > \theta)}{\prod c(\theta) f(Y_i)\, \mathbb{1}(Y_i > \theta)}$$

$$= \frac{\prod f(X_i)}{\prod f(Y_i)} \cdot \frac{\mathbb{1}(\min X_i > \theta)}{\mathbb{1}(\min Y_i > \theta)}$$

É constante em relação a θ se e só se $\min X_i = \min Y_i$, logo $\min X_i$ é uma estatística suficiente minimal.

(7) Seja Z_1 e Z_2 normais independentes $\mu=0$ e $\sigma^2=1$ com densidade comum $\phi(x) = \exp\{-x^2/2\}/\sqrt{2\pi}$, e suponha X e Y são variaveis relacionadas com essas da seguinte forma

$$X = Z_1 \quad e \quad Y = (X+Z_2)\theta$$

onde $\theta > 0$ e desconhecido.

(a) Encontre a distribuição conjunta de X e Y.

(b) Suponha um conjunto de dados $(X_1,Y_1),\dots,(X_n,Y_n)$ iid com distribuição comum com X e Y da parte (a),

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim \begin{pmatrix} X \\ Y \end{pmatrix},$$

encontre a estatística minima suficiente.

(a) Sabemos que $X = Z_1$, $Z_1 \sim N(0,1)$, então $X \sim N(0,1)$. Então a densidade de X é

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Para Y, dado $X=x$, $Y=(x+Z_2)\theta$. Então $Y|X=x$ tem distribuição

$$Y|X=x \sim N(x\theta, \theta^2).$$

Daí a distribuição conjunta de X e Y

$$f_{X,Y}(x,y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot \frac{1}{\sqrt{2\pi\theta^2}} \exp\left\{ -\frac{(y-x\theta)^2}{2\theta^2} \right\}$$

/ /

$$= \frac{1}{2\pi\theta} \exp\left\{ -\frac{x^2}{2} - \frac{(y-x\theta)^2}{2\theta^2} \right\}$$

(b) $$\ln \frac{L(\theta|\{x_i, y_i\})}{L(\theta|\{w_i, z_i\})} = \ln \frac{\prod \exp\left\{ -\frac{x_i^2}{2} - \frac{(y_i - x_i\theta)^2}{2} \right\}}{\prod \exp\left\{ -\frac{w_i^2}{2} - \frac{(z_i - w_i\theta)^2}{2} \right\}}$$

$$= \ln \prod \exp\left\{ \frac{w_i^2}{2} + \frac{(z_i - w_i\theta)^2}{2} - \frac{x_i^2}{2} - \frac{(y_i - x_i\theta)^2}{2} \right\}$$

$$= \sum \frac{w_i^2}{2} + \sum \frac{(z_i - w_i\theta)^2}{2} - \sum \frac{x_i^2}{2} - \sum \frac{(y_i - x_i\theta)^2}{2}$$

$$\propto \sum (z_i^2 - 2z_i w_i \theta + w_i^2\theta^2) - \sum (y_i^2 - 2x_i y_i \theta + x_i^2\theta^2)$$

$$\propto \theta^2 \left(\sum w_i^2 - x_i^2\right) + \theta\left(\sum x_i y_i - z_i w_i\right)$$

Logo, uma estatística suficiente minimal

$$T(X_1, \ldots, X_n) = \left(\sum X_i Y_i, \sum X_i^2\right)$$

/ /

(8) Use completude para a família $N(0,1)$, $\theta \in \mathbb{R}$ para encontrar uma solução essencialmente única f da seguinte equação

$$\int f(x) e^{\theta x} dx = \sqrt{2\pi}\, e^{\theta^2/2}, \quad \theta \in \mathbb{R}.$$

Supondo que f é da forma

$$f(x) = A e^{-\frac{x^2}{2}}$$

com A constante. Substituindo na integral

$$\int A e^{-\frac{x^2}{2}} e^{\theta x} dx = A \int e^{-\frac{x^2}{2} + \theta x} dx$$

$$= A \int e^{-\frac{1}{2}(x^2 - 2\theta x)} dx$$

$$= A \int e^{-\frac{1}{2}((x-\theta)^2 - \theta^2)} dx$$

$$= A e^{\theta^2/2} \int e^{-\frac{1}{2}(x-\theta)^2} dx$$

$$= A e^{\frac{\theta^2}{2}} \sqrt{2\pi} = \sqrt{2\pi}\, e^{\theta^2/2}$$

Ou seja, $A = 1$. Daí, $f(x) = e^{-\frac{x^2}{2}}$.

/ /

9) Seja $X_1, ..., X_n$ uma amostra aleatória para uma distribuição contínua com densidade

$$f_\theta(x) = \begin{cases} 2x/\theta^2, & x \in (0,\theta); \\ 0, & c.c \end{cases}$$

a) Encontre uma estatística suficiente T

b) Determine a densidade de T.

c) Mostre diretamente que T é completa

a)
$$L(\theta | X_1, ..., X_n) = \prod_i \frac{2X_i}{\theta^2} \mathbb{I}(X_i < \theta)$$
$$= \left(\frac{2}{\theta^2}\right)^n \prod X_i \quad \mathbb{I}(\max X_i < \theta).$$

$T(X_1, ..., X_n) = \max X_i$ é uma estatística suficiente

b)
$$F_T(t) = P(X_{max} < t) = P(X_1 < t, ..., X_n < t)$$
$$= (P(X_1 < t))^n = \left(\frac{t}{\theta}\right)^{2n}$$

$$f_T(t) = \frac{\partial}{\partial t}\left(\frac{t^{2n}}{\theta^{2n}}\right) = \frac{2n t^{2n-1}}{\theta^{2n}}$$

c) Supondo $E[g(t)] = 0$,

$$E[g(T)] = \int_0^\theta g(t) f_T(t)\, dt$$
$$= \int_0^\theta g(t) \frac{2n t^{2n-1}}{\theta^{2n}}\, dt$$

$$(u = t/\theta) \quad = \int_0^1 g(u\theta)\, \frac{2n\, t\, u^{2n-1}}{\theta^{2n}}\, \theta\, du$$

$$= \int_0^1 g(u\theta)\, \underbrace{\frac{2n\, u^{2n-1}}{\theta^{2n-1}}}_{\neq 0}\, du$$

Daí, $g(t) = 0$.

Digitalizado com CamScanner