Inferência e Decisão I (LMAC)

2003/04

2º Teste

26 de Novembro de 2003

LMAC

Resolução do teste

Exercício .1 Considere uma amostra aleatória de vectores bidimensionais $X_i = (Y_i, Z_i)$, nas condições enunciadas, com $\{Y_i, i = 1, \dots, n\} \sim_{i.i.d.} Y \sim Bin(m, p)$ e $\{Z_i, i = 1, \dots, n\} \sim_{i.i.d.} Z \sim Geo(p)$.

(a) Seja $\underline{x} = ((y_1, z_1), \dots, (y_n, z_n))$ uma amostra de $X = (Y, Z)$ de dimensão n, com $(y_i, z_i) \in \{0, \dots, m\} \times \mathbb{N}$, $\forall i = 1, \dots, n$. Considere-se $L(p|\underline{x})$ a função de verosimilhança definida para $p \in (0, 1)$. Então

$$\begin{aligned}
L(p|\underline{x}) &= \prod_{i=1}^{n} P(Y_i = y_i, Z_i = z_i|p) \\
&= \prod_{i=1}^{n} P(Y_i = y_i|p) P(Z_i = z_i|p) \quad Y_i \text{ e } Z_i \text{ são independentes} \\
&= \prod_{i=1}^{n} \left[\binom{m}{y_i} p^{y_i} (1-p)^{m-y_i} (1-p)^{z_i - 1} p\right] \\
&= \left[\prod_{i=1}^{n} \binom{m}{y_i}\right] p^{\sum_{i=1}^{n}(y_i+1)} (1-p)^{\sum_{i=1}^{n}(z_i - y_i + m - 1)} \\
\ln L(p|\underline{x}) &= \ln\left[\prod_{i=1}^{n} \binom{m}{y_i}\right] + \sum_{i=1}^{n}(y_i+1)\ln p + \sum_{i=1}^{n}(z_i - y_i + m - 1)\ln(1-p) \\
\frac{\mathrm{d} \ln L(p|\underline{x})}{\mathrm{d}p} &= \frac{\sum_{i=1}^{n}(y_i+1)}{p} - \frac{\sum_{i=1}^{n}(z_i - y_i + m - 1)}{1-p} \\
\frac{\mathrm{d}^2 \ln L(p|\underline{x})}{\mathrm{d}p^2} &= -\frac{\sum_{i=1}^{n}(y_i+1)}{p^2} - \frac{\sum_{i=1}^{n}(z_i - y_i + m - 1)}{(1-p)^2} \\
&= -n\left[\frac{\bar{y}+1}{p^2} + \frac{(\bar{z}-1) + (m-\bar{y})}{(1-p)^2}\right] < 0
\end{aligned}$$

$\forall p \in (0, 1), \bar{z} \in [1, \infty), \bar{y} \in [0, m], m \in \mathbb{N}$, pelo que uma estimativa de máxima verosimilhança para esta amostra é o ponto onde a primeira derivada se anula, i.e, $emv(p) = \hat{p}(\underline{x})$ é raiz da equação $\left.\frac{\sum_{i=1}^{n}(y_i+1)}{p} - \frac{\sum_{i=1}^{n}(z_i - y_i + m - 1)}{1-p}\right|_{p=\hat{p}(\underline{x})} = 0$, i.e.,

$$\hat{p}(\underline{x}) = \frac{1+\bar{y}}{\bar{z}+m}$$

donde

$$EMV(p) = \hat{p}(\underline{X}) = \frac{1+\bar{Y}}{m+\bar{Z}}$$

(b) Seja então

$$W_n = \frac{1+\bar{Y}_n}{m+\bar{Z}_n}$$

onde $\bar{Y}_n$ ($\bar{Z}_n$) designa a média amostral de $\underline{Y}$ ($\underline{Z}$) com base numa amostra de dimensão n. Para provar que a sucessão $\{W_n, n \in I\!N\}$ é uma sucessão de estimadores consistente de p temos de provar que

$$W_n \rightarrow_P p$$

i.e., temos de provar que a v.a. W_n converge em probabilidade para p. Da lei fraca dos grandes números resulta que $\bar{Y}_n \underset{\text{P}}{\rightarrow} E[Binomial(m,p)|p] = mp$ e que, pelo mesmo argumento, $\bar{Z}_n \underset{\text{P}}{\rightarrow} E[Geo(p)|p] = \frac{1}{p}$. Decorre da convergência em probabilidades que

$$\begin{aligned}
&(i) \quad \bar{Y}_n \underset{\text{P}}{\rightarrow} p \Rightarrow 1+\bar{Y}_n \underset{\text{P}}{\rightarrow} 1+p \\
&(ii) \quad \bar{Z}_n \underset{\text{P}}{\rightarrow} \frac{1}{p} \Rightarrow m+\bar{Z}_n \underset{\text{P}}{\rightarrow} m+\frac{1}{p} \neq 0 \\
&(i)+(ii) \Rightarrow \frac{1+\bar{Y}_n}{m+\bar{Z}_n} \underset{\text{P}}{\rightarrow} \frac{1+mp}{m+\frac{1}{p}} = p
\end{aligned}$$

o que prova o resultado.

(c) Seja $X = (Y, Z)$. Note-se que a f.d.p. conjunta de Y e Z, dado que as v.a. são independentes, é o produto das funções densidade de probabilidade de cada uma delas. Uma vez que o parâmetro m é conhecido (ou assume-se que é) então cada uma destas f.d.p. pertence à família exponencial pelo que a f.d.p. conjunta satisfaz as condições de regularidade necessárias para que $E\left[\left(\frac{\delta}{\delta p}\ln f(X|p)|p\right)^2\right] = -E\left[\frac{\delta^2}{\delta p^2}\ln f(X|p)|p\right]$. Seja $x = (y,z) \in \{0,\ldots,m\} \times I\!N$. Note-se que

$$\begin{aligned}
f(x|p) &= \tbinom{m}{y} p^y (1-p)^{(m-y)} (1-p)^{z-1} p \\
\ln f(x|p) &= \ln \tbinom{m}{y} + y\ln p + (m-y)\ln(1-p) + (z-1)\ln(1-p) + \ln p \\
\frac{\delta}{\delta p}\ln f(x|p) &= \frac{y+1}{p} - \frac{z-1+m-y}{1-p} \\
\frac{\delta^2}{\delta p^2}\ln f(x|p) &= -\frac{y+1}{p^2} - \frac{z-1+m-y}{(1-p)^2}
\end{aligned}$$

Logo

$$\begin{aligned}
-E\left[\frac{\delta^2}{\delta p^2}\ln f(X|p)\middle| p\right] &= E\left[\frac{Y+1}{p^2} + \frac{Z-1+m-Y}{(1-p)^2}\middle| p\right] \\
&= \frac{mp+1}{p^2} + \frac{\frac{1}{p}-1+m-mp}{(1-p)^2} \\
&= \frac{m}{p(1-p)} - \frac{1}{p^2(1-p)}
\end{aligned}$$

Logo o limite inferior da desigualdade de Frechet-Cramér-Rao para a variância de um estimador centrado de p, $T_n(\underline{X})$, baseada numa amostra aleatória $\underline{X} = ((Y_1, Z_1), \ldots, (Y_n, Z_n))$ de dimensão n é dado por

$$\begin{aligned} Var(T_n(\underline{X})|p) &\geq \frac{1}{n\left(\frac{m}{p(1-p)} - \frac{1}{p^2(1-p)}\right)} \\ &\geq \frac{p^2(1-p)}{n\,(mp-1)} \end{aligned}$$

(d) *Suponhamos* então que $W_n = \frac{1+\bar{Y}_n}{m+\bar{Z}_n}$ é um estimador centrado. Note-se que $f_{Bin(m,p),Geo(p)}(.|p)$ pertence à família exponencial (com m conhecido) uma vez que

$$\begin{aligned} f_{Bin(m,p),Geo(p)}((y,z)|p) &= \binom{m}{y} p^y(1-p)^{m-y}p(1-p)^{z-1} \\ &= \binom{m}{y} p(1-p)^{m-1}e^{y\ln p+(z-y)\ln(1-p)} \end{aligned}$$

pelo que recorrendo à definição de família exponencial vem

$$\begin{aligned} &h(y,z) = \binom{m}{y} > 0, \quad c(p) = p(1-p)^{m-1} \geq 0, \quad k=2 \\ &w_1(p) = \ln p, t_1(y,z) = y, \quad w_2(p) = \ln(1-p), \quad t_2(y,z) = (z-y) \end{aligned}$$

Do critério da factorização decorre que a estatística

$$T(\underline{X}) = T\left((Y_1, Z_1), \ldots, (Y_n, Z_n)\right) = (\bar{Y}, \bar{Z})$$

é uma estatística suficiente. Além disso é completa uma vez que para qualquer função g da estatística T tal que $E\left[g(T)|p\right] = 0$ tem-se que

$$0 = E\left[g(T)|p\right] = \sum_{y=0}^{nm}\sum_{z=n}^{\infty} g\left(T(\frac{y}{n},\frac{z}{n})\right)\binom{nm}{y} p^y(1-p)^{nm-y}\binom{z-1}{z-n} p^n(1-p)^{z-n}$$

uma vez que a soma de n $Bin(m,p)$ independentes é ainda uma $Bin(nm,p)$ e a soma de n $Geo(p)$ independentes é uma $BN(n,p)$. Note-se que a expressão anterior é uma série pelo que para que para que esta seja nula para todos os valores de p então os seus coeficientes terão de ser nulos, o que implica que $g(T(\frac{y}{n},\frac{z}{n})) = 0, \forall y \in \{0,\ldots,nm\}, z \in \{n,\ldots,\}$, i.e., $g(T) = 0$ em todos os pontos com probabilidade diferente de zero, pelo que $g(T) = 0$ com probabilidade 1.

Temos então que W_n é função de uma estatística suficiente completa. Como por hipótese é um estimador centrado de p, então decorre do teorema de Rao-Blackwell que se trata de um estimador de variância mínima.

(e) Invocando o teorema das probabilidades totais temos:

$$\begin{aligned} P(Y > Z|p) &= \sum_{z=1}^{\infty} P(Y > z|p)P(Z = z|p) \\ &= \sum_{z=1}^{m}(1 - F_{Bin(m,p)}(z))p(1-p)^{z-1} \end{aligned}$$

Decorre então do princípio da invariância dos estimadores de máxima verosimilhança que $EMV\left(P(Y > Z|p)\right) = P(Y > Z|EMV(p))$, i.e.

$$EMV\left(P(Y > Z|p)\right) = \sum_{z=1}^{m}(1 - F_{Bin(m,\frac{1+\bar{Y}_n}{m+\bar{Z}_n})}(z))\frac{1+\bar{Y}_n}{m+\bar{Z}_n}\left(1 - \frac{1+\bar{Y}_n}{m+\bar{Z}_n}\right)^{z-1}$$

Exercício .2 Seja $\underline{X}$ uma amostra aleatória de dimensão n, proveniente de uma população $N(\theta, \theta^2)$.

1. Seja $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$ e $S_{n-1} = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$ dois estimadores de θ. Vejamos se se trata de dois estimadores centrados de θ, com $\theta > 0$.

$$\begin{aligned}
E\left[\bar{X}|\theta\right] &= E\left[\frac{\sum_{i=1}^n X_i}{n}|\theta\right] \\
&= \frac{1}{n}\sum_{i=1}^n E\left[X_i|\theta\right] \\
&= \frac{1}{n}\sum_{i=1}^n \theta \quad X_i's \text{ são v.a. i.i.d.} \\
&= \theta
\end{aligned}$$

pelo que $\bar{X}$ é efectivamente um estimador centrado de θ.

Sabe-se que $V = \frac{(n-1)S_{n-1}^2}{\theta^2} \sim \chi^2_{n-1}$, pelo que

$$f_{\frac{(n-1)S_{n-1}^2}{\theta^2}}(s) = f_V(s) = \frac{s^{\frac{n-1}{2}-1} e^{-s/2}}{2^{(n-1)/2}\Gamma(\frac{n-1}{2})} I_{I\!R^+}(s)$$

Então

$$\begin{aligned}
E\left[\frac{\sqrt{n-1}S_{n-1}}{\theta}|\theta\right] &= \int_0^\infty s^{\frac{1}{2}} f_V(s) ds \\
&= \int_0^\infty s^{\frac{1}{2}} \frac{s^{\frac{n-1}{2}-1} e^{-s/2}}{2^{(n-1)/2}\Gamma(\frac{n-1}{2})} ds \\
&= \int_0^\infty \frac{s^{\frac{n}{2}-1} e^{-s/2}}{2^{(n-1)/2}\Gamma(\frac{n-1}{2})} ds \\
&= \frac{2^{n/2}\Gamma(\frac{n}{2})}{2^{(n-1)/2}\Gamma(\frac{n-1}{2})} \int_0^\infty \frac{s^{\frac{n}{2}-1} e^{-s/2}}{2^{n/2}\Gamma(\frac{n}{2})} ds \\
&= \frac{\sqrt{2}\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}
\end{aligned}$$

pelo que

$$E\left[S_{n-1}|\theta\right] = \frac{\theta}{\sqrt{n-1}} \frac{\sqrt{2}\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}$$

donde resulta que cS_{n-1}, com c dado no enunciado, é efectivamente um estimador centrado de θ.

2. Dado que a f.d.p. normal obedece às condições de regularidade de Frechet-Cramér-Rao, e como por hipótese $W(\underline{X}) = W(\bar{X}, S_{n-1})$ é um estimador centrado, então o limite inferior da desigualdade de Frechet-Cramér-Rao é atingido se e só se existir alguma função $\alpha(\theta)$ tal que

$$\alpha(\theta)[W(\underline{x}) - \tau(\theta)] = \frac{\delta}{\delta\theta} \ln L(\theta|\underline{x})$$

para qualquer amostra $\underline{x}$ de X.

Note-se que

$$
\begin{aligned}
L(\theta|\underline{x}) &= \prod_{i=1}^{n}\left(\frac{1}{\theta\sqrt{2\pi}}e^{-\frac{(x_i-\theta)^2}{2\theta^2}}\right)\\
&= \left(\frac{1}{\theta\sqrt{2\pi}}\right)^n e^{-\sum_{i=1}^{n}\frac{(x_i-\theta)^2}{2\theta^2}}\\
&= \left(\frac{1}{\theta\sqrt{2\pi}}\right)^n e^{-\frac{\sum_{i=1}^{n}x_i^2+n\theta^2-2\theta n\bar{x}}{2\theta^2}}\\
\ln L(\theta|\underline{x}) &= -n\ln\theta\sqrt{2\pi}-\frac{\sum_{i=1}^{n}x_i^2}{2\theta^2}-\frac{n}{2}+\frac{n\bar{x}}{\theta}\\
\frac{\delta}{\delta\theta}\ln L(\theta|\underline{x}) &= -\frac{n}{\theta}+\frac{\sum_{i=1}^{n}x_i^2}{2\theta^3}-\frac{n\bar{x}}{\theta^2}
\end{aligned}
$$

Logo se existir um estimador que atinja o limite inferior de Frechet-Cramér-Rao teremos a seguinte decomposição:

$$\alpha(\theta)\left[W(\underline{x})-\theta\right] = -\frac{n}{\theta}+\frac{\sum_{i=1}^{n}x_i^2}{2\theta^3}-\frac{n\bar{x}}{\theta^2}$$

Porém como $\frac{\delta}{\delta\theta}\ln L(\theta|\underline{x})$ depende de θ através de θ^2 e θ^3, não é possível decompor da forma pretendida, pelo que não existe nenhum estimador centrado de θ para o qual o limite de FCR seja atingido.

3. Considere-se então a v.a. $X\sim N(\theta,\theta^2)$. Note-se que

$$
\begin{aligned}
f_X(x-\theta) &= \frac{1}{\theta\sqrt{2\pi}}e^{-\frac{x^2}{2\theta^2}} = g_1(\theta) \qquad \Rightarrow \theta \text{ não é parâmetro de localização}\\
\theta f_X(x\theta) &= \theta\frac{1}{\theta\sqrt{2\pi}}e^{-\frac{(\theta x-\theta)^2}{2\theta^2}} = \frac{1}{\sqrt{2\pi}}e^{-\frac{(x-1)^2}{2}} \qquad \Rightarrow \theta \text{ é parâmetro de escala}
\end{aligned}
$$

4. Seja Y_i o valor inteiro mais próximo de X_i, e $\underline{Y}$ uma amostra aleatória de Y, de dimensão n. Sem perda de generalidade assumimos que caso $x_i = *.5$ então $y_i = *+1$. A função de verosimilhança correspondente a uma amostra $\underline{y}$ de dimensão n, com $y_i\in Z, \forall i=1,\ldots,n$, é dada por

$$
\begin{aligned}
L(\theta|\underline{y}) &= \prod_{i=1}^{n}P(y_i-0.5\le X_i<y_i+0.5|\theta)\\
&= \prod_{i=1}^{n}\left[\Phi(\frac{y_i+0.5-\theta}{\theta})-\Phi(\frac{y_i-0.5-\theta}{\theta})\right]\\
\ln L(\theta|\underline{y}) &= \sum_{i=1}^{n}\ln\left(\Phi(\frac{y_i+0.5-\theta}{\theta})-\Phi(\frac{y_i-0.5-\theta}{\theta})\right)
\end{aligned}
$$

Exercício .3 Considere-se a família $\mathcal{F} = (x|\theta) = \left(\frac{\theta}{2}\right)^{|x|}(1-\theta)^{1-|x|}I_{\{-1,0,1\}}(x), \theta\in(0,1)\}$ e seja $x\in\{-1,0,1\}$ uma observação proveniente deste modelo.

1. Para $\theta \in (0,1)$ e $x \in \{-1, 0.1\}$:

$$\begin{aligned}
L(\theta|x) &= \left(\frac{\theta}{2}\right)^{|x|} (1-\theta)^{1-|x|} \\
\ln L(\theta|x) &= |x|(\ln\theta - \ln 2) + (1-|x|)\ln(1-\theta) \\
\frac{\mathrm{d}\ln L(\theta|x)}{\mathrm{d}\theta} &= \frac{|x|}{\theta} - \frac{1-|x|}{1-\theta} \\
\frac{\mathrm{d}^2\ln L(\theta|x)}{\mathrm{d}\theta^2} &= -\frac{|x|}{\theta^2} - \frac{1-|x|}{(1-\theta)^2} < 0, \quad \forall x \in \{-1,0,1\}, \theta \in (0,1)
\end{aligned}$$

pelo que a estimativa de máxima verosimilhança calculada com base numa observação x é o zero da primeira derivada, i.e.,

$$\left.\frac{\mathrm{d}\ln L(\theta|x)}{\mathrm{d}\theta}\right|_{\theta=\hat{\theta}(x)} = 0$$

$$\Longleftrightarrow$$

$$\frac{|x|}{\hat{\theta}(x)} - \frac{1-|x|}{1-\hat{\theta}(x)} = 0$$

$$\Longleftrightarrow$$

$$\hat{\theta}(x) = |x|$$

pelo que o estimador de máxima verosimilhança baseado numa observação (aleatória) X é dado por

$$EMV(\theta) = \hat{\theta}(X) = |X|$$

Quanto aos estimador dos momentos, atendendo a que $E[X|\theta] = 0, \forall\theta \in (0,1)$ (a f.d.p. é simétrica em torno do ponto zero), decorre que o estimador dos momentos de θ é dado por

$$EM(\theta) = \sqrt{EM(E[X^2|\theta])} = \sqrt{X^2} = |X|$$

pelo que nesta situação o EMV e o EM do parâmetro coincidem.

2. Considere-se:

$$T(X) = 2I_{\{1\}}(X)$$

Trata-se efectivamente de uma estatística mas de acordo com a nossa definição não é um estimador, pois toma valores que não pertencem ao espaço paramétrico. Note-se que

$$E[T(X)|\theta] = 2P(X=1|\theta) = 2\left(\frac{\theta}{2}\right) = \theta, \quad \forall\theta \in (0,1)$$

3. Na alínea anterior deduzimos que, com base numa observação, $|X|$ é o estimador de máxima verosimilhança. Vejamos se este estimador é centrado.

$$\begin{aligned}
E[|X||\theta] &= P(X=1|\theta) + P(X=-1|\theta) \\
&= \frac{\theta}{2} + \frac{\theta}{2} \\
&= \theta
\end{aligned}$$

pelo que $|X|$ é efectivamente um estimador centrado. Além disso é função da estatística suficiente completa, pelo que do teorema de Karlin-Rubin decorre que

$|X|$ é o estimador UMVUE pelo que, em particular, é melhor que $T(X)$, uma vez que sendo estimador função de uma estatística suficiente completa ele é único a menos de um conjunto de medida nula (note-se que $T(X) \neq |X|$ com probabilidade $1 - P(X = 0|\theta) \neq 0$).