IND 1115 – Inferência Estatística – Semestre 2004.02
Teste 2 – 30/11/2004
GABARITO

Problema 1 (30 pontos)

Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma densidade Uniforme$(0,\theta)$ onde θ é desconhecido.

a) Mostre que o MLE (estimador de máxima verossimilhança) de θ é:
$X_{(n)} = \text{máx}(X_1, X_2, \ldots, X_n)$

b) Encontre um estimador por método de momentos de θ.
c) O MLE de θ é não tendencioso?
d) O MLE de θ é consistente?
e) Coleta-se uma amostra de tamanho 10 da Uniforme$(0,\theta)$. Os valores observados são:
1,54 0,20 2,42 3,60 3,94
3,80 0,06 1,60 3,44 0,56
Compare o MLE e o estimador por método dos momentos de θ baseados nesta amostra.

Dica:
A função de distribuição de $X_{(n)}$ é: $\Pr(X_{(n)} \le x) = \Pr(\text{máx}(X_1, X_2, \ldots, X_n) \le x) =$
$= \Pr(X_1 \le x, X_2 \le x, \ldots, X_n \le x)$

Solução
A verossimilhança é:

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} \quad \text{se } 0 \le x_i \le \theta \quad \text{para todo } x_i$$

$$= \begin{cases} \dfrac{1}{\theta^n} & \text{se } 0 \le x_{(1)} \le x_{(n)} \le \theta \\ 0 & \text{se algum } x_i \text{ fora do intervalo } [0,\theta] \end{cases}$$

Mas, a verossimilhança é decrescente e seu máximo é alcançado em $x_{(n)} = \text{máx}(x_1, x_2, \ldots, x_n)$.

Também:

$Y_i = X_i / \theta =$ são iid Unif(0,1) , o que pode ser facilmente mostrado pelo método do jacobiano.

$Y_{(n)} = \text{máx}(Y_1, Y_2, \ldots, Y_n) = \text{máx}(X_1,/\theta, X_2,/\theta, \ldots, X_n/\theta) = X_{(n)}/\theta$

Mas, $Y_{(n)}$ é Beta (n,1) e então:

$$E\left(Y_{(n)}\right)=\frac{n}{n+1}$$

$$VAR\left(Y_{(n)}\right)=\frac{n}{(n+2)(n+1)^2}$$

Logo,

$$E\left(\frac{X_{(n)}}{\theta}\right)=\frac{n}{n+1}\Rightarrow E\left(X_{(n)}\right)=\frac{n\theta}{n+1}\Rightarrow \text{o MLE e' tendencioso}$$

$$VAR\left(\frac{X_{(n)}}{\theta}\right)=\frac{1}{\theta^2}VAR\left(X_{(n)}\right)=\frac{n}{(n+2)(n+1)^2}\Rightarrow VAR\left(X_{(n)}\right)=\frac{n\theta^2}{(n+2)(n+1)^2}$$

$MSE(X_{(n)}) = VAR(X_{(n)}) + \{ BIAS(X_{(n)}) \}^2$

Mas:

$$BIAS\left(X_{(n)}\right)=E\left(X_{(n)}\right)-\theta=\frac{n}{n+1}\theta-\theta=\frac{-\theta}{n+1}$$

$$\left[BIAS\left(X_{(n)}\right)\right]^2=\frac{\theta^2}{(n+1)^2}$$

$$MSE\left(X_{(n)}\right)=\frac{n\theta^2}{(n+2)(n+1)^2}+\frac{\theta^2}{(n+1)^2}=\frac{\theta^2}{(n+1)^2}\left[\frac{n}{n+2}+1\right]$$

$$e\ \ MSE\left(X_{(n)}\right)\to 0 \quad \text{quando n } \to\infty$$

Logo, o MLE é consistente.

A média da distribuição é: $\theta/2$. Logo, o estimador por método de momentos iguala a média amostral à média da distribuição, ou seja:

$$\frac{\tilde{\theta}}{2}=\overline{X}\Rightarrow\tilde{\theta}=2\overline{X}$$

O MLE é o máximo das observações, neste caso igual a **3.94**

Neste caso temos:
Média amostral = 2.116 e então o estimador por método de momentos é **4.232**

Problema 2 (20 pontos)

Sejam X_1, X_2, ..., X_n iid Poisson(λ). Encontre o estimador de máxima verossimilhança de Pr{ X_i = 0) = $e^{-\lambda}$. Verifique que este estimador de máxima verossimilhança é tendencioso. Encontre um estimador não tendencioso de $e^{-\lambda}$.

Dica: Considere a função geradora de momentos dos X's.

Solução

A verossimilhança é:

$$L(\lambda)=\prod_{i=1}^{n}\frac{e^{-\lambda}\lambda^{x_i}}{x_i!}=\frac{e^{-n\lambda}\lambda^{\sum x_i}}{\prod_{i=1}^{n}x_i!}$$

A log-verossimilhança é:

$$l(\lambda)=-n.\lambda+\left(\ln(\lambda)\right)\sum_{i=1}^{n}x_i-\ln\left(\prod_{i=1}^{n}x_i!\right)$$

A primeira derivada da log-verossimilhança é:

$$\frac{dl}{d\lambda}=-n+\frac{1}{\lambda}\sum_{i=1}^{n}x_i$$

Igualando esta derivada a zero obtemos o estimador de máxima verossimilhança de λ.

$$\frac{dl}{d\lambda}=-n+\frac{1}{\lambda}\sum_{i=1}^{n}x_i=0\Leftrightarrow\frac{n\bar{X}}{\lambda}=n\Leftrightarrow\hat{\lambda}=\bar{X}\ \text{ e´ o estimador de max. veross. de } \lambda$$

Este estimador é não tendencioso para λ, pois $E(X_i)=E(\bar{X})=\lambda$

Pelo princípio de invariância dos estimadores de máxima verossimilhança, o estimador desejado é:

$$T(\underset{\sim}{X})=e^{-\bar{X}}=e^{-\frac{\sum x_i}{n}}$$

A média deste estimador é:

$$E\left(T(\underset{\sim}{X})\right)=E\left(e^{-\frac{\sum x_i}{n}}\right)=M\left(\frac{-1}{n}\right)$$

Onde M(.) é a função geradora de momentos de Σx_i. Mas, Σx_i é Poisson($n.\lambda$) e portanto sua função geradora de momentos é:

$$M(t) = \exp\left\{n.\lambda\left(e^t - 1\right)\right\}$$

Logo:

$$M\left(\frac{-1}{n}\right) = \exp\left\{n.\lambda\left(e^{-1/n} - 1\right)\right\}$$

Então certamente o estimador é tendencioso para Pr{ $X_i = 0$} = $e^{-\lambda}$.

Note que:

$$\left(e^{-1/n} - 1\right) = 1 + \left(\frac{-1}{n}\right) + \frac{1}{2!}\left(\frac{-1}{n}\right)^2 + \frac{1}{3!}\left(\frac{-1}{n}\right)^3 + \ldots - 1 \approx \left(\frac{-1}{n}\right) + \frac{1}{2!}\left(\frac{-1}{n}\right)^2$$

Daí, até o termo de ordem 2 da série de Taylor:

$$M\left(\frac{-1}{n}\right) \approx \exp\left\{n.\lambda\left(\frac{-1}{n} + \frac{1}{2n^2}\right)\right\} = \exp\left\{-\lambda + \frac{\lambda}{2n}\right\}$$

Construção de um estimador não tendencioso

Para i = 1,2, ..., n sejam

$$U_i = \begin{cases} 0 & \text{se } \ \mathrm{X_i} \neq 0 \\ 1 & \text{se } \ \mathrm{X_i} = 0 \end{cases}$$

Pr(U_i) = 0 é 1 – exp($-\lambda$) e U_i = 1 com probabilidade exp($-\lambda$). Logo, a média de U_i é apenas:

0.{ 1 – exp($-\lambda$)} + 1.{ exp($-\lambda$)} = exp($-\lambda$)

Considere o estimador:

$$\tilde{\lambda} = \bar{U} = \frac{1}{n}\sum U_i$$

Por construção, este estimador tem a mesma média que U_i e portanto é não tendencioso para exp($-\lambda$).

Problema 3 (20 pontos)

Sejam $X_1, X_2, ..., X_n$ iid $N(\mu, 1)$.

a) Encontre o MLE de μ.
b) Mostre que o MLE é não tendencioso.
c) Calcule a informação de Fisher.
d) Calcule o limite inferior de Cramer e Rao. O MLE é um estimador eficiente?

Solução

A verossimilhança é:

$$L(\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left\{ \frac{-(x_i - \mu)^2}{2} \right\} = (2\pi)^{-n/2} \exp\left\{ \frac{-1}{2} \sum_{i=1}^{n} (x_i - \mu)^2 \right\}$$

$$l(\mu) = \log L(\mu) = -\frac{n}{2} \log(2\pi) \frac{-1}{2} \sum_{i=1}^{n} (x_i - \mu)^2$$

$$\frac{dl}{d\mu} = 0 \Rightarrow \frac{-1}{2} \frac{d}{d\mu} \left\{ \sum x_i^2 - 2\mu \sum x_i + n\mu^2 \right\} = 0$$

]

$$\Leftrightarrow -2\sum x_i + 2n\mu = 0 \Leftrightarrow \hat{\mu} = \frac{\sum x_i}{n} = \bar{X}$$

b) A média amostral é um estimador não tendecioso para μ, logo o MLE é não tendencioso neste caso.

c) A informação de Fisher é dada por:

$$I(\mu) = -E\left[\frac{d^2 l}{d\mu^2} \right] = -E\left[\frac{d}{d\mu} \left\{ \sum x_i - n\mu \right\} \right] = n$$

Mas, $VAR(\bar{X}) = 1/n \Rightarrow I(\mu) = \dfrac{1}{VAR(\bar{X})}$

d) O CRLB é definido como $1/I(\mu) = 1/n$ neste caso, e como já vimos, igual à variância do MLE. Logo, o MLE é um estimador eficiente.

Problema 4 (10 pontos)

O consumo mensal em minutos por conta de celular numa certa região é uma v.a. com média 40 minutos e desvio padrão 12 minutos. Toma-se uma amostra de 24 usuários de celular.

a) Qual a probabilidade do tempo médio de uso na amostra exceder 45 minutos?
b) Qual a probabilidade do tempo médio de uso na amostra ser menor que 50 minutos?

Solução

A média amostral é uma variável com média e variância:

$$E\left(\bar{X}\right) = 40$$

$$VAR\left(\bar{X}\right) = \frac{(12)^2}{24} = 6$$

$$a)\Pr\left(\bar{X} > 45\right) = \Pr\left(\frac{\bar{X}-40}{\sqrt{6}} > \frac{45-40}{\sqrt{6}}\right) \approx 1-\Phi\left(\frac{45-40}{\sqrt{6}}\right) = 1-\Phi\left(2.0412\right) =$$

$$= 1-0.9794 = 0.0206$$

$$b)\Pr\left(\bar{X} < 50\right) = \Pr\left(\frac{\bar{X}-40}{\sqrt{6}} < \frac{50-40}{\sqrt{6}}\right) \approx \Phi\left(\frac{50-40}{\sqrt{6}}\right) = \Phi\left(4.0824\right) = 1$$

Problema 5 (20 pontos)

Seja Y ~ Bin(12 , 1/2).

a) Calcule Pr(Y ≥ 9) exatamente.

b) Calcule Pr(Y ≥ 9) aproximadamente pelo teorema de DeMoivre e Laplace com correção de continuidade.

c) Calcule Pr(Y = 7) exatamente.

d) Calcule Pr(Y= 7) aproximadamente pelo teorema de DeMoivre e Laplace com correção de continuidade.

Solução

a) Cálculo exato:

x	**Pr(X=x)**
9	0.0537
10	0.0161
11	0.0029
12	0.0002
soma	**0.0730**

b) Pr(Y ≥ 9) é, com a correção de continuidade, aproximadamente igual a Pr(Y ≥ 8.5), onde esta última é calculada a partir da distribuição Normal.

E(Y) = np = 6
VAR(Y) = npq = 6(1/2) = 3

$$\Pr\left(Y \geq 9\right) \approx \Pr\left(Y \geq 8.5\right) = \Pr\left(\frac{Y-6}{\sqrt{3}} \geq \frac{8.5-6}{\sqrt{3}}\right) = 1 - \Phi\left(\frac{8.5-6}{\sqrt{3}}\right) = 1 - \Phi\left(1.4434\right) =$$
$$= 1 - 0.9255 = 0.0745$$

c) $\Pr(Y = 7) = \binom{12}{7}\left(\frac{1}{2}\right)^{12} = 0.1934$

d) Pela aproximação do Teorema de DeMoivre e Laplace:

$$\Pr\left(Y = 7\right) \approx \Pr\left(6.5 \leq Y \leq 7.5\right) = \Pr\left(\frac{6.5-6}{\sqrt{3}} \leq \frac{Y-6}{\sqrt{3}} \leq \frac{7.5-6}{\sqrt{3}}\right) = \Phi\left(\frac{7.5-6}{\sqrt{3}}\right) - \Phi\left(\frac{6.5-6}{\sqrt{3}}\right) =$$
$$= \Phi\left(0.8660\right) - \Phi\left(0.2887\right) = 0.8068 - 0.6136 = 0.1932$$

Tabela – Função de Distribuição N(0,1)

z	Φ(z)		z	Φ (z)		z	Φ (z)
0,0000	50,00%		1,0000	84,13%		2,0125	97,79%
0,0200	50.80%		1,0100	84.38%		2.0200	97.83%
0,0300	51.20%		1.0167	84.54%		2.0300	97.88%
0,0400	51.60%		1.0250	84.73%		2.0400	97.93%
0,0500	51,99%		1,0500	85,31%		2.0412	97.94%
0,1000	53,98%		1,0553	85,44%		2.0500	97.98%
0,1500	55,96%		1,1000	86,43%		2,1000	98,21%
0,2000	57,93%		1,1180	86,82%		2,2000	98,61%
0,2236	58,85%		1,1475	87,44%		2,2361	98,73%
0,2500	59,87%		1,1500	87,49%		2,3000	98,93%
0,3000	61,79%		1,1553	87,60%		2,3263	99,00%
0,3015	61,85%		1,2000	88,49%		2,3333	99,02%
0,3475	63,59%		1,2060	88,61%		2,4000	99,18%
0,3492	63,65%		1,2200	88,88%		2,5000	99,38%
0,3500	63,68%		1,2500	89,44%		2,5500	99,46%
0,4000	65,54%		1.2700	89.79%		2,5628	99,48%
0.4167	66.16%		1,2816	90,00%		2,6000	99,53%
0,4307	66,67%		1,3000	90,32%		2,6500	99,60%
0,4500	67,36%		1,3333	90,88%		2,6667	99,62%
0,5000	69,15%		1,3750	91.54%		2,6833	99,64%
0,5500	70,88%		1,4000	91,92%		2,7000	99,65%
0,5774	71.81%		1,4468	92,60%		2.7500	99.70%
0,6000	72,57%		1,4500	92,65%		2,8000	99,74%
0,6250	73,40%		1,5000	93,32%		2,9000	99,81%
0,6500	74,22%		1,5500	93,94%		2,9500	99,84%
0,6667	74,75%		1,5811	94,31%		3,0000	99,87%
0,6708	74,88%		1,6000	94,52%		3,1000	99,90%
0,7000	75,80%		1,6450	95,00%		3,1500	99,92%
0,7500	77,34%		1,6667	95,22%		3,2000	99,93%
0,8000	78,81%		1,7000	95,54%			
0,8333	79.77%		1,8000	96,41%			
0,8500	80,23%		1.8500	96.78%			
0.8666	80.69%		1,9000	97,13%			
0,8944	81,45%		1.9500	97.44%			
0,9000	81,59%		1,9600	97,50%			
0,9167	82,03%		1,9800	97,61%			
0,9500	82,89%		2,0000	97,72%			
0,9500	82,89%		2,0100	97,78%			
0,9750	83,52%		2,0100	97,78%			
0,9800	83,65%						
0,9900	83,89%						