IND 1115 – Inferência Estatística – Semestre 2007.01
Teste 3 – 21/06/2007
Nome:

Escreva as respostas como frações ou 4 casas decimais.

PROBLEMA 1 (20 pontos)

Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de idades de alunos de graduação ao terminarem seu curso de Engenharia. Sabe-se que as idades são variáveis iid Uniforme (21, θ) onde θ é desconhecido.

a) Encontre o MLE (estimador de máxima verossimilhança) de θ.
b) Encontre um estimador por método de momentos de θ.
c) O MLE de θ é não tendencioso?
d) Toma-se uma amostra de 10 alunos formandos e observa-se que as idades (ordenadas) são:

21.2	22.3	22.5	22.9	23.1
23.4	23.6	23.8	24.2	24.5

Encontre o estimador de máxima verossimilhança de θ e compare-o com o estimador por método de momentos.

Solução

A densidade de cada X_i é:

$$f(x) = \frac{1}{\theta - 21} \quad \text{se } 21 \le x \le \theta$$

A verossimilhança é:

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\theta - 21} = \frac{1}{(\theta - 21)^n} \quad \text{se } 21 \le x_{(1)} = \min(x_1, x_2, \ldots, x_n) \le \ldots \le x_{(n)} = \max(x_1, x_2, \ldots, x_n) \le \theta$$

Á medida que θ se aproxima de 21, a verossimilhança cresce, mas existe um valor mínimo possível para θ, que é dado por $x_{(n)}$, o máximo da amostra. Logo, o máximo da verossimilhança ocorre quando $\theta = x_{(n)}$.

b) Um estimador de θ por método de momentos iguala a média amostral à média da distribuição, que é $(\theta+21)/2$. Logo:

$$\overline{X} = \frac{\theta + 21}{2} \Leftrightarrow \tilde{\theta} = 2\overline{X} - 21$$ é o estimador por método de momentos.

c) A função de distribuição do MLE é:

$$F(x) = \Pr\left(X_{(n)} \le x\right) = \Pr\left(X_1 \le x, \ldots, X_n \le x\right) =$$

$$= \prod_{i=1}^{n} \Pr\left(X_i \le x\right) = \left\{\int_{21}^{x} \frac{1}{\theta - 21} du\right\}^n = \left(\frac{x - 21}{\theta - 21}\right)^n$$

A densidade do MLE é apenas a derivada desta função de distribuição, ou seja:

$$g(x) = \frac{dF}{dx} = \frac{d}{dx}\left\{\left(\frac{x-21}{\theta-21}\right)^n\right\} = n\left(\frac{1}{\theta-21}\right)^n (x-21)^{n-1} = \frac{n}{\theta-21}\left(\frac{x-21}{\theta-21}\right)^{n-1} \quad \text{para } x \in [21, \theta]$$

A média do MLE é:

$$E\left(X_{(n)}\right)=\int_{21}^{\theta} x\frac{n}{(\theta-21)^n}(x-21)^{n-1}\,dx=\frac{n}{(\theta-21)^n}\int_{21}^{\theta} x(x-21)^{n-1}\,dx=$$

$$=\frac{n}{(\theta-21)^n}\int_{0}^{\theta-21}(t+21)t^{n-1}dt=\frac{n}{(\theta-21)^n}\int_{0}^{\theta-21}\left(t^n+21t^{n-1}\right)dt=$$

$$=\frac{n}{(\theta-21)^n}\left\{\frac{t^{n+1}}{n+1}+\frac{21t^n}{n}\right\}\Bigg|_0^{\theta-21}=\frac{n}{(\theta-21)^n}\left\{\frac{(\theta-21)^{n+1}}{n+1}+\frac{21(\theta-21)^n}{n}-0\right\}=$$

$$=\frac{n}{n+1}(\theta-21)+21=\frac{n\theta-21n+21n+21}{n+1}=\frac{n\theta+21}{n+1}$$

Claramente o estimador de máxima verossimilhança é tendencioso para θ.

d) A média da amostra é:

$\bar{X}=23.15$

Logo, pelos resultados do item b), o estimador pelo método de momentos é:

$\tilde{\theta}=2\bar{X}-21=2(23.15)-21=25.3$

O estimador de máxima verossimilhança é o máximo da amostra, 24.5.

PROBLEMA 2 (15 pontos)

Seja X uma variável aleatória Qui-quadrado com 32 graus de liberdade. Use o teorema central do limite para aproximar:

a) Pr(X > 35.2)
b) Pr(X < 22.4)

Solução

X pode ser encarado como uma soma de 32 variáveis independentes, cada uma com densidade Qui-quadrado com 1 grau de liberdade, e portanto podemos usar o teorema central do limite.

A média de X é 32 e sua variância é 64. Logo:

$\frac{X-32}{\sqrt{64}}=\frac{X-32}{8}$ é aproximadamente N(0,1).

a) $\Pr\left(X>35.2\right)=\Pr\left(\frac{X-32}{8}>\frac{35.2-32}{8}\right)=1-\Phi\left(0.4\right)=1-0.6554=0.3446$

b) $\Pr\left(X<22.4\right)=\Pr\left(\frac{X-32}{8}>\frac{22.4-32}{8}\right)=\Phi\left(-1.2\right)=1-\Phi(+1.2)=1-0.8849=0.1151$

Problema 3 (25 pontos)

Sejam X_1, X_2, ..., X_n iid N(μ, θ) onde θ, **a variância, é CONHECIDA!**

a) Encontre o MLE de μ.
b) Mostre que o MLE é não tendencioso.
c) Calcule a informação de Fisher.
d) Mostre que o MLE é consistente.

Solução

A verossimilhança é:

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

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

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

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

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:

$$I(\mu)=-E\left[\frac{d^2l}{d\mu^2}\right]=-E\left[\frac{d}{d\mu}\left\{\frac{1}{\theta}\left(\sum x_i-n\mu\right)\right\}\right]=\frac{n}{\theta}$$

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

d) O MLE é claramente consistente, pois seu erro quadrático médio é igual à sua variância (θ/n) (já que ele é um estimador não tendencioso), e esta tende a zero quando n tende a infinito.

Problema 4 (15 pontos)

Um computador gera 9 números aleatórios uniformemente distribuídos no intervalo (0,1).

a) Calcule a probabilidade de que o menor destes números será menor que 0.12.

b) Calcule o valor esperado do *menor* destes números.

c) Calcule a probabilidade de que o segundo menor destes números seja menor que 0.2.

Dica: você pode citar resultados dos slides, ao invés de demonstrar explicitamente todos os passos necessários aqui.

Solução

Sejam $X_{(1)}, X_{(2)}, ..., X_{(9)}$ os elementos da amostra em ordem crescente, ou seja, $X_{(1)}$ = mín($X_1, X_2, ..., X_9$), $X_{(2)}$ é o 2º. menor, etc.... Sabemos que os elementos ordenados de uma amostra Uniforme(0,1) têm distribuição Beta. Em particular, o r-ésimo elemento tem distribuição Beta com parâmetros r e n – r +1.

a) Das considerações anteriores, $X_{(1)}$ tem densidade Beta(1, 9) e assim:

$$\Pr(X_{(1)}<0.12)=\int_0^{0.12}\frac{\Gamma(10)}{\Gamma(1)\Gamma(9)}x^{1-1}(1-x)^{9-1}dx=\int_0^{0.12}9(1-x)^{9-1}dx=\int_{0.88}^{1}9y^8dy=y^9\Big|_{0.88}^{1}=1-(0.88)^9=0.6835$$

b) O valor esperado de uma distribuição Beta com parâmetros α e β é $\alpha/(\alpha+\beta)$. Logo, na situação do item a, $E(X_{(1)})$ = 1/10 = 0.1

c) $X_{(2)}$ = 2º. Menor dentre $X_1, X_2, ..., X_9$ tem densidade Beta(2,8), que é:

$$f(x) = \frac{\Gamma(10)}{\Gamma(2)\Gamma(8)} x^{2-1}(1-x)^{8-1} = \frac{9!}{1!7!} x(1-x)^7 = 72x(1-x)^7 \quad 0 < x < 1$$

A probabilidade desejada é:

$$\Pr\left(X_{(2)} < 0.2\right) = \int_0^{0.2} 72x(1-x)^7 dx = \int_{0.8}^{1} 72t^7(1-t)dt = 72\left[\frac{t^8}{8} - \frac{t^9}{9}\right]\Bigg|_{0.8}^{1} = \left(9t^8 - 8t^9\right)\Big|_{0.8}^{1} =$$

$$= 1 - 0.4362 = 0.5638$$

Problema 5 (25 pontos)

Os preços de apartamentos de dois quartos em duas cidades X e Y são variáveis Normais correlacionadas. Na cidade X, o preço médio é R$ 140 mil, e o desvio padrão dos preços é R$ 15 mil. Na cidade Y, o preço médio é R$ 180 mil, e o desvio padrão dos preços é R$ 25 mil. A correlação entre os preços é ρ = +0.8. Calcule as seguintes probabilidades:

a) De alguém pagar entre R$ 121250 e R$ 158750 por um apartamento na cidade X.
b) De alguém pagar entre R$ 121250 e R$ 158750 por um apartamento na cidade X sabendo que um apartamento "equivalente" na cidade Y custa R$ 200 mil.
c) De alguém pagar entre R$ 121250 e R$ 158750 por um apartamento na cidade X sabendo que um apartamento "equivalente" na cidade Y custa R$ 217500.
d) Qual é a distribuição condicional dos preços de apartamento na cidade Y sabendo que o preço de um apartamento equivalente na cidade X é R$ 160 mil?
e) Qual é a distribuição condicional dos preços de apartamento na cidade Y sabendo que o preço de um apartamento equivalente na cidade X é R$ 120 mil?
f) Quanto deve pagar uma pessoa na cidade X para que o seu apartamento esteja entre os 10% mais caros?

Solução

a)
$$\Pr(121250 < X < 158750) = \Pr\left(\frac{121250-140000}{15000} < \frac{X-140000}{15000} < \frac{158750-140000}{15000}\right) =$$

$$= \Pr(-1.25 < Z < 1.25) = 2.\Phi(1.25) - 1 = 2(0.8944) - 1 = 0.7887$$

b) A distribuição condicional é Normal com média:

$$140000 + \frac{0.8(15000)}{(25000)}(200000 - 180000) = 140000 + 0.48(20000) = 149600$$

e variância:

$$(15000)^2.\left(1-(0.8)^2\right) = (15000)^2(0.36) = (0.60)^2(15000)^2$$

Ou seja, o desvio padrão condicional é 0.60(15000) = R$ 9000.

$$\Pr(121250 < X < 158750) = \Pr\left(\frac{121250-149600}{9000} < \frac{X-149600}{9000} < \frac{158750-149600}{9000}\right) =$$

$$= \Pr(-3.15 < Z < 1.0167) = \Phi(1.0167) - \Phi(-3.15) = 0.8454 - 0.0008 = 0.8445$$

c) A distribuição condicional é Normal com média:

$140000+\frac{0.8(15000)}{(25000)}\left(217500-180000\right)=140000+18000=158000$ e a mesma variância que no item b). Logo:

$$\Pr\left(121250<X<158750\right)=\Pr\left(\frac{121250-158000}{9000}<\frac{X-158000}{9000}<\frac{158750-158000}{9000}\right)=$$

$$=\Pr\left(-4.0833<Z<0.0833\right)=\Phi(0.0833)-\Phi(-4.0833)=0.5332-0.0000=0.5332$$

d) A distribuição condicional de Y dado que X = 160 mil é Normal com média:

$$180000+\frac{0.8(25000)}{(15000)}\left(16000-140000\right)=180000+26667=206667$$

E variância:

$$(25000)^2\left(1-\left(0.8\right)^2\right)=\left(15000\right)^2$$

e) A distribuição condicional de Y dado que X = 120 mil é Normal com média:

$$180000+\frac{0.8(25000)}{(15000)}\left(12000-140000\right)=180000-26667=153333$$

e a mesma variância que no item anterior.

f) Quanto deve pagar uma pessoa na cidade X para que o seu apartamento esteja entre os 10% mais caros?

Para estar entre os 10% mais caros, $\frac{X-140000}{15000}\geq 1.2816\Rightarrow X\geq 159224$

Tabela – Função de Distribuição N(0,1)

z	Φ(z)		z	Φ (z)		z	Φ (z)
0,0000	50,00%		0,9800	83,65%		2,0000	97,72%
0,0200	50.80%		0,9900	83,89%		2,0100	97,78%
0,0300	51.20%		1,0000	84,13%		2,0125	97,79%
0,0400	51.60%		1,0100	84.38%		2.0200	97.83%
0,0500	51,99%		1.0167	84.54%		2.0300	97.88%
0,1000	53,98%		1.0250	84.73%		2.0400	97.93%
0,1500	55,96%		1,0500	85,31%		2.0412	97.94%
0,2000	57,93%		1,0553	85,44%		2.0500	97.98%
0,2236	58,85%		1,1000	86,43%		2,1000	98,21%
0,2500	59,87%		1,1180	86,82%		2,2000	98,61%
0,3000	61,79%		1,1475	87,44%		2,2361	98,73%
0,3015	61,85%		1,1500	87,49%		2,3000	98,93%
0,3333	63,06%		1,1553	87,60%		2,3263	99,00%
0,3475	63,59%		1,1667	87,83%		2,3333	99,02%
0,3492	63,65%		1,2000	88,49%		2,4000	99,18%
0,3500	63,68%		1,2060	88,61%		2,5000	99,38%
0,4000	65,54%		1,2200	88,88%		2,5500	99,46%
0.4167	66.16%		1,2500	89,44%		2,5628	99,48%
0,4307	66,67%		1.2700	89.79%		2,6000	99,53%
0,4500	67,36%		1,2816	90,00%		2,6500	99,60%
0,5000	69,15%		1,3000	90,32%		2,6667	99,62%
0,5500	70,88%		1,3333	90,88%		2,6833	99,64%
0,5774	71.81%		1,3750	91.54%		2,7000	99,65%
0,6000	72,57%		1,4000	91,92%		2.7500	99.70%
0,6250	73,40%		1,4468	92,60%		2,8000	99,74%
0,6500	74,22%		1,4500	92,65%		2,9000	99,81%
0,6667	74,75%		1,5000	93,32%		2,9500	99,84%
0,6708	74,88%		1,5500	93,94%		3,0000	99,87%
0,7000	75,80%		1,5811	94,31%		3,1000	99,90%
0,7500	77,34%		1,6000	94,52%		3,1500	99,92%
0,8000	78,81%		1,6450	95,00%		3,1667	99,92%
0,8333	79.77%		1,6667	95,22%		3,2000	99,93%
0,8500	80,23%		1,7000	95,54%		3,8333	99,99%
0.8666	80.69%		1,8000	96,41%		4,0833	100.00%
0,8944	81,45%		1,8333	96,66%			
0,9000	81,59%		1.8500	96.78%			
0,9167	82,03%		1,9000	97,13%			
0,9500	82,89%		1.9500	97.44%			
0,9600	83,15%		1,9600	97,50%			
0,9700	83,40%		1,9800	97,61%			
0,9722	83,45%		1,9900	97.67%			
0,9750	83,52%		1,9950	97,70%			