IND 1115 – Inferência Estatística – Semestre 2005.01 – TURMA B
Teste 3 – 17/06/2005
GABARITO

Problema 1 (15 pontos)

Em cada questão abaixo, indique se a afirmativa é **verdadeira** ou **falsa** (marque um X na alternativa correta). Não é necessário justificar a sua resposta.

		Verdadeiro	Falso
1	A densidade F é simétrica em torno de zero.		XXX
2	Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma densidade com média a e variância b. Então a soma dos X_i´s tem também média a mas sua variância é b/n.		XXX
3	Um estimador não tendencioso é consistente se sua tendência ("bias") vai a zero quando o tamanho da amostra cresce indefinidamente.		XXX
4	Sejam X_1 e X_2 variáveis aleatórias com médias 0.2 e 0.3, variâncias 4 e 1 e coeficiente de correlação 0.6. Então a variável $Y = 0.5X_1 + 0.5X_2$ tem média 0.25.	XXX	
5	Na situação da questão 4, a variância de Y é 3.65.	XXX	
6	Na situação da questão 4, o desvio padrão de Y é 2.		XXX
7	Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma densidade Normal com média μ e variância σ^2. Então: $\dfrac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{\sigma^2}$ tem densidade Qui-quadrado com n graus de liberdade.		XXX
8	O retorno aritmético está limitado abaixo em –100%.	XXX	
9	O retorno geométrico entre os dias 1 e 5 é a soma dos retornos aritméticos diários no período.	XXX	
10	Um portfolio é uma combinação linear de ativos financeiros na qual a soma dos pesos de todos os ativos é 0.		XXX
11	Na distribuição Binomial(n, p), onde n > 1, a "regra de parada" é encontrar o primeiro sucesso.		XXX
12	Uma possível aplicação da distribuição Exponencial é como modelo de preço de ações.		XXX

13	Os estimadores por máxima verossimilhança são sempre consistentes.		XXX
14	Os estimadores por método de momentos são sempre não tendenciosos.		XXX
15	O método de momentos sempre fornece estimadores iguais aos de máxima verossimilhança.		XXX

Problema 2 (35 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 $\rho = +0.8$. 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?
g) Toma-se uma amostra de 9 apartamentos comprados recentemente na cidade X. Qual a probabilidade do maior preço pago na amostra exceder R$ 145 mil?

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

$$\Pr(121250 < X < 158750) = \Pr\left(\frac{121250-158000}{9000} < \frac{X-158000}{9000} < \frac{158750-158000}{9000}\right) =$$
$$= \Pr(-4.0833 < Z < 0.0833) = \Phi(0.0833) - \Phi(-4.0833) = 0.5332 - 0.0000 = 0.5332$$

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

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

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

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

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

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, $\frac{X-140000}{15000} \geq 1.2816 \Rightarrow X \geq 159224$

g) Toma-se uma amostra de 9 apartamentos comprados recentemente na cidade X. Qual a probabilidade do maior preço pago na amostra exceder R$ 145 mil?

Seja $X_{(n)}$ o maior preço pago na amostra. $\Pr(X_{(n)} > 145000) = 1 - \Pr(X_{(n)} \leq 145000) =$
$= 1 - \{\Pr(X_1 \leq 145000)\}^9$

Mas:

$$\Pr\left(X_1 \le 145000\right) = \Pr\left(\frac{X_1 - 140000}{15000} \le \frac{145000 - 140000}{15000}\right) = \Phi\left(\frac{1}{3}\right) = \Phi\left(0.3333\right) = 0.6305$$

Logo:

$\Pr(X_{(n)} > 145000) = 1 - (0.6305)^9 = 0.9842$

Problema 3 (25 pontos)

Sejam X_1 e X_2 variáveis aleatórias independentes, onde X_1 tem densidade $N(\mu_1, \sigma^2)$ e X_2 tem densidade $N(\mu_2, \sigma^2)$. Mostre, **usando o método que você preferir**, que $Y = X_1 + X_2$ tem densidade $N(\mu_1 + \mu_2, 2\sigma^2)$.

Dica:

Se X é $N(\mu, \sigma^2)$ então E(X) = μ, VAR(X) = $\sigma 2$ e a fgm é: $\exp\left(\mu.t + \frac{\sigma^2 t^2}{2}\right)$

Solução

A demonstração pela função geradora de momentos é trivial.

A fgm de Y é:

$M_Y(t) = E\left(e^{tY}\right) = E\left(e^{t(X_1 + X_2)}\right) = E\left(e^{tX_1} e^{tX_2}\right)$ e pela independência isso é igual a:

$E\left(e^{tX_1}\right)E\left(e^{tX_2}\right) = \exp\left(\mu_1 t + \frac{\sigma^2 t^2}{2}\right)\exp\left(\mu_2 t + \frac{\sigma^2 t^2}{2}\right) = \exp\left((\mu_1 + \mu_2)t + \frac{2\sigma^2 t^2}{2}\right)$, que é a função geradora de momentos de uma variável Normal com média $\mu_1 + \mu_2$ e variância $2\sigma^2$.

Problema 4 (25 pontos)

Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma densidade Uniforme(0,θ) 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,θ). Os valores observados, em ordem crescente, são:

0.043
0.302
0.416
1.146
1.222
1.789
2.590

2.654

2.697

2.875

Compare o MLE e o estimador por método dos momentos de θ baseados nesta amostra.

Dica:

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

Solução

A densidade de cada X_i é $1/\theta$ desde que X_i esteja no intervalo $(0, \theta)$.
Logo, a verossimilhança é:

$$L(\theta) = \frac{1}{\theta^n} \quad \text{se } 0 < \text{todo } X_i < \theta \Leftrightarrow L(\theta) = \frac{1}{\theta^n} \text{ se } 0 < X_{(1)} = \min \le X_{(2)} \le \dots \le X_{(n)} = \max < \theta$$

Então, para que a verossimilhança seja não nula, é necessário que $\theta > X_{(n)} = \text{máx}(X_1, X_2, \dots, X_n)$ e portanto o MLE é: $\hat{\theta} = X_{(n)}$ pois a verossimilhança é decrescente como função de θ.

b) Um estimador por método de momentos de θ é encontrado igualando-se a média amostral à média da distribuição, ou seja:

$$\overline{X} = \frac{\theta}{2} \Rightarrow \tilde{\theta} = 2\overline{X}$$ é o estimador por método de momentos de θ.

c) A média do MLE é encontrada a partir de sua densidade. Usando a dica:

$$\Pr(X_{(n)} \le x) = \Pr(X_1 \le x, X_2 \le x, \dots, X_n \le x) = \{\Pr(X_1 \le x)\}^n$$ pois os X_i são iid.

Usando agora o fato dos X_i serem Uniforme(0, θ) temos:

$$\Pr(X_1 \le x) = \int_0^x \frac{1}{\theta} du = \frac{x}{\theta} \quad \text{se} \quad 0 < x < \theta$$

$$\Pr(X_{(n)} \le x) = \{\Pr(X_1 \le x)\}^n = \left(\frac{x}{\theta}\right)^n$$

e a densidade de $X_{(n)}$ (o MLE) é obtida derivando-se esta função de distribuição.

$$g(x)=\frac{nx^{n-1}}{\theta^n} \quad 0<\mathrm{x}<\theta$$

A média de $X_{(n)}$ é encontrada a partir desta densidade:

$E\left(X_{(n)}\right)=\int_0^\theta x\frac{nx^{n-1}}{\theta^n}dx=\frac{n}{\theta^n}\left(\frac{\theta^{n+1}}{n+1}\right)=\frac{n}{n+1}\theta$ e portanto o MLE é um estimador ***tendencioso****.*

O 2^{o}. Momento de $X_{(n)}$ é obtido de maneira análoga.

$$E\left({X_{(n)}}^2\right)=\int_0^\theta x^2\frac{nx^{n-1}}{\theta^n}dx=\frac{n}{\theta^n}\left(\frac{\theta^{n+2}}{n+2}\right)=\frac{n}{n+2}\theta^2$$

A variância de $X_{(n)}$ é então:

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

A tendência de $X_{(n)}$ é:

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

d) O erro quadrático médio de $X_{(n)}$ é:

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

O limite do erro quadrático médio quando n tende a infinito é zero, e portanto $X_{(n)}$ é um estimador ***consistente*** de θ.

e) Neste caso:

O MLE é $X_{(n)}$ = 2.875

E o estimador por método de momentos é $2\overline{X}=2(1.5734)=3.1468$

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%			