IND 1115 – Inferência Estatística – Semestre 2005.01 – TURMA A
Teste 3 – 16/06/2005
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 assimétrica em torno de zero.	XXX	
2	Seja X_1, X_2, ... , X_n uma amostra aleatória de uma densidade com média a e variância b. Então a média amostral tem também média a mas sua variância é b/n.	XXX	
3	Um estimador não tendencioso é consistente se sua variância tende 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 5.		XXX
6	Na situação da questão 4, o desvio padrão de Y é 2.		XXX
7	Suponha que, num problema de estimação, você precisa estimar k parâmetros “livres” (não relacionados entre si). Então, para resolver o problema pelo método dos momentos você terá de resolver k equações.	XXX	
8	O retorno aritmético está limitado abaixo em –100%.	XXX	
9	O retorno aritmético 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 1.	XXX	
11	Na distribuição Binomial Negativa (r, p), onde r > 1, a “regra de parada” é encontrar o primeiro sucesso.	XXX	
12	Uma possível aplicação da distribuição Gama é como modelo de preço de ações.		XXX
13	Os estimadores por máxima verossimilhança são sempre não tendenciosos.		XXX

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

Problema 2 (35 pontos)

Considere os preços do litro das gasolinas aditivada e premium nos postos. Uma amostra de postos revela que os preços por litro da gasolina **aditivada** têm distribuição Normal com média R$ 2,50 e desvio padrão R$ 0,15. Os preços por litro da gasolina **premium** têm distribuição Normal com média 2,65 e desvio padrão R$ 0,20. A correlação entre os preços das duas gasolinas é 80%. Calcule as seguintes probabilidades:

a) De alguém pagar entre R$ 2,275 e R$ 2,725 por um litro de gasolina aditivada.
b) De alguém pagar entre R$ 2,275 e R$ 2,725 por um litro de gasolina aditivada sabendo que neste posto o litro da gasolina premium custa R$ 2,85.
c) De alguém pagar entre R$ $ 2,275 e R$ 2,725 por um litro de gasolina aditivada sabendo que neste posto o litro da gasolina premium custa R$ 2,55.
d) Qual é a distribuição condicional dos preços da gasolina premium sabendo que o preço praticado da gasolina aditivada é R$ 2,725 por litro?
e) Qual é a distribuição condicional dos preços da gasolina premium sabendo que o preço praticado da gasolina aditivada é R$ 2.275 por litro?
f) Quanto um posto deve cobrar por litro para que a sua gasolina aditivada esteja entre as 5% mais baratas?
g) Toma-se uma amostra de 9 postos. Qual a probabilidade do preço médio da gasolina aditivada na amostra exceder R$ 2.575?

Solução

X (o preço da gasolina aditivada) é N(2.50, $(0.15)^2$).
Y (o preço da gasolina premium) é N(2.65, $(0.20)^2$).

a) Pr(2.275 < X < 2.725) =

$$\Pr\left(\frac{2.275-2.5}{0.15} < \frac{X-2.5}{0.15} < \frac{2.725-2.5}{0.15}\right) = \Pr(-1.5 < Z < 1.5) = \Phi(1.5) - \Phi(-1.5) =$$

$$= 2.\Phi(1.5) - 1 = 2(0.9332) - 1 = 0.8664$$

b) Pr(2.275 < X < 2.725.90 < X < 2.10 | Y = 2.85) = ?
A distribuição condicional é:

$$N\left(\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y-\mu_y), \sigma_x^2(1-\rho^2)\right) = N\left(2.50 + 0.8\frac{0.15}{0.20}(2.85-2.65), (0.15)^2(1-0.64)\right) =$$

$$= N\left(2.62, (0.15)^2(0.6)^2\right) = N\left(2.62, (0.09)^2\right)$$

$$\Pr\left(\frac{2.275-2.62}{0.09}<\frac{X-2.62}{0.09}<\frac{2.725-2.62}{0.09}\right)=\Pr(-3.8333<Z<1.1667)=\Phi(1.1667)-\Phi(-3.8333)=$$
$$=0.8783$$

c) Pr(2.275 < X < 2.725.90 < X < 2.10 | Y = 2.55) = ?
A distribuição condicional é:

$$N\left(\mu_X+\rho\frac{\sigma_X}{\sigma_Y}(y-\mu_y),\sigma_x^2(1-\rho^2)\right)=N\left(2.50+0.8\frac{0.15}{0.20}(2.55-2.65),(0.15)^2(1-0.64)\right)=$$
$$=N\left(2.44,(0.15)^2(0.6)^2\right)=N\left(2.44,(0.09)^2\right)$$
$$\Pr\left(\frac{2.275-2.44}{0.09}<\frac{X-2.44}{0.09}<\frac{2.725-2.44}{0.09}\right)=\Pr(-1.8333<Z<3.1667)=\Phi(3.1667)-\Phi(-1.8333)=$$
$$=0.9659$$

d) A distribuição condicional dos dos preços da gasolina premium sabendo que o preço praticado da gasolina aditivada é R$ 2,725 por litro é Normal com média:

$$\mu_y+\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)=2.65+0.8\frac{0.20}{0.15}(2.725-2.5)=2.89$$

e variância $\left(\sigma_y^2(1-\rho^2)\right)=(0.20)^2(1-0.64)=(0.20)^2(0.36)=(0.20)^2(0.60)^2=(0.12)^2$

e) A distribuição condicional dos dos preços da gasolina premium sabendo que o preço praticado da gasolina aditivada é R$ 2,275 por litro é Normal com média:

$$\mu_y+\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)=2.65+0.8\frac{0.20}{0.15}(2.275-2.5)=2.41$$

e variância $\left(\sigma_y^2(1-\rho^2)\right)=(0.20)^2(1-0.64)=(0.20)^2(0.36)=(0.20)^2(0.60)^2=(0.12)^2$

f) Para estar entre as 5% mais baratas, o preço da gasolina aditivada deve ser:

$$\frac{X-2.50}{0.15}<-1.645\Rightarrow X<2.5+0.15(-1.645)=2.2533$$

g) Toma-se uma amostra de 9 postos. Qual a probabilidade do preço médio da gasolina aditivada na amostra exceder R$ 2.575?

O preço médio da gasolina aditivada é uma variável Normal com média 2.50 e variância $\frac{(0.15)^2}{9}$. Então:

$\dfrac{\overline{X}-2.50}{\frac{0.15}{3}}=\dfrac{\overline{X}-2.5}{0.05}$ é N(0,1). Logo:

$$\Pr(\overline{X}>2.575)=\Pr\left(Z>\frac{0.075}{0.05}\right)=\Pr(Z>1.5)=1-\Phi(1.5)=1-0.9332=0.0668$$

Problema 3 (25 pontos)

Sejam X_1 e X_2 variáveis aleatórias independentes, onde X_1 tem função de probabilidade Binomial (n_1, p) e X_2 tem função de probabilidade Binomial(n_2, p). Mostre, **usando o método que você preferir**, que Y = $X_1 + X_2$ tem função de probabilidade Binomial($n_1 + n_2$, p).

Dica:

Se X é Binomial(n, p) então E(X) = n.p, VAR(X) = n.p.q e a fgm é: $(pe^t + q)^n$.

Solução

A demonstração pela função geradora de momentos é trivial.

A fgm de Y é:

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

$E(e^{tX_1})E(e^{tX_2}) = (pe^t + q)^{n_1}(pe^t + q)^{n_2} = (pe^t + q)^{n_1+n_2}$, que é a função geradora de momentos de uma variável Binomial($n_1 + n_2$, p).

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: $Pr(X_{(n)} \le x) = Pr(\text{máx}(X_1, X_2, \ldots, X_n) \le x) =$
$= Pr(X_1 \le x, X_2 \le x, \ldots, X_n \le x)$

Solução

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

Logo, a verossimilhança é:

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

Então, para que a verossimilhança seja não nula, é necessário que $\theta > X_{(n)} = \text{máx}(X_1, X_2, \ldots, X_n)$ 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:

$\Pr(X_{(n)} \le x) = \Pr(X_1 \le x, X_2 \le x, \ldots, X_n \le x) = \{\Pr(X_1 \le x)\}^n$ 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.

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

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

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

O 2º. 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%		2,0100	97,78%
0,0200	50.80%		0,9900	83,89%		2,0125	97,79%
0,0300	51.20%		1,0000	84,13%		2.0200	97.83%
0,0400	51.60%		1,0100	84.38%		2.0300	97.88%
0,0500	51,99%		1.0167	84.54%		2.0400	97.93%
0,1000	53,98%		1.0250	84.73%		2.0412	97.94%
0,1500	55,96%		1,0500	85,31%		2.0500	97.98%
0,2000	57,93%		1,0553	85,44%		2,1000	98,21%
0,2236	58,85%		1,1000	86,43%		2,2000	98,61%
0,2500	59,87%		1,1180	86,82%		2,2361	98,73%
0,3000	61,79%		1,1475	87,44%		2,3000	98,93%
0,3015	61,85%		1,1500	87,49%		2,3263	99,00%
0,3333	63,06%		1,1553	87,60%		2,3333	99,02%
0,3475	63,59%		1,1667	87,83%		2,4000	99,18%
0,3492	63,65%		1,2000	88,49%		2,5000	99,38%
0,3500	63,68%		1,2060	88,61%		2,5500	99,46%
0,4000	65,54%		1,2200	88,88%		2,5628	99,48%
0.4167	66.16%		1,2500	89,44%		2,6000	99,53%
0,4307	66,67%		1.2700	89.79%		2,6500	99,60%
0,4500	67,36%		1,2816	90,00%		2,6667	99,62%
0,5000	69,15%		1,3000	90,32%		2,6833	99,64%
0,5500	70,88%		1,3333	90,88%		2,7000	99,65%
0,5774	71.81%		1,3750	91.54%		2.7500	99.70%
0,6000	72,57%		1,4000	91,92%		2,8000	99,74%
0,6250	73,40%		1,4468	92,60%		2,9000	99,81%
0,6500	74,22%		1,4500	92,65%		2,9500	99,84%
0,6667	74,75%		1,5000	93,32%		3,0000	99,87%
0,6708	74,88%		1,5500	93,94%		3,1000	99,90%
0,7000	75,80%		1,5811	94,31%		3,1500	99,92%
0,7500	77,34%		1,6000	94,52%		3,1667	99,92%
0,8000	78,81%		1,6450	95,00%		3,2000	99,93%
0,8333	79.77%		1,6667	95,22%		3,8333	99,99%
0,8500	80,23%		1,7000	95,54%			
0.8666	80.69%		1,8000	96,41%			
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%			
0,9650	83,27%		1,9800	97,61%			
0,9700	83,40%		1,9900	97.67%			
0,9750	83,52%		2,0000	97,72%			