IND 1115 – Inferência Estatística – Semestre 2007.01
Teste 4 – 03/07/2007
GABARITO

Escreva as respostas como frações ou 4 casas decimais.

PROBLEMA 1 (20 pontos)

Seja X uma variável aleatória com função de probabilidade Poisson(λ), isto é: $\Pr(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$ onde x = 0,1,2, ...

Encontre a função geradora de momentos de X e, a partir dela, mostre que E(X) = λ

Solução

$$M(t) = E\left(e^{tX}\right) = \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\left(\lambda e^t\right)^x}{x!} = e^{-\lambda}.\exp\left(\lambda e^t\right) = \exp\left(-\lambda + \lambda e^t\right)$$

A primeira derivada da função geradora de momentos é:

$$\frac{dM(t)}{dt} = \frac{d}{dt}\left\{e^{-\lambda}.\exp\left(\lambda e^t\right)\right\} = e^{-\lambda}\lambda e^t \exp\left(\lambda e^t\right)$$

$$\left.\frac{dM(t)}{dt}\right|_{t=0} = e^{-\lambda}\lambda e^0 \exp\left(\lambda e^0\right) = e^{-\lambda}.\lambda.e^{+\lambda} = \lambda = E(X)$$

Problema 2 (30 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$ 2,60** e desvio padrão **R$ 0,15**. Os preços por litro da gasolina **premium** têm distribuição Normal com média **2,75** 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$ 2,45 e R$ 2,75 por um litro de gasolina aditivada.
b) De alguém pagar entre R$ 2,45 e R$ 2,75 por um litro de gasolina aditivada sabendo que neste posto o litro da gasolina premium custa R$ 2,95.
c) De alguém pagar entre R$ $ 2,45 e R$ 2,75 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$ 2,70 por litro?
e) Quanto um posto deve cobrar por litro para que a sua gasolina aditivada esteja entre as 10% mais baratas?
f) Quanto deve cobrar um posto por litro para que sua gasolina premium esteja entre as 5% mais caras?

Solução

Sejam X o preço da gasolina aditivada e Y o preço da gasolina premium. Então X é N(2.60, 0.15^2) e Y é N(2.75, 0.20^2), e a correlação entre os dois preços é 0.8.

a) A probabilidade de um litro de gasolina aditivada custar entre 2.45 e 2.75 é:

$$\Pr(2.45 < X < 2.75) = \Pr\left(\frac{2.45-2.60}{0.15} < \frac{X-2.60}{0.15} < \frac{2.75-2.60}{0.15}\right) = \Pr(-1 < Z < 1) =$$
$$= \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 = 2(0.8413) - 1 = 0.6827$$

b) A probabilidade da gasolina aditivada custar entre R$ 2,45 e R$ 2,75 por litro sabendo que neste posto o litro da gasolina premium custa R$ 2,95 é obtida a partir da densidade condicional:

$$N\left(\mu_X + \frac{\rho\sigma_X}{\sigma_Y}(y-\mu_Y), \sigma_X^2(1-\rho^2)\right) =$$
$$= N\left(2.60 + \frac{0.8(0.15)}{0.20}(2.95-2.75), (0.15)^2(1-0.8^2)\right) =$$
$$= N\left(2.72, (0.15)^2(1-0.8^2)\right) = N\left(2.72, (0.09)^2\right)$$

$$\Pr(2.45 < X < 2.75 | Y = 2.95) = \Pr\left(\frac{2.45-2.72}{0.09} < \frac{X-2.72}{0.09} < \frac{2.75-2.72}{0.09}\right) = \Pr(-3 < Z < 0.3333) =$$
$$= \Phi(0.333) - \Phi(-3) = 0.6306 - 0.0013 = 0.6292$$

c) A probabilidade da gasolina aditivada custar entre R$ 2,45 e R$ 2,75 por litro sabendo que neste posto o litro da gasolina premium custa R$ 2,55 é obtida a partir da densidade condicional:

$$N\left(\mu_X + \frac{\rho\sigma_X}{\sigma_Y}(y-\mu_Y), \sigma_X^2(1-\rho^2)\right) =$$
$$= N\left(2.60 + \frac{0.8(0.15)}{0.20}(2.55-2.75), (0.15)^2(1-0.8^2)\right) =$$
$$= N\left(2.48, (0.15)^2(1-0.8^2)\right) = N\left(2.48, (0.09)^2\right)$$

$$\Pr(2.45 < X < 2.75 | Y = 2.55) = \Pr\left(\frac{2.45-2.48}{0.09} < \frac{X-2.48}{0.09} < \frac{2.75-2.48}{0.09}\right) = \Pr(-0.3333 < Z < 3) =$$
$$= \Phi(3) - \Phi(-0.3333) = 0.9987 - 0.3694 = 0.6292$$

d) A distribuição condicional dos preços da gasolina premium sabendo que o preço praticado da gasolina aditivada é R$ 2,70 por litro é:

$$N\left(\mu_Y + \frac{\rho\sigma_Y}{\sigma_X}(x-\mu_X), \sigma_Y^2\left(1-\rho^2\right)\right) =$$
$$= N\left(2.75 + \frac{0.8(0.20)}{0.15}(2.70-2.60), (0.20)^2\left(1-0.8^2\right)\right) =$$
$$= N\left(2.8567, (0.20)^2\left(1-0.8^2\right)\right) = N\left(2.8567, (0.12)^2\right)$$

e) Quanto um posto deve cobrar por litro para que a sua gasolina aditivada esteja entre as 10% mais baratas?

Em termos da distribuição N(0,1), o percentil 10% é o ponto tal que a probabilidade de estar abaixo dele é 10%, e é um número negativo. Este é o número que procuramos, mas a tabela da N(0,1) só fornece valores positivos, logo devemos usar a simetria da N(0,1). Por simetria, procure o ponto tal que a probabilidade de estar ACIMA dele é 10% (ou seja, $\Phi(z) = 90\%$ e tome o seu inverso. Logo, z = 1.2816 e o percentil 10% é -1.2816.

Em termos da distribuição dos preços da gasolina aditivada:

$$\frac{X-2.60}{0.15} = -1.2816 \Leftrightarrow X = 2.4078$$

Logo, se o litro da gasolina aditivada estiver abaixo de R$ 2.4078, o posto será um dos 10% mais baratos.

f) Quanto deve cobrar um posto por litro para que sua gasolina premium esteja entre as 5% mais caras?

O percentil 95% de uma N(0,1) é 1.645.
Transformando para a distribuição dos preços de gasolina premium temos:
$\frac{X-2.75}{0.20} = +1.6450 \Leftrightarrow X = 3.079$ e então se o preço da gasolina premium estiver acima deste valor, o posto é um dos 5% mais caros.

PROBLEMA 3 (20 pontos)

Considere uma amostra aleatória de tamanho n da distribuição Bernoulli(p).

a) Encontre o estimador de máxima verossimilhança de p.

b) Encontre a informação de Fisher.

c) O MLE é consistente? É não tendencioso?

Solução

a) A verossimilhança é:

$$L(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i} = p^{n\overline{X}}(1-p)^{n-n\overline{X}}$$

A log-verossimilhança é:

$$l(p) = \log(L(p)) = n\overline{X}\log(p) + (n - n\overline{X})\log(1-p)$$

Derivando a log-verossimilhança com relação a p leva a:

$$\frac{dl}{dp} = \frac{n\overline{X}}{p} - \frac{n - n\overline{X}}{1-p} = 0 \Leftrightarrow \frac{n\overline{X}}{p} = \frac{n - n\overline{X}}{1-p} \Leftrightarrow \frac{1-p}{p} = \frac{n - n\overline{X}}{n\overline{X}} = \frac{1-\overline{X}}{\overline{X}}$$

$$\Rightarrow \hat{p} = \overline{X}$$

é o estimador de máxima verossimilhança de p.

b) A Informação de Fisher sobre p é dada por:

$$I(p) = -E\left(\frac{d^2 l}{dp^2}\right) = -E\left(\frac{d}{dp}\left(\frac{n\overline{X}}{p} - \frac{n - n\overline{X}}{1-p}\right)\right) = -E\left(\frac{-n\overline{X}}{p^2} - \frac{n - n\overline{X}}{(1-p)^2}\right) = +\frac{1}{p^2}E(n\overline{X}) + \frac{1}{(1-p)^2}E(n - n\overline{X})$$

Mas:

$\sum X_i = n\overline{X}$ tem distribuição Binomial(n,p) e portanto sua média é n.p e sua variância é npq.

Logo, a média e variância de $\overline{X}$ são, respectivamente, p e pq/n.

$$I(p) = +\frac{1}{p^2}E(n\overline{X}) + \frac{1}{(1-p)^2}E(n - n\overline{X}) = \frac{np}{p^2} + \frac{n - np}{(1-p)^2} = \frac{n}{p} + \frac{n}{1-p} = n\left(\frac{1}{p} + \frac{1}{q}\right) = n\left(\frac{p+q}{p.q}\right) = \left(\frac{n}{p.q}\right) = \frac{1}{VAR(\overline{X})}$$

c) O MLE é consistente? É não tendencioso?

O estimador de máxima verossimilhança é não tendencioso para p, como mencionado acima.

O erro quadrático médio deste estimador é apenas a sua variância, que é p.q/n.

Como a variância tende a zero quando n tende a infinito, o MLE é consistente.

Problema 4 (30 pontos)

Seja X_1 , X_2 , ... , X_n uma amostra aleatória da densidade N(0,θ).

a) Mostre que o estimador de máxima verossimilhança de θ é $T = \frac{1}{n}\sum X_i^2$

b) Mostre que T é não tendencioso para θ e tem variância $2\theta^2/n$. Dica: Calcule $E(X_i^4)$.

Solução

a) A verossimilhança é:

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

A log-verossimilhança é:

$$l(\theta) = \frac{-n}{2}\log(2\pi\theta) + \frac{-1}{2\theta}\sum_{i=1}^{n} X_i^2$$

A primeira derivada da log-verossimilhança é:

$$\frac{dl(\theta)}{d\theta} = \frac{-n}{2}.\frac{1}{\theta} + \frac{1}{2\theta^2}\sum_{i=1}^{n} X_i^2 = 0 \Leftrightarrow \hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

b) Note que:

$$\frac{X_i}{\sqrt{\theta}} \sim N(0,1) \Leftrightarrow \frac{X_i^{\ 2}}{\theta} \sim \chi_1^2 \Rightarrow \frac{\sum_{i=1}^{n} X_i^{\ 2}}{\theta} \sim \chi_n^2 \Rightarrow E\left(\sum_{i=1}^{n} X_i^{\ 2}\right) = n\theta \ \text{ e } \ VAR\left(\sum_{i=1}^{n} X_i^{\ 2}\right) = 2n\theta^2$$

Logo:

$E(\hat{\theta}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i^{\ 2}\right) = \theta$ e $VAR(\hat{\theta}) = VAR\left(\frac{\sum_{i=1}^{n} X_i^{\ 2}}{n}\right) = \frac{2n\theta^2}{n^2} = \frac{2\theta^2}{n}$ o que prova que o MLE é não tendencioso e consistente.

Tabela – Função de Distribuição N(0,1)

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