IND 1115 – Inferência Estatística – Semestre 2005.02
Teste 3 – 01/12/2005
GABARITO

PROBLEMA 1 (20 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) = σ2 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 2 (20 pontos)

Considere uma amostra aleatória de tamanho 16 da distribuição Poisson com média θ.

Suponha que a densidade a priori é uma Gama(a = 4, b = 2).

a) Qual a densidade a posteriori?
b) Suponha que observamos $\overline{X} = 2$. Escreva a média a posteriori como uma média ponderada deste estimador e da média da priori.
c) Repita o item b) supondo agora que a priori é uma Gama(a = 12, b = 6).

Solução

a) A função de probabilidade de cada X_i é: $f(x) = \frac{e^{-\theta}\theta^x}{x!}$ para x = 0,1,2, ...

A verossimilhança é:

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

A priori é Gama(4,2):

$$\pi(\theta)=k.\theta^{a-1}e^{-b\theta}\ =k.\theta^{4-1}e^{-2\theta}$$

A posteriori é proporcional ao produto da verossimilhança e da priori, isto é:

$$\pi(\theta|x)=k.\theta^{a+\sum_i X_i-1}e^{-b\theta}e^{-n\theta}\ =k.\theta^{4+\sum_i X_i-1}e^{-(2+n)\theta}$$

Ou seja, a posteriori é Gama$\left(4+\sum X_i, 2+n\right)$.

b) O tamanho da amostra é 16 e a média amostral é 2.

A média a posteriori é:

$$\mu_p\ =\frac{4+\sum X_i}{2+n}=\frac{4+n\overline{X}}{2+n}=\frac{4}{2+n}+\frac{n\overline{X}}{2+n}=\frac{2}{2+n}\left(\frac{4}{2}\right)+\frac{n}{2+n}\left(\overline{X}\right)=\frac{2}{2+16}\left(\frac{4}{2}\right)+\frac{16}{2+16}(2)=\frac{36}{18}=2$$

O peso da média a priori é 2/(2+n), enquanto o peso do estimador de máxima verossimilhança de θ é n/(2+n).

c) A posteriori é:

$$\pi(\theta|x)=k.\theta^{a+\sum_i X_i-1}e^{-b\theta}e^{-n\theta}\ =k.\theta^{12+\sum_i X_i-1}e^{-(6+n)\theta}$$

A média a posteriori é:

$$\mu_p\ =\frac{12+\sum X_i}{6+n}=\frac{12+n\overline{X}}{6+n}=\frac{12}{6+n}+\frac{n\overline{X}}{6+n}=\frac{6}{6+n}\left(\frac{12}{6}\right)+\frac{n}{6+n}\left(\overline{X}\right)=\frac{6}{6+16}\left(\frac{12}{6}\right)+\frac{16}{6+16}(2)=\frac{44}{22}=2$$

Neste caso a média se mantém inalterada pois os valores da média a priori e do MLE coincidem.

Problema 3 (20 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?

Solução

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

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

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

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

Problema 4 (20 pontos)

Seja X uma variável Qui-quadrado com 50 graus de liberdade. Use o teorema central do limite para aproximar as seguintes probabilidades:

a) Pr(X < 55)
b) Pr(45 < X < 55)

Solução

a) X é uma variável aleatória com média 50 e variância 100. Pelo teorema central do limite:

$\dfrac{X-50}{\sqrt{100}}=\dfrac{X-50}{10}$ é aproximadamente N(0,1).

Daí:

$$\Pr\left(X<55\right)=\Pr\left(\frac{X-50}{10}<\frac{55-50}{10}\right)\approx\Phi\left(\frac{55-50}{10}\right)=\Phi\left(0.5\right)=0.6915$$

b) $$\Pr\left(45<X<55\right)=\Pr\left(\frac{45-50}{10}<\frac{X-50}{10}<\frac{55-50}{10}\right)\approx\Phi\left(\frac{55-50}{10}\right)-\Phi\left(\frac{45-50}{10}\right)=$$

$$=\Phi\left(0.5\right)-\Phi\left(0.5\right)=2(0.6915)-1=0.3829$$

Problema 5 (20 pontos)

Um apartamento de 2 quartos numa certa região da cidade custa, em média R$ 260 mil. A dispersão entre os valores, medida pelo desvio padrão, é R$ 100 mil. Além disso, pode-se encarar os preços dos apartamentos como independentes entre si e Normalmente distribuídos.

a) Uma imobiliária pretende oferecer uma viagem de presente aos compradores de apartamentos de 2 quartos neste bairro que comprem os apartamentos situados na faixa dos 10% mais caros. A partir de quanto deve custar o seu apartamento para que você ganhe a viagem de "presente"?

b) Considere 16 compradores de apartamentos de 2 quartos neste bairro. Qual a probabilidade do preço médio pago por eles ser inferior a R$ 300 mil?

c) Dentre as 16 pessoas nesta mesma amostra, qual a probabilidade do comprador que pagou mais caro por um apartamento ter pago menos de R$ 285 mil?

d) Dentre as 16 pessoas nesta mesma amostra, qual a probabilidade do comprador do apartamento mais barato ter pago mais de R$ 235 mil?

Solução

a) O ponto da distribuição N(0,1) tal que a probabilidade de estar abaixo dele é 0.90 é, pela tabela, z = 1.2816.

Seja X o preço de um apartamento de dois quartos nesta região. Então X é Normal com média R$ 260 mil e desvio padrão R$ 100 mil.

Logo:

$$Z = \frac{X - 260000}{100000} \sim N(0,1)$$

Pr(Z < 1.2816) = 0.90 e você ganha a viagem de "presente" se:

$$Z = \frac{X - 260000}{100000} \geq 1.2816 \Leftrightarrow X \geq R\$388160,$$

a) Seja $\overline{X}$ o preço médio dos apartamentos de 2 quartos numa amostra de n compradores. Neste caso, n = 16 compradores. Sabemos que $\overline{X}$ é Normal com média R$ 260 mil e variância (100 mil)^2/16, isto é, desvio padrão (100 mil)/4 = R$ 25 mil.

$$\Pr\left(\overline{X} < 300\right) = \Pr\left(\frac{\overline{X} - 260}{25} < \frac{300 - 260}{25}\right) = \Pr\left(Z < \frac{40}{25}\right) = \Phi(1.60) = 0.9452$$

c) Seja V o comprador que pagou mais caro. Então se V < 285, todos os 16 compradores pagaram menos de R$ 285 mil.

Pr(V < 285) = Pr(X_1 < 285, X_2 < 285, ..., X_{16} < 285).

Como os X_i ´s são independentes, esta probabilidade conjunta é o produto das probabilidade dos X_i´s individuais. Além disso, os X_i´s formam uma amostra, e portanto são identicamente distribuídos, o que torna $\Pr(X_1 < 285) = \Pr(X_2 < 285) = \Pr(X_{16} < 285)$.

Então:

$\Pr(V < 285) = \{\Pr(X_1 < 285)\}^{16}$

Mas:

$$\Pr(X_1 < 285) = \Pr\left(\frac{X_1 - 260}{100} < \frac{285 - 260}{100}\right) = \Pr(Z < 0.25) = \Phi(0.25) = 0.5987$$

Daí:

$\Pr(V < 285) = \{\Pr(X_1 < 285)\}^{16} = (0.5987)^{16} = 2.7248\ 10^{-4}$

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%			