IND 1115 – Inferência Estatística – Semestre 2004.01
Teste 1 – 29/04/2004
Nome: GABARITO

Problema 1 (30 pontos)

Seja $X_1, X_2, \dots, X_n$ 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:
$X_{(n)} = \text{máx}(X_1, X_2, \dots, X_n)$

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 são:
1,53 0,40 2,39 3,60 3,54
3,83 0,06 1,63 3,45 0,55
Compare o MLE e o estimador por método dos momentos de θ baseados nesta amostra.

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

a) A verossimilhança é:

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} = \frac{1}{\theta^n} \quad \text{se } 0 \le \text{todo } x_i \le \theta$$

Ou seja, a verossimilhança será diferente de zero apenas quando: $0 \le x_{(1)} \le x_{(2)} \le \dots \le x_{(n)} \le \theta$

Em particular, esta última condição implica em:

$\theta \geq x_{(n)}$ para que $L(\theta) > 0$

Como a verossimilhança é decrescente como função de θ, o MLE é obtido no menor valor possível de θ, isto é, em $\theta = x_{(n)}$

b) O estimador por método de momentos de θ é:

$\bar{X} = E(X_i) = \frac{\theta}{2} \Rightarrow$ $\tilde{\theta} = 2\bar{X}$ é o estimador por método de momentos de θ.

c) Para responder sobre a consistência do MLE, é preciso olhar para o limite de seu erro quadrático médio quando n, o tamanho da amostra, tende a infinito.

Seja $\hat{\theta} = X_{(n)}$ o MLE de θ.

Precisamos encontrar a densidade de $\hat{\theta}$.

Usando a "dica":

$$P_r(X_{(n)} \le x) = P_r(X_1 \le x, X_2 \le x, \dots X_n \le x)$$

$$= \prod_{i=1}^{n} P_r(X_i \le x) = \left[P_r(X_1 \le x)\right]^n$$

independência

Mas: $P_r(X_1 \le x) = \int_0^x \frac{1}{\theta}\, dt = \frac{x}{\theta}$ se $0 \le x \le \theta$

Daí:

$$P_r(X_{(n)} \le x) = \left(\frac{x}{\theta}\right)^n \quad \text{se } 0 \le x \le \theta$$

A densidade de $X_{(n)}$ é apenas a derivada da função de distribuição, ou seja:

Problema 1 - continuação

$$g(x) = \frac{n x^{n-1}}{\theta^n} \quad \text{se} \quad 0 \le x \le \theta$$

A média de $X_{(n)}$ é:

$$E\{X_{(n)}\} = \int_0^\theta x \frac{n x^{n-1}}{\theta^n} dx = \frac{1}{\theta^n} \int_0^\theta n x^n dx$$

$$= \frac{1}{\theta^n} \left(\frac{n x^{n+1}}{(n+1)} \right) \Bigg|_0^\theta = \frac{n}{n+1} \theta$$

Logo, $X_{(n)}$ é um estimador tendencioso de θ.

O 2º momento de $X_{(n)}$ é:

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

$$= \frac{n}{\theta^n} \left(\frac{x^{n+2}}{n+2} \right) \Bigg|_0^\theta = \frac{n}{n+2} \theta^2$$

A variância de $X_{(n)}$ é então:

$$\text{VAR}(X_{(n)}) = E\{X_{(n)}^2\} - \{E(X_{(n)})\}^2$$

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

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

Problema 1 - continuação

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

$$VAR(X_{(n)}) = \theta^2 \left\{ \frac{n(n^2+2n+1) - n^3 - 2n^2}{(n+2)(n+1)^2} \right\}$$

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

A tendência de $X_{(n)}$ é:

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

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

$$MSE(X_{(n)}) = VAR(X_{(n)}) + \left[BIAS(X_{(n)})\right]^2$$

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

O limite quando $n \to \infty$ do MSE é zero e portanto $X_{(n)}$ é um <u>estimador consistente</u> de θ.

e) Coleta-se uma amostra de tamanho 10.

O MLE é, por inspeção, $X_{(10)} = 3.83$

O estimador por método de momentos é $2\overline{X}$.

Problema 1 - continuação

Mas : $\bar{X} = 2.098$ e então $\tilde{\theta} = 2\bar{X} = 4.196$

Problema 2 (15 pontos)
Provas ligeiramente diferentes foram aplicadas a duas turmas da mesma disciplina. Os resultados são mostrados a seguir:

	Turma 1	Turma 2
Nota média	75	80
Desvio padrão	15	25
Número de alunos	10	8

a) Encontre intervalos de confiança 90% e 95% para a diferença $\mu_1 - \mu_2$ onde μ_1 e μ_2 são, respectivamente, as médias das turmas 1 e 2. Suponha que as notas nas duas turmas são Normais e as duas amostras (turmas) são independentes.

b) Que hipótese adicional (além das indicadas no item a)) está implícita na obtenção dos IC do item a)?

b) Precisamos supor que as variâncias das 2 amostras são iguais para aplicar o ~~teste~~ intervalo do item a).

a) Turma 1 : $X_1, X_2, \dots X_{10} \overset{iid}{\sim} N(\mu_1, \sigma^2)$
Turma 2 : $Y_1, Y_2, \dots Y_8 \overset{iid}{\sim} N(\mu_2, \sigma^2)$
e as 2 amostras são indep.

s_1^2 = variância amostral das notas da turma 1
s_2^2 = " " " " " turma 2

$$s_1^2 = \frac{1}{m}\sum_{i=1}^{m}(X_i - \bar{X})^2 = \frac{1}{10}\sum_{i=1}^{10}(X_i - 75)^2$$

$s_1 = \sqrt{s_1^2} = 15$ (dado)

Analogamente, $\bar{Y} = 80$, $s_2^2 = (25)^2$, $n = 8$

Problema 2 - continuação

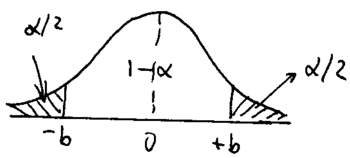
$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right)\left[\frac{m S_1^2 + n S_2^2}{m+n-2}\right]}} \sim t_{m+n-2}$$

Seja $R = \sqrt{\left(\frac{1}{m} + \frac{1}{n}\right)\left(\frac{m S_1^2 + n S_2^2}{m+n-2}\right)}$

Um intervalo de confiança $100(1-\alpha)\%$ é:

$$\left[(\bar{X} - \bar{Y}) - bR \;,\; (\bar{X} - \bar{Y}) + bR\right]$$

Onde $\underline{b}$ é encontrado a partir da distribuição t_{m+n-2} tal que:



Aqui: $R = \sqrt{\left(\frac{1}{10} + \frac{1}{8}\right)\left(\frac{10(15)^2 + 8(25)^2}{16}\right)} = \sqrt{101.953}$

$$\boxed{R = 10.097}$$

Se $1-\alpha = 90\% \Rightarrow \alpha/2 = 5\% \Rightarrow b = 1.746$
$1-\alpha = 95\% \Rightarrow \alpha/2 = 2.5\% \Rightarrow b = 2.120$

da tabela t com 16 graus de liberdade

O IC 90% para $\mu_1 - \mu_2$ é:

$$\left[(75-80) - (1.746)(10.097) \;,\; (75-80) + (1.746)(10.097)\right]$$

Problema 2 - continuação

Ou seja, o IC 90% é: $[-22.629, +12,629]$

O IC 95% é:

$$\left[-5-(2.120)(10.097),\ -5+(2.120)(10.097)\right] =$$

$$= \left[-26.405,\ 16.405\right]$$

Problema 3 (30 pontos)

Considere uma amostra aleatória de tamanho n da distribuição Geométrica(p), ou seja, a função de probabilidade é dada por:

$$\Pr(X = x) = f(x) = q^{x-1}p = (1-p)^{x-1}p \quad \text{para } x = 1,2,3.....$$

a) Encontre o estimador de máxima verossimilhança de p.
b) Encontre o MLE de Pr(X > 2).
c) Encontre a informação de Fisher.
~~d) Ache o limite inferior de Cramer-Rao para um estimador de p.~~

Dica: Série Geométrica

$$\sum_{x=1}^{\infty} a^x = \frac{a}{1-a} \quad \text{se} |a| < 1$$

a) $$L(p) = \prod_{i=1}^{n} (1-p)^{x_i-1} p = p^n (1-p)^{-n} (1-p)^{\Sigma x_i}$$

$$L(p) = p^n (1-p)^{n\bar{X}-n}$$

$$\ell(p) = \log L(p) = n \log p + (n\bar{X}-n)\log(1-p)$$

$$\frac{d\ell}{dp} = \frac{n}{p} + \frac{(n\bar{X}-n)(-1)}{(1-p)} = 0$$

$$\Rightarrow \frac{n}{p} = \frac{n\bar{X}-n}{1-p}$$

Problema 3- continuação

$$\frac{1}{p} = \frac{\bar{X}-1}{1-p}$$

$$\Rightarrow \quad 1-\cancel{p} = p\bar{X}-\cancel{p} \implies 1 = p\bar{X} \implies \hat{p} = \frac{1}{\bar{X}}$$

é o MLE de p.

Ou seja: $\hat{p} = \dfrac{1}{\bar{X}} = \dfrac{n}{\sum X_i}$

b) $Pr(X>2) = Pr(X \geq 3) = \sum_{x=3}^{\infty} q^{x-1} p$

$$= \frac{p}{q} \sum_{x=3}^{\infty} q^x = \frac{p}{q}\left\{ q^3 + q^4 + q^5 + \dots \right\}$$

$$= \frac{p}{q}\left\{ \sum_{x=1}^{\infty} q^x - q - q^2 \right\} = \frac{p}{q}\left\{ \frac{q}{1-q} - q - q^2 \right\}$$

$$= \frac{p}{q}\left\{ \frac{q}{p} - q - q^2 \right\} = 1 - p - p^2$$

Pelo princípio da invariância do MLE, o estimador de máx. verossimilhança de $Pr(X>2)$ é:

$$1 - \hat{p} - \hat{p}^2 = 1 - \frac{1}{\bar{X}} - \frac{1}{\bar{X}^2}$$

c) $\dfrac{d^2\ell}{dp^2} = \dfrac{d}{dp}\left\{ n p^{-1} + (-1)(n\bar{X}-n)(1-p)^{-1} \right\}$

$$= n(-1)p^{-2} + (-1)(n\bar{X}-n)(-1)(1-p)^{-2}(-1)$$

Problema 3 - continuação

$$\frac{d^2\ell}{dp^2} = -np^{-2} - (n\bar{x}-n)(1-p)^{-2}$$

$$\frac{d^2\ell}{dp^2} = -\frac{n}{p^2} - \frac{(n\bar{X}-n)}{(1-p)^2}$$

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

Mas:

$$X_i \overset{iid}{\sim} geom(p) \Rightarrow E(X_i) = 1/p$$

$$\Rightarrow E(\bar{X}) = E(X_i) = 1/p$$

Logo:

$$E(n\bar{X}-n) = \frac{n}{p} - n$$

$$I(p) = \frac{n}{p^2} + \frac{1}{q^2}\left[\frac{n}{p} - n\right] = \frac{n}{p^2} + \frac{1}{pq^2}(n - np)$$

$$I(p) = \frac{n}{p^2} + \frac{n}{pq^2}(1-p) = \frac{n}{p^2} + \frac{nq}{pq^2}$$

$$I(p) = \frac{n}{p^2} + \frac{n}{pq}$$

$$I(p) = \frac{n}{p}\left\{\frac{1}{p} + \frac{1}{q}\right\}$$

10

IND 1115 – Inferência Estatística – Semestre 2004.01
Teste 1 – 29/04/2004
Nome: GABARITO

Problema 4 (10 pontos)

Aproxime, com base no Teorema Central do Limite, as seguintes probabilidades:

$$a)\ \Pr\left(\chi^2_{20} \leq 22\right)$$
$$b)\ \Pr\left(\chi^2_{30} > 36\right)$$

Se $U \sim \chi^2_n$ então: $E(U) = n$ e $VAR(U) = 2n$

Além disso, U pode ser pensada como a soma de $\underline{n}$ v.a. independentes com densidade χ^2_1 e portanto, se n for "grande" pode-se aproximar probabilidades envolvendo U através do teorema central do limite.

a) $$\Pr\left(\chi^2_{20} \leq 22\right) = \Pr\left\{\frac{\chi^2_{20} - 20}{\sqrt{40}} \leq \frac{22-20}{\sqrt{40}}\right\}$$

$$\approx \Phi\left(\frac{22-20}{\sqrt{40}}\right) = \Phi(0.316) = \boxed{0.624}$$

b) $$\Pr\left(\chi^2_{30} > 36\right) = 1 - \Pr\left(\chi^2_{30} \leq 36\right) = 1 - \Pr\left\{\frac{\chi^2_{30} - 30}{\sqrt{60}} \leq \frac{36-30}{\sqrt{60}}\right\}$$

$$\approx 1 - \Phi\left(\frac{36-30}{\sqrt{60}}\right) = 1 - \Phi(0.775) = 1 - 0.781$$

$$= \boxed{0.219}$$