

IND 1115 – Inferência Estatística – Semestre 2005.02
Teste 2 – 22/11/2005
GABARITO

PROBLEMA 1 (20 pontos)

Considere uma variável Binomial com $n = 32$ e $p = \frac{1}{2}$.

a) Calcule exatamente:

x	Pr(X = x)
15	0.1317
16	0.1399
17	0.1317
Pr(15 ≤ X ≤ 17)	0.4034

b) Calcule usando o Teorema de DeMoivre e Laplace **com** correção de continuidade.

x	Pr(X = x)
15	0.1319
16	0.1403
17	0.1319
Pr(15 ≤ X ≤ 17)	0.4041

PROBLEMA 2 (25 pontos)

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

- Encontre o estimador de máxima verossimilhança de p .
- Encontre a informação de Fisher.
- O MLE é consistente? É não tendencioso?
- Ache o limite inferior de Cramer e Rao para um estimador de p .

Solução

a) A verossimilhança é:

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

A log-verossimilhança é:

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

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

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

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

é o estimador de máxima verossimilhança de p .

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

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

Mas:

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

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

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

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 $p \cdot q/n$.

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

d) O limite inferior de Cramér e Rao para p é:

$$CRLB = \frac{1}{I(p)} = \frac{pq}{n} = VAR(\bar{X})$$

Problema 3 (10 pontos)

A mesma prova foi aplicada em duas turmas, com os resultados descritos a seguir.

	Turma A	Turma B
Média	64	69
Desvio Padrão	15	20
Número de Alunos	40	32

Encontre um intervalo de confiança 95% para a diferença das médias $\mu_B - \mu_A$. Use uma aproximação Normal, já que o número de graus de liberdade da distribuição t é grande.

Solução

O resultado fundamental aqui é:

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

Aqui podemos supor que a distribuição é Normal, pois se trata de uma t com $40 + 32 - 2 = 70$ graus de liberdade.

A variância amostral combinada é:

$$S_p^2 = \left(\frac{(m-1)S_1^2 + (n-1)S_2^2}{n+m-2}\right) = \frac{39(15)^2 + 31(20)^2}{70} = 302.50 \Rightarrow S_p = \sqrt{302.5} = 17.3925$$

O IC 95% para $\mu_B - \mu_A$ é:

$$\begin{aligned} \bar{Y} - \bar{X} \pm 1.96 \sqrt{\left(\frac{1}{n} + \frac{1}{m}\right) \left(\frac{(m-1)S_1^2 + (n-1)S_2^2}{n+m-2}\right)} &= (69 - 64) \pm 1.96 \sqrt{\left(\frac{1}{40} + \frac{1}{32}\right) 302.5} = 5 \pm 8.0850 = \\ &= [-3.085, 13.085] \end{aligned}$$

Problema 4 (20 pontos)

Considere uma amostra aleatória de tamanho 12 da distribuição Bernoulli(p).

Suponha que a densidade a priori é uma Beta(4,4).

- Qual a densidade a posteriori?
- Suponha que observamos $\bar{X} = \hat{p} = 0.6$. Escreva a média a posteriori como uma média ponderada deste estimador e da média da priori.
- Repita o item b) supondo agora que a priori é uma Beta(8,8).

Solução

a) A verossimilhança tem a forma:

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

Aqui $n = 12$.

A priori tem a forma:

$$\pi(p) = k \cdot p^{4-1} (1-p)^{4-1}$$

A densidade a posteriori é:

$$\pi(p|x) = k \cdot p^{4-1+12\bar{X}} (1-p)^{4-1+12-12\bar{X}} = k \cdot p^{12\bar{X}+4-1} (1-p)^{16-12\bar{X}-1}$$

Logo, a posteriori é uma beta com parâmetros: $12\bar{X} + 4$ e $16 - 12\bar{X}$

b) A média das posteriori é:

$$\mu_p = \frac{12\bar{X} + 4}{12\bar{X} + 4 + 16 - 12\bar{X}} = \frac{12\bar{X} + 4}{20} = \frac{12}{20}\bar{X} + \frac{8}{20}\left(\frac{4}{8}\right)$$

onde $4/8$ é a média da priori.

Como $\bar{X} = \hat{p} = 0.6$, a média da posteriori será:

$$\mu_p = \frac{12\bar{X} + 4}{12\bar{X} + 4 + 16 - 12\bar{X}} = \frac{12\bar{X} + 4}{20} = \frac{12}{20}(0.6) + \frac{8}{20}(0.5) = \frac{11.2}{20} = 0.56$$

c) Agora a priori é uma Beta(8,8), e então sua média continua sendo $\frac{1}{2}$. A posteriori é:

$$\pi(p|x) = k \cdot p^{8-1+12\bar{X}} (1-p)^{8-1+12-12\bar{X}} = k \cdot p^{12\bar{X}+8-1} (1-p)^{20-12\bar{X}-1}$$

Logo, a posteriori é uma beta com parâmetros: $12\bar{X} + 8$ e $20 - 12\bar{X}$

A média das posteriori é:

$$\mu_p = \frac{12\bar{X} + 8}{12\bar{X} + 8 + 20 - 12\bar{X}} = \frac{12\bar{X} + 8}{28} = \frac{12}{28}\bar{X} + \frac{16}{28}\left(\frac{8}{16}\right)$$

Ao observarmos: $\bar{X} = \hat{p} = 0.6$, a média da posteriori torna-se:

$$\mu_p = \frac{12\bar{X} + 8}{28} = \frac{15.2}{28} = 0.5429 \text{ Ou seja, é mais próxima da média da priori que no caso anterior,}$$

indicando que a priori Beta(8,8) é mais "informativa" que a priori Beta(4,4).

Problema 5 (25 pontos)

Seja X_1, X_2, \dots, X_n uma amostra aleatória da densidade $N(0, \theta)$.

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:

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

b) Note que:

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

Logo:

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

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%		2.0100	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%
0.1768	57.02%		1.0553	85.44%		2.1000	98.21%
0.2000	57.93%		1.0800	85.99%		2.1500	98.42%
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%		1.2000	88.49%		2.4000	99.18%
0.3492	63.65%		1.2060	88.61%		2.5000	99.38%
0.3500	63.68%		1.2200	88.88%		2.5500	99.46%
0.4000	65.54%		1.2500	89.44%		2.5628	99.48%
0.4167	66.16%		1.2700	89.79%		2.6000	99.53%
0.4307	66.67%		1.2816	90.00%		2.6500	99.60%
0.4500	67.36%		1.3000	90.32%		2.6667	99.62%
0.5000	69.15%		1.3333	90.88%		2.6750	99.63%
0.5303	70.21%		1.3500	91.15%		2.6833	99.64%
0.5500	70.88%		1.3750	91.54%		2.7000	99.65%
0.5774	71.81%		1.4000	91.92%		2.7500	99.70%
0.6000	72.57%		1.4468	92.60%		2.8000	99.74%
0.6250	73.40%		1.4500	92.65%		2.9000	99.81%
0.6500	74.22%		1.5000	93.32%		2.9500	99.84%
0.6667	74.75%		1.5500	93.94%		3.0000	99.87%
0.6708	74.88%		1.5811	94.31%		3.1000	99.90%
0.7000	75.80%		1.6000	94.52%		3.1500	99.92%
0.7500	77.34%		1.6450	95.00%		3.2000	99.93%
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%			
0.9700	83.40%		2.0000	97.72%			