DEPARTAMENTO DE MATEMÁTICA DA UNIVERSIDADE DE COIMBRA

Exame de Estatística

Duração: 2h 30m 20-06-03

Observação: A resolução completa das questões inclui a justificação do raciocínio utilizado e a apresentação dos cálculos efectuados.

I. Considere um espaço de probabilidade $(\Omega, \mathcal{A}, P)$ e seja A um acontecimento em $\mathcal{A}$ de probabilidade p ($p \in]0,1[$). Sejam X a variável aleatória real (v.a.r.) $\mathbb{1}_A$ e $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória de tamanho n ($n \in \mathbb{N}$) de X.

1. Indique o modelo estatístico associado a $(X_1, X_2, \ldots, X_n)$.

2. Prove que a média da amostra, $\overline{X}_n$, é um estimador da máxima verosimilhança do valor médio de X.

3. Deduza um estimador da variância de X, convergente em probabilidade, que seja apenas função de $\overline{X}_n$.

4. Obtenha a lei limite da v.a.r.

$$Z_n = \sqrt{n}\frac{\overline{X}_n - p}{\sqrt{\overline{X}_n\left(1 - \overline{X}_n\right)}}.$$

5. A fim de efectuar um controlo *anti-doping* no final de uma maratona internacional, foram efectuados testes a 200 atletas, seleccionados aleatoriamente entre os que terminaram a prova, para averiguar se tinham tomado esteróides. Sabe-se que 24 dos testes realizados deram resultado positivo. Com base na amostra observada, obtenha um intervalo real que contenha, com uma confiança de 98%, a proporção de atletas que realizaram a maratona sob o efeito de esteróides.

II. Seja $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória de tamanho n ($n \in \mathbb{N}$) de uma v.a.r. X, definida sobre um espaço de probabilidade $(\Omega, \mathcal{A}, P)$, de lei de probabilidade Q_θ absolutamente contínua de densidade f_θ, com $\theta \in \Theta = \{\theta_0, \theta_1\}$.

1. Prove que $P\left(\prod_{i=1}^{n} f_{\theta_0}(X_i) > 0\right) = 1$.

2. Considere a v.a.r. $Y = \prod_{i=1}^{n} \frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)}$ e a função $u(t) = P(Y > t)$. Mostre que

$$\forall y \in]0,1[,\ \exists t_0 \geq 0 : u(t_0) = y.$$

3. Deduza das alíneas anteriores que, dado $\alpha \in]0,1[$, é possível determinar $k(\alpha)$ de modo que o teste de região crítica

$$\mathcal{C} = \left\{(x_1, x_2, ..., x_n) \in \mathbb{R}^n : \prod_{i=1}^{n} f_{\theta_1}(x_i) > k(\alpha) \prod_{i=1}^{n} f_{\theta_0}(x_i)\right\}$$

seja de nível de significância α para testar $H_0 : \theta = \theta_0$ contra $H_1 : \theta = \theta_1$.

v.s.f.f.

4. Suponha agora que X descreve o rendimento relativo aos créditos concedidos diariamente por determinada agência bancária. Sabe-se que X segue a lei de Pareto de parâmetro β ($\beta > 1$), isto é a lei de densidade

$$f_\beta(x) = \frac{\beta}{x^{\beta+1}} \mathbb{I}_{]1,+\infty[}(x), \; x \in \mathbb{R},$$

e valor médio $E_\beta(X) = \frac{\beta}{\beta-1}$.

a) Mostre que os logaritmos de tais rendimentos diários são bem descritos por uma v.a.r. seguindo a lei exponencial de parâmetro β.

b) Considere a estatística $T_n = \frac{1}{n}\sum_{i=1}^{n} \log X_i$. Obtenha uma função de β da qual T_n seja um estimador cêntrico e convergente.

c) Estudos efectuados sobre a população levaram à conclusão de que o valor do rendimento médio relativo aos referidos créditos corresponde a $\beta = 1.5$. Construa um teste que permita decidir, com base numa qualquer amostra de tamanho n, entre as hipóteses

$$H_0 : \beta = 1.5 \quad \text{contra} \quad H_1 : \beta < 1.5$$

ao nível de significância 0.05.

d) Sabendo que foram registados aleatoriamente, ao longo de 6 meses, 100 valores daquele rendimento, $x_1, ..., x_{100}$, para os quais $\sum_{i=1}^{100} \log x_i = 88$ u.m. (unidades monetárias), que pode concluir, com base no teste anterior, sobre o rendimento médio dos créditos concedidos diariamente por tal agência bancária?

Cotação

I 9.0 valores
II 11.0 valores