LISBOA SCHOOL OF ECONOMICS & MANAGEMENT	**Estatística** **Licenciatura MAEG**	2º semestre – 2018/2019 EN Duração: 2h10

Todas as respostas devem ser dadas na folha de exame e detalhadamente justificadas.

BOA SORTE!

PARTE I

1. Uma equipa de médicos considera que o tempo, X, em minutos, até um determinado medicamento para a dor de cabeça fazer efeito segue uma distribuição de Weibull invertida, $WI(\theta,\tau)$, ou seja

$$F(x|\theta,\tau) = e^{-(\theta/x)^\tau} \qquad \text{e} \qquad f(x|\theta,\tau) = \tau\theta^\tau x^{-(\tau+1)} e^{-(\theta/x)^\tau}, \quad x>0, \quad \theta,\tau>0.$$

(a) Considere uma amostra casual extraída da população X. Mostre que a distribuição do máximo amostral também tem distribuição Weibull invertida, indicando os seus parâmetros. (1.0)

R: $F_n(x) = P(X_{(n)} \leqslant x) = [F(x)]^n = e^{-\left(\frac{n^{1/\tau}\theta}{x}\right)^\tau} \sim WI(n^{1/\tau}\theta,\tau)$

(b) Mostre que $X^{-\tau}$ tem distribuição exponencial de média $\theta^{-\tau}$ e deduza a quantidade de informação de Fisher para θ contida numa amostra casual de dimensão n recolhida da população X quando **τ é conhecido**. (2.0)

R: $I_{X_1,\dots,X_n}(\theta) = n\frac{\tau^2}{\theta^2}$

(c) Considere agora $\tau = 3$.

i) Sabendo que $E[X] = \theta\, c_1$ e $E[X^2] = \theta^2\, c_2$, com c_1 e c_2 constantes, determine o estimador dos momentos para θ e analise o seu enviesamento e consistência em média quadrática. (1.5)

R: Estimador dos momentos: $\tilde{\theta} = \overline{X}/c_1$. $E[\tilde{\theta}] = \theta$ (centrado).
$V[\tilde{\theta}] = \frac{\theta^2(c_2-c_1^2)}{nc_1^2}$. $\lim_{n\to\infty} EQM[\tilde{\theta}] = 0$ (consistente em média quadrática).

ii) Deduza o estimador de máxima verosimilhança para θ e indique, justificando, uma estatística suficiente para θ. (1.5)

O estimador de MV é $\hat{\theta} = \left(\dfrac{n}{\sum_{i=1}^{n}(X_i^{-3})}\right)^{1/3}$.

R: O estimador de MV para θ, existindo, é função de uma estatística suficiente para θ. Logo, uma estatística suficiente para θ é $\hat{\theta}$ ou $T = \sum_{i=1}^{n} X_i^{-3}$

iii) De uma amostra casual de 100 indivíduos a estimativa de máxima verosimilhança observada foi $\hat{\theta} = 23.97$. Deduza um intervalo de confiança aproximado a 99% para θ. (1.5)

R:
$(21.9118, 26.0282)$ usando a V.F. $\sqrt{nI_X(\hat{\theta})}(\hat{\theta}-\theta) \sim^a N(0,1)$
$(22.0745, 26.2216)$ usando a V.F. $\sqrt{nI_X(\theta)}(\hat{\theta}-\theta) \sim^a N(0,1)$
$(21.8861, 26.00088)$ usando a V.F. $2\theta^3 \sum_{i=1}^{n} X_i^{-3} \sim \chi^2_{(2n)}$

iv) Qual a estimativa de máxima verosimilhança para a proporção de indivíduos a quem o medicamento leva mais do que uma hora a fazer efeito? (1.0)

R: Usando a propriedade da invariância dos estimadores de MV, a estimativa pedida é 0.06177.

2. Considere uma moeda equilibrada, em que, quando atirada ao ar, a probabilidade de sair coroa é igual à probabilidade de sair cara. Mostre que para qualquer $\varepsilon > 0$, é possível garantir com probabilidade pelo menos 95% que a proporção de caras na amostra se encontra entre $0.5 - \varepsilon$ e $0.5 + \varepsilon$, desde que se lance a moeda um número suficiente de vezes. (1.5)

R: Usando o TLC, $\forall \varepsilon > 0 \quad P(|\overline{X} - 0.5| < \varepsilon) \geqslant 0.95$ desde que $n \geqslant 0.25 \dfrac{1.96^2}{\varepsilon^2}$.

PARTE II

1. Assume-se que o tempo, X, em segundos, que os utilizadores de um portal do governo esperam até entrar segue a seguinte função densidade

$$f(x|\beta) = \frac{2xe^{-x^2/\beta}}{\beta}, \qquad x > 0, \qquad \beta > 0.$$

Sabe-se que X^2 segue uma distribuição exponencial de média β.

(a) Pretende-se testar $H_0 : \beta = \beta_0$ contra $H_1 : \beta = \beta_1$. Sabe-se que, para uma amostra de dimensão n, o teste mais potente rejeita a hipótese nula quando $\sum_{i=1}^n x_i^2 < c$. Diga, justificando, qual a relação entre β_0 e β_1? (1.5)

R: Usando o Lema de Neyman-Pearson, $\beta_1 < \beta_0$.

(b) Pretende-se testar as hipóteses $H_0 : \beta \leqslant 7$ contra $H_1 : \beta > 7$, recorrendo a uma amostra casual de 15 utilizadores. Determine, justificando, a região crítica uniformemente mais potente, considerando um nível de significância de 5%. (2.0)

R: Pelo teorema de Karlin-Rubin a região crítica UMP é $W_\alpha = \left\{ (x_1, \ldots, x_n) : \sum\limits_{i=1}^n x_i^2 > c \right\}$.

Usando a definição de nível de significânica e $\dfrac{2}{\beta} \sum\limits_{i=1}^n X_i^2 \sim \chi^2_{(2n)}$ conclui-se que

$c = \dfrac{7}{2} F^{-1}_{\chi^2_{(2n)}}(0.95) = 153.2055.$

(c) Da observação dos 15 utilizadores do portal do governo, registou-se $\sum_{i=1}^n x_i^2 = 155.43$. Neste caso o valor-p é inferior ou superior a 5%? Justifique. Usando um nível de significância de 5%, qual a decisão a tomar? (1.0)

R: $\sum_{i=1}^n x_i^2 = 155.43 \in W_{0.05}$, logo rejeita-se H_0 ao nível de 5% e o valor-p é inferior a 5%. A probabilidade de observar valores tão ou mais desfavoráveis do que $\sum_{i=1}^n x_i^2 = 155.43$ é inferior à probabilidade de observar valores tão ou mais desfavoráveis do que $\sum_{i=1}^n x_i^2 = 153.2055$ (valor crítico a 5%).

(d) Pretende-se agora fazer uma análise bayesiana do problema.

i) Mostre que a quantidade de informação de Fisher para β contida em X é $I_X(\beta) = \beta^{-2}$ e deduza a distribuição *a priori* de Jeffreys para o parâmetro β. (1.5)

R: $\pi^J(\beta) \propto \sqrt{I_X(\beta)} \propto \beta^{-1}$.

ii) Mostre que, usando a distribuição *a priori* de Jeffreys, a distribuição *a posteriori* pertence à família da distribuição gama invertida, ou seja $\beta|x_1,\dots x_n \sim GI(a,b)$, e identifique a e b. (1.5)

Note que $Y \sim GI(a,b)$ se

$$f(y|a,b) = \frac{(b/y)^a e^{-(b/y)}}{y\Gamma(a)}, \qquad y > 0, \qquad a,b > 0.$$

em que $E[Y] = \frac{b}{a-1}$ e $Var[Y] = \frac{b^2}{(a-1)^2(a-2)}$.

R: $\pi(\beta|x_1,\dots,x_n) \propto L(\beta|x_1,\dots,x_n)\pi^J(\beta) \propto \beta^{-(n+1)} e^{-\frac{\sum_{i=1}^n x_i^2}{\beta}} \sim GI\left(\beta|a = n, b = \sum_{i=1}^n x_i^2\right)$.

iii) Deduza a estimativa bayesiana para β que minimiza a função perda quadrática e compare-a com a estimativa de máxima verosimilhança, dada por $\hat{\beta} = \frac{\sum_{i=1}^n x_i^2}{n}$. Utilize a amostra da alínea *c*) para obter valores concretos para ambas as estimativas de β, bayesiana e frequencista. (1.5)

R: A estimativa baeysiana para β que minimiza a função perda quadrática é a média da distribuição *a posteriori*: $E\left[\beta|x_1,\dots,x_n\right] = \dfrac{\sum_{i=1}^n x_i^2}{n-1}$.

iv) Comente a afirmação "a distribuição *a priori* de Jeffreys é não informativa". Caso tenha resolvido as alíneas anteriores, faça o comentário à luz dos resultados obtidos. (1.0)