USP – ICMC – SME0806 - Estatística Computacional
3º trabalho – 1º/2015

1. A distribuição gama com parâmetros $\alpha > 0$ e $\beta > 0$ tem função densidade
$$f(x;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \text{ se } x > 0; \quad f(x;\alpha,\beta) = 0, \text{ caso contrário,}$$
em que $\Gamma(\cdot)$ denota a função gama. A matriz de informação de Fisher de $\boldsymbol{\theta} = (\alpha, \beta)$ é dada por
$$\boldsymbol{I}_F(\boldsymbol{\theta}) = \begin{bmatrix} \psi'(\alpha) & 1/\beta \\ 1/\beta & \alpha/\beta^2 \end{bmatrix},$$
em que $\psi(\cdot)$ denota a função digama, dada por $\psi(x) = \Gamma'(x)/\Gamma(x)$, para $x > 0$.

   Realize um estudo de simulação para avaliar algumas propriedades do estimador de máxima verossimilhança do parâmetro de forma α. Diferentes valores de α e diferentes tamanhos amostrais devem ser usados.

   Notas 1. Em linguagem R, a função $\psi'(\cdot)$ está implementada na função `trigamma`.
   2. A distribuição assintótica do estimador de máxima versossimilhança de $\boldsymbol{\theta}$ é normal com vetor de médias $\boldsymbol{\theta}$ e matriz de covariâncias $[n\boldsymbol{I}_F(\boldsymbol{\theta})]^{-1}$. Portanto, a distribuição assintótica do estimador de máxima versossimilhança de α é normal com média α e variância $\text{var}(\widehat{\alpha})$ dada pelo elemento $(1,1)$ da matriz $[n\boldsymbol{I}_F(\boldsymbol{\theta})]^{-1}$.
   3. Em linguagem R, estimativas de máxima verossimilhança de $\boldsymbol{\theta}$ e seus respectivos erros padrão podem ser obtidos com a função `fitdistr` do pacote `MASS`.

2. Uma forma de quantificar a diferença entre duas variáveis aleatórias X e Y é dada pelo desvio quadrático médio Δ^2, com expressão $\Delta^2 = \text{E}[(X-Y)^2]$. Supomos que a distribuição conjunta de (X,Y) tem segundos momentos finitos em que $\text{E}(X) = \mu_X$, $\text{E}(Y) = \mu_Y$, $\text{Var}(X) = \sigma_X^2$, $\text{Var}(Y) = \sigma_Y^2$ e $\text{Cov}(X,Y) = \sigma_{XY}$.

   (a) Prove que $\Delta^2 = (\mu_Y - \mu_X)^2 + \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}$.

   (b) Com base em uma amostra aleatória de tamanho n de (X,Y), obtemos o estimador $\widehat{\Delta^2} = (\overline{Y} - \overline{X})^2 + S_X^2 + S_Y^2 - 2S_{XY}$, em que $\overline{Y}$, $\overline{X}$, S_X^2, S_Y^2 e S_{XY} denotam médias, variâncias e covariância amostrais.
   Pode ser provado que $\log(\widehat{\Delta^2})$ tem distribuição normal assintótica com média $\log(\Delta^2)$ e variância $2[1 - (\mu_Y - \mu_X)^4/\Delta^4]/(n-2)$, para $n > 2$.

   (c) Foi coletada uma amostra aleatória de pares (x,y) com valores dados abaixo.

   `x: 70,00 55,43 18,87 40,41 57,43 31,14 70,10 137,56 221,20 276,43 316,00`
   `75,56 277,30 331,50 133,74 221,50 132,93 85,38 142,34 294,63 262,52`
   `183,56 86,12 226,55`
   `y: 46,57 31,77 36,07 19,49 127,28 51,25 7,51 141,92 13,95 138,61 175,70`
   `83,59 113,64 210,47 172,71 129,14 70,27 117,02 110,94 231,69 194,52 185,62`
   `154,70 192,51`

   Compare as estimativas do erro padrão de $\log(\widehat{\Delta^2})$ utilizando o resultado assintótico do item 2b e reamostragem.