MI 602 A - Métodos Computacionais em Estatística
Primeiro semestre de 2012
Lista de Exercícios I
Data da entrega: 02/04/2012

1. Seja $X \sim \text{Poisson}(\lambda)$, truncada à direita do zero, ou seja $f_X(x;\lambda) = \dfrac{e^{-\lambda}\lambda^x}{(1-e^{-\lambda})x!}\mathbb{1}_{\{1,2,\dots\}}(x)$. Responda os itens:

   a) Encontre uma relação entre $P(X = x+1)$ e $P(X = x)$, $x = 1, 2, 3.....$

   b) Através da relação encontrada no item a), proponha uma algoritmo para gerar uma amostra aleatória de tamanho, da distribuição em questão, através do método da transformada inversa.

   c) Implemente o algoritmo proposto no item b) na linguagem R.

   d) Considere $\lambda = 7$. Gere uma amostra da distribuição em questão, usando o programa desenvolvido no item c), para cada um dos seguintes valores de n = {20,30,50,100}. Verifique se os valores simulados correspondem à distribuição de interesse.

2. Seja $\boldsymbol{X} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, $\boldsymbol{X} = (X_1, ..., X_p)$, ou seja, um vetor aleatório com distribuição t de Student multivariada com vetor de médias $\boldsymbol{\mu}$, matriz de covariâncias $\boldsymbol{\Sigma}$ e ν graus de liberdade, com a seguinte fdp.

$$f_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu) = \frac{\Gamma[(\nu+p)/2]}{\Gamma(\nu/2)\nu^{p/2}\pi^{p/2}|\boldsymbol{\Sigma}|^{1/2}\left[1+\frac{1}{\nu}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\right]^{(\nu+p)/2}}\mathbb{1}_{\mathcal{R}^p}(\boldsymbol{x})$$

   Responda os itens.

   a) Escreva a respresentação estocástica do vetor aleatório acima em termos de uma mistura de Normais multivariadas em função da distribuição gama.

   b) Proponha um algoritmo para gerar um amostra aleatória de tamanho n da distribuição em questão usando o resultado do item a).

   c) Implemente o algoritmo proposto no item b) na linguagem R. Obs: para gerar da distribuição gama você pode usar funções do R mas, para gerar da Normal Multivariada você deve usar o algoritmo visto em sala.

   d) Considere $p = 3$, $\boldsymbol{\mu} = (-1, 0, 2)$, $\boldsymbol{\Sigma} = \begin{bmatrix} 2 & 1 & 0.8 \\ 1 & 5 & 2 \\ 0.8 & 2 & 3 \end{bmatrix}$ e $\nu = 8$. Gere uma amostra da distribuição em questão, usando o programa desenvolvido no item c), para n = 100. Verifique se os valores simulados correspondem à distribuição de interesse, em termos das distribuições marginais e de alguma distribuição univariada gerada a partir do vetor aleatório.

3. Seja $X \sim NA(0,1,\lambda)$, em que $NA(\mu,\psi,\lambda), \mu \in (-\infty,\infty), \psi \in (0,\infty), \lambda \in (-\infty,\infty)$, denota uma distribuição normal assimétrica com parâmetro de localização μ, de escala ψ e de assimetria λ. Pesquisa a representaçãpo estocástica da $NA(0,1,\lambda)$ em termos das distribuições $N(0,1)$ e $HN(0,1)$ (normal (0,1) truncada à esquerda do zero). A fdp de uma va $X \sim NA(\mu,\psi,\lambda)$ é dada por:

$$f_X(x) = \frac{2}{\sqrt{\psi}}\phi\left(\frac{x-\mu}{\sqrt{\psi}}\right)\Phi\left[\lambda\left(\frac{x-\mu}{\sqrt{\psi}}\right)\right]\mathbb{1}_{(-\infty,\infty)}(x)$$

em que ϕ e Φ denotam, respectivamente, a fdp e fda de uma distribuição N(0,1). Responda os itens.

a) Escreva a respresentação estocástica mencionada.

b) Proponha um algoritmo para gerar um amostra aleatória de tamanho n da distribuição em questão usando o resultado do item a).

c) Implemente o algoritmo proposto no item b) na linguagem R. Obs: para gerar da $N(0,1)$ você pode usar funções do R. Para gerar da $HN(0,1)$ você deve usar o algoritmo da rejeição.

d) Considere $\lambda = -4$. Gere uma amostra da distribuição em questão, usando o programa desenvolvido no item c), para cada um dos seguintes valores de n ={20,30,50,100}. Verifique se os valores simulados correspondem à distribuição de interesse.

4. Considere o modelo de regressão normal linear visto em classe. Defina a matriz "hat" como sendo $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. Responda os itens:

a) Considere a decomposição $\boldsymbol{QR}$ da matriz $\boldsymbol{X} = \boldsymbol{QR}$. Simplifique, em termos da decomposição $\boldsymbol{QR}$ a expressão da matriz "hat".

b) A medida de leverage é defina como sendo diag($\boldsymbol{H}$). Implemente um programa que calcule essa medida baseada na expressão original e na expressão reduzida usando a decomposição $\boldsymbol{QR}$.

c) Simule $R = 200$ vezes matrizes de planejamento $\boldsymbol{X}$ com n = 5000 linhas e p = 300 colunas. Para cada uma das R réplicas, calcule a medida leverage usando a definição e calcule quando tempo o processo levou. Repita a operação usando a decomposição $\boldsymbol{QR}$ para calcular a medidade leverage. Qual dos dois processos consumiu menos tempo? Sugestão: procure sempre apagar os objetos criados usando o comando rm(nome do objeto).