MI 602 A - Métodos Computacionais em Estatística
Primeiro semestre de 2012
Lista de Exercícios III
Data da entrega: 31/05/2012 (todas as questões)

1. 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. Pesquise 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). Desejamos estimar λ com base em uma amostra aleatória de X. Responda os itens.

   a) Usando a representação estocástica mencionada para a normal assimétrica e uma reparametrização adequada para λ (se ncessário), desenvolva o algoritmo EM, para estimar λ por máxima verossimilhança, de modo que o passo M seja analítico. Implemente o algoritmo desenvolvido na linguagem R.

   b Considere $\lambda \in (-4, 0, 4)$. Gere R = 500 réplicas da distribuição em questão, para cada um dos seguintes valores de n ={20,30,50,100}. Você pode escolher o método de simulação e, inclusive, usar algum pacote do R para simular. Descreva o comportamento da distribuição empírica, do vício, da variância e do EQM do estimador, em função dos valores verdadeiros do parâmetro bem como dos tamanhos da amostra, calculadas em função das réplicas.

   c) Comente sobre os possíveis problemas de convergência encontrados nas simulações/estimações do item anterior.

2. Em relação à questão 1), considere agora que os três parâmetros são desconhecidos. Refaça os itens a),b) e c), considerando os mesmos tamanhos de amostra e valores de λ. Utilize $\mu = 2$ e $\sigma = 4$.

3. Seja $Y_1, ..., Y_n$ uma amostra de $Y_i \sim$ Bernoulli$(p_i), i = 1, .., n$, mutuamente independentes. Considere que $p_i = F_\nu(\boldsymbol{X}_i'\boldsymbol{\beta})$, em que $F_\nu(.)$ denota a fda de uma distribuição t padrão com ν graus de liberdade (conhecido), $\boldsymbol{X}_i = (1, X_{i1}, ..., X_{ip})$ é um vetor $p \times 1$ de covariáveis conhecidas e $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_p)$ é um vetor com $p + 1$ parâmetros (desconhecidos). Responda os itens:

   a) Escreva a verossimilhança aumentada, em função da representação estocástica da distribuição t padrão, em função das distribuições normal e gama.

   b) Desenvolva o algoritmo CADEM para estimar $\boldsymbol{\beta}$, em que os dois tipos de variáveis aumentadas são aqueles definidos no item b). Implemente o algoritmo desenvolvido na linguagem R.

   c) Considere $\nu = 7$ e $\beta = (1; 0, 3; -2)$. Simule R = 500 réplicas do modelo acima, para cada um dos seguintes valores de n ={20,30,50,100}. Você pode escolher o método de simulação e, inclusive, usar algum pacote do R para simular. Descreva o comportamento das distribuições empíricas marginais, dos vícios, das variâncias e dos EQMs dos estimadores, em função dos valores verdadeiros dos parâmetros bem como dos tamanhos da amostra, calculadas em função das réplicas. Você pode fixar os valores das covariáveis ou simulá-las através de $U(0,1)$.

4. Seja $Y_1, ..., Y_n$ uma amostra de $Y_i \sim$ gama$(\mu_i, \phi), i = 1, .., n$, mutuamente independentes, em que $\mathcal{E}(Y_i|b_i) = \mu_i = \exp[\boldsymbol{X}_i'\boldsymbol{\beta} + b_i]$, $b_i \sim N(0, \sigma^2)$, mutuamente independentes, ϕ é o parâmetro de dispersão, e as outras componentes são como na questão 3. Ou seja, tem-se um modelo de regressão linear generalizado misto, com distribuição gama e função de ligação canônica.

    b) Desenvolva o algoritmo ECEM para estimar $\boldsymbol{\beta}$ e ϕ. Implemente o algoritmo desenvolvido na linguagem R.

    c) Considere $\psi = 50$ e $\beta = (1; 0,3; -2)$. Simule R = 500 réplicas do modelo acima, para cada um dos seguintes valores de n ={20,30,50,100}. Você pode escolhar o método de simulação e, inclusive, usar algum pacote do R para simular. Descreva o comportamento das distribuições empíricas marginais, dos vícios, das variâncias e dos EQMs dos estimadores, em função dos valores verdadeiros dos parâmetros bem como dos tamanhos da amostra, calculadas em função das réplicas. Você pode fixar os valores das covariáveis ou simulá-las através de $U(0,1)$.