MI 602A - Métodos computacionais em Estatística
Primeiro semestre de 2012
Prova I
Data: 16/04/2012

Nome: ____ RA: ____________________

Leia atentamente as instruções abaixo:

- **Coloque seu nome completo e RA em todas as folhas que você recebeu, inclusive nesta.**
- **Utilize somente um dos lados de cada folha.**
- **Leia atentamente cada uma das questões.**
- **Enuncie, claramente, todos os resultados que você utilizar.**
- **Justifique, adequadamente, seus desenvolvimentos, sem, no entanto, escrever excessivamente.**
- **O(a) aluno(a) só poderá sair da sala após as 10h30, mesmo que já tenha finalizado a prova. Após a saída do(a) primeiro(a) aluno(a) não será permitido a entrada de nenhum(a) outro(a) aluno(a).**
- **Não é permitdo empréstimo de material.**
- **Não serão dirimidas dúvidas de quaisquer natureza, após os 20 minutos iniciais.**
- **Resolva a prova, preferencialmente, à caneta, e procure ser organizado(a). Se fizer à lápis, destaque, à caneta, sua resposta.**
- **O(a) aluno(a) deverá portar sua carteira de estudante e apresentá-la, quando for solicitada sua assinatura.**
- **Contestações a respeito da nota, só serão consideradas se estiverem por escrito.**
- **A nota do aluno(a) será $\frac{NP}{NT} \times 10$, em que NP é o número de pontos obtidos na prova e NT é o numero total de pontos da prova.**
- **Use somente um lado de cada folha de resolução.**
- **Os resultados numéricos finais devem ser apresentados com duas casas decimais, apenas.**
- **A prova terá duração de 120 minutos, das 10h00 às 12h00, improrrogáveis.**

Faça uma excelente Prova!!

Parte a ser resolvida em sala

1. Seja $X \sim t_{[a,b]}(\nu), \nu > 0, a < b$, ou seja, uma variável aleatória com distribuição t de Student padrão truncada no intervalo $[a, b]$. Responda os itens.

   a) Prove que a função distribuição acumulada (fda) de X é dada por $F_X(x) = \dfrac{F_Y(x) - F_Y(a)}{F_Y(b) - F_Y(a)}$, em que F_Y é a fda de uma distribuição t de Student padrão não truncada (60 pontos).

   b) Obtenha a função densidade de probabilidade (fdp) de X (20 pontos).

   c) Para simular uma amostra aleatória (aa) de tamanho n de X, proponha um algoritmo da transformada inversa (120 pontos).

2. Considere que $X \sim$ beta-binomial$(m, a, b), m \in I\!N, a, b > 0$. Responda os itens.

   a) Prove que $f_X(x) = \int_0^1 f(x|y)f(y)dy$, em que $X|y \sim$ Binomial(m, y) e $Y \sim$ beta(a, b) (50 pontos).

   b) Obtenha $\mathcal{E}(X)$ e $\mathcal{V}(X)$ usando as propriedades de esperança e variância condicionais (90 pontos).

   c) Para simular uma aa de tamanho n de X, proponha um algoritmo baseado na representação estocástica do item a). Para simular das duas variáveis envolvidas, escolha qualquer um dos métodos vistos (60 pontos).

3. Considere a seguinte integral $h = \int_0^1 \left(\int_0^1 x^2 y^4 (1 - x - y)^3 dx \right) dy$. Responda os itens

   a) Escreva h como $\mathcal{E}[w(x, y)] = \int_0^1 \left(\int_0^1 w(x, y)f(x)f(y)dx \right) dy$, em que X e Y são duas variáveis aleatórias independentes ambas com distribuição beta (20 pontos).

   b) Para calcular a integral acima, proponha um algoritmo de integração de Monte Carlo, usando o resultado obtido no item a) (90 pontos).

   c) Para calular a integral acima, proponha um algoritmo de integração por amostragem por importância considerando que $g(x, y) \sim$ Dirichlet$_2(3, 5, 2)$, ou seja

   $$g(x, y) = \frac{\Gamma(10)}{\Gamma(3)\Gamma(5)\Gamma(2)} x^2 y^4 (1-x-y) \mathbb{1}_{(0,1)}(x) \mathbb{1}_{(0,1-x)}(y) = \frac{\Gamma(10)}{\Gamma(3)\Gamma(5)\Gamma(2)} x^2 y^4 (1-x-y) \mathbb{1}_{(0,1-y)}(x) \mathbb{1}_{(0,1)}(y),$$

   usando o resultado do item a) (90 pontos).

4. Considere $X_1, ..., X_n$ uma amostra aleatória de $X \sim \exp(\lambda), \lambda > 0$, em que $\lambda = e^\theta, \theta \in (-\infty, \infty)$. Desejamos estimar θ, com base nessa amostra, via máxima verossimilhança. Responda os itens.

   a) Obtenha a log-verossimilhança, a função escore, a função Hessiana e a informação de Fisher (100 pontos).

   b) Descreva os algoritmos de Newton-Raphson e Escore de Fisher, para a obtenção da estimativa de máxima verossimilhança (80 pontos).

   c) Como o erro-padrão assintótico associado à estimativa pode ser obtido? (20 pontos).

Formulário

1. Se $X \sim t(\nu)$ então $f_X(x) = \dfrac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\dfrac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} \mathbb{1}_{(-\infty,\infty)}(x)$

2. Se $X \sim \text{binomial}(m,\theta), m \in I\!N, \theta \in (0,1)$ então, $f_X(x) = \dfrac{m!}{x!(m-x)!}\theta^x(1-\theta)^{m-x}\mathbb{1}_{\{0,1,2,\ldots,m\}}(x)$, $\mathcal{E}(X) = m\theta$, $\mathcal{V}(X) = m\theta(1-\theta)$.

3. Se $X \sim \text{gama}(r,\lambda), r > 0, \lambda > 0$, então $f_X(x) = \frac{1}{\theta^r\Gamma(r)}e^{-\frac{x}{\lambda}}x^{r-1}\mathbb{1}_{(0,\infty)}(x), \mathcal{E}(X) = r\lambda, \mathcal{V}(X) = r\lambda^2$.

4. Fazendo-se $r = 1$, tem-se que $X \sim \exp(\lambda)$.

5. Se $X \sim \text{beta}(a,b), a > 0, b > 0$, então $f_X(x) = \frac{1}{\beta(a,b)}x^{a-1}(1-x)^{b-1}\mathbb{1}_{(0,1)}(x), \beta(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, $\mathcal{E}(X) = \frac{a}{a+b}, \mathcal{V}(X) = \frac{ab}{(a+b)^2(a+b+1)}$.

6. Se $\boldsymbol{X}|(a,b,c) = (X_1,X_2)|(a,b,c) \sim \text{Dirichlet}_2(a,b,c), a,b,c > 0$, então

$$\begin{aligned} p(\boldsymbol{x}|a,b,c) &= \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)}x_1^{a-1}x_2^{b-1}(1-x_1-x_2)^{c-1}\mathbb{1}_{(0,1)}(x_1)\mathbb{1}_{(0,1-x_1)}(x_2) \\ &= \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)}x_1^{a-1}x_2^{b-1}(1-x_1-x_2)^{c-1}\mathbb{1}_{(0,1)}(x_2)\mathbb{1}_{(0,1-x_2)}(x_1). \end{aligned}$$

7. Relativo à distribuição de Dirichlet temos que $X_1|(a,b,c) \sim \text{Beta}(a,b+c)$ e $X_2|(a,b,c) \sim \text{Beta}(b,a+c)$.

8. $\text{beta}(a,b) = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \displaystyle\int_0^1 x^{a-1}(1-x)^{b-1}dx$.

9. $\Gamma(r) = \int_0^\infty x^{r-1}e^{-x}dx$, $\Gamma(r) = (r-1)\Gamma(r-1)$. Se $r \in I\!N, \Gamma(r) = (r-1)!$

Parte a ser resolvida em casa: entrega até dia 23/04/2012 às 18h00, junto com a resolução das questões acima. As resoluções das questão acima poderão ser entregues manuscritas. A parte a ser resolvida em casa deverá ser digitada e entregue impressa e por e-mail em PDF.

10. Com relação à Questão 1 acima (200 pontos):

    a) Implemente o algoritmo proposto na linguagem R.

    b) Considere $\nu = 8$. Gere uma amostra da distribuição em questão, usando o programa desenvolvido no item a), para cada um dos seguintes valores de n = {20,30,50,100}. Verifique se os valores simulados correspondem à distribuição de interesse comparando os histogramas com as curvas verdadeiras, bem como as médias e variâncias empíricas com os verdadeiros valores.

11. Com relação à Questão 2 acima (200 pontos):

    a) Implemente o algoritmo proposto na linguagem R.

    b) Considere $m = 9, a = 4, b = 0,1$. Gere uma amostra da distribuição em questão, usando o programa desenvolvido no item a), para cada um dos seguintes valores de n = {20,30,50,100}. Verifique se os valores simulados correspondem à distribuição de interesse comparando os histogramas com as curvas verdadeiras, bem como as médias e variâncias empíricas com os verdadeiros valores.

12. Com relação à Questão 3 acima (200 pontos):

    a) Implemente os algoritmo propostos na linguagem R.

    b) Compare os valores obtidos da integral, através dos dois algoritmos, entre si e com o verdadeiro valor para $m \in \{1000, 2000, 3000, ..., 10.000\}$

13. Com relação à Questão 4 acima (200 pontos):

    a) Implemente os algoritmos propostos na linguagem R.

    b) Implemente o algoritmo BFGS usando a função *optim* da linguagem R.

    c) Considere que $n \in \{10, 30, 50, 200\}$ e $\theta \in \{-2, 0, 2\}$. Para cada um dos tamanhos amostrais e valores de θ, simule R = 500 conjuntos de respostas com base nos 3 algoritmos construídos, obtenha as estimativas de máxima verossimilhança e os respectivos erros-padrão assintóticos. Faça um histograma para cada um dos conjuntos de estimativas, para cada um dos algoritmos. O que ocorre com a distribuição do emv, sua média e seu erro-padrão? Qual dos algoritmos gera as melhores estimativas?

OBS: Será atribuida uma nota entre 0 e 10 para cada parte. As partes feitas em classe e em casa terão o mesmo peso na composição final da nota da prova.