Trabalho de Inferência Bayesiana (2019/02)

Professora: Renata Souza Bueno

Instruções:

• O trabalho deverá ser feito individualmente ou em dupla.

• O formato do trabalho deverá ser um *script* em R.

• As contas necessárias para implementação do algoritmo devem ser entregues (impressas ou no *script* do R - formato Markdown).

• O aluno deverá me enviar por email (caso as contas estejam no papel, me entregar pessoalmente) até o dia 07/12/2019.

• Só considere que o trabalho foi recebido via email quado receber um email de confirmação enviado por mim.

• Você deve escolher uma das opções abaixo para implementar.

• O trabalho se refere a um estudo simulado, ou seja, você deve fixar os valores dos parâmetros e simular os dados. Simule uma amostra de $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ do modelo escolhido considerando $n = 1000$. Depois de gerada a amostra, suponha que os parâmetros deste modelo são desconhecidos e utilize a inferência bayesiana para estimá-los.

• As opções 2 e 4 valem um ponto extra.

Situação 1: Distribuição t-Student

Uma variável aleatória Y tem distribuição t-Student se sua função de densidade é dada por

$$f(y|\mu,\sigma,\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}\sigma}\left(1+\frac{1}{\nu}\left(\frac{y-\mu}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}, \quad y \in \mathbb{R},$$

em que $\mu \in \mathbb{R}$ é um parâmetro de locação, $\sigma > 0$ é um parâmetro de escala e $\nu > 0$ é o parâmetro chamado graus de liberdade. Uma representação estocástica que pode facilitar o uso desta distribuição é dada por:

$$Y = \mu + \sigma\frac{V}{\sqrt{U}}$$

onde $V \sim N(0,1)$ e $U \sim Gama(\frac{\nu}{2},\frac{\nu}{2})$. U e V são independentes. Seja $\boldsymbol{\theta} = (\mu,\sigma^2,\nu)$ e a distribuição a priori para $\boldsymbol{\theta}$:

$$p(\boldsymbol{\theta}) = p(\mu)p(\sigma^2)p(\nu),$$

onde $\mu \sim N(a,b)$, $\sigma^2 \sim GI(c,d)$ e $\nu \sim Gama(e,f)$.

Sugestão: Simule os dados considerando $\nu > 2$.

Opção 1: Considerar o parâmetro ν fixo.

Opção 2: Estimar o parâmetro ν.

Situação 2: Distribuição Normal Assimétrica

Uma variável aleatória Y tem distribuição normal assimétrica se sua função de densidade é dada por

$$f(y|\mu,\sigma,\lambda) = \frac{2}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right)\Phi\left(\lambda\,\frac{y-\mu}{\sigma}\right), \quad y \in \mathbb{R},$$

em que $\phi(\cdot)$ e $\Phi(\cdot)$ indicam a função de densidade e a função de distribuição acumulada de uma normal padrão, respectivamente. Notação: $Y \sim NA(\mu,\sigma,\lambda)$, onde $\mu \in \mathbb{R}$ é um parâmetro de locação, $\sigma > 0$ é um parâmetro de escala e $\lambda \in \mathbb{R}$ é o parâmetro de assimetria.

Uma representação estocástica que pode facilitar o uso desta distribuição é dada por:

$$Y = \mu + \delta\sigma U + (1-\delta^2)^{1/2}\sigma V$$

onde $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}} \in [-1,1]$, $V \sim N(0,1)$ e $U \sim N_+(0,1)$. U e V são independentes. Faça $\sigma = 1$. Seja $\boldsymbol{\theta} = (\mu,\lambda)$ e a distribuição a priori para $\boldsymbol{\theta}$:

$$p(\boldsymbol{\theta}) = p(\mu)p(\lambda),$$

onde $\mu \sim N(a,b)$ e $\lambda \sim N(c,d)$.

Opção 3: Considerar o parâmetro λ fixo.

Opção 4: Estimar o parâmetro λ.

Itens obrigatórios do trabalho

1. **Simulação dos dados:** Como os dados foram simulados e quais valores foram considerados para o vetor paramétrico.

2. **Atribuição dos hiperparâmetros:** Especificar todos os hiperparâmetros e justificar a escolha dos mesmos.

3. **Contas:** Mostrar o cálculo do núcleo da distribuição a posteriori bem como todas as distribuições condicionais completas.

4. **Distribuição proposta:** Caso tenha escolhido as opções 2 e 4 falar qual a distribuição proposta utilizada no passo de Metropolis e a taxa de aceitação obtida. Comente a taxa.

5. **Algoritmo:** Explicite o algoritmo usado, especificando cada passo (código em R).

6. **Tratamento da cadeia:** Retirar o aquecimento, o espaçamento e verificar a convergência usando o método de Gelman e Rubin(1992). Comente os resultados.

7. **Traços das cadeia:** Fazer os gráficos da cadeias adicionando uma linha para o valor verdadeiro. Faça um gráfico comparando a distribuição a priori e a distribuição a posteriori. Compare as diferentes crenças.

8. **Estimativas a posteriori:** Calcular as estimativas pontuais e intervalares a posteriori e comente comparando com os valores verdadeiros que foram usados na geração dos dados.