Inferência Bayesiana
Lista 2 de Exercícios
Prof. Bruno Santos

1. Seja $X_1, \dots, X_n$ uma amostra aleatória da distribuição Exponencial(θ).

   a) Mostre como construir um intervalo de credibilidade para θ usando uma distribuição a priori conjugada e também com a priori de Jeffreys.

   b) Suponha que queremos testar $H_0 = \theta = 1$ contra $H_1 : \theta \neq 1$. Considere a priori $\theta | H_1 \sim$ Gama$(3, 3)$. Para $n = 4$, $\bar{x} = 1,4$ e $p_0 = 1/2$, calcule $P(H_0|x)$ e o fator de Bayes a favor de H_0.

2. Pretende-se testar $H_0 : \mu = \mu_0$ contra $H_1 : \mu \neq \mu_0$ considerando n observações do modelo $N(\mu, \sigma^2)$, com σ^2 conhecido.

   i) $P\{\mu = \mu_0\} = p_0$;

   ii) $\mu | H_1 \sim N(\mu_0, b^2)$.

   Encontre o fator de Bayes a favor da hipótese nula.

3. Reconsidere o exercício 9 da Lista 1.

   a) Para a distribuição a priori considerada no item a), obtenha $P(\theta_A < \theta_B | y_A, y_B)$ considerando amostras de Monte Carlo. Apresente o programa computacional utilizado.

   b) Para alguns valores de n_0, obtenha

   $$P(\theta_A < \theta_B | y_A, y_B)$$

   para $\theta_A \sim$ Gama$(120, 10)$ e $\theta_B \sim$ Gama$(12 \times n_0, n_0)$. Descreva o quão sensível as conclusões sobre o evento $\{\theta_A > \theta_B\}$ são com respeito à distribuição a priori em θ_B.

4. Os dados "CoalDisast.csv" dizem respeito a desastres verificados em minas de carvão na Grã-Bretanha de 1851 a 1962. Suponha que Y_i representa o número de desastres no ano t_i e que haja uma quebra no comportamento das taxas de ocorrência no k-ésimo ano dos dados, expressado por $Y_i \sim Poisson(\theta)$ para $i = 1, 2, ..., k$ e $Y_i \sim Poisson(\lambda)$ para $i = k + 1, k + 2, ..., n$.

   Admita que os parâmetros do modelo θ, λ, k são independentes a priori, k com distribuição uniforme {1,2,...,n}, $\theta \sim Gama(a_1, b_1)$, $\lambda \sim Gama(a_2, b_2)$. Para os hiperparâmetros assuma uma estrutura hierárquica, em que b_1 e b_2 são independentes com $b_1 \sim$ Gama$(0,0001; d_1)$ e $b_2 \sim$ Gama$(0,0001; d_2)$ e $a_1 = a_2 = 0,5$. Considere d_1 e d_2 conhecidos e iguais a 1.

   Usando o amostrador de Gibbs, estime os parâmetros do modelo e obtenha representações gráficas das distribuições a posteriori.

5. Os dados "Wage.txt" correspondem a $n = 1.217$ observações de uma pesquisa feita nos Estados Unidos (*National Longitudinal Survey of Youth (NLSY)*). A primeira coluna é o logaritmo do salário por hora (Y) de cada indivíduo e a segunda coluna apresenta os anos de educação (X).

   Considere um modelo de regressão linear simples, em que

   $$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i,$$

   em que $\epsilon_i \sim N(0, \sigma^2)$. Considere que $\sigma^2 = 0,2668$ e assuma uma distribuição Normal a priori para cada β_i, com média $b_0 = 0$ e variância $B_0 = 10$.

   a) Obtenha amostras da distribuição a priori de $\beta = (\beta_1, \beta_0)$, utilizando amostrador de Gibbs. Obtenha uma estimativa do estimador de Bayes para cada parâmetro, considerando a perda quadrática. E construa um intervalo de credibilidade de 95% para cada parâmetro.

   b) Refaça o item a), porém centralizando a variável preditora (X), isto é, subtraindo todas as observações pela média amostral. Compare as cadeias geradas em cada um itens e comente as diferenças.

6. Considere duas variáveis aleatórias, X e Y, em que exista o interesse em modelar a correlação ρ entre essas duas variáveis. Suponha que essas variáveis possam ser modeladas com uma distribuição Normal bivariada

   $$\begin{pmatrix} X \\ Y \end{pmatrix} | \rho \sim N(\mu, \Sigma) \tag{1}$$

   em que $\mu = (\mu_x \quad \mu_y)^t$ e

   $$\Sigma = \begin{bmatrix} \sigma_{xx} & \rho \\ \rho & \sigma_{yy} \end{bmatrix}$$

   Assuma que $\mu_x = \mu_y = 0$, $\sigma_{xx} = \sigma_{yy} = 1$.

   a) Escreva a verossimilhança considerando uma amostra de pares (X_i, Y_i) com n observações.

   b) Considerando a priori de Jeffreys,

   $$\pi(\mu, \Sigma) = |\Sigma|^{-3/2},$$

   escreva a distribuição a posteriori de ρ, a menos de uma constante de proporcionalidade.

   c) Gere 30 valores da distribuição Normal em (1) e escreva um algoritmo de Metropolis-Hastings para obter 1000 amostras da distribuição a posteriori de ρ, considerando as seguintes propostas.

   - $\rho^{\text{prop}} \sim \text{Uniforme}(\rho^{i-1} - \delta_u, \rho^{i-1} + \delta_u)$;
   - $\rho^{\text{prop}} \sim \text{Normal}(\rho^{i-1}, \delta_n)$;
   - $\rho^{\text{prop}} \sim \text{Gama}(\rho^{i-1} * \delta_g, \delta_g)$.

   Ajuste os valores $\delta_u, \delta_n, \delta_g$ para obter taxas de aceitação do algoritmo entre 0,15 e 0,50.