ME 705 A - Inferência Bayesiana
Segundo semestre de 2023
Lista de Exercícios I

OBS:

- A menos que o contrário seja mencionado, nos exercícios você deve considerar uma amostra aleatória $X_1|\theta, ..., X_n|\theta$ de $X|\theta$.
- A menos que o contrário seja mencionado, variância do estimador Bayesiano deve ser calculada sob a ótica frequentista.
- Sempre construa gráficos das posteriores (para valores específicos dos hiperparâmetros e para alguma amostra). Caso tais valores não sejam fornecidos, escolha-os, usando algum critério.
- Não se esqueça de fazer uma análise descritiva apropriada, inclusive consgtruindo QQ-plots com envelope, quando pertinente.

1. Resolva os exercícios deixados em sala.

2. Seja $X|\theta \sim \text{bin}(m, \theta), \theta \in (0, 1)$, m conhecido. Considere $\theta \sim \text{beta}(a, b)$, (a,b) conhecidos. Responda os itens

   a) Considere m=1. Qual seria a estimativa de máxima verossimilhança de θ quando $x_i = 0, \forall i$? E quanto $x_i = 1, \forall i$?
   b) Encontre a distribuição $\theta|\boldsymbol{x}$ calculando diretamente a constante de normalização.
   c) Obtenha $\widehat{\theta}_{EAP}$, $\widehat{\theta}_{Mo}$ e $\widehat{\theta}_{Md}$ e suas respectivas variâncias.
   d) Como fica a posteriori considerando o cenário do item a)? É possível obter estimativas dentro do espaço paramétrico?
   e) Considere que $a = b = 1$. Qual seria a conjectura à priori com relação à θ neste caso?
   f) Compare os estimadores de máxima verossimilhança, EAP e MAP, através de seus EQM's. Considerando $a = b = 1$, qual dos estimadores você escolheria? Justifique, adequadamente, sua resposta.
   g) Determine a família conjugada natural para o o modelo probabilístico em questão.
   h) Suponha que, a priori, você acredite que $\theta \in \{\theta_0, \theta_1, \theta_2\}$, em que $\theta_i \in (0, 1), i = 0, 1, 2,$. Obtenha a posteriori e calcule o EAP e o MAP neste caso.

3. Seja $X|\theta \sim \text{gama}(r, \lambda), \lambda > 0$, em que $\mathcal{E}[X|\lambda] = r\lambda$, r conhecido.

   a) Determine a família conjugada natural para o modelo em questão.
   b) Ache a distribuição a posteriori de λ com base na priori determinada no item a).
   c) Obtenha os estimadores EAP e MAP de λ e suas variâncias.
   d) Compare os estimadores de máxima verossimilhaça, EAP e MAP, através de seus EQM's. Qual dos estimadores você escolheria? Justifique, adequadamente, sua resposta.

e) A priori encontrada no item a) depende de dois parâmetros. Considerando que os dois são iguais a 1, repita a compração feita no item d).

f) Considere que $p(\lambda) \propto \mathbb{1}_{(0,\infty)}(\lambda)$. Verifique se a posteriori é própria.

4. Seja $X|\theta \sim \text{Poisson}(\theta), \theta > 0$.

   a) Determine a família conjugada natural para o modelo me questão.

   b) Ache a distribuição a posteriori de θ com base na priori encontrada no item a).

   c) Obtenha os estimadores EAP e MAP de θ e suas variâncias.

   d) Compare os estimadores de máxima verossimilhança, EAP e MAP, através de seus EQM's. Qual dos estimadores você escolheria? Justifique, adequadamente, sua resposta.

   e) A priori encontrada no item a) depende de dois parâmetros. Considerando que os dois são iguais a 1, repita a comparação feita no item d).

   f) Considere que $p(\theta) \propto \mathbb{1}_{(0,\infty)}(\theta)$. Verifique se a posteriori é própria.

5. Aplique os resultados da Questão 4 anterior nos dados da Tabela 2.2 pag 69 do livro Bayesian data analysis, second edition. Considere, primeiramente, o número de acidentes fatais como variável resposta e, depois, o número de passageiros mortos. Resolva os exercício tanto com priori do item a) quanto a do item f).

6. Considere $X \sim f_X(x;\theta)$ em que

$$p(x|\theta) = \frac{1}{\theta} x^{\frac{1}{\theta}-1} \mathbb{1}_{(0,1)}(x)$$

   a) Determine a família conjugada natural para o o modelo.

   b) Ache a distribuição a posteriori de θ com base na priori encontrada no item a).

   c) Obtenha os estimadores EAP e MAP de θ e suas variâncias.

   d) Compare os estimadores de máxima verossimilhança, EAP e MAP, através de seus EQM's. Qual dos estimadores você escolheria? Justifique, adequadamente, sua resposta.

   e) A priori encontrada no item a) depende de dois parâmetros. Considerando que os dois são iguais a 1, repita a comparação feita no item d).

   f) Considere que $p(\theta) \propto \mathbb{1}_{(0,\infty)}(\theta)$. Verifique se a posteriori é própria.

7. Considere $X \sim U_{[0,\theta]}, \theta > 0$.

   a) Determine a família conjugada natural para o o modelo.

   b) Ache a distribuição a posteriori de θ com base na priori encontrada no item a).

   c) Obtenha os estimadores EAP, MdAP e MAP de θ e suas variâncias.

   d) Compare os estimadores de máxima verossimilhança, EAP, MdAP e MAP, através de seus EQM's. Qual dos estimadores você escolheria? Justifique, adequadamente, sua resposta.

e) A priori encontrada no item a) depende de dois parâmetros. Considerando que os dois são iguais a 1, repita a comparação feita no item d).

f) Considere que $p(\theta) \propto \mathbb{1}_{(0,\infty)}(\theta)$. Verifique se a posteriori é própria.

8. Como seria a fórmula geral da posteriori se o espaço paramétrico fosse finito? E se fosse infinito enumerável?

9. Considere a fórmula da atualização da posteriori. Como ela ficaria se $\boldsymbol{X}'|\theta \perp \boldsymbol{X}''\theta$?

10. Ainda em relação à formula da atualização da posteriori, considere:

a) Considere uma amostra aleatória de tamanho n de $p(x|\theta) = x_i\theta^{x_i-1}\mathbb{1}_{(1,\infty)}(x), \theta \in (0,1)$ e $p(\theta) =$ beta(a,b), a,b conhecidos. Encontre: a verossimilhança associada ao modelo e a posteriori de θ. Suponha que uma novo conjunto de observações $X_{n+1},...,X_{n+m}$ está disponível. Como ficaria posteriori atualizada?

b) Considere o modelo $Y_i = \theta Y_{i-1} + \xi_i, i = 2,...,n,\ Y_1 \sim N(0,1), \xi_i \sim N(0,1), Y_{i-1} \perp \xi_i, \theta \in (-1,1)$. Assuma que $\theta \sim U_{[-1,1]}$. Encontre: a verossimilhança associada ao modelo e a posteriori de θ. Suponha que uma novo conjunto de observações $Y_{n+1},...,Y_{n+m}$ está disponível. Como ficaria posteriori atualizada?

11. Os dados disponíveis no site (DadosLampada1IB2023.txt) dizem respeito ao tempo de vida de n = 60 lâmpadas, de uma nova versão de um modelo comercial, medidos em horas. Além disso, traz informações sobre as médias dos tempos de vida de 30 (DadosLampada2IB2023.txt) outras versões anteriores da lâmpada em questão. Considere que o tempo de vida de cada lâmpada (da nova versão) possa ser modelado por uma distribuição exponencial $\mathcal{E}(X) = \theta$. Desejamos estimar θ.

a) Faça uma análise descritiva completa dos dados (inclusive calculando medidas descritivas). Você acredita que a suposição de distribuição exponencial para os dados é razoável? Justifique, adequadamente, sua resposta.

b) Proponha uma priori, com base nas informações das 20 estimativas. Justifique, adequadamente, sua escolha.

c) Como base nessa priori, encontre a posteriori e forneça uma estimativa Bayesiana para θ.

12. Um laboratório que pesquisa o câncer está estimando a taxa de tumorigênese em duas espécias de ratos (A e B). Eles tem dados de contagem de tumor para 10 ratos da espécie A e 13 da espécie B. A literatura sugere que a tumorigênese para ratos do tipo A pode ser adequadamente modelada usando uma distribuição de Poisson(θ_A), enquanto que para os ratos da espécie B a distribuição (BN)Binomial-Negativa(r,θ_B), com r conhecido, é mais apropriada. Se $X|\theta \sim BN(r,\theta)$, então $P(X = x|\theta) = \binom{x+r-1}{x}(1-\theta)^x\theta^r\mathbb{1}_{\{0,1,2,...\}}(x)$, condicionado em θ_A e θ_B, as observações são independentes inter e intra grupos. Os dados se encontram abaixo:

$$y_A = (12; 9; 12; 14; 13; 13; 15; 8; 15; 6)$$

$$y_B = (11; 11; 10; 9; 9; 8; 7; 10; 6; 8; 8; 9; 7)$$

Responda os itens:

a) Obtenha a verossimilhança conjunta de (y_A, y_B).

b) Considerando, à priori, $\theta_A \sim gama(a, b)$ e $\theta_B \sim Beta(c, d)$ em que $\theta_A \perp \theta_B$ e (a,b,c,d) conhecidos. Encontre a posteriori conjunta e as marginais.

c) Encontre a distribuição a posteriori de $\lambda = \frac{\theta_A}{\theta_A - \ln \theta_B}$ com base na posteriori conjunta de $(\theta_A, \theta_B)'$.

d) Utilizando a distribuição à posteriori obtida no item c), os dados fornecidos e considerando $r = 15$, em sua opinião, qual das duas espécies apresenta maior taxa de tumorigênese? Justifique, adequadamente, sua resposta.