

Introdução à Estatística Bayesiana - Lista 1

1) Suponha que a proporção de itens defeituosos em um lote seja θ (desconhecido), em uma amostra de 20 itens, encontrou-se apenas 1 item com problema. Supondo que $\theta \sim B(5, 10)$, encontre o estimador Bayesiano para θ usando (a) Perda quadrática, (b) Perda absoluta e (c) Perda zero-um. *Dica: para (b) use o R pra identificar a mediana associada à distribuição a posteriori.*

2) Em um conjunto de rolos de filmes, o número de defeitos nos rolos é uma v.a. X com média desconhecida θ . Uma amostra de cinco rolos foi selecionada e observou-se o seguinte número de defeitos: 2, 2, 6, 0 e 3. Supondo que $\theta \sim G(3, 1)$. Encontre o estimador bayesiano baseado da perda quadrática.

3) Suponha que as alturas (em cm) dos indivíduos de uma população tenha distribuição normal com média θ e variância 10. Suponha também que $\theta \sim N(172, 4)$. Dez pessoas são selecionadas ao acaso, a altura média encontrada foi $\bar{x} = 176$. Encontre os estimadores de Bayes para θ sob perda quadrática e absoluta?

4) Mostre que modelos envolvendo as seguintes distribuições formam uma conjugação. Para cada um deles encontre a distribuição a posteriori.

- a) Binomial-Beta
- b) Binomial Negativa - Beta
- c) Poisson-Gama
- d) Multinomial-Dirichlet
- e) Geometric-Beta

5) Suponha que uma a.a. $x_1, \dots, x_n | \theta \sim U(0, \theta)$ iid e que $\theta \sim \text{Pareto}(x_m, k)$. Encontre a moda a posteriori de θ .

6) Suponha que uma a.a. $x_1, \dots, x_n | r, \theta \sim \text{BinNeg}(r, \theta)^1$ iid, onde r : conhecido e $0 < \theta < 1$ é o parâmetro de interesse.

- a) Suponha que à priori $\theta \sim \text{Beta}(\alpha, \beta)$ (α, β : conhecidos), encontre a distribuição a posterior de θ .
- b) Agora considere que não existe informação a priori disponível, encontre a distribuição a posteriori de θ , usando a priori de Jeffrey.
- c) Sob perda absoluta, qual o estimador Bayesiano da distribuição a posteriori obtida no item (b)?

7) Na questão anterior, um pesquisador tem absoluta certeza de que θ pode assumir qualquer valor no intervalo (a, b) , onde $0 < a, b < 1$ e $a < b$. Encontre a distribuição a posteriori de θ .

8) Seja $x_1, \dots, x_n | \beta \sim G(\alpha, \beta)$ iid, com α : conhecido. Encontre uma distribuição conjugada para o modelo. Especifique a distribuição a posteriori resultante.

¹Binomial Negativa: $P(X_i = x_i) = \binom{x_i + r - 1}{x_i} \theta^{x_i} (1 - \theta)^r$

9) Considere um modelo conjugado Normal-Gama Inversa, i. é

$$\begin{cases} x_1, \dots, x_{10} | \sigma^2 \sim \text{Normal}(0, \sigma^2) \text{ iid} \\ \sigma^2 \sim \text{Gama-Inversa}(a, b), \end{cases}$$

Sabendo-se que a distribuição a posteriori $\sigma^2 | \mathbf{x} \sim \text{Gama} - \text{Inversa}(20, 30)$, encontre os valores de a e b que levaram a esta distribuição a posteriori. Use o fato de que $\sum x_i^2 = 3.5$.

10) Suponha que uma a.a. $x_1, \dots, x_n | \mu, \sigma^2 \sim \text{LogNormal}(\mu, \sigma^2)$ iid.

- Encontre a distribuição a priori de Jeffrey para o modelo.
- Mostre que usando a priori de Jeffrey,

$$p(\mu, \sigma^2 | \mathbf{x}) \propto (\sigma^2)^{-(n+3)/2} \exp\left\{-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\mu - \bar{z})^2]\right\},$$

onde $z_i = \log(x_i)$, $\bar{z} = \sum z_i / n$ e $(n-1)s^2 = \sum (z_i - \bar{z})^2$.

11) Ainda em relação à Questão 10:

- Mostre que a distribuição marginal a posterior de μ é t de Student, especifique os parâmetros da distribuição.
- Estime μ , usando perda quadrática, absoluta e 0-1.
- Qual o desvio-padrão associado ao estimador?

12) Ainda em relação à Questão 10:

- Encontre o estimador Bayesiano para σ^2 , sob perda quadrática e 0-1.
- Reconsidere o modelo da Questão 1, agora usando uma distribuição a priori $p(\mu, \sigma^2) \propto 1$. Compare o estimador (sob perda quadrática) com aquele obtido no item (a). O que ocorre com os dois estimadores quando n é grande?
- Usando o modelo com distribuição a priori de Jeffrey, estime $g(\sigma^2) = 1/\sigma^2$.

13) Suponha que uma a.a. 18.36, 16.17, 10.21, 11.07, 18.32, 6.03, 18.42, 8.04, 13.53 $|\mu, \sigma^2 \sim \text{Normal}(\mu, \sigma^2)$ iid, onde $\sigma^2 = 4$. Sendo que $\bar{x} = 13.35$ e $S^2 = 22.54$. Seja $\mu \sim N(8, 9)$.

- Qual das fontes de informação (dados e à priori) carrega menos incerteza sobre μ ? Justifique.
- Encontre a distribuição a posterior para o modelo. *Dica: Não precisa necessariamente deduzir a expressão, basta escrever usando os valores dados.*
- O que acontece com a incerteza sobre μ após observar os dados?
- O estimador Bayesiano sob perda quadrática será mais influenciado pela informação amostral ou à priori? Justifique.

14) Seja $x_1, x_2, \dots, x_n | \theta \sim \text{Poisson}(\theta)$ uma a.a. iid. Informações a priori reunidas sobre θ apontam que $\theta \sim \text{Gama}(a, b)$, com θ em torno de 2, com desvio-padrão 1. Sabe-se que $\mathbf{x} = (0, 2, 3, 1, 0, 0, 3)$.

- a) Encontre a distribuição a posteriori de θ e seus estimadores sob perdas quadráticas e 0–1.
- b) O que se pode dizer sobre o estimador Bayesiano sob perda absoluta (mediana)?
- c) Mostre que o estimador Bayesiano sob perda quadrática é assintoticamente não-viesado.

15) Considere $x_1, \dots, x_k | \theta_1, \dots, \theta_k \sim \text{Multinomial}(n; \theta_1, \theta_2, \dots, \theta_k)$, $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$ são desconhecidos.

- a) Encontre a priori de Jeffreys para o modelo.
- b) Proponha uma priori conjugada e encontre a distribuição a posteriori.
- c) Encontre o coeficiente de correlação de Pearson para θ_1 e θ_2 . Assuma: $k = 3$, $n = 10$, $\sum x_{i1} = 8$, $\sum x_{i2} = 7$ e $\sum x_{i3} = 2$.

16) Seja θ um parâmetro que pode assumir apenas dois valores: θ_1 e θ_2 . Uma v.a. Y assume valores 1 e 2, tendo distribuição:

$$P(Y = y | \theta_1) = \begin{cases} 0.2 & ; \text{ se } y = 1 \\ 0.8 & ; \text{ se } y = 2 \\ 0; \text{ c.c.} \end{cases} \text{ e } P(Y = y | \theta_2) = \begin{cases} 0.5 & ; \text{ se } y = 1 \\ 0.5 & ; \text{ se } y = 2 \\ 0; \text{ c.c.} \end{cases}$$

Assuma que $p(\theta_1) = p(\theta_2) = 0.5$ e uma amostra de tamanho $n = 1$.

- a) Encontre a distribuição a posteriori de θ .
- b) Quão mais provável é um valor de θ em relação ao outro, antes de observar o dado y ?
- c) Dado que o valor observado foi $y = 2$, quão mais provável é um valor de θ em relação ao outro?
- d) Dado que o valor observado foi $y = 1$, quão mais provável é um valor de θ em relação ao outro?

17) Suponha que uma a.a. $\mathbf{x} = (24, 21, 10, 2.5, 9.8) | \mu, \sigma^2 \sim \text{Log-Normal}(\mu, \sigma^2)$ iid, onde $\sigma^2 = 1$. Seja $\mu \sim N(8, 1)$.

- a) Proponha um modelo conjugado e obtenha a distribuição a posteriori.
- b) Obtenha os estimadores bayesianos sob perdas quadrática, absoluta e 0–1.
- c) De acordo com todas as fontes de informação, a distribuição a posteriori resultante carrega muita incerteza sobre μ ? Justifique.

Nota: Distribuição LogNormal: $f(x | \mu, \sigma^2) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$. De forma que se uma v.a. $W \sim N(\mu, \sigma^2)$ então $e^W \sim \text{Log-Normal}(\mu, \sigma^2)$.

18) Suponha que uma amostra $x_1, \dots, x_n | \lambda \sim \text{Poisson}(\lambda)$ iid. $\lambda \sim G(\alpha, \beta)$.

- a) Determine a distribuição a posteriori.

b) Encontre a distribuição preditiva de uma nova observação \tilde{x} .

19) Considere a questão anterior. Suponha que $\sum_{i=1}^7 x_i = 20$, $\alpha = 5$ e $\beta = 8$.

a) Qual o valor previsto para a oitava observação?

b) Qual o desvio padrão associado a este valor?

c) Se alguém afirmasse que a nova observação será $\tilde{x} = 10$, você endossaria tal afirmação?

20) Suponha que uma a.a. $x|\theta \sim \text{Bin}(n, \theta)$ iid, onde $0 < \theta < 1$ é o parâmetro de interesse, onde $\theta \sim \text{Beta}(\alpha, \beta)$ (α, β : conhecidos).

a) Determine a distribuição preditiva para uma nova observação \tilde{x} .

b) Mostre que o valor estimado para nova observação é $\hat{\tilde{x}} = \mathbb{E}[\tilde{x}|x] = n \frac{a_1}{a_1 + b_1}$, onde a_1, b_1 : parâmetros da distribuição a posteriori.

c) Mostre que $\text{Var}[\tilde{x}|x] = a_1 b_1 (n^2 + a_1 + b_1) / [(a_1 + b_1)^2 (a_1 + b_1 + 1)]$, onde a_1, b_1 : parâmetros da distribuição a posteriori.

21) Considere uma v.a. com distribuição normal inversa, dada por

$$f(x|\lambda, \mu) = \left[\frac{\lambda}{2\pi x^3} \right]^{1/2} \exp \left\{ \frac{-\lambda(x - \mu)^2}{2\mu^2 x} \right\}, \quad x > 0.$$

a) Encontre a priori de Jeffreys para o modelo.

b) Sugira uma priori conjugada para λ , considerando μ : constante.

22) Suponha que uma a.a. $\mathbf{x} = (24, 21, 10, 2.5, 9.8) | \mu, \sigma^2 \sim \text{Log-Normal}(\mu, \sigma^2)$ iid, onde $\sigma^2 = 1$. Seja $\mu \sim N(8, 1)$.

a) Proponha um modelo conjugado e obtenha a distribuição a posteriori.

b) Obtenha os estimadores bayesianos sob perdas quadrática, absoluta e 0-1.

c) De acordo com todas as fontes de informação, a distribuição a posteriori resultante carrega muita incerteza sobre μ ? Justifique.

Nota: Distribuição LogNormal: $f(x|\mu, \sigma^2) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$. De forma que se uma v.a. $W \sim N(\mu, \sigma^2)$ então $e^W \sim \text{Log-Normal}(\mu, \sigma^2)$.

23) Suponha que uma amostra $x_1, \dots, x_n | \lambda \sim \text{Poisson}(\lambda)$ iid. $\lambda \sim G(\alpha, \beta)$.

a) Determine a distribuição a posteriori.

b) Encontre a distribuição preditiva de uma nova observação \tilde{x} .

24) Considere a questão anterior. Suponha que $\sum_{i=1}^7 x_i = 20$, $\alpha = 5$ e $\beta = 8$.

a) Qual o valor previsto para a oitava observação?

b) Qual o desvio padrão associado a este valor?

c) Se alguém afirmasse que a nova observação será $\tilde{x} = 10$, você endossaria tal afirmação?

25) Suponha que uma a.a. $x|\theta \sim \text{Bin}(n, \theta)$ iid, onde $0 < \theta < 1$ é o parâmetro de interesse, onde $\theta \sim \text{Beta}(\alpha, \beta)$ (α, β : conhecidos).

a) Determine a distribuição preditiva para uma nova observação \tilde{x} .

b) Mostre que o valor estimado para nova observação é $\hat{\tilde{x}} = \mathbb{E}[\tilde{x}|x] = n \frac{a_1}{a_1 + b_1}$, onde a_1, b_1 : parâmetros da distribuição a posteriori.

c) Mostre que $\text{Var}[\tilde{x}|x] = a_1 b_1 (n^2 + a_1 + b_1) / [(a_1 + b_1)^2 (a_1 + b_1 + 1)]$, onde a_1, b_1 : parâmetros da distribuição a posteriori.

26) Considere uma v.a. com distribuição normal inversa, dada por

$$f(x|\lambda, \mu) = \left[\frac{\lambda}{2\pi x^3} \right]^{1/2} \exp \left\{ \frac{-\lambda(x - \mu)^2}{2\mu^2 x} \right\}, \quad x > 0.$$

a) Encontre a priori de Jeffreys para o modelo.

b) Sugira uma priori conjugada para λ , considerando μ : constante.

27) Suponha que uma a.a. $x_1, \dots, x_n | \theta \sim \text{Ber}(\theta)$ iid. A distribuição à priori para θ é dada pela combinação (mistura) de duas distribuições Beta com os mesmos pesos, isto é

$$p(\theta) = 0.5 \times p_1(\theta|a_1, b_1) + 0.5 \times p_2(\theta|a_2, b_2),$$

onde $p_i(\theta|a_i, b_i) = \frac{\Gamma(a_i + b_i)}{\Gamma(a_i)\Gamma(b_i)} \theta^{a_i-1} (1 - \theta)^{b_i-1}$ ($i = 1, 2$). Dados: $\sum_{i=1}^{12} x_i = 9$, $a_1 = b_1 = 5$, $a_2 = 8$ e $b_2 = 1$.

a) Encontre a distribuição a posteriori para θ . Existe conjugação?

b) Obtenha o estimador bayesiano sob perda 0–1.

28) Encontre a priori de Jeffreys para os seguintes modelos:

a) $X \sim \text{Gama-Inversa}(a, b)$, onde a : conhecido.

b) $X \sim \text{Kumaraswamy}(a, b)$, onde b : conhecido.

c) $X \sim \text{Geometrica}(\theta)$.

29) Considere uma a.a. $x_1, \dots, x_n | k \sim \text{Pareto}(x_m, k)$ iid, onde x_m : conhecido. Suponha que $k \sim \text{Gama}(\alpha, \beta)$. (a) Mostre que o modelo é conjugado e encontre a distribuição a posteriori. (b) Encontre o estimador Bayesiano sob perda quadrática.

30) Seja y_i o número de acidentes fatais em uma rodovia no ano i ($i = 1, \dots, 10$). Seja θ o número médio de acidentes por ano. Use um modelo Poisson–Gama para o problema. Dados: $\mathbf{y} = (3, 2, 3, 1, 2, 4, 3, 1, 0, 3)$. Assuma uma priori não-informativa para θ : $\theta \sim G(0, 0)$.

a) Encontre a distribuição a posterior e uma estimativa a posterior para θ sob perda 0–1 (moda).

b) Encontre a variância a posterior de θ .

c) O que se aprende sobre a taxa de acidentes por ano, após observar os dados?