Resolução dos Exercícios das Listas

Inferência Bayesiana | 1º/2024

Tailine J. S. Nonato

July 7, 2024

Lista 2

Questão 1

O seu professor chega na sala de aula e mostra uma moeda. Você suspeita que a moeda possa ser falsa e ter duas caras. Considere a priori probabilidades iguais para os eventos da moeda ser falsa ou ser honesta (i.e. uma moeda bem equilibrada).

(i) Calcule a sua probabilidade de obter cara num lançamento dessa moeda.

Seja F o evento da moeda ser falsa e C o evento de obter cara. Tem-se que:

Pela regra da probabilidade total, tem-se:

$$P(C) = P(C|F)P(F) + P(C|F^c)P(F^c) = 1 \cdot 0.5 + 0.5 \cdot 0.5 = 0.75.$$

(ii) Se o professor lançar a moeda e o resultado for cara, qual é agora a probabilidade dela ser falsa?

Pela regra de Bayes, tem-se:

$$P(F|C) = \frac{P(C|F)P(F)}{P(C)} = \frac{1 \cdot 0.5}{0.75} = \frac{2}{3}.$$

(iii) Se o professor lançar a moeda n vezes e obter n caras, qual é a probabilidade dela ser falsa? Estude o comportamento desta probabilidade para n grande.

Seja C_n o evento de obter cara em n lançamentos. Pela regra de Bayes, tem-se:

$$P(F|C_n) = \frac{P(C_n|F)P(F)}{P(C_n)} = \frac{1^n \cdot 0.5}{0.75^n} = \left(\frac{2}{3}\right)^n.$$

Assim, a probabilidade de a moeda ser falsa decresce exponencialmente com o número de lançamentos.

(iv) Se o professor lançar a moeda uma vez e obter cara, qual é a probabilidade do próximo lançamento ser cara?

Seja C_1 o evento de obter cara no primeiro lançamento e C_2 o evento de obter cara no segundo lançamento. Pela regra da probabilidade total, tem-se:

$$P(C_2) = P(C_2|F)P(F) + P(C_2|F^c)P(F^c) = 1 \cdot 0.5 + 0.5 \cdot 0.5 = 0.75.$$

Pela regra de Bayes, tem-se:

$$P(C_2|C_1) = \frac{P(C_1|C_2)P(C_2)}{P(C_1)} = \frac{P(C_1|C_2)P(C_2)}{P(C_1|C_2)P(C_2) + P(C_1|C_2^c)P(C_2^c)} = \frac{1 \cdot 0.75}{1 \cdot 0.75 + 0.5 \cdot 0.25} = \frac{3}{4}.$$

Portanto, a probabilidade do próximo lançamento ser cara é de 3/4.

(v) Explique porque é falso neste contexto a afirmação "os dois lançamentos da moeda são independentes", e explique qual seria a afirmação correta.

Os lançamentos da moeda não são independentes, pois o resultado de um lançamento afeta o resultado do próximo lançamento. A afirmação correta seria "os lançamentos da moeda são condicionalmente independentes dado o estado da moeda". Ou seja, se soubermos se a moeda é falsa ou honesta, os lançamentos da moeda são independentes.

Questão 2

Seja $y_1, y_2, ..., y_n$ uma amostra da distribuição de Bernoulli com probabilidade de sucesso θ e considere uma distribuição a priori uniforme para θ.

(i) Ache a distribuição a posteriori de θ e a sua média e variância.

Entende-se que:

$s = \sum_{i=1}^{n} y_i \sim Binomial(n, \theta)$

$P(\theta) \sim Uniforme(0, 1)$

Assim, a priori é dada por:

$P(\theta) = \frac{1}{b-a} = \frac{1}{1-0} = 1, \qquad 0 < \theta < 1$

A posteriori pode ser obtida por:

$\theta|x \propto P(x|\theta)P(\theta)$

$\theta|x \propto \theta^{s}(1-\theta)^{n-s} * 1$

Manipulando a expressão, tem-se:

$\theta|x \propto \theta^{s+1-1}(1-\theta)^{n-s+1-1}$

Assim, a posteriori é dada por:

$\theta|x \sim Beta(s+1, n-s+1)$

Logo, pela distribuição, a média e a variância da posteriori são dadas por:

$E(\theta|x) = \frac{s+1}{s+1+n-s+1} = \frac{s+1}{n+2}$

$Var(\theta|x) = \frac{(s+1)(n-s+1)}{(s+1+n-s+1)^2(s+1+n-s+1+1)} = \frac{(s+1)(n-s+1)}{(n+2)^2(n+3)}$

(ii) Mostre que é possível expressar a esperança a posteriori de θ da forma $(1-w)E(\theta) + w\hat{\theta}$, onde $E(\theta)$ e $\hat{\theta}$ são respectivamente a esperança a priori e a estimativa máximo verossímil de θ, e interprete este resultado.

A esperança a priori de θ é dada por:

$E(\theta) = \frac{1}{2}$

A estimativa máximo verossímil de θ é dada por:

$\hat{\theta} = \frac{s}{n}$

Assim, a esperança a posteriori de θ pode ser expressa como:

$E(\theta|x) = \frac{s+1}{n+2} = \frac{n}{n+2} \cdot \frac{s}{n} + \frac{2}{n+2} \cdot \frac{1}{2}$

$E(\theta|x) = \frac{n}{n+2}\hat{\theta} + \frac{2}{n+2}E(\theta)$

Sendo $w = \frac{n}{n+2}$, tem-se:

$E(\theta|x) = (1-w)E(\theta) + w\hat{\theta}$

(iii) Se y_{n+1} é uma observação futura deste processo de Bernoulli, ache a distribuição preditiva $p(y_{n+1}|y_1, ..., y_n)$.

A distribuição preditiva é dada por:

$p(y_{n+1}|y_1, ..., y_n) = \int p(y_{n+1}|\theta)p(\theta|y_1, ..., y_n)d\theta$

$p(y_{n+1}|y_1, ..., y_n) = \int \theta^{y_{n+1}}(1-\theta)^{1-y_{n+1}} \frac{s+1}{n+2}\theta^{s}(1-\theta)^{n-s}d\theta$

$p(y_{n+1}|y_1, ..., y_n) = \frac{s+1}{n+2} \int \theta^{s+y_{n+1}}(1-\theta)^{n-s+1-y_{n+1}}d\theta$

$p(y_{n+1}|y_1, ..., y_n) = \frac{s+1}{n+2} \frac{\Gamma(s+y_{n+1}+1)\Gamma(n-s+1-y_{n+1}+1)}{\Gamma(n+2)}$

Questão 4

No exercício 2, calcule

(i) a estimativa bayesiana para Perda Quadrática

- A perda quadrática é dada por:

$L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$

- Assim, a estimativa bayesiana para perda quadrática é dada por:

$E[L(\theta, \hat{\theta})|x] = E[(\theta - \hat{\theta})^2|x]$

$E[L(\theta, \hat{\theta})|x] = E[(\theta - \hat{\theta})^2|x]$

$E[L(\theta, \hat{\theta})|x] = E[\theta^2 - 2\theta\hat{\theta} + \hat{\theta}^2|x]$

$E[L(\theta, \hat{\theta})|x] = E[\theta^2|x] - 2E[\theta\hat{\theta}|x] + E[\hat{\theta}^2|x]$

$E[L(\theta, \hat{\theta})|x] = E[\theta^2|x] - 2\hat{\theta}E[\theta|x] + \hat{\theta}^2$

$E[L(\theta, \hat{\theta})|x] = Var(\theta|x) + E[\theta|x]^2 - 2\hat{\theta}E[\theta|x] + \hat{\theta}^2$

$E[L(\theta, \hat{\theta})|x] = \frac{(s+1)(n-s+1)}{(n+2)^2(n+3)} + \left(\frac{s+1}{n+2}\right)^2 - 2\hat{\theta}\frac{s+1}{n+2} + \hat{\theta}^2$

$E[L(\theta, \hat{\theta})|x] = \frac{(s+1)(n-s+1)}{(n+2)^2(n+3)} + \left(\frac{s+1}{n+2}\right)^2 - 2\frac{s}{n}\frac{s+1}{n+2} + \left(\frac{s}{n}\right)^2$

(ii) o limite da estimativa bayesiana para Perda Zero-Um quando $\epsilon \to \theta$.

- A perda zero-um é dada por:

$L(\theta, \hat{\theta}) = I(\theta \neq \hat{\theta})$

- Assim, o limite da estimativa bayesiana para perda zero-um é dado por:

$E[L(\theta, \hat{\theta})|x] = E[I(\theta \neq \hat{\theta})|x]$

$E[L(\theta, \hat{\theta})|x] = P(\theta \neq \hat{\theta}|x)$

$E[L(\theta, \hat{\theta})|x] = 1 - P(\theta = \hat{\theta}|x)$

No caso especial que $n = 12, s = \sum_{i=1}^{12} y_i = 9$, calcule

(iii) a estimativa bayesiana sob Perda Absoluta

- A perda absoluta é dada por:

$L(\theta, \hat{\theta}) = |\theta - \hat{\theta}|$

- Assim, a estimativa bayesiana para perda absoluta é dada por:

$E[L(\theta, \hat{\theta})|x] = E[|\theta - \hat{\theta}|^2|x]$

(iv) um intervalo HPD com nível 99%

- O intervalo HPD é dado por:

$HPD = [\theta_{(1-\alpha)/2}, \theta_{(1+\alpha)/2}]$

- Assim, o intervalo HPD com nível 99% é dado por:

$HPD = [\theta_{0.005}, \theta_{0.995}]$

Questão 8

Suponha que (x_1, x_2, x_3) dado p_1, p_2, p_3 segue uma distribuição Multinomial com parâmetros n e (p_1, p_2, p_3), onde $p_i \geq 0$ e $p_1 + p_2 + p_3 = 1$, e que, a priori, (p_1, p_2, p_3) segue uma distribuição de Dirichlet com parâmetros $(\alpha_1, \alpha_2, \alpha_3)$.

(i) Ache a distribuição a posteriori de p_1, p_2, p_3 e as distribuições a posteriori marginais de p_i $(i = 1, 2, 3)$

- Entende-se que:

$(x_1, x_2, x_3) \sim Multinomial(n, (p_1, p_2, p_3))$

$P(p_1, p_2, p_3) \sim Dirichlet(\alpha_1, \alpha_2, \alpha_3)$

- Assim, a priori é dada por:

$P(p_1, p_2, p_3) = \frac{1}{B(\alpha_1, \alpha_2, \alpha_3)} p_1^{\alpha_1 - 1} p_2^{\alpha_2 - 1} p_3^{\alpha_3 - 1}$

- A posteriori pode ser obtida por:

$p_1, p_2, p_3|x \propto P(x|p_1, p_2, p_3)P(p_1, p_2, p_3)$

$p_1, p_2, p_3|x \propto p_1^{x_1} p_2^{x_2} p_3^{x_3} p_1^{\alpha_1 - 1} p_2^{\alpha_2 - 1} p_3^{\alpha_3 - 1}$

$p_1, p_2, p_3|x \propto p_1^{x_1 + \alpha_1 - 1} p_2^{x_2 + \alpha_2 - 1} p_3^{x_3 + \alpha_3 - 1}$

- Assim, a posteriori é dada por:

$p_1, p_2, p_3|x \sim Dirichlet(x_1 + \alpha_1, x_2 + \alpha_2, x_3 + \alpha_3)$

- As distribuições a posteriori marginais de p_i são dadas por:

$p_i|x = \int p_1, p_2, p_3|x dp_1 dp_2 dp_3$

$p_i|x = \int p_1, p_2, p_3|x \prod_{j \neq i} dp_j$

$p_i|x = \int p_1^{x_1+\alpha_1-1} p_2^{x_2+\alpha_2-1} p_3^{x_3+\alpha_3-1} dp_1 dp_2 dp_3$

$p_i|x = \int p_i^{x_i+\alpha_i-1} (1-p_i)^{n-x_i+\alpha_{-i}-1} dp_i$

- Logo, é possível observar que:

$p_i|x = \frac{B(x_i+\alpha_i, n-x_i+\alpha_{-i})}{B(\alpha_i, \alpha_{-i})}$

- Então, as distribuições a posteriori marginais de p_i são dadas por:

$p_i|x \sim Beta(x_i + \alpha_i, n - x_i + \alpha_{-i})$

(ii) Calcule as estimativas bayesianas de p_i e de $p_j - p_i$ sob Perda Quadrática ($i, j = 1, 2, 3, i \neq j$).

- A estimativa bayesiana de p_i sob perda quadrática é dada por:

$E[L(p_i, \hat{p_i})|x] = E[(p_i - \hat{p_i})^2|x]$

$E[L(p_i, \hat{p_i})|x] = E[(p_i - E(p_i|x))^2|x]$

$E[L(p_i, \hat{p_i})|x] = Var(p_i|x)$

$E[L(p_i, \hat{p_i})|x] = \frac{(x_i+\alpha_i)(n-x_i+\alpha_{-i})}{(n+\alpha_{-i})^2(n+\alpha)}$

- A estimativa bayesiana de $p_j - p_i$ sob perda quadrática é dada por:

$E[L(p_j - p_i, \hat{p_j} - \hat{p_i})|x] = E[(p_j - p_i - \hat{p_j} + \hat{p_i})^2|x]$

$E[L(p_j - p_i, \hat{p_j} - \hat{p_i})|x] = E[(p_j - \hat{p_j} - p_i + \hat{p_i})^2|x]$

$E[L(p_j - p_i, \hat{p_j} - \hat{p_i})|x] = E[(p_j - \hat{p_j})^2 + (p_i - \hat{p_i})^2 - 2(p_j - \hat{p_j})(p_i - \hat{p_i})|x]$

$E[L(p_j - p_i, \hat{p_j} - \hat{p_i})|x] = E[(p_j - \hat{p_j})^2|x] + E[(p_i - \hat{p_i})^2|x] - 2E[(p_j - \hat{p_j})(p_i - \hat{p_i})|x]$

$E[L(p_j - p_i, \hat{p_j} - \hat{p_i})|x] = Var(p_j|x) + Var(p_i|x) - 2Cov(p_j, p_i|x)$

$E[L(p_j - p_i, \hat{p_j} - \hat{p_i})|x] = \frac{(x_j+\alpha_j)(n-x_j+\alpha_{-j})}{(n+\alpha_{-j})^2(n+\alpha)} + \frac{(x_i+\alpha_i)(n-x_i+\alpha_{-i})}{(n+\alpha_{-i})^2(n+\alpha)} - 2\frac{\alpha_{ij}}{(n+\alpha_{-i})(n+\alpha_{-j})}$

- Sendo $\alpha_{ij} = \alpha_i \alpha_j / (\alpha + 1)$, tem-se:

$E[L(p_j - p_i, \hat{p_j} - \hat{p_i})|x] = \frac{(x_j+\alpha_j)(n-x_j+\alpha_{-j})}{(n+\alpha_{-j})^2(n+\alpha)} + \frac{(x_i+\alpha_i)(n-x_i+\alpha_{-i})}{(n+\alpha_{-i})^2(n+\alpha)} - 2\frac{\alpha_i \alpha_j}{(n+\alpha_{-i})(n+\alpha_{-j})(n+\alpha)}$

Questão 11

É conhecido que 25% dos pacientes de um certo grupo que sofrem de enxaqueca melhoram após duas horas de serem tratados com um placebo. Para verificar se uma droga nova é melhor que o placebo, $n = 20$ pacientes foram tratados com o placebo e verificou-se que após duas horas $s = 8$ deles relataram ter melhorado. Seja θ a probabilidade de um paciente tratado com a droga nova melhorar após duas horas.

(i) Especifique a hipótese nula H_0 e a alternativa H_1;

$$\begin{cases} H_0 : \theta \leq 0.25 \\ H_1 : \theta > 0.25 \end{cases}$$

(ii) Usando a distribuição a priori "não informativa" $\theta \sim Uniforme(0, 1)$, calcule as chances relativas a priori e a posteriori de H_1 e o correspondente Fator de Bayes;

- Entende-se que:

$\sum X_i \sim^{iid} Binomial(20, \theta)$

$P(\theta) \sim Uniforme(0, 1)$

- Assim, a priori é dada por:

$P(\theta) = \frac{1}{b-a} = \frac{1}{1-0} = 1, \qquad 0 < \theta < 1$

- Logo, as probabilidades a priori são dadas por:

$P(H_0) = P(\theta \leq 0.25) = \int_0^{0.25} 1 d\theta = 0.25$

$P(H_1) = P(\theta > 0.25) = \int_{0.25}^{1} 1 d\theta = 0.75$

- Odds a priori

$odds(\theta \leq 0.25) = \frac{P(\theta \leq 0.25)}{P(\theta > 0.25)} = \frac{0.25}{0.75} = \frac{1}{3}$

$odds(\theta > 0.25) = \frac{P(\theta > 0.25)}{P(\theta \leq 0.25)} = \frac{0.75}{0.25} = 3$

- Pode-se concluir assim que a priori a chance de H_1 é 3 vezes maior que a chance de H_0.
- Sabe-se que:

$\theta | x \propto P(x|\theta)P(\theta)$

$\theta | x \propto \theta^s (1 - \theta)^{n-s} * 1$

- Manipulando a expressão, tem-se:

$\theta|x \propto \theta^{s+1-1}(1-\theta)^{n-s+1-1}$

- Assim, a posteriori é dada por:

$\theta|x \sim Beta(s+1, n-s+1)$

- Com $s = 8$ e $n = 20$, tem-se:

$\theta|x \sim Beta(9, 13)$

- Logo, as probabilidades a posteriori são dadas por:

$P(H_0|x) = P(\theta \leq 0.25|s = 8)$

```
ph0<-round(pbeta(0.25, 9, 13),3)
ph0
```

```
[1] 0.056
```

$P(H_1|x) = P(\theta > 0.25|s = 8)$

```
ph1<-round(1 - pbeta(0.25, 9, 13),3)
ph1
```

```
[1] 0.944
```

- Odds a posteriori

$odds(\theta \leq 0.25) = 0.056/\ 0.944 = 0.059$

$odds(\theta > 0.25) = 0.944/\ 0.056 = 16.857$

- Pode-se concluir assim que a posteriori a chance de H_1 é 16.86 vezes maior que a chance de H_0.

- Fator de Bayes

$\beta_{1,0} = \frac{odds(\theta_1|x)}{odds(\theta_1)} = 16.8/3 = 5.6$ em favor de H_1.

$\beta_{0,1} = \frac{odds(\theta_0|x)}{odds(\theta_0)} = 3/16.8 = 0.18$ em favor de H_0.

- Assim, a interpretação do Fator de Bayes é que a evidência a favor de H_1 é 5.6 vezes mais forte que a evidência a favor de H_0.

(iii) Seja $d = 1$ a decisão de rejeitar H_0 e $d = 0$ a de não rejeitar. Considere a função de perda de Neyman para a qual é 5 vezes mais custoso rejeitar H_0 quando ela é verdadeira do que não rejeitar quando ela é falsa [isto é, $L(d = 1, \theta \in H_0) = 5L(d = 0, \theta \notin H_0)$, $L(d = 1, \theta \notin H_0) = L(d = 0, \theta \in H_0) = 0$]. Calcule a decisão ótima a posteriori;

$$d = \begin{cases} 1 & \text{se } L(d=1|x) < L(d=0|x) \\ 0 & \text{se } L(d=1|x) > L(d=0|x) \end{cases}$$

- Perda a posteriori

$L(d = 1|x) = E[L(d = 1|x, \theta)]$

$L(d = 0|x) = E[L(d = 0|x, \theta)]$

- Com $L(d = 1, \theta \in H_0) = 5L(d = 0, \theta \notin H_0)$, tem-se:

$L(d = 1|x) = 5L(d = 0|x)$

- Assim, a decisão ótima a posteriori é dada por:

$$\begin{cases} d = 1, & se \quad 5P(H_0|x) < P(H_1|x) \\ d = 0, & se \quad 5P(H_0|x) > P(H_1|x) \end{cases}$$

(iv) É razoável chamar essa distribuição a priori de "não informativa" nesse problema? Se a sua resposta for negativa, sugira uma outra distribuição a priori e refaça os cálculos anteriores.

- Não é razoável, visto que ao assumir uma distribuição a priori uniforme, supõe-se que todas as probabilidades são igualmente prováveis, gerando um viés na análise. Uma distribuição a priori mais adequada seria a distribuição Beta(0.5, 0.5), que assume que a probabilidade de sucesso é igualmente provável de ser maior ou menor que 0.5.
- Refazendo os cálculos anteriores, tem-se:
- A priori é dada por:

$P(\theta) = \frac{\theta^{0.5-1}(1-\theta)^{0.5-1}}{B(0.5,0.5)}$

- Logo, as probabilidades a priori são dadas por:

$P(H_0) = P(\theta \leq 0.25) = \int_0^{0.25} \frac{\theta^{0.5-1}(1-\theta)^{0.5-1}}{B(0.5,0.5)} d\theta$

```
ph0<-round(pbeta(0.25, 0.5, 0.5),3)
ph0
```

```
[1] 0.333
```

$P(H_1) = P(\theta > 0.25) = \int_{0.25}^{1} \frac{\theta^{0.5-1}(1-\theta)^{0.5-1}}{B(0.5,0.5)} d\theta$

```
ph1<-round(1 - pbeta(0.25, 0.5, 0.5),3)
ph1
```

```
[1] 0.667
```

- Odds a priori

$odds(\theta \leq 0.25) = \frac{P(\theta \leq 0.25)}{P(\theta > 0.25)} = \frac{0.25}{0.75} = \frac{1}{3}$

$odds(\theta > 0.25) = \frac{P(\theta > 0.25)}{P(\theta \leq 0.25)} = \frac{0.75}{0.25} = 3$

- Pode-se concluir assim que a priori a chance de H_1 é 3 vezes maior que a chance de H_0.
- Sabe-se que:

$\theta|x \propto P(x|\theta)P(\theta)$

$\theta|x \propto \theta^s(1-\theta)^{n-s} * \frac{\theta^{0.5-1}(1-\theta)^{0.5-1}}{B(0.5,0.5)}$

- Manipulando a expressão, tem-se:

$\theta|x \propto \theta^{s+0.5-1}(1-\theta)^{n-s+0.5-1}$

- Assim, a posteriori é dada por:

$\theta|x \sim Beta(s+0.5, n-s+0.5)$

- Com $s = 8$ e $n = 20$, tem-se:

$\theta|x \sim Beta(8.5, 12.5)$

- Logo, as probabilidades a posteriori são dadas por:

$P(H_0|x) = P(\theta \leq 0.25|s = 8)$

```r
ph0<-round(pbeta(0.25, 8.5, 12.5),3)
ph0
```

```
[1] 0.066
```

$P(H_1|x) = P(\theta > 0.25|s = 8)$

```r
ph1<-round(1 - pbeta(0.25, 8.5, 12.5),3)
ph1
```

```
[1] 0.934
```

- Odds a posteriori

$odds(\theta \leq 0.25) = 0.066/\ 0.934 = 0.071$

$odds(\theta > 0.25) = 0.934/\ 0.066 = 14.152$

- Fator de Bayes

$\beta_{1,0} = \frac{odds(\theta_1|x)}{odds(\theta_1)} = 14.15/0.071 = 199.3$ em favor de H_1.

$\beta_{0,1} = \frac{odds(\theta_0|x)}{odds(\theta_0)} = 0.071/14.15 = 0.005$ em favor de H_0.

- Conclui-se assim que a evidência a favor de H_1 é 199.3 vezes mais forte que a evidência a favor de H_0. Logo, a escolha da distribuição a priori gera resultados diferentes, de forma que a distribuição Beta(0.5, 0.5) fornece uma evidência mais forte a favor de H_1 e é mais adequada para o problema.