Aluno: Cleiton Moya de Almeida
EST5104 - Inferência Bayesiana
Prof. Josemar Rodrigues
07/04/2025

Lista 2

Enunciado

- Modelo estatístico: $X_i|\theta \sim \text{Ber}(\theta)$, $i = 1, 2, \ldots$, v.a.i.;
- Priori: $\theta \sim \text{Beta}(a, b)$;
- Item 1: Determine a *posterior predictive distribution*;
- Item 2: Determine $E[\tilde{X} = \tilde{x}|\boldsymbol{X} = \boldsymbol{x}]$, $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^\intercal$;
- Item 3: Sugira um procedimento para determinar os hiperparâmetros a e b;
- Item 4: Obter o estimador de Bayes de $1/\theta$.

Resolução

Item 1

Seja $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^\intercal$ o vetor dos dados observados e $\tilde{x}$ uma observação futura. Para calcularmos a distribuição preditiva a posteriori $p(\tilde{x}|\boldsymbol{x})$, necessitamos primeiramente calcular a distribuição a posteriori de θ, $p(\theta|\boldsymbol{x})$.

Distribuição a posteriori de θ

A distribuição a posteriori $p(\theta|\boldsymbol{x})$ é obtida pela fórmula de Bayes:

$$p(\theta|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\theta)p(\theta)}{p(\boldsymbol{x})}, \tag{1}$$

onde $p(\boldsymbol{x}|\theta)$ é a função de verossimilhança, $p(\theta)$ é a distribuição a priori e $p(\boldsymbol{x})$ é a probabilidade total a priori (*prior predictive probability*).

A função de verossimilhança é obtida à partir do modelo estatístico:

$$p(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^(1 - x_i) \tag{2}$$

Para facilitar os cálculos, seja $T = \sum_{i=1}^{n} x_i$. Podemos, re-escrever a eq. (4) como

$$p(\boldsymbol{x}|\theta) = \theta^T (1-\theta)^{(n-T)}. \tag{3}$$

Por sua vez, a função de densidade da priori é

$$p(\theta) = c\theta^{a-1}(1-\theta)^{b-1} \tag{4}$$

sendo $c = B(a, b)$ a constante de normalização, onde $B(a, b)$ é a função beta. Com isso, a posteriori fica

$$p(\theta|\boldsymbol{x}) = \left(\frac{c}{p(\boldsymbol{x})}\right)\theta^{T+a-1}(1-\theta)^{n-T+b-1} \tag{5}$$

$$\tag{6}$$

Calculando agora $p(\boldsymbol{x})$:

$$\begin{aligned}
p(\boldsymbol{x}) &= \int_0^1 p(\boldsymbol{x}|\theta)p(\theta)d\theta &(7)\\
&= \int_0^1 \left[\theta^T(1-\theta)^{n-T}\right]\left[c\theta^{a-1}(1-\theta)^{b-1}\right]d\theta &(8)\\
&= c\int_0^1 \theta^{T+a-1}(1-\theta)^{n-T+b-1}d\theta &(9)\\
&= cB(T+a, n-T+b) &(10)
\end{aligned}$$

Então, a distribuição a posteriori $p(\theta|\boldsymbol{x})$ fica

$$p(\theta|\boldsymbol{x}) = \frac{\theta^{T+a-1}(1-\theta)^{n-T+b-1}}{B(T+a, n-T+b)} \tag{11}$$

De onde notamos que trata-se da própria distribuição Beta, mas com parâmetros atualizados:

$$\theta|\boldsymbol{x} \sim \text{Beta}(T+a, n-T+b). \tag{12}$$

Distribuição preditiva a posteriori

Seja $\tilde{x}$ uma observação futura (um único ponto). Observe que $\tilde{x}$ é uma v.a. de Bernoulli. A distribuição preditiva a posteriori é dada por:

$$\begin{aligned}
p(\tilde{x}=1|\boldsymbol{x}) &= \int_0^1 p(\tilde{x}=1|\theta,\boldsymbol{x})p(\theta|\boldsymbol{x})d\theta &(13)\\
p(\tilde{x}=0|\boldsymbol{x}) &= \int_0^1 p(\tilde{x}=0|\theta,\boldsymbol{x})p(\theta|\boldsymbol{x})d\theta &(14)\\
&= 1 - p(\tilde{x}=1|\boldsymbol{x}) &(15)
\end{aligned}$$

Para calcularmos $p(\tilde{x}=1|\boldsymbol{x})$, adotamo a premissa de que $\tilde{x}$ é condicionalmente independente de $\boldsymbol{x}$ na presença de θ ($\tilde{x} \perp \boldsymbol{x} \mid \theta$). Então:

$$
\begin{aligned}
p(\tilde{x}=1|\boldsymbol{x}) &= \int_0^1 p(\tilde{x}=1|\theta)p(\theta|\boldsymbol{x})d\theta & (16)\\
&= \int_0^1 [\theta]\left[\frac{\theta^{T+a-1}(1-\theta)^{n-T+b-1}}{B(T+a,n-T+b)}\right]d\theta & (17)\\
&= \frac{1}{B(T+a,n-T+b)}\int_0^1 \theta^{T+a}(1-\theta)^{n-T+b-1}d\theta & (18)\\
&= \frac{B(T+a+1,\ n-T+b)}{B(T+a,\ n-T+b)} & (19)\\
& & (20)
\end{aligned}
$$

Item 2

Como $\tilde{x}$ é uma v.a. de Bernoulli, a esperança de $\tilde{x}|\boldsymbol{x}$ é dada pelo próprio valor de $p(\tilde{x}=1|\boldsymbol{x})$, o qual já foi calculado no item anterior:

$$
\begin{aligned}
E[\tilde{x} \mid \boldsymbol{x}] &= 1 \times p(\tilde{x}=1|\boldsymbol{x}) + 0 \times p(\tilde{x}=0|\boldsymbol{x}) & (21)\\
&= p(\tilde{x}=1|\boldsymbol{x}) & (22)\\
&= \frac{B(T+a-1,\ n-T+b)}{B(T+a,\ n-T+b)} & (23)\\
& & (24)
\end{aligned}
$$

Item 3

De acordo com Lehman e Casella (seção 4.1, pg. 232), embora o estatístico bayesiano subjetivo talvez não consiga expressar a priori precisamente, ele geralmente possui uma ideia do formato da função densidade: ela pode ser simétrica, assimétrica, com formato “L” ou “U”. Então, os parâmetros a e b podem ser ajustados com o intuito de buscar esta forma da função densidade. Além da forma, se o estatístico tiver uma ideia do valor esperado e da variância da distribuição, é possível ajustar a e b de forma que a distribuição tenha estes valores.

Uma ideia similar, sugerida por Berger (Cap. 3, pg. 64) é selecionar o(s) ponto(s) com valores mais plausíveis"e "menos plausíveis", e então determinar, quais as chances (*odds*) relativas destes pontos, buscando novamente esboçar uma forma para a distribuição a priori.

Item 4

Vamos obter o estimador de Bayes de $1/\theta$ através da definição de valor esperado:

$$
\begin{aligned}
E[1/\theta] &= \int_0^1 \frac{1}{\theta}\ p(\theta|\boldsymbol{x})d\theta && (25)\\
&= \int_0^1 \frac{1}{\theta}\frac{\theta^{T+a-1}(1-\theta)^{n-T+b-1}}{B(T+a,n-T+b)}d\theta && (26)\\
&= \frac{1}{B(T+a,n-T+b)}\int_0^1 \theta^{T+a-2}(1-\theta)^{n-T+b-1} && (27)\\
&= \frac{B(T+a-1,n-T+b)}{B(T+a,n-T+b)} && (28)
\end{aligned}
$$