ALUNO: CLEITON MOYA DE ALMEIDA
EST5104 - INFERÊNCIA BAYESIANA
PROF. JOSEMAR RODRIGUES
14/04/2025

Lista 3 (continuação da Lista 2)

Enunciado

- Modelo estatístico: $X_i|\theta \sim \text{Ber}(\theta)$, $i = 1, 2, \ldots$, v.a.i.;
- Priori: $\theta \sim \text{Beta}(a, b)$;
- Seja $\tilde{T} = \sum_1^{50} \tilde{X}_i$, $T = \sum_1^{100} X_i = 8$, $a = 9$, $b = 90$;
- Item 1: Determine a densidade preditiva de $\tilde{T}$;
- Item 2: Determine o EMV de θ e a moda a posteriori;
- Item 3: Compare graficamente a densidade preditiva de $\tilde{T}$ com $\tilde{T} \sim \text{Bin}(50, \hat{\theta})$, onde $\hat{\theta}$: EMV (clássica). Comente esta comparação.
- Item 4: O que acontece quando utilizamos $a = b = 1$?

Resolução

Item 1

Vamos chamar $n = 100$ e $\tilde{n} = 50$. Para calcularmos a densidade preditiva, necessitamos novamente da densidade a posteriori de θ. Esta densidade já foi determinada na lista anterior (Lista 2):

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

Calculamos a distribuição preditiva através da seguinte integral:

$$p(\tilde{T} = \tilde{t}|\boldsymbol{x}) = \int_0^1 p(\tilde{T} = \tilde{t}|\theta, \boldsymbol{x})p(\theta|\boldsymbol{x})d\theta. \tag{2}$$

Novamente usando a suposição de que $\tilde{T}$ é condicionalmente independente de $\boldsymbol{X}$ na presença de θ ($\tilde{T} \perp \boldsymbol{X} \mid \theta$), temos:

$$p(\tilde{T}=\tilde{t}|\boldsymbol{x}) = \int_0^1 p(\tilde{T}=\tilde{t}|\theta)p(\theta|\boldsymbol{x})d\theta. \tag{3}$$

Agora, $\tilde{T}$ conta o número de sucessos em $\boldsymbol{X}$ ($\tilde{T} = \sum \tilde{X}_i$), então $\tilde{T}$ segue uma distribuição Binomial:

$$p(\tilde{T}=\tilde{t}|\theta) = \binom{\tilde{n}}{\tilde{t}}\theta^{\tilde{t}}(1-\theta)^{\tilde{n}-\tilde{t}}. \tag{4}$$

Então, a distribuição preditiva fica

$$\begin{aligned}
p(\tilde{T}=\tilde{t}|\boldsymbol{x}) &= \int_0^1 p(\tilde{T}=\tilde{t}|\theta)p(\theta|\boldsymbol{x})d\theta && (5)\\
&= \int_0^1 \left[\binom{\tilde{n}}{\tilde{t}}\theta^{\tilde{t}}(1-\theta)^{\tilde{n}-\tilde{t}}\right]\left[\frac{\theta^{T+a-1}(1-\theta)^{n-T+b-1}}{B(T+a,n-T+b)}\right]d\theta && (6)\\
&= \binom{\tilde{n}}{\tilde{t}}\int_0^1 \frac{\theta^{T+\tilde{t}+a-1}(1-\theta)^{n+\tilde{n}-(T+\tilde{t})+b-1}}{B(T+a,n-T+b)}d\theta && (7)\\
&= \binom{\tilde{n}}{\tilde{t}}\frac{B(T+\tilde{t}+a,\ n+\tilde{n}-(T+\tilde{t})+b)}{B(T+a,\ n-T+b)}. && (8)
\end{aligned}$$

Item 2

Estimador EMV

A função de verossimilhança é dada por

$$\begin{aligned}
\mathcal{L}(\theta|\boldsymbol{x}) &= \prod_{i=1}^{n}\theta^{x_i}(1-\theta)^{1-x_i} && (9)\\
&=\theta^{T}(1-\theta)^{n-T} && (10)
\end{aligned}$$

Como X_i segue a distribuição de Bernoulli, sabemos que o estimador de máxima verossimilhança (EMV) é dado por

$$\hat{\theta}_{\text{emv}} = \frac{T}{n} = \frac{8}{100} = 0.08 \tag{11}$$

Moda a posteriori

Como $\theta|\boldsymbol{x} \sim \text{Beta}(\alpha,\beta)$, com $\alpha = T+a$ e $\beta = n-T+b$, temos

$$\begin{aligned}
\hat{\theta}_{\text{map}} &= \frac{\alpha-1}{\alpha+\beta-2} && (12)\\
&= \frac{T+a-1}{n+a+b-2} = \frac{16}{197} \approxeq 0.0812 && (13)
\end{aligned}$$

Verificamos que os dois estimadores (EMV e MAP) possuem valores bem próximos, sendo que $\hat{\theta}_{\text{map}}$ está ligeiramente à direita de $\hat{\theta}_{\text{emv}}$. Isto ocorre devido à influência da priori.

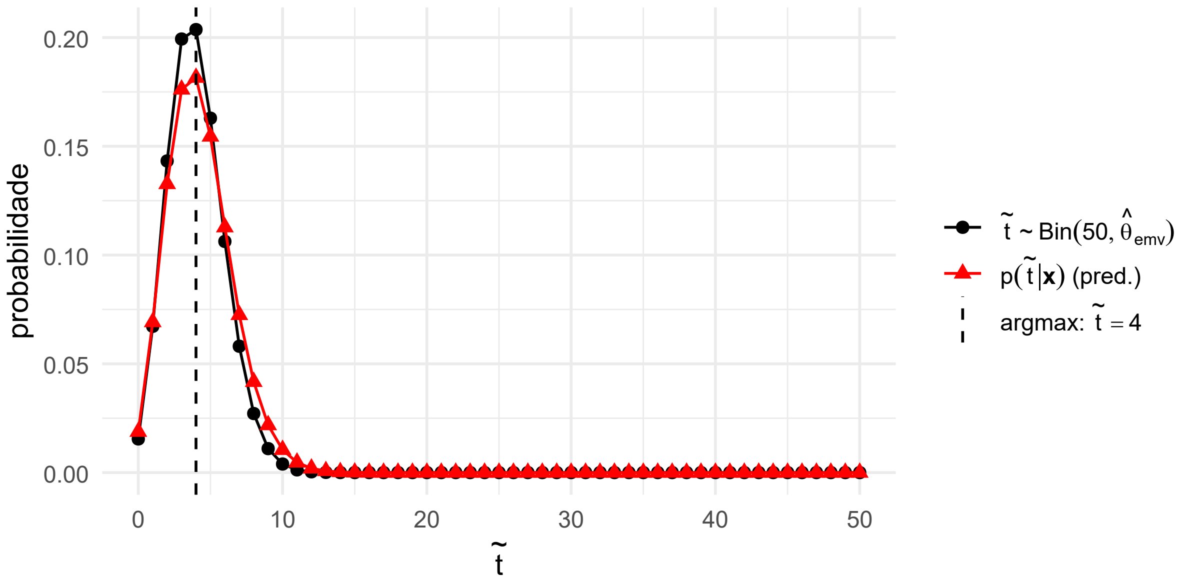

Figura 1: Distribuição preditiva a posteriori (em vermelho) e distribuição binomial com $\hat{\theta}_{emv}$ (em preto). Observe que ambas distribuições possuem o mesmo máximo em $\tilde{t} = 4$, porém a binomial é um pouco mais concentrada neste ponto.

Item 3

Na fig. 1, plotamos a distribuição preditiva a posteriori $p(\tilde{T} = \tilde{t}|\boldsymbol{x})$ (em vermelho) e a distribuição binomial com $\hat{\theta}_{emv}$ (em preto) para 50 observações futuras. Observamos que as duas distribuições apresentam formatos bastante semelhantes, além do mesmo ponto de máximo ($\tilde{t} = 4$). Este ponto previsto de máxima probabilidade era esperado, pois a proporção estimada de sucessos (tanto pelo EMV quando MAP) foi em torno de 0.08 (no caso do EMV, exatamente 0.08). Portanto, em uma janela de 50 observações futuras, esperamos $50 \times 0.08 = 4$ ocorrências de $X = 1$.

Porém, a distribuição binomial apresenta a massa um pouco mais concentrada (pico um pouco maior) em torno da moda. Isto significa que a escolha da priori acrescentou uma incerteza (variância) adicional em torno da moda.

Item 4

No item 4, analisamos os resultados quando utilizarmos os hiperparâmetros $a = b = 1$. Notamos que, com esta configuração, a distribuição beta é equivalente à distribuição uniforme com intervalo $[0, 1]$.

Para melhor entendemos e visualizarmos os resultados, adotamos 3 configurações de prioris:

- Priori 1: $a = 9$, $b = 90$ (moda: $\theta = 8/97 \approxeq 0.082$);
- Priori 2: $a = 21$, $b = 81$ (moda: $\theta = 0.2$);
- Priori 3: $a = 1$, $b = 1$ (distribuição uniforme).

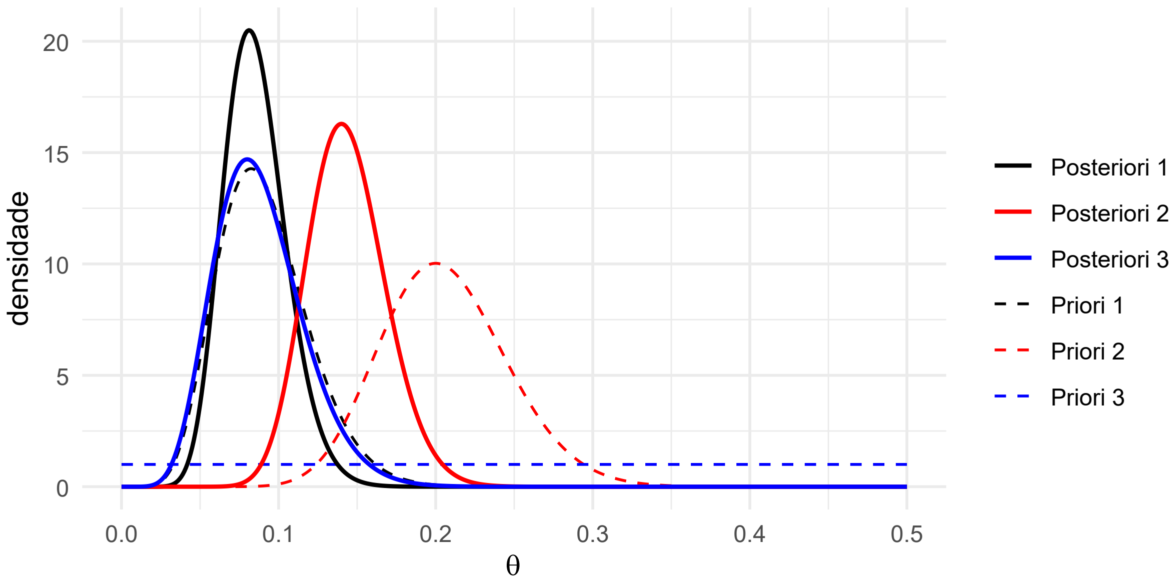

Figura 2: Distribuições a priori e respectivas posteriores. Priori 1: $a = 9$, $b = 90$; Priori 2: $a = 21$, $b = 81$; Priori 3: $a = 1$, $b = 1$. Note que as posteriores 1 e 3 são semelhantes, enquanto a Posteriori 2 possui a moda "puxada" para a direita devido à influência da priori.

Primeiramente, notamos que, com $a = 1$, $b = 1$, o valor da moda a posteriori coincide com o valor da estimativa de máxima verossimilhança:

$$\hat{\theta}_{\text{map}_3} = \frac{T + a - 1}{n + a + b - 2} = \frac{T}{n} = 0.08. \tag{14}$$

Plotamos então, na fig. 2 o gráfico das distribuições a priori e posteriori com estas 3 configurações de hiperparâmetros. No caso da Priori 1, como ela possui moda próxima ao valor de EMV, a Posteriori 1 ficou bastante similar à Posteriori 3. Já a Priori 2 está distante da EMV. Então, ela exerce maior influência na inferência "puxando" a Posteriori 2 para a direita (em relação às posteriores 1 e 3).