Estimador Bayesiano e Funções de Perda

Parte das notas de minha aula didática[1]

Concurso de Professor Assistente

UFRJ - 1995

Hedibert F. Lopes

Professor Titular de Estatística e Econometria

Insper Instituto de Educação e Pesquisa

21 de maio de 2026

[1] As notas escritas à mão viraram esses slides via assistente IA Claude for Mac.

Sumário

Ingredientes Bayesianos

Combinando a informação **a priori** com a informação dos dados (via função de verossimilhança), pelo **Teorema de Bayes**:

$$\pi(\theta \mid x) = \frac{\pi(\theta)\ \ell(\theta;\, x)}{p(x)}$$

Como $p(x) = \int_{\Theta} \pi(\theta)\, \ell(\theta;\, x)\, d\theta$ não depende de θ, podemos escrever:

Proporcionalidade

$$\pi(\theta \mid x)\ \propto\ \pi(\theta)\ \ell(\theta)$$

Exemplo: Modelo Exponencial-Gama

Conjunta $p(x_1, \ldots, x_n, \theta)$

$$x_1, \ldots, x_n | \theta \sim \mathrm{Exp}(\theta), \qquad \theta \sim \mathrm{Gama}(\alpha_0, \beta_0).$$

Núcleos:

$$\pi(\theta) \propto \theta^{\alpha_0 - 1} \, e^{-\beta_0 \, \theta}$$

$$\ell(\theta;\, x) \propto \theta^{n} \, e^{-\theta \sum x_i}$$

Logo:

$$\pi(\theta \mid x) \;\propto\; \theta^{(\alpha_0 + n) - 1} \, e^{-\theta(\beta_0 + \sum x_i)}$$

que é o núcleo da $\mathrm{Gama}(\alpha_0 + n,\ \beta_0 + \sum x_i)$.

Exemplo: Média a Posteriori

A média da distribuição a posteriori $\mathrm{Gama}(\alpha_0 + n,\ \beta_0 + \sum x_i)$ é:

Estimativa de Bayes (perda quadrática - veja mais para frente)

$$E(\theta \mid x) = \frac{\alpha_0 + n}{\beta_0 + \sum x_i}$$

Esta será a base para comparar o estimador Bayesiano com o estimador de máxima verossimilhança (EMV).

Função de Perda

Para cada $\theta \in \Theta$ e $a \in A$ (estimativa), existe um número $L(\theta, a)$ que mede a **perda** gerada quando:

- o verdadeiro valor do parâmetro é θ, e
- a estimativa adotada é a.

Perda Esperada a Priori (para cada a):

$$E[L(\theta, a)] = \int_{\Theta} L(\theta, a)\, \pi(\theta)\, d\theta$$

Perda Esperada a Posteriori

Suponha que x é o valor observado de X antes de estimar θ, e $\pi(\theta \mid x)$ é a posteriori.

Para toda estimativa a, a **perda esperada a posteriori** é:

Definição

$$E[L(\theta, a) \mid x] = \int_{\Theta} L(\theta, a)\, \pi(\theta \mid x)\, d\theta$$

Objetivo

Encontrar a tal que $E(L(\theta, a) \mid x)$ seja **mínima**.

Estimador de Bayes

Define-se o **estimador de Bayes** para θ, denotado $\delta^*(x)$, por:

Estimador de Bayes

$$\delta^*(x) \;=\; \arg\min_{a \in \Theta} \; E[L(\theta, a) \mid x]$$

O estimador ótimo depende da **escolha da função de perda** $L(\theta, a)$.

(1) Perda Quadrática

$$L(\theta, a) = (\theta - a)^2$$

Estimador ótimo

$$\delta^*(x) = E(\theta \mid x) \quad \text{(média a posteriori)}$$

A perda quadrática penaliza erros grandes de forma quadrática — é a função de perda mais utilizada na prática.

(2) Perda Módulo (Valor Absoluto)

$$L(\theta, a) = |\theta - a|$$

Estimador ótimo

$$\delta^*(x) = \text{mediana de } \pi(\theta \mid x)$$

Mais robusta que a perda quadrática — penaliza erros de forma linear e é menos sensível a valores extremos.

(3) Perda 0–1

$$L(\theta, a) = \begin{cases} 0 & \text{se } |\theta - a| < \varepsilon \\ 1 & \text{se } |\theta - a| \geq \varepsilon \end{cases} \qquad \forall\, \varepsilon > 0$$

Estimador ótimo

$$\delta^*(x) = \text{moda de } \pi(\theta \mid x) \quad \text{(MAP)}$$

O estimador de máximo a posteriori (**MAP**) é ótimo sob a perda 0–1.

Resumo das Funções de Perda

Perda	$L(\theta, a)$	$\delta^*(x)$
Quadrática	$(\theta - a)^2$	Média a posteriori
Módulo	$\lvert\theta - a\rvert$	Mediana a posteriori
0–1	$\mathbf{1}[\lvert\theta - a\rvert \geq \varepsilon]$	Moda a posteriori (MAP)

Observação 1: Vício do Estimador de Bayes

O estimador de Bayes é em geral **viciado**.

Exemplo (continuação): sob perda quadrática,

$$\delta^*(x) = E(\theta \mid x) = \frac{\alpha_0 + n}{\beta_0 + \sum x_i} \neq \frac{n}{\sum x_i} = \hat{\theta}_{\mathsf{EMV}}$$

Enquanto o EMV é não-viciado por construção, o estimador de Bayes incorpora a informação a priori e, em geral, é viciado.

Observação 2: Eficiência Assintótica

Já vimos que, para $n \to \infty$:

$$\pi(\theta \mid x) \ \to \ \ell(\theta)$$

e então o estimador Bayesiano é **assintoticamente eficiente**.

Convergência

$$E(\theta \mid x) \ \xrightarrow{n\to\infty} \ \hat{\theta}_{\mathrm{EMV}}$$

Para **grandes amostras**, o efeito da priori desaparece.

Comparação: EMV vs. Bayes

EMV (Estimador de Máxima Verossimilhança):

$$\hat{\theta} = \frac{n}{\sum x_i} = \frac{1}{\bar{x}}$$

Estimador de Bayes (perda quadrática):

$$E(\theta \mid x) = \frac{\alpha_0 + n}{\beta_0 + \sum x_i}$$

Priori vaga: $\alpha_0 \to 0,\ \beta_0 \to 0$

$$E(\theta \mid x) \approx \hat{\theta}_{\mathrm{EMV}}$$

Com priori vaga, os dois estimadores coincidem aproximadamente.

Conclusão

- O **estimador de Bayes** minimiza a perda esperada a posteriori.
- A escolha da **função de perda** determina se o estimador é a média, mediana ou moda da distribuição a posteriori.
- O estimador de Bayes é em geral **viciado**, mas **assintoticamente eficiente**.
- Com **priori vaga** ou **amostras grandes**, o estimador Bayesiano converge para o EMV.