Inferência: Métodos *Bayesianos*

Abraão D. C. Nascimento

Universidade Federal de Pernambuco, Brasil

Pós-graduação em Estatística
03 de Julho de 2025

Índice

Índice

Introdução aos Métodos *Bayesianos*

@.1 Até então, em Inferência Clássica, tinhamos o seguinte padrão:

> *"Métodos (estimação pontual/intervalar e teste de hipóteses) dependentes da função de verossimilhança* $L(\boldsymbol{\theta}; \boldsymbol{x}) = \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta})$*, em que* $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^p$ *é desconhecido (mas fixado) e* $\boldsymbol{x} = (x_1, \ldots, x_n)^\top$ *é uma realização de uma amostra aleatória n-dimensional de X com fdp*[1] *(ou de fmp*[2]*) dada por* $f(x; \boldsymbol{\theta})$*."*

@.2 Na *Inferência Bayesiana*, o parâmetro desconhecido é uma variável aleatória–diga-se $\boldsymbol{v}$ com distribuição (denominada de *distribuição a priori de* $\boldsymbol{v}$) sobre o suporte $\boldsymbol{\Theta}$ e fdp (ou fmp) $h(\boldsymbol{v})$.

> Nos capítulos anteriores $f(x; \boldsymbol{\theta})$ foi denotada como fdp (ou fmp) de X, mas agora $f(x; \boldsymbol{\theta})$ será denotada como a fdp (fmp) de $\{X \mid \boldsymbol{v} = \boldsymbol{\theta}\}$.

[1] 1: fdp: função densidade de probabilidade

[2] 2: fdp: função de massa de probabilidade

@.3 A distribuição a priori $h(\boldsymbol{v})$ frequentemente reflete a crença subjetiva do experimentador, considerando quais valores de $\boldsymbol{v}$ são mais ou menos prováveis.

@.4 O paradigma dos Métodos Bayesianos é:

> *"combinar a evidência sobre* $\boldsymbol{v}$ *da distribuição a priori com a função de verossimilhança por meio do Teorema de Bayes, resultando na distribuição a posteriori"*

$$\left\{[\boldsymbol{v}=\boldsymbol{\theta}]\,,\,[X\mid \boldsymbol{v}=\boldsymbol{\theta}]\,,\,[T(X)\mid \boldsymbol{v}=\boldsymbol{\theta}]\right\}\Rightarrow[\boldsymbol{v}=\boldsymbol{\theta}\mid T(X)=t]$$

Teorema de Bayes

Assuma que os eventos $A_1, A_2, \ldots, A_k$ formam uma partição do espaço amostral Ω e B é um outro evento tal que $\Pr(B) > 0$. Então

$$\Pr(A_j \mid B) = \frac{\Pr(A_j)\Pr(B \mid A_j)}{\sum_{i=1}^{k}\Pr(A_i)\Pr(B \mid A_i)},$$

para $j = 1, 2, \ldots, k$ fixado.

Índice

Distribuições a priori e a posteriori

@.5 Na abordagem Bayesiana, $\boldsymbol{v}$ substitue o parâmetro e sua distribuição é a distribuição a priori. Neste curso, adotaremos que v (ao invés de $\boldsymbol{v}$) assume um modelo contínuo de valor real com fdp $h(v)$, em que $v \in \Theta \subset \mathbb{R}$.

Elementos da Inferência Bayesiana:

(1) **Distribuição a priori:** $h(\theta)$ para $\theta \in \Theta$.

(2) $\boldsymbol{x} = (X_1, \ldots, X_n)^\top$ como uma amostra aleatória de $[\, X \mid V = \theta \,]$.

(3) $T \triangleq T(\boldsymbol{x})$ como uma estatística suficiente (mínima) que assume frequentemente valor real. A fdp ou fmp de $[\, T \mid V = \theta \,]$ será denotada por

$$g(t\,;\,\theta) \triangleq f_{T|V=\theta}(t|\theta),$$

para $t \in \mathcal{T} \subset \mathbb{R}$ e θ fixado.

Elementos da Inferência Bayesiana:

(4) A fdp (ou fmp) do par aleatório (T, V) é dada por

$$f_{T,V}(t,\theta) = f_{T|V=\theta}(t \mid \theta)\, h(\theta) = g(t\,;\,\theta)\, h(\theta),$$

para $t \in \mathcal{T}$ e $\theta \in \Theta$.

(5) A fdp (ou fmp) marginal de T é obtida integrando $f_{T,V}(t,\theta)$ em termos de θ:

$$m(t) \triangleq f_T(t) = \int_{\theta\in\Theta} f_{T,V}(t,\theta)\,\mathrm{d}\theta = \int_{\theta\in\Theta} g(t;\theta)\, h(\theta)\,\mathrm{d}\theta,$$

para todo $t \in \mathcal{T}$.

Elementos da Inferência Bayesiana:

(6) Finalmente, a fdp (ou fmp) de $[\, V \mid T = t\,]$ é dada por

$$k(\theta; t) \triangleq f_{V|T=t}(\theta \mid t) = \frac{g(t;\theta)\, h(\theta)}{m(t)},$$

para todo $t \in \mathcal{T}$ fixado e $\theta \in \Theta$ tal que $m(t) > 0$. Esta fdp (ou fmp) é chamada de distribuição a posteriori.

Nota: A tratabilidade analítica de $k(\theta; t)$ depende da obtenção de $m(t)$. Em alguns casos, as distribuições marginal e a posteriori podem ser obtidas numericamente.

Exercício-1.a

Sejam $X_1, \ldots, X_n$ uma amostra aleatória de $[X \mid V = \theta] \sim$ Bernoulli(θ) tal que $\theta \in (0, 1)$, em que V é uma probabilidade de sucesso tal que $0 < V < 1$. Assumindo a distribuição a priori como $V \sim U(0, 1)$, encontre:

(a) A densidade marginal da estatística suficiente para θ (diga-se $T \triangleq T(\boldsymbol{x})$), em que $\boldsymbol{x} = (X_1, \ldots, X_n)^\top$;

(b) A densidade da distribuição a posteriori de $[V \mid T = t]$.

Exercício-1.b

Sejam $X_1, \ldots, X_n$ uma amostra aleatória de $[X \mid V = \theta] \sim$ Poisson(θ) tal que $\theta > 0$, em que V é o número esperado de ocorrências em um dado intervalo de tempo. Assumindo a distribuição a priori como $V \sim \Gamma(\alpha, \beta)$, encontre:

(a) A densidade marginal da estatística suficiente para θ (diga-se $T \triangleq T(\boldsymbol{x})$), em que $\boldsymbol{x} = (X_1, \ldots, X_n)^\top$;

(b) A densidade da distribuição a posteriori de $[V \mid T = t]$.

Índice

Priori Conjugada

@.6 Se a priori $h_V(\theta) \triangleq h(\theta)$ de V (variável que descreve o parâmetro) é tal que a marginal $m(t) = f_T(t)$ de T no par aleatório (T, V)–em que T é a estatística suficiente minimal para θ–é analiticamente intratável, então não será possível derivar a posteriori $k(\theta; t) = f_{V|T=t}(\theta \mid t)$.

@.7 Para muitas verossimilhanças, podemos formular um tipo especial de priori $h(\theta)$ tal que se tenha simplicidade analítica, chamada *priori conjugada*.

Priori conjugada

Suponha que a fdp à priori de V dada por $h(\theta)$ pertence a uma família de fdp's $\mathbb{P}$. Então $h(\theta)$ é chamada de priori conjugada para V se só se a posteriori $k(\theta; t)$ também pertence a família $\mathbb{P}$.

Exercício-2.a

Sejam $X_1, X_2, \ldots, X_n$ uma amostra aleatória de $[X \mid V = \theta] \sim \text{Bernoulli}(\theta)$ tal que $\theta \in (0, 1)$. Encontre distribuição a priori conjugada.

Exercício-2.b

Sejam $X_1, X_2, \ldots, X_n$ uma amostra aleatória de $[X \mid V = \theta] \sim \text{Poisson}(\theta)$ tal que $\theta > 0$, em que a variável V descreve a média populacional. Encontre distribuição a priori conjugada.

Índice

Motivação a Estimação Bayesiana

@.8 Neste ponto, exploramos como

"estimar θ sob uma função perda particular."

@.9 <u>Relembrando o contexto:</u> Temos $X_1, \ldots, X_n$ como uma amostra aleatória de $[X \mid V = \theta]$ e T como uma estatística suficiente (minimal) para θ. Seja ainda $\mathcal{T}$ domínio de t.

@.10 Como antes, ao invés de considerar a função de verossimilhança, consideremos a fdp (ou fmp) $g(t;\theta)$ da estatística suficiente T no ponto $T = t$ dado $V = \theta$, i.e., $[T \mid V = \theta]$. Adicionalmente, seja V a distribuição a priori como fdp (fmp) $h(\theta)$ para $\theta \in \Theta$.

@.11 Seja $\delta \equiv \delta(T)$ um estimador arbitrário para $V = \theta$, que assume valor $\delta(t)$ quando se observa $T = t$, $t \in \mathcal{T}$.

@.12 Assumamos que a *perda* em se estimar $V = \theta$ por $\delta(T)$ é dada por

$$L^*(\theta, \delta) \equiv [\delta(T) - \theta]^2,$$

que se chama de *perda erro quadrado*.

@.13 Com base nesta medida é possível definir o erro quadrático médio que corresponde à média ponderada com respeito à fdp (ou fmp) de $[T \mid V = \theta]$, $g(t; \theta)$. A função risco é dada por

$$R^*(\theta, \delta) \equiv \mathrm{E}_{[T|V=\theta]}[L^*(\theta, \delta)] = \int_{\mathcal{T}} L^*(\theta, \delta(t))\, g(t; \theta)\, \mathrm{d}t.$$

Esta medida é chamada de risco frequentista.

@.14 Na prática, dados dois estimadores $\delta_1 \equiv \delta_1(T)$ e $\delta_2 \equiv \delta_2(T)$, se $R^*(\theta, \delta_1) < R^*(\theta, \delta_2)$ para $\theta \in \Theta^* \subset \Theta$ então δ_1 é melhor do que δ_2 neste sub espaço paramétrico.

@.15 Como alternativa ao risco frequentista, tem-se o risco bayesiano:

$$r^*(\theta,\delta) \equiv \mathrm{E}_V[R^*(V,\delta)] = \int_\Theta R^*(\theta,\delta)\,h(\theta)\,\mathrm{d}\theta.$$

@.16 Suponha que D seja uma classe de todos os estimadores de θ cujos riscos bayesianos são finitos. Assim o melhor estimador sob o paradigma Bayesiano será δ^* em D tal que

$$r^*(\theta,\delta^*) = \inf_{\delta\in D} r^*(\theta,\delta).$$

@.17 Este estimador é chamado de *estimador Bayesiano*.

Teorema

O estimador de Bayes $\delta^* \equiv \delta^*(T)$ é determinado tal que o risco a posteriori de $\delta^*(T)$ é o menor possível:

$$\int_\Theta L^*(\theta, \delta^*(t))\, k(\theta; t)\, \mathrm{d}\theta \;=\; \inf_{\delta \in D} \int_\theta L^*(\theta, \delta(t))\, k(\theta; t)\, \mathrm{d}\theta,$$

para todo $t \in \mathcal{T}$.

Teorema

No caso da função de perda erro quadrado, a estimativa de Bayes $\delta^* \equiv \delta^*(t)$ é a média da distribuição a posteriori com fdp (ou fmp) $k(\theta; t)$; isto é,

$$\delta^*(t) \;=\; \int_\Theta \theta\, k(\theta; t)\, \mathrm{d}\theta \equiv \mathrm{E}_{V|T=t}[V],$$

para todo $t \in \mathcal{T}$.

Definição: Estimador Bayesiano

Estimador Bayesiano

Sejam $X_1, X_2, \ldots, X_n$ uma amostra aleatória de $[X \mid V = \theta]$ tendo fdp (ou fmp) $f_{X|V=\theta}$. Seja ainda V uma variável aleatória representando a distribuição a priori com fdp (ou fmp) $h(\theta)$. O estimador bayesiano para $\tau(\theta)$ é definido por

$$\widehat{\tau}_{\text{Bayes}}(\theta) = \mathrm{E}_{V|T=t}[\,\tau(V) \mid T(\boldsymbol{x}) = t\,] = \int_\Theta \tau(\theta)\, k(\theta \mid t)\, \mathrm{d}\theta$$
$$= \int_\Theta \tau(\theta)\, g(t \mid \theta)\, h(\theta)\, \mathrm{d}\theta,$$

em que $T(\boldsymbol{x}) \equiv T(X_1, \ldots, X_n)$ é uma estatística suficiente (minimal) para θ.

Exercício-3

Sejam $X_1, X_2, \ldots, X_n$ uma amostra aleatória de $[X \mid V = \theta] \sim$ Bernoulli(θ) tal que $\theta \in (0,1)$ e $V \sim U(0,1)$. Encontre o estimador Bayesiano para θ e para $\theta(1-\theta)$.

Exercício-4

Sejam $X_1, X_2, \ldots, X_n$ uma amostra aleatória de $[X \mid V = \theta] \sim$ Bernoulli(θ) tal que $\theta \in (0,1)$ e $V \sim$ Beta(α, β). Encontre o estimador Bayesiano para θ.