Material Auxiliar: Métodos Bayesianos

Inferência Estatística - Prof. Abraão D. C. Nascimento

Preparação para Prova

19 de novembro de 2025

Sumário

1 Introdução aos Métodos Bayesianos

1.1 Comparação: Inferência Clássica vs. Inferência Bayesiana

Na **Inferência Clássica** (ou Frequentista), o paradigma tradicional segue o seguinte padrão:

- Os métodos de estimação pontual, intervalar e teste de hipóteses dependem fundamentalmente da **função de verossimilhança**:

$$L(\theta; \mathbf{x}) = \prod_{i=1}^{n} f(x_i; \theta)$$

- O parâmetro $\theta \in \Theta \subset \mathbb{R}$ é tratado como um valor **desconhecido, mas fixo**
- $\mathbf{x} = (x_1, \ldots, x_n)^\top$ é uma realização de uma amostra aleatória n-dimensional de X
- A função $f(x; \theta)$ representa a função densidade de probabilidade (fdp) ou função de massa de probabilidade (fmp) de X

Na **Inferência Bayesiana**, ocorre uma mudança fundamental de paradigma:

- O parâmetro desconhecido θ é tratado como uma **variável aleatória** V
- V possui uma distribuição de probabilidade sobre o suporte Θ, chamada de **distribuição a priori** com fdp (ou fmp) $h(v)$
- A notação $f(x; \theta)$ agora representa a fdp (ou fmp) de $\{X|V = \theta\}$, ou seja, a distribuição condicional de X dado que $V = \theta$
- A distribuição a priori $h(v)$ frequentemente reflete a **crença subjetiva** do experimentador sobre quais valores de V são mais ou menos prováveis

Diferença Fundamental

A principal diferença conceitual é que na abordagem clássica, θ é um valor fixo desconhecido, enquanto na abordagem bayesiana, θ é uma variável aleatória com distribuição de probabilidade.

1.2 O Paradigma Bayesiano

O paradigma dos Métodos Bayesianos pode ser resumido como:

> *Combinar a evidência sobre V da distribuição a priori com a função de verossimilhança por meio do Teorema de Bayes, resultando na distribuição a posteriori.*

Simbolicamente, o processo pode ser representado como:

$$[V = \theta], \quad [X|V = \theta], \quad [T(X)|V = \theta] \quad \Rightarrow \quad [V = \theta|T(X) = t]$$

onde:

- $[V = \theta]$: distribuição a priori de V
- $[X|V = \theta]$: distribuição condicional dos dados dado o parâmetro
- $[T(X)|V = \theta]$: distribuição da estatística suficiente dado o parâmetro
- $[V = \theta|T(X) = t]$: distribuição a posteriori de V dado os dados observados

1.3 Teorema de Bayes

O Teorema de Bayes é a ferramenta fundamental que permite combinar informação a priori com os dados observados.

Teorema 1.1 (Teorema de Bayes - Caso Discreto)**.** 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

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

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

Para o caso contínuo (que é o foco deste curso), temos:

Teorema 1.2 (Teorema de Bayes - Caso Contínuo)**.** Se V é uma variável aleatória contínua com fdp a priori $h(\theta)$ e T é uma estatística com fdp condicional $g(t|\theta) = f_{T|V=\theta}(t|\theta)$, então a fdp a posteriori de V dado $T = t$ é:

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

onde $m(t) = \int_\Theta g(t;\theta)h(\theta)d\theta$ é a fdp marginal de T.

Exemplo 1.1 (Interpretação Intuitiva)**.** Imagine que você quer estimar a probabilidade de chuva amanhã (θ). Você tem:

- **Priori**: Sua crença inicial baseada em experiência (ex: 30% de chance)
- **Dados**: Informação meteorológica observada hoje
- **Posteriori**: Sua crença atualizada combinando sua crença inicial com os dados observados

O Teorema de Bayes combina essas duas fontes de informação de forma matematicamente rigorosa.

2 Distribuições a Priori e a Posteriori

2.1 Elementos da Inferência Bayesiana

Na abordagem Bayesiana, adotamos que V (ao invés de v) assume um modelo contínuo de valor real com fdp $h(v)$, em que $v \in \Theta \subset \mathbb{R}$.

Os **seis elementos fundamentais** da Inferência Bayesiana são:

2.1.1 Elemento 1: Distribuição a Priori

A **distribuição a priori** $h(\theta)$ para $\theta \in \Theta$ representa nosso conhecimento ou crença sobre o parâmetro antes de observar os dados.

Observação 2.1. A escolha da priori é crucial e pode ser:

- **Informativa**: Baseada em conhecimento prévio ou estudos anteriores
- **Não-informativa**: Representa ignorância ou neutralidade (ex: uniforme)
- **Conjugada**: Escolhida por conveniência matemática (ver Seção 3)

2.1.2 Elemento 2: Amostra Aleatória

$\mathbf{x} = (X_1, \ldots, X_n)^\top$ é uma amostra aleatória de $[X|V = \theta]$, ou seja, cada X_i tem distribuição condicional $f(x_i|\theta)$ dado que $V = \theta$.

2.1.3 Elemento 3: Estatística Suficiente

$T \triangleq T(\mathbf{x})$ é uma estatística suficiente (mínima) que assume frequentemente valor real. A fdp ou fmp de $[T|V = \theta]$ será denotada por:

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

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

Por que usar Estatística Suficiente?

O uso de estatística suficiente reduz a dimensão dos dados sem perder informação sobre θ, simplificando os cálculos bayesianos.

2.1.4 Elemento 4: Distribuição Conjunta

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

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

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

Esta é a distribuição conjunta que combina a informação dos dados (através de T) com a informação a priori (através de $h(\theta)$).

2.1.5 Elemento 5: Distribuição Marginal

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

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

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

Observação 2.2. A distribuição marginal $m(t)$ é frequentemente chamada de **distribuição preditiva** ou **verossimilhança marginal**. Ela representa a probabilidade de observar $T = t$ considerando todas as possíveis valores de θ ponderados pela priori.

2.1.6 Elemento 6: Distribuição a Posteriori

Finalmente, a fdp (ou fmp) de $[V|T = t]$ é dada por:

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

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** e representa nossa crença atualizada sobre θ após observar os dados.

Interpretação da Posteriori

A distribuição a posteriori combina:

- **Numerador**: $g(t;\theta)h(\theta)$ - informação dos dados vezes informação a priori
- **Denominador**: $m(t)$ - constante de normalização que garante que a posteriori seja uma distribuição válida

Observação 2.3. 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 através de métodos computacionais como MCMC (Markov Chain Monte Carlo).

2.2 Exercícios Resolvidos

Exercício 2.1 (Exercício 1.a)**.** Sejam $X_1, \ldots, X_n$ uma amostra aleatória de $[X|V = \theta] \sim \text{Bernoulli}(\theta)$ 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 $T \triangleq T(\mathbf{x})$), em que $\mathbf{x} = (X_1, \ldots, X_n)^\top$;

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

Solução do Exercício 1.a

Passo 1: Identificar a estatística suficiente

Para uma amostra de n observações de uma distribuição Bernoulli(θ), a estatística suficiente mínima é:

$$T = \sum_{i=1}^{n} X_i$$

que representa o número total de sucessos na amostra.

Passo 2: Determinar a distribuição de T dado θ

Como $X_i \sim$ Bernoulli(θ) e são independentes, temos:

$$T|V = \theta \sim \text{Binomial}(n, \theta)$$

Portanto, a fmp de T dado θ é:

$$g(t;\theta) = \binom{n}{t}\theta^t(1-\theta)^{n-t}, \quad t = 0, 1, \ldots, n$$

Passo 3: Especificar a distribuição a priori

A priori uniforme em $(0, 1)$ tem fdp:

$$h(\theta) = \begin{cases} 1 & \text{se } 0 < \theta < 1 \\ 0 & \text{caso contrário} \end{cases}$$

Passo 4: Calcular a densidade marginal $m(t)$

A densidade marginal é:

$$m(t) = \int_0^1 g(t;\theta)h(\theta)d\theta = \int_0^1 \binom{n}{t}\theta^t(1-\theta)^{n-t} \cdot 1\, d\theta$$

Fatorando a constante:

$$m(t) = \binom{n}{t}\int_0^1 \theta^t(1-\theta)^{n-t}d\theta$$

A integral $\int_0^1 \theta^t(1-\theta)^{n-t}d\theta$ é uma função Beta. Lembrando que:

$$B(a,b) = \int_0^1 u^{a-1}(1-u)^{b-1}du = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

Temos:

$$\int_0^1 \theta^t(1-\theta)^{n-t}d\theta = B(t+1, n-t+1) = \frac{\Gamma(t+1)\Gamma(n-t+1)}{\Gamma(n+2)}$$

Como $\Gamma(k+1) = k!$ para k inteiro não-negativo:

$$\int_0^1 \theta^t(1-\theta)^{n-t}d\theta = \frac{t!(n-t)!}{(n+1)!}$$

Portanto:

$$m(t) = \binom{n}{t}\frac{t!(n-t)!}{(n+1)!} = \frac{n!}{t!(n-t)!}\cdot\frac{t!(n-t)!}{(n+1)!} = \frac{1}{n+1}$$

Resultado (a): A densidade marginal de T é:

$$m(t) = \frac{1}{n+1}, \quad t = 0, 1, \ldots, n$$

Ou seja, T tem distribuição uniforme discreta sobre $\{0, 1, \ldots, n\}$.
Passo 5: Calcular a distribuição a posteriori
A distribuição a posteriori é:

$$k(\theta; t) = \frac{g(t;\theta)h(\theta)}{m(t)} = \frac{\binom{n}{t}\theta^t(1-\theta)^{n-t}\cdot 1}{\frac{1}{n+1}}$$

Simplificando:

$$k(\theta; t) = (n+1)\binom{n}{t}\theta^t(1-\theta)^{n-t}$$

Reescrevendo:

$$k(\theta; t) = \frac{\Gamma(n+2)}{\Gamma(t+1)\Gamma(n-t+1)}\theta^t(1-\theta)^{n-t}$$

Resultado (b): A densidade a posteriori de V dado $T = t$ é:

$$k(\theta; t) = \frac{\Gamma(n+2)}{\Gamma(t+1)\Gamma(n-t+1)}\theta^t(1-\theta)^{n-t}, \quad 0 < \theta < 1$$

Ou seja, $V|T = t \sim \text{Beta}(t+1, n-t+1)$.

Observação Importante

Note que começamos com uma priori uniforme (que é um caso especial de Beta(1,1)) e obtivemos uma posteriori Beta. Isso ilustra o conceito de priori conjugada que será estudado na próxima seção!

Exercício 2.2 (Exercício 1.b)**.** Sejam $X_1, \ldots, X_n$ uma amostra aleatória de $[X|V = \theta] \sim \text{Poisson}(\theta)$ 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 $T \triangleq T(\mathbf{x})$), em que $\mathbf{x} = (X_1, \ldots, X_n)^\top$;

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

Solução do Exercício 1.b

Passo 1: Identificar a estatística suficiente

Para uma amostra de n observações de uma distribuição Poisson(θ), a estatística suficiente mínima é:

$$T = \sum_{i=1}^{n} X_i$$

que representa o número total de ocorrências na amostra.

Passo 2: Determinar a distribuição de T dado θ

Como $X_i \sim \text{Poisson}(\theta)$ e são independentes, temos:

$$T|V = \theta \sim \text{Poisson}(n\theta)$$

Portanto, a fmp de T dado θ é:

$$g(t;\theta) = \frac{(n\theta)^t e^{-n\theta}}{t!}, \quad t = 0, 1, 2, \ldots$$

Passo 3: Especificar a distribuição a priori

A priori Gama(α, β) tem fdp:

$$h(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}, \quad \theta > 0$$

onde $\alpha > 0$ e $\beta > 0$ são os parâmetros da distribuição Gama.

Passo 4: Calcular a densidade marginal $m(t)$

A densidade marginal é:

$$\begin{aligned} m(t) &= \int_0^\infty g(t;\theta) h(\theta) d\theta \\ &= \int_0^\infty \frac{(n\theta)^t e^{-n\theta}}{t!} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} d\theta \end{aligned} \tag{1}$$

Reorganizando os termos:

$$m(t) = \frac{n^t \beta^\alpha}{t!\Gamma(\alpha)} \int_0^\infty \theta^t \theta^{\alpha-1} e^{-n\theta} e^{-\beta\theta} d\theta$$

Simplificando:

$$m(t) = \frac{n^t \beta^\alpha}{t!\Gamma(\alpha)} \int_0^\infty \theta^{t+\alpha-1} e^{-(n+\beta)\theta} d\theta$$

A integral $\int_0^\infty \theta^{t+\alpha-1} e^{-(n+\beta)\theta} d\theta$ pode ser resolvida usando a função Gama. Fazendo a substituição $u = (n+\beta)\theta$, temos $du = (n+\beta)d\theta$, então:

$$\begin{aligned} \int_0^\infty \theta^{t+\alpha-1} e^{-(n+\beta)\theta} d\theta &= \int_0^\infty \left(\frac{u}{n+\beta}\right)^{t+\alpha-1} e^{-u} \frac{du}{n+\beta} \\ &= \frac{1}{(n+\beta)^{t+\alpha}} \int_0^\infty u^{t+\alpha-1} e^{-u} du \end{aligned}$$

$$= \frac{\Gamma(t+\alpha)}{(n+\beta)^{t+\alpha}} \tag{2}$$

Portanto:

$$m(t) = \frac{n^t \beta^\alpha}{t!\Gamma(\alpha)} \cdot \frac{\Gamma(t+\alpha)}{(n+\beta)^{t+\alpha}}$$

Simplificando:

$$m(t) = \frac{\Gamma(t+\alpha)}{t!\Gamma(\alpha)} \cdot \frac{n^t \beta^\alpha}{(n+\beta)^{t+\alpha}}$$

$$m(t) = \binom{t+\alpha-1}{t} \left(\frac{\beta}{n+\beta}\right)^\alpha \left(\frac{n}{n+\beta}\right)^t$$

Resultado (a): A densidade marginal de T é:

$$m(t) = \frac{\Gamma(t+\alpha)}{t!\Gamma(\alpha)} \cdot \frac{n^t \beta^\alpha}{(n+\beta)^{t+\alpha}}, \quad t = 0, 1, 2, \ldots$$

Ou seja, T tem distribuição Binomial Negativa com parâmetros α e $\frac{\beta}{n+\beta}$.

Passo 5: Calcular a distribuição a posteriori

A distribuição a posteriori é:

$$k(\theta; t) = \frac{g(t;\theta)h(\theta)}{m(t)}$$

Substituindo as expressões:

$$k(\theta; t) = \frac{\frac{(n\theta)^t e^{-n\theta}}{t!} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}}{\frac{\Gamma(t+\alpha)}{t!\Gamma(\alpha)} \cdot \frac{n^t \beta^\alpha}{(n+\beta)^{t+\alpha}}}$$

Simplificando:

$$\begin{aligned} k(\theta; t) &= \frac{n^t \theta^t e^{-n\theta} \beta^\alpha \theta^{\alpha-1} e^{-\beta\theta} \Gamma(\alpha)(n+\beta)^{t+\alpha}}{\Gamma(\alpha)\Gamma(t+\alpha) n^t \beta^\alpha} \\ &= \frac{\theta^{t+\alpha-1} e^{-(n+\beta)\theta} (n+\beta)^{t+\alpha}}{\Gamma(t+\alpha)} \\ &= \frac{(n+\beta)^{t+\alpha}}{\Gamma(t+\alpha)} \theta^{t+\alpha-1} e^{-(n+\beta)\theta} \end{aligned} \tag{3}$$

Resultado (b): A densidade a posteriori de V dado $T = t$ é:

$$k(\theta; t) = \frac{(n+\beta)^{t+\alpha}}{\Gamma(t+\alpha)} \theta^{t+\alpha-1} e^{-(n+\beta)\theta}, \quad \theta > 0$$

Ou seja, $V | T = t \sim \Gamma(t+\alpha, n+\beta)$.

Observação Importante

Note que começamos com uma priori Gama(α, β) e obtivemos uma posteriori Gama$(t + \alpha, n + \beta)$. Isso confirma que a distribuição Gama é conjugada para a distribuição Poisson!

3 Priori Conjugada

3.1 Motivação

Como visto na seção anterior, a obtenção da distribuição a posteriori requer o cálculo da distribuição marginal $m(t)$:

$$m(t) = \int_{\Theta} g(t; \theta) h(\theta) d\theta$$

Se a priori $h_V(\theta) \triangleq h(\theta)$ de V 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 $k(\theta; t) = f_{V|T=t}(\theta|t)$ de forma fechada.

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

3.2 Definição de Priori Conjugada

Definição 3.1 (Priori Conjugada)**.** Suponha que a fdp à priori de V dada por $h(\theta)$ pertence a uma família de fdp's $\mathcal{P}$. Então $h(\theta)$ é chamada de **priori conjugada** para V se e somente se a posteriori $k(\theta; t)$ também pertence à família $\mathcal{P}$.

Vantagens das Prioris Conjugadas

- **Tratabilidade analítica**: Permitem obter a posteriori em forma fechada
- **Interpretação intuitiva**: Os parâmetros da posteriori são atualizações simples dos parâmetros da priori
- **Eficiência computacional**: Evitam a necessidade de métodos numéricos complexos

3.3 Exercícios Resolvidos

Exercício 3.1 (Exercício 2.a)**.** Sejam $X_1, X_2, \ldots, X_n$ uma amostra aleatória de $[X|V = \theta] \sim \text{Bernoulli}(\theta)$ tal que $\theta \in (0, 1)$. Encontre a distribuição a priori conjugada.

Solução do Exercício 2.a

Passo 1: Identificar a verossimilhança

Para uma amostra de n observações de uma distribuição Bernoulli(θ), a estatística suficiente é $T = \sum_{i=1}^{n} X_i$ e sua distribuição é:

$$T|V = \theta \sim \text{Binomial}(n, \theta)$$

A fmp é:

$$g(t; \theta) = \binom{n}{t}\theta^t(1-\theta)^{n-t}$$

Passo 2: Procurar uma família de distribuições conjugada

Queremos encontrar uma família de distribuições $\mathcal{P}$ tal que:

- A priori $h(\theta) \in \mathcal{P}$
- A posteriori $k(\theta; t) \in \mathcal{P}$

A posteriori é proporcional a:

$$k(\theta; t) \propto g(t; \theta)h(\theta) = \binom{n}{t}\theta^t(1-\theta)^{n-t}h(\theta)$$

$$\propto \theta^t(1-\theta)^{n-t}h(\theta)$$

Para que a posteriori tenha a mesma forma funcional da priori, precisamos que $h(\theta)$ seja proporcional a $\theta^a(1-\theta)^b$ para alguns $a, b > -1$.

Passo 3: Identificar a família Beta

A distribuição Beta tem fdp:

$$h(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}, \quad 0 < \theta < 1$$

onde $\alpha > 0$ e $\beta > 0$ são os parâmetros.

Passo 4: Verificar que Beta é conjugada

Se $h(\theta) \sim \text{Beta}(\alpha, \beta)$, então:

$$k(\theta; t) \propto \theta^t(1-\theta)^{n-t} \cdot \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$\propto \theta^{t+\alpha-1}(1-\theta)^{n-t+\beta-1}$$

Esta é exatamente a forma de uma distribuição Beta com parâmetros $(t+\alpha, n-t+\beta)$.

Portanto:

$$V|T = t \sim \text{Beta}(t+\alpha, n-t+\beta)$$

Resultado: A distribuição a priori conjugada para Bernoulli(θ) é a **distribuição Beta(α, β)**.

Interpretação dos Parâmetros

Os parâmetros da priori Beta podem ser interpretados como:

- α: "número de sucessos anteriores"(pseudo-observações)
- β: "número de fracassos anteriores"(pseudo-observações)
- $\alpha + \beta$: "tamanho da amostra anterior"

A posteriori atualiza esses valores somando os sucessos e fracassos observados.

Exercício 3.2 (Exercício 2.b)**.** Sejam $X_1, X_2, \ldots, X_n$ uma amostra aleatória de $[X|V = \theta] \sim \text{Poisson}(\theta)$ tal que $\theta > 0$, em que a variável V descreve a média populacional. Encontre a distribuição a priori conjugada.

Solução do Exercício 2.b

Passo 1: Identificar a verossimilhança

Para uma amostra de n observações de uma distribuição Poisson(θ), a estatística suficiente é $T = \sum_{i=1}^{n} X_i$ e sua distribuição é:

$$T|V = \theta \sim \text{Poisson}(n\theta)$$

A fmp é:

$$g(t;\theta) = \frac{(n\theta)^t e^{-n\theta}}{t!}$$

Passo 2: Procurar uma família de distribuições conjugada

A posteriori é proporcional a:

$$k(\theta;t) \propto g(t;\theta)h(\theta) = \frac{(n\theta)^t e^{-n\theta}}{t!} h(\theta)$$

$$\propto \theta^t e^{-n\theta} h(\theta)$$

Para que a posteriori tenha a mesma forma funcional da priori, precisamos que $h(\theta)$ seja proporcional a $\theta^a e^{-b\theta}$ para alguns $a > -1$ e $b > 0$.

Passo 3: Identificar a família Gama

A distribuição Gama tem fdp:

$$h(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}, \quad \theta > 0$$

onde $\alpha > 0$ e $\beta > 0$ são os parâmetros.

Passo 4: Verificar que Gama é conjugada

Se $h(\theta) \sim \Gamma(\alpha, \beta)$, então:

$$k(\theta;t) \propto \theta^t e^{-n\theta} \cdot \theta^{\alpha-1} e^{-\beta\theta}$$

$$\propto \theta^{t+\alpha-1}e^{-(n+\beta)\theta}$$

Esta é exatamente a forma de uma distribuição Gama com parâmetros $(t+\alpha, n+\beta)$.

Portanto:

$$V|T = t \sim \Gamma(t+\alpha, n+\beta)$$

Resultado: A distribuição a priori conjugada para Poisson(θ) é a **distribuição Gama(α, β)**.

Interpretação dos Parâmetros

Os parâmetros da priori Gama podem ser interpretados como:

- α: "número de ocorrências anteriores"(pseudo-observações)
- β: "escala do intervalo de tempo anterior"
- A média da priori é α/β

A posteriori atualiza esses valores somando as ocorrências observadas e o tamanho da amostra.

3.4 Tabela de Prioris Conjugadas Comuns

Tabela 1: Prioris Conjugadas para Distribuições Comuns

Distribuição dos Dados	**Priori Conjugada**	**Posteriori**
Bernoulli(θ)	Beta(α, β)	Beta($t+\alpha, n-t+\beta$)
Binomial(n, θ)	Beta(α, β)	Beta($t+\alpha, n^2-t+\beta$)
Poisson(θ)	Gama(α, β)	Gama($t+\alpha, n+\beta$)
Normal(μ, σ^2) conhecido σ^2	Normal(μ_0, τ_0^2)	Normal(μ_n, τ_n^2)
Normal(μ, σ^2) conhecido μ	Gama Inversa(α, β)	Gama Inversa($\alpha+n/2, \beta+S/2$)
Exponencial(θ)	Gama(α, β)	Gama($\alpha+n, \beta+t$)

Observação 3.1. Na tabela acima, t representa a estatística suficiente e n o tamanho da amostra. Os parâmetros da posteriori são atualizações dos parâmetros da priori com base nos dados observados.

4 Estimadores Bayesianos

4.1 Motivação à Estimação Bayesiana

Até agora, exploramos como obter a distribuição a posteriori de V dado os dados observados. Neste ponto, exploramos como **estimar θ sob uma função de perda particular**.

4.2 Contexto e Notação

Relembrando o contexto:

- $X_1, \ldots, X_n$ é uma amostra aleatória de $[X|V = \theta]$
- T é uma estatística suficiente (minimal) para θ
- $\mathcal{T}$ é o domínio de t
- $g(t; \theta)$ é a fdp (ou fmp) de $[T|V = \theta]$
- $h(\theta)$ é a fdp (fmp) a priori de V para $\theta \in \Theta$
- $k(\theta; t)$ é a fdp (fmp) a posteriori de $[V|T = t]$

4.3 Função de Perda e Risco

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

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**.

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

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

Esta medida é chamada de **risco frequentista**.

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 subespaço paramétrico.

4.4 Risco Bayesiano

Como alternativa ao risco frequentista, tem-se o **risco bayesiano**:

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

O risco bayesiano integra o risco frequentista sobre toda a distribuição a priori, fornecendo uma medida única de desempenho do estimador.

Suponha que $\mathcal{D}$ 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 $\mathcal{D}$ tal que:

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

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

4.5 Teoremas Fundamentais

Teorema 4.1. O estimador de Bayes $\delta^* \equiv \delta^*(T)$ é determinado tal que o risco a posteriori de $\delta^*(T)$ é o menor possível:

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

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

Teorema 4.2 (Estimador Bayesiano sob Perda Quadrática)**.** 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,

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

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

Interpretação

Sob perda quadrática, o estimador Bayesiano é simplesmente a **média da distribuição a posteriori**. Isso faz sentido intuitivamente: a média minimiza o erro quadrático médio.

4.6 Definição Geral de Estimador Bayesiano

Definição 4.1 (Estimador Bayesiano)**.** Sejam $X_1, X_2, \ldots, X_n$ uma amostra aleatória de $[X|V = \theta]$ 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:

$$\begin{aligned}\hat{\tau}_{\text{Bayes}}(\theta) = E_{V|T=t}[\tau(V)|T(\mathbf{x}) = t] &= \int_\Theta \tau(\theta)k(\theta|t)d\theta \\ &= \frac{\int_\Theta \tau(\theta)g(t|\theta)h(\theta)d\theta}{\int_\Theta g(t|\theta)h(\theta)d\theta},\end{aligned}$$

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

Observação 4.1. Note que o estimador Bayesiano para uma função $\tau(\theta)$ é simplesmente a esperança de $\tau(V)$ com respeito à distribuição a posteriori. Isso generaliza o resultado do teorema anterior.

4.7 Exercícios Resolvidos

Exercício 4.1 (Exercício 3)**.** Sejam $X_1, X_2, \ldots, X_n$ uma amostra aleatória de $[X|V = \theta] \sim \text{Bernoulli}(\theta)$ tal que $\theta \in (0, 1)$ e $V \sim U(0, 1)$. Encontre o estimador Bayesiano para θ e para $\theta(1 - \theta)$.

Solução do Exercício 3

Parte 1: Estimador Bayesiano para θ

Do Exercício 1.a, sabemos que:

$$V|T = t \sim \text{Beta}(t+1, n-t+1)$$

A fdp a posteriori é:

$$k(\theta; t) = \frac{\Gamma(n+2)}{\Gamma(t+1)\Gamma(n-t+1)}\theta^t(1-\theta)^{n-t}, \quad 0 < \theta < 1$$

O estimador Bayesiano para θ sob perda quadrática é a média da distribuição a posteriori:

$$\hat{\theta}_{\text{Bayes}} = E_{V|T=t}[V] = \int_0^1 \theta k(\theta; t) d\theta$$

$$= \int_0^1 \theta \cdot \frac{\Gamma(n+2)}{\Gamma(t+1)\Gamma(n-t+1)}\theta^t(1-\theta)^{n-t} d\theta$$

$$= \frac{\Gamma(n+2)}{\Gamma(t+1)\Gamma(n-t+1)} \int_0^1 \theta^{t+1}(1-\theta)^{n-t} d\theta$$

A integral é uma função Beta:

$$\int_0^1 \theta^{t+1}(1-\theta)^{n-t} d\theta = B(t+2, n-t+1) = \frac{\Gamma(t+2)\Gamma(n-t+1)}{\Gamma(n+3)}$$

Portanto:

$$\hat{\theta}_{\text{Bayes}} = \frac{\Gamma(n+2)}{\Gamma(t+1)\Gamma(n-t+1)} \cdot \frac{\Gamma(t+2)\Gamma(n-t+1)}{\Gamma(n+3)}$$

$$= \frac{\Gamma(n+2)\Gamma(t+2)}{\Gamma(t+1)\Gamma(n+3)}$$

Usando $\Gamma(k+1) = k\Gamma(k)$:

$$\Gamma(t+2) = (t+1)\Gamma(t+1)$$

$$\Gamma(n+3) = (n+2)\Gamma(n+2)$$

Portanto:

$$\hat{\theta}_{\text{Bayes}} = \frac{\Gamma(n+2)(t+1)\Gamma(t+1)}{\Gamma(t+1)(n+2)\Gamma(n+2)} = \frac{t+1}{n+2}$$

Resultado para θ:

$$\hat{\theta}_{\text{Bayes}} = \frac{t+1}{n+2} = \frac{T+1}{n+2}$$

onde $T = \sum_{i=1}^n X_i$ é o número de sucessos.

Parte 2: Estimador Bayesiano para $\theta(1-\theta)$

O estimador Bayesiano para $\tau(\theta) = \theta(1-\theta)$ é:

$$\widehat{\theta(1-\theta)}_{\text{Bayes}} = E_{V|T=t}[\theta(1-\theta)] = \int_0^1 \theta(1-\theta)k(\theta;t)d\theta$$

$$= \int_0^1 \theta(1-\theta) \cdot \frac{\Gamma(n+2)}{\Gamma(t+1)\Gamma(n-t+1)}\theta^t(1-\theta)^{n-t}d\theta$$

$$= \frac{\Gamma(n+2)}{\Gamma(t+1)\Gamma(n-t+1)} \int_0^1 \theta^{t+1}(1-\theta)^{n-t+1}d\theta$$

A integral é:

$$\int_0^1 \theta^{t+1}(1-\theta)^{n-t+1}d\theta = B(t+2, n-t+2) = \frac{\Gamma(t+2)\Gamma(n-t+2)}{\Gamma(n+4)}$$

Portanto:

$$\widehat{\theta(1-\theta)}_{\text{Bayes}} = \frac{\Gamma(n+2)}{\Gamma(t+1)\Gamma(n-t+1)} \cdot \frac{\Gamma(t+2)\Gamma(n-t+2)}{\Gamma(n+4)}$$

Usando propriedades da função Gama:

$$\Gamma(t+2) = (t+1)\Gamma(t+1)$$

$$\Gamma(n-t+2) = (n-t+1)\Gamma(n-t+1)$$

$$\Gamma(n+4) = (n+3)(n+2)\Gamma(n+2)$$

Portanto:

$$\widehat{\theta(1-\theta)}_{\text{Bayes}} = \frac{\Gamma(n+2)(t+1)\Gamma(t+1)(n-t+1)\Gamma(n-t+1)}{\Gamma(t+1)\Gamma(n-t+1)(n+3)(n+2)\Gamma(n+2)}$$

$$= \frac{(t+1)(n-t+1)}{(n+2)(n+3)}$$

Resultado para $\theta(1-\theta)$:

$$\widehat{\theta(1-\theta)}_{\text{Bayes}} = \frac{(T+1)(n-T+1)}{(n+2)(n+3)}$$

Observação

Note que $\theta(1-\theta)$ é a variância de uma variável Bernoulli(θ). O estimador Bayesiano leva em conta tanto a informação dos dados quanto a informação a priori.

Exercício 4.2 (Exercício 4)**.** Sejam $X_1, X_2, \ldots, X_n$ uma amostra aleatória de $[X|V = \theta] \sim$ Bernoulli(θ) tal que $\theta \in (0,1)$ e $V \sim$ Beta(α, β). Encontre o estimador Bayesiano para θ.

Solução do Exercício 4

Passo 1: Determinar a distribuição a posteriori

Como visto no Exercício 2.a, se $V \sim \text{Beta}(\alpha, \beta)$ e $T|V = \theta \sim \text{Binomial}(n, \theta)$, então:

$$V|T = t \sim \text{Beta}(t + \alpha, n - t + \beta)$$

A fdp a posteriori é:

$$k(\theta; t) = \frac{\Gamma(n + \alpha + \beta)}{\Gamma(t + \alpha)\Gamma(n - t + \beta)}\theta^{t+\alpha-1}(1 - \theta)^{n-t+\beta-1}, \quad 0 < \theta < 1$$

Passo 2: Calcular o estimador Bayesiano

O estimador Bayesiano para θ sob perda quadrática é a média da distribuição a posteriori. Para uma distribuição $\text{Beta}(a, b)$, a média é:

$$E[\text{Beta}(a, b)] = \frac{a}{a + b}$$

Portanto:

$$\hat{\theta}_{\text{Bayes}} = E_{V|T=t}[V] = \frac{t + \alpha}{t + \alpha + n - t + \beta} = \frac{t + \alpha}{n + \alpha + \beta}$$

Resultado:

$$\hat{\theta}_{\text{Bayes}} = \frac{T + \alpha}{n + \alpha + \beta}$$

onde $T = \sum_{i=1}^{n} X_i$ é o número de sucessos.

Interpretação

O estimador pode ser reescrito como:

$$\hat{\theta}_{\text{Bayes}} = \frac{\alpha}{\alpha + \beta} \cdot \frac{\alpha + \beta}{n + \alpha + \beta} + \frac{T}{n} \cdot \frac{n}{n + \alpha + \beta}$$

Isso mostra que o estimador é uma média ponderada entre:

- A média da priori: $\frac{\alpha}{\alpha+\beta}$
- A média amostral: $\frac{T}{n}$

Os pesos dependem do "tamanho efetivo" da priori $(\alpha + \beta)$ e do tamanho da amostra (n).

Verificação alternativa usando a definição:

$$\hat{\theta}_{\text{Bayes}} = \int_0^1 \theta k(\theta; t) d\theta$$

$$= \int_0^1 \theta \cdot \frac{\Gamma(n + \alpha + \beta)}{\Gamma(t + \alpha)\Gamma(n - t + \beta)}\theta^{t+\alpha-1}(1 - \theta)^{n-t+\beta-1} d\theta$$

$$
\begin{aligned}
&= \frac{\Gamma(n+\alpha+\beta)}{\Gamma(t+\alpha)\Gamma(n-t+\beta)} \int_0^1 \theta^{t+\alpha}(1-\theta)^{n-t+\beta-1} d\theta \\
&= \frac{\Gamma(n+\alpha+\beta)}{\Gamma(t+\alpha)\Gamma(n-t+\beta)} \cdot \frac{\Gamma(t+\alpha+1)\Gamma(n-t+\beta)}{\Gamma(n+\alpha+\beta+1)} \\
&= \frac{\Gamma(n+\alpha+\beta)(t+\alpha)\Gamma(t+\alpha)}{\Gamma(t+\alpha)(n+\alpha+\beta)\Gamma(n+\alpha+\beta)} \\
&= \frac{t+\alpha}{n+\alpha+\beta}
\end{aligned}
$$

Confirmando o resultado anterior.

5 Resumo de Fórmulas Importantes

5.1 Fórmulas Fundamentais

Tabela 2: Resumo de Fórmulas Bayesianas

Conceito	**Fórmula**
Distribuição Conjunta	$f_{T,V}(t,\theta) = g(t;\theta)h(\theta)$
Distribuição Marginal	$m(t) = \int_\Theta g(t;\theta)h(\theta)d\theta$
Distribuição a Posteriori	$k(\theta;t) = \frac{g(t;\theta)h(\theta)}{m(t)}$
Estimador Bayesiano (perda quadrática)	$\hat{\theta}_{\text{Bayes}} = E_{V\|T=t}[V]$
Estimador Bayesiano (geral)	$\hat{\tau}_{\text{Bayes}} = E_{V\|T=t}[\tau(V)]$
Risco Frequentista	$R^*(\theta,\delta) = E_{[T\|V=\theta]}[L^*(\theta,\delta)]$
Risco Bayesiano	$r^*(\theta,\delta) = E_V[R^*(V,\delta)]$

5.2 Propriedades da Função Gama e Beta

- $\Gamma(k+1) = k\Gamma(k)$ para $k > 0$
- $\Gamma(n+1) = n!$ para n inteiro não-negativo
- $B(a,b) = \int_0^1 u^{a-1}(1-u)^{b-1}du = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$
- Se $X \sim \text{Beta}(a,b)$, então $E[X] = \frac{a}{a+b}$
- Se $X \sim \Gamma(\alpha,\beta)$, então $E[X] = \frac{\alpha}{\beta}$

6 Observações Finais

Pontos-Chave para a Prova

1. Entenda a diferença fundamental entre Inferência Clássica e Bayesiana
2. Domine os 6 elementos da Inferência Bayesiana e como eles se relacionam
3. Saiba identificar e usar prioris conjugadas
4. Lembre-se: sob perda quadrática, o estimador Bayesiano é a média da posteriori
5. Pratique o cálculo de integrais envolvendo funções Beta e Gama
6. Entenda a interpretação dos parâmetros das prioris conjugadas

Observação 6.1. Este material foi desenvolvido como auxílio para estudo. Recomenda-se também revisar as notas de aula originais e resolver exercícios adicionais para consolidar o aprendizado.