

Universidade Federal do Rio Grande do Norte
Instituto Metr pole Digital
IMD0601 - Bioestat stica

Infer ncia Bayesiana

Prof. Dr. Tetsu Sakamoto
Instituto Metr pole Digital - UFRN
Sala A224, ramal 182
Email: tetsu@imd.ufrn.br



Baixe a aula (e os arquivos)

- Para aqueles que não clonaram o repositório:

```
> git clone https://github.com/tetsufmbio/IMD0601.git
```

- Para aqueles que já tem o repositório local:

```
> cd /path/to/IMD0601
```

```
> git pull
```

Essência do algoritmo bayesiano

$$P(\textit{Modelo}|\textit{dados}) \propto P(\textit{Dados}|\textit{modelo})P(\textit{Modelo})$$

$$\textit{Posteriori} \propto \textit{Verossimilhança} \cdot \textit{Priori}$$

“Modelo subjetivo a priori é atualizado por um modelo de dados (verossimilhança) para produzir uma distribuição posteriori, combinando a subjetividade do conhecimento do pesquisador com a objetividade dos dados observados.”

Teorema de Bayes

Uma simples regra dentro da probabilidade.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

$$P(A|B) \cdot P(B) = P(A \cap B)$$

$$P(B|A) \cdot P(A) = P(B \cap A)$$

$$P(A \cap B) = P(B \cap A)$$

$$P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

**Teorema de
Bayes**

Teorema de Bayes

Variações do teorema de Bayes

$$P(A|B) = \frac{P(B|A).P(A)}{P(B)}$$

$$P(A|B) \propto P(B|A).P(A) \quad P(B) \text{ constante, } \propto = \text{“proporcional a”}$$

$$P(H|E) = \frac{P(E|H).P(H)}{P(E)}$$

E = Evidência

H = Hipótese

P(H) = probabilidade a priori antes da evidência

P(H|E) = probabilidade a posteriori depois da evidência

P(E|H) = probabilidade da evidência dada hipótese

Estendendo o teorema de Bayes para distribuição

$$P(H|E) \propto P(E|H) \cdot P(H)$$

Voltando ao exemplo da moeda.

$P(H)$ → Probabilidade a priori → Crença subjetiva da proporção de caras (vamos supor que é de $\frac{1}{4}$);

Distribuição beta → modelar distribuição de probabilidades com dados de proporção;

$$f(x) = \frac{1}{\beta(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, 0 < x < 1$$

Achar valores de alfa e beta onde a distribuição fique centrada em 0.25;

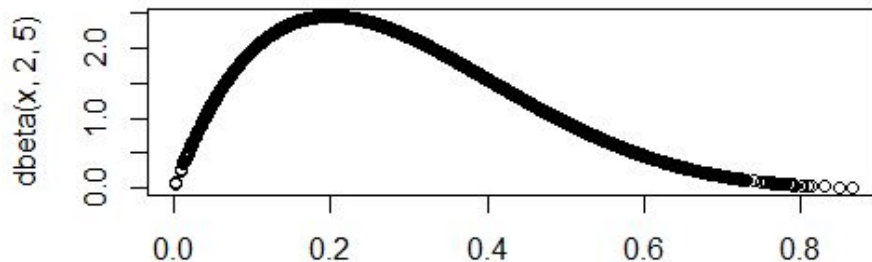
Estendendo o teorema de Bayes para distribuição

Determinar alfa e beta onde a distribuição fique centrada em 0.25;

$$\mu = \frac{\alpha}{\alpha + \beta}$$

$$\sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

Alfa = 2, Beta = 5



Estendendo o teorema de Bayes para distribuição

$$P(H|E) \propto P(E|H) \cdot P(H)$$

$P(E|H)$ → Probabilidade da evidência dado uma hipótese;

Para obter esta probabilidade precisamos de **evidência** e evidência são os **dados**.

Para obter dados, precisamos realizar **experimento**: Jogadas de 10 moedas:

- cada jogada é um Bernoulli {cara, coroa}.
- Combinação das 10 jogadas → distribuição binomial.

$$P(k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Estendendo o teorema de Bayes para distribuição

Suponha que das 10 jogadas, 5 deram cara.

- Bayesiana $\rightarrow P(\text{cara})=p$ é uma variável aleatória \rightarrow infinitos modelos plausíveis, um para cada valor de p .

$$P(5 \text{ caras em } 10|p) = \binom{10}{5} p^5 (1 - p)^5$$

Probabilidade de um resultado dado um $p \rightarrow$ **Verossimilhança(p)**

Estendendo o teorema de Bayes para distribuição

$$P(H|E) \propto P(E|H) \cdot P(H)$$

$$P(H) = \frac{1}{\beta(\alpha, \beta)} p^{\alpha-1} (1-p)^{\beta-1}$$

$$P(E|H) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$P(H|E) \propto \binom{n}{k} p^k (1-p)^{n-k} \cdot \frac{1}{\beta(\alpha, \beta)} p^{\alpha-1} (1-p)^{\beta-1}$$

Estendendo o teorema de Bayes para distribuição

$$P(H|E) \propto P(E|H). P(H)$$

$$P(H|E) \propto \binom{n}{k} p^k (1-p)^{n-k} \cdot \frac{1}{\beta(\alpha, \beta)} p^{\alpha-1} (1-p)^{\beta-1}$$

$$a^m \cdot a^n = a^{m+n}$$

$$P(H|E) \propto \binom{n}{k} \frac{1}{\beta(\alpha, \beta)} p^{k+\alpha-1} (1-p)^{n-k+\beta-1}$$

Considerar constante como c

$$P(H|E) \propto c p^{k+\alpha-1} (1-p)^{n-k+\beta-1}$$

Estendendo o teorema de Bayes para distribuição

$$P(H|E) \propto cp^{k+\alpha-1} (1-p)^{n-k+\beta-1}$$

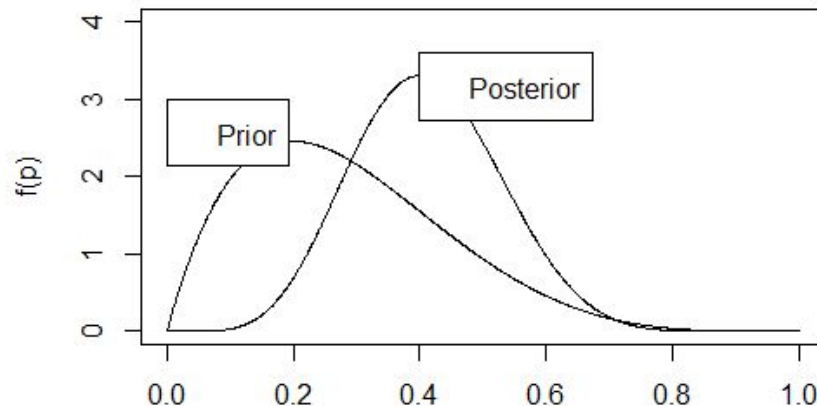
Segue uma distribuição beta:

$$f(x) = \frac{1}{\beta(\alpha,\beta)} x^{\alpha-1} (1-x)^{\beta-1}, 0 < x < 1$$

onde:

- **alfa2 = k + alfa** → 5 + 2 = 7
- **beta2 = n - k + beta** → 10 - 5 + 5 = 10

Distribuição posteriori = beta(7, 10)



Nesta aula...

Apresentar mais detalhes sobre cada um dos elementos da inferência bayesiana:

- Função verossimilhança;
- Distribuição a Prior
- Distribuição posteriori

$$P(H|E) \propto P(E|H). P(H)$$

Função de verossimilhança

Modelo de distribuição

$$P(k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

$$P(k|n, p) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Probabilidade de um dado resultado acontecer
(dado n repetição e p probabilidade de sucesso)

Função de verossimilhança

$$L(p|k, n) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Probabilidade do parâmetro p dado que um
resultado de n repetições aconteceu.

Máxima verossimilhança → o valor máximo que
 $L(p|k, n)$ pode assumir → o p que torna $L(p|k, n)$
máximo seria o valor mais plausível para p;

Priors

- Fonte de controvérsia e distinção;
- Introduz ao modelo:
 - Informação subjetiva
 - Conhecimentos anteriores → atualização do conhecimento;
- Não existe uma regra definida do que se pode dizer de prior correto;
- Escolha dos priors para um modelo
 - Necessidade para realizar uma inferência bayesiana
 - Disciplina;
- Nesta aula → modelos simples e formas padrões;
 - Priors conjugados (beta e binomial);
 - Priors não informativos:

Priors conjugados (priors informativos)

Faz a distribuição posteriori seguir a mesma forma paramétrica da distribuição a priori, através do uso de uma forma particular da função verossimilhança.

Propriedade: **Conjugação**;

Exemplo: combinando prior com distribuição beta e a verossimilhança binomial produz uma distribuição posteriori beta;

Distribuição Discreta X Contínua;

Outros priors conjugados

	Prior	Verossimilhança	Posterior
Modelos univariadas	Beta	Binomial	Beta
	Gamma	Poisson	Gamma
	Gamma	Exponencial	Gamma
Modelos multivariadas	Normal	Normal	Normal
	Dirichlet	Multinomial	Dirichlet
	Normal multivariada	Normal multivariada	Normal multivariada

Poisson e Gamma

Poisson

$$P(x|\theta) = \frac{\theta^x e^{-\theta}}{x!}$$

$$P(dados|modelo) = P(x_1, x_2, \dots, x_n|\theta) = \prod \frac{\theta^{x_i} e^{-\theta}}{x_i!}$$

$$P(dados|modelo) \propto \theta^{\sum x_i} e^{-n\theta}$$

Gamma

$$f(\theta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \theta^{\alpha-1} e^{-\theta\beta}$$

$$P(\theta) \propto \theta^{\alpha-1} e^{-\theta\beta}$$

Medindo a taxa de mutação de um gene

Taxa de mutação $\rightarrow \theta$

Conhecimento prévio $\rightarrow \theta = 0,4$ para cada 1000 pessoas;

Modelo a priori \rightarrow Distribuição gamma(alfa = 2, beta = 5)

Novos dados

- $n = 10$ amostras de 1000 pessoas cada;
- Número de mutações (0,2,1,0,1,2,0,1,1,0)

Medindo a taxa de mutação de um gene

$$P(\text{Modelo}|\text{dados}) \propto P(\text{Dados}|\text{modelo})P(\text{Modelo})$$

$$P(\theta|x) \propto P(x|\theta) * P(\theta)$$

$$P(\theta|x) \propto \theta^{\sum x_i} e^{-n\theta} * \theta^{\alpha-1} e^{-\theta\beta}$$

$$P(\theta|x) \propto \theta^{\alpha-1+\sum x_i} e^{-\theta(n+\beta)}$$

Medindo a taxa de mutação de um gene

$$P(\theta|x) \propto \theta^{\alpha-1+\sum x_i} e^{-\theta(n+\beta)}$$

- Dados = (0,2,1,0,1,2,0,1,1,0) $\rightarrow \sum x_i = 8$;
- newAlpha = $\sum x_i + \alpha = 10$
- newBeta = $n + \beta = 15$

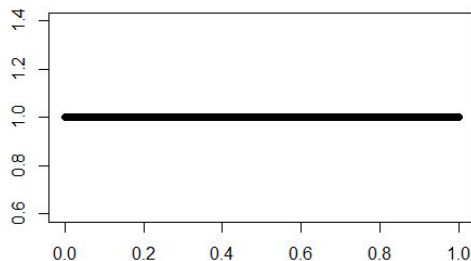
Distribuição posteriori $\rightarrow \text{Gamma}(\alpha=10, \beta=15)$

Priors não informativos

O que fazer se queremos utilizar a inferência bayesiana mas não sabemos nada sobre a distribuição a priori?

Utilizar **priors não informativos** (prior de referência, prior vago, prior difuso, flat prior)

Exemplo: Se não conhecemos o prior de algum dado relativo a proporção de algo (exemplo: proporção de nucleotídeos em uma sequência) → $\text{beta}(1,1)$



Avaliando a distribuição posteriori

Distribuição posteriori → Influência da **função verossimilhança** e da **distribuição a priori**

Voltando para o exemplo da moeda:

- Prior $p = 0,25$ (beta(2,5));
- Distribuição binomial $n = 10$; $k = 5$;

Se ao invés de jogar 10 moedas, mas jogássemos 100 moedas, e obtivéssemos 50 caras, como se comportaria a distribuição posteriori?

Avaliando a distribuição posteriori

$$P(H|E) \propto P(E|H) \cdot P(H)$$

$$P(H) = \frac{1}{\beta(\alpha, \beta)} p^{\alpha-1} (1-p)^{\beta-1}$$

$$P(E|H) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$P(H|E) \propto \binom{n}{k} p^k (1-p)^{n-k} \cdot \frac{1}{\beta(\alpha, \beta)} p^{\alpha-1} (1-p)^{\beta-1}$$

$$P(H|E) \propto c p^{k+\alpha-1} (1-p)^{n-k+\beta-1}$$

Avaliando a distribuição posteriori

Distribuição posteriori → Influência da **função verossimilhança** e da **distribuição a priori**

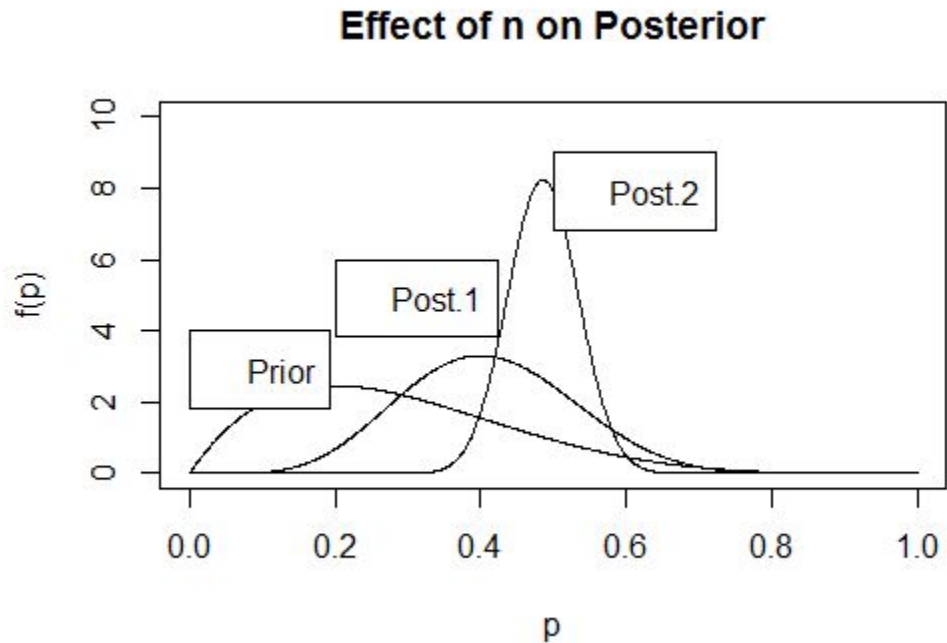
Voltando para o exemplo da moeda:

- Prior $p = 0,25$ (beta(2,5));
- Distribuição binomial $n = 10$; $k = 5$;

Se ao invés de jogar 10 moedas, mas jogássemos 100 moedas, e obtivéssemos 50 caras, como se comportaria a distribuição posteriori?

- $\alpha_{\text{New}} = k + \alpha = 52$
- $\beta_{\text{New}} = n - k + \beta = 55$

Avaliando a distribuição posteriori



Avaliando a distribuição posteriori

Distribuição posteriori → Influência da **função verossimilhança** e da **distribuição a priori**

Voltando para o exemplo da moeda:

- Prior $p = 0,25$ (beta(2,5));
- Distribuição binomial $n = 10$; $k = 5$;

Se ao invés de considerar um prior de 0,25, deixarmos o prior vago (prior não informativo), o que aconteceria com a distribuição posteriori?

Avaliando a distribuição posteriori

Distribuição posteriori → Influência da **função verossimilhança** e da **distribuição a priori**

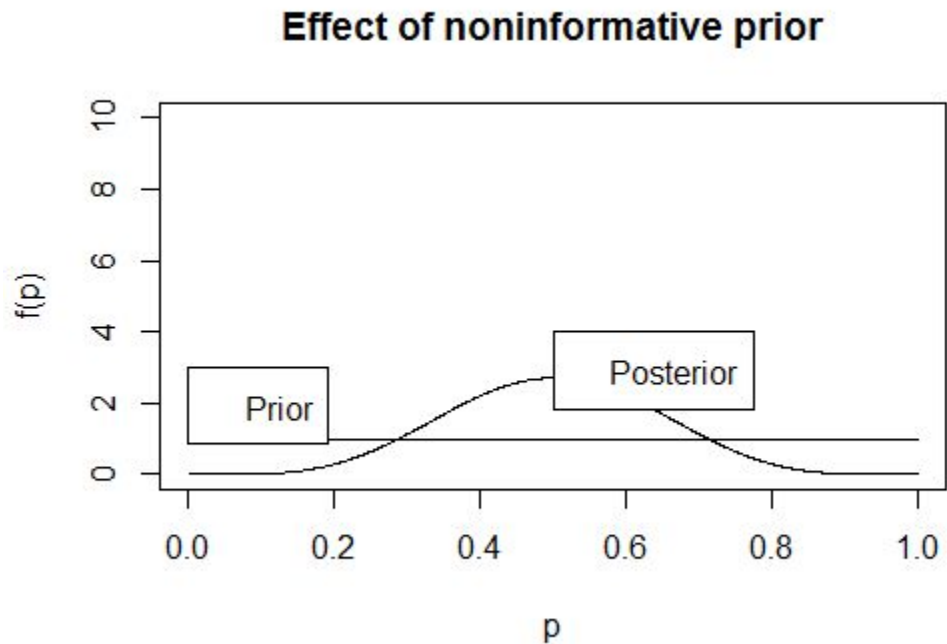
Voltando para o exemplo da moeda:

- Prior $p = 0,25$ ($\text{beta}(2,5)$);
- Distribuição binomial $n = 10$; $k = 5$;

Se ao invés de considerar um prior de 0,25, deixarmos o prior vago (prior não informativo, $\text{beta}(1,1)$), o que aconteceria com a distribuição posteriori?

- $\alpha_{\text{New}} = k + \alpha = 6$
- $\beta_{\text{New}} = n - k + \beta = 6$

Avaliando a distribuição posteriori



Avaliando a distribuição posteriori

Efeito da função verossimilhança e da distribuição a priori

Grande volume de conjunto de dados → os dados dominam a forma da distribuição posteriori;

Pequeno volume de dados e um prior preciso → o conhecimento a priori domina a forma da distribuição posteriori.

Constante normalizador

Na maioria dos exemplos mostrados trabalhamos com modelos utilizando o fator de proporcionalidade:

$$P(\textit{Modelo}|\textit{dados}) \propto P(\textit{Dados}|\textit{modelo})P(\textit{Modelo})$$
$$P(H|E) \propto P(E|H). P(H)$$

Esta forma de representação não define explicitamente o modelo, e para isso devemos considerar a constante que deixamos de lado.

$$P(H|E) = c * P(E|H) * P(H)$$

Necessário para que a integral da distribuição seja 1, mas não influencia na forma da distribuição.