MAT02034 - Métodos bayesianos para análise de dados

Introdução ao raciocínio bayesiano

Rodrigo Citton P. dos Reis
citton.padilha@ufrgs.br

Universidade Federal do Rio Grande do Sul
Instituto de Matemática e Estatística
Departamento de Estatística

Porto Alegre, 2022

Introdução

Introdução

- Nesta aula, os elementos básicos da abordagem inferencial bayesiana são introduzidos por meio do problema básico de aprender sobre uma proporção populacional.
- Antes de coletar os dados, temos crenças sobre o valor da proporção e modelamos nossas crenças em termos de uma **distribuição *a priori***.
- Após a **observação dos dados**, atualiza-se a crença sobre a proporção calculando a **distribuição *a posteriori***.
- **Resumimos** esta distribuição de probabilidade para realizar **inferências**.
- Além disso, pode-se estar interessado em **prever os resultados** prováveis de uma nova amostra retirada da população.

Aprendendo sobre uma proporção

Aprendendo sobre uma proporção

- Suponha que uma pessoa esteja interessada em aprender sobre os **hábitos de sono** dos estudantes universitários.
- Ela ouve que os médicos recomendam **oito horas de sono** para um adulto médio.
- **Que proporção de estudantes universitários dormem pelo menos oito horas?**

- Aqui pensamos em uma **população** composta por todos os estudantes universitários e p representa a proporção dessa população que dorme (em uma noite típica durante a semana) pelo menos oito horas.
- Estamos interessados em aprender sobre p.

Aprendendo sobre uma proporção

- O valor de p é desconhecido.
- Sob o ponto de vista bayesiano, as crenças de uma pessoa sobre a incerteza de p são representadas por uma distribuição de probabilidade colocada nesse parâmetro.
- Essa distribuição reflete a opinião subjetiva *a priori* da pessoa sobre valores plausíveis de p.

- Uma **amostra aleatória** de estudantes de uma determinada universidade será coletada para aprender sobre essa proporção.
- Mas primeiro o pesquisador faz algumas pesquisas para aprender sobre os hábitos de sono dos estudantes universitários.
  - Esta pesquisa a ajudará na construção de uma distribuição *a priori*.

Aprendendo sobre uma proporção

- Com base nessa pesquisa, **a pessoa** que faz o estudo **acredita que os estudantes** universitários geralmente **dormem menos de oito horas** e, portanto, p (a proporção que dorme pelo menos oito horas) é provavelmente menor que 0.5.
- Após alguma reflexão, seu melhor palpite para o valor de p é 0.3.
  - Mas é muito plausível que essa proporção possa ser qualquer valor no intervalo de 0 a 0.5.

- Uma amostra com 27 estudantes é coletada.
  - Destes, 11 registram que dormiram pelo menos oito horas na noite anterior.
- Com base nas informações anteriores e nesses dados observados, o pesquisador está interessado em estimar a proporção p.

Aprendendo sobre uma proporção

- Denote a densidade *a priori* para p por $g(p)$.
- Se considerarmos um **"sucesso"** como dormir pelo menos oito horas e tomarmos uma amostra aleatória com s sucessos e f falhas, então a **função de verossimilhança** é dada por

$$L(p) \propto p^s(1-p)^f,\ 0 < p < 1.$$

- A densidade *a posteriori* para p, pela **regra de Bayes**, é obtida, a menos de uma constante de proporcionalidade, multiplicando a densidade *a priori* pela verossimilhança:

$$g(p|\text{dados}) \propto g(p)L(p).$$

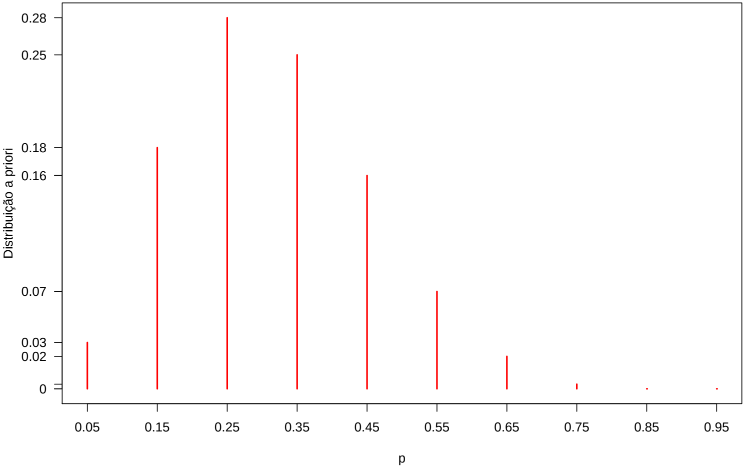
Utilizando uma priori discreta

Utilizando uma priori discreta

- Uma **abordagem simples** para avaliar uma distribuição priori para p é escrever uma **lista de valores** de proporção plausíveis e então atribuir **pesos** (w) a esses valores.

p_i	0.05	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85	0.95
$w(p_i)$	1	5.2	8	7.2	4.6	2.1	0.7	0.1	0	0
$g(p_i)$	0.035	0.180	0.277	0.249	0.159	0.073	0.024	0.003	0	0

Utilizando uma priori discreta



Utilizando uma priori discreta

- Lembrando que em nosso exemplo, 11 de 27 alunos dormem um número suficiente de horas ($\hat{p} = 0.40$), então $s = 11$ e $f = 16$, e a função de verossimilhança[1] é

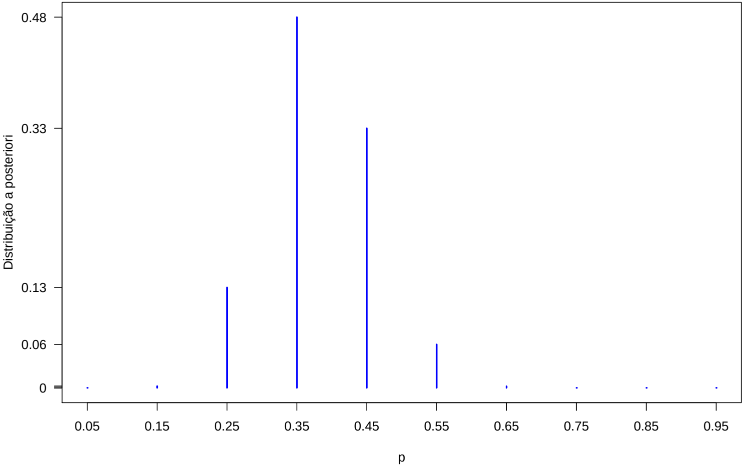
$$L(p) \propto p^{11}(1-p)^{16},\ 0 < p < 1.$$

- A distribuição *a posteriori* é obtida, no caso discreto, da seguinte forma

$$g(p_i|dados) = \frac{g(p_i)L(p_i)}{\sum_i g(p_i)L(p_i)},\ p_i \in \{0.05, 0.15, \ldots, 0.95\}.$$

[1] Note que a verossimilhança é núcleo de uma densidade beta com parâmetros $s+1 = 12$ e $f+1 = 17$.

Utilizando uma priori discreta



Utilizando uma priori discreta

p_i	0.25	0.35	0.45	0.55	0.65	0.75
$g(p_i)$	0.277	0.249	0.159	0.073	0.024	0.003
$L(p_i) \times 10^{12}$	2389.573	9803.072	10743.374	3939.035	443.747	9.834
$g(p_i \| dados)$	0.129	0.477	0.334	0.056	0.002	0.000

- Notamos que a maior parte da probabilidade *a posteriori* está concentrada nos valores $p = 0.35$ e $p = 0.45$.
- Se combinarmos as probabilidades para os três valores mais prováveis, podemos dizer que a probabilidade *a posteriori* de que p caia no conjunto $0.25, 0.35, 0.45$ é igual a 0.940.

Utilizando uma priori beta

Utilizando uma priori beta

- Como a p é um parâmetro contínuo, uma abordagem alternativa é construir uma densidade $g(p)$ no intervalo $(0, 1)$ que represente as crenças iniciais do pesquisador.
- Suponha que ele acredite que a proporção tem a mesma probabilidade de ser menor ou maior que $p = 0.3$.
  - $\Pr(p \leq 0.3) = \Pr(p \geq 0.3)$.
- Além disso, ela está 90% confiante de que p é menor que 0.5.
  - $\Pr(p \leq 0.5) = 0.90$.

Utilizando uma priori beta

- Uma **família** conveniente **de densidades** para uma proporção é a **beta**

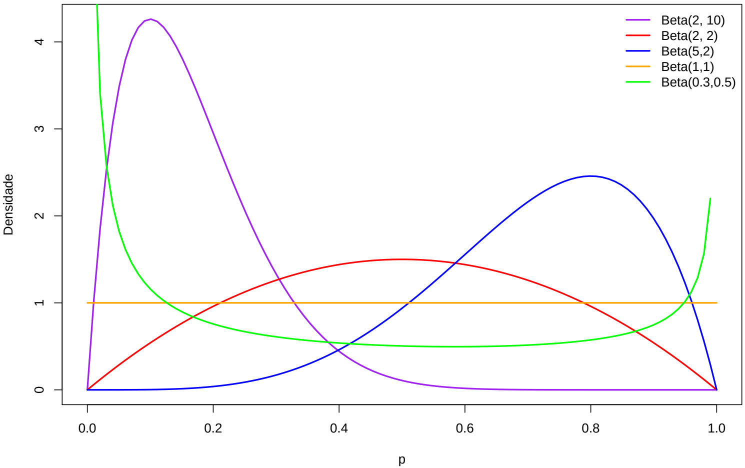
$$
\begin{aligned}
g(p) &= \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1} \\
&\propto p^{a-1}(1-p)^{b-1}, \; p \in (0,1), a > 0, b > 0,
\end{aligned}
$$

em que $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ é a chamada função beta e $\Gamma(t) = \int_0^\infty u^{t-1}e^{-u}du, t > 0$ é a função gama[2].

- Os **hiperparâmetros** a e b são escolhidos para refletir as crenças *a priori* do pesquisador sobre p.

[2] $\Gamma(a) = (a-1)!$ se $a \in \mathbb{N}$

Utilizando uma priori beta



Utilizando uma priori beta

- A **média** de uma distribuição *a priori* beta é $m = a/(a+b)$ e a **variância** *a priori* é $v = m(1-m)/(a+b+1)$
  - Na prática é difícil para um pesquisador avaliar os valores de m e v para obter os valores dos parâmetros da beta a e b.
  - É mais fácil obter a e b indiretamente por meio de declarações sobre os percentis da distribuição ($p_{0.5} = 0.3$ e $p_{0.9} = 0.5$ corresponde à dist. beta com $a = 3.26$ e $b = 7.19$)[3].

[3] Veja a função `beta.select` do pacote `LearnBayes` do R.

Utilizando uma priori beta

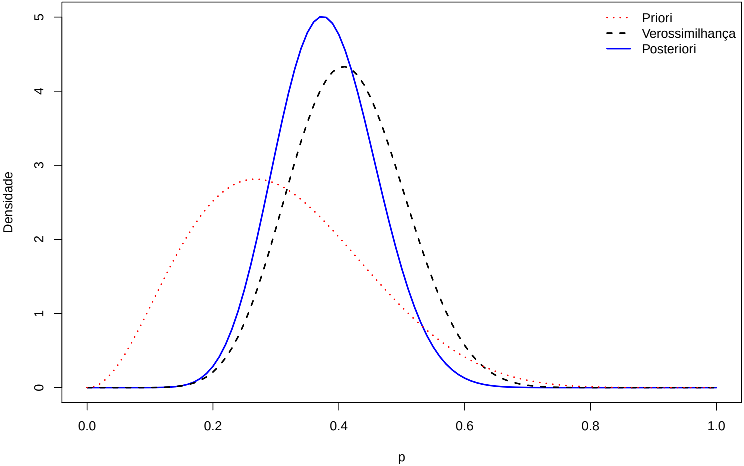
- Combinando este dist. *a priori* beta com a função de verossimilhança, temos a dist. *a posteriori* **também é beta**[4] com parâmetros $a+s$ e $b+f$

$$g(p|\text{dados}) =\propto p^{a+s-1}(1-p)^{b+f-1}, \;\; p \in (0,1),$$

em que $a+s = 3.26 + 11$ e $b+f = 7.19 + 16$.

[4]Este é um exemplo de **análise conjugada**, em que as densidades *a priori* e *a posteriori* têm a mesma forma funcional.

Utilizando uma priori beta



Utilizando uma priori beta

- Existem diferentes maneiras de resumir a distribuição beta *a posteriori* **para fazer inferências sobre a proporção de dorminhocos** p.
- Podemos utilizar a função distribuição acumulad da beta, e a sua inversa, para calcular probabilidades *a posteriori* e construir estimativas intervalares para p.

Utilizando uma priori beta

- É provável que a proporção de pessoas com sono pesado seja maior que 0.5?
  - Isso é respondido calculando a probabilidade *a posteriori* $\Pr(p \geq 0.5 | \text{dados})$:

```
a <- 3.26; b <- 7.19; s <- 11; f <- 16

pbeta(q = 0.5,
      shape1 = a+s, shape2 = b+f,
      lower.tail = FALSE)
```

```
## [1] 0.0690226
```

Utilizando uma priori beta

- Uma estimativa de intervalo de 90% para p é encontrada calculando os percentis 5 e 95 da densidade beta:

```
qbeta(p = c(0.05, 0.95),
      shape1 = a+s, shape2 = b+f)
```

```
## [1] 0.2555267 0.5133608
```

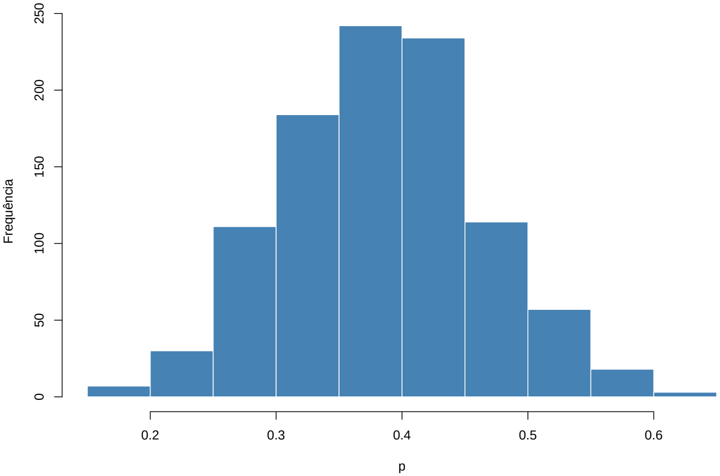
- Um **intervalo de credibilidade** *a posteriori* de 90% para a proporção é (0.256, 0.513).

Utilizando uma priori beta

- Esses resumos são "exatos" porque são baseados em funções R para a densidade beta *a posteriori*.
- Um método alternativo de sumarização de uma densidade *a posteriori* é baseado em **simulação**.
- Nesse caso, podemos simular um grande número de valores da densidade beta *a posteriori* e resumir a **amostra** (saída) **simulada**.

```
amostra.post <- rbeta(n = 1000,
                      shape1 = a+s, shape2 = b+f)
```

Utilizando uma priori beta



Utilizando uma priori beta

- A probabilidade de que a proporção seja maior que 0.5 é aproximada (estimada[5] usando a proporção de valores simulados neste intervalo

$$\frac{1}{M}\sum_{j=1}^{M} I(p^{(j)} \geq 0.5),$$

em que M é o número de valores simulados, $p^{(j)}$ é j-ésimo valor simulado da dist. beta ($j = 1, \ldots, M$) com parâmetros $a + s$ e $b + f$ e $I(\cdot)$ é **função indicadora**.

- Note que para M "grande"

$$\Pr(p \geq 0.5|\text{dados}) = \mathrm{E}\left[I(p \geq 0.5)|\text{dados}\right] \approx \frac{1}{M}\sum_{j=1}^{M} I(p^{(j)} \geq 0.5).$$

[5] estimativa estocástica, ou de Monte Carlo.

Utilizando uma priori beta

- No R:

```
sum(amostra.post >= 0.5)/1000
```

```
## [1] 0.078
```

- Uma estimativa de intervalo de 90% pode ser estimada pelos 5º e 95º quantis amostrais da amostra simulada:

```
quantile(x = amostra.post, probs = c(0.05, 0.95))
```

```
##        5%       95%
## 0.2603226 0.5167994
```

Utilizando uma priori beta

- Observe que esses resumos da densidade *a posteriori* para p baseados em simulação são aproximadamente iguais aos valores exatos baseados em cálculos da distribuição beta.
- A aproximação melhora conforme aumentamos o número de valores simulados M.

Predição

Predição

- Nós nos concentramos em aprender sobre a proporção populacional de pessoas com sono pesado p.
- Suponha que nosso pesquisador também esteja interessado em **prever** o número de dorminhocos $\tilde{y}$ em uma amostra futura de $m = 20$ alunos.

Preditiva a posteriori

- Depois de observarmos os dados e obtermos a distribuição *a posteriori* dos parâmetros, podemos agora usar a distribuição *a posteriori* para gerar **dados futuros** do modelo.
- Em outras palavras, dada a distribuição *a posteriori* dos parâmetros do modelo, a \structure{distribuição preditiva *a posteriori* nos dá uma indicação de como os dados futuros podem parecer, condicional aos dados (y) e o modelo ($L(p)$ e $g(p)$).

Predição

- Uma vez que temos a distribuição posteriori $g(p|$dados, podemos derivar as previsões $(\tilde{y})$ com base nessa distribuição:

$$f(\tilde{y}|y) = \int f(\tilde{y}, p|y)dp = \int f(\tilde{y}|p, y)g(p|y)dp.$$

- Assumindo que as observações passadas e futuras são **condicionalmente independentes**, dado p, ou seja, $f(\tilde{y}|p, y) = f(\tilde{y}|p)$, nós podemos escrever:

$$f(\tilde{y}|y) = \int f(\tilde{y}|p)g(p|y)dp. \tag{1}$$

Predição

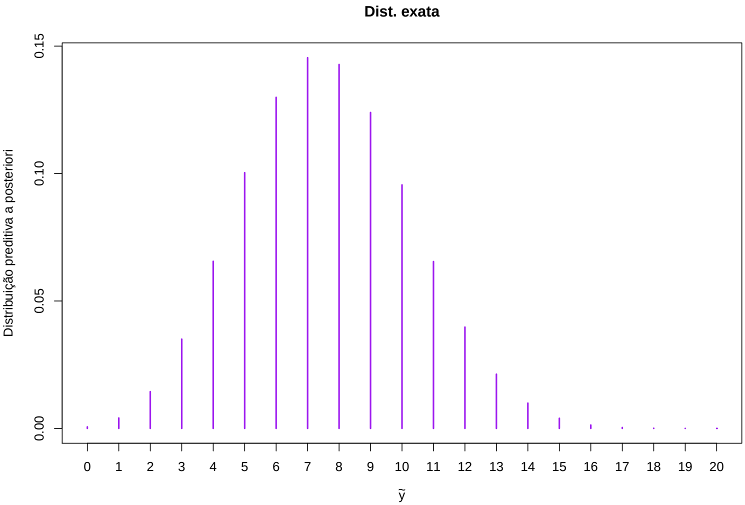
- Na equação (1), estamos condicionando $\tilde{y}$ apenas em y, não condicionamos o que não conhecemos (p); integramos os parâmetros desconhecidos.
- Considerando o modelo anterior (*priori* beta), e um modelo adequado para $f(\tilde{y}|p)$ a distribuição binomial, de parâmetros $m = 20$ (ensaios/entrevistados) e p (probabilidade de ter sono "pesado").
  - Neste caso, podemos integrar analiticamente a equação (1) e obter uma expressão de forma fechada para a densidade preditiva,

$$f(\tilde{y}|y) = \binom{m}{\tilde{y}} \frac{B(a+s+\tilde{y}, b+f+m-\tilde{y})}{B(a+s, b+f)}, \tilde{y} = 0, \ldots, m, \qquad (2)$$

em que $B(a, b)$ é a função beta.

- A distribuição em (2) é conhecida como **beta-binomial**.

Predição


Dist. exata
Distribuição preditiva a posteriori
0.00
0.05
0.10
0.15
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
ỹ

Predição

- Uma maneira conveniente de calcular uma densidade preditiva para qualquer distribuição *a priori* é por simulação. Note que (1) pode ser aproximada por uma média amostral

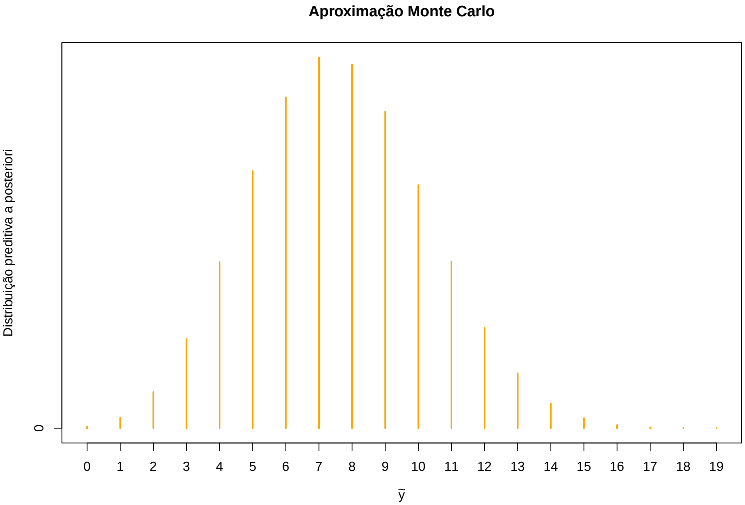
$$f(\tilde{y}|y) = \int f(\tilde{y}|p)g(p|y)dp = \mathrm{E}\left[f(\tilde{y}|p)|y\right] \approx \frac{1}{M}\sum_{j}^{M} f(\tilde{y}|p^{(j)}).$$

- Para obter $\tilde{y}$, primeiro simulamos, digamos, $p^{(j)}$ a partir de $g(p|y)$, e então simulamos $\tilde{y}$ a partir da distribuição binomial $f(\tilde{y}|p^{(j)})$.

```
M <- 1000000

p <- rbeta(n = M, shape1 = a + s, shape2 = b + f)
ytilde.sim <- rbinom(n = M, size = m, prob = p)
```

Predição


Aproximação Monte Carlo
Distribuição preditiva a posteriori
0
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
$\tilde{y}$

Próxima aula

- Modelos de parâmetro único;
- Modelos de múltiplos parâmetros.

Por hoje é só!

Sejam tod@s bem-vind@s!