Introduction to Probability (Joe Blitzstein)

Fernando Náufel

(versão de 17/03/2022)

Sumário

Apresentação

- Página do curso (com *link* para pdf do livro): https://projects.iq.harvard.edu/stat110/home
- *Playlist*: https://www.youtube.com/playlist?list=PL2SOU6wwxB0uwwH80KTQ6ht66KWxbzTIo
- Soluções não oficiais (de alguns capítulos): https://fifthist.github.io/Introduction-To-Probability-Blitzstein-Solutions/

01: Probabilidade e contagem

Vídeo

https://youtu.be/KbB0FjPg0mw

Pascal e Fermat

- Ver artigo DEVLIN (2010).
- Ver originais em francês de toda a correspondência de Pascal.

R

Fatoriais e combinações

- Qual o maior valor de n para o qual o R calcula `factorial(n)`?

  No meu computador, $n = 170$:

  ```
  factorial(170:171)
  ## [1] 7,257416e+306           Inf
  ```

- Para valores maiores, podemos usar `lfactorial(n)` para calcular $\ln n!$:

```r
lfactorial(170:171)
## [1] 706,5731 711,7147
```

- Da mesma forma, `lchoose(n, k)` calcula ln $\binom{n}{k}$.

Tabulando dados: `tabulate` x `table`

```r
b <- sample(1:365,23,replace=TRUE)
tabulate(b)
##   [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0
##  [34] 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##  [67] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [100] 0 0 0 0 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0
## [133] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [166] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [199] 0 0 0 1 0 0 0 0 0 0 0 3 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [232] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 1 0 0 0 0 0 0 0 1 0 0 0
## [265] 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
## [298] 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [331] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2
```

```r
table(b)
## b
##  15  23  28  40  42 110 121 123 148 202 210 247 252 253 261 267 288
##   1   1   1   1   1   2   1   1   1   1   3   1   1   1   1   1   1
## 308 348
##   1   2
```

Funções para o problema do aniversário

```r
pbirthday(23)
## [1] 0,5072972
```

```r
qbirthday(.5)
## [1] 23
```

```r
qbirthday(1)
## [1] 366
```

Para no mínimo 3 no mesmo dia:

```r
qbirthday(.5, coincident = 3)
## [1] 88
```

Exercícios (cap. 1)

---

13. Superbaralho

> A certain casino uses 10 standard decks of cards mixed together into one big deck, which we will call a superdeck. Thus, the superdeck has $52 \cdot 10 = 520$ cards, with 10 copies of each card.
>
> How many different 10-card hands can be dealt from the superdeck? The order of the cards does not matter, nor does it matter which of the original 10 decks the cards came from. Express your answer as a binomial coefficient.

- Usando a notação de OLIVEIRA MORGADO et al. (2004), onde CR_k^n é o número de combinações completas de n elementos de k tipos diferentes, a resposta é

$$\text{CR}_{52}^{10} = \binom{52+10-1}{10} = \binom{61}{10} = 90.177.170.226$$

- Só foi possível usar combinações completas porque a mão tem 10 cartas, o que faz com que haja, essencialmente, um número infinito de cópias de cada um dos 52 tipos de carta. Se a mão tivesse 11 ou mais cartas, seria impossível que todas as cartas fossem iguais, e este raciocínio não poderia ser usado.

---

42. Norepeat words

> A *norepeatword* is a sequence of at least one (and possibly all) of the usual 26 letters a, b, c, ...,z, with repetitions not allowed.
>
> For example, "course" is a *norepeatword*, but "statistics" is not.
>
> Order matters, e.g., "course" is not the same as "source".
>
> A *norepeatword* is chosen randomly, with all *norepeatwords* equally likely. Show that the probability that it uses all 26 letters is very close to $1/e$.

- O denominador vai ser o total de todas as *norepeatwords* (NRW), que é a soma de
  - NRW de 1 letra: 26

  - NRW de 2 letras: $26 \cdot 25$
  - NRW de 3 letras: $26 \cdot 25 \cdot 24$
  - …
  - NRW de 24 letras: $26 \cdot 25 \cdot 24 \cdot \dots \cdot 3$
  - NRW de 25 letras: $26 \cdot 25 \cdot 24 \cdot \dots \cdot 2$
  - NRW de 26 letras: $26 \cdot 25 \cdot 24 \cdot \dots \cdot 1$

- Ou seja,

$$\sum_{k=0}^{25} \frac{26!}{k!}$$

- Que é igual a

$$26! \left(1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{25!}\right)$$

- O total de NRW que usam as 26 letras é $26!$.
- A probabilidade procurada é

$$\begin{aligned} P &= \frac{26!}{26!\left(1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{25!}\right)} \\ &= \frac{1}{1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{25!}} \\ &= \frac{1}{e} \end{aligned}$$

- A última igualdade se justifica porque a série de Taylor para e^x é

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

- Numericamente:

```r
1 / exp(1)
## [1] 0,3678794
```

```r
1 / sum(1 / factorial(0:25))
## [1] 0,3678794
```

02: Histórias e axiomas

Vídeo

https://youtu.be/FJd_1H3rZGg

Exercícios do livro (cap. 1)

17

$$\sum_{k=0}^{n} \binom{n}{k}^2 = \binom{2n}{n}$$

- Quero escolher n pessoas dentre $2n$ pessoas (lado direito).
- Divido as $2n$ pessoas em dois grupos de n cada.
- Para $k \in \{0, 1, \dots n\}$:
  - Escolho k pessoas do primeiro grupo — $\binom{n}{k}$ — para entrar.
  - Escolho k pessoas do segundo grupo — $\binom{n}{k}$ — para não entrar.
  - Para este valor de k, tenho $\binom{n}{k}^2$ maneiras de selecionar n pessoas, com k pessoas do primeiro grupo e $n - k$ pessoas do segundo.

18

$$\sum_{k=1}^{n} k \binom{n}{k}^2 = n \binom{2n-1}{n-1}$$

- Usamos o mesmo raciocínio do exercício anterior, com uma modificação.
- Temos $2n$ pessoas, divididas em dois grupos de n, como antes.
- Como antes, quero escolher n pessoas dentre as $2n$.
- Mas agora, para cada escolha, quero designar uma das n pessoas do primeiro grupo como chefe (i.e., sempre haverá pelo menos uma pessoa do primeiro grupo). Isto corresponde ao fator n no lado direito.
- Escolhido o chefe, preciso escolher $n-1$ pessoas dentre as $2n-1$ restantes ($n-1$ do primeiro grupo, n do segundo). Isto corresponde ao segundo fator do lado direito.
- Do lado esquerdo, para $k \in \{1, 2, \dots, n\}$:
  - Vou escolher k pessoas do primeiro grupo — $\binom{n}{k}$ — para entrar.
  - Dentre elas, vou escolher um chefe — k.
  - Vou escolher k pessoas do segundo grupo — $\binom{n}{k}$ — para não entrar.

19

$$\sum_{k=2}^{n} \binom{k}{2} \binom{n-k+2}{2} = \binom{n+3}{5}, \quad \forall n \geq 2$$

- Um exemplo concreto: $n = 3$
- Queremos escolher 5 elementos dentre 6.
- Para $k \in \{2, 3\}$:
  1. Escolhemos sempre o elemento $k+1$:

$$1 \quad 2 \quad \underbrace{3}_{k=2} \quad 4 \quad 5 \quad 6$$
$$1 \quad 2 \quad 3 \quad \underbrace{4}_{k=3} \quad 5 \quad 6$$

     Este é o elemento que vai estar no meio do grupo dos escolhidos.

2. Escolhemos 2 dentre os k primeiros elementos:

$$\underbrace{1 \quad 2}_{k=2} \quad 3 \quad 4 \quad 5 \quad 6$$
$$\underbrace{1 \quad 2 \quad 3}_{k=3} \quad 4 \quad 5 \quad 6$$

3. Escolhemos 2 dentre os $n - k + 2$ últimos elementos:

$$1 \quad 2 \quad 3 \quad \underbrace{4 \quad 5 \quad 6}_{k=2}$$
$$1 \quad 2 \quad 3 \quad 4 \quad \underbrace{5 \quad 6}_{k=3}$$

---

20. Teorema das colunas

(a) Mostre que

$$\binom{k}{k} + \binom{k+1}{k} + \binom{k+2}{k} + \cdots + \binom{n}{k} = \binom{n+1}{k+1}$$

- O lado direito significa escolher $k+1$ pessoas dentre $n+1$ pessoas.
- O truque é ordenar as pessoas de algum modo.
- Um exemplo concreto, com $n = 4$ e $k = 2$, mostrando que

$$\binom{5}{3} = \binom{4}{2} + \binom{3}{2} + \binom{2}{2}$$

1. Vamos chamar as $n+1$ pessoas de $1, 2, 3, 4, 5$.
2. Grupos de $k+1$ pessoas onde o menor número é 1:
   - $1, 2, 3$
   - $1, 2, 4$
   - $1, 2, 5$
   - $1, 3, 4$
   - $1, 3, 5$
   - $1, 4, 5$
3. Grupos de $k+1$ pessoas onde o menor número é 2:
   - $2, 3, 4$
   - $2, 3, 5$
   - $2, 4, 5$

4. Grupos de $k+1$ pessoas onde o menor número é 3:
    - $3, 4, 5$

- No caso geral, vamos ordenar as $n+1$ pessoas, rotulando-as como

$$a_0, a_1, a_2, \dots, a_n$$

- Como a ordem dentro de cada grupo não importa, vamos escolher primeiro um elemento para ser o de menor índice do grupo e escolher os restantes k elementos dentre os elementos de índice maior do que o primeiro.
- Se escolhermos a_0 como o de menor índice, temos $\binom{n}{k}$ modos de escolher os restantes.
- Se escolhermos a_1 como o de menor índice, temos $\binom{n-1}{k}$ modos de escolher os restantes.
- …
- Se escolhermos $a_{n-(k+1)}$ como o de menor índice, temos $\binom{k+1}{k}$ modos de escolher os restantes.
- Se escolhermos a_{n-k} como o de menor índice, temos $\binom{k}{k}$ modos de escolher os restantes.

(b) Suppose that a large pack of Haribo gummi bears can have anywhere between 30 and 50 gummi bears. There are 5 delicious flavors. How many possibilities are there for the composition of such a pack of gummi bears?

- Usando a notação de OLIVEIRA MORGADO et al. (2004), onde CR_k^n é o número de combinações completas de n elementos de k tipos diferentes, a resposta é

$$\begin{aligned}
\text{CR}_5^{30} + \text{CR}_5^{31} + \cdots + \text{CR}_5^{50} &= \binom{34}{4} + \binom{35}{4} + \cdots + \binom{54}{4} \\
&= \binom{55}{5} - \left[\binom{33}{4} + \binom{32}{4} + \cdots + \binom{4}{4}\right] \\
&= \binom{55}{5} - \binom{34}{5}
\end{aligned}$$

22. Soma de cubos

Para provar

$$1^3 + 2^3 + \cdots + n^3 = (1 + 2 + \cdots + n)^2$$

a. Vamos provar

$$1 + 2 + \cdots + n = \binom{n+1}{2}$$

b. E vamos provar

$$1^3 + 2^3 + \cdots + n^3 = 6\binom{n+1}{4} + 6\binom{n+1}{3} + \binom{n+1}{2}$$

a. Considere $n+1$ times, numerados de 0 a n, num campeonato onde cada time joga com todos os outros exatamente uma vez. O total de jogos será

$$\binom{n+1}{2}$$

- O time 0 joga com n times com número maior que ele.
- O time 1 joga com $n-1$ times com número maior que ele.
- O time 2 joga com $n-2$ times com número maior que ele.
- …
- O time $n-2$ joga com 2 times com número maior que ele.
- O time $n-1$ joga com 1 time com número maior que ele.
- O time n joga com 0 times com número maior que ele.

b. Hint: Imagine choosing a number between 1 and n and then choosing 3 numbers between 0 and n smaller than the original number, with replacement. Then consider cases based on how many distinct numbers were chosen.

- Vamos chamar o número escolhido de k.
- Para $k = 1$, só temos o 0 para escolher. É 1 possibilidade.
- Para $k = 2$, temos 2 números. São 2^3 possibilidades.
- Para $k = 3$, temos 3 números. São 3^3 possibilidades.
- No geral, para $k = n$, são n^3 possibilidades.

- A soma de todos os casos é a soma dos cubos.
- Vamos examinar o caso em que k foi o número escolhido. De quantas maneiras podemos escolher 3 números entre 0 e $k-1$, com reposição?
  - Com 1 número, basta escolher o número: $\binom{k}{1}$ possibilidades.
  - Com 2 números distintos:
    * Escolhemos os 2 números: $\binom{k}{2}$ possibilidades;
    * Escolhemos qual dos 2 aparecerá 2 vezes: 2 possibilidades;
    * Escolhemos a posição do número que aparece uma vez: 3 possibilidades.
    * São $6\binom{k}{2}$ possibilidades.
  - Com 3 números distintos:
    * Escolhemos os 3 números: $\binom{k}{3}$ possibilidades;
    * Escolhemos a ordem dos 3 números: 6 possibilidades;
    * São $6\binom{k}{3}$ possibilidades.
- Para todos os valores de k, o total de possibilidades é

$$\sum_{k=1}^{n}\left[\binom{k}{1}+6\binom{k}{2}+6\binom{k}{3}\right]$$

- Usando o teorema das colunas, chegamos ao resultado

$$\binom{n+1}{2}+6\binom{n+1}{3}+6\binom{n+1}{4}$$

03: Problema do aniversário, propriedades

Vídeo

https://youtu.be/LZ5Wergp_PA

Exercícios do livro (cap. 1)

26

Amostra com reposição de tamanho 1000 a partir de uma população de tamanho 1 milhão. Cada pessoa tem a mesma probabilidade de ser escolhida.

Qual a probabilidade de pelo menos uma pessoa ser escolhida mais de uma vez?

- Usando $N = 1.000.000$ e $n = 1.000$:

$$\begin{aligned} P &= 1 - P(\text{ninguém escolhido mais de uma vez}) \\ &= 1 - \frac{\binom{N}{n} \cdot n!}{N^n} \\ &= 1 - \frac{N}{N} \cdot \frac{N-1}{N} \cdot \frac{N-2}{N} \cdot \ldots \cdot \frac{N-n+1}{N} \end{aligned}$$

- Na verdade, esta é a mesma distribuição do problema dos aniversários, com N dias e n pessoas.

N	n	P
1.000.000	1.000.000	1,0000000
1.000.000	100.000	1,0000000
1.000.000	10.000	1,0000000
1.000.000	1.000	0,3932670
1.000.000	100	0,0049379
1.000.000	10	0,0000450
1.000.000	1	0,0000000

- Para N grande e n pequeno, a probabilidade P de pelo menos uma pessoa ser escolhida mais de uma vez se aproxima de 0:

27

> - Função de *hash* $h(x)$.
> - k telefones armazenados em n posições, todas com a mesma probabilidade.
> - Qual a probabilidade de colisão?

- De novo, `pbirthday(k, n)`.

57

> - Existem 10^{44} moléculas na atmosfera.
> - No seu último suspiro, César respirou 10^{22} delas (sem reposição).
> - Você respira 10^{22} moléculas agora (com reposição).
> - Qual a probabilidade de que pelo menos uma molécula sua tenha sido de César também?

- A resposta é $1 - P(\text{nenhuma molécula sua foi de César})$.
- Vamos calcular $P(\text{nenhuma molécula sua foi de César})$.
- Todas as respiradas possíveis (com reposição, com ordem): $(10^{44})^{(10^{22})}$.
- Todas as respiradas sem moléculas de César (com reposição, com ordem): $(10^{44} - 10^{22})^{(10^{22})}$.
- Daí,

$$
\begin{aligned}
P(\text{nenhuma molécula sua foi de César}) &= \frac{(10^{44} - 10^{22})^{(10^{22})}}{(10^{44})^{(10^{22})}} \\
&= \left(\frac{10^{44} - 10^{22}}{10^{44}}\right)^{(10^{22})} \\
&= \left(1 - \frac{1}{10^{22}}\right)^{(10^{22})} \\
&\approx e^{-1}
\end{aligned}
$$

- Logo, a probabilidade procurada é aproximadamente

$$1 - e^{-1} \approx 0{,}63$$

- Formulação original em PAULOS (1988):

> Now for better news of a kind of immortal persistence. First, take a deep breath. Assume Shakespeare's account is accurate and Julius Caesar gasped "You too, Brutus" before breathing his last. What are the chances you just inhaled a molecule which Caesar exhaled in his dying breath? The surprising answer is that, with probability better than 99 percent, you did just inhale such a molecule.
>
> For those who don't believe me: I'm assuming that after more than two thousand years the exhaled molecules are uniformly spread about the world and the vast majority are still free in the atmosphere. Given these reasonably valid assumptions, the problem of determining the relevant probability is straightforward. If there are N molecules of air in the world and Caesar exhaled A of them, then the probability that any given molecule you inhale is from Caesar is A/N. The probability that any given molecule you inhale is not from Caesar is thus 1 – A/N. By the multiplication principle, if you inhale three molecules, the probability that none of these three is from Caesar is $[1 - A/N]^3$. Similarly, if you inhale B molecules, the probability that none of them is from Caesar is approximately $[1 - A/N]^B$. Hence, the probability of the complementary event, of your inhaling at least one of his exhaled molecules, is $1 - [1 - A/N]^B$. A, B (each about 1/30th of a liter, or 2.2×10^{22}), and N (about 10^{44} molecules) are such that this probability is more than .99. It's intriguing that we're all, at least in this minimal sense, eventually part of one another.

61

- n passageiros para n assentos.
- Passageiro k inicialmente alocado no assento k.
- MAS Passageiro 1 decide escolher assento ao acaso (cada assento com a mesma probabilidade).
- Então, cada passageiro seguinte se senta no assento inicialmente alocado para ele, se disponível; caso contrário, escolhe um assento ao acaso.
- Qual a probabilidade de que o último passageiro se sente no assento alocado para ele?

Possibilidades

- O último passageiro só pode se sentar no lugar 1 ou no lugar n.
- Aliás, isto é um caso específico de um fenômeno mais geral: qualquer que seja o valor de n:
  - O passageiro 3 nunca fica no assento 2: ou o passageiro 1 pegou, ou ficou livre para o passageiro 2 (que é obrigado a pegá-lo).
  - O passageiro 4 nunca fica nos assentos 2 nem 3: ou o passageiro 1 pegou o assento 3, ou aconteceu um dos casos acima e o assento 3 ficou livre para o passageiro 3 (que é obrigado a pegá-lo).
  - O passageiro 5 nunca fica nos assentos 2 nem 3 nem 4: ou o passageiro 1 pegou o assento 4, ou aconteceu um dos casos acima e o assento 4 ficou livre para o passageiro 4 (que é obrigado a pegá-lo).
  - etc.
- Seguindo o raciocínio, chegamos à conclusão de que o passageiro n só pode ocupar o assento 1 ou o assento n.

Probabilidades

- Examinando exemplos com $n \in \{3, 4, 5\}$, a probabilidade parece ser $1/2$.
- Por quê?
- O assento do passageiro n depende da pergunta (no momento em que o passageiro n vai se sentar)

  "O assento 1 já foi tomado?"

  cuja resposta é oposta à da pergunta

  "O assento n já foi tomado?"

- Com que probabilidade a resposta a esta última pergunta é sim?
- Em todos os passos anteriores, como explicado acima em “possibilidades”, os assentos 1 e n sempre estão disponíveis para qualquer passageiro que vá escolher um assento ao acaso.
- Como todos os assentos têm a mesma probabilidade de ser escolhidos, a probabilidade de o assento 1 estar tomado é igual à probabilidade de o assento n estar tomado.
- Logo, a probabilidade de o passageiro n acabar no assento n é $1/2$.

62

Problema do aniversário com probabilidades diferentes:

- Seja $\vec{p} = (p_1, p_2, \dots, p_{365})$ o vetor das probabilidades de alguém nascer em algum dos dias do ano.
- Seja

$$e_k(x_1, \dots, x_n) = \sum_{1 \leq j_1 < j_2 < \cdots < j_k \leq n} x_{j_1} \dots x_{j_k}$$

o k-ésimo polinômio simétrico elementar sobre as variáveis $x_1, \dots, x_n$.

Por exemplo,

$$\begin{aligned} e_1(x_1, x_2, x_3) &= x_1 + x_2 + x_3 \\ e_2(x_1, x_2, x_3) &= x_1x_2 + x_1x_3 + x_2x_3 \\ e_3(x_1, x_2, x_3) &= x_1x_2x_3 \end{aligned}$$

a. Seja $k \geq 2$. Qual a probabilidade de pelo menos uma coincidência de aniversários em termos de $\vec{p}$ e de um polinômio simétrico elementar?

- A probabilidade de não haver coincidência é a soma das probabilidades de todos os eventos da forma

  As pessoas $1, 2, \dots k$ nasceram em dias diferentes $j_1, j_2, \dots, j_k$.
- Por exemplo, para $k = 3$, a soma será

$$3! \cdot (p_1p_2p_3 + p_1p_2p_4 + \cdots + p_{363}p_{364}p_{365})$$

- Isto é igual a

$$k! \cdot e_k(\vec{p})$$

- A probabilidade de pelo menos uma coincidência é

$$1 - k! \cdot e_k(\vec{p})$$

b. Quando $p_j = 1/365$ para todo j, esta probabilidade é mínima.

- No caso em que $p_j = p = 1/365$ para todo j, temos

$$\begin{aligned} 1 - k! \cdot e_k(\vec{p}) &= 1 - k! \cdot \binom{365}{k} \cdot p^k \\ &= 1 - \frac{365 \cdot 364 \cdot \dots \cdot (365 - k + 1)}{365^k} \end{aligned}$$

- Neste caso, o valor de $e_k(\vec{p})$ é máximo.

c. Considere a desigualdade

$$\frac{x+y}{2} \geq \sqrt{xy}$$

que vem do raciocínio

$$\begin{aligned}(x+y)^2 - 4xy = (x-y)^2 \geq 0 &\implies (x+y)^2 - 4xy \geq 0 \\ &\implies (x+y)^2 \geq 4xy \\ &\implies \frac{(x+y)^2}{4} \geq xy \\ &\implies \frac{x+y}{2} \geq \sqrt{xy}\end{aligned}$$

Defina $\vec{r} = (r_1, \dots, r_{365})$ tal que

- $r_1 = r_2 = (p_1 + p_2)/2$,
- $r_j = p_j$ para $3 \leq j \leq 365$.

Verifique que

$$\begin{aligned}e_k(x_1, \dots, x_n) = x_1 x_2 e_{k-2}(x_3, \dots x_n) + \\ (x_1 + x_2) e_{k-1}(x_3, \dots x_n) + \\ e_k(x_3, \dots x_n)\end{aligned}$$

e use a desigualdade para mostrar que, quando $\vec{p} \neq \vec{r}$,

$$P(\text{coincidência} \mid \vec{p}) > P(\text{coincidência} \mid \vec{r})$$

- Observe que

$$\frac{x+y}{2} = \sqrt{xy} \implies x = y$$

Ou seja, quando $x \neq y$, a desigualdade é estrita.

- Para o vetor $\vec{p}$:

$$\begin{aligned}&P(\text{coincidência} \mid \vec{p}) \\ &= 1 - k! \cdot e_k(\vec{p}) \\ &= 1 - k! \cdot [p_1 p_2 \cdot e_{k-2}(p_3, \dots p_n) + (p_1 + p_2) \cdot e_{k-1}(p_3, \dots p_n) + e_k(p_3, \dots p_n)]\end{aligned}$$

- Para o vetor $\vec{r}$:

$$
\begin{aligned}
&P(\text{coincidência} \mid \vec{r}) \\
&= 1 - k! \cdot e_k(\vec{r}) \\
&= 1 - k! \cdot \left[\frac{(p_1 + p_2)^2}{4} \cdot e_{k-2}(p_3, \dots p_n) + (p_1 + p_2) \cdot e_{k-1}(p_3, \dots p_n) + e_k(p_3, \dots p_n)\right]
\end{aligned}
$$

- A única diferença está no primeiro fator dentro dos colchetes.
- Pela desigualdade:

$$\frac{p_1 + p_2}{2} \geq \sqrt{p_1 p_2} \implies \frac{(p_1 + p_2)^2}{4} \geq p_1 p_2$$

- Conclusão: para todo vetor $\vec{p}$ de probabilidades, é possível preservar ou diminuir a probabilidade de uma coincidência substituindo quaisquer 2 de suas componentes pelo valor da média aritmética delas. Assim, o vetor que minimiza a probabilidade precisa ter todas as suas componentes iguais.

04: Probabilidade condicional

Vídeo

https://youtu.be/P7NE4WF8j-Q

Exercícios do livro (cap. 2)

2

- Uma mulher está grávida de meninos gêmeos.
- Gêmeos podem ser idênticos ou fraternos.
- $1/3$ dos gêmeos são idênticos.
- Gêmeos idênticos têm 50% de chance de serem ambos meninos, e 50% de chance de serem ambos meninas.
- Para gêmeos fraternos, cada gêmeo tem, independentemente, 50% de chance de ser menino, e 50% de chance de ser menina.
- Qual a probabilidade de a mulher estar grávida de gêmeos idênticos?

- Eventos:

$$
\begin{aligned}
I &= \text{os gêmeos são idênticos} \\
M &= \text{os gêmeos são dois meninos}
\end{aligned}
$$

- Probabilidades dadas:

$$
\begin{aligned}
P(I) &= 1/3 \\
P(M \mid I) &= 1/2 \\
P(M \mid \neg I) &= 1/4
\end{aligned}
$$

- Queremos calcular $P(I \mid M)$. Usamos Bayes:

$$
P(I \mid M) = \frac{P(M \mid I) \cdot P(I)}{P(M)}
$$

- Só falta o valor de $P(M)$, que calculamos usando probabilidade total:

$$
\begin{aligned}
P(M) &= P(M \mid I)P(I) + P(M \mid \neg I)P(\neg I) \\
&= 1/2 \cdot 1/3 + 1/4 \cdot 2/3 \\
&= 1/3
\end{aligned}
$$

- Concluímos

$$
\begin{aligned}
P(I \mid M) &= \frac{P(M \mid I) \cdot P(I)}{P(M)} \\
&= \frac{1/2 \cdot 1/3}{1/3} \\
&= 1/2
\end{aligned}
$$

4

- Um aluno está resolvendo uma questão de múltipla escolha com n opções.
- K é o evento que corresponde a ele saber a resposta correta.
- R é o evento que corresponde a ele acertar (sabendo ou chutando).
- Se ele sabe a resposta, ele acerta.
- Se ele não sabe, ele chuta uma opção ao acaso.
- Considere $P(K) = p$.

(a) Ache $P(K \mid R)$.

(b) Mostre que $P(K \mid R) \geq p$. Por que isto faz sentido? Quando a igualdade é exata?

a. Vamos usar Bayes, supondo $p = P(K) > 0$:

$$\begin{aligned} P(K \mid R) &= \frac{P(R \mid K) \cdot P(K)}{P(R)} \\ &= \frac{1 \cdot p}{P(R \mid K) \cdot P(K) + P(R \mid \neg K) \cdot P(\neg K)} \\ &= \frac{p}{1 \cdot p + \frac{1}{n} \cdot (1-p)} \\ &= \frac{np}{np + (1-p)} \end{aligned}$$

b. Isto equivale a mostrar

$$\frac{np}{np + (1-p)} \geq p$$

Como todos os valores são positivos, isto equivale a

$$np \geq np^2 + (1-p)p$$

Dividindo ambos os lados por $p > 0$:

$$n \geq np + 1 - p$$

Daí,

$$
\begin{aligned}
n \geq np + 1 - p &\iff n - np \geq 1 - p \\
&\iff n(1-p) \geq 1 - p \\
&\iff n \geq 1 \qquad\qquad \text{(supondo } p < 1\text{)}
\end{aligned}
$$

O que é verdade, pois existe pelo menos uma opção (a correta).

Se $p = 1$, a igualdade vale, e $P(K \mid R) = 1$, pois o aluno sempre sabe a resposta e sempre acerta.

Se $n = 1$, a igualdade vale, e $P(K \mid R) = p$, pois o aluno acertar (ele sempre acerta!) e o aluno saber são independentes.

Imagine $n > 1$.

Lembre-se de que $P(K \mid R) = \frac{P(K \cap R)}{P(R)}$.

O numerador é a probabilidade do evento "o aluno sabe e o aluno acerta". Esta probabilidade é exatamente p, pois "o aluno sabe" é sub-evento de "o aluno acerta".

O numerador $P(R)$ é um número tal que $0 < P(R) < 1$. Vamos chamar este número de $1/x$, para $x > 1$.

Então,

$$P(K \mid R) = \frac{P(K \cap R)}{P(R)} = \frac{p}{1/x} = xp > p$$

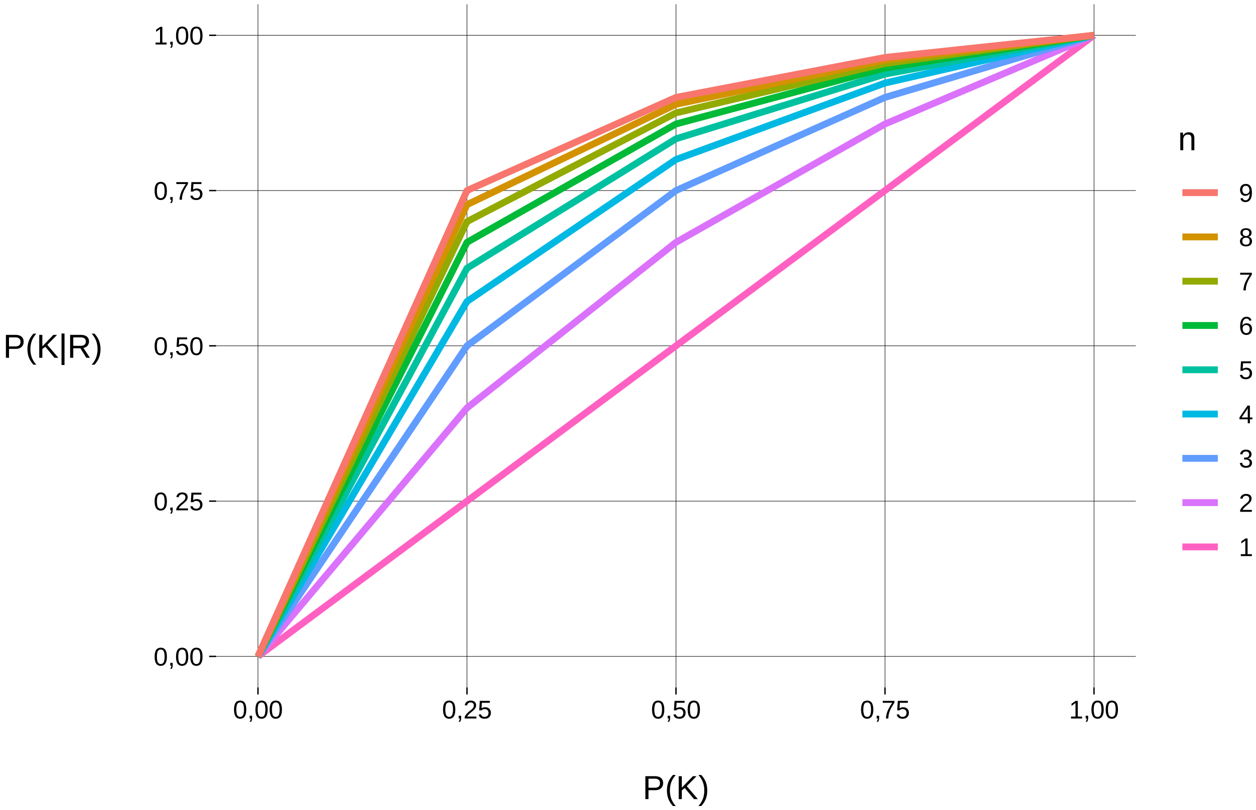
Em R

Valores teóricos:

```r
n <- 1:9
p <- seq(0, 1, .25)

df <- expand.grid(n = n, p = p) %>%
  mutate(
    pkr_teorico = n * p / (p * (n - 1) + 1)
  ) %>%
  arrange(n)

df
## # A tibble: 45 x 3
##       n     p pkr_teorico
##   <int> <dbl>       <dbl>
## 1     1  0           0
## 2     1  0.25        0.25
## 3     1  0.5         0.5
## 4     1  0.75        0.75
## 5     1  1           1
```

```
## 6      2  0            0
## # ... with 39 more rows
```

Gráfico:



Conclusão óbvia: quanto maior o número n de opções, maior a probabilidade $P(K \mid R)$ de o aluno ter acertado sabendo, em oposição a ter acertado chutando:

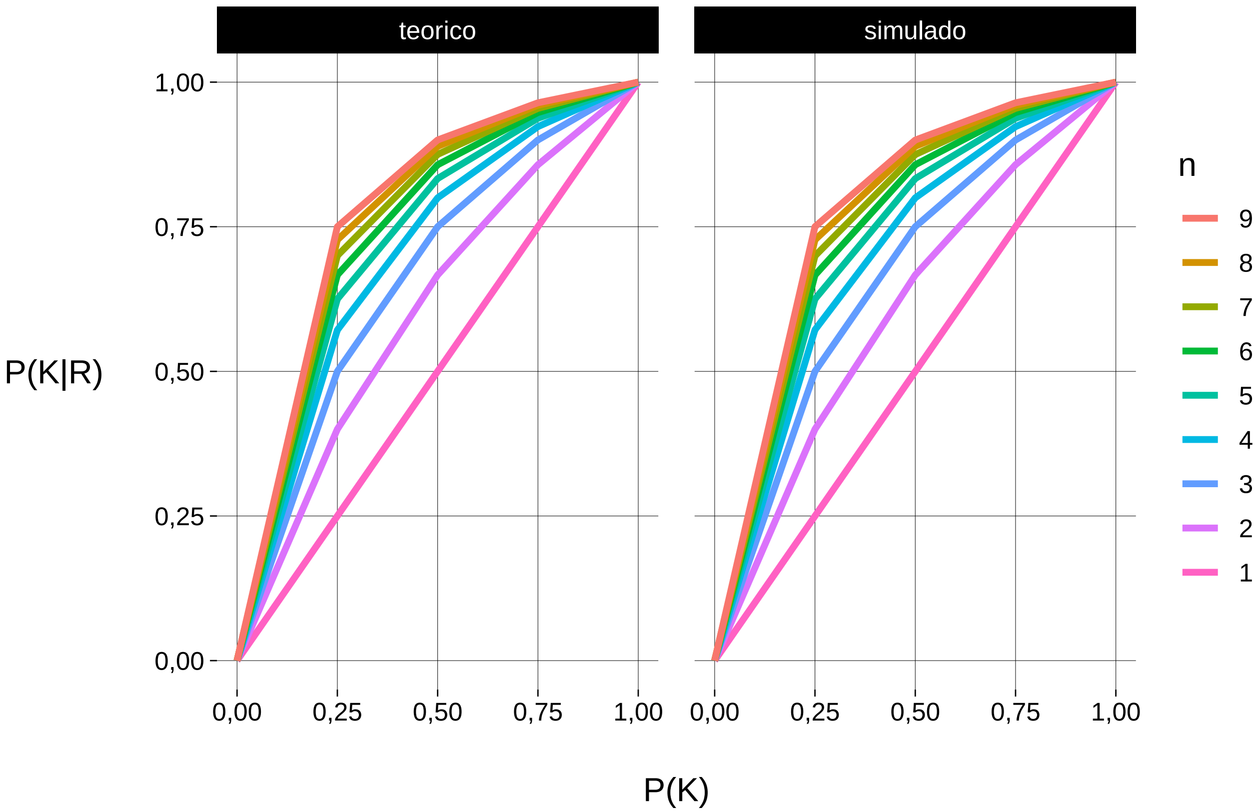
$$
\begin{aligned}
\lim_{n\to\infty} P(K \mid R) &= \lim_{n\to\infty} \frac{np}{np+1-p} \\
&= \lim_{n\to\infty} \frac{p}{p} \\
&= 1
\end{aligned}
$$

Simulação:

```
sim <- function(n, p, reps = 1e7) {

  sabe <- sample(
    c(TRUE, FALSE),
    reps,
    replace = TRUE,
```

```
    c(p, 1 - p)
  )

  acerta <- sabe
  nao_sabe <- sum(!sabe)

  acerta[which(!sabe)] <-
    sample(
      c(TRUE, FALSE),
      nao_sabe,
      replace = TRUE,
      prob = c(1/n, 1 - 1/n)
    )

  sum(sabe) / sum(acerta)

}

sim <- Vectorize(sim)
```

```
df <- df %>%
  mutate(
    pkr_simulado = sim(n, p)
  )

df
## # A tibble: 45 x 4
##       n     p pkr_teorico pkr_simulado
##   <int> <dbl>       <dbl>        <dbl>
## 1     1  0           0            0
## 2     1  0.25        0.25         0.250
## 3     1  0.5         0.5          0.500
## 4     1  0.75        0.75         0.750
## 5     1  1           1            1
## 6     2  0           0            0
## # ... with 39 more rows
```



12

> a. Alice está tentando enviar uma mensagem codificada em binário para Bob.
> - Ela envia um *bit*: 0 ou 1 com probabilidades iguais.
> - Se ela envia 0, há probabilidade 5% de erro.
> - Se ela envia 1, há probabilidade 10% de erro.
> - Dado que Bob recebeu 1, qual a probabilidade de Alice ter enviado 1?

- Eventos e probabilidades:

p_0 = prob. of $0 \to 1$ error $q_0 = 1 - p_0$ = prob. of 0 transmitted OK.
p_1 = prob. of $1 \to 0$ error $q_1 = 1 - p_1$ = prob. of 1 transmitted OK.

A = Alice transmits 1 $\neg A$ = Alice transmits 0 $P(A) = 1/2$
B = Bob receives 1 $\neg B$ = Bob receives 0

- Queremos achar $P(A \mid B)$. Usando Bayes:

a) $P(A|B) = ?$

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

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)\cdot q_1}{P(A)} = q_1$$

$$P(B) = P(B|A)P(A) + P(B|\neg A)P(\neg A)$$

$$= q_1 P(A) + \frac{P(\neg A)\cdot p_0}{P(\neg A)} \cdot P(\neg A)$$

$$= P(A)q_1 + P(\neg A)p_0$$

$$\boxed{\begin{array}{l}\text{If } P(A) = P(\neg A) = 1/2 \\ \text{then } P(B) = \frac{1}{2}(q_1 + p_0)\end{array}}$$

```
p0 <- .05
p1 <- .1
q0 <- 1 - p0
q1 <- 1 - p1

pA <- 1/2
pBIA <- q1
pB <- (q1 + p0) / 2
pAIB <- pBIA * pA / pB

pAIB
## [1] 0,9473684
```

- Para sermos completos, vamos calcular outras probabilidades, com $P(A) = P(\neg A) = 1/2$:

- Simulação:

$$P(A|B) = \frac{q_1 \cdot P(A)}{P(A)q_1 + P(\neg A)\,p_0}$$

If $P(A) = P(\neg A) = 1/2$:

$$P(A|B) = \frac{q_1}{q_1 + p_0}$$

$$P(B|A) = q_1$$

$$\boxed{\frac{p_0}{q_1 + p_0}}$$

$$P(\neg A|B) = 1 - \frac{q_1}{q_1 + p_0}$$

$$P(\neg B|A) = 1 - q_1 = p_1$$

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

$$= \frac{p_1 \cdot P(A)}{1 - \frac{(q_1 + p_0)}{2}}$$

$$\boxed{\frac{p_1}{p_1 + q_0}} = \frac{p_1}{2 - q_1 - p_0}$$

$$P(\neg A|\neg B) = 1 - \frac{p_1}{2 - q_1 - p_0}$$

$$= \frac{2 - q_1 - p_0 - p_1}{2 - q_1 - p_0}$$

$$\frac{p_0}{p_1 + q_0} \leftarrow = \frac{1 - q_1 - p_1 + 1 - p_0}{1 - q_1 + 1 - p_0}$$

- Simulação:

```r
reps <- 1e7
alice_envia <- sample(c(0, 1), reps, replace = TRUE)
bob_recebe <- alice_envia

bob_recebe[which(alice_envia == 0)] <-
  sample(
    c(0, 1),
    length(which(alice_envia == 0)),
    replace = TRUE,
    prob = c(95/100, 5/100)
  )

bob_recebe[which(alice_envia == 1)] <-
  sample(
    c(0, 1),
    length(which(alice_envia == 1)),
    replace = TRUE,
    # Atenção: aqui, erro é 1 virar 0:
    prob = c(10/100, 90/100)
  )

pab <- sum(alice_envia & bob_recebe) / sum(bob_recebe)
pab
## [1] 0,9473928
```

b. Agora, eles usam um código com repetição:

- Alice envia 000 para representar 0 e 111 para representar 1.
- Bob decodifica a mensagem tomando o *bit* que está em maioria.
- As probabilidades de erro são como antes, e os erros em *bits* diferentes são independentes.
- Dado que Bob recebe 110, qual a probabilidade de que Alice tenha enviado 111?

- Eventos:

$$\begin{aligned} AAA &= \text{Alice envia } 111 \\ BB\neg B &= \text{Bob recebe } 110 \end{aligned}$$

- $P(AAA) = 1/2$, pois Alice envia somente 111 ou 000.
- Usando Bayes:

$$P(AAA \mid BB\neg B) = \frac{P(BB\neg B \mid AAA) \cdot P(AAA)}{P(BB\neg B)}$$

- $P(BB\neg B \mid AAA) = q_1 \cdot q_1 \cdot p_1$, pois erros em *bits* diferentes são independentes.
- Pela lei da probabilidade total:

$$
\begin{aligned}
P(BB\neg B) &= P(BB\neg B \mid AAA) \cdot P(AAA) + P(BB\neg B \mid \neg(AAA))P(\neg(AAA)) \\
&= \frac{1}{2} \cdot (q_1 \cdot q_1 \cdot p_1 + p_0 \cdot p_0 \cdot q_0)
\end{aligned}
$$

- Daí,

$$
\begin{aligned}
P(AAA \mid BB\neg B) &= \frac{P(BB\neg B \mid AAA) \cdot P(AAA)}{P(BB\neg B)} \\
&= \frac{q_1 \cdot q_1 \cdot p_1 \cdot 1/2}{(q_1 \cdot q_1 \cdot p_1 + p_0 \cdot p_0 \cdot q_0) \cdot 1/2} \\
&= \frac{q_1 \cdot q_1 \cdot p_1}{(q_1 \cdot q_1 \cdot p_1 + p_0 \cdot p_0 \cdot q_0)}
\end{aligned}
$$

- Numericamente:

```
q1 * q1 * p1 / (q1 * q1 * p1 + p0 * p0 * q0)
## [1] 0,9715142
```

14

Se $P(A), P(B) \in (0, 1)$, então

$$
P(A \mid B) > P(A \mid \neg B) \iff P(B \mid A) > P(B \mid \neg A)
$$

???

15

A e B são eventos com

$$0 < P(A \cap B) < P(A) < P(B) < P(A \cup B) < 1$$

Você está torcendo para que A e B ambos ocorram.

O que você ficaria mais feliz em observar?

- Que A ocorreu?
- Que B ocorreu?
- Que $A \cup B$ ocorreu?

- Queremos observar o evento E tal que $P(A \cap B \mid E)$ seja máximo.

$$\begin{aligned}
P(A \cap B \mid A) &= \frac{P(A \cap B)}{P(A)} \\
P(A \cap B \mid B) &= \frac{P(A \cap B)}{P(B)} \\
P(A \cap B \mid A \cup B) &= \frac{P((A \cap B) \cap (A \cup B))}{P(A \cup B)} \\
&= \frac{P(A \cap B)}{P(A \cup B)}
\end{aligned}$$

- Como $P(A)$ é o menor denominador, observar A maximiza $P(A \cap B \mid E)$.
- Intuitivamente, como A é o evento menos provável dos dois, saber que A ocorreu nos deixa mais próximo da ocorrência dos dois eventos do que saber que B (ou que algum dos dois) ocorreu.

16

$$P(A \mid B) \leq P(A) \implies P(A \mid \neg B) \geq P(A)$$

- Se a ocorrência de B torna A menos provável, então a não-ocorrência de B torna A mais provável.
- Pela lei da probabilidade total:

$$
\begin{aligned}
P(A) &= P(A \mid B)P(B) + P(A \mid \neg B)P(\neg B) \\
\therefore P(A \mid \neg B) &= \frac{P(A) - P(A \mid B)P(B)}{P(\neg B)}
\end{aligned}
$$

- Daí,

$$
\begin{aligned}
P(A \mid \neg B) &\geq \frac{P(A) - P(A)P(B)}{P(\neg B)} \quad \text{pois } P(A \mid B) \leq P(A) \\
&= \frac{P(A)(1 - P(B))}{P(\neg B)} \\
&= P(A)
\end{aligned}
$$

17

> Em lógica determinística, $A \to B \iff \neg B \to \neg A$.
>
> Em probabilidades?
>
> Considere eventos A e B com $P(A), P(B) \notin \{0, 1\}$.
>
> a. Mostre que $P(B \mid A) = 1 \implies P(\neg A \mid \neg B) = 1$.

- A está contido em B:

$$
P(B \mid A) = 1 \iff \frac{P(A \cap B)}{P(A)} = 1 \iff P(A \cap B) = P(A)
$$

- Vamos mostrar que $P(\neg A \mid \neg B) = 1$:

$$
\begin{aligned}
P(\neg A \mid \neg B) &= \frac{P(\neg A \cap \neg B)}{P(\neg B)} \\
&= \frac{P(\neg(A \cup B))}{P(\neg B)} \\
&= \frac{1 - P(A \cup B)}{P(\neg B)} \\
&= \frac{1 - [P(A) + P(B) - P(A \cap B)]}{P(\neg B)} \\
&= \frac{1 - P(B)}{P(\neg B)} \\
&= \frac{P(\neg B)}{P(\neg B)} \\
&= 1
\end{aligned}
$$

b. Mostre que, se "$=$" for substituído por "$\approx$", o resultado não vale. Ache um exemplo em que $P(B \mid A)$ seja quase 1, mas $P(\neg A \mid \neg B)$ seja quase 0.

- Valores de exemplo, com A e B independentes:

$$
\begin{aligned}
P(A) &= 80/100 \\
P(B) &= 90/100 \\
P(\neg A) &= 20/100 \\
P(\neg B) &= 10/100 \\
P(A \cap B) &= P(A) \cdot P(B) = 72/100 \\
P(A \cup B) &= P(A) + P(B) - P(A \cap B) = 98/100 \\
P(B \mid A) &= \frac{P(A \cap B)}{P(A)} = 72/80 \approx 1 \\
P(\neg A \mid \neg B) &= \frac{P(\neg A \cap \neg B)}{P(\neg B)} = \frac{1 - P(A \cup B)}{P(\neg B)} = 2/10 \approx 0
\end{aligned}
$$

Vamos criar uma medida de independência. Se $P(A) \neq 0$ e $P(B) \neq 0$, definimos

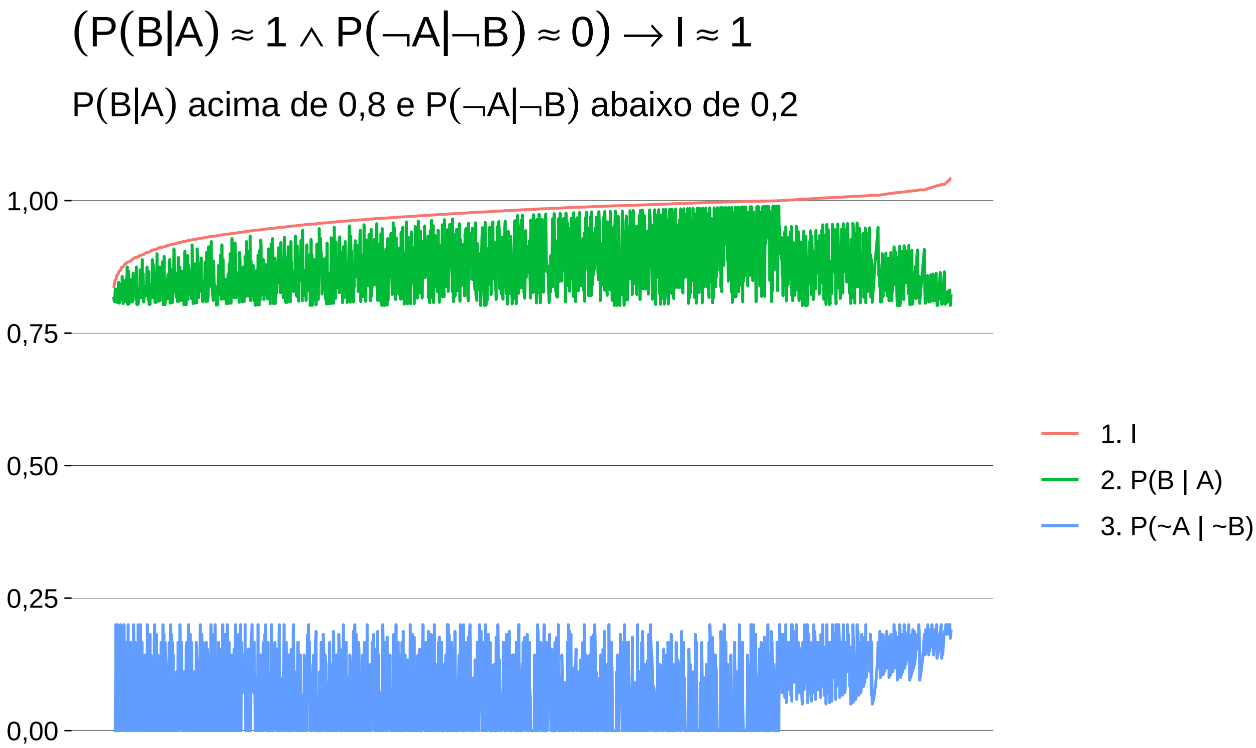
$$I = \frac{P(A \cap B)}{P(A)P(B)}$$

Com isso,

- A e B são disjuntos $\iff I = 0$
- A e B são independentes $\iff I = 1$
- A e B se atrapalham $\iff 0 < I < 1$.
  I.e., $P(A) > P(A \mid B)$ e $P(B) > P(B \mid A)$.
- A e B se ajudam $\iff I > 1$.
  I.e., $P(A) < P(A \mid B)$ e $P(B) < P(B \mid A)$.

Valores de I para P(A), P(B), P(A, B) ∈ {0,01, … , 0,99}

(grid 100 x 100)



22

> Este problema foi proposto pela primeira vez por Lewis Carroll em 1893.
>
> - Uma bolsa contém uma bola, que é ou azul, ou verde, com probabilidades iguais.
> - Uma bola verde é colocada na bolsa; agora, há 2 bolas na bolsa.
> - Uma bola é retirada da bolsa ao acaso.
> - A bola retirada é verde.
> - Qual é a probabilidade de que a bola que sobrou na bolsa seja verde?

- Antes de mais nada, vamos definir os eventos:

$$\begin{aligned} O &= \text{bola original é verde} \\ R &= \text{bola retirada é verde} \\ S &= \text{bola que sobrou é verde} \end{aligned}$$

- O importante é perceber que o enunciado diz que o evento R aconteceu, mas as probabilidades devem ser calculadas pensando em todos os resultados possíveis, antes de o experimento acontecer.

- Ou seja, em vez de tomar $P(R) = 1$ — o que seria errado — vamos calcular $P(S \mid R)$: a probabilidade de que a bola que sobrou seja verde, sabendo que a bola retirada foi verde.
- Começamos com a lei da probabilidade total, condicionando aos dois casos possíveis:

$$P(S \mid R) = \underbrace{P(S \mid R, O) \cdot P(O \mid R)}_{\text{caso 1: bola original verde}} + \underbrace{P(S \mid R, \neg O) \cdot P(\neg O \mid R)}_{\text{caso 2: bola original azul}}$$

- No caso 1:

$$\begin{aligned} P(S \mid R, O) \cdot P(O \mid R) &= 1 \cdot P(O \mid R) \\ &= P(O \mid R) \end{aligned}$$

- No caso 2:

$$\begin{aligned} P(S \mid R, \neg O) \cdot P(\neg O \mid R) &= 0 \cdot P(\neg O \mid R) \\ &= 0 \end{aligned}$$

  Isto faz sentido: se a bola original era azul, não há como a bola que sobrou ser verde.

- Só precisamos calcular a probabilidade do caso 1, que é $P(O \mid R)$. Vamos usar Bayes:

$$\begin{aligned} P(O \mid R) &= \frac{P(R \mid O) \cdot P(O)}{P(R)} \\ &= \frac{1 \cdot 1/2}{P(R)} \end{aligned}$$

- Para calcular $P(R)$, lei da probabilidade total de novo, condicionando sobre a bola original:

$$\begin{aligned} P(R) &= P(R \mid O) \cdot P(O) + P(R \mid \neg O) \cdot P(\neg O) \\ &= 1 \cdot 1/2 + 1/2 \cdot 1/2 \\ &= 3/4 \end{aligned}$$

- Chegamos a

$$
\begin{aligned}
P(O \mid R) &= \frac{P(R \mid O) \cdot P(O)}{P(R)} \\
&= \frac{1 \cdot 1/2}{P(R)} \\
&= \frac{1/2}{3/4} \\
&= \frac{2}{3}
\end{aligned}
$$

- Outra maneira de calcular $P(S \mid R)$ seria aplicar Bayes primeiro:

$$P(S \mid R) = \frac{P(R \mid S) \cdot P(S)}{P(R)}$$

  mas as probabilidades do numerador são mais difíceis de calcular! Teríamos que usar a lei da probabilidade total duas vezes para o numerador (além de uma vez para o denominador).

29

Uma família tem 2 crias.

Cada cria tem a mesma probabilidade de ser menino ou menina, e os sexos delas são independentes.

Cada cria tem a característica C com probabilidade p, independentemente uma da outra e do sexo.

Mostre que a probabilidade de serem duas meninas, dado que pelo menos uma das crias é uma menina com a característica C, é

$$\frac{2-p}{4-p}$$

Observe:

- Se $p = 1$, então a probabilidade é $1/3$, como no exemplo 2.2.5.
- Se $p \to 0$, então a probabilidade tende a $1/2$ pela esquerda, como no exemplo 2.2.7.

- Eventos:

$$
\begin{aligned}
AA &= \text{As duas são meninas} \\
AC &= \text{Pelo menos uma é menina e tem } C
\end{aligned}
$$

- Vamos usar Bayes:

$$P(AA \mid AC) = \frac{P(AC \mid AA) \cdot P(AA)}{P(AC)}$$

- A probabilidade de ambas serem meninas é

$$P(AA) = 1/4$$

- A probabilidade $P(AC)$ de pelo menos uma ser menina e ter C é a soma das probabilidades de
  - Ambas serem meninas, ambas terem C: $\frac{1}{4} \cdot p^2$.
  - Ambas serem meninas, só uma ter C (a primeira ou a segunda): $2 \cdot \frac{1}{4} \cdot p \cdot (1-p)$.
  - Uma ser menina com C, a outra ser menino (com ou sem C): $\frac{1}{2} \cdot p$.
  - Logo,
$$P(AC) = \frac{p \cdot (4-p)}{4}$$

- A probabilidade $P(AC \mid AA)$ de uma ser menina com C, dado que ambas são meninas, é

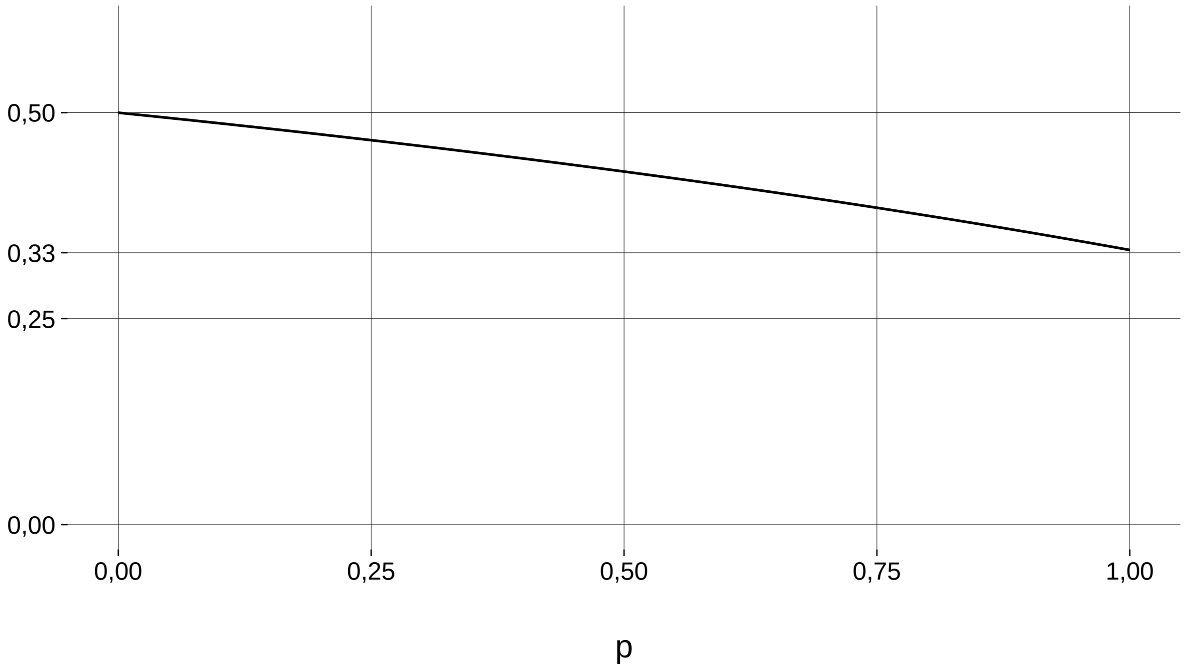
$$
\begin{aligned}
P(AC \mid AA) &= 1 - (1-p)^2 \\
&= p \cdot (2-p)
\end{aligned}
$$

- Juntando tudo:

$$
\begin{aligned}
P(AA \mid AC) &= \frac{P(AC \mid AA)P(AA)}{P(AC)} \\
&= \frac{p \cdot (2-p) \cdot 1/4}{p \cdot (4-p) \cdot 1/4} \\
&= \frac{2-p}{4-p}
\end{aligned}
$$

- Um gráfico:

Valor de $\frac{2-p}{4-p}$ em função de p



05: Probabilidade condicional (continuação)

Vídeo

https://youtu.be/JzDvVgNDxo8

Exercícios do livro (cap. 2)

30

Uma família tem 3 filhos: A, B e C.

a. O evento "A é mais velho que B" é independente de "A é mais velho que C"?

- Intuitivamente:

  Não, pois A ser mais velho que B torna mais provável que A seja mais velho que C.

b. Qual a probabilidade de que A é mais velho que B, dado que A é mais velho que C?

- Queremos achar $P(A > B \mid A > C)$.
- Esta probabilidade é igual a

$$\frac{P(A > B, A > C)}{P(A > C)}$$

- O numerador é a probabilidade de A ser o mais velho.
- Se todas as 6 ordens de nascimento tiverem a mesma probabilidade,

$$P(A > B, A > C) = 2/6 = 1/3$$

- O denominador é

$$P(A > C) = 3/6 = 1/2$$

- Daí, $P(A > B \mid A > C) = \frac{1/3}{1/2} = \frac{2}{3}$.
- De fato, a probabilidade condicional $P(A > B \mid A > C) = 2/3$ é maior do que a probabilidade não-condicional $P(A > B) = 1/2$.

31

Um evento pode ser independente de si mesmo?

- Chamando este evento de A, é preciso que

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

- Isto só é possível se $P(A) = 0$ ou se $P(A) = 1$.

---

32. Dados de Efron

Considere 4 dados não-padrão (*dados de Efron*), cujos lados são rotulados da seguinte forma (cada lado tem a mesma probabilidade):

$$\begin{aligned} A &: 4,4,4,4,0,0 \\ B &: 3,3,3,3,3,3 \\ C &: 6,6,2,2,2,2 \\ D &: 5,5,5,1,1,1 \end{aligned}$$

Cada dado é lançado uma vez. Cada letra representa o resultado do dado correspondente.

a. Ache $P(A > B)$, $P(B > C)$, $P(C > D)$, e $P(D > A)$.

- Eventos equivalentes:

$$\begin{aligned} A > B &\iff A = 4 \\ B > C &\iff C = 2 \\ C > D &\iff C = 6 \cup (C = 2 \cap D = 1) \\ D > A &\iff D = 5 \cup (D = 1 \cap A = 0) \end{aligned}$$

- $P(A > B) = 2/3.$
- $P(B > C) = 2/3.$
- $P(C > D) = 1/3 + 2/3 \cdot 1/2 = 2/3.$
- $P(D > A) = 1/2 + 1/2 \cdot 2/3 = 2/3.$

b. O evento $A > B$ é independente de $B > C$?
O evento $B > C$ é independente de $C > D$?

- Sim:

$$\begin{aligned} P(A > B \cap B > C) &= P(A = 4) \cdot P(C = 2) \\ &= P(A > B) \cdot P(B > C) \end{aligned}$$

Intuitivamente: como o resultado de B não importa para $A > B$ nem para $B > C$, os eventos são independentes.

- Não:

$$
\begin{aligned}
P(B > C \cap C > D) &= P(C = 2 \cap [C = 6 \cup (C = 2 \cap D = 1)]) \\
&= P((C = 2 \cap C = 6) \cup (C = 2 \cap D = 1)) \\
&= P(C = 2 \cap D = 1) \\
&= 2/3 \cdot 1/2 \\
&= 1/3
\end{aligned}
$$

mas

$$
P(B > C) \cdot P(C > D) = 4/9
$$

33

- Alice, Bob, e mais 100 pessoas vivem em uma cidade.
- C é o conjunto das outras 100 pessoas.
- $A \subseteq C$ é o conjunto de amigos de Alice.
- $B \subseteq C$ é o conjunto de amigos de Bob.
- Para cada pessoa em C, a probabilidade de Alice ser amiga da pessoa é $1/2$.
- Idem para Bob.
- As amizades são independentes.

a. Seja $D \subseteq C$. Achar $P(A = D)$.

- Para cada $x \in C$:
  1. $x \in A \land x \in D$, com probabilidade $\frac{1}{2} \cdot \frac{|D|}{|C|}$,

     ou (exclusivo)
  2. $x \notin A \land x \notin D$, com probabilidade $\frac{1}{2} \cdot \left(1 - \frac{|D|}{|C|}\right)$
- Somando as probabilidades dos casos, para cada $x \in C$, a probabilidade de x estar em A e em D, ou de x não estar nem em A, nem em D, é

$$
\frac{1}{2} \cdot \frac{|D|}{|C|} + \frac{1}{2} \cdot \left(1 - \frac{|D|}{|C|}\right) = \frac{1}{2}
$$

- Convenientemente, $|D|$ desapareceu, mas só porque $p = 1/2$.

- A probabilidade de $A = D$ é a probabilidade de, para todo $x \in C$, acontecer de $x \in A \wedge x \in D$ ou $x \in A \wedge x \in D$. Como os eventos são independentes, temos

$$P(A = D) = \frac{1}{2^{|C|}}$$

- Vamos simular a situação. Estamos supondo que os elementos de D são escolhidos segundo uma amostragem uniforme. Veja explicações mais detalhadas abaixo.

```
simular <- function(p, n) {

  A <- (1:n)[runif(n) <= p]
  D <- (1:n)[runif(n) <= 1/2]

  if (length(A) != length(D)) {
    FALSE
  } else {
    all(A == D)
  }

}
```

- Para as probabilidades não ficarem tão pequenas, vamos usar um universo C com 10 elementos apenas:

```
p <- 1/2
n <- 10
nsims <- 1e6

resultado <- mean(
  1:nsims %>%
    map_lgl(~simular(p, n))
)

cat('P(A = D) simulado = ', resultado)
```

```
## P(A = D) simulado =  0,000991
```

```
cat('P(A = D) teórico  = ', 1/2^n)
```

```
## P(A = D) teórico  =  0,0009765625
```

p **qualquer**

- Vamos resolver de outra maneira, pensando em p genérico.
- ??? Substituir $\frac{|D|}{n}$ por $\frac{1}{2}$ na fórmula.
- Simular com outros valores de p.

Extra: amostragem uniforme (sem reposição)

- Na resolução acima, usamos a igualdade

$$P(x \in D) = \frac{|D|}{n}$$

  para representar a probabilidade de que um elemento qualquer de um universo com n elementos pertença a um conjunto D fixo.
- Mas e quando não sabemos qual é o conjunto D, nem qual o seu tamanho?
- Neste caso, precisamos supor algo sobre a distribuição dos subconjuntos.
- O mais comum é supor que cada um dos 2^n subconjuntos tem a mesma probabilidade de ser o resultado da amostragem.
- Com esta suposição, como calculamos $P(x \in D)$?
- Vamos condicionar ao tamanho do conjunto D e marginalizar:

$$P(x \in D) = \sum_{k=0}^{n} P(x \in D \mid |D| = k) \cdot P(|D| = k)$$

- Calculando a primeira probabilidade dentro do somatório:

$$\begin{aligned} P(x \in D \mid |D| = k) &= \frac{\binom{n-1}{k-1}}{\binom{n}{k}} \\ &= \frac{k}{n} \end{aligned}$$

  Na primeira linha, o numerador é a quantidade de subconjuntos de k elementos que contêm x (basta escolher os outros $k-1$ elementos dentre os $n-1$ outros elementos do universo); o denominador é o total de subconjuntos de k elementos.

  Perceba que aqui usamos a suposição de que todos os subconjuntos têm a mesma probabilidade de ser amostrados.

  Perceba também que, como antes, esta probabilidade é igual a $\frac{|D|}{n}$.

- Calculando a segunda probabilidade dentro do somatório:

$$P(|D| = k) = \frac{\binom{n}{k}}{2^n}$$

  De novo, usando a suposição de que todos os subconjuntos são equiprováveis, temos a quantidade de subconjuntos de k elementos sobre o total de subconjuntos.

- Juntando tudo:

$$\begin{aligned}
P(x \in D) &= \sum_{k=0}^{n} P(x \in D \mid |D| = k) \cdot P(|D| = k) \\
&= \sum_{k=0}^{n} \frac{1}{2^n} \binom{n}{k} \frac{k}{n} \\
&= \frac{1}{2^n} \sum_{k=0}^{n} \binom{n-1}{k-1} \\
&= \frac{1}{2^n} \cdot 2^{n-1} \\
&= \frac{1}{2}
\end{aligned}$$

> Fazer amostragem uniforme significa que todos os subconjuntos são equiprováveis, ou, equivalentemente, que cada elemento do universo tem 50% de chance de estar na amostra.

- Por isso, na simulação, geramos o conjunto D com o comando

```
D <- (1:n)[runif(n) <= 1/2]
```

35. Xadrez

- Você vai jogar 2 partidas de xadrez contra um adversário desconhecido.
- O nível do seu adversário pode ser novato, intermediário, ou avançado, com probabilidades iguais.
- As probabilidades de você vencer uma partida são, dependendo do nível do adversário, respectivamente, 90%, 50%, e 30%.

a. Qual a probabilidade de você vencer a primeira partida?

b. Parabéns, você venceu a primeira partida. Dada esta informação, qual a probabilidade de que você também vença a segunda partida (suponha que, dado o nível do seu adversário, os resultados das partidas são independentes)?

c. Explique a diferença entre

   1. supor que os resultados das partidas são independentes, e
   2. supor que os resultados das partidas são independentes, dado o nível do seu adversário.

   Qual destas suposições parece mais razoável? Por quê?

- Antes de mais nada, vamos definir os eventos:

$$
\begin{aligned}
N &= \text{o adversário é novato} \\
I &= \text{o adversário é intermediário} \\
A &= \text{o adversário é avançado} \\
V_1 &= \text{você vence a primeira partida} \\
V_2 &= \text{você vence a segunda partida}
\end{aligned}
$$

- O enunciado dá as probabilidades

$$
\begin{aligned}
P(N) &= 1/3 \\
P(I) &= 1/3 \\
P(A) &= 1/3 \\
P(V_1 \mid N) &= 9/10 \\
P(V_1 \mid I) &= 5/10 \\
P(V_1 \mid A) &= 3/10
\end{aligned}
$$

- Para resolver (a), basta usar probabilidade total, condicionando sobre o nível do adversário:

$$
\begin{aligned}
P(V_1) &= P(V_1 \mid N) \cdot P(N) \;+\; P(V_1 \mid I) \cdot P(I) \;+\; P(V_1 \mid A) \cdot P(A) \\
&= \frac{9}{10} \cdot \frac{1}{3} + \frac{5}{10} \cdot \frac{1}{3} + \frac{3}{10} \cdot \frac{1}{3} \\
&= \frac{1}{3} \cdot \left(\frac{9}{10} + \frac{5}{10} + \frac{3}{10} \right) \\
&= \frac{17}{30}
\end{aligned}
$$

Faz sentido. Como as probabilidades dos níveis possíveis do adversário são iguais, a probabilidade de vencer é a média aritmética das probabilidades de vencer contra cada nível.

Se as probabilidades dos níveis do adversário fossem diferentes, seria a média ponderada.

- Para (b), queremos calcular $P(V_2 \mid V_1)$.

  Dizer que V_1 e V_2 são independentes dado o nível do adversário é dizer

$$
P(V_2 \mid V_1, N) = P(V_1 \mid V_2, N) = P(V_2 \mid N) = P(V_1 \mid N)
$$

  e analogamente para probabilidades condicionadas a I e a A.

  Ou seja, dado um nível específico do adversário, saber que V_1 ocorreu não altera a probabilidade de V_2 ocorrer, e vice-versa.

  Vamos calcular $P(V_2 \mid V_1)$ usando probabilidade total, condicionando ao nível:

$$
\underbrace{P(V_2 \mid V_1, N) \cdot P(N \mid V_1)}_{\text{novato}} \;+\; \underbrace{P(V_2 \mid V_1, I) \cdot P(I \mid V_1)}_{\text{intermediário}} \;+\; \underbrace{P(V_2 \mid V_1, A) \cdot P(A \mid V_1)}_{\text{avançado}}
$$

  Para o lado esquerdo de cada produto, a independência condicional diz que

$$
\begin{aligned}
P(V_2 \mid V_1, N) &= P(V_1 \mid N) = 9/10 \\
P(V_2 \mid V_1, I) &= P(V_1 \mid I) = 5/10 \\
P(V_2 \mid V_1, A) &= P(V_1 \mid A) = 3/10
\end{aligned}
$$

  Para o lado direito de cada produto, usamos Bayes. Todas as probabilidades envolvidas já foram calculadas.

  Novato:

$$
\begin{aligned}
P(N \mid V_1) &= \frac{P(V_1 \mid N) \cdot P(N)}{P(V_1)} \\
&= \frac{9/10 \cdot 1/3}{17/30} \\
&= \frac{9}{17}
\end{aligned}
$$

Intermediário:

$$
\begin{aligned}
P(I \mid V_1) &= \frac{P(V_1 \mid I) \cdot P(I)}{P(V_1)} \\
&= \frac{5/10 \cdot 1/3}{17/30} \\
&= \frac{5}{17}
\end{aligned}
$$

Avançado:

$$
\begin{aligned}
P(A \mid V_1) &= \frac{P(V_1 \mid A) \cdot P(A)}{P(V_1)} \\
&= \frac{3/10 \cdot 1/3}{17/30} \\
&= \frac{3}{17}
\end{aligned}
$$

A resposta final é

$$
\frac{9}{10} \cdot \frac{9}{17} + \frac{5}{10} \cdot \frac{5}{17} + \frac{3}{10} \cdot \frac{3}{17} = \frac{23}{34}
$$

- Para responder (c):

  Dizer que V_1 e V_2 são incondicionalmente independentes seria dizer que

$$
P(V_2 \mid V_1) = P(V_2)
$$

Ainda mais, como o nível do adversário não muda de uma partida para outra, teríamos também

$$
P(V_2) = P(V_1)
$$

Ou seja, cada partida seria uma prova de Bernoulli com a mesma probabilidade de sucesso.

Considerando independência condicional, saber que vencemos a primeira partida nos dá informação sobre o nível do adversário, e esta informação é considerada para calcular a probabilidade de vencer a segunda partida.

De fato, usando independência condicional, temos

$$P(V_1) \approx 0{,}57$$

e

$$P(V_2 \mid V_1) \approx 0{,}68$$

---

36. Paradoxo de Berkson

Se

- A e B são independentes, e
- $P(A \cap B) > 0$, e
- $P(A \cup B) < 1$, e
- $C = A \cup B$

então A e B são condicionalmente dependentes, dado C, com

$$P(A \mid B, C) < P(A \mid C)$$

Exemplo: universidade admite candidatos que são bons jogadores de basquete (A) ou que são bons em Matemática (B) — supondo que estes são eventos independentes, o que é meio duvidoso.[a]

[a]Na verdade, em vez de independência entre A e B, basta que um evento atrapalhe o outro para termos o paradoxo. Em termos da nossa medida de independência I, basta $0 < I < 1$.

- De $P(A \cap B) > 0$, temos que $P(A) > 0$ e que $P(B) > 0$.
- Daí, $P(C) = P(A \cup B) > 0$.
- Para o lado esquerdo:

$$
\begin{aligned}
P(A \mid B, C) &= \frac{P(A \cap B \cap C)}{P(B \cap C)} \\
&= \frac{P(A \cap B)}{P(B)} \\
&= \frac{P(A)P(B)}{P(B)} \\
&= P(A)
\end{aligned}
$$

- Para o lado direito:

$$
\begin{aligned}
P(A \mid C) &= \frac{P(A \cap C)}{P(C)} \\
&= \frac{P(A)}{P(C)} \\
&= \frac{P(A)}{P(A \cup B)}
\end{aligned}
$$

- Como o denominador $P(A \cup B) < 1$, esta probabilidade é maior do que $P(A)$.

37

- Quem tem a doença D_1 ou a doença D_2 (ou ambas) tem o sintoma estranho W.
- D_1 e D_2 são independentes, com $P(D_j) = p_j$, e com $q_j = 1 - p_j$.
- $0 < p_j < 1$.
- Uma pessoa sadia tem o sintoma W com probabilidade w_0.

a. Achar $P(W)$.

- Por probabilidade total:

$$
\begin{aligned}
P(W) &= P(W \mid D_1 \cup D_2) \cdot P(D_1 \cup D_2) + P(W \mid \neg(D_1 \cup D_2)) \cdot P(\neg(D_1 \cup D_2)) \\
&= 1 \cdot (p_1 + p_2 - p_1 p_2) + w_0 \cdot (1 - p_1 - p_2 + p_1 p_2) \\
&= p_1 + p_2 - p_1 p_2 + w_0 q_1 q_2
\end{aligned}
$$

b. Achar $P(D_1 \mid W)$, $P(D_2 \mid W)$, e $P(D_1 \cap D_2 \mid W)$.

- Por Bayes:

$$
\begin{aligned}
P(D_1 \mid W) &= \frac{P(W \mid D_1) \cdot P(D_1)}{P(W)} \\
&= \frac{1 \cdot p_1}{p_1 + p_2 - p_1 p_2 + w_0 q_1 q_2} \\
&= \frac{p_1}{p_1 + p_2 - p_1 p_2 + w_0 q_1 q_2}
\end{aligned}
$$

- Analogamente,

$$
P(D_2 \mid W) = \frac{p_2}{p_1 + p_2 - p_1 p_2 + w_0 q_1 q_2}
$$

- Finalmente,

$$
\begin{aligned}
P(D_1 \cap D_2 \mid W) &= \frac{P(D_1 \cap D_2 \cap W)}{P(W)} \\
&= \frac{P(D_1 \cap D_2)}{P(W)} && (\text{pois } D_1 \cap D_2 \subseteq W) \\
&= \frac{p_1 p_2}{p_1 + p_2 - p_1 p_2 + w_0 q_1 q_2}
\end{aligned}
$$

c. D_1 e D_2 são condicionalmente independentes, dado W?

- Basta verificar se

$$
P(D_1 \mid W) \cdot P(D_2 \mid W) = P(D_1 \cap D_2 \mid W)
$$

- Isto equivale a

$$
\frac{p_1 p_2}{(p_1 + p_2 - p_1 p_2 + w_0 q_1 q_2)^2} = \frac{p_1 p_2}{p_1 + p_2 - p_1 p_2 + w_0 q_1 q_2}
$$

- Esta igualdade só é verdade se

  1. $w_0 = 0$ e $P(D_1 \cup D_2) = 0$ (o que é proibido pelo enunciado). Ninguém estaria doente, e ninguém teria o sintoma.

     ou

2. $P(D_1 \cup D_2) = 1$. Aqui, todos estariam doentes, e todos teriam o sintoma.

- Lembramos que

$$w_0 = P(W \mid \neg(D_1 \cup D_2)) = \frac{P(W \cap \neg(D_1 \cup D_2))}{P(\neg(D_1 \cup D_2))}$$

- No caso (1) acima, $w_0 = P(W) = 0$, e não faz mais sentido condicionar sobre W.
- No caso (2) acima, $P(\neg(D_1 \cup D_2)) = 0$, e a própría definição de w_0 deixa de fazer sentido.

d. Suponha que $w_0 = 0$. Neste caso, D_1 e D_2 são condicionalmente independentes, dado W?

- Por causa do discutido no item (c), vamos supor que

$$0 < P(D_1 \cup D_2) < 1$$

para que possamos condicionar sobre W e para que a definição de w_0 faça sentido.

- Como no item (c), verificamos se

$$P(D_1 \mid W) \cdot P(D_2 \mid W) = P(D_1 \cap D_2 \mid W)$$

- Que equivale a

$$\frac{p_1 p_2}{(p_1 + p_2 - p_1 p_2)^2} = \frac{p_1 p_2}{p_1 + p_2 - p_1 p_2}$$

- Isto é impossível, se $0 < P(D_1 \cup D_2) < 1$.
- Intuitivamente, D_1 e D_2 seriam condicionalmente dependentes, dado W, pois, dado que uma pessoa tem o sintoma W, saber que ela não tem a doença D_1 imediatamente nos diz que ela tem a doença D_2.

Quando $w_0 = 0$ e $0 < P(D_1 \cup D_2) < 1$, a consequência é que

$$D_1 \cup D_2 = W$$

como eventos.

Daí, com as outras condições do enunciado, temos uma instância do paradoxo de Berkson, com $A = D_1$ e $B = D_2$.

38. Naïve Bayes

- Temos uma lista de 100 palavras.
- Evento $W_j =$ a palavra j aparece no *email*.
- Evento spam $=$ o *email* é *spam*.
- $p = P(\text{spam})$.
- $p_j = P(W_j \mid \text{spam})$.
- $r_j = P(W_j \mid \neg\text{spam})$.
- A *naïveté* do algoritmo consiste nas seguintes suposições — não-realistas, mas úteis:
  - Os W_j são condicionalmente independentes, dado spam.
  - Os W_j são condicionalmente independentes, dado ¬spam.
- Um novo *email* é recebido. Contém as palavras 23, 64, e 65, e nenhuma das outras.
- Vamos chamar de $\vec{W}$ a lista de eventos

$$\begin{array}{l} \neg W_1, \ldots, \neg W_{22}, \\ W_{23}, \\ \neg W_{24}, \ldots, \neg W_{63}, \\ W_{64}, W_{65}, \\ \neg W_{66}, \ldots, \neg W_{100} \end{array}$$

- Calcular $P\left(\text{spam} \,\middle|\, \vec{W}\right)$.

- Por Bayes:

$$P\left(\text{spam} \,\middle|\, \vec{W}\right) = \frac{P\left(\vec{W} \,\middle|\, \text{spam}\right)}{P\left(\vec{W}\right)}$$

- Com a suposição *naïve*, podemos multiplicar as probabilidades condicionais:

$$P\left(\vec{W} \,\middle|\, \text{spam}\right) = p_{23} \cdot p_{64} \cdot p_{65} \cdot \prod_{\substack{1 \le j \le 100. \\ j \notin \{23, 64, 65\}}} (1 - p_j)$$

Vamos chamar este valor de m.

- Para calcular $P\left(\vec{W}\right)$, usamos probabilidade total:

$$\begin{aligned} P\left(\vec{W}\right) &= P\left(\vec{W} \,\middle|\, \text{spam}\right) \cdot P(\text{spam}) \;+\; P\left(\vec{W} \,\middle|\, \neg\text{spam}\right) \cdot P(\neg\text{spam}) \\ &= m \cdot p \;+\; m' \cdot (1-p) \end{aligned}$$

onde, de maneira análoga ao cálculo de m,

$$m' = r_{23} \cdot r_{64} \cdot r_{65} \cdot \prod_{\substack{1 \leq j \leq 100. \\ j \notin \{23, 64, 65\}}} (1 - r_j)$$

- O valor procurado é, então,

$$P\left(\text{spam} \,\middle|\, \vec{W}\right) = \frac{P\left(\vec{W} \,\middle|\, \text{spam}\right)}{P\left(\vec{W}\right)} = \frac{mp}{mp + m'(1-p)}$$

Referências

DEVLIN, K. The Pascal-Fermat Correspondence: How Mathematics Is Really Done. **The Mathematics Teacher**, v. 103, n. 8, p. 579–582, abr. 2010.

OLIVEIRA MORGADO, A. C. DE et al. **Análise combinatória e probabilidade.** Rio de Janeiro: Impa / Vitae, 2004.

PAULOS, J. A. **Innumeracy: Mathematical Illiteracy and Its Consequences**. 1. ed. New York: Hill; Wang, 1988.