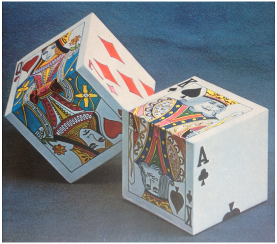
Fundamentos Estatísticos para Ciência dos Dados
Introdução à Probabilidade

Renato Martins Assunção

DCC, UFMG



Aleatoriedade

- Vamos lidar com fenômenos não determinísticos, probabilísticos, aleatórios.

Aleatoriedade

- Vamos lidar com fenômenos não determinísticos, probabilísticos, aleatórios.
- O modelo matemático para **qualquer** fenômeno probabilístico é o espaço de probabilidade.

Aleatoriedade

- Vamos lidar com fenômenos não determinísticos, probabilísticos, aleatórios.
- O modelo matemático para **qualquer** fenômeno probabilístico é o espaço de probabilidade.
- Espaço de probabilidade é uma 3-upla constituída por três elementos satisfazendo os três axiomas de Kolmogorov (1903-1987).

Espaço de probabilidade

(a) Um espaço amostral Ω
Ω é um conjunto com **todos** os resultados possíveis do fenômeno.

Espaço de probabilidade

(a) Um espaço amostral Ω
Ω é um conjunto com **todos** os resultados possíveis do fenômeno.

(b) Uma σ-álgebra $\mathcal{A}$ de sub-conjuntos de Ω, os sub-conjuntos aos quais vamos atribuir probabilidades.

Espaço de probabilidade

(a) Um espaço amostral Ω
Ω é um conjunto com **todos** os resultados possíveis do fenômeno.

(b) Uma σ-álgebra $\mathcal{A}$ de sub-conjuntos de Ω, os sub-conjuntos aos quais vamos atribuir probabilidades.

(c) Uma função matemática atribuindo probabilidades aos sub-conjuntos de $\mathcal{A}$

$$\begin{aligned} \mathbb{P} : \mathcal{A} &\longrightarrow [0,1] \\ A &\longrightarrow \mathbb{P}(A) \end{aligned}$$

Espaço de probabilidade

(a) Um espaço amostral Ω
Ω é um conjunto com **todos** os resultados possíveis do fenômeno.

(b) Uma σ-álgebra $\mathcal{A}$ de sub-conjuntos de Ω, os sub-conjuntos aos quais vamos atribuir probabilidades.

(c) Uma função matemática atribuindo probabilidades aos sub-conjuntos de $\mathcal{A}$

$$\begin{aligned} \mathbb{P} : \mathcal{A} &\longrightarrow [0, 1] \\ A &\longrightarrow \mathbb{P}(A) \end{aligned}$$

(d) Esta função $\mathbb{P}$ deve satisfazer os três axiomas de Kolmogorov

Vamos ver cada um desses elementos com mais detalhes.

O espaço amostral Ω

- $\Omega =$ conjunto representando **todos** os resultados possíveis do fenômeno.
- Em IA, falamos de todos os possíveis "estados do mundo".

O espaço amostral Ω

- $\Omega =$ conjunto representando **todos** os resultados possíveis do fenômeno.
- Em IA, falamos de todos os possíveis "estados do mundo".
- Cada resultado possível deve ser completamente especificado e único em Ω (não pode haver dois elementos em Ω representando o mesmo resultado possível).

O espaço amostral Ω

- $\Omega =$ conjunto representando **todos** os resultados possíveis do fenômeno.
- Em IA, falamos de todos os possíveis "estados do mundo".
- Cada resultado possível deve ser completamente especificado e único em Ω (não pode haver dois elementos em Ω representando o mesmo resultado possível).
- A todo estado do mundo corresponde um, e somente um, elemento $\omega \in \Omega$.

O espaço amostral Ω

- $\Omega =$ conjunto representando **todos** os resultados possíveis do fenômeno.
- Em IA, falamos de todos os possíveis "estados do mundo".
- Cada resultado possível deve ser completamente especificado e único em Ω (não pode haver dois elementos em Ω representando o mesmo resultado possível).
- A todo estado do mundo corresponde um, e somente um, elemento $\omega \in \Omega$.
- Ω pode ter mais elementos que estados do mundo (pode ter elementos que representam resultados IMpossíveis.

O espaço amostral Ω

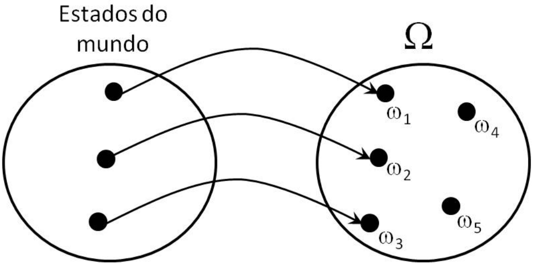


Figura: O espaço amostral Ω

Exemplos de Ω

- Observa-se o lançamento de uma moeda
- $\Omega = \{\text{cara, coroa}\}$

Exemplos de Ω

- Observa-se o lançamento de uma moeda
- $\Omega = \{\text{cara, coroa}\}$
- ou $\Omega = \{\text{c}, \tilde{\text{c}}\}$
- ou $\Omega = \{0, \quad 1\}$
- ou $\Omega = \{T, \quad F\}$
- Precisamos de um conjunto com pelo menos dois elementos.

Exemplos de Ω: moeda

- Poderíamos também definir

$$\begin{cases} \Omega = \{\text{c}, \tilde{\text{c}}, \text{ reis de paus}\} \\ \text{ou} \\ \Omega = \{\text{c}, \tilde{\text{c}}, \text{ céu azul, cárie}\} \end{cases}$$

Exemplos de Ω: moeda

- Poderíamos também definir

$$\begin{cases} \Omega = \{\text{c}, \tilde{\text{c}}, \text{ reis de paus}\} \\ \text{ou} \\ \Omega = \{\text{c}, \tilde{\text{c}}, \text{ céu azul, cárie}\} \end{cases}$$

- *Reis de paus*, *céu azul*, e *cárie* representam resultados impossíveis no mundo que se limita apenas a observar o resultado de lançar uma moeda.

Exemplos de Ω: moeda

- Poderíamos também definir

$$\begin{cases} \Omega = \{\text{c}, \tilde{\text{c}}, \text{ reis de paus}\} \\ \text{ou} \\ \Omega = \{\text{c}, \tilde{\text{c}}, \text{ céu azul, cárie}\} \end{cases}$$

- *Reis de paus*, *céu azul*, e *cárie* representam resultados impossíveis no mundo que se limita apenas a observar o resultado de lançar uma moeda.
- Podemos "corrigir" este excesso de elementos em Ω atribuindo probabilidade zero aos elementos em "excesso".

Mais moedas

- Observa-se três lançamentos sucessivos de uma moeda

Mais moedas

- Observa-se três lançamentos sucessivos de uma moeda

$$\Omega = \{ccc, cc\tilde{c}, c\tilde{c}c, ..., \tilde{c}\tilde{c}\tilde{c}\}$$

- Ω tem 8 elementos

Mais moedas

- Observa-se três lançamentos sucessivos de uma moeda

$$\Omega = \{ccc, cc\tilde{c}, c\tilde{c}c, ..., \tilde{c}\tilde{c}\tilde{c}\}$$

- Ω tem 8 elementos
- O mundo deste segundo exemplo é mais amplo que aquele do primeiro observador-exemplo.

Mais moedas

- Observa-se três lançamentos sucessivos de uma moeda

$$\Omega = \{ccc, cc\tilde{c}, c\tilde{c}c, ..., \tilde{c}\tilde{c}\tilde{c}\}$$

- Ω tem 8 elementos
- O mundo deste segundo exemplo é mais amplo que aquele do primeiro observador-exemplo.
- Neste mundo podemos calcular a probabilidade do segundo lançamento da moeda ser *cara*.

Mais moedas

- Observa-se três lançamentos sucessivos de uma moeda

$$\Omega = \{ccc, cc\tilde{c}, c\tilde{c}c, ..., \tilde{c}\tilde{c}\tilde{c}\}$$

- Ω tem 8 elementos
- O mundo deste segundo exemplo é mais amplo que aquele do primeiro observador-exemplo.
- Neste mundo podemos calcular a probabilidade do segundo lançamento da moeda ser *cara*.
- No mundo do primeiro observador não podemos calcular a probabilidades referentes ao segundo ou terceiro lançamentos da moeda pois eles não pertencem ao Ω daquele mundo.

Espaço de imagens

- Imagem com 512×512 pixels, cada pixel tem um tom de cinza.
- Tom de cinza de cada pixel é codificado com um inteiro entre 0 e 255.
- 8 bits, $2^8 = 256$ tons possíveis: 0 é preto e 255 é branco.

Espaço de imagens

- Imagem com 512×512 pixels, cada pixel tem um tom de cinza.
- Tom de cinza de cada pixel é codificado com um inteiro entre 0 e 255.
- 8 bits, $2^8 = 256$ tons possíveis: 0 é preto e 255 é branco.
- Ω é o conjunto:
    - de todas as matrizes M,

Espaço de imagens

- Imagem com 512×512 pixels, cada pixel tem um tom de cinza.
- Tom de cinza de cada pixel é codificado com um inteiro entre 0 e 255.
- 8 bits, $2^8 = 256$ tons possíveis: 0 é preto e 255 é branco.
- Ω é o conjunto:
  - de todas as matrizes M,
  - de dimensão 512×512

Espaço de imagens

- Imagem com 512×512 pixels, cada pixel tem um tom de cinza.
- Tom de cinza de cada pixel é codificado com um inteiro entre 0 e 255.
- 8 bits, $2^8 = 256$ tons possíveis: 0 é preto e 255 é branco.
- Ω é o conjunto:
  - de todas as matrizes M,
  - de dimensão 512×512
  - e com $M(i,j) \in \{0, 1, \ldots, 255\}$.

Espaço de imagens

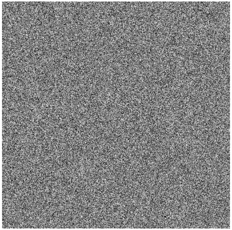
- Imagem com 512×512 pixels, cada pixel tem um tom de cinza.
- Tom de cinza de cada pixel é codificado com um inteiro entre 0 e 255.
- 8 bits, $2^8 = 256$ tons possíveis: 0 é preto e 255 é branco.
- Ω é o conjunto:
  - de todas as matrizes M,
  - de dimensão 512×512
  - e com $M(i,j) \in \{0, 1, \ldots, 255\}$.
- Ω é um conjunto finito com 256^{512^2} elementos.

Espaço de imagens

- Dois elementos de Ω, duas imagens 512×512 em tons de cinza.

Espaço de imagens

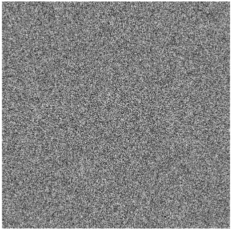
- Dois elementos de Ω, duas imagens 512×512 em tons de cinza.



- A imagem da esquerda é uma imagem "estruturada"
- A da direita é uma imagem em que cada pixel é um número aleatório entre 0 e 255: ruído puro.

Espaço de imagens

- Dois elementos de Ω, duas imagens 512×512 em tons de cinza.



- A imagem da esquerda é uma imagem "estruturada"
- A da direita é uma imagem em que cada pixel é um número aleatório entre 0 e 255: ruído puro.
- Modelos para imagens podem atribuir probabilidades maiores a imagens como a da esquerda.

Movimento Browniano

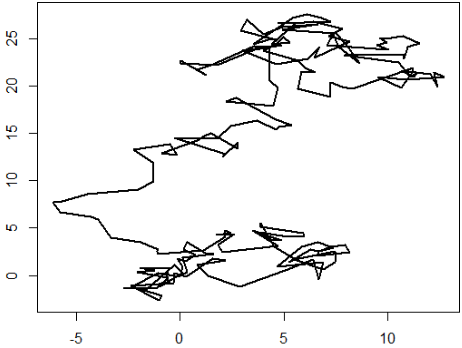


Figura: Movimento errático de um grão de pólen na superfície da água observado a cada 1 segundo.

Movimento Browniano

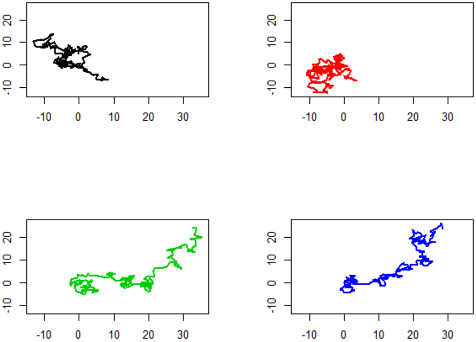


Figura: Quatro realizações com pólen partindo da origem $(0, 0)$ no tempo $t = 0$.

Movimento Browniano

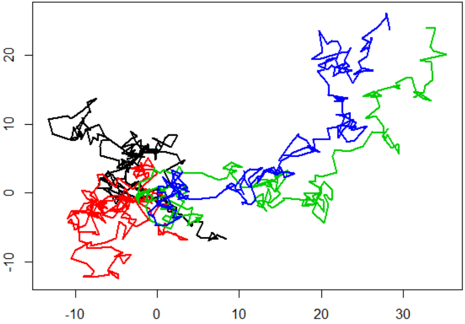


Figura: Quatro realizações com pólen partindo da origem $(0, 0)$ no tempo $t = 0$.

Movimento Browniano

- Este exemplo é importante para motivar a abstração e complicação matemática dos espaços amostrais.

Movimento Browniano

- Este exemplo é importante para motivar a abstração e complicação matemática dos espaços amostrais.
- Não veremos este exemplo no resto do curso...

Movimento Browniano

- Este exemplo é importante para motivar a abstração e complicação matemática dos espaços amostrais.
- Não veremos este exemplo no resto do curso...
- mas ele é um exemplo típico de processo estocástico, um assunto crucial em probabilidade mais avançada.

Movimento Browniano

- Este exemplo é importante para motivar a abstração e complicação matemática dos espaços amostrais.
- Não veremos este exemplo no resto do curso...
- mas ele é um exemplo típico de processo estocástico, um assunto crucial em probabilidade mais avançada.
- Tem importância histórico: Einstein publicou em 1905 um paper fundamental explicando o movimento browniano como efeito da movimentação atômica.

Movimento Browniano

- Este exemplo é importante para motivar a abstração e complicação matemática dos espaços amostrais.
- Não veremos este exemplo no resto do curso...
- mas ele é um exemplo típico de processo estocástico, um assunto crucial em probabilidade mais avançada.
- Tem importância histórico: Einstein publicou em 1905 um paper fundamental explicando o movimento browniano como efeito da movimentação atômica.
- $\Omega =$ conjunto de TODAS AS CURVAS no plano da forma (x_t, y_t) para $t = 1, 2, 3, \ldots$.

Movimento Browniano

- Este exemplo é importante para motivar a abstração e complicação matemática dos espaços amostrais.
- Não veremos este exemplo no resto do curso...
- mas ele é um exemplo típico de processo estocástico, um assunto crucial em probabilidade mais avançada.
- Tem importância histórico: Einstein publicou em 1905 um paper fundamental explicando o movimento browniano como efeito da movimentação atômica.
- $\Omega =$ conjunto de TODAS AS CURVAS no plano da forma (x_t, y_t) para $t = 1, 2, 3, \ldots$.
- Mostramos apenas 4 dessas curvas na figura anterior.

Movimento Browniano

- Este exemplo é importante para motivar a abstração e complicação matemática dos espaços amostrais.
- Não veremos este exemplo no resto do curso...
- mas ele é um exemplo típico de processo estocástico, um assunto crucial em probabilidade mais avançada.
- Tem importância histórico: Einstein publicou em 1905 um paper fundamental explicando o movimento browniano como efeito da movimentação atômica.
- $\Omega =$ conjunto de TODAS AS CURVAS no plano da forma (x_t, y_t) para $t = 1, 2, 3, \ldots$.
- Mostramos apenas 4 dessas curvas na figura anterior.
- Ω é um conjunto infinito.

$\{0,1\}^\infty$

- Joga-se uma moeda para cima *indefinidamente* (infinitas vezes).
- A probabilidade de sair cara num lançamento vai decaindo.

$\{0, 1\}^\infty$

- Joga-se uma moeda para cima *indefinidamente* (infinitas vezes).
- A probabilidade de sair cara num lançamento vai decaindo.
- No primeiro lançamento, a probbilidade de sair cara é p_1.
- No segundo, ela diminui para um valor $p_2 < p_1$.

$\{0,1\}^\infty$

- Joga-se uma moeda para cima *indefinidamente* (infinitas vezes).
- A probabilidade de sair cara num lançamento vai decaindo.
- No primeiro lançamento, a probbilidade de sair cara é p_1.
- No segundo, ela diminui para um valor $p_2 < p_1$.
- No terceiro fica menor ainda: $p_3 < p_2 < p_1$.

$\{0, 1\}^\infty$

- Joga-se uma moeda para cima *indefinidamente* (infinitas vezes).
- A probabilidade de sair cara num lançamento vai decaindo.
- No primeiro lançamento, a probbilidade de sair cara é p_1.
- No segundo, ela diminui para um valor $p_2 < p_1$.
- No terceiro fica menor ainda: $p_3 < p_2 < p_1$.
- E assim por diante, com p_n sempre maior que zero mas ...

$\{0, 1\}^\infty$

- Joga-se uma moeda para cima *indefinidamente* (infinitas vezes).
- A probabilidade de sair cara num lançamento vai decaindo.
- No primeiro lançamento, a probbilidade de sair cara é p_1.
- No segundo, ela diminui para um valor $p_2 < p_1$.
- No terceiro fica menor ainda: $p_3 < p_2 < p_1$.
- E assim por diante, com p_n sempre maior que zero mas ...
- indo a zero: $p_n \to 0$.

$\{0, 1\}^\infty$

- Por exemplo, podemos ter $p_n = 1/n$ ou $p_n = (1/2)^n$ ou ainda $p_n = 1/\log(n)$.

$\{0,1\}^\infty$

- Por exemplo, podemos ter $p_n = 1/n$ ou $p_n = (1/2)^n$ ou ainda $p_n = 1/\log(n)$.
- Queremos calcular probabilidades deste tipo:
- qual a probabilidade de que, depois de certo n, nunca mais vejamos uma cara.

$\{0, 1\}^\infty$

- Por exemplo, podemos ter $p_n = 1/n$ ou $p_n = (1/2)^n$ ou ainda $p_n = 1/\log(n)$.
- Queremos calcular probabilidades deste tipo:
- qual a probabilidade de que, depois de certo n, nunca mais vejamos uma cara.
- Isto é, qual a probabilidade de que o número total de caras seja finito?

$\{0, 1\}^\infty$

- Por exemplo, podemos ter $p_n = 1/n$ ou $p_n = (1/2)^n$ ou ainda $p_n = 1/\log(n)$.
- Queremos calcular probabilidades deste tipo:
- qual a probabilidade de que, depois de certo n, nunca mais vejamos uma cara.
- Isto é, qual a probabilidade de que o número total de caras seja finito?
- Qual a probabilidade de que caras sejam extintas depois de certo tempo?
- O que sua intuição diz?

Teorema de Borel-Cantelli

- Calcule $S = \sum_{n=1}^{\infty} p_n$.

Teorema de Borel-Cantelli

- Calcule $S = \sum_{n=1}^{\infty} p_n$.
- Por exemplo, podemos ter $S = \sum_{n=1}^{\infty} (\frac{1}{2})^n = 2$ ou $S = \sum_{n=1}^{\infty} \frac{1}{n} = \infty$.

Teorema de Borel-Cantelli

- Calcule $S = \sum_{n=1}^{\infty} p_n$.
- Por exemplo, podemos ter $S = \sum_{n=1}^{\infty} (\frac{1}{2})^n = 2$ ou $S = \sum_{n=1}^{\infty} \frac{1}{n} = \infty$.
- Se $S < \infty$, então apenas um número finito de caras vai ocorrer na sequência de infinitos lançamentos da moeda.

Teorema de Borel-Cantelli

- Calcule $S = \sum_{n=1}^{\infty} p_n$.
- Por exemplo, podemos ter $S = \sum_{n=1}^{\infty} (\frac{1}{2})^n = 2$ ou $S = \sum_{n=1}^{\infty} \frac{1}{n} = \infty$.
- Se $S < \infty$, então apenas um número finito de caras vai ocorrer na sequência de infinitos lançamentos da moeda.
- Depois de certo n, elas com certeza vão desaparecer. Se esperarmos um tempo longo o suficiente, elas somem de vez.

Teorema de Borel-Cantelli

- Calcule $S = \sum_{n=1}^{\infty} p_n$.
- Por exemplo, podemos ter $S = \sum_{n=1}^{\infty} (\frac{1}{2})^n = 2$ ou $S = \sum_{n=1}^{\infty} \frac{1}{n} = \infty$.
- Se $S < \infty$, então apenas um número finito de caras vai ocorrer na sequência de infinitos lançamentos da moeda.
- Depois de certo n, elas com certeza vão desaparecer. Se esperarmos um tempo longo o suficiente, elas somem de vez.
- Mas se $S = \infty$, vai aparecer um número infinito de caras.
- Por exemplo, se $p_n = \frac{1}{n}$, veremos caras cada vez mais raramente mas elas *nunca* desaparecem completamente.

Qual é o Ω?

- Joga-se uma moeda para cima independentemente *indefinidamente*.
- Vamos representar por 0 e 1 os dois resultados possíveis de um lançamento.

Qual é o Ω?

- Joga-se uma moeda para cima independentemente *indefinidamente*.
- Vamos representar por 0 e 1 os dois resultados possíveis de um lançamento.
- Como não existe um limite para o número de lançamentos da moeda, o espaço amostral será composto por elementos da forma

$$\omega = (a_1, a_2, a_3, \ldots)$$

- onde $a_i = 0$ ou $a_i = 1$.

Qual é o Ω?

- Joga-se uma moeda para cima independentemente *indefinidamente*.
- Vamos representar por 0 e 1 os dois resultados possíveis de um lançamento.
- Como não existe um limite para o número de lançamentos da moeda, o espaço amostral será composto por elementos da forma

$$\omega = (a_1, a_2, a_3, \ldots)$$

- onde $a_i = 0$ ou $a_i = 1$.
- Isto é, um vetor de comprimento infinito onde cada entrada é 0 ou 1.

Qual é o Ω?

- Joga-se uma moeda para cima independentemente *indefinidamente*.
- Vamos representar por 0 e 1 os dois resultados possíveis de um lançamento.
- Como não existe um limite para o número de lançamentos da moeda, o espaço amostral será composto por elementos da forma

$$\omega = (a_1, a_2, a_3, \ldots)$$

- onde $a_i = 0$ ou $a_i = 1$.
- Isto é, um vetor de comprimento infinito onde cada entrada é 0 ou 1.
- O espaço amostral Ω é composto pelos infinitos elementos ω desta forma: strings infinitos compostos de 0's e 1's .

Qual é o Ω?

- Joga-se uma moeda para cima independentemente *indefinidamente*.
- Vamos representar por 0 e 1 os dois resultados possíveis de um lançamento.
- Como não existe um limite para o número de lançamentos da moeda, o espaço amostral será composto por elementos da forma

$$\omega = (a_1, a_2, a_3, \ldots)$$

- onde $a_i = 0$ ou $a_i = 1$.
- Isto é, um vetor de comprimento infinito onde cada entrada é 0 ou 1.
- O espaço amostral Ω é composto pelos infinitos elementos ω desta forma: strings infinitos compostos de 0's e 1's .
- Curiosidade: $\Omega = [0, 1]$ pois a expansão de um número real em $[0, 1]$ na base 2 é um desses ω.

A σ-álgebra $\mathcal{A}$

- O 2^o elemento do espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$ é a σ-álgebra $\mathcal{A}$.

A σ-álgebra $\mathcal{A}$

- O 2^o elemento do espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$ é a σ-álgebra $\mathcal{A}$.
- Queremos atribuir probabilidades $\mathbb{P}(\omega)$ a elementos $\omega \in \Omega$.

A σ-álgebra $\mathcal{A}$

- O 2^o elemento do espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$ é a σ-álgebra $\mathcal{A}$.
- Queremos atribuir probabilidades $\mathbb{P}(\omega)$ a elementos $\omega \in \Omega$.
- Mas queremos também atribuir probabilidades a *subconjuntos* de elementos de Ω
- Se $A \subset \Omega$ queremos calcular $\mathbb{P}(A)$ de alguma forma.
- Todo e qualquer evento para qual queremos calcular uma probabilidade será um sub-conjunto $A \subseteq \Omega$.

A σ-álgebra $\mathcal{A}$

- Por exemplo, rolar um dado e observar a sua face: $\Omega = \{1, 2, 3, 4, 5, 6\}$

A σ-álgebra $\mathcal{A}$

- Por exemplo, rolar um dado e observar a sua face: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Queremos calcular $\mathbb{P}(\text{Sair um 4}) = \mathbb{P}(\{4\}) = \mathbb{P}(4)$.

A σ-álgebra $\mathcal{A}$

- Por exemplo, rolar um dado e observar a sua face: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Queremos calcular $\mathbb{P}(\text{Sair um 4}) = \mathbb{P}(\{4\}) = \mathbb{P}(4)$.
- Queremos também calcular:
  - $\mathbb{P}(\text{Sair uma face maior que 4}) = \mathbb{P}(\{5, 6\})$.

A σ-álgebra $\mathcal{A}$

- Por exemplo, rolar um dado e observar a sua face: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Queremos calcular $\mathbb{P}(\text{Sair um } 4) = \mathbb{P}(\{4\}) = \mathbb{P}(4)$.
- Queremos também calcular:
    - $\mathbb{P}(\text{Sair uma face maior que } 4) = \mathbb{P}(\{5, 6\})$.
    - $\mathbb{P}(\text{Sair face ímpar}) = \mathbb{P}(\{1, 3, 5\})$.

A σ-álgebra $\mathcal{A}$

- Por exemplo, rolar um dado e observar a sua face: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Queremos calcular $\mathbb{P}(\text{Sair um 4}) = \mathbb{P}(\{4\}) = \mathbb{P}(4)$.
- Queremos também calcular:
  - $\mathbb{P}(\text{Sair uma face maior que 4}) = \mathbb{P}(\{5, 6\})$.
  - $\mathbb{P}(\text{Sair face ímpar}) = \mathbb{P}(\{1, 3, 5\})$.
- Jargão: dizemos que $\mathbb{P}(A)$ é a probabilidade de ocorrer o evento A. O que isto quer dizer?

A σ-álgebra $\mathcal{A}$

- Por exemplo, rolar um dado e observar a sua face: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Queremos calcular $\mathbb{P}(\text{Sair um 4}) = \mathbb{P}(\{4\}) = \mathbb{P}(4)$.
- Queremos também calcular:
  - $\mathbb{P}(\text{Sair uma face maior que 4}) = \mathbb{P}(\{5, 6\})$.
  - $\mathbb{P}(\text{Sair face ímpar}) = \mathbb{P}(\{1, 3, 5\})$.
- Jargão: dizemos que $\mathbb{P}(A)$ é a probabilidade de ocorrer o evento A. O que isto quer dizer?
- Ao observarmos o fenômeno aleatório que estamos modelando, aparece por acaso um único resultado $\omega \in \Omega$.

A σ-álgebra $\mathcal{A}$

- Por exemplo, rolar um dado e observar a sua face: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Queremos calcular $\mathbb{P}(\text{Sair um 4}) = \mathbb{P}(\{4\}) = \mathbb{P}(4)$.
- Queremos também calcular:
  - $\mathbb{P}(\text{Sair uma face maior que 4}) = \mathbb{P}(\{5, 6\})$.
  - $\mathbb{P}(\text{Sair face ímpar}) = \mathbb{P}(\{1, 3, 5\})$.
- Jargão: dizemos que $\mathbb{P}(A)$ é a probabilidade de ocorrer o evento A. O que isto quer dizer?
- Ao observarmos o fenômeno aleatório que estamos modelando, aparece por acaso um único resultado $\omega \in \Omega$.
- Então $\mathbb{P}(A)$ é a probabilidade de que este resultado ω seja um elemento do conjunto A.
- $\mathbb{P}(A) = \mathbb{P}(\omega \in A)$

A σ-álgebra $\mathcal{A}$

- Por exemplo, rolar um dado e observar a sua face: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Queremos calcular $\mathbb{P}(\text{Sair um 4}) = \mathbb{P}(\{4\}) = \mathbb{P}(4)$.
- Queremos também calcular:
    - $\mathbb{P}(\text{Sair uma face maior que 4}) = \mathbb{P}(\{5, 6\})$.
    - $\mathbb{P}(\text{Sair face ímpar}) = \mathbb{P}(\{1, 3, 5\})$.
- Jargão: dizemos que $\mathbb{P}(A)$ é a probabilidade de ocorrer o evento A. O que isto quer dizer?
- Ao observarmos o fenômeno aleatório que estamos modelando, aparece por acaso um único resultado $\omega \in \Omega$.
- Então $\mathbb{P}(A)$ é a probabilidade de que este resultado ω seja um elemento do conjunto A.
- $\mathbb{P}(A) = \mathbb{P}(\omega \in A)$
- Ainda não definimos probabilidade!
- Se $A \subset \Omega$ queremos calcular $\mathbb{P}(A)$ de alguma forma.

A σ-álgebra $\mathcal{A}$

- A σ-álgebra $\mathcal{A}$ é o conjunto dos sub-conjuntos A de Ω para os quais podemos calcular $\mathbb{P}(A)$.

A σ-álgebra $\mathcal{A}$

- A σ-álgebra $\mathcal{A}$ é o conjunto dos sub-conjuntos A de Ω para os quais podemos calcular $\mathbb{P}(A)$.
- Os sub-conjuntos $A \in \mathcal{A}$ são chamados de *eventos*.

A σ-álgebra $\mathcal{A}$

- A σ-álgebra $\mathcal{A}$ é o conjunto dos sub-conjuntos A de Ω para os quais podemos calcular $\mathbb{P}(A)$.
- Os sub-conjuntos $A \in \mathcal{A}$ são chamados de *eventos*.
- Sub-conjuntos $A = \{\omega\}$, com um único elemento de Ω são chamados de *eventos atômicos*.

A σ-álgebra $\mathcal{A}$

- Idealmente, queremos calcular $\mathbb{P}(A)$ para TODO E QUALQUER subconjunto $A \subset \Omega$.
- Infelizmente, *em alguns casos*, não podemos calcular $\mathbb{P}(A)$ para TODO E QUALQUER subconjunto $A \subset \Omega$.

A σ-álgebra $\mathcal{A}$

- Idealmente, queremos calcular $\mathbb{P}(A)$ para TODO E QUALQUER subconjunto $A \subset \Omega$.
- Infelizmente, *em alguns casos*, não podemos calcular $\mathbb{P}(A)$ para TODO E QUALQUER subconjunto $A \subset \Omega$.
- A σ-álgebra $\mathcal{A}$ é simplesmente a classe dos sub-conjuntos de Ω para os quais podemos calcular $\mathbb{P}(A)$.

A σ-álgebra $\mathcal{A}$

- Idealmente, queremos calcular $\mathbb{P}(A)$ para TODO E QUALQUER subconjunto $A \subset \Omega$.
- Infelizmente, *em alguns casos*, não podemos calcular $\mathbb{P}(A)$ para TODO E QUALQUER subconjunto $A \subset \Omega$.
- A σ-álgebra $\mathcal{A}$ é simplesmente a classe dos sub-conjuntos de Ω para os quais podemos calcular $\mathbb{P}(A)$.
- Entretanto, *qualquer* evento que você conceber e que seja útil na prática, mesmo que muito complicado, fará parte da σ-álgebra $\mathcal{A}$.
- Existe uma discussão um pouquinho mais longa sobre isto nas notas de aula.

A função de probabilidade $\mathbb{P}$

- O 3^o elemento do espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$ é a atribuição de probabilidades aos eventos $A \subset \Omega$.

A função de probabilidade $\mathbb{P}$

- O 3^o elemento do espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$ é a atribuição de probabilidades aos eventos $A \subset \Omega$.
- Já definimos quais são os resultados possíveis do fenêmeno aleatório: são os elementos $\omega \in \Omega$.

A função de probabilidade $\mathbb{P}$

- O 3^o elemento do espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$ é a atribuição de probabilidades aos eventos $A \subset \Omega$.
- Já definimos quais são os resultados possíveis do fenêmeno aleatório: são os elementos $\omega \in \Omega$.
- Já definimos também quais são os sub-conjuntos $A \subset \Omega$ para os quais podemos calcular uma probabilidade $\mathbb{P}(A)$: qualquer sub-conjunto de Ω que pudermos conceber na prática.

A função de probabilidade $\mathbb{P}$

- O 3^o elemento do espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$ é a atribuição de probabilidades aos eventos $A \subset \Omega$.
- Já definimos quais são os resultados possíveis do fenêmeno aleatório: são os elementos $\omega \in \Omega$.
- Já definimos também quais são os sub-conjuntos $A \subset \Omega$ para os quais podemos calcular uma probabilidade $\mathbb{P}(A)$: qualquer sub-conjunto de Ω que pudermos conceber na prática.
- Precisamos agora definir $\mathbb{P}(A)$ para todo *A de forma consistente*.

A função de probabilidade $\mathbb{P}$

- O 3^o elemento do espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$ é a atribuição de probabilidades aos eventos $A \subset \Omega$.
- Já definimos quais são os resultados possíveis do fenêmeno aleatório: são os elementos $\omega \in \Omega$.
- Já definimos também quais são os sub-conjuntos $A \subset \Omega$ para os quais podemos calcular uma probabilidade $\mathbb{P}(A)$: qualquer sub-conjunto de Ω que pudermos conceber na prática.
- Precisamos agora definir $\mathbb{P}(A)$ para todo A *de forma consistente*.
- Quais as propriedades que esta atribuição deve ter para que a gente não chegue a resultados inconsistentes ou contraditórios?

A função de probabilidade $\mathbb{P}$

- O 3^o elemento do espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$ é a atribuição de probabilidades aos eventos $A \subset \Omega$.
- Já definimos quais são os resultados possíveis do fenêmeno aleatório: são os elementos $\omega \in \Omega$.
- Já definimos também quais são os sub-conjuntos $A \subset \Omega$ para os quais podemos calcular uma probabilidade $\mathbb{P}(A)$: qualquer sub-conjunto de Ω que pudermos conceber na prática.
- Precisamos agora definir $\mathbb{P}(A)$ para todo A *de forma consistente*.
- Quais as propriedades que esta atribuição deve ter para que a gente não chegue a resultados inconsistentes ou contraditórios?
- Quais os requisitos mínimos que esta atribuição de probabilidades deve satisfazer?

A função de probabilidade $\mathbb{P}$

- $\mathbb{P}$ é QUALQUER função:

$$\mathbb{P} : \mathcal{A} \longrightarrow \mathbb{R}$$
$$A \longrightarrow \mathbb{P}(A)$$

A função de probabilidade $\mathbb{P}$

- $\mathbb{P}$ é QUALQUER função:

$$\mathbb{P} : \mathcal{A} \longrightarrow \mathbb{R}$$
$$A \longrightarrow \mathbb{P}(A)$$

que obedeça aos seguintes três axiomas de Kolmogorov:

- Axioma 1: $\mathbb{P}(A) \geq 0 \quad \forall \quad A \in \mathcal{A}$
- Axioma 2: $\mathbb{P}(\Omega) = 1$

A função de probabilidade $\mathbb{P}$

- $\mathbb{P}$ é QUALQUER função:

$$\mathbb{P} : \mathcal{A} \longrightarrow \mathbb{R}$$
$$A \longrightarrow \mathbb{P}(A)$$

  que obedeça aos seguintes três axiomas de Kolmogorov:
  - Axioma 1: $\mathbb{P}(A) \geq 0 \quad \forall \quad A \in \mathcal{A}$
  - Axioma 2: $\mathbb{P}(\Omega) = 1$
- Estes dois primeiros axiomas estão apenas fixando uma escala para a probabilidade:
  - a probabilidade de um evento é um número maior que zero
  - a probabilidade de que ocorra algum elemento de Ω é $\mathbb{P}(\Omega) = 1$.
  - Vamos ver que isto implica que $\mathbb{P}(A) \in [0, 1]$ para qualquer evento A.
- O importante mesmo é o terceiro axioma que discutiremos a seguir.

O terceiro axioma: um caso particular

- Antes de ver o axioma 3 em toda sua generalidade, vamos considerar um caso particular.
- O axioma 3 afirma que a função de probabilidade seja aditiva para pares eventos disjuntos.
- Isto é, se $A \cap B = \emptyset$ então $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.
- Vamos entender o que este axioma está afirmando ignorando casos extremos:
- O evento $A \cup B$ é maior que o evento A sozinho. Esperamos então que $\mathbb{P}(A \cup B) \geq \mathbb{P}(A)$.
- O mesmo vale para B de modo que esperamos $\mathbb{P}(A \cup B) \geq \mathbb{P}(B)$.
- Mas quanto é este acréscimo que devemos dar à $\mathbb{P}(A)$ para chegarmos a $\mathbb{P}(A \cup B)$?

O terceiro axioma: um caso particular

- O axioma 3 afirma: basta somar as probabilidades dos dois eventos *se eles forem disjuntos*:

$$\mathbb{P}\left(A \cup B\right) = \mathbb{P}(A) + \mathbb{P}(B)$$

  se $A \cap B = \emptyset$.
- Assim, se olhamos o resultado de lançar um dado com 6 faces com $\Omega = \{1, 2, 3, 4, 5, 6\}$ com
- $A = \text{ face é par } = \{2, 4, 6\}$ e $B = \{5\}$, então
- $A \cup B = \text{ face é par ou 5 }$ e

$$\begin{aligned} \mathbb{P}\left(A \cup B\right) &= \mathbb{P}\left(\text{ face é par ou 5 } \right) \\ &= \mathbb{P}\left(\text{ face é par } \right) + \mathbb{P}\left(\text{ face é 5 } \right) \\ &= \mathbb{P}(A) + \mathbb{P}(B) \end{aligned}$$

O terceiro axioma: um caso particular

- O terceiro axioma cobre também o caso de mais de um evento disjunto.
- Se A, B, C são três eventos mutuamente exclusivos (todos disjuntos)
- então

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C)$$

- Esta é uma condição para que uma função $\mathbb{P}$ possa ser chamada de probabilidade.
- O axioma 3 precisa ser um pouco mais geral: ele precisa valer para qualquer lista enumerável de eventos disjuntos.

O terceiro axioma: um caso particular

- Isto vale apenas se os eventos A e B forem disjuntos.
- Se $A = \text{ face é par } = \{2, 4, 6\}$
- mas $B = \text{ face é menor que 4 } = \{1, 2, 3\}$,
- então $A \cap B \neq \emptyset$ e

$$\mathbb{P}(A \cup B) \neq \mathbb{P}(A) + \mathbb{P}(B)$$

O terceiro axioma

- Axioma 3:

$$\mathbb{P}\left(A_1 \cup A_2 \cup A_3 \cup \ldots\right) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) + \ldots$$

  se os eventos $A_1, A_2, A_3, \ldots$ forem todos disjuntos (isto é, mutuamente exclusivos).
- Jargão: probabilidade é uma função σ-aditiva.
- Se os A_i's são disjuntos então

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$$

O terceiro axioma

- Para entender isto um pouco melhor, vamos considerar o caso em que:
- $A_1 = A$, $A_2 = B$, $A_3 = C$
- e $A_n = \emptyset$ para $n = 4, 5, \ldots$
- Assim,

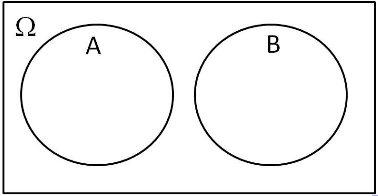
$$\begin{aligned} \bigcup_{n=1}^{\infty} A_n &= A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5 \cup A_6 \cup \ldots \\ &= A \cup B \cup C \cup \emptyset \cup \emptyset \ldots \\ &= A \cup B \cup C \end{aligned}$$

$\mathbb{P}(A \cup B)$

- Pelo axioma 3, se $A \cap B = \varnothing$ e $A \cap C = \varnothing$ e $B \cap C = \varnothing$ então

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C)$$

- A probabilidade é uma função aditiva sobre a união de eventos (ou conjuntos) disjuntos.
- O Axioma 3 pede que esta propriedade aditiva valha com somas infinitas (enumeráveis).



Consequências

- Basta que $\mathbb{P}$ satisfaça aos três axiomas de Kolmogorov para que $\mathbb{P}$ seja uma atribuição de probabilidades válida.

Consequências

- Basta que $\mathbb{P}$ satisfaça aos três axiomas de Kolmogorov para que $\mathbb{P}$ seja uma atribuição de probabilidades válida.
- TODO o restante do cálculo de propriedades é decorrente destes três axiomas de Kolmogorov. Por exemplo:

  (P1) $\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$

Consequências

- Basta que $\mathbb{P}$ satisfaça aos três axiomas de Kolmogorov para que $\mathbb{P}$ seja uma atribuição de probabilidades válida.
- TODO o restante do cálculo de propriedades é decorrente destes três axiomas de Kolmogorov. Por exemplo:

(P1) $\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$

(P2) $0 \leq \mathbb{P}(A) \leq 1$ para todo evento $A \in \mathcal{A}$.

Consequências

- Basta que $\mathbb{P}$ satisfaça aos três axiomas de Kolmogorov para que $\mathbb{P}$ seja uma atribuição de probabilidades válida.
- TODO o restante do cálculo de propriedades é decorrente destes três axiomas de Kolmogorov. Por exemplo:

(P1) $\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$

(P2) $0 \leq \mathbb{P}(A) \leq 1$ para todo evento $A \in \mathcal{A}$.

(P3) se $A_1 \subset A_2 \Longrightarrow \mathbb{P}(A_1) \leq \mathbb{P}(A_2)$

Consequências

- Basta que $\mathbb{P}$ satisfaça aos três axiomas de Kolmogorov para que $\mathbb{P}$ seja uma atribuição de probabilidades válida.
- TODO o restante do cálculo de propriedades é decorrente destes três axiomas de Kolmogorov. Por exemplo:

(P1) $\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$

(P2) $0 \leq \mathbb{P}(A) \leq 1$ para todo evento $A \in \mathcal{A}$.

(P3) se $A_1 \subset A_2 \Longrightarrow \mathbb{P}(A_1) \leq \mathbb{P}(A_2)$

(P4) $\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_i\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_i)$

Consequências

- Basta que $\mathbb{P}$ satisfaça aos três axiomas de Kolmogorov para que $\mathbb{P}$ seja uma atribuição de probabilidades válida.
- TODO o restante do cálculo de propriedades é decorrente destes três axiomas de Kolmogorov. Por exemplo:

(P1) $\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$

(P2) $0 \leq \mathbb{P}(A) \leq 1$ para todo evento $A \in \mathcal{A}$.

(P3) se $A_1 \subset A_2 \Longrightarrow \mathbb{P}(A_1) \leq \mathbb{P}(A_2)$

(P4) $\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_i\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_i)$

(P5) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

Consequências

- Basta que $\mathbb{P}$ satisfaça aos três axiomas de Kolmogorov para que $\mathbb{P}$ seja uma atribuição de probabilidades válida.
- TODO o restante do cálculo de propriedades é decorrente destes três axiomas de Kolmogorov. Por exemplo:

(P1) $\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$

(P2) $0 \leq \mathbb{P}(A) \leq 1$ para todo evento $A \in \mathcal{A}$.

(P3) se $A_1 \subset A_2 \Longrightarrow \mathbb{P}(A_1) \leq \mathbb{P}(A_2)$

(P4) $\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_i\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_i)$

(P5) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

- Esta última propriedade é o caso geral de $\mathbb{P}(A \cup B)$ quando A e B

Provar que $\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$

- Temos $\mathbb{P}(\Omega) = 1$ e $\Omega = A \cup A^c$.
- Como $A \cap A^c = \varnothing$, pelo axioma 3 temos

$$1 = \mathbb{P}\left(A \cup A^c\right) = \mathbb{P}\left(A\right) + \mathbb{P}\left(A^c\right)$$

- e portanto

$$\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$$

Provar que se $A_1 \subset A_2 \Longrightarrow \mathbb{P}(A_1) \leq \mathbb{P}(A_2)$

- Como $A_1 \subset A_2$, podemos escrever $A_2 = A_1 \cup (A_2 - A_1)$.
- Como $A_1 \cap (A_2 - A_1) = \varnothing$, pelo axioma 3 temos

$$\mathbb{P}(A_2) = \mathbb{P}(A_1 \cup (A_2 - A_1)) = \mathbb{P}(A_1) + \mathbb{P}(A_2 - A_1) \geq \mathbb{P}(A_1)$$

- pois, pelo axioma 1, $\mathbb{P}(A_2 - A_1)$ tem de ser maior ou igual a zero.

Como estabelecer uma função $\mathbb{P}$?

- OK, QUALQUER função $\mathbb{P}$ que satisfaça aos axiomas de Kolmogorov é válida.

Como estabelecer uma função $\mathbb{P}$?

- OK, QUALQUER função $\mathbb{P}$ que satisfaça aos axiomas de Kolmogorov é válida.
- Mas como escolher uma função que satisfaça o axioma 3 e, mais importante, como escolher uma dessas funções válidas num caso prático?

Como estabelecer uma função $\mathbb{P}$?

- OK, QUALQUER função $\mathbb{P}$ que satisfaça aos axiomas de Kolmogorov é válida.
- Mas como escolher uma função que satisfaça o axioma 3 e, mais importante, como escolher uma dessas funções válidas num caso prático?
- Usamos uma combinação de:
  - conveniência matemática (facilidade de manuseio).
  - com boa aproximação da realidade.

Como estabelecer uma função $\mathbb{P}$?

- OK, QUALQUER função $\mathbb{P}$ que satisfaça aos axiomas de Kolmogorov é válida.
- Mas como escolher uma função que satisfaça o axioma 3 e, mais importante, como escolher uma dessas funções válidas num caso prático?
- Usamos uma combinação de:
    - conveniência matemática (facilidade de manuseio).
    - com boa aproximação da realidade.
- Existe um trade-off entre estes dois aspectos.
- Se focarmos apenas no uso de modelos matematicamente muito simples vamos acabar com modelos que são muito distantes da realidade do fenômeno, que não o representam bem.

Como estabelecer uma função $\mathbb{P}$?

- OK, QUALQUER função $\mathbb{P}$ que satisfaça aos axiomas de Kolmogorov é válida.
- Mas como escolher uma função que satisfaça o axioma 3 e, mais importante, como escolher uma dessas funções válidas num caso prático?
- Usamos uma combinação de:
  - conveniência matemática (facilidade de manuseio).
  - com boa aproximação da realidade.
- Existe um trade-off entre estes dois aspectos.
- Se focarmos apenas no uso de modelos matematicamente muito simples vamos acabar com modelos que são muito distantes da realidade do fenômeno, que não o representam bem.
- Se insistirmos em incorporar todoss os aspectos que podem afetar um fenômeno, teremos um modelo probabilístico inviável do ponto de vista matemático e computacional.

Como estabelecer uma função $\mathbb{P}$?

- Para definir a função de probabilidade $\mathbb{P}$ devemos considerar três casos:

Como estabelecer uma função $\mathbb{P}$?

- Para definir a função de probabilidade $\mathbb{P}$ devemos considerar três casos:
    - Ω é finito: $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$

Como estabelecer uma função $\mathbb{P}$?

- Para definir a função de probabilidade $\mathbb{P}$ devemos considerar três casos:
  - Ω é finito: $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$
  - Ω é infinito enumerável: $\Omega = \{\omega_1, \omega_2, \ldots\}$, tal como $\Omega = \{0, 1, 2, 3, \ldots\}$

Como estabelecer uma função $\mathbb{P}$?

- Para definir a função de probabilidade $\mathbb{P}$ devemos considerar três casos:
  - Ω é finito: $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$
  - Ω é infinito enumerável: $\Omega = \{\omega_1, \omega_2, \ldots\}$, tal como $\Omega = \{0, 1, 2, 3, \ldots\}$
  - Ω é não-enumerável, tal como $\Omega = (0, 1)$ ou $\Omega = \mathbb{R}^2$.
- O terceiro caso tem algumas complicações a mais em relação aos outros dois.

$\mathbb{P}$ quando Ω é finito

- Seja $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ onde os ω_i são eventos atômicos distintos, indivisíveis.
- Notação: $\mathbb{P}(\{\omega_i\}) = \mathbb{P}(\omega_i)$.

$\mathbb{P}$ quando Ω é finito

- Seja $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ onde os ω_i são eventos atômicos distintos, indivisíveis.
- Notação: $\mathbb{P}(\{\omega_i\}) = \mathbb{P}(\omega_i)$.
- Atribua valores $\mathbb{P}(\omega_i) \geq 0$ arbitrariamente mas ...

$\mathbb{P}$ quando Ω é finito

- Seja $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ onde os ω_i são eventos atômicos distintos, indivisíveis.
- Notação: $\mathbb{P}(\{\omega_i\}) = \mathbb{P}(\omega_i)$.
- Atribua valores $\mathbb{P}(\omega_i) \geq 0$ arbitrariamente mas ...
- ...com a restrição de que sua soma seja igual a 1:

$$\mathbb{P}(\omega_1) + \ldots + \mathbb{P}(\omega_N) = \sum_{i=1}^{N} \mathbb{P}(\omega_i) = 1$$

- Qualquer função definida assim satisfaz os três axiomas de Kolmogorov e é válida.

Exemplo: micro data mining

- Suponha que existam apenas três produtos: A, B, e C.
- Ω é composto pelas possíveis 8 cestas de produtos:
  - 0 (ou nenhum produto),
  - apenas A, apenas B, apenas C,
  - apenas os produtos AB juntos, apenas AC juntos, apenas BC juntos,
  - os 3 produtos ABC juntos.

Exemplo: micro data mining

- Suponha que existam apenas três produtos: A, B, e C.
- Ω é composto pelas possíveis 8 cestas de produtos:
  - 0 (ou nenhum produto),
  - apenas A, apenas B, apenas C,
  - apenas os produtos AB juntos, apenas AC juntos, apenas BC juntos,
  - os 3 produtos ABC juntos.
- Vamos representar $\Omega = \{0, A, B, C, AB, AC, BC, ABC\}$

Exemplo: micro data mining

- Suponha que existam apenas três produtos: A, B, e C.
- Ω é composto pelas possíveis 8 cestas de produtos:
    - 0 (ou nenhum produto),
    - apenas A, apenas B, apenas C,
    - apenas os produtos AB juntos, apenas AC juntos, apenas BC juntos,
    - os 3 produtos ABC juntos.
- Vamos representar $\Omega = \{0, A, B, C, AB, AC, BC, ABC\}$
- Fez-se uma análise estatística do padrão de compras de vários clientes.

Exemplo: micro data mining

- Suponha que existam apenas três produtos: A, B, e C.
- Ω é composto pelas possíveis 8 cestas de produtos:
  - 0 (ou nenhum produto),
  - apenas A, apenas B, apenas C,
  - apenas os produtos AB juntos, apenas AC juntos, apenas BC juntos,
  - os 3 produtos ABC juntos.
- Vamos representar $\Omega = \{0, A, B, C, AB, AC, BC, ABC\}$
- Fez-se uma análise estatística do padrão de compras de vários clientes.
- Observou-se, por exemplo, que aproximadamente 17% dos clientes saíam com a cesta A e 3% saíam com a cesta AB.

Exemplo: micro data mining

- Isto permitiu obter aproximadamente as probabilidades as possibilidades de cada $\omega \in \Omega$.

Exemplo: micro data mining

- Isto permitiu obter aproximadamente as probabilidades as possibilidades de cada $\omega \in \Omega$.
- Por exemplo, $\mathbb{P}(A) \approx 0.17$ e $\mathbb{P}(AB) \approx 0.03$

Exemplo: micro data mining

- Isto permitiu obter aproximadamente as probabilidades as possibilidades de cada $\omega \in \Omega$.
- Por exemplo, $\mathbb{P}(A) \approx 0.17$ e $\mathbb{P}(AB) \approx 0.03$
- Assim, podemos atribuir probabilidades aos elementos atômicos de Ω:

ω	0	A	B	C	AB	BC	AC	ABC	*soma*
$\mathbb{P}(\omega)$	0.02	0.17	0.19	0.09	0.03	0.21	0.18	0.11	1.00

$\mathbb{P}$ quando Ω é finito

- Temos $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$.
- Como fica a probabilidade $\mathbb{P}(A)$ de um evento $A = \{\omega_{i_1}, \ldots \omega_{i_n}\}$?

$\mathbb{P}$ quando Ω é finito

- Temos $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$.
- Como fica a probabilidade $\mathbb{P}(A)$ de um evento $A = \{\omega_{i_1}, \ldots \omega_{i_n}\}$?
- A é um conjunto finito, sub-conjunto de Ω, com n elementos.

$\mathbb{P}$ quando Ω é finito

- Temos $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$.
- Como fica a probabilidade $\mathbb{P}(A)$ de um evento $A = \{\omega_{i_1}, \ldots \omega_{i_n}\}$?
- A é um conjunto finito, sub-conjunto de Ω, com n elementos.

- Seja $A = \{\omega_{i_1}, \ldots \omega_{i_n}\} = \bigcup_j \{\omega_{i_j}\}$

$\mathbb{P}$ quando Ω é finito

- Temos $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$.
- Como fica a probabilidade $\mathbb{P}(A)$ de um evento $A = \{\omega_{i_1}, \ldots \omega_{i_n}\}$?
- A é um conjunto finito, sub-conjunto de Ω, com n elementos.

- Seja $A = \{\omega_{i_1}, \ldots \omega_{i_n}\} = \bigcup_j \{\omega_{i_j}\}$
- A é a união de n eventos atômicos disjuntos.

$\mathbb{P}$ quando Ω é finito

- Temos $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$.
- Como fica a probabilidade $\mathbb{P}(A)$ de um evento $A = \{\omega_{i_1}, \ldots \omega_{i_n}\}$?
- A é um conjunto finito, sub-conjunto de Ω, com n elementos.

- Seja $A = \{\omega_{i_1}, \ldots \omega_{i_n}\} = \bigcup_j \{\omega_{i_j}\}$
- A é a união de n eventos atômicos disjuntos.
- Para o Axioma 3 ser válido devemos ter

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_j \{\omega_{i_j}\}\right) = \sum_{j=1}^{k} \mathbb{P}\left(\{\omega_{i_j}\}\right) = \sum_{\omega_i \in A} \mathbb{P}(\omega_i)$$

$\mathbb{P}$ quando Ω é finito

- Repetindo: atribua probabilidades (somando 1) aos eventos atômicos $\omega \in \Omega$.

$\mathbb{P}$ quando Ω é finito

- Repetindo: atribua probabilidades (somando 1) aos eventos atômicos $\omega \in \Omega$.
- Para qualquer evento $A \subset \Omega$:
    - identifique quais os elementos ω_i que pertecem a A
    - some suas probabilidades $\mathbb{P}(\omega_i)$

$$\mathbb{P}(A) = \mathbb{P}\left(\{\omega_{i_1}, \ldots \omega_{i_n}\}\right) = \sum_{\omega_i \in A} \mathbb{P}(\omega_i)$$

Exemplo: micro data mining

- Produtos A, B, e C. $\Omega = \{0, A, B, C, AB, AC, BC, ABC\}$ com as probabilidades

ω	0	A	B	C	AB	BC	AC	ABC	*soma*
$\mathbb{P}(\omega)$	0.02	0.17	0.19	0.09	0.03	0.21	0.18	0.11	1.00

Exemplo: micro data mining

- Produtos A, B, e C. $\Omega = \{0, A, B, C, AB, AC, BC, ABC\}$ com as probabilidades

ω	0	A	B	C	AB	BC	AC	ABC	*soma*
$\mathbb{P}(\omega)$	0.02	0.17	0.19	0.09	0.03	0.21	0.18	0.11	1.00

- Alguns eventos compostos e suas probabilidades:
    - E significa levar o produto A na cesta, ou $E = \{A, AB, AC, ABC\}$ e portanto $\mathbb{P}(E) = 0.17 + 0.03 + 0.18 + 0.12 = 0.49$.

Exemplo: micro data mining

- Produtos A, B, e C. $\Omega = \{0, A, B, C, AB, AC, BC, ABC\}$ com as probabilidades

ω	0	A	B	C	AB	BC	AC	ABC	*soma*
$\mathbb{P}(\omega)$	0.02	0.17	0.19	0.09	0.03	0.21	0.18	0.11	1.00

- Alguns eventos compostos e suas probabilidades:
  - E significa levar o produto A na cesta, ou $E = \{A, AB, AC, ABC\}$ e portanto $\mathbb{P}(E) = 0.17 + 0.03 + 0.18 + 0.12 = 0.49$.
  - $E =$ levar o produto A mas não o produto C. Ou seja, $E = \{A, AB\}$ e $\mathbb{P}(E) = 0.17 + 0.03 = 0.20$.

Exemplo: micro data mining

- Produtos A, B, e C. $\Omega = \{0, A, B, C, AB, AC, BC, ABC\}$ com as probabilidades

ω	0	A	B	C	AB	BC	AC	ABC	*soma*
$\mathbb{P}(\omega)$	0.02	0.17	0.19	0.09	0.03	0.21	0.18	0.11	1.00

- Alguns eventos compostos e suas probabilidades:
  - E significa levar o produto A na cesta, ou $E = \{A, AB, AC, ABC\}$ e portanto $\mathbb{P}(E) = 0.17 + 0.03 + 0.18 + 0.12 = 0.49$.
  - $E =$ levar o produto A mas não o produto C. Ou seja, $E = \{A, AB\}$ e $\mathbb{P}(E) = 0.17 + 0.03 = 0.20$.
  - $E =$ uma cesta vazia, ou $E = \{0\}$ e portanto $\mathbb{P}(E) = \mathbb{P}(0) = 0.02$.
  - $E =$ uma cesta vazia ou com pelo menos um produto. Então $E = \Omega$ e portanto $\mathbb{P}(E) = \mathbb{P}(\Omega) = 1$.
  - $E =$ cesta com 4 produtos distintos. Ops, não existe. Portanto $E = \varnothing$, o conjunto vazio, com $\mathbb{P}(\varnothing) = 1 - \mathbb{P}(\varnothing^c) = 1 - \mathbb{P}(\Omega) = 1 - 1 = 0$.

Visão frequentista

- Se o fenômeno puder ser repetido:
  - indefinidadamente,

Visão frequentista

- Se o fenômeno puder ser repetido:
  - indefinidadamente,
  - nas mesmas condições

Visão frequentista

- Se o fenômeno puder ser repetido:
    - indefinidadamente,
    - nas mesmas condições
    - de forma independente (sem que uma repetição afete outra),

Visão frequentista

- Se o fenômeno puder ser repetido:
  - indefinidadamente,
  - nas mesmas condições
  - de forma independente (sem que uma repetição afete outra),
- então $\mathbb{P}(\omega) \approx \dfrac{m}{N}$ onde m é o número de vezes que ω ocorreu e N é o número de repetiçôes.

Visão frequentista

- Se o fenômeno puder ser repetido:
  - indefinidadamente,
  - nas mesmas condições
  - de forma independente (sem que uma repetição afete outra),
- então $\mathbb{P}(\omega) \approx \frac{m}{N}$ onde m é o número de vezes que ω ocorreu e N é o número de repetiçôes.
- Podemos tomar $\mathbb{P}(\omega) = \frac{m}{N}$, ignorando a aproximação amostral embutida.

Visão frequentista

- Se o fenômeno puder ser repetido:
  - indefinidademente,
  - nas mesmas condições
  - de forma independente (sem que uma repetição afete outra),
- então $\mathbb{P}(\omega) \approx \frac{m}{N}$ onde m é o número de vezes que ω ocorreu e N é o número de repetiçôes.
- Podemos tomar $\mathbb{P}(\omega) = \frac{m}{N}$, ignorando a aproximação amostral embutida.
- Esta é chamada a visão *frequentista* de probabilidade.

Visão frequentista

- E $\mathbb{P}(A)$ para $A = \{\omega_{i_1}, \ldots \omega_{i_n}\}$?

Visão frequentista

- E $\mathbb{P}(A)$ para $A = \{\omega_{i_1}, \ldots \omega_{i_n}\}$?
- Temos duas possibilidades.

Visão frequentista

- E $\mathbb{P}(A)$ para $A = \{\omega_{i_1}, \ldots \omega_{i_n}\}$?

- Temos duas possibilidades.
- Tome $\mathbb{P}(\omega_{i_j}) = \frac{m_{i_j}}{N}$ para cada $\omega_{i_j} \in A$ e some estas probabilidades:

$$\mathbb{P}(A) = \sum_j \mathbb{P}(\omega_{i_j}) = \sum_j \frac{m_{i_j}}{N}$$

Visão frequentista

- E $\mathbb{P}(A)$ para $A = \{\omega_{i_1}, \ldots \omega_{i_n}\}$?

- Temos duas possibilidades.
- Tome $\mathbb{P}(\omega_{i_j}) = \dfrac{m_{i_j}}{N}$ para cada $\omega_{i_j} \in A$ e some estas probabilidades:

$$\mathbb{P}(A) = \sum_j \mathbb{P}(\omega_{i_j}) = \sum_j \frac{m_{i_j}}{N}$$

- A outra opção é simplesmente verificar quantas vezes o evento A ocorreu nas N repetições independentes e tomar

$$\mathbb{P}(A) = \frac{m}{N}$$

onde m é o número de vezes que o evento A ocorreu nas N repetições.

Visão frequentista

- E $\mathbb{P}(A)$ para $A = \{\omega_{i_1}, \ldots \omega_{i_n}\}$?

- Temos duas possibilidades.
- Tome $\mathbb{P}(\omega_{i_j}) = \frac{m_{i_j}}{N}$ para cada $\omega_{i_j} \in A$ e some estas probabilidades:

$$\mathbb{P}(A) = \sum_j \mathbb{P}(\omega_{i_j}) = \sum_j \frac{m_{i_j}}{N}$$

- A outra opção é simplesmente verificar quantas vezes o evento A ocorreu nas N repetições independentes e tomar

$$\mathbb{P}(A) = \frac{m}{N}$$

onde m é o número de vezes que o evento A ocorreu nas N repetições.

Frequentista no micro data mining

- No exemplo, temos três produtos: A, B, e C.
- Uma análise estatística permitiu obter aproximadamente as probabilidades:

ω	0	A	B	C	AB	BC	AC	ABC	*soma*
$\mathbb{P}(\omega)$	0.02	0.17	0.19	0.09	0.03	0.21	0.18	0.11	1.00

- Como foram obtidas?

Frequentista no micro data mining

- No exemplo, temos três produtos: A, B, e C.
- Uma análise estatística permitiu obter aproximadamente as probabilidades:

ω	0	A	B	C	AB	BC	AC	ABC	*soma*
$\mathbb{P}(\omega)$	0.02	0.17	0.19	0.09	0.03	0.21	0.18	0.11	1.00

- Como foram obtidas?
- Um grande número N de clientes foram observados: são as repetições.
- Contou-se o número de vezes m em que a cesta foi BC
- Finalmente tivemos $\mathbb{P}(BC) = m/N = 0.21$ ou 21% dos clientes.

Um olhar crítico

- Pelo argumento frequentista, para que $\mathbb{P}(BC) \approx m/N = 0.21$, deveríamos ter repetições nas mesmas condições.

Um olhar crítico

- Pelo argumento frequentista, para que $\mathbb{P}(BC) \approx m/N = 0.21$, deveríamos ter repetições nas mesmas condições.
- Talvez isto não seja razoável.

Um olhar crítico

- Pelo argumento frequentista, para que $\mathbb{P}(BC) \approx m/N = 0.21$, deveríamos ter repetições nas mesmas condições.
- Talvez isto não seja razoável.
- Alguns clientes são velhos, outros são jovens;

Um olhar crítico

- Pelo argumento frequentista, para que $\mathbb{P}(BC) \approx m/N = 0.21$, deveríamos ter repetições nas mesmas condições.
- Talvez isto não seja razoável.
- Alguns clientes são velhos, outros são jovens;
- Alguns compram no inverno e outros no verão, etc.

Um olhar crítico

- Pelo argumento frequentista, para que $\mathbb{P}(BC) \approx m/N = 0.21$, deveríamos ter repetições nas mesmas condições.
- Talvez isto não seja razoável.
- Alguns clientes são velhos, outros são jovens;
- Alguns compram no inverno e outros no verão, etc.
- As condições em que as repetições estão ocorrendo não parecem ser idênticas.

Um olhar crítico

- Pelo argumento frequentista, para que $\mathbb{P}(BC) \approx m/N = 0.21$, deveríamos ter repetições nas mesmas condições.
- Talvez isto não seja razoável.
- Alguns clientes são velhos, outros são jovens;
- Alguns compram no inverno e outros no verão, etc.
- As condições em que as repetições estão ocorrendo não parecem ser idênticas.
- Se as condições não são idênticas pode ser que as probabilidades não se mantenam constantes.

Um olhar crítico

- Uma outra suposição é que as repetições são independentes.

Um olhar crítico

- Uma outra suposição é que as repetições são independentes.
- Vamos formalizar este conceito probabilístico a seguir mas ele significa que o resultado de uma repetição não afeta nenhuma outra.

Um olhar crítico

- Uma outra suposição é que as repetições são independentes.
- Vamos formalizar este conceito probabilístico a seguir mas ele significa que o resultado de uma repetição não afeta nenhuma outra.
- Isto também pode ser questionado.

Um olhar crítico

- Uma outra suposição é que as repetições são independentes.
- Vamos formalizar este conceito probabilístico a seguir mas ele significa que o resultado de uma repetição não afeta nenhuma outra.
- Isto também pode ser questionado.
- Alguns clientes podem influenciar outros via telefone ou comentários.

Um olhar crítico

- Uma outra suposição é que as repetições são independentes.
- Vamos formalizar este conceito probabilístico a seguir mas ele significa que o resultado de uma repetição não afeta nenhuma outra.
- Isto também pode ser questionado.
- Alguns clientes podem influenciar outros via telefone ou comentários.
- Outro motivo é que se os clientes não forem todos distintos, as compras de um mesmo cliente podem ser muito semelhantes.

Um olhar crítico

- Uma outra suposição é que as repetições são independentes.
- Vamos formalizar este conceito probabilístico a seguir mas ele significa que o resultado de uma repetição não afeta nenhuma outra.
- Isto também pode ser questionado.
- Alguns clientes podem influenciar outros via telefone ou comentários.
- Outro motivo é que se os clientes não forem todos distintos, as compras de um mesmo cliente podem ser muito semelhantes.
- Para pensar numa situação limite, imagine que apenas um único cliente que compra sempre a mesma cesta tenha sido observado.

Um olhar crítico

- Uma outra suposição é que as repetições são independentes.
- Vamos formalizar este conceito probabilístico a seguir mas ele significa que o resultado de uma repetição não afeta nenhuma outra.
- Isto também pode ser questionado.
- Alguns clientes podem influenciar outros via telefone ou comentários.
- Outro motivo é que se os clientes não forem todos distintos, as compras de um mesmo cliente podem ser muito semelhantes.
- Para pensar numa situação limite, imagine que apenas um único cliente que compra sempre a mesma cesta tenha sido observado.
- Estimar as probabilidades baseado nos dados deste único cliente não é uma boa idéia.

Um olhar crítico

- Outra suposição é que as repetições podem ser feitas indefinidadamente.
- Suponha que estejamos interessados em $\Omega = \{TGG, \widetilde{TGG}\}$ onde TGG significa a chance de uma terceira grande mundial nos próximos 5 anos e $\widetilde{TGG}$ a sua não-ocorrência.

Um olhar crítico

- Outra suposição é que as repetições podem ser feitas indefinidamente.
- Suponha que estejamos interessados em $\Omega = \{TGG, \widetilde{TGG}\}$ onde TGG significa a chance de uma terceira grande mundial nos próximos 5 anos e $\widetilde{TGG}$ a sua não-ocorrência.
- Não parece razoável querer estabelecer probabilidades invocando frequencias em repetições prolongadas nas mesmas condições destes eventos.

Um olhar crítico

- Outra suposição é que as repetições podem ser feitas indefinidamente.
- Suponha que estejamos interessados em $\Omega = \{TGG, \widetilde{TGG}\}$ onde TGG significa a chance de uma terceira grande mundial nos próximos 5 anos e $\widetilde{TGG}$ a sua não-ocorrência.
- Não parece razoável querer estabelecer probabilidades invocando frequencias em repetições prolongadas nas mesmas condições destes eventos.
- A abordagem bayesiana assume que probabilidades são subjetivas e podem ser manipuladas com as regras do cálculo de probabilidade (ver na disciplina PGM: Modelos Gráficos Probabilísticos).

Um olhar crítico

- Outra suposição é que as repetições podem ser feitas indefinidamente.
- Suponha que estejamos interessados em $\Omega = \{TGG, \widetilde{TGG}\}$ onde TGG significa a chance de uma terceira grande mundial nos próximos 5 anos e $\widetilde{TGG}$ a sua não-ocorrência.
- Não parece razoável querer estabelecer probabilidades invocando frequencias em repetições prolongadas nas mesmas condições destes eventos.
- A abordagem bayesiana assume que probabilidades são subjetivas e podem ser manipuladas com as regras do cálculo de probabilidade (ver na disciplina PGM: Modelos Gráficos Probabilísticos).
- Veremos ao longo do curso que existem várias maneiras de adaptar a versão básica da abordagem frequenstista para situações mais realistas, com as repetições não precisando ser nas mesmas condições e também com dependência entre elas.

Ω é infinito enumerável

- $\Omega = \{\omega_1, \omega_2, \ldots\}$: É idêntico ao caso finito: atribua $\mathbb{P}(\omega_i) \geq 0$ tal que some 1.

Ω é infinito enumerável

- $\Omega = \{\omega_1, \omega_2, \ldots\}$: É idêntico ao caso finito: atribua $\mathbb{P}(\omega_i) \geq 0$ tal que some 1.
- Para obter $\mathbb{P}(A)$ some os valores $\mathbb{P}(\omega)$ de todos os elementos $\omega \in A$.

Ω é infinito enumerável

- $\Omega = \{\omega_1, \omega_2, \ldots\}$: É idêntico ao caso finito: atribua $\mathbb{P}(\omega_i) \geq 0$ tal que some 1.
- Para obter $\mathbb{P}(A)$ some os valores $\mathbb{P}(\omega)$ de todos os elementos $\omega \in A$.
- Por exemplo, uma moeda honesta é lançada repetidamente até observarmos a primeira coroa $\tilde{c}$.
- $\Omega = \{\tilde{c}, c\tilde{c}, cc\tilde{c}, \ldots\}$

Ω é infinito enumerável

- $\Omega = \{\omega_1, \omega_2, \ldots\}$: É idêntico ao caso finito: atribua $\mathbb{P}(\omega_i) \geq 0$ tal que some 1.
- Para obter $\mathbb{P}(A)$ some os valores $\mathbb{P}(\omega)$ de todos os elementos $\omega \in A$.
- Por exemplo, uma moeda honesta é lançada repetidamente até observarmos a primeira coroa $\tilde{c}$.
- $\Omega = \{\tilde{c}, c\tilde{c}, cc\tilde{c}, \ldots\}$
- Atribuindo probabilidades de forma intuitiva (e correta)

$$\mathbb{P}(\omega_i) = \begin{cases} \mathbb{P}(\tilde{c}) = 1/2 \\ \\ \\ \end{cases}$$

Ω é infinito enumerável

- $\Omega = \{\omega_1, \omega_2, \ldots\}$: É idêntico ao caso finito: atribua $\mathbb{P}(\omega_i) \geq 0$ tal que some 1.
- Para obter $\mathbb{P}(A)$ some os valores $\mathbb{P}(\omega)$ de todos os elementos $\omega \in A$.
- Por exemplo, uma moeda honesta é lançada repetidamente até observarmos a primeira coroa $\tilde{c}$.
- $\Omega = \{\tilde{c}, c\tilde{c}, cc\tilde{c}, \ldots\}$
- Atribuindo probabilidades de forma intuitiva (e correta)

$$\mathbb{P}(\omega_i) = \begin{cases} \mathbb{P}(\tilde{c}) = 1/2 \\ \mathbb{P}(c\tilde{c}) = (1/2)(1/2) \\ \\ \\ \end{cases}$$

Ω é infinito enumerável

- $\Omega = \{\omega_1, \omega_2, \ldots\}$: É idêntico ao caso finito: atribua $\mathbb{P}(\omega_i) \geq 0$ tal que some 1.
- Para obter $\mathbb{P}(A)$ some os valores $\mathbb{P}(\omega)$ de todos os elementos $\omega \in A$.
- Por exemplo, uma moeda honesta é lançada repetidamente até observarmos a primeira coroa $\tilde{c}$.
- $\Omega = \{\tilde{c}, c\tilde{c}, cc\tilde{c}, \ldots\}$
- Atribuindo probabilidades de forma intuitiva (e correta)

$$\mathbb{P}(\omega_i) = \begin{cases} \mathbb{P}(\tilde{c}) = 1/2 \\ \mathbb{P}(c\tilde{c}) = (1/2)(1/2) \\ \mathbb{P}(cc\tilde{c}) = (1/2)(1/2)(1/2) \\ \vdots \end{cases}$$

- Temos

$$\sum_i \mathbb{P}(\omega_i) = \sum_i \mathbb{P}(\underbrace{cc\ldots c\tilde{c}}_{i \text{ terms}}) = \sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^i = 1$$

Ω é infinito enumerável

- Ainda no exemplo, seja A o evento em que a moeda é lançada um número par de vezes:

$$A = \{c\tilde{c}, ccc\tilde{c}, ccccc\tilde{c}, \ldots\}$$

Ω é infinito enumerável

- Ainda no exemplo, seja A o evento em que a moeda é lançada um número par de vezes:

$$A = \{c\tilde{c}, ccc\tilde{c}, ccccc\tilde{c}, \ldots\}$$

- Temos

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^{2i} = \sum_{i=1}^{\infty} \left(\frac{1}{4}\right)^{i} = \frac{1}{3}$$

Ω é infinito não-enumerável

- Os conjuntos não-enumeráveis... Os conjuntos infinitos *enumeráveis* são infinitinhos.

Ω é infinito não-enumerável

- Os conjuntos não-enumeráveis... Os conjuntos infinitos *enumeráveis* são infinitinhos.
- Os conjuntos infinitos não-enumeráveis são infinitões.

Ω é infinito não-enumerável

- Os conjuntos não-enumeráveis... Os conjuntos infinitos *enumeráveis* são infinitinhos.
- Os conjuntos infinitos não-enumeráveis são infinitões.
- Existem várias dificuldades para lidar rigorosamente com eles.

Ω é infinito não-enumerável

- Os conjuntos não-enumeráveis... Os conjuntos infinitos *enumeráveis* são infinitinhos.
- Os conjuntos infinitos não-enumeráveis são infinitões.
- Existem várias dificuldades para lidar rigorosamente com eles.
- Daremos apenas um exemplo para estes conjuntos.

Ω é infinito não-enumerável

- Os conjuntos não-enumeráveis... Os conjuntos infinitos *enumeráveis* são infinitinhos.
- Os conjuntos infinitos não-enumeráveis são infinitões.
- Existem várias dificuldades para lidar rigorosamente com eles.
- Daremos apenas um exemplo para estes conjuntos.
- Selecione um número completamente ao acaso no intervalo [0,1].
- $\Omega = [0,1]$. Como atribuir probabilidades?
- Vamos tentar o mesmo procedimento do caso em que Ω é finito ou enumerável.
- Isto é, atribua um valor $\mathbb{P}(\omega)$ para cada ω e defina

$$\mathbb{P}(A) = \sum_{\omega_i \in A} \mathbb{P}(\omega_i)$$

Não vai dar certo

- Como nenhum ponto é favorecido deveremos fazer $\mathbb{P}(\omega) = \xi > 0$ para todo $\omega \in \Omega$.

Não vai dar certo

- Como nenhum ponto é favorecido deveremos fazer $\mathbb{P}(\omega) = \xi > 0$ para todo $\omega \in \Omega$.
- e então $\mathbb{P}(A) = ??$

Não vai dar certo

- Como nenhum ponto é favorecido deveremos fazer $\mathbb{P}(\omega) = \xi > 0$ para todo $\omega \in \Omega$.
- e então $\mathbb{P}(A) = ??$
- Suponha que $A = \{1/2, 1/4, 1/8, 1/16, \ldots, ...\}$.

Não vai dar certo

- Como nenhum ponto é favorecido deveremos fazer $\mathbb{P}(\omega) = \xi > 0$ para todo $\omega \in \Omega$.
- e então $\mathbb{P}(A) = ??$
- Suponha que $A = \{1/2, 1/4, 1/8, 1/16, \ldots, ...\}$.

$$\mathbb{P}(A) = \sum_{\omega_i \in A} \mathbb{P}(\omega_i) = \sum_{\omega_i \in A} \xi = \xi.\infty = \infty$$

  se $\xi > 0$
- Portanto, algo está errado.

Todo ponto em $[0, 1]$ tem probabilidade 0

- O erro é assumir $\mathbb{P}(\omega) = \xi > 0$.

Todo ponto em $[0, 1]$ tem probabilidade 0

- O erro é assumir $\mathbb{P}(\omega) = \xi > 0$.
- O correto é assumir que $\mathbb{P}(\omega) = 0$ para todo ponto $\omega \in [0, 1]$.

Todo ponto em $[0, 1]$ tem probabilidade 0

- O erro é assumir $\mathbb{P}(\omega) = \xi > 0$.
- O correto é assumir que $\mathbb{P}(\omega) = 0$ para todo ponto $\omega \in [0, 1]$.
- Mas se todo número em $[0, 1]$ tem probabilidade ZERO, como poderemos ter $\mathbb{P}(A) > 0$?

Todo ponto em $[0, 1]$ tem probabilidade 0

- O erro é assumir $\mathbb{P}(\omega) = \xi > 0$.
- O correto é assumir que $\mathbb{P}(\omega) = 0$ para todo ponto $\omega \in [0, 1]$.
- Mas se todo número em $[0, 1]$ tem probabilidade ZERO, como poderemos ter $\mathbb{P}(A) > 0$?
- Este paradoxo sempre aparece quando representamos a realidade com os números da reta real.

Todo ponto em $[0, 1]$ tem probabilidade 0

- O erro é assumir $\mathbb{P}(\omega) = \xi > 0$.
- O correto é assumir que $\mathbb{P}(\omega) = 0$ para todo ponto $\omega \in [0, 1]$.
- Mas se todo número em $[0, 1]$ tem probabilidade ZERO, como poderemos ter $\mathbb{P}(A) > 0$?
- Este paradoxo sempre aparece quando representamos a realidade com os números da reta real.
- Por exemplo, em física, a representação da realidade com números reais gera paradoxos.

Mesmo paradoxo em física elementar

- Suponha que o intervalo $[0, 1]$ represente um segmento de fio com massa de 1 grama.

Mesmo paradoxo em física elementar

- Suponha que o intervalo $[0, 1]$ represente um segmento de fio com massa de 1 grama.
- Suponha que o fio tem sua massa perfeitamente e regularmente distribuída no fio.
- Dizemos que ele tem uma densidade de massa constante.

Mesmo paradoxo em física elementar

- Suponha que o intervalo $[0, 1]$ represente um segmento de fio com massa de 1 grama.
- Suponha que o fio tem sua massa perfeitamente e regularmente distribuída no fio.
- Dizemos que ele tem uma densidade de massa constante.
- Qual a massa de um ponto $x \in [0, 1]$?

Mesmo paradoxo em física elementar

- Suponha que o intervalo $[0, 1]$ represente um segmento de fio com massa de 1 grama.
- Suponha que o fio tem sua massa perfeitamente e regularmente distribuída no fio.
- Dizemos que ele tem uma densidade de massa constante.
- Qual a massa de um ponto $x \in [0, 1]$?
- Suponha que o ponto tem uma massa $\xi > 0$.

Mesmo paradoxo em física elementar

- Suponha que o intervalo $[0, 1]$ represente um segmento de fio com massa de 1 grama.
- Suponha que o fio tem sua massa perfeitamente e regularmente distribuída no fio.
- Dizemos que ele tem uma densidade de massa constante.
- Qual a massa de um ponto $x \in [0, 1]$?
- Suponha que o ponto tem uma massa $\xi > 0$.
- Como a densidade é constante, todos os pontos devem ter a mesma massa $\xi > 0$.

Mesmo paradoxo em física elementar

- Suponha que o intervalo $[0, 1]$ represente um segmento de fio com massa de 1 grama.
- Suponha que o fio tem sua massa perfeitamente e regularmente distribuída no fio.
- Dizemos que ele tem uma densidade de massa constante.
- Qual a massa de um ponto $x \in [0, 1]$?
- Suponha que o ponto tem uma massa $\xi > 0$.
- Como a densidade é constante, todos os pontos devem ter a mesma massa $\xi > 0$.
- Como existem infinitos pontos, a massa total deveria ser $\xi \times \infty = \infty$ e não 1 grama.

Solução no modelo físico

- O modelo que representa o fio por um segmento de reta é incorreto.
- O fio possui unidades atômicas que possuem massa.
- Sua representação como uma linha contínua leva a paradoxos.

Solução no modelo físico

- O modelo que representa o fio por um segmento de reta é incorreto.
- O fio possui unidades atômicas que possuem massa.
- Sua representação como uma linha contínua leva a paradoxos.
- A solução matemática para tornar a representação **útil** é assumir que:
    - Todo ponto isolado do fio possui massa ZERO.

Solução no modelo físico

- O modelo que representa o fio por um segmento de reta é incorreto.
- O fio possui unidades atômicas que possuem massa.
- Sua representação como uma linha contínua leva a paradoxos.
- A solução matemática para tornar a representação **útil** é assumir que:
  - Todo ponto isolado do fio possui massa ZERO.
  - A massa associada com um segmento $[a, b]$ é diretamente proporcional ao seu comprimento.

Solução no modelo físico

- O modelo que representa o fio por um segmento de reta é incorreto.
- O fio possui unidades atômicas que possuem massa.
- Sua representação como uma linha contínua leva a paradoxos.
- A solução matemática para tornar a representação **útil** é assumir que:
  - Todo ponto isolado do fio possui massa ZERO.
  - A massa associada com um segmento $[a, b]$ é diretamente proporcional ao seu comprimento.
  - Como a massa total de $[0, 1]$ é 1 grama, a massa de $[a, b] \in [0, 1]$ é $b - a$.

Solução no modelo físico

- O modelo que representa o fio por um segmento de reta é incorreto.
- O fio possui unidades atômicas que possuem massa.
- Sua representação como uma linha contínua leva a paradoxos.
- A solução matemática para tornar a representação **útil** é assumir que:
    - Todo ponto isolado do fio possui massa ZERO.
    - A massa associada com um segmento $[a, b]$ é diretamente proporcional ao seu comprimento.
    - Como a massa total de $[0, 1]$ é 1 grama, a massa de $[a, b] \in [0, 1]$ é $b - a$.
    - Por exemplo, $[0, 1/2]$ tem massa $1/2$ grama, $[1/2, 3/4]$ tem massa $1/4$ grama, etc.

Solução no modelo físico

- O modelo que representa o fio por um segmento de reta é incorreto.
- O fio possui unidades atômicas que possuem massa.
- Sua representação como uma linha contínua leva a paradoxos.
- A solução matemática para tornar a representação **útil** é assumir que:
  - Todo ponto isolado do fio possui massa ZERO.
  - A massa associada com um segmento $[a, b]$ é diretamente proporcional ao seu comprimento.
  - Como a massa total de $[0, 1]$ é 1 grama, a massa de $[a, b] \in [0, 1]$ é $b - a$.
  - Por exemplo, $[0, 1/2]$ tem massa $1/2$ grama, $[1/2, 3/4]$ tem massa $1/4$ grama, etc.
  - Note que o ponto x é também o intervalo $[x, x]$ que possui massa 0 pois tem comprimento 0.

A função densidade de massa

- Uma maneira um pouco mais complicada é usar uma função densidade de massa.
- Esta função será muito útil quando a massa não estiver distribuída de maneira uniforme.

A função densidade de massa

- Uma maneira um pouco mais complicada é usar uma função densidade de massa.
- Esta função será muito útil quando a massa não estiver distribuída de maneira uniforme.
- Densidade de massa: uma função $f(x)$ definida para cada x no segmento $[0, 1]$.

A função densidade de massa

- Uma maneira um pouco mais complicada é usar uma função densidade de massa.
- Esta função será muito útil quando a massa não estiver distribuída de maneira uniforme.
- Densidade de massa: uma função $f(x)$ definida para cada x no segmento $[0, 1]$.
- Esta função é tal que a massa no segmento $[a, b]$ é a sua integral:

$$\text{massa em } [a, b] = \int_a^b f(x) dx$$

A função densidade de massa

- Se tomarmos $f(x) = 1$ para todo $x \in [0, 1]$ teremos

$$\text{massa em } [a, b] = \int_a^b f(x)dx = \int_a^b 1dx = b - a$$

A função densidade de massa

- Se tomarmos $f(x) = 1$ para todo $x \in [0, 1]$ teremos

$$\text{massa em } [a, b] = \int_a^b f(x)dx = \int_a^b 1dx = b - a$$

- Esta é a função densidade $f(x)$ para o fio com massa uniformemente distribuída em $[0, 1]$.
- A idéia é espalhar a massa total do objeto por meio da função $f(x)$.
- A massa de qualquer subconjunto é obtida por integração.

A função densidade de massa

- A massa pode não ser uniformemente distribuída.
- Por exemplo, o material é uma liga com dois elementos (cobre e zinco).

A função densidade de massa

- A massa pode não ser uniformemente distribuída.
- Por exemplo, o material é uma liga com dois elementos (cobre e zinco).
- Em certas regiões, existe mais cobre que zinco. Em outras, o zinco domina.

A função densidade de massa

- A massa pode não ser uniformemente distribuída.
- Por exemplo, o material é uma liga com dois elementos (cobre e zinco).
- Em certas regiões, existe mais cobre que zinco. Em outras, o zinco domina.
- A densidade do fio vai variar de acordo com a proporção de zinco no local.

A função densidade de massa

- A massa pode não ser uniformemente distribuída.
- Por exemplo, o material é uma liga com dois elementos (cobre e zinco).
- Em certas regiões, existe mais cobre que zinco. Em outras, o zinco domina.
- A densidade do fio vai variar de acordo com a proporção de zinco no local.
- Ela pode estar mais concentrada em algumas regiões do fio que em outras.

A função densidade de massa

- A massa pode não ser uniformemente distribuída.
- Por exemplo, o material é uma liga com dois elementos (cobre e zinco).
- Em certas regiões, existe mais cobre que zinco. Em outras, o zinco domina.
- A densidade do fio vai variar de acordo com a proporção de zinco no local.
- Ela pode estar mais concentrada em algumas regiões do fio que em outras.
- Isto fica refletido imediatamente na função densidade $f(x)$.
- Nas regiões onde a massa é mais concentrada, f será maior.

Exemplos de densidade de massa

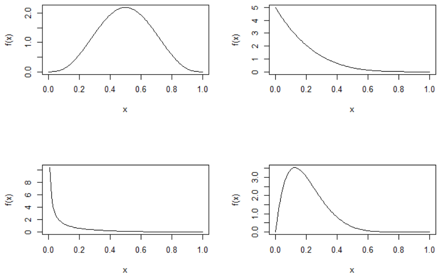


Figura: Quatro diferentes funções densidade $f(x)$.

Densidade de massa de PROBABILIDADE

- Com conjuntos Ω não-enumeráveis tais como $\Omega \subseteq \mathbb{R}^n$ adotamos o mesmo procedimento.

Densidade de massa de PROBABILIDADE

- Com conjuntos Ω não-enumeráveis tais como $\Omega \subseteq \mathbb{R}^n$ adotamos o mesmo procedimento.
- Massa total de probabilidade de Ω é 1 pois $\mathbb{P}(\Omega) = 1$.

Densidade de massa de PROBABILIDADE

- Com conjuntos Ω não-enumeráveis tais como $\Omega \subseteq \mathbb{R}^n$ adotamos o mesmo procedimento.
- Massa total de probabilidade de Ω é 1 pois $\mathbb{P}(\Omega) = 1$.
- Espalhe em Ω a massa total de probabilidade usando $f(x)$.

Densidade de massa de PROBABILIDADE

- Com conjuntos Ω não-enumeráveis tais como $\Omega \subseteq \mathbb{R}^n$ adotamos o mesmo procedimento.
- Massa total de probabilidade de Ω é 1 pois $\mathbb{P}(\Omega) = 1$.
- Espalhe em Ω a massa total de probabilidade usando $f(x)$.
- A massa de probabilidade de qualquer evento $A \subset \Omega$ é obtida por integração:

$$\mathbb{P}(A) = \int_a^b f(x) dx$$

Densidade de massa de PROBABILIDADE

- Escolhendo um ponto completamente ao acaso em $[0,1]$.
- Tome $f(x) = 1$ para $x \in [0,1]$.

Densidade de massa de PROBABILIDADE

- Escolhendo um ponto completamente ao acaso em $[0, 1]$.
- Tome $f(x) = 1$ para $x \in [0, 1]$.
- Evento: $A = [a, b]$, um intervalo.
- O experimento escolhe um único ponto em $[0, 1]$, e não intervalos.

Densidade de massa de PROBABILIDADE

- Escolhendo um ponto completamente ao acaso em $[0, 1]$.
- Tome $f(x) = 1$ para $x \in [0, 1]$.
- Evento: $A = [a, b]$, um intervalo.
- O experimento escolhe um único ponto em $[0, 1]$, e não intervalos.
- Ocorrer o evento A significa que o ponto escolhido pertence ao intervalo $A = [a, b]$.

Densidade de massa de PROBABILIDADE

- Escolhendo um ponto completamente ao acaso em $[0, 1]$.
- Tome $f(x) = 1$ para $x \in [0, 1]$.
- Evento: $A = [a, b]$, um intervalo.
- O experimento escolhe um único ponto em $[0, 1]$, e não intervalos.
- Ocorrer o evento A significa que o ponto escolhido pertence ao intervalo $A = [a, b]$.

$$\mathbb{P}(\text{Intervalo[a,b]}) = \text{Comprimento do intervalo [a,b]} \tag{1}$$

Densidade de massa de PROBABILIDADE

- Escolhendo um ponto completamente ao acaso em $[0,1]$.
- Tome $f(x) = 1$ para $x \in [0,1]$.
- Evento: $A = [a,b]$, um intervalo.
- O experimento escolhe um único ponto em $[0,1]$, e não intervalos.
- Ocorrer o evento A significa que o ponto escolhido pertence ao intervalo $A = [a,b]$.

$$\mathbb{P}(\text{Intervalo[a,b]}) = \text{Comprimento do intervalo [a,b]} \tag{1}$$

- Todo ponto isolado do fio possui probabilidade ZERO.
- (Opcional:) Em $\mathcal{A}$, todo evento pode ser aproximado com $\cup$, $\cap$, e $^{\mathrm{C}}$ de intervalos (número enumerável). Assim $\mathbb{P}(A)$ fica estabelecido $\forall$ A (Teorema de extensão de Caratheodory)
- I jump to fourth slide of next frame

Quando Ω é complicado - OPCIONAL

- Definir densidades para Ω complicados pode ser difícil.

Quando Ω é complicado - OPCIONAL

- Definir densidades para Ω complicados pode ser difícil.
- Pior: pode ser impossível!

Quando Ω é complicado - OPCIONAL

- Definir densidades para Ω complicados pode ser difícil.
- Pior: pode ser impossível!
- Não sabemos explicitar uma densidade.

Quando Ω é complicado - OPCIONAL

- Definir densidades para Ω complicados pode ser difícil.
- Pior: pode ser impossível!
- Não sabemos explicitar uma densidade.
- Novamente, os conjuntos INFINITÕES vêm nos assombrar.

Quando Ω é complicado - OPCIONAL

- Definir densidades para Ω complicados pode ser difícil.
- Pior: pode ser impossível!
- Não sabemos explicitar uma densidade.
- Novamente, os conjuntos INFINITÕES vêm nos assombrar.
- É sempre a dificuldade de lidar matematicamente com o infinito "excessivo".

Quando Ω é complicado - OPCIONAL

- Definir densidades para Ω complicados pode ser difícil.
- Pior: pode ser impossível!
- Não sabemos explicitar uma densidade.
- Novamente, os conjuntos INFINITÕES vêm nos assombrar.
- É sempre a dificuldade de lidar matematicamente com o infinito "excessivo".
- Existem situações práticas que exigem trabalhar com estes conjuntos Ω e temos de solucionar isto de alguma forma.

Quando Ω é complicado - OPCIONAL

- Por exemplo, no caso do movimento browniano: movimento errático de um grão de pólen na superfície da água observado a cada 1 segundo.

Quando Ω é complicado - OPCIONAL

- Por exemplo, no caso do movimento browniano: movimento errático de um grão de pólen na superfície da água observado a cada 1 segundo.

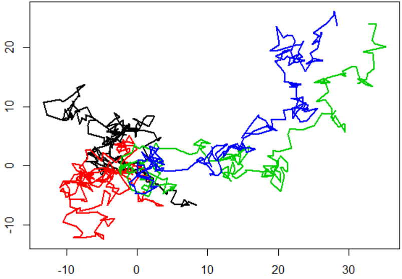


Figura: Quatro diferentes elementos $\omega \in \Omega$.

Quando Ω é complicado - OPCIONAL

- Ω é o conjunto de todas as curvas erráticas do movimento browniano.
- Como definir eventos (sub-conjuntos de Ω) aqui?

Quando Ω é complicado - OPCIONAL

- Ω é o conjunto de todas as curvas erráticas do movimento browniano.
- Como definir eventos (sub-conjuntos de Ω) aqui?
- Queremos calcular, por exemplo, a probabilidade de a trajetória da partícula não se intersecte nos primeiros 10 minutos.

Quando Ω é complicado - OPCIONAL

- Ω é o conjunto de todas as curvas erráticas do movimento browniano.
- Como definir eventos (sub-conjuntos de Ω) aqui?
- Queremos calcular, por exemplo, a probabilidade de a trajetória da partícula não se intersecte nos primeiros 10 minutos.
- Este evento corresponde a uma imenso conjunto de curvas de Ω.

Quando Ω é complicado - OPCIONAL

- Ω é o conjunto de todas as curvas erráticas do movimento browniano.
- Como definir eventos (sub-conjuntos de Ω) aqui?
- Queremos calcular, por exemplo, a probabilidade de a trajetória da partícula não se intersecte nos primeiros 10 minutos.
- Este evento corresponde a uma imenso conjunto de curvas de Ω.
- Qual a probabilidade de sua ocorrência?

Quando Ω é complicado - OPCIONAL

- Ω é o conjunto de todas as curvas erráticas do movimento browniano.
- Como definir eventos (sub-conjuntos de Ω) aqui?
- Queremos calcular, por exemplo, a probabilidade de a trajetória da partícula não se intersecte nos primeiros 10 minutos.
- Este evento corresponde a uma imenso conjunto de curvas de Ω.
- Qual a probabilidade de sua ocorrência?
- Como atribuir probabilidades de forma consistente a TODOS estes eventos?

Quando Ω é complicado - OPCIONAL

- Ω é o conjunto de todas as curvas erráticas do movimento browniano.
- Como definir eventos (sub-conjuntos de Ω) aqui?
- Queremos calcular, por exemplo, a probabilidade de a trajetória da partícula não se intersecte nos primeiros 10 minutos.
- Este evento corresponde a uma imenso conjunto de curvas de Ω.
- Qual a probabilidade de sua ocorrência?
- Como atribuir probabilidades de forma consistente a TODOS estes eventos?

Quando Ω é complicado - OPCIONAL

- O lançamento de uma moeda honesta indefinidamente tem $\Omega = \{0, 1\}^{\infty}$.

Quando Ω é complicado - OPCIONAL

- O lançamento de uma moeda honesta indefinidamente tem $\Omega = \{0, 1\}^{\infty}$.
- Como definir probabilidades de forma consistente para TODOS os eventos?

Quando Ω é complicado - OPCIONAL

- O lançamento de uma moeda honesta indefinidamente tem $\Omega = \{0, 1\}^\infty$.
- Como definir probabilidades de forma consistente para TODOS os eventos?
- Eventos devem levar em conta os infinitos lançamentos.

Quando Ω é complicado - OPCIONAL

- O lançamento de uma moeda honesta indefinidamente tem $\Omega = \{0, 1\}^{\infty}$.
- Como definir probabilidades de forma consistente para TODOS os eventos?
- Eventos devem levar em conta os infinitos lançamentos.
- Por exempo:
  - Seja f_n a proporção de 1's nos primeiros n lançamentos.

Quando Ω é complicado - OPCIONAL

- O lançamento de uma moeda honesta indefinidamente tem $\Omega = \{0, 1\}^{\infty}$.
- Como definir probabilidades de forma consistente para TODOS os eventos?
- Eventos devem levar em conta os infinitos lançamentos.
- Por exempo:
    - Seja f_n a proporção de 1's nos primeiros n lançamentos.
    - Monitore f_n ao longo de uma sequência $\omega \in \Omega$ tal como $(0, 1, 0, 0, 0, 1, 0, \ldots)$.
    - O que acontece com f_n quando n cresce?

Quando Ω é complicado - OPCIONAL

- O lançamento de uma moeda honesta indefinidamente tem $\Omega = \{0, 1\}^\infty$.
- Como definir probabilidades de forma consistente para TODOS os eventos?
- Eventos devem levar em conta os infinitos lançamentos.
- Por exempo:
    - Seja f_n a proporção de 1's nos primeiros n lançamentos.
    - Monitore f_n ao longo de uma sequência $\omega \in \Omega$ tal como $(0, 1, 0, 0, 0, 1, 0, \ldots)$.
    - O que acontece com f_n quando n cresce?
    - Com base na nossa experiência, esperamos ver $f_n \to 1/2$.
    - Mas isto acontece com certeza? Com probabilidade 1?

Quando Ω é complicado - OPCIONAL

- O lançamento de uma moeda honesta indefinidamente tem $\Omega = \{0, 1\}^\infty$.
- Como definir probabilidades de forma consistente para TODOS os eventos?
- Eventos devem levar em conta os infinitos lançamentos.
- Por exempo:
  - Seja f_n a proporção de 1's nos primeiros n lançamentos.
  - Monitore f_n ao longo de uma sequência $\omega \in \Omega$ tal como $(0, 1, 0, 0, 0, 1, 0, \ldots)$.
  - O que acontece com f_n quando n cresce?
  - Com base na nossa experiência, esperamos ver $f_n \to 1/2$.
  - Mas isto acontece com certeza? Com probabilidade 1?
  - Ou existe alguma chance, por mínima que seja, de que f_n não convirja para $1/2$?

Quando Ω é complicado - OPCIONAL

- O lançamento de uma moeda honesta indefinidamente tem $\Omega = \{0,1\}^{\infty}$.
- Como definir probabilidades de forma consistente para TODOS os eventos?
- Eventos devem levar em conta os infinitos lançamentos.
- Por exempo:
  - Seja f_n a proporção de 1's nos primeiros n lançamentos.
  - Monitore f_n ao longo de uma sequência $\omega \in \Omega$ tal como $(0,1,0,0,0,1,0,\ldots)$.
  - O que acontece com f_n quando n cresce?
  - Com base na nossa experiência, esperamos ver $f_n \to 1/2$.
  - Mas isto acontece com certeza? Com probabilidade 1?
  - Ou existe alguma chance, por mínima que seja, de que f_n não convirja para $1/2$?

Quando Ω é complicado - OPCIONAL

- Ou, quem sabe, f_n não convirja para valor algum, oscilando no intervalo $(0, 1)$ sem estabilizar-se permanentemente em torno de nenhum valor.

Quando Ω é complicado - OPCIONAL

- Ou, quem sabe, f_n não convirja para valor algum, oscilando no intervalo $(0, 1)$ sem estabilizar-se permanentemente em torno de nenhum valor.
- Afinal, podemos pensar em muitas (infinitas!) sequências $\omega \in \Omega$ tais que $f_n \nrightarrow 1/2$.

Quando Ω é complicado - OPCIONAL

- Ou, quem sabe, f_n não convirja para valor algum, oscilando no intervalo $(0, 1)$ sem estabilizar-se permanentemente em torno de nenhum valor.
- Afinal, podemos pensar em muitas (infinitas!) sequências $\omega \in \Omega$ tais que $f_n \nrightarrow 1/2$.
- Por exemplo, $\omega = (0, 1, 0, 1, 1, 1, 1, 1, 1, 1, \ldots)$

Quando Ω é complicado - OPCIONAL

- Ou, quem sabe, f_n não convirja para valor algum, oscilando no intervalo $(0,1)$ sem estabilizar-se permanentemente em torno de nenhum valor.
- Afinal, podemos pensar em muitas (infinitas!) sequências $\omega \in \Omega$ tais que $f_n \nrightarrow 1/2$.
- Por exemplo, $\omega = (0,1,0,1,1,1,1,1,1,1,\ldots)$
- Ou $\omega = (1,0,1,0,0,1,0,0,0,1,0,0,0,0,1,\ldots)$

Quando Ω é complicado - OPCIONAL

- Ou, quem sabe, f_n não convirja para valor algum, oscilando no intervalo $(0, 1)$ sem estabilizar-se permanentemente em torno de nenhum valor.
- Afinal, podemos pensar em muitas (infinitas!) sequências $\omega \in \Omega$ tais que $f_n \nrightarrow 1/2$.
- Por exemplo, $\omega = (0, 1, 0, 1, 1, 1, 1, 1, 1, 1, \ldots)$
- Ou $\omega = (1, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, \ldots)$
- Qual a probabilidade de ocorra uma dessas infnitas seqências com $f_n \nrightarrow 1/2$?
- A resposta é ...

Quando Ω é complicado - OPCIONAL

- Ou, quem sabe, f_n não convirja para valor algum, oscilando no intervalo $(0, 1)$ sem estabilizar-se permanentemente em torno de nenhum valor.
- Afinal, podemos pensar em muitas (infinitas!) sequências $\omega \in \Omega$ tais que $f_n \nrightarrow 1/2$.
- Por exemplo, $\omega = (0, 1, 0, 1, 1, 1, 1, 1, 1, 1, \ldots)$
- Ou $\omega = (1, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, \ldots)$
- Qual a probabilidade de ocorra uma dessas infnitas seqências com $f_n \nrightarrow 1/2$?
- A resposta é ... a probabilidade é igual a ZERO (teorema Lei Forte dos Grandes Números)

Quando Ω é complicado - OPCIONAL

- Ou, quem sabe, f_n não convirja para valor algum, oscilando no intervalo $(0, 1)$ sem estabilizar-se permanentemente em torno de nenhum valor.
- Afinal, podemos pensar em muitas (infinitas!) sequências $\omega \in \Omega$ tais que $f_n \nrightarrow 1/2$.
- Por exemplo, $\omega = (0, 1, 0, 1, 1, 1, 1, 1, 1, 1, \ldots)$
- Ou $\omega = (1, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, \ldots)$
- Qual a probabilidade de ocorra uma dessas infnitas seqências com $f_n \nrightarrow 1/2$?
- A resposta é ... a probabilidade é igual a ZERO (teorema Lei Forte dos Grandes Números)
- Numa sequência infinita de lançamentos da moeda honesta a probabilidade de que f_n não convirja para $1/2$ é zero.

Quando Ω é complicado - OPCIONAL

- Ou, quem sabe, f_n não convirja para valor algum, oscilando no intervalo $(0, 1)$ sem estabilizar-se permanentemente em torno de nenhum valor.
- Afinal, podemos pensar em muitas (infinitas!) sequências $\omega \in \Omega$ tais que $f_n \nrightarrow 1/2$.
- Por exemplo, $\omega = (0, 1, 0, 1, 1, 1, 1, 1, 1, 1, \ldots)$
- Ou $\omega = (1, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, \ldots)$
- Qual a probabilidade de ocorra uma dessas infnitas seqências com $f_n \nrightarrow 1/2$?
- A resposta é ... a probabilidade é igual a ZERO (teorema Lei Forte dos Grandes Números)
- Numa sequência infinita de lançamentos da moeda honesta a probabilidade de que f_n não convirja para $1/2$ é zero.
- Mas existem infinitas sequências desse tipo não-convergente em Ω...

Quando Ω é complicado - OPCIONAL

- E se a moeda tiver a probabilidade de cara bem pequena, digamos $\theta \approx 0$.

Quando Ω é complicado - OPCIONAL

- E se a moeda tiver a probabilidade de cara bem pequena, digamos $\theta \approx 0$.
- Qual a probabilidade de que f_n não convirja para este θ bem próximo de zero?

Quando Ω é complicado - OPCIONAL

- E se a moeda tiver a probabilidade de cara bem pequena, digamos $\theta \approx 0$.
- Qual a probabilidade de que f_n não convirja para este θ bem próximo de zero?
- Zero.

Quando Ω é complicado - OPCIONAL

- E se a moeda tiver a probabilidade de cara bem pequena, digamos $\theta \approx 0$.
- Qual a probabilidade de que f_n não convirja para este θ bem próximo de zero?
- Zero.
- Está divertido... vamos ver um outro evento...

Quando Ω é complicado - OPCIONAL

- E se a moeda tiver a probabilidade de cara bem pequena, digamos $\theta \approx 0$.
- Qual a probabilidade de que f_n não convirja para este θ bem próximo de zero?
- Zero.
- Está divertido... vamos ver um outro evento...
- Pegue o comprimento da mais longa sequência de 1's ininterruptos na série infinita de lançamentos da moeda honesta.

Quando Ω é complicado - OPCIONAL

- E se a moeda tiver a probabilidade de cara bem pequena, digamos $\theta \approx 0$.
- Qual a probabilidade de que f_n não convirja para este θ bem próximo de zero?
- Zero.
- Está divertido... vamos ver um outro evento...
- Pegue o comprimento da mais longa sequência de 1's ininterruptos na série infinita de lançamentos da moeda honesta.
- Qual a probabilidade de que este comprimento seja pelo menos 2000?

Quando Ω é complicado - OPCIONAL

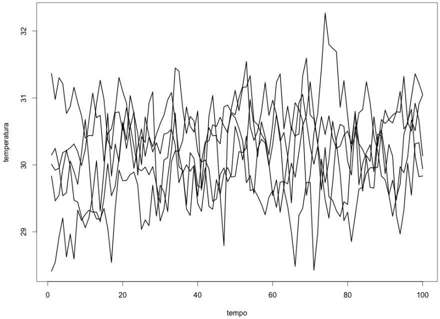
- E se a moeda tiver a probabilidade de cara bem pequena, digamos $\theta \approx 0$.
- Qual a probabilidade de que f_n não convirja para este θ bem próximo de zero?
- Zero.
- Está divertido... vamos ver um outro evento...
- Pegue o comprimento da mais longa sequência de 1's ininterruptos na série infinita de lançamentos da moeda honesta.
- Qual a probabilidade de que este comprimento seja pelo menos 2000?
- Curioso?

Quando Ω é complicado - OPCIONAL

- E se a moeda tiver a probabilidade de cara bem pequena, digamos $\theta \approx 0$.
- Qual a probabilidade de que f_n não convirja para este θ bem próximo de zero?
- Zero.
- Está divertido... vamos ver um outro evento...
- Pegue o comprimento da mais longa sequência de 1's ininterruptos na série infinita de lançamentos da moeda honesta.
- Qual a probabilidade de que este comprimento seja pelo menos 2000?
- Curioso?
- A probabilidade é igual a 1... (uma aplicação do Teorema de Borel-Cantelli).

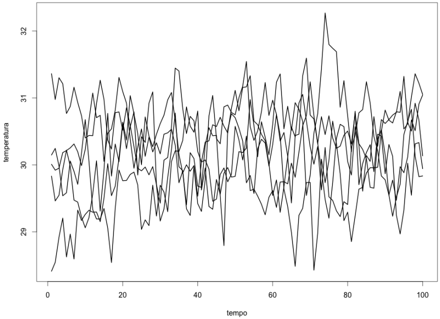
Quando Ω é complicado - OPCIONAL

- Se Ω é um conjunto formado por todas as funções contínuas:



Quando Ω é complicado - OPCIONAL

- Se Ω é um conjunto formado por todas as funções contínuas:



- O experimento é observar uma curva contínua de temperatura durante 24 horas.

Quando Ω é complicado - OPCIONAL

- ω é uma das infinitas curvas possíveis.
- Eventos são sub-conjuntos de curvas deste conjunto Ω.

Quando Ω é complicado - OPCIONAL

- ω é uma das infinitas curvas possíveis.
- Eventos são sub-conjuntos de curvas deste conjunto Ω.
- Como atribuir probabilidades de forma consistente a TODOS os eventos possíveis?

Quando Ω é complicado - OPCIONAL

- ω é uma das infinitas curvas possíveis.
- Eventos são sub-conjuntos de curvas deste conjunto Ω.
- Como atribuir probabilidades de forma consistente a TODOS os eventos possíveis?
- Por exemplo, se A e B são dois eventos (dois conjuntos de curvas) tais que $A \subset B$ então devemos ter $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Quando Ω é complicado - OPCIONAL

- ω é uma das infinitas curvas possíveis.
- Eventos são sub-conjuntos de curvas deste conjunto Ω.
- Como atribuir probabilidades de forma consistente a TODOS os eventos possíveis?
- Por exemplo, se A e B são dois eventos (dois conjuntos de curvas) tais que $A \subset B$ então devemos ter $\mathbb{P}(A) \leq \mathbb{P}(B)$.
- O que poderia ser uma densidade de probabilidade neste conjunto Ω de curvas contínuas?

Quando Ω é complicado - OPCIONAL

- ω é uma das infinitas curvas possíveis.
- Eventos são sub-conjuntos de curvas deste conjunto Ω.
- Como atribuir probabilidades de forma consistente a TODOS os eventos possíveis?
- Por exemplo, se A e B são dois eventos (dois conjuntos de curvas) tais que $A \subset B$ então devemos ter $\mathbb{P}(A) \leq \mathbb{P}(B)$.
- O que poderia ser uma densidade de probabilidade neste conjunto Ω de curvas contínuas?
- Como integrar neste conjunto? Precisamos de uma noção de integral mais complexa que a integral de Riemann.

Quando Ω é complicado - OPCIONAL

- ω é uma das infinitas curvas possíveis.
- Eventos são sub-conjuntos de curvas deste conjunto Ω.
- Como atribuir probabilidades de forma consistente a TODOS os eventos possíveis?
- Por exemplo, se A e B são dois eventos (dois conjuntos de curvas) tais que $A \subset B$ então devemos ter $\mathbb{P}(A) \leq \mathbb{P}(B)$.
- O que poderia ser uma densidade de probabilidade neste conjunto Ω de curvas contínuas?
- Como integrar neste conjunto? Precisamos de uma noção de integral mais complexa que a integral de Riemann.
- Isto é assunto de cursos avançados de probabilidade e processos estocásticos.

Quando Ω é complicado - OPCIONAL

- Vamos evitar TODAS estas complicações.

Quando Ω é complicado - OPCIONAL

- Vamos evitar TODAS estas complicações.
- Na prática da análise de dados NÃO trabalhamos diretamente com Ω.

Quando Ω é complicado - OPCIONAL

- Vamos evitar TODAS estas complicações.
- Na prática da análise de dados NÃO trabalhamos diretamente com Ω.
- Reduzimos o fenêmeno estocástico a algumas poucas características numéricas com as quais descrevemos o experimento aleatório.

Quando Ω é complicado - OPCIONAL

- Vamos evitar TODAS estas complicações.
- Na prática da análise de dados NÃO trabalhamos diretamente com Ω.
- Reduzimos o fenêmeno estocástico a algumas poucas características numéricas com as quais descrevemos o experimento aleatório.
- Estas características são chamadas *variáveis aleatórias*.

Quando Ω é complicado - OPCIONAL

- Vamos evitar TODAS estas complicações.
- Na prática da análise de dados NÃO trabalhamos diretamente com Ω.
- Reduzimos o fenêmeno estocástico a algumas poucas características numéricas com as quais descrevemos o experimento aleatório.
- Estas características são chamadas *variáveis aleatórias*.
- Na prática isto vai significar que, no "pior caso", teremos Ω equivalente a subconjuntos de $\mathbb{R}^n$, para os quais podemos definir densidades de probabilidade.

Densidade de probabilidade quando Ω é contínuo

- Variáveis aleatórias, na prática, fazem com que $\Omega \subset \mathbb{R}^n$

Densidade de probabilidade quando Ω é contínuo

- Variáveis aleatórias, na prática, fazem com que $\Omega \subset \mathbb{R}^n$
- E este caso é muito fácil.

Densidade de probabilidade quando Ω é contínuo

- Variáveis aleatórias, na prática, fazem com que $\Omega \subset \mathbb{R}^n$
- E este caso é muito fácil.
- A densidade $f(\omega)$ pode ser QUALQUER função

$$
\begin{aligned}
f : \Omega \subset \mathbb{R}^n &\longrightarrow \mathbb{R} \\
\omega &\longrightarrow f(\omega)
\end{aligned}
$$

tal que:

- $f(\omega) \geq 0$, (para não obter uma probabilidade negativa).

Densidade de probabilidade quando Ω é contínuo

- Variáveis aleatórias, na prática, fazem com que $\Omega \subset \mathbb{R}^n$
- E este caso é muito fácil.
- A densidade $f(\omega)$ pode ser QUALQUER função

$$
\begin{array}{rcl}
f : \Omega \subset \mathbb{R}^n & \longrightarrow & \mathbb{R} \\
\omega & \longrightarrow & f(\omega)
\end{array}
$$

tal que:

- $f(\omega) \geq 0$, (para não obter uma probabilidade negativa).
- $\int_\Omega f(\omega)\, d\omega = 1$

$\Omega = [0, 1]$

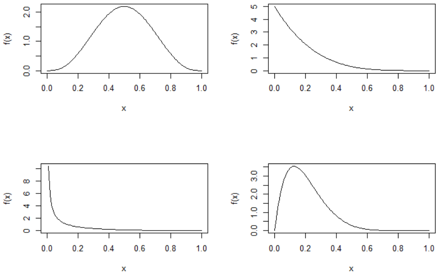


Figura: Quatro diferentes funções densidade de massa de probabilidade $f(x)$.

Observações

- Devemos ter $f(\omega) \geq 0$, um limite inferior.

Observações

- Devemos ter $f(\omega) \geq 0$, um limite inferior.
- Mas podemos ter $f(\omega) > 1$: não há um limite superior.

Observações

- Devemos ter $f(\omega) \geq 0$, um limite inferior.
- Mas podemos ter $f(\omega) > 1$: não há um limite superior.
- A restrição fundamental é que a integral sobre todo Ω deve ser 1.

Observações

- Devemos ter $f(\omega) \geq 0$, um limite inferior.
- Mas podemos ter $f(\omega) > 1$: não há um limite superior.
- A restrição fundamental é que a integral sobre todo Ω deve ser 1.
- Não se exige que o valor $f(\omega)$ em cada $\omega \in \Omega$ seja menor que 1.

Observações

- Devemos ter $f(\omega) \geq 0$, um limite inferior.
- Mas podemos ter $f(\omega) > 1$: não há um limite superior.
- A restrição fundamental é que a integral sobre todo Ω deve ser 1.
- Não se exige que o valor $f(\omega)$ em cada $\omega \in \Omega$ seja menor que 1.
- Para obter a probabilidade de um evento $A \subset \Omega$ basta integrar $f(x)$ sobre a região A:

$$\mathbb{P}(A) = \int_A f(x)dx$$

- Assim, uma proabilidade $\mathbb{P}(A)$ é área (ou volume) sob uma curva (ou superfície) de densidade.

Exemplo

- Suponha que dardos são atirados num alvo circular de raio 1.

Exemplo

- Suponha que dardos são atirados num alvo circular de raio 1.
- Um jogador possui uma habilidadefaz com que a chance de acertar numa região A próxima do centro é maior que se esta mesma região estiver próxima da borda.

Exemplo

- Suponha que dardos são atirados num alvo circular de raio 1.
- Um jogador possui uma habilidadefaz com que a chance de acertar numa região A próxima do centro é maior que se esta mesma região estiver próxima da borda.
- Esta habilidade está representada pela densidade

$$f(x, y) = c \ \left(\sqrt{x^2 + y^2} - 1\right)^2$$

para x, y no disco unitário

Exemplo

- Suponha que dardos são atirados num alvo circular de raio 1.
- Um jogador possui uma habilidadefaz com que a chance de acertar numa região A próxima do centro é maior que se esta mesma região estiver próxima da borda.
- Esta habilidade está representada pela densidade

$$f(x,y) = c \ \left(\sqrt{x^2+y^2} - 1\right)^2$$

para x, y no disco unitário

- c é uma constante para garantir que $\int_\Omega f(x,y)dxdy = 1$.
- OBS: Encontra-se que $c = 14\pi/12$

Exemplo

- Seja $r = \sqrt{x^2 + y^2}$, a distância de (x, y) até a origem.

Exemplo

- Seja $r = \sqrt{x^2 + y^2}$, a distância de (x, y) até a origem.
- Então podemos reescrever a densidade anterior

$$f(x, y) = c \left(\sqrt{x^2 + y^2} - 1\right)^2 = c(r - 1)^2$$

Exemplo

- Seja $r = \sqrt{x^2 + y^2}$, a distância de (x, y) até a origem.
- Então podemos reescrever a densidade anterior

$$f(x,y) = c \left(\sqrt{x^2+y^2} - 1\right)^2 = c(r-1)^2$$

- Isto torna mais simples a visualização da densidade: mapa de calor ou curvas de nível.

Exemplo

- Seja $r = \sqrt{x^2 + y^2}$, a distância de (x, y) até a origem.
- Então podemos reescrever a densidade anterior

$$f(x, y) = c \left(\sqrt{x^2 + y^2} - 1\right)^2 = c(r - 1)^2$$

- Isto torna mais simples a visualização da densidade: mapa de calor ou curvas de nível.
- Para uma região A qualquer dentro do disco, temos

$$\mathbb{P}(A) = \int_A f(x, y) dxdy$$

Exemplo

- Indivíduos com habilidades diferentes terão sua densidade diferente.

Exemplo

- Indivíduos com habilidades diferentes terão sua densidade diferente.
- A densidade deverá expressar quais as regiões mais prováveis de serem atingidas.

Exemplo

- Indivíduos com habilidades diferentes terão sua densidade diferente.
- A densidade deverá expressar quais as regiões mais prováveis de serem atingidas.
- Como seria um mapa de calor da densidade $f(x, y)$ de um jogador "cego´´?

Exemplo

- Indivíduos com habilidades diferentes terão sua densidade diferente.
- A densidade deverá expressar quais as regiões mais prováveis de serem atingidas.
- Como seria um mapa de calor da densidade $f(x, y)$ de um jogador "cego´´?
- E um jogador extremamente habilidoso?

Exemplo

- Indivíduos com habilidades diferentes terão sua densidade diferente.
- A densidade deverá expressar quais as regiões mais prováveis de serem atingidas.
- Como seria um mapa de calor da densidade $f(x, y)$ de um jogador "cego´´?
- E um jogador extremamente habilidoso?
- E um jogador que tem um viés par a direita, que tende a jogar o dardo deslocado para a direita?

Exemplo

- Interesse no tempo de espera pelo primeiro comentário após a postagem de um vídeo do YouTube do canal de Whindersson Nunes.
- Espaço amostral Ω?

Exemplo

- Interesse no tempo de espera pelo primeiro comentário após a postagem de um vídeo do YouTube do canal de Whindersson Nunes.
- Espaço amostral Ω?
- $\Omega = (0, \infty)$

Exemplo

- Interesse no tempo de espera pelo primeiro comentário após a postagem de um vídeo do YouTube do canal de Whindersson Nunes.
- Espaço amostral Ω?
- $\Omega = (0, \infty)$
- Densidade $f(x)$?

Exemplo

- Interesse no tempo de espera pelo primeiro comentário após a postagem de um vídeo do YouTube do canal de Whindersson Nunes.
- Espaço amostral Ω?
- $\Omega = (0, \infty)$
- Densidade $f(x)$?
- Várias alternativas para $f(x)$ - ver no gráfico a seguir.

Esperando comentário

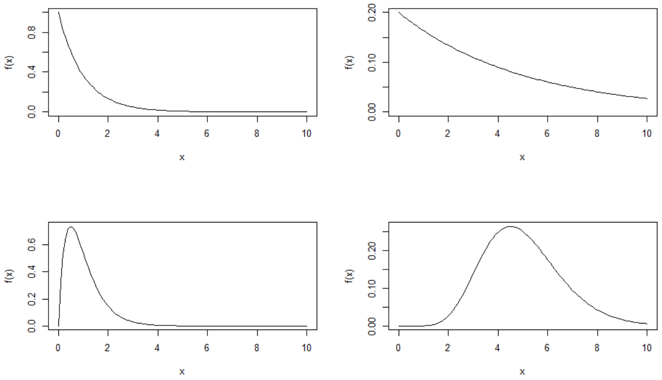


Figura: Quatro possíveis modelos de densidade de probabilidade $f(x)$ para tempo de espera. VISUALMENTE¡ veja como são diferentes as probabilidades $\mathbb{P}(A) = \mathbb{P}((2, 4))$.

Probabilidade Condicional

- Seja B um evento em Ω e $\mathbb{P}(B)$ sua probabilidade de ocorrência.

Probabilidade Condicional

- Seja B um evento em Ω e $\mathbb{P}(B)$ sua probabilidade de ocorrência.
- Sem poder ver o resultado do experimento diretamente, somos informados apenas que outro evento A ocorreu.

Probabilidade Condicional

- Seja B um evento em Ω e $\mathbb{P}(B)$ sua probabilidade de ocorrência.
- Sem poder ver o resultado do experimento diretamente, somos informados apenas que outro evento A ocorreu.
- Isto muda a probabilidade de B ocorrer?

Probabilidade Condicional

- Seja B um evento em Ω e $\mathbb{P}(B)$ sua probabilidade de ocorrência.
- Sem poder ver o resultado do experimento diretamente, somos informados apenas que outro evento A ocorreu.
- Isto muda a probabilidade de B ocorrer?
- Por exemplo, dois dados bem equilibrados são lançados em sequência.

Probabilidade Condicional

- Seja B um evento em Ω e $\mathbb{P}(B)$ sua probabilidade de ocorrência.
- Sem poder ver o resultado do experimento diretamente, somos informados apenas que outro evento A ocorreu.
- Isto muda a probabilidade de B ocorrer?
- Por exemplo, dois dados bem equilibrados são lançados em sequência.
- Você aposta na ocorrência de B: o primeiro dado vai resultar num 6.

Probabilidade Condicional

- Seja B um evento em Ω e $\mathbb{P}(B)$ sua probabilidade de ocorrência.
- Sem poder ver o resultado do experimento diretamente, somos informados apenas que outro evento A ocorreu.
- Isto muda a probabilidade de B ocorrer?
- Por exemplo, dois dados bem equilibrados são lançados em sequência.
- Você aposta na ocorrência de B: o primeiro dado vai resultar num 6.
- Se você souber que a soma dos dois dados foi menor que 8 (evento A) e pudesse rever sua aposta, você colocaria mais fichas na ocorrência de B? Ou menos fichas?

Probabilidade Condicional

- De posse da informação de que certo evento A ocorreu, queremos recalcular as chances de outros eventos $B_1, B_2, \ldots$

Probabilidade Condicional

- De posse da informação de que certo evento A ocorreu, queremos recalcular as chances de outros eventos $B_1, B_2, \ldots$
- Chamamos a isto de probabilidade de B condicionada à ocorrência do evento A,

Probabilidade Condicional

- De posse da informação de que certo evento A ocorreu, queremos recalcular as chances de outros eventos $B_1, B_2, \ldots$
- Chamamos a isto de probabilidade de B condicionada à ocorrência do evento A,
- ou de probabilidade de B dado que A ocorreu,

Probabilidade Condicional

- De posse da informação de que certo evento A ocorreu, queremos recalcular as chances de outros eventos $B_1, B_2, \ldots$
- Chamamos a isto de probabilidade de B condicionada à ocorrência do evento A,
- ou de probabilidade de B dado que A ocorreu,
- ou, mais curto ainda, probabilidade de B dado A.

Probabilidade Condicional

- De posse da informação de que certo evento A ocorreu, queremos recalcular as chances de outros eventos $B_1, B_2, \ldots$
- Chamamos a isto de probabilidade de B condicionada à ocorrência do evento A,
- ou de probabilidade de B dado que A ocorreu,
- ou, mais curto ainda, probabilidade de B dado A.
- Notação: $\mathbb{P}(B|A)$

Probabilidade Condicional

- De posse da informação de que certo evento A ocorreu, queremos recalcular as chances de outros eventos $B_1, B_2, \ldots$
- Chamamos a isto de probabilidade de B condicionada à ocorrência do evento A,
- ou de probabilidade de B dado que A ocorreu,
- ou, mais curto ainda, probabilidade de B dado A.
- Notação: $\mathbb{P}(B|A)$
- A imensa maioria das técnicas de Aprendizagem de Máquina (ou Machine Learning, ML) são algoritmos para fazer cálculos de probabilidade condicional.

Probabilidade Condicional e ML

- Momento do diagnóstico de um câncer de estômago para um paciente qualquer.

Probabilidade Condicional e ML

- Momento do diagnóstico de um câncer de estômago para um paciente qualquer.
- B é o evento em que o paciente terá pelo menos mais 1 ano de vida.

Probabilidade Condicional e ML

- Momento do diagnóstico de um câncer de estômago para um paciente qualquer.
- B é o evento em que o paciente terá pelo menos mais 1 ano de vida.
- Suponha que $\mathbb{P}(B) = 0.70$.

Probabilidade Condicional e ML

- Momento do diagnóstico de um câncer de estômago para um paciente qualquer.
- B é o evento em que o paciente terá pelo menos mais 1 ano de vida.
- Suponha que $\mathbb{P}(B) = 0.70$.
- Usando a idéia frequentista, dentre todos os pacientes observados em situação semelhante no passado, 70% deles viveu mais de um ano a partir do diagnóstico.

Probabilidade Condicional e ML

- Momento do diagnóstico de um câncer de estômago para um paciente qualquer.
- B é o evento em que o paciente terá pelo menos mais 1 ano de vida.
- Suponha que $\mathbb{P}(B) = 0.70$.
- Usando a idéia frequentista, dentre todos os pacientes observados em situação semelhante no passado, 70% deles viveu mais de um ano a partir do diagnóstico.
- Seja A o evento em que um paciente de câncer de estômago tenha uma biópsia confirmando que o tumor é benigno.

Probabilidade Condicional e ML

- Momento do diagnóstico de um câncer de estômago para um paciente qualquer.
- B é o evento em que o paciente terá pelo menos mais 1 ano de vida.
- Suponha que $\mathbb{P}(B) = 0.70$.
- Usando a idéia frequentista, dentre todos os pacientes observados em situação semelhante no passado, 70% deles viveu mais de um ano a partir do diagnóstico.
- Seja A o evento em que um paciente de câncer de estômago tenha uma biópsia confirmando que o tumor é benigno.
- Imaginamos que $\mathbb{P}(B|A)$ seja maior que $\mathbb{P}(B) = 0.70$. Como recalcular a probabilidade da ocorrência de B?

Probabilidade Condicional e ML

- Momento do diagnóstico de um câncer de estômago para um paciente qualquer.
- B é o evento em que o paciente terá pelo menos mais 1 ano de vida.
- Suponha que $\mathbb{P}(B) = 0.70$.
- Usando a idéia frequentista, dentre todos os pacientes observados em situação semelhante no passado, 70% deles viveu mais de um ano a partir do diagnóstico.
- Seja A o evento em que um paciente de câncer de estômago tenha uma biópsia confirmando que o tumor é benigno.
- Imaginamos que $\mathbb{P}(B|A)$ seja maior que $\mathbb{P}(B) = 0.70$. Como recalcular a probabilidade da ocorrência de B?
- Se tivermos um grande número de pacientes inicialmente diagnosticados e com biópsia posterior indicando benigno, contamos a proporção desses indivíduos que sobrevivem mais de um ano. Isto Será uma boa aproximação para $\mathbb{P}(B|A)$.

Probabilidade Condicional e ML

- O problema fica mais complicado se o evento A representar a seguinte informação:
  - biópsia indica benigno,

Probabilidade Condicional e ML

- O problema fica mais complicado se o evento A representar a seguinte informação:
  - biópsia indica benigno,
  - paciente tem 45 anos de idade,

Probabilidade Condicional e ML

- O problema fica mais complicado se o evento A representar a seguinte informação:
  - biópsia indica benigno,
  - paciente tem 45 anos de idade,
  - é homem,

Probabilidade Condicional e ML

- O problema fica mais complicado se o evento A representar a seguinte informação:
  - biópsia indica benigno,
  - paciente tem 45 anos de idade,
  - é homem,
  - sempre morou em Santa Catarina,

Probabilidade Condicional e ML

- O problema fica mais complicado se o evento A representar a seguinte informação:
    - biópsia indica benigno,
    - paciente tem 45 anos de idade,
    - é homem,
    - sempre morou em Santa Catarina,
    - é fumante

Probabilidade Condicional e ML

- O problema fica mais complicado se o evento A representar a seguinte informação:
  - biópsia indica benigno,
  - paciente tem 45 anos de idade,
  - é homem,
  - sempre morou em Santa Catarina,
  - é fumante
  - e sempre come salames e salsichas defumadas.

Probabilidade Condicional e ML

- O problema fica mais complicado se o evento A representar a seguinte informação:
  - biópsia indica benigno,
  - paciente tem 45 anos de idade,
  - é homem,
  - sempre morou em Santa Catarina,
  - é fumante
  - e sempre come salames e salsichas defumadas.
- Não haverá uma amostra muito grande de pacientes nestas condições exatas.

Probabilidade Condicional e ML

- O problema fica mais complicado se o evento A representar a seguinte informação:
  - biópsia indica benigno,
  - paciente tem 45 anos de idade,
  - é homem,
  - sempre morou em Santa Catarina,
  - é fumante
  - e sempre come salames e salsichas defumadas.
- Não haverá uma amostra muito grande de pacientes nestas condições exatas.
- Talvez apenas 2, 1 ou até zero pessoas tenham sido observadas nestas condições.

Probabilidade Condicional e ML

- O problema fica mais complicado se o evento A representar a seguinte informação:
  - biópsia indica benigno,
  - paciente tem 45 anos de idade,
  - é homem,
  - sempre morou em Santa Catarina,
  - é fumante
  - e sempre come salames e salsichas defumadas.
- Não haverá uma amostra muito grande de pacientes nestas condições exatas.
- Talvez apenas 2, 1 ou até zero pessoas tenham sido observadas nestas condições.
- Isto impede usar a simples frequência ocorrida nestes pouquíssimos casos para aproximar $\mathbb{P}(B|A)$.

Probabilidade Condicional e ML

- O problema fica mais complicado se o evento A representar a seguinte informação:
  - biópsia indica benigno,
  - paciente tem 45 anos de idade,
  - é homem,
  - sempre morou em Santa Catarina,
  - é fumante
  - e sempre come salames e salsichas defumadas.
- Não haverá uma amostra muito grande de pacientes nestas condições exatas.
- Talvez apenas 2, 1 ou até zero pessoas tenham sido observadas nestas condições.
- Isto impede usar a simples frequência ocorrida nestes pouquíssimos casos para aproximar $\mathbb{P}(B|A)$.
- Ferramentas de ML calculam estas probabilidades usando vários truques. Elas procuram extrair o máximo de informação dos dados.

Probabilidade Condicional e ML

- De maneira geral, dadas as características representadas por A, como fica a chance de ocorrer B?

Probabilidade Condicional e ML

- De maneira geral, dadas as características representadas por A, como fica a chance de ocorrer B?
- Dado que os sensores do robô dizem que ocorreu A, qual a chance de que ele esteja na região B?

Probabilidade Condicional e ML

- De maneira geral, dadas as características representadas por A, como fica a chance de ocorrer B?
- Dado que os sensores do robô dizem que ocorreu A, qual a chance de que ele esteja na região B?
- Dado que o usuário comprou o conjunto A de itens nas últimas visitas, qual a chance dele comprar o item B agora?

Probabilidade Condicional e ML

- De maneira geral, dadas as características representadas por A, como fica a chance de ocorrer B?
- Dado que os sensores do robô dizem que ocorreu A, qual a chance de que ele esteja na região B?
- Dado que o usuário comprou o conjunto A de itens nas últimas visitas, qual a chance dele comprar o item B agora?
- Dado certo comportamento da ação nos últimos 3 anos, qual a probabilidade de que ela suba 10% ou mais dentro de 30 dias?

Probabilidade Condicional e ML

- De maneira geral, dadas as características representadas por A, como fica a chance de ocorrer B?
- Dado que os sensores do robô dizem que ocorreu A, qual a chance de que ele esteja na região B?
- Dado que o usuário comprou o conjunto A de itens nas últimas visitas, qual a chance dele comprar o item B agora?
- Dado certo comportamento da ação nos últimos 3 anos, qual a probabilidade de que ela suba 10% ou mais dentro de 30 dias?
- Dadas certas características A de um e-mail, qual a chance dele ser um spam?

Probabilidade Condicional e ML

- De maneira geral, dadas as características representadas por A, como fica a chance de ocorrer B?
- Dado que os sensores do robô dizem que ocorreu A, qual a chance de que ele esteja na região B?
- Dado que o usuário comprou o conjunto A de itens nas últimas visitas, qual a chance dele comprar o item B agora?
- Dado certo comportamento da ação nos últimos 3 anos, qual a probabilidade de que ela suba 10% ou mais dentro de 30 dias?
- Dadas certas características A de um e-mail, qual a chance dele ser um spam?
- Probabilidade condicional é:
  - extremamente importante em teoria.
  - é mais importante ainda na prática de análise de dados.
  - pode ser difícil de calcular: é a fonte de quase todos os paradoxos em probabilidade.

Probabilidade Condicional

- Primeira questão: como passar de $\mathbb{P}(B)$ para $\mathbb{P}(B|A)$?

Probabilidade Condicional

- Primeira questão: como passar de $\mathbb{P}(B)$ para $\mathbb{P}(B|A)$?
- Qual a relação entre $\mathbb{P}(B)$ e $\mathbb{P}(B|A)$?

Probabilidade Condicional

- Primeira questão: como passar de $\mathbb{P}(B)$ para $\mathbb{P}(B|A)$?
- Qual a relação entre $\mathbb{P}(B)$ e $\mathbb{P}(B|A)$?
- Podemos ter $\mathbb{P}(B) = \mathbb{P}(B|A)$?

Probabilidade Condicional

- Primeira questão: como passar de $\mathbb{P}(B)$ para $\mathbb{P}(B|A)$?
- Qual a relação entre $\mathbb{P}(B)$ e $\mathbb{P}(B|A)$?
- Podemos ter $\mathbb{P}(B) = \mathbb{P}(B|A)$?
- Veremos que, em alguns casos sim. NESTE CASOS, a ocorrência de A não afeta as chances da ocorrência de B.

Probabilidade Condicional

- Primeira questão: como passar de $\mathbb{P}(B)$ para $\mathbb{P}(B|A)$?
- Qual a relação entre $\mathbb{P}(B)$ e $\mathbb{P}(B|A)$?
- Podemos ter $\mathbb{P}(B) = \mathbb{P}(B|A)$?
- Veremos que, em alguns casos sim. NESTE CASOS, a ocorrência de A não afeta as chances da ocorrência de B.
- Muitas vezes, teremos $\mathbb{P}(B) \neq \mathbb{P}(B|A)$.

Probabilidade Condicional

- Primeira questão: como passar de $\mathbb{P}(B)$ para $\mathbb{P}(B|A)$?
- Qual a relação entre $\mathbb{P}(B)$ e $\mathbb{P}(B|A)$?
- Podemos ter $\mathbb{P}(B) = \mathbb{P}(B|A)$?
- Veremos que, em alguns casos sim. NESTE CASOS, a ocorrência de A não afeta as chances da ocorrência de B.
- Muitas vezes, teremos $\mathbb{P}(B) \neq \mathbb{P}(B|A)$.
- Por exemplo, queremos saber quando teremos $\mathbb{P}(B) < \mathbb{P}(B|A)$.

Probabilidade Condicional

- Primeira questão: como passar de $\mathbb{P}(B)$ para $\mathbb{P}(B|A)$?
- Qual a relação entre $\mathbb{P}(B)$ e $\mathbb{P}(B|A)$?
- Podemos ter $\mathbb{P}(B) = \mathbb{P}(B|A)$?
- Veremos que, em alguns casos sim. NESTE CASOS, a ocorrência de A não afeta as chances da ocorrência de B.
- Muitas vezes, teremos $\mathbb{P}(B) \neq \mathbb{P}(B|A)$.
- Por exemplo, queremos saber quando teremos $\mathbb{P}(B) < \mathbb{P}(B|A)$.
- Mais do que isto, queremos uma fórmula que nos permita calcular de maneira exata $\mathbb{P}(B|A)$ em qualquer situação.

Um caso óbvio

- Alguns casos são fáceis de calcular pois eles são casos extremos.

Um caso óbvio

- Alguns casos são fáceis de calcular pois eles são casos extremos.
- Por exemplo, lançar um dado bem equilibrado e anotar a face: $\Omega = \{1, 2, \ldots, 6\}$.

Um caso óbvio

- Alguns casos são fáceis de calcular pois eles são casos extremos.
- Por exemplo, lançar um dado bem equilibrado e anotar a face: $\Omega = \{1, 2, \ldots, 6\}$.
- Seja $B = \{4, 5, 6\}$ com $\mathbb{P}(B) = 3/6$.

Um caso óbvio

- Alguns casos são fáceis de calcular pois eles são casos extremos.
- Por exemplo, lançar um dado bem equilibrado e anotar a face: $\Omega = \{1, 2, \ldots, 6\}$.
- Seja $B = \{4, 5, 6\}$ com $\mathbb{P}(B) = 3/6$.
- Vamos considerar um evento $A \subset B$. Por exemplo, $A = \{5, 6\}$.

Um caso óbvio

- Alguns casos são fáceis de calcular pois eles são casos extremos.
- Por exemplo, lançar um dado bem equilibrado e anotar a face: $\Omega = \{1, 2, \ldots, 6\}$.
- Seja $B = \{4, 5, 6\}$ com $\mathbb{P}(B) = 3/6$.
- Vamos considerar um evento $A \subset B$. Por exemplo, $A = \{5, 6\}$.
- Intuitivamente, o que deveria ser $\mathbb{P}(B|A)$?

Um caso óbvio

- Alguns casos são fáceis de calcular pois eles são casos extremos.
- Por exemplo, lançar um dado bem equilibrado e anotar a face: $\Omega = \{1, 2, \ldots, 6\}$.
- Seja $B = \{4, 5, 6\}$ com $\mathbb{P}(B) = 3/6$.
- Vamos considerar um evento $A \subset B$. Por exemplo, $A = \{5, 6\}$.
- Intuitivamente, o que deveria ser $\mathbb{P}(B|A)$?
- Qual a probabilidade de que a face seja 4, 5 ou 6 sabendo que saiu 5 ou 6?

Um caso óbvio

- Alguns casos são fáceis de calcular pois eles são casos extremos.
- Por exemplo, lançar um dado bem equilibrado e anotar a face: $\Omega = \{1, 2, \ldots, 6\}$.
- Seja $B = \{4, 5, 6\}$ com $\mathbb{P}(B) = 3/6$.
- Vamos considerar um evento $A \subset B$. Por exemplo, $A = \{5, 6\}$.
- Intuitivamente, o que deveria ser $\mathbb{P}(B|A)$?
- Qual a probabilidade de que a face seja 4, 5 ou 6 sabendo que saiu 5 ou 6?
- Ao saber que um evento $\omega \in A$ ocorreu, automaticamente inferimos que B também ocorreu pois $A \subset B$.

Um caso óbvio

- Alguns casos são fáceis de calcular pois eles são casos extremos.
- Por exemplo, lançar um dado bem equilibrado e anotar a face: $\Omega = \{1, 2, \ldots, 6\}$.
- Seja $B = \{4, 5, 6\}$ com $\mathbb{P}(B) = 3/6$.
- Vamos considerar um evento $A \subset B$. Por exemplo, $A = \{5, 6\}$.
- Intuitivamente, o que deveria ser $\mathbb{P}(B|A)$?
- Qual a probabilidade de que a face seja 4, 5 ou 6 sabendo que saiu 5 ou 6?
- Ao saber que um evento $\omega \in A$ ocorreu, automaticamente inferimos que B também ocorreu pois $A \subset B$.
- Assim, devemos ter $\mathbb{P}(B|A) = 1 > \mathbb{P}(B) = 3/6$.

Outro caso óbvio

- Outro caso óbvio: $A \cap B = \emptyset$.

Outro caso óbvio

- Outro caso óbvio: $A \cap B = \emptyset$.
- Intuitivamente, o que deveria ser $\mathbb{P}(B|A)$?

Outro caso óbvio

- Outro caso óbvio: $A \cap B = \emptyset$.
- Intuitivamente, o que deveria ser $\mathbb{P}(B|A)$?
- Se o evento que ocorreu está em A, ele não pode estar em B (pois A e B são disjuntos).

Outro caso óbvio

- Outro caso óbvio: $A \cap B = \emptyset$.
- Intuitivamente, o que deveria ser $\mathbb{P}(B|A)$?
- Se o evento que ocorreu está em A, ele não pode estar em B (pois A e B são disjuntos).
- Ao saber que um evento $\omega \in A$ ocorreu, automaticamente inferimos que B não ocorreu.

Outro caso óbvio

- Outro caso óbvio: $A \cap B = \emptyset$.
- Intuitivamente, o que deveria ser $\mathbb{P}(B|A)$?
- Se o evento que ocorreu está em A, ele não pode estar em B (pois A e B são disjuntos).
- Ao saber que um evento $\omega \in A$ ocorreu, automaticamente inferimos que B não ocorreu.
- Assim, devemos ter $\mathbb{P}(B|A) = 0 \leq \mathbb{P}(B)$.

Outro caso óbvio: um exemplo

- Exemplo do dado com $\Omega = \{1, 2, \ldots, 6\}$.
- Seja B o evento *FACE PAR*.

Outro caso óbvio: um exemplo

- Exemplo do dado com $\Omega = \{1, 2, \ldots, 6\}$.
- Seja B o evento *FACE PAR*.
- Isto é, $B = \{2, 4, 6\}$ com $P(B) = \frac{1}{2}$.

Outro caso óbvio: um exemplo

- Exemplo do dado com $\Omega = \{1, 2, \ldots, 6\}$.
- Seja B o evento *FACE PAR*.
- Isto é, $B = \{2, 4, 6\}$ com $P(B) = \frac{1}{2}$.
- Seja $A = \{5\}$. É claro que $A \cap B = \emptyset$.

Outro caso óbvio: um exemplo

- Exemplo do dado com $\Omega = \{1, 2, \ldots, 6\}$.
- Seja B o evento *FACE PAR*.
- Isto é, $B = \{2, 4, 6\}$ com $P(B) = \frac{1}{2}$.
- Seja $A = \{5\}$. É claro que $A \cap B = \emptyset$.
- Intuitivamente, se ocorreu a face 5, qual a chance de ocorrer uma face par?

Outro caso óbvio: um exemplo

- Exemplo do dado com $\Omega = \{1, 2, \ldots, 6\}$.
- Seja B o evento *FACE PAR*.
- Isto é, $B = \{2, 4, 6\}$ com $P(B) = \frac{1}{2}$.
- Seja $A = \{5\}$. É claro que $A \cap B = \emptyset$.
- Intuitivamente, se ocorreu a face 5, qual a chance de ocorrer uma face par?
- Esta chance é zero.

Outro caso óbvio: um exemplo

- Exemplo do dado com $\Omega = \{1, 2, \ldots, 6\}$.
- Seja B o evento *FACE PAR*.
- Isto é, $B = \{2, 4, 6\}$ com $P(B) = \frac{1}{2}$.
- Seja $A = \{5\}$. É claro que $A \cap B = \emptyset$.
- Intuitivamente, se ocorreu a face 5, qual a chance de ocorrer uma face par?
- Esta chance é zero.
- Ou você apostaria na ocorrência de B neste caso?

Os outros casos

- Assim, dois casos intuitivamente óbvios são:
  - Se $A \subset B$ então $\mathbb{P}(B|A) = 1$.
  - Se $A \cap B = \emptyset$ então $\mathbb{P}(B|A) = 0$.
- E o caso geral?

Os outros casos

- Assim, dois casos intuitivamente óbvios são:
  - Se $A \subset B$ então $\mathbb{P}(B|A) = 1$.
  - Se $A \cap B = \emptyset$ então $\mathbb{P}(B|A) = 0$.
- E o caso geral?
- Sejam A e B dois eventos com $P(A) > 0$ e com $A \cap B \neq \emptyset$.

Os outros casos

- Assim, dois casos intuitivamente óbvios são:
  - Se $A \subset B$ então $\mathbb{P}(B|A) = 1$.
  - Se $A \cap B = \emptyset$ então $\mathbb{P}(B|A) = 0$.
- E o caso geral?
- Sejam A e B dois eventos com $P(A) > 0$ e com $A \cap B \neq \emptyset$.
- Como calcular $\mathbb{P}(B|A)$?

Definição

- Sejam A e B dois eventos com $P(A) > 0$.

Definição

- Sejam A e B dois eventos com $P(A) > 0$.
- Então, por definição,

Definição

- Sejam A e B dois eventos com $P(A) > 0$.
- Então, por definição,

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

Definição

- Sejam A e B dois eventos com $P(A) > 0$.
- Então, por definição,

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

- Assim, para calcular a probabilidade de que A ocorreu DADO QUE B ocorreu:

Definição

- Sejam A e B dois eventos com $P(A) > 0$.
- Então, por definição,

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

- Assim, para calcular a probabilidade de que A ocorreu DADO QUE B ocorreu:
  - Calcule a probabilidade $\mathbb{P}(A \cap B)$ de que A e B tenham ambos ocorrido
  - Aumente esta probabilidade multiplicando-a por $1/\mathbb{P}(B)$, que é maior que 1.

Exemplo

- Considere o lançamento de uma moeda honesta 5 vezes seguidas:

$$\Omega = \{CCCCC, CCCC\tilde{C}, \ldots, \tilde{C}\tilde{C}\tilde{C}\tilde{C}\tilde{C}\}$$
$$\hookrightarrow 32 \text{ elementos}$$

Exemplo

- Considere o lançamento de uma moeda honesta 5 vezes seguidas:

$$\Omega = \{CCCCC, CCCC\tilde{C}, \ldots, \tilde{C}\tilde{C}\tilde{C}\tilde{C}\tilde{C}\}$$
$$\hookrightarrow 32 \text{ elementos}$$

- Temos $\mathbb{P}(\omega) = 1/32$, igualmente prováveis.

Exemplo

- Considere o lançamento de uma moeda honesta 5 vezes seguidas:

$$\Omega = \{CCCCC, CCCC\tilde{C}, \ldots, \tilde{C}\tilde{C}\tilde{C}\tilde{C}\tilde{C}\}$$
$$\hookrightarrow 32 \text{ elementos}$$

- Temos $\mathbb{P}(\omega) = 1/32$, igualmente prováveis.
- Seja $B = \{\omega \in \Omega \ ; \ 1^o \text{ elemento é } C\}$

Exemplo

- Considere o lançamento de uma moeda honesta 5 vezes seguidas:

$$\Omega = \{CCCCC, CCCC\tilde{C}, \ldots, \tilde{C}\tilde{C}\tilde{C}\tilde{C}\tilde{C}\}$$
$$\hookrightarrow 32 \text{ elementos}$$

- Temos $\mathbb{P}(\omega) = 1/32$, igualmente prováveis.
- Seja $B = \{\omega \in \Omega \; ; \; 1^o \text{ elemento é } C\}$
- Temos $\mathbb{P}(B) = 1/2$.

Exemplo

- Considere o lançamento de uma moeda honesta 5 vezes seguidas:

$$\Omega = \{CCCCC, CCCC\tilde{C}, \ldots, \tilde{C}\tilde{C}\tilde{C}\tilde{C}\tilde{C}\}$$
$$\hookrightarrow 32 \text{ elementos}$$

- Temos $\mathbb{P}(\omega) = 1/32$, igualmente prováveis.
- Seja $B = \{\omega \in \Omega \; ; \; 1^o \text{ elemento é } C\}$
- Temos $\mathbb{P}(B) = 1/2$. OK?

Exemplo

- Considere o lançamento de uma moeda honesta 5 vezes seguidas:

$$\Omega = \{CCCCC, CCCC\tilde{C}, \ldots, \tilde{C}\tilde{C}\tilde{C}\tilde{C}\tilde{C}\}$$
$$\hookrightarrow 32 \text{ elementos}$$

- Temos $\mathbb{P}(\omega) = 1/32$, igualmente prováveis.
- Seja $B = \{\omega \in \Omega \ ; \ 1^o \text{ elemento é } C\}$
- Temos $\mathbb{P}(B) = 1/2$. OK?
- 16 elementos em Ω tem C na 1^a posição. Como são igualmente prováveis, $\mathbb{P}(B) = 16/32 = 1/2$.

Exemplo

- Considere o lançamento de uma moeda honesta 5 vezes seguidas:

$$\Omega = \{CCCCC, CCCC\tilde{C}, \ldots, \tilde{C}\tilde{C}\tilde{C}\tilde{C}\tilde{C}\}$$
$$\hookrightarrow 32 \text{ elementos}$$

- Temos $\mathbb{P}(\omega) = 1/32$, igualmente prováveis.
- Seja $B = \{\omega \in \Omega \; ; \; 1^o \text{ elemento é } C\}$
- Temos $\mathbb{P}(B) = 1/2$. OK?
- 16 elementos em Ω tem C na 1^a posição. Como são igualmente prováveis, $\mathbb{P}(B) = 16/32 = 1/2$.
- É fornecida a seguinte informação: ocorreu $A = \{ \text{ Houve apenas uma coroa nos 5 lançamentos } \}$.

Exemplo

- Considere o lançamento de uma moeda honesta 5 vezes seguidas:

$$\Omega = \{CCCCC, CCCC\tilde{C}, \ldots, \tilde{C}\tilde{C}\tilde{C}\tilde{C}\tilde{C}\}$$
$$\hookrightarrow 32 \text{ elementos}$$

- Temos $\mathbb{P}(\omega) = 1/32$, igualmente prováveis.
- Seja $B = \{\omega \in \Omega \ ; \ 1^o \text{ elemento é } C\}$
- Temos $\mathbb{P}(B) = 1/2$. OK?
- 16 elementos em Ω tem C na 1^a posição. Como são igualmente prováveis, $\mathbb{P}(B) = 16/32 = 1/2$.
- É fornecida a seguinte informação: ocorreu $A = \{ \text{ Houve apenas uma coroa nos 5 lançamentos } \}$.
- Intuitivamente, $\mathbb{P}(B|A) > \mathbb{P}(B) = 1/2$.

Exemplo: usando a definição

- Calculando $\mathbb{P}(B|A)$ pela definição:

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{4/32}{5/32} = \frac{4}{5}$$

Exemplo: usando a definição

- Calculando $\mathbb{P}(B|A)$ pela definição:

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{4/32}{5/32} = \frac{4}{5}$$

- Assim a probabilidade mudou bastante ao sabermos que A ocorreu:

$$\mathbb{P}(B) = \frac{1}{2} \rightsquigarrow \mathbb{P}(B|A) = \frac{4}{5}$$

Exemplo: usando a definição

- Calculando $\mathbb{P}(B|A)$ pela definição:

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{4/32}{5/32} = \frac{4}{5}$$

- Assim a probabilidade mudou bastante ao sabermos que A ocorreu:

$$\mathbb{P}(B) = \frac{1}{2} \rightsquigarrow \mathbb{P}(B|A) = \frac{4}{5}$$

- Saber que ocorreu apenas uma coroa em cinco lançamentos torna altamente provável que a 1^a posição seja cara.

ML e Condicional

- Este é um dos grandes objetivos gerais de ML:

ML e Condicional

- Este é um dos grandes objetivos gerais de ML:
- quando tivermos um sistema complexo, envolvendo vários fatores, obtemos a um custo baixo algumas informações.

ML e Condicional

- Este é um dos grandes objetivos gerais de ML:
- quando tivermos um sistema complexo, envolvendo vários fatores, obtemos a um custo baixo algumas informações.
- Estas informações são representadas por A.

ML e Condicional

- Este é um dos grandes objetivos gerais de ML:
- quando tivermos um sistema complexo, envolvendo vários fatores, obtemos a um custo baixo algumas informações.
- Estas informações são representadas por A.
- Usamos estas informaçãões de baixo custo para recalcular as probabilidades de coisas que não sabemos: $\mathbb{P}(B|A)$.

ML e Condicional

- Este é um dos grandes objetivos gerais de ML:
- quando tivermos um sistema complexo, envolvendo vários fatores, obtemos a um custo baixo algumas informações.
- Estas informações são representadas por A.
- Usamos estas informaçãões de baixo custo para recalcular as probabilidades de coisas que não sabemos: $\mathbb{P}(B|A)$.
- Não sabemos porque estão no futuro ou porque são caras para observar ou porque são impossíveis ou é anti-ético conhecer, etc.

ML e Condicional

- Este é um dos grandes objetivos gerais de ML:
- quando tivermos um sistema complexo, envolvendo vários fatores, obtemos a um custo baixo algumas informações.
- Estas informações são representadas por A.
- Usamos estas informaçãões de baixo custo para recalcular as probabilidades de coisas que não sabemos: $\mathbb{P}(B|A)$.
- Não sabemos porque estão no futuro ou porque são caras para observar ou porque são impossíveis ou é anti-ético conhecer, etc.
- Com estas probabilidades recalculadas podemos tomar decisões.

Intuição para a definição

- Vimos a definição $\mathbb{P}(B|A) = \mathbb{P}(A \cap B)/\mathbb{P}(A)$.

Intuição para a definição

- Vimos a definição $\mathbb{P}(B|A) = \mathbb{P}(A \cap B)/\mathbb{P}(A)$.
- Por quê a fórmula acima?

Intuição para a definição

- Vimos a definição $\mathbb{P}(B|A) = \mathbb{P}(A \cap B)/\mathbb{P}(A)$.
- Por quê a fórmula acima? Por quê esta definição, e não outra tal como

$$\frac{\mathbb{P}(A \cup B)}{\mathbb{P}(A)} \quad \text{ou} \quad \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad ??$$

Intuição para a definição

- Vimos a definição $\mathbb{P}(B|A) = \mathbb{P}(A \cap B)/\mathbb{P}(A)$.
- Por quê a fórmula acima? Por quê esta definição, e não outra tal como

$$\frac{\mathbb{P}(A \cup B)}{\mathbb{P}(A)} \quad \text{ou} \quad \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad ??$$

- Resposta: para ser consistente com a experiência empírica.

Intuição para a definição

- Vimos a definição $\mathbb{P}(B|A) = \mathbb{P}(A \cap B)/\mathbb{P}(A)$.
- Por quê a fórmula acima? Por quê esta definição, e não outra tal como
$$\frac{\mathbb{P}(A \cup B)}{\mathbb{P}(A)} \quad \text{ou} \quad \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad ??$$
- Resposta: para ser consistente com a experiência empírica.
- Para ver isto, vamos encontrar $\mathbb{P}(B|A)$ de duas formas distintas num caso simples de simular no computador.

Intuição para a definição

- Vimos a definição $\mathbb{P}(B|A) = \mathbb{P}(A \cap B)/\mathbb{P}(A)$.
- Por quê a fórmula acima? Por quê esta definição, e não outra tal como

$$\frac{\mathbb{P}(A \cup B)}{\mathbb{P}(A)} \quad \text{ou} \quad \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad ??$$

- Resposta: para ser consistente com a experiência empírica.
- Para ver isto, vamos encontrar $\mathbb{P}(B|A)$ de duas formas distintas num caso simples de simular no computador.
- Uma das forma será através da contagem do evento B dentre aqueles casos em que A ocorre.
- Esta é a forma natural de estimar probabilidades: pela frequência relativa.

Intuição para a definição

- Vimos a definição $\mathbb{P}(B|A) = \mathbb{P}(A \cap B)/\mathbb{P}(A)$.
- Por quê a fórmula acima? Por quê esta definição, e não outra tal como
$$\frac{\mathbb{P}(A \cup B)}{\mathbb{P}(A)} \quad \text{ou} \quad \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad ??$$
- Resposta: para ser consistente com a experiência empírica.
- Para ver isto, vamos encontrar $\mathbb{P}(B|A)$ de duas formas distintas num caso simples de simular no computador.
- Uma das forma será através da contagem do evento B dentre aqueles casos em que A ocorre.
- Esta é a forma natural de estimar probabilidades: pela frequência relativa.
- A segunda forma será pela definição $\mathbb{P}(B|A) = \mathbb{P}(A \cap B)/\mathbb{P}(A)$.

Intuição para a definição

- Vimos a definição $\mathbb{P}(B|A) = \mathbb{P}(A \cap B)/\mathbb{P}(A)$.
- Por quê a fórmula acima? Por quê esta definição, e não outra tal como
$$\frac{\mathbb{P}(A \cup B)}{\mathbb{P}(A)} \quad \text{ou} \quad \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad ??$$
- Resposta: para ser consistente com a experiência empírica.
- Para ver isto, vamos encontrar $\mathbb{P}(B|A)$ de duas formas distintas num caso simples de simular no computador.
- Uma das forma será através da contagem do evento B dentre aqueles casos em que A ocorre.
- Esta é a forma natural de estimar probabilidades: pela frequência relativa.
- A segunda forma será pela definição $\mathbb{P}(B|A) = \mathbb{P}(A \cap B)/\mathbb{P}(A)$.
- Veremos que as duas coincidem e portanto que a definição é o que tem de ser.

Intuição para a definição

- Role um dado duas vezes

$$\Omega = \{(1,1),(1,2),\ldots,(6,6)\} \qquad P(\omega) = \frac{1}{36}$$

Intuição para a definição

- Role um dado duas vezes

$$\Omega = \{(1,1),(1,2),\ldots,(6,6)\} \qquad P(\omega) = \frac{1}{36}$$

- Seja $B = [1^o \text{ dado é um } 6]$

Intuição para a definição

- Role um dado duas vezes

$$\Omega = \{(1,1),(1,2),\ldots,(6,6)\} \qquad P(\omega) = \frac{1}{36}$$

- Seja $B = [1^o \text{ dado é um } 6]$
- $B = \{(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}$ e $\mathbb{P}(B) = 6/36 = 1/6$.

Intuição para a definição

- Role um dado duas vezes

$$\Omega = \{(1,1), (1,2), \dots, (6,6)\} \qquad P(\omega) = \frac{1}{36}$$

- Seja $B = [1^o \text{ dado é um } 6]$
- $B = \{(6,1), (6,2), (6,3), (6,4), (6,5), (6,6)\}$ e $\mathbb{P}(B) = 6/36 = 1/6$.
- Seja $A = [\text{ Soma das faces é maior que } 8 \ ]$

Intuição para a definição

- Role um dado duas vezes

$$\Omega = \{(1,1),(1,2),\dots,(6,6)\} \qquad P(\omega) = \frac{1}{36}$$

- Seja $B = [1^o \text{ dado é um } 6]$
- $B = \{(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}$ e $\mathbb{P}(B) = 6/36 = 1/6$.
- Seja $A = [\text{ Soma das faces é maior que } 8 \text{ }]$
- $A = \{(3,6),(4,5),(4,6),(5,4),(5,5),(5,6),(6,3),(6,4),(6,5),(6,6)\}$ e $\mathbb{P}(A) = 10/36 = 0.28$.

Intuição para a definição

- Role um dado duas vezes

$$\Omega = \{(1,1),(1,2),\ldots,(6,6)\} \qquad P(\omega) = \frac{1}{36}$$

- Seja $B = [1^o \text{ dado é um } 6]$
- $B = \{(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}$ e $\mathbb{P}(B) = 6/36 = 1/6$.
- Seja $A = [\text{ Soma das faces é maior que } 8 \text{ }]$
- $A = \{(3,6),(4,5),(4,6),(5,4),(5,5),(5,6),(6,3),(6,4),(6,5),(6,6)\}$ e $\mathbb{P}(A) = 10/36 = 0.28$.
- Quanto é $P(B|A)$? Devemos esperar que seja maior ou menor que $\mathbb{P}(B)$?

Intuição para a definição

- B : primeiro dado é um 6. A : soma das faces é maior que 8.
- A soma das faces varia de 2 a 12.

Intuição para a definição

- B : primeiro dado é um 6. A : soma das faces é maior que 8.
- A soma das faces varia de 2 a 12.
- Ser > 8 quer dizer que é um valor alto e que podemos esperar que as duas faces sejam pelo menos moderadas.

Intuição para a definição

- B : primeiro dado é um 6. A : soma das faces é maior que 8.
- A soma das faces varia de 2 a 12.
- Ser >8 quer dizer que é um valor alto e que podemos esperar que as duas faces sejam pelo menos moderadas.
- De fato, usando a fórmula,

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} = \frac{4/36}{10/36} = 0.4 > 1/6 = 0.17 = \mathbb{P}(B)$$

Intuição para a definição

- B : primeiro dado é um 6. A : soma das faces é maior que 8.
- A soma das faces varia de 2 a 12.
- Ser > 8 quer dizer que é um valor alto e que podemos esperar que as duas faces sejam pelo menos moderadas.
- De fato, usando a fórmula,

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} = \frac{4/36}{10/36} = 0.4 > 1/6 = 0.17 = \mathbb{P}(B)$$

- Vamos calcular $P(B|A)$ simulando os dados num computador.

Intuição para a definição

- B : primeiro dado é um 6. A : soma das faces é maior que 8.
- A soma das faces varia de 2 a 12.
- Ser > 8 quer dizer que é um valor alto e que podemos esperar que as duas faces sejam pelo menos moderadas.
- De fato, usando a fórmula,

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} = \frac{4/36}{10/36} = 0.4 > 1/6 = 0.17 = \mathbb{P}(B)$$

- Vamos calcular $P(B|A)$ simulando os dados num computador.
- Replique os lançamentos duplos um grande número N de vezes (por exemplo, $N = 100$ mil)

Intuição para a definição

- B : primeiro dado é um 6. A : soma das faces é maior que 8.

Repetição	1	2	3	4	5	6	7	8	9	10	...
Dado 1	2	5	5	2	6	4	2	1	6	6	...
Dado 2	1	5	1	3	1	5	3	6	4	3	...
B ocorreu?	N	N	N	N	Y	N	N	N	Y	Y	...
A ocorreu?	N	Y	N	N	N	Y	N	N	Y	Y	...

Tabela: Lançamentos duplos de dados

Intuição para a definição

- B : primeiro dado é um 6. A : soma das faces é maior que 8.

Repetição	1	2	3	4	5	6	7	8	9	10	...
Dado 1	2	5	5	2	6	4	2	1	6	6	...
Dado 2	1	5	1	3	1	5	3	6	4	3	...
B ocorreu?	N	N	N	N	Y	N	N	N	Y	Y	...
A ocorreu?	N	Y	N	N	N	Y	N	N	Y	Y	...

Tabela: Lançamentos duplos de dados

- Considere apenas as vezes em que A ocorreu: 13886 vezes

Intuição para a definição

- B : primeiro dado é um 6. A : soma das faces é maior que 8.

Repetição	1	2	3	4	5	6	7	8	9	10	...
Dado 1	2	5	5	2	6	4	2	1	6	6	...
Dado 2	1	5	1	3	1	5	3	6	4	3	...
B ocorreu?	N	N	N	N	Y	N	N	N	Y	Y	...
A ocorreu?	N	Y	N	N	N	Y	N	N	Y	Y	...

Tabela: Lançamentos duplos de dados

- Considere apenas as vezes em que A ocorreu: 13886 vezes
- *Dentre estas 13886 ocorrências*, verifique quantas vezes o evento B ocorreu: 5623 vezes

Intuição para a definição

- B : primeiro dado é um 6. A : soma das faces é maior que 8.

Repetição	1	2	3	4	5	6	7	8	9	10	...
Dado 1	2	5	5	2	6	4	2	1	6	6	...
Dado 2	1	5	1	3	1	5	3	6	4	3	...
B ocorreu?	N	N	N	N	Y	N	N	N	Y	Y	...
A ocorreu?	N	Y	N	N	N	Y	N	N	Y	Y	...

Tabela: Lançamentos duplos de dados

- Considere apenas as vezes em que A ocorreu: 13886 vezes
- *Dentre estas 13886 ocorrências*, verifique quantas vezes o evento B ocorreu: 5623 vezes
- É natural esperarmos $\mathbb{P}(B|A) \approx 5623/13886 = 0.405$. Por quê?

Intuição para a definição

- Repetindo: $\mathbb{P}(B|A) \approx \frac{5623}{13886} = 0.405$. Por quê?

Intuição para a definição

- Repetindo: $\mathbb{P}(B|A) \approx \frac{5623}{13886} = 0.405$. Por quê?
- Considerando apenas as 13886 vezes em que A ocorreu, verificamos qual a proporção de vezes que ocorreu B.

Intuição para a definição

- Repetindo: $\mathbb{P}(B|A) \approx \frac{5623}{13886} = 0.405$. Por quê?
- Considerando apenas as 13886 vezes em que A ocorreu, verificamos qual a proporção de vezes que ocorreu B.
- Esta é a maneira de estimar empiricamente, apenas com dados, o valor de $\mathbb{P}(B|A)$.

Intuição para a definição

- Repetindo: $\mathbb{P}(B|A) \approx \frac{5623}{13886} = 0.405$. Por quê?
- Considerando apenas as 13886 vezes em que A ocorreu, verificamos qual a proporção de vezes que ocorreu B.
- Esta é a maneira de estimar empiricamente, apenas com dados, o valor de $\mathbb{P}(B|A)$.
- Pela frequência relativa da ocorrência do evento B DADO QUE O EVENTO A ocorreu.

Intuição para a definição

- Repetindo: $\mathbb{P}(B|A) \approx \frac{5623}{13886} = 0.405$. Por quê?
- Considerando apenas as 13886 vezes em que A ocorreu, verificamos qual a proporção de vezes que ocorreu B.
- Esta é a maneira de estimar empiricamente, apenas com dados, o valor de $\mathbb{P}(B|A)$.
- Pela frequência relativa da ocorrência do evento B DADO QUE O EVENTO A ocorreu.
- Vamos agora estimar $\mathbb{P}(B|A)$ de outra forma: considerando o numerador e o denominador da definição.
- $\mathbb{P}(B|A) = \mathbb{P}(A \cap B)/\mathbb{P}(A)$.

Intuição para a definição

- Sabemos que $\mathbb{P}(A) \approx \frac{\text{n}^{\text{o}} \text{ de vezes que } A \text{ ocorreu}}{N}$

Intuição para a definição

- Sabemos que $\mathbb{P}(A) \approx \frac{\text{n}^\text{o} \text{ de vezes que } A \text{ ocorreu}}{N}$

$$\Rightarrow \mathbb{P}(A) \approx \frac{13886}{N} \qquad \text{ou} \qquad 13886 \approx N\,\mathbb{P}(A)$$

Intuição para a definição

- Sabemos que $\mathbb{P}(A) \approx \frac{\text{n}^{\text{o}} \text{ de vezes que } A \text{ ocorreu}}{N}$

$$\Rightarrow \mathbb{P}(A) \approx \frac{13886}{N} \qquad \text{ou} \qquad 13886 \approx N \, \mathbb{P}(A)$$

- Do mesmo modo, pela interpretação de probabilidade como frequência em longas repetições,

$$\mathbb{P}(A \cap B) \approx \frac{\text{n}^{\text{o}} \text{ de vezes em que } A \text{ e } B \text{ ocorrem}}{N}$$

Intuição para a definição

- Sabemos que $\mathbb{P}(A) \approx \frac{\text{n}^{\text{o}} \text{ de vezes que } A \text{ ocorreu}}{N}$

$$\Rightarrow \mathbb{P}(A) \approx \frac{13886}{N} \qquad \text{ou} \qquad 13886 \approx N\,\mathbb{P}(A)$$

- Do mesmo modo, pela interpretação de probabilidade como frequência em longas repetições,

$$\mathbb{P}(A \cap B) \approx \frac{\text{n}^{\text{o}} \text{ de vezes em que } A \text{ e } B \text{ ocorrem}}{N}$$

- Mas A e B ocorrem 5623 vezes em N
  (Separamos os 13886 casos em que A ocorreu e depois contamos os B dentro destes 13886 casos)

Intuição para a definição

- Sabemos que $\mathbb{P}(A) \approx \frac{\text{n}^{\text{o}} \text{ de vezes que } A \text{ ocorreu}}{N}$

$$\Rightarrow \mathbb{P}(A) \approx \frac{13886}{N} \qquad \text{ou} \qquad 13886 \approx N\,\mathbb{P}(A)$$

- Do mesmo modo, pela interpretação de probabilidade como frequência em longas repetições,

$$\mathbb{P}(A \cap B) \approx \frac{\text{n}^{\text{o}} \text{ de vezes em que } A \text{ e } B \text{ ocorrem}}{N}$$

- Mas A e B ocorrem 5623 vezes em N
(Separamos os 13886 casos em que A ocorreu e depois contamos os B dentro destes 13886 casos)
- Então $\mathbb{P}(A \cap B) \approx \frac{5623}{N} \Rightarrow N\,\mathbb{P}(A \cap B) \approx 5623$

Intuição para a definição

- Desse modo,

$$\mathbb{P}(B|A) \approx \frac{5623}{13886} \approx \frac{N\ \mathbb{P}(A \cap B)}{N\ \mathbb{P}(A)} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$$

Intuição para a definição

- Desse modo,

$$\mathbb{P}(B|A) \approx \frac{5623}{13886} \approx \frac{N\ \mathbb{P}(A \cap B)}{N\ \mathbb{P}(A)} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$$

- Nossa conclusão é que:
  se quisermos manter intacta nossa idéia de que a probabilidade de um evento é aproximadamente igual à sua frequência relativa numa longa série de repetições independentes,

Intuição para a definição

- Desse modo,

$$\mathbb{P}(B|A) \approx \frac{5623}{13886} \approx \frac{N\ \mathbb{P}(A \cap B)}{N\ \mathbb{P}(A)} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$$

- Nossa conclusão é que:
  se quisermos manter intacta nossa idéia de que a probabilidade de um evento é aproximadamente igual à sua frequência relativa numa longa série de repetições independentes,
- então a definição da probabilidade condicional $\mathbb{P}(B|A)$ TEM DE SER $\mathbb{P}(A \cap B)/\mathbb{P}(A)$.

Intuição para a definição

- Desse modo,

$$\mathbb{P}(B|A) \approx \frac{5623}{13886} \approx \frac{N\ \mathbb{P}(A \cap B)}{N\ \mathbb{P}(A)} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$$

- Nossa conclusão é que:
  se quisermos manter intacta nossa idéia de que a probabilidade de um evento é aproximadamente igual à sua frequência relativa numa longa série de repetições independentes,
- então a definição da probabilidade condicional $\mathbb{P}(B|A)$ TEM DE SER $\mathbb{P}(A \cap B)/\mathbb{P}(A)$.
- Nenhuma outra definição vai gerar resultados consistentes com os experimentos que fizemos.

Diagrama de Venn

- É comum representar eventos em diagramas de conjuntos de Venn.

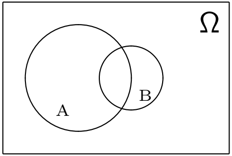
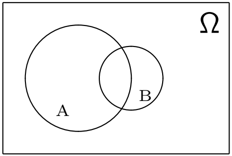


Diagrama de Venn

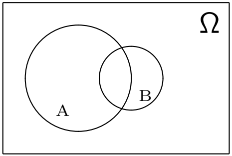
- É comum representar eventos em diagramas de conjuntos de Venn.



- Ω é o retângulo maior envolvente. Os eventos são figuras com tamanhos proporcionais à sua probabilidade.

Diagrama de Venn

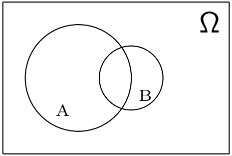
- É comum representar eventos em diagramas de conjuntos de Venn.



- Ω é o retângulo maior envolvente. Os eventos são figuras com tamanhos proporcionais à sua probabilidade.
- Qual o valor aproximado de $\mathbb{P}(A)$? USAR CORES NOS EVENTOS
  - $\mathbb{P}(A) \approx 0.90$?
  - $\mathbb{P}(A) \approx 1/4$?
  - $\mathbb{P}(A) \approx 1/8$?
  - $\mathbb{P}(A) \approx 0.01$?

Diagrama de Venn

- É comum representar eventos em diagramas de conjuntos de Venn.



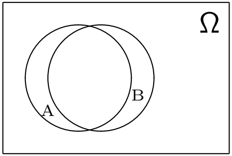
- Ω é o retângulo maior envolvente. Os eventos são figuras com tamanhos proporcionais à sua probabilidade.
- Qual o valor aproximado de $\mathbb{P}(A)$? USAR CORES NOS EVENTOS
  - $\mathbb{P}(A) \approx 0.90$?
  - $\mathbb{P}(A) \approx 1/4$?
  - $\mathbb{P}(A) \approx 1/8$?
  - $\mathbb{P}(A) \approx 0.01$?
- Com as mesmas opções, qual o valor aproximado de $\mathbb{P}(B)$?

Probabilidade condicional no diagrama de Venn

- Como enxergar a definição $\mathbb{P}(B|A) = \mathbb{P}(A \cap B)/\mathbb{P}(A)$ neste diagrama?

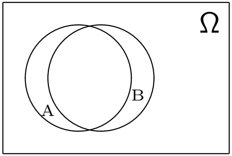
Probabilidade condicional no diagrama de Venn

- Como enxergar a definição $\mathbb{P}(B|A) = \mathbb{P}(A \cap B)/\mathbb{P}(A)$ neste diagrama?
- $\mathbb{P}(B|A)$ é o tamanho de $A \cap B$ *relativamente ao tamanho de* A.



Probabilidade condicional no diagrama de Venn

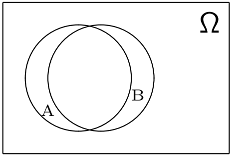
- Como enxergar a definição $\mathbb{P}(B|A) = \mathbb{P}(A \cap B)/\mathbb{P}(A)$ neste diagrama?
- $\mathbb{P}(B|A)$ é o tamanho de $A \cap B$ *relativamente ao tamanho de* A.



- Qual o valor aproximado de $\mathbb{P}(B|A)$?
  - $\mathbb{P}(B|A) \approx 0.85$?
  - $\mathbb{P}(B|A) \approx 1/3$?
  - $\mathbb{P}(B|A) \approx 1/8$?
  - $\mathbb{P}(B|A) \approx 0.05$?

Probabilidade condicional no diagrama de Venn

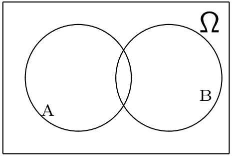
- Como enxergar a definição $\mathbb{P}(B|A) = \mathbb{P}(A \cap B)/\mathbb{P}(A)$ neste diagrama?
- $\mathbb{P}(B|A)$ é o tamanho de $A \cap B$ *relativamente ao tamanho de* A.



- Qual o valor aproximado de $\mathbb{P}(B|A)$?
  - $\mathbb{P}(B|A) \approx 0.85$?
  - $\mathbb{P}(B|A) \approx 1/3$?
  - $\mathbb{P}(B|A) \approx 1/8$?
  - $\mathbb{P}(B|A) \approx 0.05$?
- Temos $\mathbb{P}(B|A)$ bem maior que $\mathbb{P}(B) \approx 1/3$.

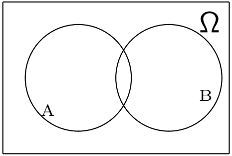
Probabilidade condicional no diagrama de Venn

- $\mathbb{P}(B|A) = \mathbb{P}(A \cap B)/\mathbb{P}(A)$ é o tamanho de $A \cap B$ *relativamente ao tamanho de* A.



Probabilidade condicional no diagrama de Venn

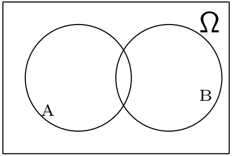
- $\mathbb{P}(B|A) = \mathbb{P}(A \cap B)/\mathbb{P}(A)$ é o tamanho de $A \cap B$ *relativamente ao tamanho de* A.



- Qual o valor aproximado de $\mathbb{P}(B|A)$?
  - $\mathbb{P}(B|A) \approx 0.85$?
  - $\mathbb{P}(B|A) \approx 1/3$?
  - $\mathbb{P}(B|A) \approx 1/8$?
  - $\mathbb{P}(B|A) \approx 0.05$?

Probabilidade condicional no diagrama de Venn

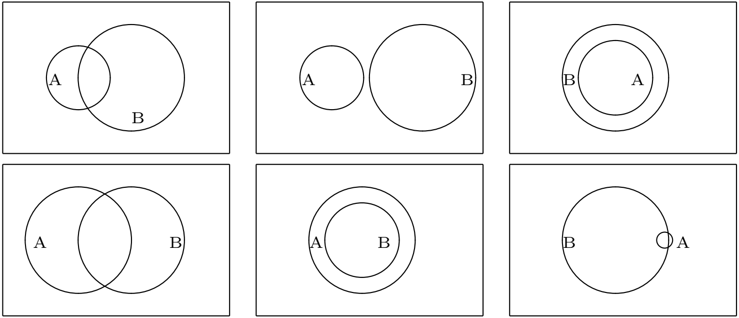
- $\mathbb{P}(B|A) = \mathbb{P}(A \cap B)/\mathbb{P}(A)$ é o tamanho de $A \cap B$ *relativamente ao tamanho de* A.



- Qual o valor aproximado de $\mathbb{P}(B|A)$?
  - $\mathbb{P}(B|A) \approx 0.85$?
  - $\mathbb{P}(B|A) \approx 1/3$?
  - $\mathbb{P}(B|A) \approx 1/8$?
  - $\mathbb{P}(B|A) \approx 0.05$?
- Temos $\mathbb{P}(B|A)$ bem menor que $\mathbb{P}(B) \approx 1/3$.

Probabilidade condicional no diagrama de Venn

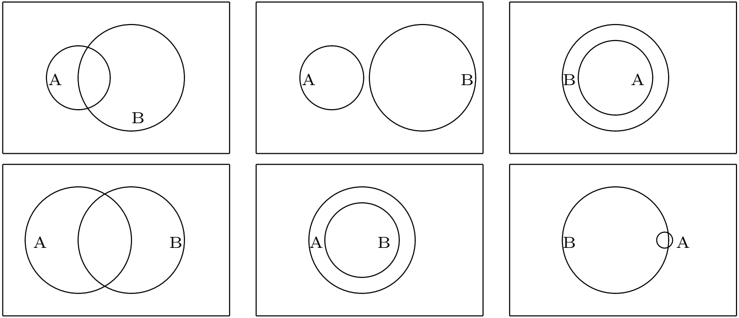
- Em todos os casos abaixo temos $\mathbb{P}(B) \approx 1/5$.
- Obtenha $\mathbb{P}(B|A)$ aproximadamente em cada caso:



- Diagrama $(1, 1)$ tem $\mathbb{P}(B|A) \approx \ldots$

Probabilidade condicional no diagrama de Venn

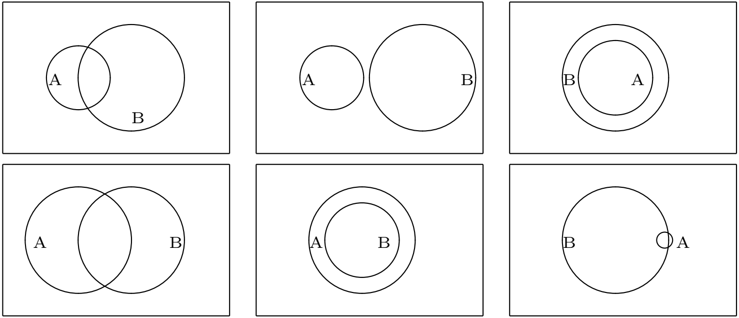
- Em todos os casos abaixo temos $\mathbb{P}(B) \approx 1/5$.
- Obtenha $\mathbb{P}(B|A)$ aproximadamente em cada caso:



- Diagrama $(1,1)$ tem $\mathbb{P}(B|A) \approx \ldots 0.40$.

Probabilidade condicional no diagrama de Venn

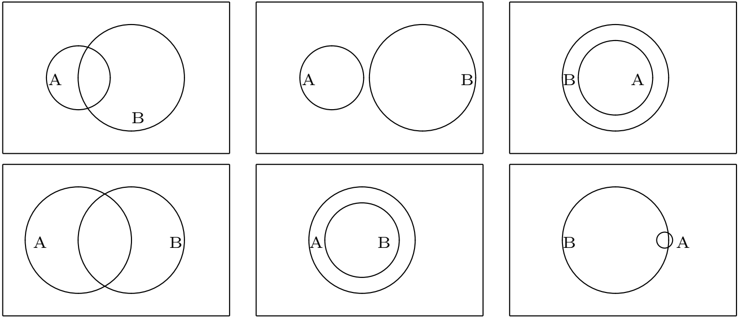
- Em todos os casos abaixo temos $\mathbb{P}(B) \approx 1/5$.
- Obtenha $\mathbb{P}(B|A)$ aproximadamente em cada caso:



- Diagrama $(1,1)$ tem $\mathbb{P}(B|A) \approx \ldots 0.40$.
- $(1,2)$: $\mathbb{P}(B|A) \ldots$

Probabilidade condicional no diagrama de Venn

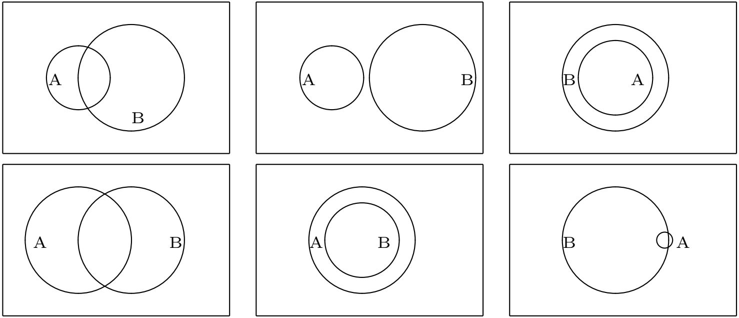
- Em todos os casos abaixo temos $\mathbb{P}(B) \approx 1/5$.
- Obtenha $\mathbb{P}(B|A)$ aproximadamente em cada caso:



- Diagrama $(1,1)$ tem $\mathbb{P}(B|A) \approx \ldots 0.40$.
- $(1,2) : \mathbb{P}(B|A) \ldots \ = 0$

Probabilidade condicional no diagrama de Venn

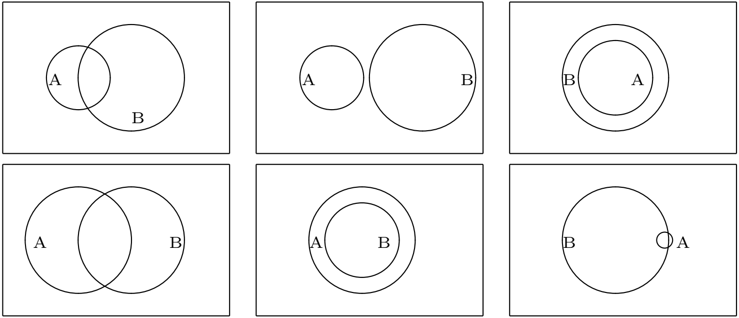
- Em todos os casos abaixo temos $\mathbb{P}(B) \approx 1/5$.
- Obtenha $\mathbb{P}(B|A)$ aproximadamente em cada caso:



- Diagrama $(1,1)$ tem $\mathbb{P}(B|A) \approx \ldots 0.40$.
- $(1,2) : \mathbb{P}(B|A) \ldots = 0$ $\quad$ $(1,3) : \mathbb{P}(B|A) \ldots$

Probabilidade condicional no diagrama de Venn

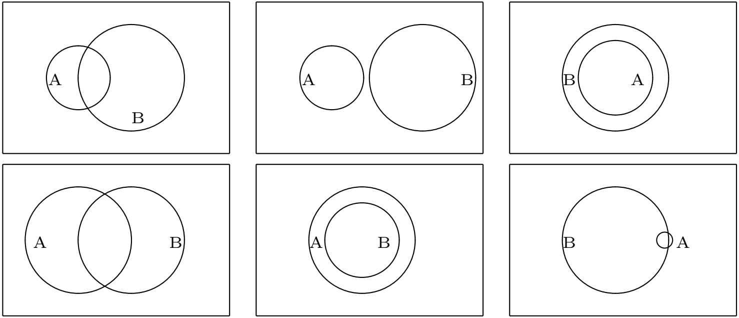
- Em todos os casos abaixo temos $\mathbb{P}(B) \approx 1/5$.
- Obtenha $\mathbb{P}(B|A)$ aproximadamente em cada caso:



- Diagrama $(1,1)$ tem $\mathbb{P}(B|A) \approx \ldots 0.40$.
- $(1,2)$: $\mathbb{P}(B|A) \ldots = 0$ $\quad$ $(1,3)$: $\mathbb{P}(B|A) \ldots = 1.0$.

Probabilidade condicional no diagrama de Venn

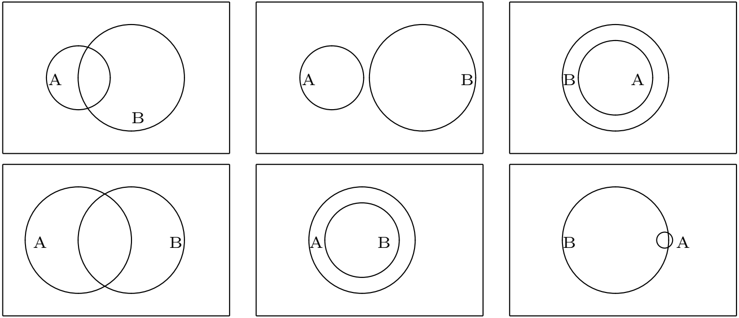
- Em todos os casos abaixo temos $\mathbb{P}(B) \approx 1/5$.
- Obtenha $\mathbb{P}(B|A)$ aproximadamente em cada caso:



- Diagrama $(1,1)$ tem $\mathbb{P}(B|A) \approx \ldots 0.40$.
- $(1,2) : \mathbb{P}(B|A) \ldots = 0$ $\quad$ $(1,3) : \mathbb{P}(B|A) \ldots = 1.0$.
- Diagrama $(2,1)$ tem $\mathbb{P}(B|A) \approx \ldots$

Probabilidade condicional no diagrama de Venn

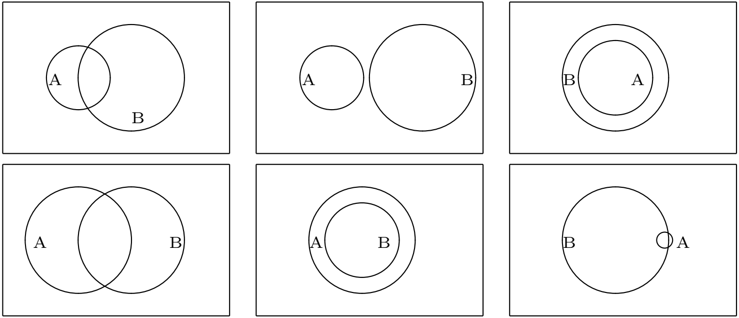
- Em todos os casos abaixo temos $\mathbb{P}(B) \approx 1/5$.
- Obtenha $\mathbb{P}(B|A)$ aproximadamente em cada caso:



- Diagrama $(1,1)$ tem $\mathbb{P}(B|A) \approx \ldots 0.40$.
- $(1,2) : \mathbb{P}(B|A) \ldots = 0$ $\quad (1,3) : \mathbb{P}(B|A) \ldots = 1.0$.
- Diagrama $(2,1)$ tem $\mathbb{P}(B|A) \approx \ldots 0.40$, como no diagrama $(1,1)$.

Probabilidade condicional no diagrama de Venn

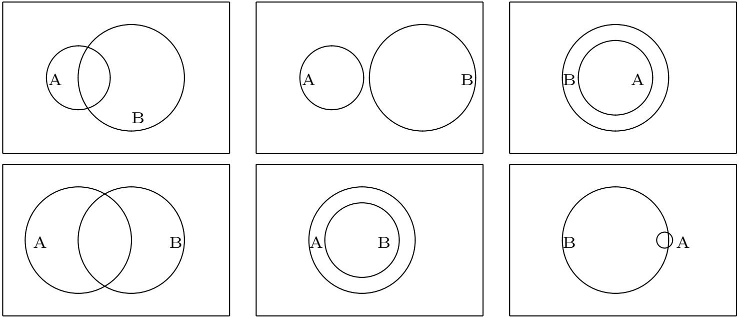
- Em todos os casos abaixo temos $\mathbb{P}(B) \approx 1/5$.
- Obtenha $\mathbb{P}(B|A)$ aproximadamente em cada caso:



- Diagrama $(1,1)$ tem $\mathbb{P}(B|A) \approx \ldots 0.40$.
- $(1,2)$: $\mathbb{P}(B|A) \ldots = 0$ $\qquad$ $(1,3)$: $\mathbb{P}(B|A) \ldots = 1.0$.
- Diagrama $(2,1)$ tem $\mathbb{P}(B|A) \approx \ldots 0.40$, como no diagrama $(1,1)$.
- $(2,2)$: $\mathbb{P}(B|A) \ldots$

Probabilidade condicional no diagrama de Venn

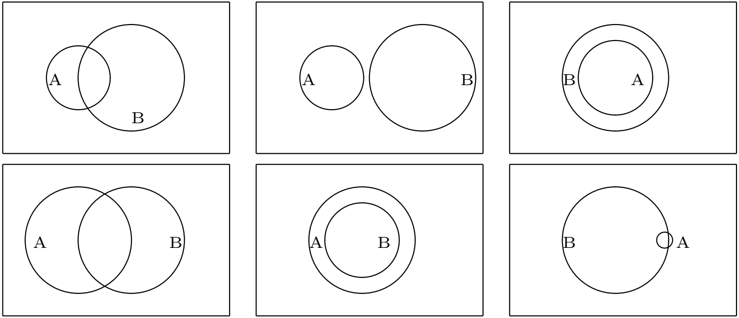
- Em todos os casos abaixo temos $\mathbb{P}(B) \approx 1/5$.
- Obtenha $\mathbb{P}(B|A)$ aproximadamente em cada caso:



- Diagrama $(1,1)$ tem $\mathbb{P}(B|A) \approx \ldots 0.40$.
- $(1,2) : \mathbb{P}(B|A) \ldots = 0$ $\quad$ $(1,3) : \mathbb{P}(B|A) \ldots = 1.0$.
- Diagrama $(2,1)$ tem $\mathbb{P}(B|A) \approx \ldots 0.40$, como no diagrama $(1,1)$.
- $(2,2) : \mathbb{P}(B|A) \ldots 0.85$

Probabilidade condicional no diagrama de Venn

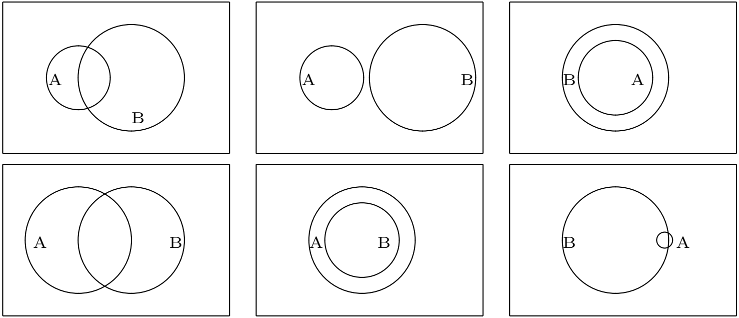
- Em todos os casos abaixo temos $\mathbb{P}(B) \approx 1/5$.
- Obtenha $\mathbb{P}(B|A)$ aproximadamente em cada caso:



- Diagrama $(1,1)$ tem $\mathbb{P}(B|A) \approx \ldots 0.40$.
- $(1,2)$: $\mathbb{P}(B|A) \ldots = 0$ $(1,3)$: $\mathbb{P}(B|A) \ldots = 1.0$.
- Diagrama $(2,1)$ tem $\mathbb{P}(B|A) \approx \ldots 0.40$, como no diagrama $(1,1)$.
- $(2,2)$: $\mathbb{P}(B|A) \ldots 0.85$ $(2,3)$: $\mathbb{P}(B|A) \ldots$

Probabilidade condicional no diagrama de Venn

- Em todos os casos abaixo temos $\mathbb{P}(B) \approx 1/5$.
- Obtenha $\mathbb{P}(B|A)$ aproximadamente em cada caso:



- Diagrama $(1,1)$ tem $\mathbb{P}(B|A) \approx \ldots 0.40$.
- $(1,2) : \mathbb{P}(B|A) \ldots = 0$ $\quad$ $(1,3) : \mathbb{P}(B|A) \ldots = 1.0$.
- Diagrama $(2,1)$ tem $\mathbb{P}(B|A) \approx \ldots 0.40$, como no diagrama $(1,1)$.
- $(2,2) : \mathbb{P}(B|A) \ldots 0.85$ $\quad$ $(2,3) : \mathbb{P}(B|A) \ldots 0.9$.

$\mathbb{P}(B|A)$ e $\mathbb{P}(B)$

- Se $A \subset B$ então $\mathbb{P}(B|A) = 1$. A informação de que A ocorreu torna *certa* a ocorrência de um resultado $\omega \in B$.

$\mathbb{P}(B|A)$ e $\mathbb{P}(B)$

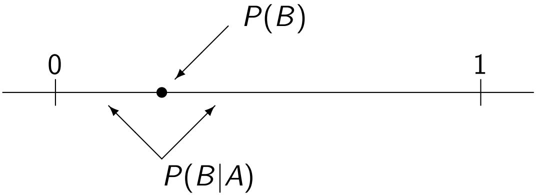
- Se $A \subset B$ então $\mathbb{P}(B|A) = 1$. A informação de que A ocorreu torna *certa* a ocorrência de um resultado $\omega \in B$.
- Se $A \cap B = \emptyset$ então $\mathbb{P}(B|A) = 0$. A informação de que A ocorreu torna *impossível* a ocorrência de qualquer $\omega \in B$

$\mathbb{P}(B|A)$ e $\mathbb{P}(B)$

- Se $A \subset B$ então $\mathbb{P}(B|A) = 1$. A informação de que A ocorreu torna *certa* a ocorrência de um resultado $\omega \in B$.
- Se $A \cap B = \emptyset$ então $\mathbb{P}(B|A) = 0$. A informação de que A ocorreu torna *impossível* a ocorrência de qualquer $\omega \in B$
- Estas são *situações extremas*: saber que A ocorreu leva a um conhecimento sem incerteza sobre a ocorrência de B.

$\mathbb{P}(B|A)$ e $\mathbb{P}(B)$

- Se $A \subset B$ então $\mathbb{P}(B|A) = 1$. A informação de que A ocorreu torna *certa* a ocorrência de um resultado $\omega \in B$.
- Se $A \cap B = \emptyset$ então $\mathbb{P}(B|A) = 0$. A informação de que A ocorreu torna *impossível* a ocorrência de qualquer $\omega \in B$
- Estas são *situações extremas*: saber que A ocorreu leva a um conhecimento sem incerteza sobre a ocorrência de B.
- Na maioria das vezes, saber que A ocorreu não vai eliminar a incerteza sobre a ocorrencia de B. Teremos $0 < P(B|A) < 1$



- Podemos ter $\mathbb{P}(B|A) > \mathbb{P}(B)$ ou $\mathbb{P}(B|A) < \mathbb{P}(B)$.

Eventos independentes

- Há um outro caso importante:

Eventos independentes

- Há um outro caso importante:
- quando saber que A ocorreu não tem qualquer influência na incerteza sobre a ocorrência de B.

Eventos independentes

- Há um outro caso importante:
- quando saber que A ocorreu não tem qualquer influência na incerteza sobre a ocorrência de B.
- Isto é, existem casos em que $\mathbb{P}(B|A) = \mathbb{P}(B)$

Eventos independentes

- Há um outro caso importante:
- quando saber que A ocorreu não tem qualquer influência na incerteza sobre a ocorrência de B.
- Isto é, existem casos em que $\mathbb{P}(B|A) = \mathbb{P}(B)$
- Dizemos que A e B são *eventos independentes*.

Eventos independentes

- Independência se, e somente se, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\ \mathbb{P}(B)$ pois

Eventos independentes

- Independência se, e somente se, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$ pois

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

Definição de Probabilidade Condicional

Se forem independentes

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

Como surge independência?

- Independência pode surgir de duas formas distintas.

Como surge independência?

- Independência pode surgir de duas formas distintas.
- Ela pode surgir porque nós *supomos* que o eventos são independentes.

Como surge independência?

- Independência pode surgir de duas formas distintas.
- Ela pode surgir porque nós *supomos* que o eventos são independentes.
- Por exemplo, pensando sobre o mecanismo físico envolvido, supomos que lançamentos sucessivos de uma moeda são independentes: a moeda não tem memória do que aconteceu.

Como surge independência?

- Independência pode surgir de duas formas distintas.
- Ela pode surgir porque nós *supomos* que o eventos são independentes.
- Por exemplo, pensando sobre o mecanismo físico envolvido, supomos que lançamentos sucessivos de uma moeda são independentes: a moeda não tem memória do que aconteceu.
- Assim $\mathbb{P}(\text{ Cara no 2o. } \mid \text{Cara no 1o. }) = \mathbb{P}(\text{ Cara no 2o. }) = 1/2$

Como surge independência?

- Independência pode surgir de duas formas distintas.
- Ela pode surgir porque nós *supomos* que o eventos são independentes.
- Por exemplo, pensando sobre o mecanismo físico envolvido, supomos que lançamentos sucessivos de uma moeda são independentes: a moeda não tem memória do que aconteceu.
- Assim $\mathbb{P}(\text{ Cara no 2o. } \mid \text{ Cara no 1o. }) = \mathbb{P}(\text{ Cara no 2o. }) = 1/2$
- A outra forma é quando *verificamos* matematicamente que $\mathbb{P}(A \cap B) = \mathbb{P}(A)\ \mathbb{P}(B)$ ou que $\mathbb{P}(B|A) = \mathbb{P}(B)$.

Como surge independência?

- Independência pode surgir de duas formas distintas.
- Ela pode surgir porque nós *supomos* que o eventos são independentes.
- Por exemplo, pensando sobre o mecanismo físico envolvido, supomos que lançamentos sucessivos de uma moeda são independentes: a moeda não tem memória do que aconteceu.
- Assim $\mathbb{P}(\text{ Cara no 2o. } \mid \text{ Cara no 1o. }) = \mathbb{P}(\text{ Cara no 2o. }) = 1/2$
- A outra forma é quando *verificamos* matematicamente que $\mathbb{P}(A \cap B) = \mathbb{P}(A)\ \mathbb{P}(B)$ ou que $\mathbb{P}(B|A) = \mathbb{P}(B)$.
- As vezes, não podemos intuir facilmente que A e B são independentes.

Como surge independência?

- Independência pode surgir de duas formas distintas.
- Ela pode surgir porque nós *supomos* que o eventos são independentes.
- Por exemplo, pensando sobre o mecanismo físico envolvido, supomos que lançamentos sucessivos de uma moeda são independentes: a moeda não tem memória do que aconteceu.
- Assim $\mathbb{P}(\text{ Cara no 2o. } | \text{ Cara no 1o. }) = \mathbb{P}(\text{ Cara no 2o. }) = 1/2$
- A outra forma é quando *verificamos* matematicamente que $\mathbb{P}(A \cap B) = \mathbb{P}(A)\, \mathbb{P}(B)$ ou que $\mathbb{P}(B|A) = \mathbb{P}(B)$.
- As vezes, não podemos intuir facilmente que A e B são independentes.
- Nestes casos, calculamos $\mathbb{P}(B|A)$ e $\mathbb{P}(B)$ e, voilá: se as probabilidades forem iguais, os eventos são independentes.

Exemplos

- Alguns exemplos óbvios de independência: em repetições de certos experimentos.

Exemplos

- Alguns exemplos óbvios de independência: em repetições de certos experimentos.
- Rolar um dado duas vezes e anotar o resultado

Exemplos

- Alguns exemplos óbvios de independência: em repetições de certos experimentos.
- Rolar um dado duas vezes e anotar o resultado
- Eventos relacionados apenas ao primeiro lançamento devem ser independentes de eventos relacionados apenas ao segundo lançamento do dado.

Exemplos

- Alguns exemplos óbvios de independência: em repetições de certos experimentos.
- Rolar um dado duas vezes e anotar o resultado
- Eventos relacionados apenas ao primeiro lançamento devem ser independentes de eventos relacionados apenas ao segundo lançamento do dado.
- Isto é intuitivo a partir de nossa experiência com este tipo de situação.

Exemplos

- Alguns exemplos óbvios de independência: em repetições de certos experimentos.
- Rolar um dado duas vezes e anotar o resultado
- Eventos relacionados apenas ao primeiro lançamento devem ser independentes de eventos relacionados apenas ao segundo lançamento do dado.
- Isto é intuitivo a partir de nossa experiência com este tipo de situação.
- As probabilidades devem se manter as mesmas: rolar um dado não o modifica fisicamente a ponto de afetar as probabilidades das 6 faces.

Exemplos

- Alguns exemplos óbvios de independência: em repetições de certos experimentos.
- Rolar um dado duas vezes e anotar o resultado
- Eventos relacionados apenas ao primeiro lançamento devem ser independentes de eventos relacionados apenas ao segundo lançamento do dado.
- Isto é intuitivo a partir de nossa experiência com este tipo de situação.
- As probabilidades devem se manter as mesmas: rolar um dado não o modifica fisicamente a ponto de afetar as probabilidades das 6 faces.
- Além disso, o dado não tem memória do que saiu antes de forma que um lançamento não afeta o seguinte.

Exemplos

- Alguns exemplos óbvios de independência: em repetições de certos experimentos.
- Rolar um dado duas vezes e anotar o resultado
- Eventos relacionados apenas ao primeiro lançamento devem ser independentes de eventos relacionados apenas ao segundo lançamento do dado.
- Isto é intuitivo a partir de nossa experiência com este tipo de situação.
- As probabilidades devem se manter as mesmas: rolar um dado não o modifica fisicamente a ponto de afetar as probabilidades das 6 faces.
- Além disso, o dado não tem memória do que saiu antes de forma que um lançamento não afeta o seguinte.
- Mas podemos verificar matematicamente esta intuição checando a validade da condição $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$.

Exemplos

- $\Omega = \{(1,1), \ldots, (6,6)\}$ com $\mathbb{P}(\omega) = 1/36$.

Exemplos

- $\Omega = \{(1,1), \ldots, (6,6)\}$ com $\mathbb{P}(\omega) = 1/36$.
- A: primeira rolagem é par.

Exemplos

- $\Omega = \{(1,1), \ldots, (6,6)\}$ com $\mathbb{P}(\omega) = 1/36$.
- A: primeira rolagem é par.
- $A = \{(2,1), \ldots, (2,6), (4,1), \ldots, (4,6), (6,1), \ldots, (6,6)\}$

Exemplos

- $\Omega = \{(1,1), \ldots, (6,6)\}$ com $\mathbb{P}(\omega) = 1/36$.
- A: primeira rolagem é par.
- $A = \{(2,1), \ldots, (2,6), (4,1), \ldots, (4,6), (6,1), \ldots, (6,6)\}$
- $\mathbb{P}(A) = 18/36 = 1/2$.
- B: segunda rolagem é divisível por 3.

Exemplos

- $\Omega = \{(1,1), \ldots, (6,6)\}$ com $\mathbb{P}(\omega) = 1/36$.
- A: primeira rolagem é par.
- $A = \{(2,1), \ldots, (2,6), (4,1), \ldots, (4,6), (6,1), \ldots, (6,6)\}$
- $\mathbb{P}(A) = 18/36 = 1/2$.
- B: segunda rolagem é divisível por 3.
- $B = \{(1,3), \ldots, (6,3), (1,6), \ldots, (6,6)\}$.

Exemplos

- $\Omega = \{(1,1), \ldots, (6,6)\}$ com $\mathbb{P}(\omega) = 1/36$.
- A: primeira rolagem é par.
- $A = \{(2,1), \ldots, (2,6), (4,1), \ldots, (4,6), (6,1), \ldots, (6,6)\}$
- $\mathbb{P}(A) = 18/36 = 1/2$.
- B: segunda rolagem é divisível por 3.
- $B = \{(1,3), \ldots, (6,3), (1,6), \ldots, (6,6)\}$.
- $\mathbb{P}(B) = 12/36 = 1/3$.

Exemplos

- $\Omega = \{(1,1), \ldots, (6,6)\}$ com $\mathbb{P}(\omega) = 1/36$.
- A: primeira rolagem é par.
- $A = \{(2,1), \ldots, (2,6), (4,1), \ldots, (4,6), (6,1), \ldots, (6,6)\}$
- $\mathbb{P}(A) = 18/36 = 1/2$.
- B: segunda rolagem é divisível por 3.
- $B = \{(1,3), \ldots, (6,3), (1,6), \ldots, (6,6)\}$.
- $\mathbb{P}(B) = 12/36 = 1/3$.
- $A \cap B = \{(2,3), (4,3), (6,3), (2,6), (4,6), (6,6)\}$

Exemplos

- $\Omega = \{(1,1), \ldots, (6,6)\}$ com $\mathbb{P}(\omega) = 1/36$.
- A: primeira rolagem é par.
- $A = \{(2,1), \ldots, (2,6), (4,1), \ldots, (4,6), (6,1), \ldots, (6,6)\}$
- $\mathbb{P}(A) = 18/36 = 1/2$.
- B: segunda rolagem é divisível por 3.
- $B = \{(1,3), \ldots, (6,3), (1,6), \ldots, (6,6)\}$.
- $\mathbb{P}(B) = 12/36 = 1/3$.
- $A \cap B = \{(2,3), (4,3), (6,3), (2,6), (4,6), (6,6)\}$
- Como esperado, A e B são independnetes:

$$\mathbb{P}(A \cap B) = 6/36 = 1/6 = \mathbb{P}(A)\ \mathbb{P}(B)$$

Exemplo menos óbvio

- Rola-se um dado bem equilibrado uma única vez.

Exemplo menos óbvio

- Rola-se um dado bem equilibrado uma única vez.
- Seja $A = \{2, 4, 6\}$ e $B = \{1, 2, 3, 4\}$.

Exemplo menos óbvio

- Rola-se um dado bem equilibrado uma única vez.
- Seja $A = \{2, 4, 6\}$ e $B = \{1, 2, 3, 4\}$.
- $\mathbb{P}(A) = 1/2$ e $\mathbb{P}(B) = 2/3$.

Exemplo menos óbvio

- Rola-se um dado bem equilibrado uma única vez.
- Seja $A = \{2, 4, 6\}$ e $B = \{1, 2, 3, 4\}$.
- $\mathbb{P}(A) = 1/2$ e $\mathbb{P}(B) = 2/3$.
- $A \cap B = \{2, 4\}$

Exemplo menos óbvio

- Rola-se um dado bem equilibrado uma única vez.
- Seja $A = \{2, 4, 6\}$ e $B = \{1, 2, 3, 4\}$.
- $\mathbb{P}(A) = 1/2$ e $\mathbb{P}(B) = 2/3$.
- $A \cap B = \{2, 4\}$
- $\mathbb{P}(A \cap B) = 1/3$.

Exemplo menos óbvio

- Rola-se um dado bem equilibrado uma única vez.
- Seja $A = \{2, 4, 6\}$ e $B = \{1, 2, 3, 4\}$.
- $\mathbb{P}(A) = 1/2$ e $\mathbb{P}(B) = 2/3$.
- $A \cap B = \{2, 4\}$
- $\mathbb{P}(A \cap B) = 1/3$.
- Como

$$\mathbb{P}(A \cap B) = \frac{1}{3} = \frac{1}{2}\frac{2}{3} = \mathbb{P}(A)\ \mathbb{P}(B)\,,$$

Exemplo menos óbvio

- Rola-se um dado bem equilibrado uma única vez.
- Seja $A = \{2, 4, 6\}$ e $B = \{1, 2, 3, 4\}$.
- $\mathbb{P}(A) = 1/2$ e $\mathbb{P}(B) = 2/3$.
- $A \cap B = \{2, 4\}$
- $\mathbb{P}(A \cap B) = 1/3$.
- Como
$$\mathbb{P}(A \cap B) = \frac{1}{3} = \frac{1}{2}\frac{2}{3} = \mathbb{P}(A)\,\mathbb{P}(B)\,,$$
- Os eventos A e B são independentes. É muito difícil alguém conseguir enxergar isto intuitivamente.

Independência no diagrama de Venn

- Se A e B são eventos disjuntos num diagrama de Venn, eles não são independentes.

Independência no diagrama de Venn

- Se A e B são eventos disjuntos num diagrama de Venn, eles não são independentes.
  - Se são disjuntos então $A \cap B = \emptyset$ e portanto $\mathbb{P}(A \cap B) = 0$.

Independência no diagrama de Venn

- Se A e B são eventos disjuntos num diagrama de Venn, eles não são independentes.
  - Se são disjuntos então $A \cap B = \emptyset$ e portanto $\mathbb{P}(A \cap B) = 0$.
  - Se $\mathbb{P}(A)$ e $\mathbb{P}(B)$ são > 0 então $0 = \mathbb{P}(A \cap B) \neq \mathbb{P}(A)\,\mathbb{P}(B)$.

Independência no diagrama de Venn

- Se A e B são eventos disjuntos num diagrama de Venn, eles não são independentes.
  - Se são disjuntos então $A \cap B = \emptyset$ e portanto $\mathbb{P}(A \cap B) = 0$.
  - Se $\mathbb{P}(A)$ e $\mathbb{P}(B)$ são > 0 então $0 = \mathbb{P}(A \cap B) \neq \mathbb{P}(A)\,\mathbb{P}(B)$.
- Se $A \subset B$ então $\mathbb{P}(A \cap B) = \mathbb{P}(A) \neq \mathbb{P}(A)\,\mathbb{P}(B)$ e portanto A e B não são independentes.

Independência no diagrama de Venn

- Se A e B são eventos disjuntos num diagrama de Venn, eles não são independentes.
  - Se são disjuntos então $A \cap B = \emptyset$ e portanto $\mathbb{P}(A \cap B) = 0$.
  - Se $\mathbb{P}(A)$ e $\mathbb{P}(B)$ são > 0 então $0 = \mathbb{P}(A \cap B) \neq \mathbb{P}(A)\, \mathbb{P}(B)$.
- Se $A \subset B$ então $\mathbb{P}(A \cap B) = \mathbb{P}(A) \neq \mathbb{P}(A)\, \mathbb{P}(B)$ e portanto A e B não são independentes.
- Exceto nestes dois casos, é difícil verificar visualmente se A e B são independentes num diagrama de Venn.

Independência no diagrama de Venn

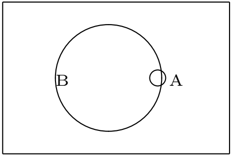
- Se A e B são eventos disjuntos num diagrama de Venn, eles não são independentes.
  - Se são disjuntos então $A \cap B = \emptyset$ e portanto $\mathbb{P}(A \cap B) = 0$.
  - Se $\mathbb{P}(A)$ e $\mathbb{P}(B)$ são > 0 então $0 = \mathbb{P}(A \cap B) \neq \mathbb{P}(A)\,\mathbb{P}(B)$.
- Se $A \subset B$ então $\mathbb{P}(A \cap B) = \mathbb{P}(A) \neq \mathbb{P}(A)\,\mathbb{P}(B)$ e portanto A e B não são independentes.
- Exceto nestes dois casos, é difícil verificar visualmente se A e B são independentes num diagrama de Venn.
- Teríamos de ser capazes de ver que o tamanho de $A \cap B$ relativamente a Ω é igual ao produto das proporções dos tamanhos de A e de B

Independência no diagrama de Venn

- Se A e B são eventos disjuntos num diagrama de Venn, eles não são independentes.
  - Se são disjuntos então $A \cap B = \emptyset$ e portanto $\mathbb{P}(A \cap B) = 0$.
  - Se $\mathbb{P}(A)$ e $\mathbb{P}(B)$ são > 0 então $0 = \mathbb{P}(A \cap B) \neq \mathbb{P}(A)\, \mathbb{P}(B)$.
- Se $A \subset B$ então $\mathbb{P}(A \cap B) = \mathbb{P}(A) \neq \mathbb{P}(A)\, \mathbb{P}(B)$ e portanto A e B não são independentes.
- Exceto nestes dois casos, é difícil verificar visualmente se A e B são independentes num diagrama de Venn.
- Teríamos de ser capazes de ver que o tamanho de $A \cap B$ relativamente a Ω é igual ao produto das proporções dos tamanhos de A e de B
- Mas...se $\mathbb{P}(B|A)$ for muito diferente de $\mathbb{P}(B)$ poderemos dizer com segurança que A e B não são independentes.

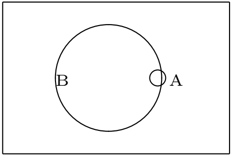
Independência no diagrama de Venn

- Por exemplo, sem fazer nenhuma conta podemos dizer que A e B não são independentes neste caso:



Independência no diagrama de Venn

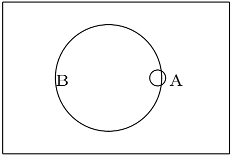
- Por exemplo, sem fazer nenhuma conta podemos dizer que A e B não são independentes neste caso:



- Visualmente é óbvio que $\mathbb{P}(B) \approx 1/3$ mas que $\mathbb{P}(B|A) \approx 1$.

Independência no diagrama de Venn

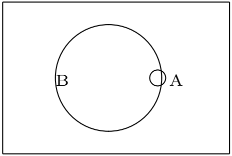
- Por exemplo, sem fazer nenhuma conta podemos dizer que A e B não são independentes neste caso:



- Visualmente é óbvio que $\mathbb{P}(B) \approx 1/3$ mas que $\mathbb{P}(B|A) \approx 1$.
- Portanto, a ocorrência de A aumenta as chances da ocorrência de B.

Independência no diagrama de Venn

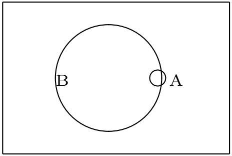
- Por exemplo, sem fazer nenhuma conta podemos dizer que A e B não são independentes neste caso:



- Visualmente é óbvio que $\mathbb{P}(B) \approx 1/3$ mas que $\mathbb{P}(B|A) \approx 1$.
- Portanto, a ocorrência de A aumenta as chances da ocorrência de B.
- Explicação: A é um evento raro pois $\mathbb{P}(A) \approx 0$. Entretanto, a maior parte de A está em B. Se o raro evento A ocorrer, é altamente provável que seja um dos $\omega \in A \cap B$.

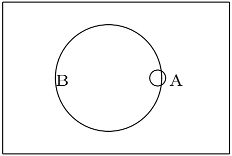
$\mathbb{P}(A|B)$ e $\mathbb{P}(B|A)$

- $\mathbb{P}(A|B)$ e $\mathbb{P}(B|A)$ podem ser completamente diferentes.



$\mathbb{P}(A|B)$ e $\mathbb{P}(B|A)$

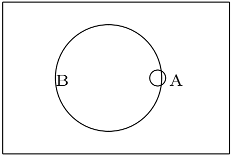
- $\mathbb{P}(A|B)$ e $\mathbb{P}(B|A)$ podem ser completamente diferentes.



- $\mathbb{P}(B|A) \approx 1$ mas $\mathbb{P}(A|B) \approx 1/25$.

$\mathbb{P}(A|B)$ e $\mathbb{P}(B|A)$

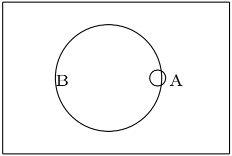
- $\mathbb{P}(A|B)$ e $\mathbb{P}(B|A)$ podem ser completamente diferentes.



- $\mathbb{P}(B|A) \approx 1$ mas $\mathbb{P}(A|B) \approx 1/25$.
- $\mathbb{P}(\text{ ser Drácula } | \text{ não dorme a noite }) \approx 0$

$\mathbb{P}(A|B)$ e $\mathbb{P}(B|A)$

- $\mathbb{P}(A|B)$ e $\mathbb{P}(B|A)$ podem ser completamente diferentes.



- $\mathbb{P}(B|A) \approx 1$ mas $\mathbb{P}(A|B) \approx 1/25$.
- $\mathbb{P}(\text{ ser Drácula } | \text{ não dorme a noite }) \approx 0$
- $\mathbb{P}(\text{ não dorme a noite } | \text{ ser Drácula }) \approx 1$

Regra de Bayes

- Existe uma relação matemática muito simples entre $\mathbb{P}(A|B)$ e $\mathbb{P}(B|A)$.

$$
\begin{aligned}
\mathbb{P}(B|A) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \Rightarrow \mathbb{P}(A \cap B) = \mathbb{P}(B|A)\ \mathbb{P}(A) \\
\mathbb{P}(A|B) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \Rightarrow \mathbb{P}(A \cap B) = \mathbb{P}(A|B)\ \mathbb{P}(B)
\end{aligned}
$$

Regra de Bayes

- Existe uma relação matemática muito simples entre $\mathbb{P}(A|B)$ e $\mathbb{P}(B|A)$.

$$\begin{aligned}\mathbb{P}(B|A) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \Rightarrow \mathbb{P}(A \cap B) = \mathbb{P}(B|A)\ \mathbb{P}(A) \\ \mathbb{P}(A|B) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \Rightarrow \mathbb{P}(A \cap B) = \mathbb{P}(A|B)\ \mathbb{P}(B)\end{aligned}$$

- Igualando as duas expressões para $\mathbb{P}(A \cap B)$ temos

$$\mathbb{P}(B|A)\ \mathbb{P}(A) = \mathbb{P}(A|B)\ \mathbb{P}(B)$$

Regra de Bayes

- Existe uma relação matemática muito simples entre $\mathbb{P}(A|B)$ e $\mathbb{P}(B|A)$.

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

- Igualando as duas expressões para $\mathbb{P}(A \cap B)$ temos

$$\mathbb{P}(B|A)\ \mathbb{P}(A) = \mathbb{P}(A|B)\ \mathbb{P}(B)$$

- Ou ainda

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

Regra de Bayes

- O principal uso da regra de Bayes é quando temos uma das probabilidades condicionais, digamos $\mathbb{P}(A|B)$, e queremos calcular a inversa: $\mathbb{P}(B|A)$.

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

Regra de Bayes

- O principal uso da regra de Bayes é quando temos uma das probabilidades condicionais, digamos $\mathbb{P}(A|B)$, e queremos calcular a inversa: $\mathbb{P}(B|A)$.

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

- Para isto precisamos também das probabilidades não-condicionais $\mathbb{P}(A)$ e $\mathbb{P}(B)$.

Teste diagnóstico

- A população brasileira deve ser rastreada com um teste para detecção do vírus HIV?

Teste diagnóstico

- A população brasileira deve ser rastreada com um teste para detecção do vírus HIV?
- Acredita-se que 0.1% são HIV+ $\Rightarrow \approx$ 200 mil dentre 200 milhões.

Teste diagnóstico

- A população brasileira deve ser rastreada com um teste para detecção do vírus HIV?
- Acredita-se que 0.1% são HIV+ $\Rightarrow \approx$ 200 mil dentre 200 milhões.
- Teste diagnóstico para testar a presença do vírus: não é perfeito.

Teste diagnóstico

- A população brasileira deve ser rastreada com um teste para detecção do vírus HIV?
- Acredita-se que 0.1% são HIV+ $\Rightarrow \approx$ 200 mil dentre 200 milhões.
- Teste diagnóstico para testar a presença do vírus: não é perfeito.
- Tabela de confusão:

	Resultado do Teste	
Vírus?	$T+$	$T-$
$V+$	ok	erro
$V-$	erro	ok

Teste diagnóstico

- A população brasileira deve ser rastreada com um teste para detecção do vírus HIV?
- Acredita-se que 0.1% são HIV+ $\Rightarrow \approx$ 200 mil dentre 200 milhões.
- Teste diagnóstico para testar a presença do vírus: não é perfeito.
- Tabela de confusão:

Vírus?	Resultado do Teste $T+$	$T-$
$V+$	ok	erro
$V-$	erro	ok

- Um paciente recebe o resultados $T+$. A questão é: ele é $V+$ ou aconteceu um erro?

Teste diagnóstico

- A população brasileira deve ser rastreada com um teste para detecção do vírus HIV?
- Acredita-se que 0.1% são HIV+ $\Rightarrow \approx$ 200 mil dentre 200 milhões.
- Teste diagnóstico para testar a presença do vírus: não é perfeito.
- Tabela de confusão:

	Resultado do Teste	
Vírus?	$T+$	$T-$
$V+$	ok	erro
$V-$	erro	ok

- Um paciente recebe o resultados $T+$. A questão é: ele é $V+$ ou aconteceu um erro?
  - O principal problema médico é calcular $\mathbb{P}(V+ \mid T+)$.

Teste diagnóstico

- A população brasileira deve ser rastreada com um teste para detecção do vírus HIV?
- Acredita-se que 0.1% são HIV+ $\Rightarrow \approx$ 200 mil dentre 200 milhões.
- Teste diagnóstico para testar a presença do vírus: não é perfeito.
- Tabela de confusão:

Vírus?	Resultado do Teste $T+$	$T-$
$V+$	ok	erro
$V-$	erro	ok

- Um paciente recebe o resultados $T+$. A questão é: ele é $V+$ ou aconteceu um erro?
  - O principal problema médico é calcular $\mathbb{P}(V+|T+)$.
- Como obter isto? Pela regra de Bayes pois temos $\mathbb{P}(T+|V+)$

Sensitividade e Especificidade

- Aplica-se o teste diagnóstico em dois grandes grupos de indivíduos:

Sensitividade e Especificidade

- Aplica-se o teste diagnóstico em dois grandes grupos de indivíduos:
  - um em que sabidamente todos possuem o vírus HIV

Sensitividade e Especificidade

- Aplica-se o teste diagnóstico em dois grandes grupos de indivíduos:
  - um em que sabidamente todos possuem o vírus HIV
  - outro em que que sabidamente eles não possuem o vírus HIV

Sensitividade e Especificidade

- Aplica-se o teste diagnóstico em dois grandes grupos de indivíduos:
  - um em que sabidamente todos possuem o vírus HIV
  - outro em que que sabidamente eles não possuem o vírus HIV
- A probabilidade de qualquer evento A ocorrera no primeiro grupo será $\mathbb{P}(A|V+)$.

Sensitividade e Especificidade

- Aplica-se o teste diagnóstico em dois grandes grupos de indivíduos:
  - um em que sabidamente todos possuem o vírus HIV
  - outro em que que sabidamente eles não possuem o vírus HIV
- A probabilidade de qualquer evento A ocorrera no primeiro grupo será $\mathbb{P}(A|V+)$.
- No segundo grupo será $\mathbb{P}(A|V-)$.

Sensitividade e Especificidade

- Aplica-se o teste diagnóstico em dois grandes grupos de indivíduos:
  - um em que sabidamente todos possuem o vírus HIV
  - outro em que que sabidamente eles não possuem o vírus HIV
- A probabilidade de qualquer evento A ocorrera no primeiro grupo será $\mathbb{P}(A|V+)$.
- No segundo grupo será $\mathbb{P}(A|V-)$.
- Com base na frequência daqueles que respondem $T+$ em cada grupo o laboratório que produz o teste estima:
- $\boxed{Sensitividade}$: $\mathbb{P}(T+|V+) \approx 0.99\%$

Sensitividade e Especificidade

- Aplica-se o teste diagnóstico em dois grandes grupos de indivíduos:
  - um em que sabidamente todos possuem o vírus HIV
  - outro em que que sabidamente eles não possuem o vírus HIV
- A probabilidade de qualquer evento A ocorrera no primeiro grupo será $\mathbb{P}(A|V+)$.
- No segundo grupo será $\mathbb{P}(A|V-)$.
- Com base na frequência daqueles que respondem $T+$ em cada grupo o laboratório que produz o teste estima:
- $\boxed{Sensitividade}$: $\mathbb{P}(T+|V+) \approx 0.99\%$
- $\boxed{Especificidade}$: $\mathbb{P}(T-|V-) \approx 0.95\%$

Sensitividade e Especificidade

- Repetindo:
- $\boxed{\textit{Sensitividade}}$: $\mathbb{P}(T+|V+) \approx 0.99\%$
- $\boxed{\textit{Especificidade}}$: $\mathbb{P}(T-|V-) \approx 0.95\%$

Sensitividade e Especificidade

- Repetindo:
- $\boxed{\textit{Sensitividade}}$: $\mathbb{P}(T+|V+) \approx 0.99\%$
- $\boxed{\textit{Especificidade}}$: $\mathbb{P}(T-|V-) \approx 0.95\%$
- Quanto maior, melhor. Idealmente, gostaríamos que fossem iguais a 1. Mas na prática, testes de diagnósticos podem cometer erros.

Sensitividade e Especificidade

- Repetindo:
- $\boxed{\textit{Sensitividade}}$: $\mathbb{P}(T + | V+) \approx 0.99\%$
- $\boxed{\textit{Especificidade}}$: $\mathbb{P}(T - | V-) \approx 0.95\%$
- Quanto maior, melhor. Idealmente, gostaríamos que fossem iguais a 1. Mas na prática, testes de diagnósticos podem cometer erros.
- Motivo para os nomes:
  - O teste é sensível à presença do vírus: se o vírus estiver presente, o teste é $+$?
  - O teste é específico para o vírus HIV: se o paciente tiver qualquer outra coisa que não seja o vírus, o teste não deveria dar positivo.

Falso positivo e falso negativo

- $\mathbb{P}(T+|V+) = 0.99$ (sensibilidade) e $\mathbb{P}(T-|V-) = 0.95$ (especificidade)

Falso positivo e falso negativo

- $\mathbb{P}(T+|V+) = 0.99$ (sensibilidade) e $\mathbb{P}(T-|V-) = 0.95$ (especificidade)
- As probabilidades complementares estão associadas a erros de diagnóstico e os médicos usam dois termos para eles:
  - Falso positivo (FP): $T+$ para um paciente que é $V-$
  - Falso negativo (FN): $T-$ para um paciente que é $V+$
- As probabilidades de FP e FN são obtidas diretamente da sensibilidade e especificidade.

Falso positivo e falso negativo

- $\mathbb{P}(T+|V+) = 0.99$ (sensibilidade) e $\mathbb{P}(T-|V-) = 0.95$ (especificidade)
- As probabilidades complementares estão associadas a erros de diagnóstico e os médicos usam dois termos para eles:
  - Falso positivo (FP): $T+$ para um paciente que é $V-$
  - Falso negativo (FN): $T-$ para um paciente que é $V+$
- As probabilidades de FP e FN são obtidas diretamente da sensibilidade e especificidade.
- $\mathbb{P}(FP) = \mathbb{P}(T+|V-) = 0.05 = 1 - 0.95 = 1 - \text{especificidade}$

Falso positivo e falso negativo

- $\mathbb{P}(T+|V+) = 0.99$ (sensibilidade) e $\mathbb{P}(T-|V-) = 0.95$ (especificidade)
- As probabilidades complementares estão associadas a erros de diagnóstico e os médicos usam dois termos para eles:
  - Falso positivo (FP): $T+$ para um paciente que é $V-$
  - Falso negativo (FN): $T-$ para um paciente que é $V+$
- As probabilidades de FP e FN são obtidas diretamente da sensibilidade e especificidade.
- $\mathbb{P}(FP) = \mathbb{P}(T+|V-) = 0.05 = 1 - 0.95 = 1 - \text{especificidade}$
- $\mathbb{P}(FN) = \mathbb{P}(T-|V+) = 0.01 = 1 - 0.99 = 1 - \text{sensibilidade}$

Falso positivo e falso negativo

- A partir das frequências na tabela de confusão, podemos estimar essas probabilidades:

Vírus?	Resultado do Teste $T+$	$T-$	Total
$V+$	sens $\mathbb{P}(T+\mid V+)$		1.0
$V-$		esp $\mathbb{P}(T-\mid V-)$	1.0

Falso positivo e falso negativo

- A partir das frequências na tabela de confusão, podemos estimar essas probabilidades:

Vírus?	Resultado do Teste $T+$	$T-$	Total
$V+$	sens $\mathbb{P}(T+ \mid V+)$	1 - sens $\mathbb{P}(FN) = \mathbb{P}(T- \mid V+)$	1.0
$V-$	1-esp $\mathbb{P}(FP) = \mathbb{P}(T+ \mid V-)$	esp $\mathbb{P}(T- \mid V-)$	1.0

Probabilidades inversas

- Mas não queremos apenas $\mathbb{P}(FP) = \mathbb{P}(T + | V-)$ e $\mathbb{P}(FN) = \mathbb{P}(T - | V+)$

Probabilidades inversas

- Mas não queremos apenas $\mathbb{P}(FP) = \mathbb{P}(T + | V-)$ e $\mathbb{P}(FN) = \mathbb{P}(T - | V+)$
- Mais importante é calcular as probabilidades inversas.

Probabilidades inversas

- Mas não queremos apenas $\mathbb{P}(FP) = \mathbb{P}(T + |V-)$ e $\mathbb{P}(FN) = \mathbb{P}(T - |V+)$
- Mais importante é calcular as probabilidades inversas.
- O médico tem em mãos o resultado $T+$ do exame.

Probabilidades inversas

- Mas não queremos apenas $\mathbb{P}(FP) = \mathbb{P}(T + | V-)$ e $\mathbb{P}(FN) = \mathbb{P}(T - | V+)$
- Mais importante é calcular as probabilidades inversas.
- O médico tem em mãos o resultado $T+$ do exame.
- Dado que ele tem este resultado $T+$, qual a probabilidade de que o paciente tenha o vírus?

Probabilidades inversas

- Mas não queremos apenas $\mathbb{P}(FP) = \mathbb{P}(T + | V-)$ e $\mathbb{P}(FN) = \mathbb{P}(T - | V+)$
- Mais importante é calcular as probabilidades inversas.
- O médico tem em mãos o resultado $T+$ do exame.
- Dado que ele tem este resultado $T+$, qual a probabilidade de que o paciente tenha o vírus?
- Isto é, qual o valor de $\mathbb{P}(V + | T+)$?

Probabilidades inversas

- Mas não queremos apenas $\mathbb{P}(FP) = \mathbb{P}(T + | V-)$ e $\mathbb{P}(FN) = \mathbb{P}(T - | V+)$
- Mais importante é calcular as probabilidades inversas.
- O médico tem em mãos o resultado $T+$ do exame.
- Dado que ele tem este resultado $T+$, qual a probabilidade de que o paciente tenha o vírus?
- Isto é, qual o valor de $\mathbb{P}(V + | T+)$?
- Do mesmo modo, queremos saber $\mathbb{P}(V - | T-)$

Probabilidades inversas

- Mas não queremos apenas $\mathbb{P}(FP) = \mathbb{P}(T + | V-)$ e $\mathbb{P}(FN) = \mathbb{P}(T - | V+)$
- Mais importante é calcular as probabilidades inversas.
- O médico tem em mãos o resultado $T+$ do exame.
- Dado que ele tem este resultado $T+$, qual a probabilidade de que o paciente tenha o vírus?
- Isto é, qual o valor de $\mathbb{P}(V + | T+)$?
- Do mesmo modo, queremos saber $\mathbb{P}(V - | T-)$
- De posse de uma estimativa de $\mathbb{P}(V+)$, usamos a regra de Bayes para obter estas probabilidades inversas.

Probabilidades inversas

- Temos $\mathbb{P}(V+) = 0.001$, uma estimativa grosseira. Esta é a estimativa da prevalência do vírus na população em geral.

Probabilidades inversas

- Temos $\mathbb{P}(V+) = 0.001$, uma estimativa grosseira. Esta é a estimativa da prevalência do vírus na população em geral.
- Se não soubermos este valor, podemos calcular as probabilidades com diversos cenários plausíveis para $\mathbb{P}(V+)$ e ver como as probabilidades se modificam (talvez elas não mudem muito).

Probabilidades inversas

- Temos $\mathbb{P}(V+) = 0.001$, uma estimativa grosseira. Esta é a estimativa da prevalência do vírus na população em geral.
- Se não soubermos este valor, podemos calcular as probabilidades com diversos cenários plausíveis para $\mathbb{P}(V+)$ e ver como as probabilidades se modificam (talvez elas não mudem muito).
- Pela regra de Bayes:

$$\mathbb{P}(V+|T+) = \frac{\mathbb{P}(T+|V+)\mathbb{P}(V+)}{\mathbb{P}(T+)} = \frac{0.99 * 0.001}{\mathbb{P}(T+)}$$

Probabilidades inversas

- Temos $\mathbb{P}(V+) = 0.001$, uma estimativa grosseira. Esta é a estimativa da prevalência do vírus na população em geral.
- Se não soubermos este valor, podemos calcular as probabilidades com diversos cenários plausíveis para $\mathbb{P}(V+)$ e ver como as probabilidades se modificam (talvez elas não mudem muito).
- Pela regra de Bayes:

$$\mathbb{P}(V + | T+) = \frac{\mathbb{P}(T + | V+)\mathbb{P}(V+)}{\mathbb{P}(T+)} = \frac{0.99 * 0.001}{\mathbb{P}(T+)}$$

- Para obter $\mathbb{P}(T+)$ usamos um truque muito útil baseado em interseção de conjuntos.

O evento $T+$

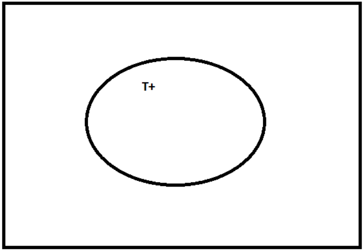


Figura: Evento $T+$ em Ω

O evento $T+$ e a decomposição de Ω

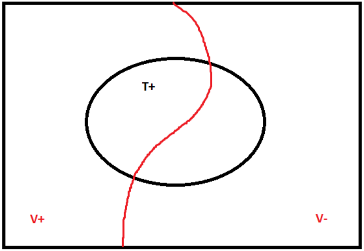


Figura: Evento $T+$ e Ω decomposto como $\Omega = V + \cup V-$.

O evento $T+$ decomposto

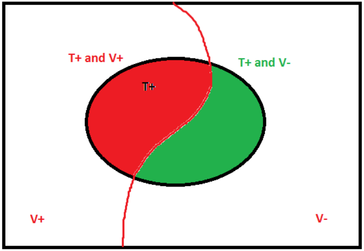


Figura: Decompondo o evento $T+ = (T + \cap V+) \bigcup (T + \cap V-)$.

$\mathbb{P}(T+)$

$$
\begin{aligned}
\mathbb{P}(T+) &= \mathbb{P}(T+ \;\cap\; (V+ \;\cup\; V-)) \\
&= \mathbb{P}((T+ \cap V+) \;\cup\; (T+ \cap V-)) \\
&= \mathbb{P}(T+ \cap V+) + \mathbb{P}(T+ \cap V-) \\
&= \mathbb{P}(T+|V+) * \mathbb{P}(V+) + \mathbb{P}(T+|V-) * \mathbb{P}(V-) \\
&= \quad 0.99 \;*\; 0.001 + \quad 0.05 \;*\; 0.999 \\
&= 0.05094
\end{aligned}
$$

Falso positivo e falso negativo

- Finalizando:

$$\mathbb{P}(V+|T+) = \frac{0.99 * 0.001}{0.05094} = 0.019$$

Falso positivo e falso negativo

- Finalizando:

$$\mathbb{P}(V+|T+) = \frac{0.99 * 0.001}{0.05094} = 0.019$$

- Assim, se tivermos um $T+$, será muito alta a chance do paciente ser $V-$ ou não ter o vírus.

Falso positivo e falso negativo

- Finalizando:

$$\mathbb{P}(V+|T+) = \frac{0.99 * 0.001}{0.05094} = 0.019$$

- Assim, se tivermos um $T+$, será muito alta a chance do paciente ser $V-$ ou não ter o vírus.
- Apenas 2% dos indivíduos com teste positivo ($T+$) possuem o vírus de fato.

Falso positivo e falso negativo

- Finalizando:

$$\mathbb{P}(V+|T+) = \frac{0.99 * 0.001}{0.05094} = 0.019$$

- Assim, se tivermos um $T+$, será muito alta a chance do paciente ser $V-$ ou não ter o vírus.
- Apenas 2% dos indivíduos com teste positivo ($T+$) possuem o vírus de fato.
- Idem, calculamos a outra probabilidade inversa:

Falso positivo e falso negativo

- Finalizando:

$$\mathbb{P}(V+|T+) = \frac{0.99 * 0.001}{0.05094} = 0.019$$

- Assim, se tivermos um $T+$, será muito alta a chance do paciente ser $V-$ ou não ter o vírus.
- Apenas 2% dos indivíduos com teste positivo ($T+$) possuem o vírus de fato.
- Idem, calculamos a outra probabilidade inversa:

$$\begin{aligned}\mathbb{P}(V-|T-) &= \frac{\mathbb{P}(T-|V-)\mathbb{P}(V-)}{\mathbb{P}(T-)} \\ &= \frac{0.95 * (1-0.001)}{1-0.05094} = 0.9999895\end{aligned}$$

Falso positivo e falso negativo

- Finalizando:

$$\mathbb{P}(V+|T+) = \frac{0.99 * 0.001}{0.05094} = 0.019$$

- Assim, se tivermos um $T+$, será muito alta a chance do paciente ser $V-$ ou não ter o vírus.
- Apenas 2% dos indivíduos com teste positivo ($T+$) possuem o vírus de fato.
- Idem, calculamos a outra probabilidade inversa:

$$\begin{aligned}\mathbb{P}(V-|T-) &= \frac{\mathbb{P}(T-|V-)\mathbb{P}(V-)}{\mathbb{P}(T-)} \\ &= \frac{0.95 * (1-0.001)}{1-0.05094} = 0.9999895\end{aligned}$$

- Se o teste for negativo, é praticamente certo que o indivíduo será $V-$.

Rastreamento em massa?

- Estes cálculos mostram por que não fazemos um rastreamento em massa na população brasileira.

Rastreamento em massa?

- Estes cálculos mostram por que não fazemos um rastreamento em massa na população brasileira.
- $\mathbb{P}(V+|T-) \approx 0 \Rightarrow$ se teste não detecta, a chance de estar infectado é baixa (Ok, ótimo!)

Rastreamento em massa?

- Estes cálculos mostram por que não fazemos um rastreamento em massa na população brasileira.
- $\mathbb{P}(V+\mid T-) \approx 0 \Rightarrow$ se teste não detecta, a chance de estar infectado é baixa (Ok, ótimo!)
- Mas $\mathbb{P}(V-\mid T+) \approx 1 \Rightarrow$ quase todos detectados pelo teste não estão infectados.

Rastreamento em massa?

- Estes cálculos mostram por que não fazemos um rastreamento em massa na população brasileira.
- $\mathbb{P}(V+|T-) \approx 0 \Rightarrow$ se teste não detecta, a chance de estar infectado é baixa (Ok, ótimo!)
- Mas $\mathbb{P}(V-|T+) \approx 1 \Rightarrow$ quase todos detectados pelo teste não estão infectados.
- Quantas pessoas dariam positivas (falsamente ou corretamente) pelo teste? Isto é, quantos teriam $T+$?

Rastreamento em massa?

- Estes cálculos mostram por que não fazemos um rastreamento em massa na população brasileira.
- $\mathbb{P}(V+ \mid T-) \approx 0 \Rightarrow$ se teste não detecta, a chance de estar infectado é baixa (Ok, ótimo!)
- Mas $\mathbb{P}(V- \mid T+) \approx 1 \Rightarrow$ quase todos detectados pelo teste não estão infectados.
- Quantas pessoas dariam positivas (falsamente ou corretamente) pelo teste? Isto é, quantos teriam $T+$?
- Aproximadamente 200 milhões $\times \mathbb{P}(T+) \approx 10$ milhões, um número enorme.

Rastreamento em massa?

- Estes cálculos mostram por que não fazemos um rastreamento em massa na população brasileira.
- $\mathbb{P}(V+ \mid T-) \approx 0 \Rightarrow$ se teste não detecta, a chance de estar infectado é baixa (Ok, ótimo!)
- Mas $\mathbb{P}(V- \mid T+) \approx 1 \Rightarrow$ quase todos detectados pelo teste não estão infectados.
- Quantas pessoas dariam positivas (falsamente ou corretamente) pelo teste? Isto é, quantos teriam $T+$?
- Aproximadamente 200 milhões $\times \mathbb{P}(T+) \approx 10$ milhões, um número enorme.
- Destes, 98% (ou 9.8 milhões) não têm HIV: a imensa maioria de um número enorme de pessoas.

Rastreamento em massa?

- Estes cálculos mostram por que não fazemos um rastreamento em massa na população brasileira.
- $\mathbb{P}(V+ \mid T-) \approx 0 \Rightarrow$ se teste não detecta, a chance de estar infectado é baixa (Ok, ótimo!)
- Mas $\mathbb{P}(V- \mid T+) \approx 1 \Rightarrow$ quase todos detectados pelo teste não estão infectados.
- Quantas pessoas dariam positivas (falsamente ou corretamente) pelo teste? Isto é, quantos teriam $T+$?
- Aproximadamente 200 milhões $\times \mathbb{P}(T+) \approx 10$ milhões, um número enorme.
- Destes, 98% (ou 9.8 milhões) não têm HIV: a imensa maioria de um número enorme de pessoas.
- As dificuldades de garantir um teste em todos e o custo envolvido leva a outra estratégia: fazer uma busca ativa entre pessoas de grupos de risco (que teriam $\mathbb{P}(V+)$ bem maior).

Regra da probabilidade total

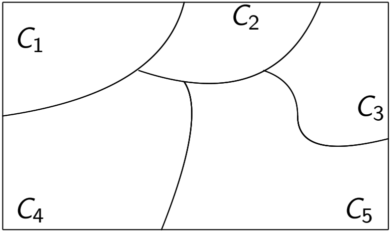
- Na regra de Bayes, derivamos uma fórmula muito útil, chamada fórmula da probabilidade total. Vamos ver o caso geral.

Regra da probabilidade total

- Na regra de Bayes, derivamos uma fórmula muito útil, chamada fórmula da probabilidade total. Vamos ver o caso geral.
- Espaço amostral Ω é particionado nos eventos $C_1, C_2, \ldots, C_k$.

Regra da probabilidade total

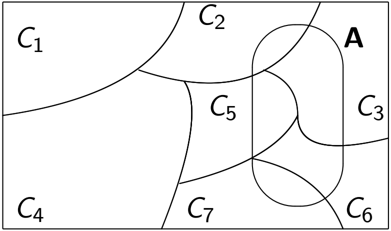
- Na regra de Bayes, derivamos uma fórmula muito útil, chamada fórmula da probabilidade total. Vamos ver o caso geral.
- Espaço amostral Ω é particionado nos eventos $C_1, C_2, \ldots, C_k$.



$$\Omega = C_1 \cup C_2 \cup \ldots \cup C_k$$
e
$$C_i \cap C_j = \emptyset \quad se\ i \neq j$$

Regra da probabilidade total

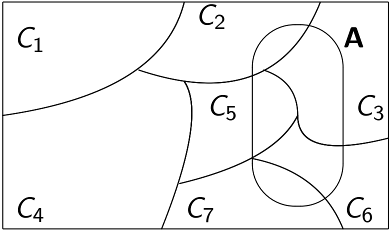
- Para qualquer evento A temos



$$\begin{aligned} A &= A \cap \Omega \\ &= A \cap (C_1 \cup C_2 \cup \ldots \cup C_k) \\ &= (A \cap C_1) \cup (A \cap C_2) \cup \ldots \cup (A \cap C_k) \end{aligned}$$

Regra da probabilidade total

- Para qualquer evento A temos



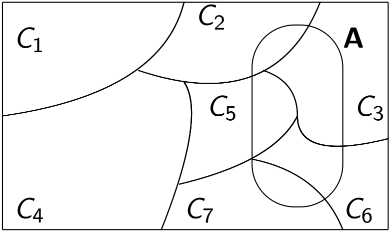
$$\begin{aligned} A &= A \cap \Omega \\ &= A \cap (C_1 \cup C_2 \cup \ldots \cup C_k) \\ &= (A \cap C_1) \cup (A \cap C_2) \cup \ldots \cup (A \cap C_k) \end{aligned}$$

- Temos então

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A \cap C_1) + \ldots + \mathbb{P}(A \cap C_k) \\ &= \mathbb{P}(A|C_1)\mathbb{P}(C_1) + \ldots + \mathbb{P}(A|C_k)\mathbb{P}(C_k) \end{aligned}$$

Extensão da Regra de Bayes

- Espaço amostral Ω é particionado nos eventos $C_1, \ldots, C_k$:



$$\Omega = C_1 \cup \ldots \cup C_k$$
e
$$C_i \cap C_j = \emptyset$$

Regra de Bayes

- Temos

$$
\begin{aligned}
\mathbb{P}(C_i|A) &= \frac{\mathbb{P}(A|C_i)\mathbb{P}(C_i)}{\mathbb{P}(A)} \\
&= \frac{\mathbb{P}(A|C_i)\mathbb{P}(C_i)}{\mathbb{P}(A \cap C_1) + \ldots + \mathbb{P}(A \cap C_k)} \\
&= \frac{\mathbb{P}(A|C_i)\mathbb{P}(C_i)}{\mathbb{P}(A|C_1)\mathbb{P}(C_1) + \ldots + \mathbb{P}(A|C_k)\mathbb{P}(C_k)}
\end{aligned}
$$

Regra de Bayes

- Temos

$$\begin{aligned}
\mathbb{P}(C_i|A) &= \frac{\mathbb{P}(A|C_i)\mathbb{P}(C_i)}{\mathbb{P}(A)} \\
&= \frac{\mathbb{P}(A|C_i)\mathbb{P}(C_i)}{\mathbb{P}(A \cap C_1) + \ldots + \mathbb{P}(A \cap C_k)} \\
&= \frac{\mathbb{P}(A|C_i)\mathbb{P}(C_i)}{\mathbb{P}(A|C_1)\mathbb{P}(C_1) + \ldots + \mathbb{P}(A|C_k)\mathbb{P}(C_k)}
\end{aligned}$$

- Isto é,

$$\mathbb{P}(C_i|A) = \frac{\mathbb{P}(A|C_i)\mathbb{P}(C_i)}{\sum_{j=1}^{k} \mathbb{P}(A|C_j)\mathbb{P}(C_j)}$$

Regra de Bayes

- Temos

$$\begin{aligned}\mathbb{P}(C_i|A) &= \frac{\mathbb{P}(A|C_i)\mathbb{P}(C_i)}{\mathbb{P}(A)} \\ &= \frac{\mathbb{P}(A|C_i)\mathbb{P}(C_i)}{\mathbb{P}(A \cap C_1) + \ldots + \mathbb{P}(A \cap C_k)} \\ &= \frac{\mathbb{P}(A|C_i)\mathbb{P}(C_i)}{\mathbb{P}(A|C_1)\mathbb{P}(C_1) + \ldots + \mathbb{P}(A|C_k)\mathbb{P}(C_k)}\end{aligned}$$

- Isto é,

$$\mathbb{P}(C_i|A) = \frac{\mathbb{P}(A|C_i)\mathbb{P}(C_i)}{\sum_{j=1}^{k} \mathbb{P}(A|C_j)\mathbb{P}(C_j)}$$

- Esta é a forma geral da regra de Bayes.

Exemplo: Regra de Bayes

- Website produz artigos: em três tópicos:
  - Política (P), Esportes (E), e Cultura (C)
- Porcentagens usuais de artigos:
  - $P : 50\%$ $\quad E : 40\%$ $\quad C : 10\%$
- Assim, $\mathbb{P}(P) = 0.50$, $\mathbb{P}(E) = 0.40$ e $\mathbb{P}(C) = 0.10$.
- Classificador automático de textos recebe como entrada um artigo e verifica as palavras presentes.
- Objetivo: classificar o texto em uma das 3 categorias.

Exemplo: Regra de Bayes

- A palavra *arcabouço* aparece em:
  - 50% dos textos de cultura,
  - em 30% dos textos de política
  - e em somente 5% dos textos de esportes.
- Veja que as probabilidades acima não somam 1.
- Elas representam as seguintes probabilidades condicionais:
  - $\mathbb{P}(\text{arcabouço}|C) = 0.50$,
  - $\mathbb{P}(\text{arcabouço}|P) = 0.30$,
  - $\mathbb{P}(\text{arcabouço}|E) = 0.05$,
- $A =$ o evento de que a palavra *arcabouço* está presente num dado artigo.
- Qual a probabilidade de que este texto com a palavra *arcabouço* seja do tópico Cultura?
- Isto é, quanto é $\mathbb{P}(C|A)$? Nós temos a probabilidade reversa: $\mathbb{P}(A|C) = 0.50$

Exemplo: Regra de Bayes

Queremos

$$\mathbb{P}(C|A) = \mathbb{P}(C|\text{arcabouço})$$

tendo

$$\mathbb{P}(A|C) = \mathbb{P}(\text{arcabouço}|C) = 0.50$$

Pela regra de Bayes,

$$\begin{aligned}
\mathbb{P}(C|A) &= \frac{\mathbb{P}(A|C)\mathbb{P}(C)}{\mathbb{P}(A|C)\mathbb{P}(C) + \mathbb{P}(A|E)\mathbb{P}(E) + \mathbb{P}(A|P)\mathbb{P}(P)} \\
&= \frac{0.50 \cdot 0.10}{0.50 \cdot 0.10 + 0.30 \cdot 0.50 + 0.05 \cdot 0.40} \\
&= 0.23
\end{aligned}$$

Parece que a classe C (cultura) não é a mais provável (apenas 0.23 de probabilidade).

Exemplo: Regra de Bayes

Vamos calcular a probabilidade do texto pertencer ás outra possíveis classes com a regra de Bayes:

$$\mathbb{P}(P|A), \mathbb{P}(E|A), \text{ sabendo que } \mathbb{P}(C|A) = 0.23$$

$$\mathbb{P}(P|A) = \frac{\mathbb{P}(A|P)\mathbb{P}(P)}{\mathbb{P}(A|C)\mathbb{P}(C) + \mathbb{P}(A|E)\mathbb{P}(E) + \mathbb{P}(A|P)\mathbb{P}(P)} == \frac{0.30}{0.50 \cdot 0.10 + 0.30}$$

$$\mathbb{P}(E|A) = \frac{\mathbb{P}(A|P)\mathbb{P}(P)}{\mathbb{P}(A|C)\mathbb{P}(C) + \mathbb{P}(A|E)\mathbb{P}(E) + \mathbb{P}(A|P)\mathbb{P}(P)} == \frac{0.05}{0.50 \cdot 0.10 + 0.30}$$

Exemplo: Regra de Bayes

- A classe mais provável desse texto é *Política*.
- Veja que a palavra *arcabouço* é mais comum nos textos de *cultura*.
- A maior presença de textos de política na coleção fez a balança pender para a classe *política*.
- Métodos de classificação de textos todas as palavras do dicionário, não apenas uma delas.
- Um desses métodos é o Naive Bayes, que veremos mais à frente.

Exemplo: Regra de Bayes

- Em uma urna, existem 6 bolas de cores desconhecidas.

Exemplo: Regra de Bayes

- Em uma urna, existem 6 bolas de cores desconhecidas.
- Três bolas são retiradas sem reposição e são pretas.

Exemplo: Regra de Bayes

- Em uma urna, existem 6 bolas de cores desconhecidas.
- Três bolas são retiradas sem reposição e são pretas.
- Ache a probabilidade de que não restam bolas pretas na urna.

Exemplo: Regra de Bayes

- Em uma urna, existem 6 bolas de cores desconhecidas.
- Três bolas são retiradas sem reposição e são pretas.
- Ache a probabilidade de que não restam bolas pretas na urna.
- $A = 3$ bolas pretas são retiradas

Exemplo: Regra de Bayes

- Em uma urna, existem 6 bolas de cores desconhecidas.
- Três bolas são retiradas sem reposição e são pretas.
- Ache a probabilidade de que não restam bolas pretas na urna.
- $A =$ 3 bolas pretas são retiradas
- $C_i =$ existem i bolas pretas na urna

Exemplo: Regra de Bayes

- Em uma urna, existem 6 bolas de cores desconhecidas.
- Três bolas são retiradas sem reposição e são pretas.
- Ache a probabilidade de que não restam bolas pretas na urna.
- $A =$ 3 bolas pretas são retiradas
- $C_i =$ existem i bolas pretas na urna
- Urna com 6 bolas $\rightarrow$ retira-se 3 bolas

Exemplo: Regra de Bayes

- Em uma urna, existem 6 bolas de cores desconhecidas.
- Três bolas são retiradas sem reposição e são pretas.
- Ache a probabilidade de que não restam bolas pretas na urna.
- $A =$ 3 bolas pretas são retiradas
- $C_i =$ existem i bolas pretas na urna
- Urna com 6 bolas $\rightarrow$ retira-se 3 bolas
- $C_i =$ existem i bolas pretas na urna $(i = 0, 1, \ldots, 6)$

Exemplo: Regra de Bayes

- Em uma urna, existem 6 bolas de cores desconhecidas.
- Três bolas são retiradas sem reposição e são pretas.
- Ache a probabilidade de que não restam bolas pretas na urna.
- $A =$ 3 bolas pretas são retiradas
- $C_i =$ existem i bolas pretas na urna
- Urna com 6 bolas $\rightarrow$ retira-se 3 bolas
- $C_i =$ existem i bolas pretas na urna $(i = 0, 1, \ldots, 6)$
- $A =$ 3 bolas pretas são retiradas

Exemplo: Regra de Bayes

- Em uma urna, existem 6 bolas de cores desconhecidas.
- Três bolas são retiradas sem reposição e são pretas.
- Ache a probabilidade de que não restam bolas pretas na urna.
- $A =$ 3 bolas pretas são retiradas
- $C_i =$ existem i bolas pretas na urna
- Urna com 6 bolas $\rightarrow$ retira-se 3 bolas
- $C_i =$ existem i bolas pretas na urna $(i = 0, 1, \ldots, 6)$
- $A =$ 3 bolas pretas são retiradas
- $P(A|C_i) =?$

Regra de Bayes

- $P(A|C_i) = ?$

$$\begin{cases} \mathbb{P}(A|C_0) = \frac{\mathbb{P}(A \cap C_0)}{\mathbb{P}(C_0)} = \frac{0}{\mathbb{P}(C_0)} = 0 \\ \\ \\ \\ \end{cases}$$

Regra de Bayes

- $P(A|C_i) = ?$

$$\begin{cases} \mathbb{P}(A|C_0) = \frac{\mathbb{P}(A \cap C_0)}{\mathbb{P}(C_0)} = \frac{0}{\mathbb{P}(C_0)} = 0 \\ \\ \mathbb{P}(A|C_1) = 0 \\ \\ \\ \\ \\ \end{cases}$$

Regra de Bayes

- $P(A|C_i) = ?$

$$\begin{cases} \mathbb{P}(A|C_0) = \frac{\mathbb{P}(A \cap C_0)}{\mathbb{P}(C_0)} = \frac{0}{\mathbb{P}(C_0)} = 0 \\ \\ \mathbb{P}(A|C_1) = 0 \\ \mathbb{P}(A|C_2) = 0 \\ \\ \\ \\ \\ \end{cases}$$

Regra de Bayes

- $P(A|C_i) =?$

$$\begin{cases} \mathbb{P}(A|C_0) = \frac{\mathbb{P}(A \cap C_0)}{\mathbb{P}(C_0)} = \frac{0}{\mathbb{P}(C_0)} = 0 \\ \\ \mathbb{P}(A|C_1) = 0 \\ \mathbb{P}(A|C_2) = 0 \\ \mathbb{P}(A|C_3) = 1/20 \quad \text{(Hipergeométrica)} \\ \\ \\ \\ \end{cases}$$

Regra de Bayes

- $P(A|C_i) = ?$

$$
\begin{cases}
\mathbb{P}(A|C_0) = \frac{\mathbb{P}(A \cap C_0)}{\mathbb{P}(C_0)} = \frac{0}{\mathbb{P}(C_0)} = 0 \\
\\
\mathbb{P}(A|C_1) = 0 \\
\mathbb{P}(A|C_2) = 0 \\
\mathbb{P}(A|C_3) = 1/20 \quad \text{(Hipergeométrica)} \\
\\
\mathbb{P}(A|C_4) = 1/5 \\
\\
\\
\end{cases}
$$

Regra de Bayes

- $P(A|C_i) =?$

$$\begin{cases} \mathbb{P}(A|C_0) = \frac{\mathbb{P}(A \cap C_0)}{\mathbb{P}(C_0)} = \frac{0}{\mathbb{P}(C_0)} = 0 \\ \\ \mathbb{P}(A|C_1) = 0 \\ \mathbb{P}(A|C_2) = 0 \\ \mathbb{P}(A|C_3) = 1/20 \quad \text{(Hipergeométrica)} \\ \\ \mathbb{P}(A|C_4) = 1/5 \\ \mathbb{P}(A|C_5) = 1/2 \\ \end{cases}$$

Regra de Bayes

- $P(A|C_i) = ?$

$$
\begin{cases}
\mathbb{P}(A|C_0) = \frac{\mathbb{P}(A \cap C_0)}{\mathbb{P}(C_0)} = \frac{0}{\mathbb{P}(C_0)} = 0 \\
\\
\mathbb{P}(A|C_1) = 0 \\
\mathbb{P}(A|C_2) = 0 \\
\mathbb{P}(A|C_3) = 1/20 \quad \text{(Hipergeométrica)} \\
\\
\mathbb{P}(A|C_4) = 1/5 \\
\mathbb{P}(A|C_5) = 1/2 \\
\mathbb{P}(A|C_6) = 1
\end{cases}
$$

- Queremos calcular $\mathbb{P}(C_3|A)$

Regra de Bayes

- Temos

$$\mathbb{P}(C_3|A) = \frac{\mathbb{P}(A|C_3) * \mathbb{P}(C_3)}{\sum_{j=0}^{6} \mathbb{P}(A|C_j) * \mathbb{P}(C_j)}$$

$$= \frac{\frac{1}{20} * \mathbb{P}(C_3)}{0 + 0 + 0 + \frac{1}{20}\mathbb{P}(C_3) + \frac{1}{5}\mathbb{P}(C_4) + \frac{1}{2}\mathbb{P}(C_5) + 1\mathbb{P}(C_6)}$$

$$= ??$$

Regra de Bayes

- Temos

$$\mathbb{P}(C_3|A) = \frac{\mathbb{P}(A|C_3) * \mathbb{P}(C_3)}{\sum_{j=0}^{6} \mathbb{P}(A|C_j) * \mathbb{P}(C_j)}$$

$$= \frac{\frac{1}{20} * \mathbb{P}(C_3)}{0 + 0 + 0 + \frac{1}{20}\mathbb{P}(C_3) + \frac{1}{5}\mathbb{P}(C_4) + \frac{1}{2}\mathbb{P}(C_5) + 1\mathbb{P}(C_6)}$$

$$= ??$$

- Precisamos estabelecer o valor de $\mathbb{P}(C_j)$, a probabilidade de que existam j bolas pretas na urna.

Regra de Bayes

- Temos

$$\mathbb{P}(C_3|A) = \frac{\mathbb{P}(A|C_3) * \mathbb{P}(C_3)}{\sum_{j=0}^{6} \mathbb{P}(A|C_j) * \mathbb{P}(C_j)}$$

$$= \frac{\frac{1}{20} * \mathbb{P}(C_3)}{0 + 0 + 0 + \frac{1}{20}\mathbb{P}(C_3) + \frac{1}{5}\mathbb{P}(C_4) + \frac{1}{2}\mathbb{P}(C_5) + 1\mathbb{P}(C_6)}$$

$$= ??$$

- Precisamos estabelecer o valor de $\mathbb{P}(C_j)$, a probabilidade de que existam j bolas pretas na urna.
- Depende do mecanismo que colocou bolas na urna e isto não foi explicado no problema.

Regra de Bayes

- Vamos mostrar algumas possibilidades para $\mathbb{P}(C_j)$.

Regra de Bayes

- Vamos mostrar algumas possibilidades para $\mathbb{P}(C_j)$.
- Qualquer número de bolas pretas entre 0 e 6 tem a mesma probabilidade. Então $\mathbb{P}(C_j) = \frac{1}{7}$ para todo j?

Regra de Bayes

- Vamos mostrar algumas possibilidades para $\mathbb{P}(C_j)$.
- Qualquer número de bolas pretas entre 0 e 6 tem a mesma probabilidade. Então $\mathbb{P}(C_j) = \frac{1}{7}$ para todo j?
- Bolas são escolhidas preferencialmente de uma única cor. Então os valores de $\mathbb{P}(C_j)$ para $j = 0$ e $j = 6$ seriam os maiores, com o valor mínimo com $j = 3$.

Regra de Bayes

- Vamos mostrar algumas possibilidades para $\mathbb{P}(C_j)$.
- Qualquer número de bolas pretas entre 0 e 6 tem a mesma probabilidade. Então $\mathbb{P}(C_j) = \frac{1}{7}$ para todo j?
- Bolas são escolhidas preferencialmente de uma única cor. Então os valores de $\mathbb{P}(C_j)$ para $j = 0$ e $j = 6$ seriam os maiores, com o valor mínimo com $j = 3$.
- Por exemplo, $\mathbb{P}(C_j) = \frac{1}{28}(j-3)^2$

Regra de Bayes

- Vamos mostrar algumas possibilidades para $\mathbb{P}(C_j)$.
- Qualquer número de bolas pretas entre 0 e 6 tem a mesma probabilidade. Então $\mathbb{P}(C_j) = \frac{1}{7}$ para todo j?
- Bolas são escolhidas preferencialmente de uma única cor. Então os valores de $\mathbb{P}(C_j)$ para $j = 0$ e $j = 6$ seriam os maiores, com o valor mínimo com $j = 3$.
- Por exemplo, $\mathbb{P}(C_j) = \frac{1}{28}(j-3)^2$
- Outra opção: Existem 10 cores distintas e a cor de cada bola é escolhida ao acaso.

Regra de Bayes

- Vamos mostrar algumas possibilidades para $\mathbb{P}(C_j)$.
- Qualquer número de bolas pretas entre 0 e 6 tem a mesma probabilidade. Então $\mathbb{P}(C_j) = \frac{1}{7}$ para todo j?
- Bolas são escolhidas preferencialmente de uma única cor. Então os valores de $\mathbb{P}(C_j)$ para $j = 0$ e $j = 6$ seriam os maiores, com o valor mínimo com $j = 3$.
- Por exemplo, $\mathbb{P}(C_j) = \frac{1}{28}(j-3)^2$
- Outra opção: Existem 10 cores distintas e a cor de cada bola é escolhida ao acaso.
- A chance de colocar uma bola preta na urna é $1/10$.

Regra de Bayes

- Vamos mostrar algumas possibilidades para $\mathbb{P}(C_j)$.
- Qualquer número de bolas pretas entre 0 e 6 tem a mesma probabilidade. Então $\mathbb{P}(C_j) = \frac{1}{7}$ para todo j?
- Bolas são escolhidas preferencialmente de uma única cor. Então os valores de $\mathbb{P}(C_j)$ para $j = 0$ e $j = 6$ seriam os maiores, com o valor mínimo com $j = 3$.
- Por exemplo, $\mathbb{P}(C_j) = \frac{1}{28}(j-3)^2$
- Outra opção: Existem 10 cores distintas e a cor de cada bola é escolhida ao acaso.
- A chance de colocar uma bola preta na urna é $1/10$.
- A chance de colocar j bolsas pretas na urna de 6 bolas é

$$\mathbb{P}(C_j) = \binom{6}{j}(0.1)^j(0.9)^{6-j}$$

Independência de eventos em geral

- Falamos da independência de dois eventos A e B.

Independência de eventos em geral

- Falamos da independência de dois eventos A e B.
- Eles são eventos independentes se

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)\,,$$

Independência de eventos em geral

- Falamos da independência de dois eventos A e B.
- Eles são eventos independentes se

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\ \mathbb{P}(B)\,,$$

ou, equivalentemente,

$$\mathbb{P}(A|B) = \mathbb{P}(A)$$

Independência de eventos em geral

- Falamos da independência de dois eventos A e B.
- Eles são eventos independentes se

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\ \mathbb{P}(B)\,,$$

ou, equivalentemente,

$$\mathbb{P}(A|B) = \mathbb{P}(A)$$

ou ainda

$$\mathbb{P}(B|A) = \mathbb{P}(B)$$

Independência de eventos em geral

- Falamos da independência de dois eventos A e B.
- Eles são eventos independentes se

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\, \mathbb{P}(B)\,,$$

ou, equivalentemente,

$$\mathbb{P}(A|B) = \mathbb{P}(A)$$

ou ainda

$$\mathbb{P}(B|A) = \mathbb{P}(B)$$

- E quando tivermos vários eventos $E_1, E_2, \ldots, E_n$?

Independência de eventos em geral

- Falamos da independência de dois eventos A e B.
- Eles são eventos independentes se

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\ \mathbb{P}(B)\ ,$$

ou, equivalentemente,

$$\mathbb{P}(A|B) = \mathbb{P}(A)$$

ou ainda

$$\mathbb{P}(B|A) = \mathbb{P}(B)$$

- E quando tivermos vários eventos $E_1, E_2, \ldots, E_n$?
- Infelizmente, não basta olhar os pares de eventos e verificar a definição acima.

Independência de eventos em geral

- Os eventos $E_1, E_2, \ldots, E_n$ são eventos independentes se toda combinação de eventos satisfazer a regra do produto:

$$\mathbb{P}(E_{i_1} \ \cap \ E_{i_2} \ \cap \ \ldots \ \cap \ E_{i_m}) = \mathbb{P}(E_{i_1}) \ldots \mathbb{P}(E_{i_m})$$

Independência de eventos em geral

- Os eventos $E_1, E_2, \ldots, E_n$ são eventos independentes se toda combinação de eventos satisfazer a regra do produto:

$$\mathbb{P}(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_m}) = \mathbb{P}(E_{i_1}) \ldots \mathbb{P}(E_{i_m})$$

para toda seleção de índices $i_1, i_2, \ldots, i_m$ e para todo m entre 2 e n.

Independência de eventos em geral

- Os eventos $E_1, E_2, \ldots, E_n$ são eventos independentes se toda combinação de eventos satisfazer a regra do produto:

$$\mathbb{P}(E_{i_1} \ \cap \ E_{i_2} \ \cap \ \ldots \ \cap \ E_{i_m}) = \mathbb{P}(E_{i_1}) \ldots \mathbb{P}(E_{i_m})$$

  para toda seleção de índices $i_1, i_2, \ldots, i_m$ e para todo m entre 2 e n.
- Estes eventos são chamados *mutuamente independentes*.

Independência de eventos em geral

- Os eventos $E_1, E_2, \ldots, E_n$ são eventos independentes se toda combinação de eventos satisfazer a regra do produto:

$$\mathbb{P}(E_{i_1} \ \cap \ E_{i_2} \ \cap \ \ldots \ \cap \ E_{i_m}) = \mathbb{P}(E_{i_1}) \ldots \mathbb{P}(E_{i_m})$$

  para toda seleção de índices $i_1, i_2, \ldots, i_m$ e para todo m entre 2 e n.
- Estes eventos são chamados *mutuamente independentes*.
- Podemos deduzir que se A, B, e C são independentes então C é também independente de $A \cap B$, de $A \cap B^c$, de $A \cup B$, de B^c etc. (ver lista de exercícios).

Independência mútua e independência par a par

- Se os eventos são mutuamente independentes então qualquer par de eventos é independente.

Independência mútua e independência par a par

- Se os eventos são mutuamente independentes então qualquer par de eventos é independente.
- Um resultado curioso é que a conversa não é verdade.

Independência mútua e independência par a par

- Se os eventos são mutuamente independentes então qualquer par de eventos é independente.
- Um resultado curioso é que a conversa não é verdade.
- Podemos ter eventos independentes par a par mas que não são mutuamente independentes.

Independência mútua e independência par a par

- Se os eventos são mutuamente independentes então qualquer par de eventos é independente.
- Um resultado curioso é que a conversa não é verdade.
- Podemos ter eventos independentes par a par mas que não são mutuamente independentes.
- Por exemplo, podemos ter A e B indep, A e C indep, e B e C indep mas A, B, C dependentes.

Independência mútua e independência par a par

- Se os eventos são mutuamente independentes então qualquer par de eventos é independente.
- Um resultado curioso é que a conversa não é verdade.
- Podemos ter eventos independentes par a par mas que não são mutuamente independentes.
- Por exemplo, podemos ter A e B indep, A e C indep, e B e C indep mas A, B, C dependentes.
- Um uso prático desta distinção aparece numa técnica para compartilhar senhas em criptografia (ver Davis, 2012, seção 8.9.2: leitura opcional na página da disciplina).