

Aula 2

Roteiro

- Espaço amostral
- Probabilidade
- Eventos
- Independência
- Exclusão mútua
- Probabilidade total
- Regra de Bayes
- Variável aleatória
- Função de distribuição
- Distribuição de Bernoulli

Aula passada

- Três (tipos de) problemas
- *Monte Carlo to the Rescue*
- História da origem do nome
- Monte Carlo na Computação

Espaço Amostral

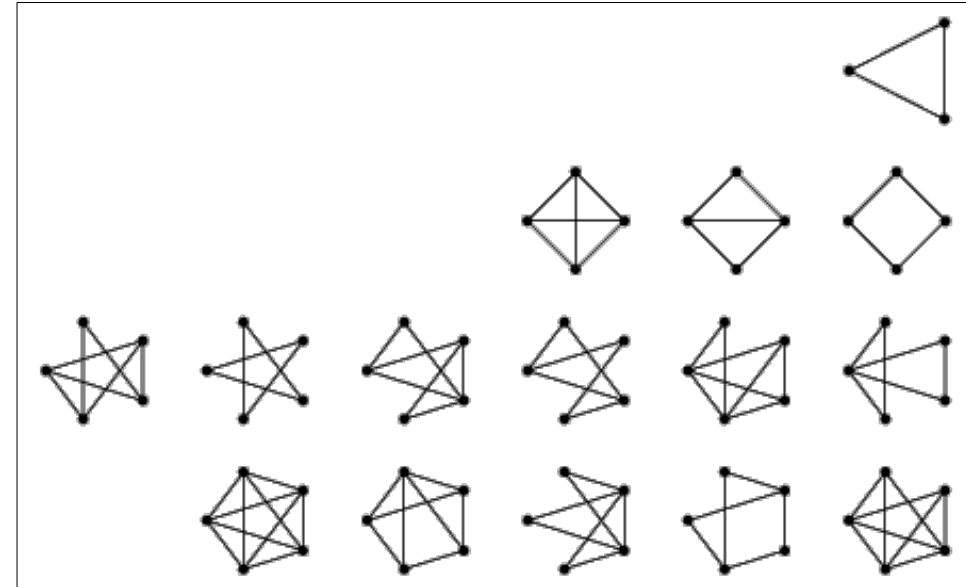
- Conjunto de objetos, enumerável (pode contar)



Frutas



Letras



Grafos

- S é o conjunto
 - ex. $S = \{a, b, c, \dots, z\}$
- $|S|$ é sua cardinalidade
 - ex. $|S| = 26$

Probabilidade

- Função que associa a cada elemento de S um valor entre 0 e 1
 - $p : S \rightarrow [0, 1]$
- Restrição: soma sobre todos elementos vale 1



Letras

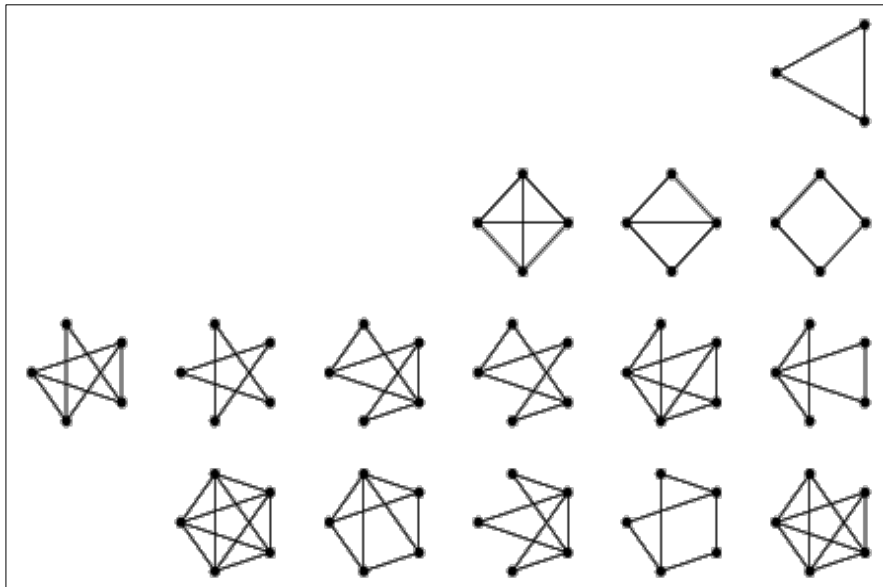
- $S = \text{alfabeto}, |S| = 26$
- $p_x = 1/26$ para qualquer letra x
- $p_x = 1/10$ para qualquer x vogal, e $1/42$ para qualquer x consoante

Eventos

- Subconjunto do espaço amostral



- $S = \text{alfabeto}, |S| = 26$
- $A = \{a, b, c, d\}$
- $B = \text{todas as consoantes}$
- $C = \text{todas as letras depois de Q}$



- $S = \text{grafos ao lado}, |S| = 15$
- $A = \text{todas as cliques}$
- $B = \text{todos os grafos com menos de quatro vértices}$
- $C = \text{todas as árvores}$

Probabilidade de Eventos

- Soma das probabilidades dos elementos que definem o evento

$$P[A] = \sum_{e \in A} p_e$$



- $S = \text{alfabeto}, |S| = 26, p_x = 1/26$
- $A = \{a, b, c, d\}$
- $B = \text{todas as consoantes}$
- $C = \text{todas as letras depois de Q}$
- $P[A] = 2/13$
- $P[B] = 21/26$
- $P[C] = 9/26$

Manipulando Eventos

- Eventos são conjuntos, podem ser manipulados usando lógica de conjuntos
- Operações básicas
 - união, interseção, complemento



- $A = \{a, b, c, d\}$

- $B =$ todas as consoantes

- $C =$ todas as letras depois de Q

$$A \cup B = ?$$

$$A \cap B = ?$$

$$B^c = ?$$

$$A \cap B \cap C = ?$$

Manipulando Eventos

- Probabilidade de eventos resultantes seguem a mesma definição

- ex. $Y = A \cap B \longrightarrow P[Y] = \sum_{e \in Y} p_e$



- $S =$ alfabeto, $|S| = 26$, $p_x = 1/26$
- $A = \{a, b, c, d\}$
- $B =$ todas as consoantes
- $C =$ todas as letras depois de Q

$$P[A \cup B] = ?$$

$$P[A \cap B \cap C] = ? \quad P[B^c] = ?$$

Ponto de Confusão

- Operadores lógicos são usados no lugar de operadores de conjunto
 - “ou”, “+” ao invés de “união”
 - “e”, “.” ao invés de “interseção”
 - “negação”, “not” ao invés de “complemento”
- Exemplos

$$A \cup B = A + B = A \vee B$$

$$A \cap B = A . B = A \wedge B$$

$$B^c = \overline{B} = \neg B$$

**Iremos usar esta notação,
que é a mais usual**

Independência

- Dois eventos A e B são independentes sse

$$P[A \cap B] = P[A \wedge B] = P[A]P[B]$$

- Ou seja, a probabilidade da conjunção (interseção) é igual ao produto das probabilidades



- $S =$ alfabeto, $|S| = 26$, $p_x = 1/26$
- $A = \{a, b, c, d\}$
- $B =$ todas as consoantes
- A e B são independentes?
- $A =$ todas as letras antes de N
- $B = \{a, z\}$
- A e B são independentes?

Exclusão Mútua

- Dois eventos A e B são mutuamente exclusivos sse

$$P[A \cup B] = P[A \vee B] = P[A] + P[B]$$

- Ou seja, a probabilidade da disjunção (união) é igual a soma das probabilidades



- $S =$ alfabeto, $|S| = 26$, $p_x = 1/26$
- $A = \{a, b, c, d\}$
- $B =$ todas as consoantes
- A e B são mutuamente exclusivos?
- $A =$ todas as letras antes de N
- $B = \{x, y, z\}$
- A e B são mutuamente exclusivos?

Probabilidade Condicional

- Probabilidade do evento A dado a “ocorrência” do evento B
 - novo espaço amostral para evento A passa a ser o conjunto B

$$P[A|B] = \frac{P[A \wedge B]}{P[B]}$$



- $S =$ alfabeto, $|S| = 26$, $p_x = 1/26$

- $A = \{a, b, c, d\}$

- $B =$ todas as consoantes

- $C = \{x, y, z\}$

$P[A|B] = ?$ $P[B|A] = ?$ $P[C|B] = ?$

Lei da Probabilidade Total

- Partição de um conjunto S
 - conjunto de subconjuntos de S tal que todo elemento de S aparece em exatamente um subconjunto
 - ex. $S =$ alfabeto: $O_1 =$ todas vogais, $O_2 =$ todas consoantes
- Relação entre probabilidade de um evento e probabilidade condicional com eventos de uma partição

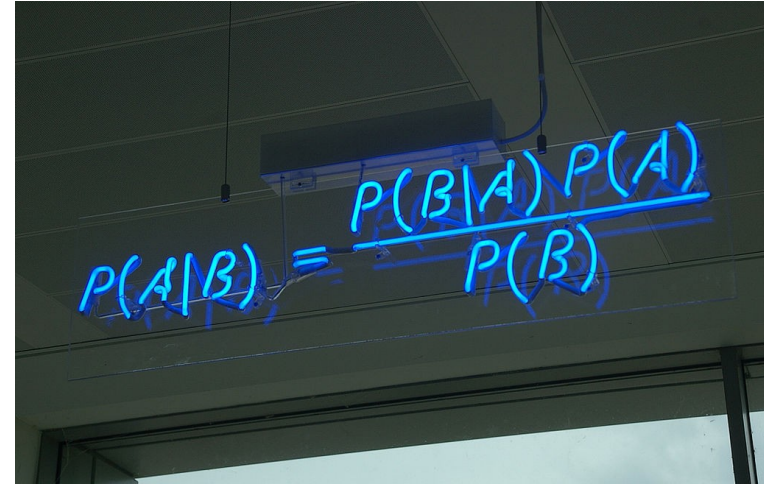
$$P[A] = \sum_i P[A \wedge B_i] = \sum_i P[A|B_i] P[B_i]$$

- onde B_i é qualquer partição do espaço amostral
- Regra muito usada para calcular probabilidade de um evento

Regra de Bayes

- Relação fundamental entre probabilidades condicionais

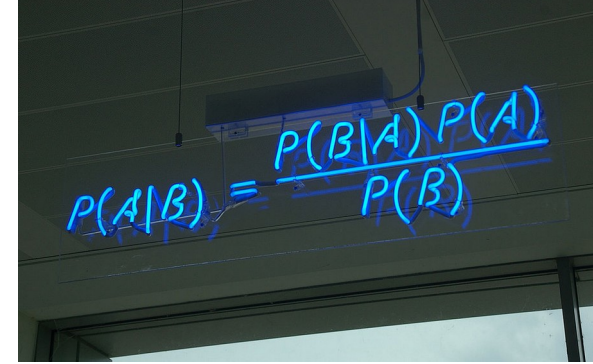
$$P[A|B] = \frac{P[B|A] P[A]}{P[B]}$$



A photograph of a whiteboard with the Bayes' Rule formula written in blue marker. The formula is $P(A|B) = \frac{P(B|A) P(A)}{P(B)}$. The whiteboard is mounted on a wall, and the lighting is somewhat dim, with the whiteboard being the primary light source.

- Talvez a mais importante relação em probabilidade
 - usada em muitos problemas, inclusive aprendizado
- Razão
 - queremos $P[A|B]$, mas calcular isto pode ser difícil
 - calcular $P[B|A]$ pode ser mais fácil, assim como $P[A]$ e $P[B]$
 - e muitas vezes calcular $P[B]$ não é necessário

Exemplo



A photograph of a chalkboard with the formula for Bayes' theorem written in blue chalk: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$.

- Filtro de spam usando regra de Bayes
- S = todos os e-mails de um grande universo
- M = subconjunto de e-mails que são spam
- V = subconjunto de e-mails que possuem a palavra “viagra”

$P[M|V]$ ← Probabilidade de um e-mail ser spam
dado que possui a palavra “viagra”

- Como calcular esta probabilidade? Regra de Bayes!

$$P[M|V] = \frac{P[V|M]P[M]}{P[V]}$$

- $P[V|M]$ é mais fácil de calcular empiricamente,
assim como $P[V]$ e $P[M]$

Variável Aleatória

- Variável aleatória X é uma função que mapeia o espaço amostral nos inteiros (v.a. discreta)
 - permite trabalhar com números, independente dos elementos do espaço amostral

$$X : S \rightarrow \mathbb{Z}$$

- Cuidado: X não é nem variável e nem é aleatória!
- Usaremos X para definir eventos em função de seus valores e elementos do espaço amostral

$$A = \{ X > 5 \} = \{ e \in S : X(e) > 5 \}$$

↖
Evento A: subconjunto do espaço amostral que X mapeia para valores maiores que 5

Exemplo de Variável Aleatória



- $S = \text{alfabeto}$, $|S| = 26$, $p_x = 1/26$
- X é uma v.a. tal que $X(a)=1$, $X(b)=2$, $X(c)=3$, ..., $X(z)=26$
- Y é uma v.a. tal que $Y(\text{vogal})=1$, $Y(\text{consoante})=2$
- Z é uma v.a. tal que $Z(\text{primeiras 10 letras}) = 1$, $Z(\text{últimas 10 letras}) = 2$, $Z(\text{outras letras}) = 3$
- Usamos v.a. para definir eventos
- $\{X > 13\} = \{n, o, \dots, z\}$, $\{X \text{ ímpar}\} = \{a, c, e, g, \dots, y\}$
- $\{Y = 1\} = \{a, e, i, o, u\}$
- $\{Z = 3\} = \{k, l, m, n, o, p\}$

Probabilidade de V.A.

- Probabilidade associada é dada pela probabilidade do evento definido pela v.a.
- $\{X > 13\} = \{n, o, \dots, z\}$, $\{X \text{ ímpar}\} = \{a, c, e, g, \dots, y\}$
- $\{Y = 1\} = \{a, e, i, o, u\}$
- $\{Z = 3\} = \{k, l, m, n, o, p\}$

- $P[X > 13] = ?$
- $P[Y = 1] = ?$
- $P[Y = 1 \text{ e } Z = 3] = ?$
são independentes

Cuidado! Y e Z são v.a. diferentes mas eventos não necessariamente são independentes

Manipulando V.A.

- Variáveis aleatórias podem ser manipuladas algebricamente
- Multiplicação por um escalar
 - ex. se X tem valores 1 e 2, então $2X$ tem valores 2 e 4
- Atribuição para definir uma nova v.a.
 - novo mapeamento do espaço amostral
 - ex. $Y = 2X + 1$
 - ex. $Z = X + Y$ (X e Y definidas no mesmo espaço amostral)
- Simplificação de eventos
 - ex. $\{2X > 4\} = \{X > 2\}$
 - ex. $\{2XY = 0\} = \{\{X = 0\} + \{Y = 0\}\}$

V.A. Indicadora

- A mais comum e importante variável aleatória
 - assume apenas dois valores (1 ou 0), indica elementos de interesse
 - $X = 1$, subconjunto do espaço amostral com os elementos de interesse
 - $X = 0$, caso contrário
- Exemplo: S = todos os grafos com n vértices
 - $X = 1$: todos os grafos conexos, $X = 0$ c.c.
 - $Y_k = 1$: todos os grafos com diâmetro k , $Y_k = 0$ c.c.
 - $Z_k = 1$: todos os grafos com k componentes conexas
 - $P[X = 1] = P[X]$ = probabilidade associada ao evento “ser conexo”

Novas Regras!

- Até agora: probabilidade associada ao valor de uma v.a. é definida pela probabilidade do respectivo evento
 - $P[X = 1] = P [\{e \mid X(e) = 1\}]$
- **Novas regras!** Iremos definir a probabilidade associada ao valor de uma v.a. diretamente
 - sem considerar o respectivo evento
 - $P[X = 1] = p_1$
- Iremos determinar o valor p_1 a partir de uma função
 - função de probabilidade (ou função de massa de probabilidade ou função de distribuição)

Função de Distribuição

- Seja X uma v.a. e x um de seus possíveis valores

$$f_X(x) = P[X = x] \quad \longleftarrow \text{Função de probabilidade}$$

$$F_X(x) = P[X \leq x] \quad \longleftarrow \text{Função cumulativa}$$

- Uma pode ser definida usando a outra

$$F_X(x) = \sum_{y \leq x} f_X(y) \quad \longleftarrow \text{Soma por exclusão mútua dos eventos } X = x$$

- Restrição

$$0 \leq f_X(x) \leq 1, \forall x \in O_X \quad \sum_{x \in O_X} f_X(x) = 1$$

- O_X é a imagem da v.a. X (valores usados no mapeamento)
- O_X é o domínio da função f_X (valores que ela pode assumir)

Distribuição de Bernoulli

- Uma v.a. X que possui distribuição de Bernoulli assume apenas dois valores (v.a. indicadora)

$$f_X(1) = P[X=1] = p$$

$$f_X(0) = P[X=0] = 1 - p$$

- Possui um único parâmetro $0 \leq p \leq 1$
- Usada como distribuição de qualquer v.a. indicadora
 - ex. cara ou coroa, sim ou não, verdade ou falso
- Notação: $X \sim \text{Bernoulli}(p)$
 - X é uma v.a. que possui distribuição de Bernoulli com parâmetro p