Noções de probabilidade
Parte 1

Prof. Eduardo Vargas Ferreira

Curso de Especialização em
Data Science & Big Data
Universidade Federal do Paraná

16 de março de 2018

Aplicação e *Toy Models*

- Probabilidade e Estatística são amplamente utilizadas nas engenharias, medicina, ciências sociais, economia, ciência da computação etc.

- A lista de aplicações é essencialmente infinita como: probabilidades e estratégias de jogo, previsão econômica, epidemiologia, dentre outros;

- Dadas tantas aplicações interessantes, você pode se perguntar "por que vamos passar tanto tempo pensando em *toy models* como moedas e dados?"

- **Resposta:** a fim de desenvolveremos boas percepções para a essência simples dentro de muitos problemas complexos do mundo real.

Teoria de conjuntos

- **Elemento:** escrevemos $x \in \Omega$ para significar que o elemento x está no conjunto Ω;
- **Subconjunto:** Dizemos que o conjunto A é um subconjunto de Ω se todos os seus elementos estiverem em Ω. Nós escrevemos isso como $A \subset \Omega$;
- **Complemento:** O complemento de A em Ω é o conjunto de elementos de Ω que não estão em A. Nós escrevemos isso como A^c ou $\Omega – A$;
- **União:** A união de A e B é o conjunto de todos os elementos em A ou B (ou ambos). Nós escrevemos isso como $A \cup B$;
- **Intersecção:** A interseção de A e B é o conjunto de todos os elementos em A e B. Escrevemos isto como $A \cap B$;

Teoria de conjuntos

- **Disjunto:** A e B são disjuntos se não tiverem elementos comuns. Ou seja, se $A \cap B = \emptyset$;
- **Diferença:** a diferença de A e B é o conjunto de elementos em A que não estão em B. Nós escrevemos isso como $A–B$;

Exemplo: Sejam A, B e C eventos associados a um experimento aleatório. Expresse em notação de conjuntos e faça os diagramas de Venn:

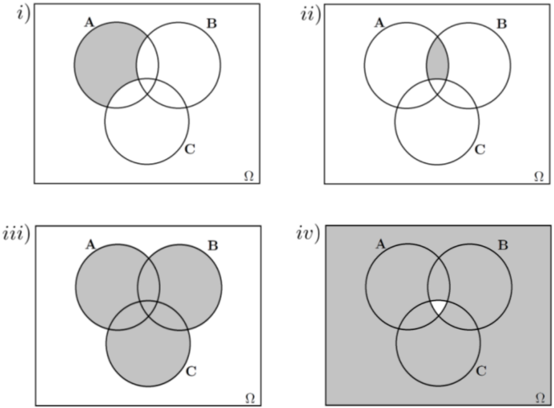
i Somente A ocorre;

ii A e B ocorrem, mas C não;

iii Pelo menos um deles ocorre;

iv Não mais que dois deles ocorrem.

Teoria de conjuntos

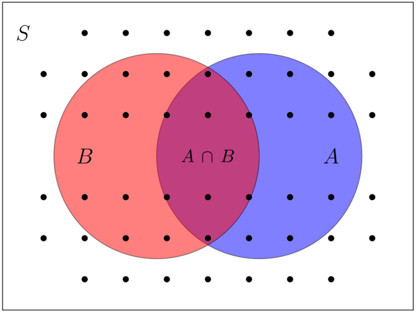


Princípio da inclusão-exclusão

- O Princípio da inclusão-exclusão diz:

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

- A figura abaixo exemplifica essa questão. *S* representa todos os pontos, *A* são os pontos no interior do círculo azul, e *B* no interior do vermelho.



Exemplo

- Uma banda é composta de cantores e guitarristas:
- 7 pessoas cantam;
- 4 tocam guitarra;
- 2 fazem ambos.
- Quantas pessoas estão na banda?

Resposta: seja C o conjunto dos cantores e G o conjunto dos guitarristas. O tamanho da banda é dado por:

$$\text{Tamanho da banda} = |C \cup G| = |C| + |G| - |C \cap G| = 7 + 4 - 2 = 9.$$

Produto cartesiano

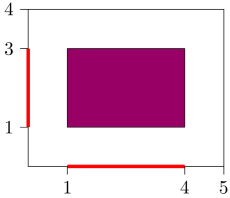
- O produto cartesiano de *R* e *S* é dado pelo par ordenado:

$$S \times T = \{(s,t) | s \in S, t \in T\}$$

Exemplo

$\times$	1	2	3	4
1	(1,1)	(1,2)	(1,3)	(1,4)
2	(2,1)	(2,2)	(2,3)	(2,4)
3	(3,1)	(3,2)	(3,3)	(3,4)

$\{1,2,3\} \times \{1,2,3,4\}$



$[1,4] \times [1,3] \subset [0,5] \times [0,4]$

Regra do produto

- A Regra do Produto diz:

  "Se houver *n* maneiras de executar a ação 1 e *m* maneiras de executar a ação 2, então existem $n \times m$ maneiras para executar a ação 1 seguida da ação 2."

- Também chamaremos isso de regra de multiplicação.

Exemplo: Se você tem 3 camisas e 4 calças, então você pode fazer $3 \times 4 = 12$ combinações de roupas.

Exemplo: Existem 5 competidores na final de 100*m* nas Olimpíadas. De quantas formas as medalhas de ouro, prata e bronze podem ser distribuídas? $5 \times 4 \times 3$

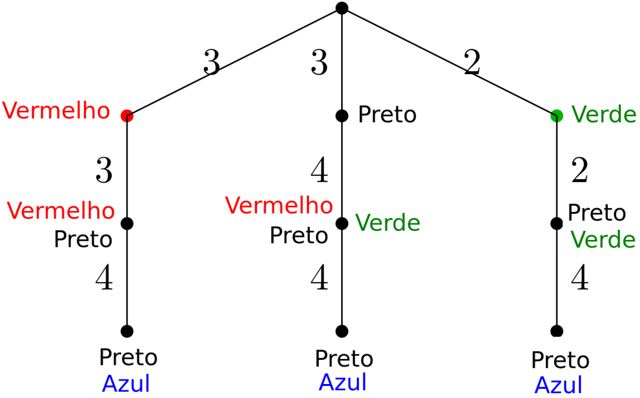
Exemplo

- Eu não usarei verde e vermelho juntos. E acho que o preto ou o azul vão bem com qualquer coisa. Abaixo está o meu guarda-roupa.
  - Camisas: 3 pretas, 3 vermelhas e 2 verdes;
  - Blusas: 1 preta, 2 vermelhas, 1 verde;
  - Calças: 2 azuis, 2 pretas.

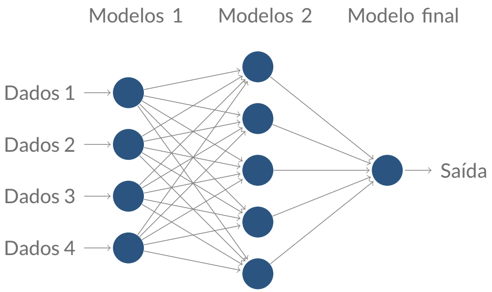
- Quantas combinações de roupa posso utilizar?

Exemplo

$$\text{Número de possibilidades} = (3 \times 3 \times 4) + (3 \times 4 \times 4) + (2 \times 2 \times 4) = 100$$



Combinando vários modelos



- As camadas de modelos envolvem algoritmos do tipo:
  - Regressão Linear;
  - Regressão logística;
  - KNN;
  - Gradiente Boosting;
  - Naive Bayes;
  - Redes Neurais Artificiais;
  - Árvores de decisão;
  - Random Forests etc.

Cáculo de Probabilidades

- A probabilidade é uma quantidade que pode ser utilizada para se medir a incerteza sobre certos eventos ou características de interesse;

- Tais eventos, em geral, estão associados a **experimentos aleatórios** (experimento para qual não se tem certeza sobre seu resultados, a priori);

- As probabilidades geralmente são baseadas em:

    - **Procedimento empírico:** calculada com base nos valores observados.

    - **Procedimento teórico:** proposto pelo pesquisador para representar a distribuição de frequência populacional.

Exemplo

- **Exemplo:** estudar as probabilidades de ocorrência das faces de um dado.
  - **Procedimento empírico:** lançar o dado um certo número de vezes e contar quantas vezes a face $i = 1, 2, 3, 4, 5, 6$ ocorreu;
    - $f_i = \frac{n_i}{n}$ é a distribuição empírica das probabilidades;
    - Para diferentes vezes que esse experimento for realizado, a distribuição de frequência terá resultados diferentes;
  - **Procedimento teórico:** construir a distribuição de frequências populacionais (probabilidades) através de suposições teóricas.

Distribuição de Frequências

- Exemplo: Salário (variável quantitativa contínua)

Faixa salarial	Frequência (n_i)	Proporção (f_i)	% ($100 \times f_i$)
4.00-8.00	10	0.2778	27.78
8.00-12.00	12	0.3333	33.33
12.00-16.00	8	0.2222	22.22
16.00-20.00	5	0.1389	13.89
20.00-24.00	1	0.0278	2.78
Total	36	1.0000	100.00

Faixa salarial	Capital	Interior	Outro	Marginal
$[4-8)$	4	3	3	10
$[8-12)$	3	4	6	13
$[12-16)$	1	3	3	7
$[16-20)$	3	1	1	5
$[20-24]$	0	1	0	1
Marginal	11	12	13	36

Espaço Amostral

- São todos os resultados possíveis do experimento (aleatório), denotado por $\Omega = \{\omega_1, \omega_2, ...\}$;

  Exemplo: lançar uma moeda duas vezes:

$$C = cara \qquad K = coroa$$

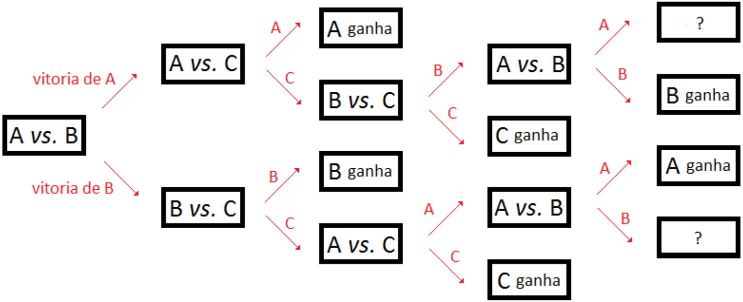
  - $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$
    $\omega_1 = (C, C); \omega_2 = (C, K); \omega_3 = (K, C); \omega_4 = (K, K);$
  - Considerando que a moeda é honesta: $P(\omega_i) = \frac{1}{4}, \forall\, i = 1, 2, 3, 4.$
  - Seja o evento $A = \{\omega_1, \omega_4\} =$ obter duas faces iguais
    $P(A) = P(\{\omega_1, \omega_4\}) = P(\omega_1) + P(\omega_4) = \frac{1}{4} + \frac{1}{4} = \frac{2}{4} = \frac{1}{2}.$

Exemplo

- Três jogadores **A**, **B** e **C** disputam um torneio de tênis;
- Inicialmente, **A** joga com **B** e o vencedor joga com **C**, e assim por diante;
- O torneio termina quando um jogador ganha duas vezes seguidas ou quando são disputadas, ao todo, quatro partidas;
- Quais são os resultados possíveis do torneio?

Fonte: Morettin & Bussab, Estatística Básica 5^a edição, pág 105.

Solução



$$\Omega = \{AA, BB, ACC, BCC, ACBA, ACBB, BCAA, BCAB\}$$

Exercício

- Defina o espaço amostral dos seguintes experimentos aleatórios:

  (i) Numa linha de produção, conta-se o número de peças defeituosas num intervalo de uma hora;

  (ii) Investigam-se famílias com três crianças, anotando-se a configuração segundo o sexo;

  (iii) Numa entrevista telefônica com 250 assinantes, anota-se se o proprietário tem ou não máquina de secar roupa;

  (iv) Uma lâmpada é retirada de um lote e é medido seu tempo de vida antes de queimar.