Universidade de São Paulo

Instituto de Ciências Matemáticas e de Computação

# Fundamentos de Probabilidade e Estatística para Ciência de Dados

Resumo das aulas do Prof. Dr. Francisco Rodrigues

Bruna Zamith Santos

Agosto de 2025

# Sumário

# 1 Teoria dos Conjuntos

Sejam os conjuntos:

$$A = \{1, 2, 4, 9\}, \quad B = \{3, 7, 9\}$$

- União: $A \cup B = \{1, 2, 3, 4, 7, 9\}$
- Interseção: $A \cap B = \{7, 9\}$
- Complementar de $B$: $B^C = \{1, 2, 4\}$
- Complementar de $A$: $A^C = \{3, 7\}$
- Espaço amostral ($\Omega$): É o conjunto de todos os resultados possíveis de um experimento aleatório. Exemplo: $\Omega = \{1, 2, 3, 4, 5, 6\}$, ao lançar um dado.
- Evento ($A$): É um subconjunto do espaço amostral. Exemplo: $A = \{2, 4, 6\}$
- Evento impossível ($\emptyset$): É um evento que nunca ocorre.
- Evento certo ($\Omega$): É o evento que sempre ocorre.
- $A \cup B$: É o evento que ocorre se $A$ ou $B$ (ou ambos) ocorrerem.
- $A \cap B$: É o evento que ocorre se $A$ e $B$ ocorrerem ao mesmo tempo.
- $A^C$: É o evento que ocorre se $A$ não ocorre.
- Eventos mutuamente exclusivos: Quando $A \cap B = \emptyset$.

# 2 Experimento Aleatório

Um experimento aleatório é um experimento que pode ser repetido inúmeras vezes sob as mesmas condições, sendo o seu resultado incerto.

# 3 Conceitos de Probabilidade

Sejam $\Omega$ o espaço amostral e $A$ um evento em $\Omega$. Então, uma função $P(\cdot)$ é denominada probabilidade se satisfaz:

- $0 \leq P(A) \leq 1$, $\forall A \in \Omega$
- $P(\Omega) = 1$
- Se $A_1, A_2, \ldots$ forem eventos mutuamente exclusivos, isto é, $A_i \cap A_j = \emptyset$, $\forall i \neq j$, então:
$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Se um experimento aleatório tiver $n(\Omega)$ resultados mutuamente exclusivos e igualmente possíveis, e se um evento $A$ conter $n(A)$ desses resultados, a probabilidade de ocorrência desse evento é definida por:

$$P(A) = \frac{n(A)}{n(\Omega)} = \frac{|A|}{|\Omega|}$$

Sejam $A$ e $B$ eventos em um mesmo espaço amostral, então:

- $P(\emptyset) = 0$
- $P(A) = 1 - P(A^C)$
- Se $A \subseteq B$, então $P(A) \leq P(B)$

## 3.1 Probabilidade Frequentista

A probabilidade de um evento é igual à sua frequência de ocorrência em um grande número de experimentos:

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n}$$

, onde $n_A$ é o número de vezes que o evento $A$ ocorre em $n$ experimentos.

## 3.2 Probabilidade de União de Dois Eventos

Para dois eventos $A$ e $B$ em um mesmo espaço amostral:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

## 3.3 Probabilidade Condicional

Sejam dois eventos $A$ e $B$ em um mesmo espaço amostral $\Omega$. A probabilidade condicional de $A$ dado que $B$ ocorreu é definida por:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad \text{com } P(B) > 0$$

Assim, $A$ e $B$ são eventos independentes se, e somente se:

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

Ou equivalentemente:

$$P(A \mid B) = P(A) \quad \text{e} \quad P(B \mid A) = P(B)$$

### 3.4 Partições do Espaço Amostral

Os eventos $B_1, B_2, \ldots, B_n$ formam uma partição do espaço amostral $\Omega$ se:

- $B_i \cap B_j = \emptyset$, para $i \neq j$, com $i, j = 1, \ldots, n$
- $\bigcup_{i=1}^{n} B_i = \Omega$
- $P(B_i) \geq 0$, para $i = 1, \ldots, n$

Seja $A$ um evento no espaço amostral $\Omega$ e seja $B_1, \ldots, B_n$ uma partição amostral de $\Omega$. Podemos escrever $A$ considerando tal partição:

$$A = \bigcup_{i=1}^{n} (A \cap B_i)$$

$$P(A) = P\left(\bigcup_{i=1}^{n} A \cap B_i\right) = \sum_{i=1}^{n} P(A \cap B_i)$$

Sejam $B_1, B_2, \ldots, B_n$ uma partição do espaço amostral $\Omega$. Então, qualquer evento $A \subseteq \Omega$ pode ser escrito como:

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i) \cdot P(B_i)$$

# 4 Teorema de Bayes

Sejam $B_1, B_2, \ldots, B_n$ uma partição do espaço amostral $\Omega$, e $A$ um evento com $P(A) > 0$, então:

$$P(B_i \mid A) = \frac{P(A \mid B_i) \cdot P(B_i)}{\sum_{j=1}^{n} P(A \mid B_j) \cdot P(B_j)}$$

E assim podemos definir:

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

# 5 Variáveis Aleatórias

Suponha que lancemos dois dados. O espaço amostral associado ao experimento, sendo os eventos $C$: "sai uma cara" e $R$: "sai uma coroa", é dado por:

$$\Omega = \{CC, CR, RC, RR\}$$

Uma possível variável aleatória associada ao experimento é definida por:

$$X = \text{“número de caras obtido no experimento”}$$

- Representamos variáveis aleatórias por letras maiúsculas $(X, Y, Z)$, enquanto usamos letras minúsculas para indicar os valores das variáveis $(x, y, z)$.
- Se o número de valores possíveis de uma variável aleatória for finito ou infinito enumerável, dizemos que é uma variável aleatória discreta.
- Caso contrário, é uma variável aleatória contínua.

A função que atribui a cada valor da variável aleatória sua respectiva probabilidade é chamada de distribuição de probabilidade:

$$P(X = x_i) = p(x_i) = p_i, \quad i = 1, 2, 3$$

A distribuição de probabilidade também é chamada de função massa de probabilidade. E temos que:

$$\sum_{i=1}^{n} P(X = x_i) = 1$$

Dizemos que $X$ é uma variável aleatória contínua se existir uma função $f$ denominada função densidade de probabilidade (fdp) que satisfaz:

- $f(x) \geq 0, \quad \forall x \in \mathbb{R}$
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$
- $P(a \leq X \leq b) = \int_a^b f(x)\,dx, \ -\infty < a < b < \infty$
- $f(x)$ é uma função com valores positivos e área unitária.

Seja $X$ uma variável aleatória discreta ou contínua. A probabilidade condicional de que $X \in S$ dado que $X \in V$ é:

$$P(X \in S \mid X \in V) = \frac{P(X \in S \cap V)}{P(X \in V)}$$

, onde $S$ e $V$ são subconjuntos do espaço da variável.

# 6 Função de Distribuição

A função distribuição acumulada ou simplesmente função de distribuição de uma variável aleatória $X$ é definida por:

$$F(x) = P(X \leq x)$$

Se discreta:

$$F(x) = \sum_{x_i \leq x} P(X = x_i)$$

Se contínua:

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

Propriedades da função de distribuição:

- $0 \leq F(x) \leq 1, \quad F(x)$ é não decrescente,
- $\lim_{x\to-\infty} F(x) = 0, \quad \lim_{x\to+\infty} F(x) = 1$
- Caso discreto: $P(a < X \leq b) = F(b) - F(a)$
- Caso contínuo: $f(x) = \frac{dF(x)}{dx}$

# 7 Esperança

## 7.1 Variável Aleatória Discreta

Seja $X$ uma variável aleatória discreta com distribuição de probabilidade $P(X = x_i)$. O valor esperado (ou esperança matemática) é:

$$E[X] = \sum_{i=1}^{n} x_i \cdot P(X = x_i)$$

## 7.2 Variável Aleatória Contínua

Seja $X$ uma variável aleatória contínua com função densidade de probabilidade $f(x)$, então:

$$E[X] = \int_{-\infty}^{+\infty} x \cdot f(x)\,dx$$

## 7.3 Função de uma Variável Aleatória

Seja $g(X)$ uma função de uma variável aleatória discreta $X$. Então:

$$E[g(X)] = \sum_{i=1}^{n} g(x_i) \cdot P(X = x_i)$$

Seja $g(X)$ uma função de variável contínua com densidade $f(x)$. Então:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f(x)\,dx$$

## 7.4 Propriedades

- Se $X = c$, onde $c$ é constante, então: $E[X] = E[c] = c$
- Se $c$ é constante: $E[cX] = c \cdot E[X]$
- Então: $E[aX + b] = a \cdot E[X] + b$

# 8 Momento

## 8.1 Momento Estatístico

Seja $X$ uma variável aleatória discreta com valores $x_1, x_2, \ldots, x_k$. O momento de ordem $n$ de $X$ é:

$$E[X^n] = \sum_{i=1}^{k} x_i^n \cdot P(X = x_i)$$

Se $X$ for contínua:

$$E[X^n] = \int_{-\infty}^{\infty} x^n f(x)\, dx$$

## 8.2 Momento Central

Seja $X$ uma variável aleatória.

- Se $X$ é discreta, o momento central de ordem $n$ ($n > 0$) de $X$ é:

$$\mu_n = E\left[(X - E[X])^n\right] = \sum_{x_i} (x_i - E[X])^n \cdot P(X = x_i)$$

- Se $X$ é contínua, então:

$$\mu_n = E\left[(X - E[X])^n\right] = \int_{-\infty}^{\infty} (x - E[X])^n \cdot f(x)\, dx$$

# 9 Variância

A variância de uma variável aleatória $X$ é definida por:

$$V(X) = \sigma^2 = E\left[(X - E(X))^2\right]$$

O desvio padrão é igual à raiz quadrada da variância:

$$\sigma = \sqrt{V(X)}$$

Temos a propriedade de que:

$$V(X) = E[X^2] - (E[X])^2$$

Seja $g(X)$ uma função da variável aleatória $X$. Então,

$$V[g(X)] = E[g(X)^2] - \left(E[g(X)]\right)^2$$

Seja $X$ uma variável aleatória e $a$ e $b$ constantes. Então,

$$V(aX + b) = a^2 \cdot V(X)$$

# 10 Modelos Probabilísticos (Estocásticos) Discretos

- Os resultados de cada experimento parecem imprevisíveis, mas quando um grande número de experimentos é analisado, surge um padrão.
- Não podemos determinar o valor exato do resultado de um experimento, mas sim as probabilidades de cada resultado possível.

## 10.1 Distribuição Uniforme Discreta

Seja $X$ uma variável aleatória discreta assumindo os $n$ valores $\{a, a+c, a+2c, \ldots, b-c, b\}$, com $a, b \in \mathbb{R}$, $c \in \mathbb{R}_{>0}$ e $a < b$.

Dizemos que $X$ segue o modelo uniforme discreto se atribuímos a mesma probabilidade $1/n$ a cada um desses valores. Isto é, sua distribuição de probabilidade é dada por:

$$P(X = x) = \frac{1}{n}, \quad x = a, a+c, a+2c, \ldots, b$$

onde:

$$n = 1 + \frac{b-a}{c}$$

Então,

$$\mathbb{E}[X] = \frac{a+b}{2},$$

$$V(X) = \frac{c^2(n^2-1)}{12}$$

A Figura 1 apresenta um exemplo de distribuição uniforme discreta.

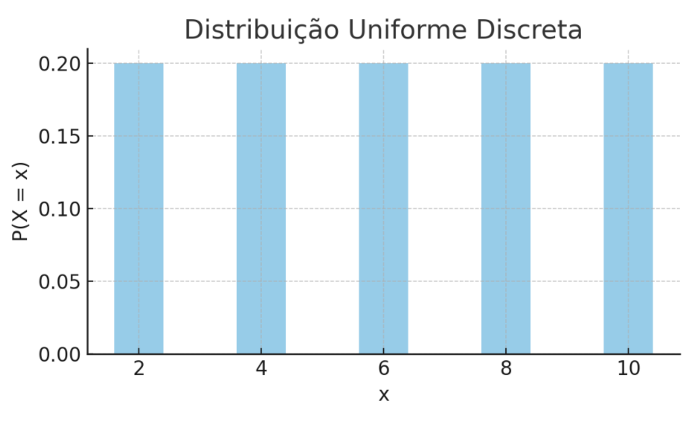


Figura 1: Distribuição uniforme discreta.

## 10.2 Distribuição de Bernoulli

Dizemos que a variável aleatória $X$ segue o modelo de Bernoulli se atribuímos 0 à ocorrência de um fracasso ou 1 à ocorrência de um sucesso, com $p$ representando a probabilidade de sucesso, $0 \leq p \leq 1$, e $1-p$ a probabilidade de fracasso.

A distribuição de probabilidade é dada por:

$$P(X=k) = p^k(1-p)^{1-k}, \quad k = 0, 1$$

| $X$ | 0 | 1 |
|---|---|---|
| $P(X=k)$ | $1-p$ | $p$ |

Então,

$$\mathbb{E}[X] = p,$$

$$V(X) = p \cdot (1-p)$$

A Figura 2 apresenta um exemplo de distribuição de Bernoulli discreta.

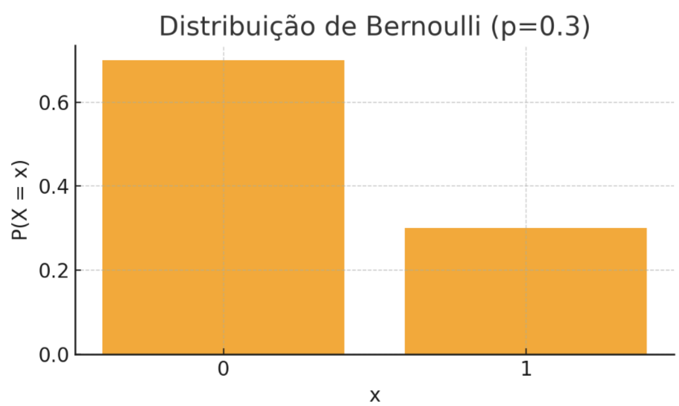


Figura 2: Distribuição de Bernoulli discreta.

## 10.3 Distribuição Binomial

O processo estocástico de Bernoulli possui as seguintes propriedades:

- O experimento consiste de $n$ tentativas repetidas.
- Cada tentativa gera um resultado que pode ser classificado como sucesso ou falha.
- A probabilidade de sucesso $p$ se mantém constante de tentativa para tentativa.
- As tentativas são feitas de forma independente uma da outra.

Seja $X$ uma variável aleatória baseada em $n$ repetições de um processo de Bernoulli. Então a probabilidade de obtermos $k$ sucessos em $n$ repetições é dada por:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \ldots, n$$

, onde:

$$C_n^k = \binom{n}{k} = \frac{n!}{(n-k)!\,k!}$$

é uma combinação de $n$ elementos tomados de $k$ em $k$.

Então,

$$\mathbb{E}[X] = n \cdot p,$$

$$V(X) = n \cdot p \cdot (1-p)$$

A Figura 3 apresenta um exemplo de distribuição de binomial discreta.

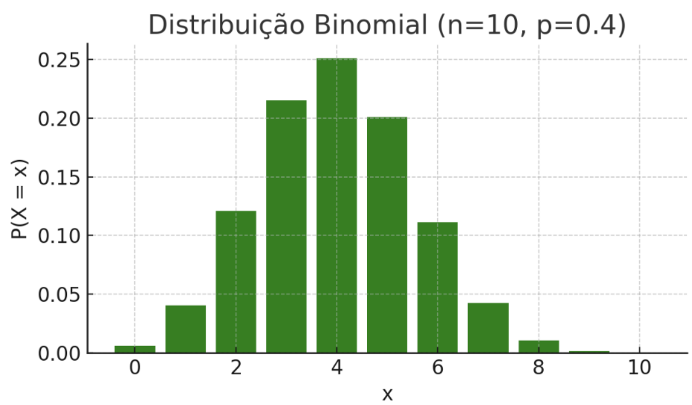


Figura 3: Distribuição binomial discreta.

## 10.4 Distribuição de Poisson

Uma variável aleatória discreta $X$ segue o modelo de Poisson com taxa $\lambda > 0$ se:

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, 2, \ldots$$

Então,

$$\mathbb{E}[X] = \lambda,$$

$$V(X) = \lambda$$

A Figura 4 apresenta um exemplo de distribuição de Poisson discreta.

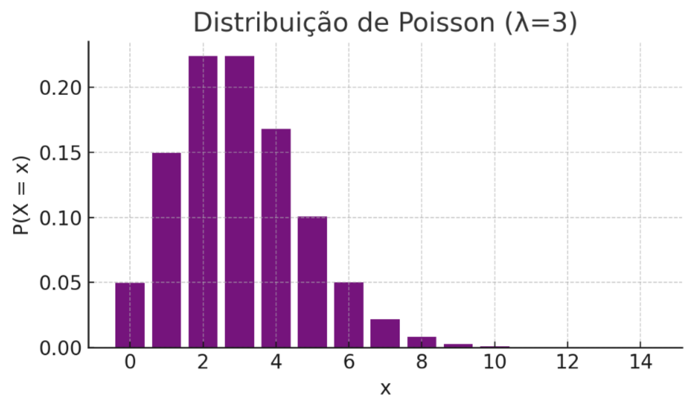


Figura 4: Distribuição de Poisson discreta.

## 10.5 Lei dos Eventos Raros

Seja $X$ uma variável aleatória com distribuição binomial e $p$ a probabilidade de sucesso. Então,

$$\lim_{n\to\infty} \binom{n}{k} p^k (1-p)^{n-k} = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots,$$

onde $\lambda = np$ é constante.

## 10.6 Distribuição Geométrica

Dizemos que a variável aleatória discreta $X$ segue uma distribuição geométrica se:

$$P(X = k) = p(1-p)^{k-1}, \quad k = 1, 2, \ldots$$

Então,

$$\mathbb{E}[X] = \frac{1}{p},$$

$$V(X) = \frac{1-p}{p^2}$$

A Figura 7 apresenta um exemplo de distribuição geométrica discreta.

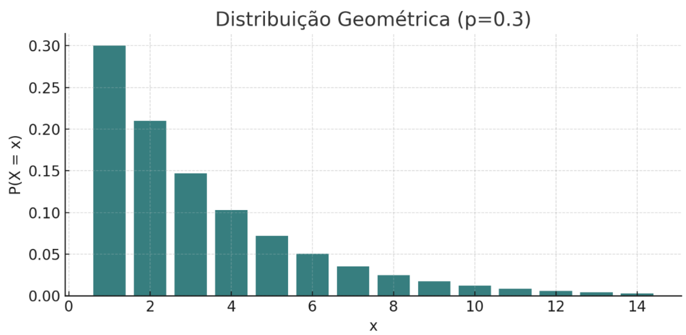


Figura 5: Distribuição geométrica discreta.

## 10.7 Distribuição Binomial Negativa

Seja $X$ o número de repetições necessárias a fim de que ocorram exatamente $r$ sucessos, de modo que o $r$-ésimo sucesso ocorra na $k$-ésima tentativa. Então,

$$P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \quad k = r, r+1, \ldots$$

A Figura 6 apresenta um exemplo de distribuição binomial negativa discreta.

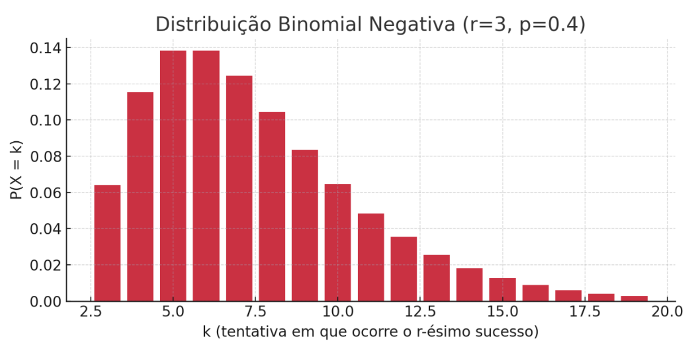


Figura 6: Distribuição binomial negativa discreta.

## 10.8 Distribuição Hipergeométrica

Considere um conjunto de $N$ objetos, dos quais $N_1$ são do tipo 1 e $N_2 = N - N_1$ são do tipo 2. Para um sorteio de $n$ objetos ($n < N$), sem reposição, seja $X$ a variável aleatória que define o número de objetos do tipo 1 sorteados. Então, a probabilidade de sortearmos $k$ objetos do tipo 1 é:

$$P(X = k) = \frac{\binom{N_1}{k}\binom{N-N_1}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, \ldots, n$$

A Figura 7 apresenta um exemplo de distribuição hipergeométrica discreta.

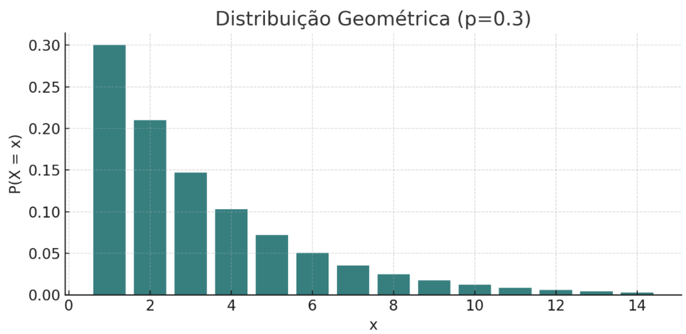


Figura 7: Distribuição hipergeométrica discreta.