

# Cálculo de Probabilidade 1

Caio Tomás de Paula

14 de setembro de 2022

## Sumário

<b>1</b>	<b>Espaço de probabilidade</b>	<b>3</b>
1.1	Definição e exemplos . . . . .	3
1.2	Condicionamento e independência . . . . .	11
1.3	Exercícios . . . . .	15
<b>2</b>	<b>Noções de Análise Combinatória</b>	<b>21</b>
2.1	Amostras ordenadas . . . . .	21
2.2	Amostras desordenadas e sem reposição (combinações) . . . . .	22
2.3	Permutações com elementos repetidos . . . . .	22
2.4	Partições — problema da urna . . . . .	23
<b>3</b>	<b>Variáveis aleatórias</b>	<b>24</b>
3.1	Variáveis aleatórias discretas . . . . .	26
3.2	Exemplos clássicos de v.a.'s discretas . . . . .	27
3.3	Exercícios - combinatória . . . . .	35
3.4	Exercícios - v.a.'s discretas . . . . .	38
3.5	Vetores aleatórios discretos (variáveis aleatórias multidimensionais) . . . . .	41
3.6	Variáveis aleatórias independentes . . . . .	47
3.7	Funções de variáveis aleatórias . . . . .	49
3.8	Distribuição condicional de variáveis aleatórias discretas . . . . .	51
3.9	Exercícios . . . . .	52
<b>4</b>	<b>Esperança de variáveis aleatórias discretas</b>	<b>57</b>
4.1	Esperança de função de variável aleatória . . . . .	59
4.2	Momentos . . . . .	62
4.3	Exercícios - esperança e funções de v.a.'s discretas . . . . .	67
4.4	Exercícios - variância e covariância . . . . .	69
<b>5</b>	<b>Variáveis aleatórias contínuas</b>	<b>71</b>
5.1	Exemplos clássicos de distribuições contínuas . . . . .	75
5.2	Funções de variáveis aleatórias contínuas . . . . .	79
<b>6</b>	<b>Vetores aleatórios contínuos</b>	<b>82</b>
6.1	Definições gerais . . . . .	82
6.2	Distribuições marginais e independência . . . . .	82
6.3	Função de distribuição $n$ -dimensional . . . . .	86
6.4	Funções de vetores aleatórios . . . . .	86
<b>7</b>	<b>Esperança de variáveis aleatórias contínuas</b>	<b>94</b>
7.1	Esperança de função de variável aleatória contínua . . . . .	95
7.2	Momentos de uma variável aleatória contínua . . . . .	96

7.3	Exercícios - v.a.'s contínuas . . . . .	98
7.4	Exercícios - funções de v.a.'s contínuas . . . . .	99
7.5	Exercícios - soma e quociente de v.a.'s contínuas . . . . .	101
<b>8</b>	<b>Função Geradora de Momentos, Lei dos Grandes Números e o Teorema do Limite Central</b>	<b>105</b>
8.1	Função Geradora de Momentos . . . . .	105
8.2	Desigualdades de Chebyshev-Markov e Lei dos Grandes Números . . . . .	106
8.3	Teorema do Limite Central . . . . .	108
8.4	Funções Características . . . . .	112
8.5	Fórmulas de inversão e o Teorema da Continuidade . . . . .	114
8.6	Demonstração da Lei Fraca dos Grandes Números e do Teorema do Limite Central .	116
8.7	Exercícios - esperança e momentos de v.a.'s contínuas . . . . .	118

# Introdução

Essas notas consistem de um compilado, com algumas adições, das notas de aula do curso de Cálculo de Probabilidade 1 ministrado de agosto a novembro de 2021 (1º semestre letivo de 2021 da Universidade de Brasília) pela prof<sup>a</sup> Daniele da Silva Baratela Martins Neto. Qualquer dúvida, comentário ou sugestão é bem-vinda: basta entrar em contato pelo e-mail caiotomas6@gmail.com.

## 1 Espaço de probabilidade

### 1.1 Definição e exemplos

A **Teoria da Probabilidade** é um ramo da Matemática que estuda, através de modelos matemáticos, fenômenos (ou experimentos) aleatórios, isto é, experimentos que, quando repetidos sob condições semelhantes, produz resultados diferentes em geral.

| **Exemplo.** Lançar uma moeda e observar o resultado (cara ou coroa).

Vemos, então, que experimentos aleatórios diferem bastante de experimentos **determinísticos**, isto é, experimentos que, quando repetidos sob condições semelhantes, produzem resultados idênticos.

| **Exemplo.** Água aquecida a 100°C ao nível do mar entra em ebulição.

Algumas características de um experimento aleatório são:

- i. pode ser repetido indefinidamente sob condições inalteradas;
- ii. permite descrever todos os possíveis resultados do experimento;
- iii. quando executado um número grande de vezes, os resultados apresentam certa regularidade (e isso nos permite construir modelos matemáticos para o experimento).

| **Exemplos.** Alguns exemplos são:

- 1) Lançar um dado e observar seu resultado
- 2) Jogar uma moeda duas vezes e observar seus resultados
- 3) Tem-se duas caixas:
  - a caixa I contém 5 bolas numeradas de 1 a 5;
  - a caixa II contém 6 bolas numeradas de 1 a 6.

Retirar uma bola de cada caixa e observar os números de cada bola consiste em um experimento aleatório

- 4) Jogar uma moeda até que apareça cara pela primeira vez
- 5) Escolher, ao acaso, um número real entre 0 e 1
- 6) Medir, em horas, o tempo de vida útil de uma lâmpada

Nosso objetivo aqui é construir um modelo matemático, que recebe o nome de **espaço de probabilidade**, para uma análise matemática destes experimentos aleatórios. Para tanto, seja  $\mathcal{E}$  um experimento aleatório.

**Definição 1.1.** Seja  $\mathcal{E}$  um experimento aleatório cujo conjunto de todos os resultados possíveis são conhecidos. Esse conjunto, denotado  $\Omega$ , é chamado **espaço amostral** de  $\mathcal{E}$ . Um ponto  $\omega \in \Omega$  é chamado **ponto amostral**.

| **Observações.** Note que o espaço amostral é um conjunto amostral. Além disso,

- (i)  $\Omega$  pode conter mais pontos que o necessário, mas não pode excluir nenhum resultado possível
- (ii)  $\Omega$  pode ser finito,

infinito enumerável,

$$\Omega = \{x_1, x_2, \dots, x_N\},$$

$$\Omega = \{x_1, x_2, \dots\}$$

ou, ainda, infinito não-enumerável,

$\Omega \subset \mathbb{R}$  não enumerável, e.g., um intervalo.

**Exemplos.** Considere os 5 exemplos de experimentos aleatórios anteriores e denote por  $\Omega_i$  o espaço amostral associado ao  $i$ -ésimo experimento anterior. Temos

- 1)  $\Omega_1 = \{1, 2, 3, 4, 5, 6\}$
- 2)  $\Omega_2 = \{cc, c\hat{c}, \hat{c}c, \hat{c}\hat{c}\}$ , sendo  $c$  cara e  $\hat{c}$  coroa
- 3)  $\Omega_3 = \{(1, 1), (1, 2), \dots, (1, 6), (2, 1), (2, 2), \dots, (5, 6)\} = \{(i, j) : i \in \{1, 2, \dots, 5\}, j \in \{1, 2, \dots, 6\}\}$
- 4)  $\Omega_4 = \{c, \hat{c}c, \hat{c}\hat{c}c, \hat{c}\hat{c}\hat{c}c, \dots\} = \{1, 2, 3, 4, 5, \dots\}$ , onde  $\omega = j$  se, e só se, a primeira cara saiu no  $j$ -ésimo lançamento, ou seja,  $\underbrace{\hat{c}\hat{c}\dots\hat{c}}_{j-1}c$ .
- 5)  $\Omega_5 = [0, 1] = \{x \in \mathbb{R} : 0 \leq x \leq 1\}$
- 6)  $\Omega_6 = [0, +\infty) = \{t \in \mathbb{R} : t \geq 0\}$

**Definição 1.2.** Um subconjunto  $A \subseteq \Omega$  do espaço amostral de um experimento  $\mathcal{E}$  é chamado **evento**.

Na prática, os eventos são os subconjuntos do espaço amostral para os quais desejamos saber o quão provável será a sua ocorrência quando o experimento for realizado. Dito de outro modo, são os subconjuntos de  $\Omega$  para os quais desejamos atribuir probabilidades.

**Exemplos.** Considere os mesmos experimentos acima.

- 1) Para  $\mathcal{E}_1$ ,  $A = \{\text{sair um número ímpar}\} = \{1, 3, 5\}$  e  $B = \{\text{sair 1 ou 6}\} = \{1, 6\}$  são eventos
- 2) Para  $\mathcal{E}_3$ ,  
 $A = \{\text{soma dos resultados é par}\} = \{1, 3, 5\} = \{(1, 1), (1, 3), (1, 5), (2, 2), (2, 4), \dots, (5, 5)\}$   
 $= \{(i, j) : i, j \in \{1, 2, 3, 4, 5, 6\}, i \neq 6 \text{ e } i + j \text{ é par}\}$   
é um evento
- 3) Para  $\mathcal{E}_6$ ,  $A = \{\text{lâmpada queima após 10 horas de uso}\} = \{t \in \mathbb{R} : t > 10\} = (10, +\infty)$  é um evento

**Observações.** Note que dois exemplos triviais (mas importantes!) de eventos são  $\emptyset$  e  $\Omega$ . Além disso, perceba que o espaço amostral de um mesmo experimento pode ser representado de maneiras distintas.

**Definição 1.3.** Uma coleção  $\mathcal{A}$  de subconjuntos de  $\Omega$  tal que

- (i)  $\mathcal{A} \neq \emptyset$
- (ii)  $A \in \mathcal{A} \implies A^C = \Omega \setminus A \in \mathcal{A}$ , isto é, se  $A$  é um evento de  $\Omega$ , então  $A^C$  também é
- (iii)  $A_1, A_2, \dots \in \mathcal{A} \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ , isto é, a união enumerável de eventos de  $\Omega$  é também um evento de  $\Omega$

é chamada  **$\sigma$ -álgebra** de eventos/subconjuntos de  $\Omega$ .

**Proposição 1.4.** Se  $\mathcal{A}$  é uma  $\sigma$ -álgebra de subconjuntos de  $\Omega$ , então

- (a)  $A, B \in \mathcal{A} \implies A \cup B \in \mathcal{A}$
- (b)  $\emptyset, \Omega \in \mathcal{A}$
- (c)  $A, B \in \mathcal{A} \implies A \cap B \in \mathcal{A}$
- (d)  $A_1, A_2, \dots \in \mathcal{A} \implies \bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$

**Demonstração.** Demonstramos cada item.

- (a) Tomando  $A_1 = A, A_2 = B, A_i = \emptyset, i \geq 3$ , segue da definição que  $\bigcup_{i=1}^{\infty} A_i = A \cup B \in \mathcal{A}$ .

- (b) Como  $\mathcal{A} \neq \emptyset$ , existe  $A \in \mathcal{A}$ . Logo, por definição de  $\sigma$ -álgebra,  $A^C \in \mathcal{A}$ . Ora, como  $\Omega = A \cup A^C$ , então  $\Omega \in \mathcal{A}$  e, daí,  $\Omega^C = \emptyset \in \mathcal{A}$ .
- (c) Como  $A, B \in \mathcal{A}$ , então  $A^C, B^C \in \mathcal{A}$ . Daí, por (b),  $A^C \cup B^C \in \mathcal{A}$  e, portanto,  $(A^C \cup B^C)^C = A \cap B \in \mathcal{A}$ .
- (d) Se  $A_1, A_2, \dots \in \mathcal{A}$ , então  $A_1^C, A_2^C, \dots \in \mathcal{A}$ . Logo,  $\bigcup_{i=1}^{\infty} A_i^C \in \mathcal{A}$  e, portanto,  $\left(\bigcup_{i=1}^{\infty} A_i^C\right)^C = \bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$

□

**Observações.** Dado um conjunto  $\Omega$  qualquer, temos

- (i)  $\mathcal{A} = \{\emptyset, \Omega\}$  é a menor  $\sigma$ -álgebra de  $\Omega$ , denominada  **$\sigma$ -álgebra trivial**
- (ii)  $\mathcal{A} = \mathcal{P}(\Omega) = \{A : A \subset \Omega\}$  é a maior  $\sigma$ -álgebra de  $\Omega$ , denominada  **$\sigma$ -álgebra das partes**.

Os termos “maior” e “menor” se referem à inclusão de conjuntos.

**Exemplo.** Considere  $\mathcal{E}$  : jogar um dado honesto. Como vimos, nesse caso podemos tomar  $\Omega = \{1, 2, 3, 4, 5, 6\}$ . Daí,  $\mathcal{A} = \{\emptyset, \Omega, \{1, 2, 3\}, \{4, 5, 6\}\}$  é uma  $\sigma$ -álgebra de subconjuntos de  $\Omega$ . Note que  $A = \{\text{sair um número par}\} = \{2, 4, 6\} \notin \mathcal{A}$ . Isso ilustra a necessidade de considerarmos as maiores (novamente, no sentido de inclusão de conjuntos)  $\sigma$ -álgebras possíveis.

Listamos as três  $\sigma$ -álgebras mais usuais.

- (a) Se  $\Omega = \{\omega_1, \omega_2, \dots, \omega_N\}$  (finito) ou  $\Omega = \{\omega_1, \omega_2, \dots\}$  (infinito enumerável), tomamos  $\mathcal{A} = \mathcal{P}(\Omega)$ , ou seja, a  $\sigma$ -álgebra das partes de  $\Omega$ . Note que  $|\mathcal{A}| = 2^N$  se  $\Omega$  é finito.

**Exemplo.** Se  $\Omega = \{1, 2, 3\}$ , então a  $\sigma$ -álgebra das partes de  $\Omega$  é  $\mathcal{A} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \Omega\}$ .

- (b) Se  $\Omega \subseteq \mathbb{R}$  é não-enumerável, tomamos  $\mathcal{A} = \mathcal{B}(\Omega)$ : a  **$\sigma$ -álgebra de Borel** de  $\Omega$ . Esse conjunto é a coleção de todos os subconjuntos de  $\Omega$  “gerados” por intervalos, e é a menor  $\sigma$ -álgebra contendo todos os intervalos de  $\Omega$ . Intuitivamente, um conjunto é “gerado” por intervalos se ele pode ser obtido a partir de uma quantidade enumerável de intervalos aplicando-se as operações de união, interseção e complementar.

Chamamos ainda  $A \in \mathcal{B}(\Omega)$  de **boreliano** de  $\Omega$  e denotamos também a  $\sigma$ -álgebra de Borel na reta por  $\mathcal{B}(\mathbb{R}) = \mathcal{B}$ .

**Observação.** Subconjuntos de  $\mathbb{R}$  que não são “gerados” por intervalos são “raríssimos”.

- (c) Subindo uma dimensão, se  $\Omega \subseteq \mathbb{R}^2$  é não-enumerável, tomamos  $\mathcal{A} = \mathcal{B}^2(\Omega)$ , também chamada  **$\sigma$ -álgebra de Borel** de  $\Omega$ . Analogamente ao caso anterior, esse conjunto é a coleção de todos os subconjuntos de  $\Omega$  “gerados” por retângulos, e é a menor  $\sigma$ -álgebra contendo todos os retângulos de  $\Omega$ .

Chamamos ainda  $A \in \mathcal{B}^2(\Omega)$  de **boreliano** de  $\Omega$  e denotamos também a  $\sigma$ -álgebra de Borel no plano por  $\mathcal{B}^2(\mathbb{R}) = \mathcal{B}^2$ .

**Observação.** Definições e notações análogas valem para dimensão maior que 2. Por exemplo,  $\mathcal{B}^n = \mathcal{B}^n(\mathbb{R})$  é a  $\sigma$ -álgebra de Borel no  $\mathbb{R}^n$ .

Agora que já definimos o que é um espaço amostral e uma  $\sigma$ -álgebra, queremos definir um **espaço de probabilidade**, i.e., um espaço no qual possamos, de fato, calcular as probabilidades desejadas. Para tanto, seja  $\mathcal{E}$  um experimento aleatório. Associamos, acima, um par  $(\Omega, \mathcal{A})$  a  $\mathcal{E}$ , sendo  $\Omega$  um espaço amostral e  $\mathcal{A}$  uma  $\sigma$ -álgebra de eventos de  $\Omega$ .

Queremos, dado  $A \in \mathcal{A}$ , associar um número que indique o quão provável é a ocorrência de  $A$  a cada realização de  $\mathcal{E}$ : a **probabilidade de  $A$** .

**Definição 1.5.** Considere  $\Omega$  um espaço amostral e  $\mathcal{A}$  uma  $\sigma$ -álgebra de eventos de  $\Omega$ . Uma **medida de probabilidade** sobre  $(\Omega, \mathcal{A})$  é uma função  $P : \mathcal{A} \rightarrow \mathbb{R}$  que satisfaz os seguintes axiomas:

- (A1)  $P(\Omega) = 1$

(A2)  $\forall A \in \mathcal{A}, P(A) \geq 0$

(A3) se  $A_1, A_2, \dots \in \mathcal{A}$  são tais que  $A_i \cap A_j = \emptyset \forall i \neq j$ , então

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \quad (\sigma\text{-aditividade})$$

Nesse caso,  $P(A)$  é dita **probabilidade de  $A$** .

A partir da definição, temos a seguinte proposição.

**Proposição 1.6.** Se  $P$  é uma medida de probabilidade sobre  $(\Omega, \mathcal{A})$ , então

(i)  $P(\emptyset) = 0$

(ii) se  $A_1, \dots, A_n \in \mathcal{A}$  são tais que  $A_i \cap A_j = \emptyset \forall i \neq j$ , então

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i),$$

ou seja,  $P$  é **finitamente aditiva**.

**Demonstração.**

(i) Tome  $A_1 = \Omega, A_n = \emptyset \forall n \geq 2$ . Da  $\sigma$ -aditividade de  $P$  segue que

$$1 = 1 + \sum_{n=2}^{\infty} P(A_n) \iff P(\emptyset) = 0.$$

(ii) Sejam  $A_1, A_2, \dots, A_n, A_{n+1}, \dots \in \mathcal{A}$  tais que  $A_i \cap A_j = \emptyset \forall i \neq j$  e tome  $A_k = \emptyset \forall k \geq n+1$ . Logo,

$$P\left(\bigcup_{i=1}^n A_i\right) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^n P(A_i),$$

pois  $P(\emptyset) = 0$ .

□

**Definição 1.7.** Um **espaço de probabilidade** é um trio  $(\Omega, \mathcal{A}, P)$ , com  $\Omega$  espaço amostral,  $\mathcal{A}$  uma  $\sigma$ -álgebra de eventos de  $\Omega$  e  $P$  medida de probabilidade sobre  $(\Omega, \mathcal{A})$ .

**Exemplo.** Considere os 6 experimentos  $\mathcal{E}_i$  dados anteriormente. Temos

$\mathcal{E}_1$ : nesse caso, vimos que podíamos tomar  $\Omega = \{1, 2, 3, 4, 5, 6\}$ . Daí, tomando  $\mathcal{A} = \mathcal{P}(\Omega)$ , definimos

$$P(\{\omega\}) = \frac{1}{6}, \forall \omega \in \Omega.$$

Então,  $\forall A \in \mathcal{A}$ , é intuitivo definir (só se  $\Omega$  for finito!) que

$$P(A) = \sum_{\omega: \omega \in A} P(\{\omega\}) = \frac{|A|}{|\Omega|}.$$

Mas será que essa  $P$  é de fato uma medida de probabilidade? Dito de outro modo, se  $0 < |\Omega| < \infty$ , então  $P: \mathcal{A} \rightarrow \mathbb{R}$  tal que, dado  $A \in \mathcal{A}$ ,  $P(A) = |A|/|\Omega|$  é medida de probabilidade? A resposta é sim!

**Demonstração.** Note que  $P(\Omega) = |\Omega|/|\Omega| = 1$  e que,  $\forall A \in \mathcal{A}, P(A) = |A|/|\Omega| \geq 0$  pois  $|A| \geq 0$ . Por fim, se  $A_1, A_2, \dots \in \mathcal{A}$  são tais que  $A_i \cap A_j = \emptyset \forall i \neq j$ , então

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \frac{\left|\bigcup_{i=1}^{\infty} A_i\right|}{|\Omega|} \stackrel{\text{união disjunta}}{=} \frac{1}{|\Omega|} \cdot \sum_{i=1}^{\infty} |A_i| = \sum_{i=1}^{\infty} \frac{|A_i|}{|\Omega|} = \sum_{i=1}^{\infty} P(A_i),$$

logo  $P$  é medida de probabilidade.

□

$\mathcal{E}_4$ : para esse experimento, tomamos  $\Omega = \{1, 2, 3, \dots\}$ . Sendo  $\mathcal{A} = \mathcal{P}(\Omega)$ , é intuitivo tomar  $P: \mathcal{A} \rightarrow \mathbb{R}$  tal que

$$P(\{j\}) = \frac{1}{2^j}, \forall j \in \Omega$$

e

$$P(A) = \sum_{j \in A} P(\{j\}) = \sum_{j \in A} \frac{1}{2^j}.$$

Vamos mostrar que essa  $P$  é uma medida de probabilidade.

**Demonstração.** Note que

$$P(\Omega) = \sum_{j \in \Omega} P(\{j\}) = \sum_{j \in \Omega} \frac{1}{2^j} = \sum_{j=1}^{\infty} \frac{1}{2^j} = \frac{1/2}{1 - 1/2} = 1.$$

Além disso, dado  $A \in \mathcal{A}$  temos

$$P(A) = \sum_{j \in A} \frac{1}{2^j} \geq 0$$

pois  $1/2^j > 0, \forall j \in A$ . Por fim, se  $A_1, A_2, \dots \in \mathcal{A}$  são tais que  $A_i \cap A_k = \emptyset \forall i \neq k$ , então

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{j \in \bigcup A_i} \frac{1}{2^j} = \sum_{i=1}^{\infty} \sum_{j \in A_i} \frac{1}{2^j} = \sum_{i=1}^{\infty} P(A_i),$$

logo  $P$  é medida de probabilidade.  $\square$

$\mathcal{E}_5$ : nesse último caso, tínhamos  $\Omega = [0, 1]$ . Tomando  $\mathcal{A} = \mathcal{B}(\Omega)$ , é intuitivo definir (considerando um espaço de probabilidade uniforme e  $\Omega$  de “comprimento” finito)  $P : \mathcal{A} \rightarrow \mathbb{R}$  tal que

$$P(A) = \frac{\text{“comprimento” de } A}{\text{“comprimento de” } \Omega} = \frac{\int_A dx}{\int_{\Omega} dx}.$$

Vamos mostrar que para  $\Omega \subset \mathbb{R}$  tal que  $0 < \int_{\Omega} dx < \infty$ , a função  $P : \mathcal{A} \rightarrow \mathbb{R}$  definida acima é uma medida de probabilidade.

**Demonstração.** Da monotonicidade da integral, segue que  $P(A) \geq 0, \forall A \in \mathcal{A}$ . Além disso,

$$P(\Omega) = \int_{\Omega} dx / \int_{\Omega} dx = 1.$$

Por último, se  $A_1, A_2, \dots \in \mathcal{A}$  são tais que  $A_i \cap A_k = \emptyset \forall i \neq k$ , então

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \int_{\bigcup A_i} dx / \int_{\Omega} dx \stackrel{\text{união disjunta}}{=} \sum_{i=1}^{\infty} \int_{A_i} dx / \int_{\Omega} dx = \sum_{i=1}^{\infty} P(A_i),$$

logo  $P$  é medida de probabilidade.  $\square$

No contexto de espaços de probabilidade, há dois casos especiais que receberão nossa atenção: os espaços discretos e os espaços contínuos.

**Espaço de probabilidade discreto.** Seja  $(\Omega, \mathcal{A}, P)$  um espaço de probabilidade tal que  $\Omega = \{\omega_1, \omega_2, \dots, \omega_N\}$  ou  $\Omega = \{\omega_1, \omega_2, \dots\}$  ( $\Omega$  enumerável).

**Observação.**  $\Omega$  também poderia ser um espaço amostral qualquer contendo um subconjunto enumerável  $\{\omega_1, \omega_2, \dots\} \subset \Omega$  tal que

$$\sum_{i \geq 1} P(\{\omega_i\}) = 1.$$

Nesse caso, podemos tomar  $\mathcal{A} = \mathcal{P}(\Omega)$  e definir,  $\forall A \in \mathcal{A}$ ,

$$P(A) = \sum_{i: \omega_i \in A} P(\{\omega_i\}).$$

Então, para calcular  $P(A)$ , basta conhecer  $P(\{\omega_i\}) = p_i, i = 1, 2, \dots$ . Note que  $p_i \geq 0 \forall i = 1, 2, \dots$  e  $\sum_{i \geq 1} p_i = 1$ .

No caso particular de  $\Omega$  finito de cardinalidade  $N$  e resultados **equiprováveis**, i.e.,  $p_i = P(\{\omega_i\}) = p \forall i = 1, 2, \dots$ , pode-se mostrar (Lista 1 - Exercício 1) que  $p = 1/N$  e, dado  $A \in \mathcal{A}$ ,

$$P(A) = \sum_{i: \omega_i \in A} p = p|A| = \frac{|A|}{N} = \frac{|A|}{|\Omega|}.$$

Nesse caso,  $(\Omega, \mathcal{A}, P)$  é chamado **espaço de probabilidade uniforme discreto**. A principal dificuldade em espaços desse tipo é determinar as cardinalidades dos conjuntos envolvidos, e para isso serão

necessárias ferramentas e métodos de contagem e análise combinatória, vistos mais adiante.

**Espaço de probabilidade contínuo.** Aqui lidaremos com dois casos principais.

(I) Seja  $\Omega \subset \mathbb{R}$  não-enumerável tal que  $\int_{\Omega} dx > 0$  e tome  $\mathcal{A} = \mathcal{B}(\Omega)$ . Em geral, a probabilidade de um evento  $A \in \mathcal{A}$  é

$$P(A) = \int_A f(x) dx,$$

com  $f : \mathbb{R} \rightarrow \mathbb{R}$  tal que  $f(x) \geq 0 \forall x \in \mathbb{R}$  e  $\int_{\Omega} f(x) dx = 1$ .

Um caso particular é o espaço uniforme: suponha  $0 < \int_{\Omega} dx < \infty$ . Para cada  $A \in \mathcal{A}$ ,

$$P(A) = \frac{\text{“comprimento” de } A}{\text{“comprimento de” } \Omega} = \frac{\int_A dx}{\int_{\Omega} dx} = \int_A f(x) dx,$$

sendo

$$f(x) = \begin{cases} \frac{1}{\int_{\Omega} dx}, & x \in \Omega \\ 0, & x \notin \Omega \end{cases}.$$

**Exemplo.** Seja  $\Omega = [a, b] \subset \mathbb{R}$ . Então o “comprimento” de  $\Omega$  é  $\int_a^b dx = b - a$  e, nesse caso,

$$f(x) = \begin{cases} \frac{1}{b - a}, & x \in [a, b] \\ 0, & x \notin [a, b] \end{cases}.$$

(II) Seja  $\Omega \subset \mathbb{R}^2$  não-enumerável tal que  $\iint_{\Omega} dx dy > 0$  e tome  $\mathcal{A} = \mathcal{B}^2(\Omega)$ . Em geral, a probabilidade de um evento  $A \in \mathcal{A}$  é

$$P(A) = \iint_A f(x) dx,$$

com  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  tal que  $f(x, y) \geq 0 \forall (x, y) \in \mathbb{R}^2$  e  $\iint_{\Omega} f(x, y) dx dy = 1$ .

Um caso particular é o espaço uniforme (em dimensão 2): suponha  $0 < \iint_{\Omega} dx dy < \infty$ . Para cada  $A \in \mathcal{A}$ ,

$$P(A) = \frac{\text{“área” de } A}{\text{“área” de } \Omega} = \frac{\iint_A dx dy}{\iint_{\Omega} dx dy} = \iint_A f(x, y) dx dy,$$

sendo

$$f(x, y) = \begin{cases} \frac{1}{\iint_{\Omega} dx dy}, & (x, y) \in \Omega \\ 0, & (x, y) \notin \Omega \end{cases}.$$

**Exemplo.** Considere o experimento de escolher, ao acaso, um ponto no retângulo  $[-1, 1] \times [-1, 1]$ . Nesse caso,  $\Omega = \{(x, y) \in \mathbb{R}^2 : -1 \leq x \leq 1, -1 \leq y \leq 1\}$ . Então a “área” de  $\Omega$  é  $\int_{-1}^1 \int_{-1}^1 dx dy = 4$  e, nesse caso,

$$f(x, y) = \begin{cases} \frac{1}{4}, & (x, y) \in [-1, 1] \times [-1, 1] \\ 0, & (x, y) \notin [-1, 1] \times [-1, 1] \end{cases}.$$

Por exemplo, se  $A = \{\text{o ponto escolhido dista menos de } 1/2 \text{ da origem}\} = \{(x, y) \in \Omega : x^2 + y^2 < (1/2)^2\}$ , então

$$P(A) = \frac{\text{“área” de } A}{\text{“área” de } \Omega} = \frac{\iint_A dx dy}{\iint_{\Omega} dx dy} = \iint_A f(x, y) dx dy = \iint_A \frac{1}{4} dx dy = \frac{\pi}{4} = \frac{\pi}{16}.$$

Dados os exemplos e feitas as observações, vamos ver algumas propriedades da medida de probabilidade.

**Proposição 1.8.** Sejam  $(\Omega, \mathcal{A}, P)$  um espaço de probabilidade e  $A, B, C \in \mathcal{A}$ . Temos

(P1)  $P(\emptyset) = 0$



(P2)  $P(A) = 1 - P(A^C)$

(P3) se  $A \subset B$ , então

(i)  $P(B - A) = P(B) - P(A)$

(ii)  $P(A) \leq P(B)$  (monotonicidade)

(P4)  $P(A) \leq 1$

(P5)  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$  e, em geral,

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} \cap A_{i_2}) + \sum_{1 \leq i_1 < i_2 < i_3 \leq n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) + \dots + (-1)^{n-1} P(A_1 \cap \dots \cap A_n)$$

(P6)  $P(A \cup B) \leq P(A) + P(B)$  e, em geral,

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i).$$

(P7)  $P(A \cap B) \geq 1 - P(A^C) - P(B^C).$

### Demonstração.

(P1) Provada acima.

(P2) Da Proposição anterior, temos

$$P(\Omega) = 1 = P(A \cup A^C) = P(A) + P(A^C) \iff P(A) = 1 - P(A^C).$$

(P3) (i) Temos

$$P(B - A) = P(B \cap A^C) \stackrel{(P2)}{=} 1 - P(B^C \cup A) = 1 - P(B^C) - P(A) \stackrel{(P2)}{=} P(B) - P(A),$$

pois  $A \subset B$ , logo  $A \cap B^C = \emptyset$ .

(ii) Temos

$$P(B - A) = P(B) - P(A) \stackrel{(A2)}{=} 0 \iff P(B) \geq P(A).$$

(P4) Como  $A \subset \Omega$ , segue da (P3) que  $P(A) \leq P(\Omega) = 1$ .

(P5) Mostramos o resultado para dois conjuntos. Para o caso geral, veja [1] p.50-51. Temos

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

pois  $A \cup B = A \cup (B - A)$  (conjuntos disjuntos). Como  $B = (A \cap B) \cup (B - A)$  (eventos disjuntos), então

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

isto é,

$$P(B - A) = P(B) - P(A \cap B).$$

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

(P6) Mostramos o resultado para dois conjuntos. Para o caso geral, veja [1] p.50-51. Temos

$$P(A \cup B) = P(A) + P(B) - \underbrace{P(A \cap B)}_{\geq 0} \leq P(A) + P(B).$$

(P7) Temos

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 1 - P(A^C) - P(B^C) + \underbrace{1 - P(A \cup B)}_{\geq 0} \geq 1 - P(A^C) - P(B^C).$$

□

Enunciamos, a seguir, uma das propriedades mais importantes de uma medida de probabilidade: sua continuidade.

**Teorema 1.9** (Cont. da Probabilidade). Sejam  $(\Omega, \mathcal{A}, P)$  um espaço de probabilidade e  $A_1, A_2, \dots \in \mathcal{A}$ . Então,

(a) se  $A_1 \subset A_2 \subset \dots$  e

$$\lim_{n \rightarrow \infty} A_n := \bigcup_{n=1}^{\infty} A_n \in \mathcal{A},$$

vale

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\lim_{n \rightarrow \infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n).$$

(b) se  $A_1 \supset A_2 \supset \dots$  e

$$\lim_{n \rightarrow \infty} A_n := \bigcap_{n=1}^{\infty} A_n \in \mathcal{A},$$

vale

$$P\left(\bigcap_{n=1}^{\infty} A_n\right) = P\left(\lim_{n \rightarrow \infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n).$$

### Demonstração.

(a) Defina  $B_n, n \geq 1$ , por

$$B_1 = A_1 \text{ e } B_n = A_n - \bigcup_{i=1}^{n-1} A_i = A_n - A_{n-1}, n > 1.$$

Note que  $B_1 \cap B_k = \emptyset, \forall k \neq 1$ , pois, nesse caso,  $B_1 \cap B_k = B_1 \cap (A_k - A_{k-1})$  e, como  $A_1 \subset A_{k-1}$ , não existe  $b \in B_1$  tal que  $b_1 \in A_k - A_{k-1}$ . De maneira análoga, temos  $B_i \cap B_j = \emptyset, \forall i > j > 1$ . Ademais, temos

$$\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i \text{ e } \bigcup_{i=1}^n B_i = \bigcup_{i=1}^n A_i, \forall n \geq 1.$$

Portanto,

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} A_i\right) &= P\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} P(B_i) = \lim_{n \rightarrow \infty} \sum_{i=1}^n P(B_i) \\ &= \lim_{n \rightarrow \infty} P\left(\bigcup_{i=1}^n B_i\right) \\ &= \lim_{n \rightarrow \infty} P\left(\bigcup_{i=1}^n A_i\right) \\ &= \lim_{n \rightarrow \infty} P(A_n). \end{aligned}$$

(b) Como  $A_1 \supset A_2 \supset \dots$ , então  $A_1^C \subset A_2^C \subset \dots$ . Do item anterior, segue que

$$P\left(\lim_{n \rightarrow \infty} A_i^C\right) = P\left(\bigcup_{i=1}^{\infty} A_i^C\right) = \lim_{n \rightarrow \infty} P(A_n^C).$$

Mas como

$$\bigcap_{i=1}^{\infty} A_i = \left(\bigcup_{i=1}^{\infty} A_i^C\right)^C,$$

então

$$P\left(\left(\bigcap_{i=1}^{\infty} A_i\right)^C\right) = \lim_{n \rightarrow \infty} P(A_n^C),$$

ou seja,

$$1 - P\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n \rightarrow \infty} [1 - P(A_n)] = 1 - \lim_{n \rightarrow \infty} P(A_n),$$

e segue o resultado. □

Vamos agora mostrar que dado  $\Omega$  tal que  $0 < |\Omega| < \infty$  e assumindo resultados equiprováveis, temos que  $P : \mathcal{A} = \mathcal{P}(\Omega) \rightarrow \mathbb{R}$  dada por  $P(A) = |A|/|\Omega|, \forall A \in \mathcal{A}$ , define uma medida de probabilidade.

**Demonstração.** É claro que  $\forall A \in \mathcal{A}$  temos  $P(A) = |A|/|\Omega| > 0$ . Ademais,  $P(\Omega) = |\Omega|/|\Omega| = 1$ . Por fim, note que como  $\mathcal{A} = \mathcal{P}(\Omega)$  e  $|\Omega| < \infty$ , então  $\mathcal{A}$ ; daí, a única forma de termos uma sequência

$A_1, A_2, \dots \in \mathcal{A}$  de elementos dois a dois disjuntos é se  $A_m = \emptyset \forall m \geq N \in \mathbb{N}$ . Assim,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P(\Omega) = 1 = \sum_{i=1}^{N-1} P(A_i) = \sum_{i=1}^{\infty} P(A_i).$$

Outra maneira seria apenas pegar uma tal sequência acima e argumentar que

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \frac{|\bigcup_{i=1}^{\infty} A_i|}{|\Omega|} = \frac{\sum_{i=1}^{\infty} |A_i|}{|\Omega|} = \sum_{i=1}^{\infty} P(A_i).$$

□

Aproveitando o ensejo, mostremos também que dado  $\Omega = \mathbb{N}$  e sendo  $\mathcal{A} = \mathcal{P}(\mathbb{N})$ , temos que  $P : \mathcal{A} \rightarrow \mathbb{R}$  dada por

$$P(\{j\}) = \frac{1}{2^j}, \{j\} \in \Omega \text{ e } P(A) = \sum_{j \in A} P(\{j\}) = \sum_{j \in A} \frac{1}{2^j}, \forall A \in \mathcal{A}$$

define uma medida de probabilidade.

**Demonstração.** Note que dado  $A \in \mathcal{A}$ , temos  $P(A) = \sum_{j \in A} \frac{1}{2^j} \geq 0$ . Ademais,  $P(\Omega) = \sum_{j=1}^{\infty} \frac{1}{2^j} = 1$ .

Por fim, dados  $A_1, A_2, \dots \in \mathcal{A}$  dois a dois disjuntos, temos

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{j \in \bigsqcup A_i} \frac{1}{2^j} = \sum_{j \in A_1} \frac{1}{2^j} + \sum_{j \in A_2} \frac{1}{2^j} + \dots = \sum_{i=1}^{\infty} P(A_i).$$

□

Para finalizar as demonstrações dessas medidas de probabilidade usuais, mostremos que se  $\Omega \subset \mathbb{R}$  é não-enumerável tal que  $0 < \int_{\Omega} dx < \infty$  e sendo  $\mathcal{A} = \mathcal{B}(\Omega)$ , então  $P : \mathcal{A} \rightarrow \mathbb{R}$  dada por

$$P(A) = \frac{\int_A dx}{\int_{\Omega} dx}, \forall A \in \mathcal{A} \quad (1)$$

define uma medida de probabilidade.

**Demonstração.** Segue da monotonicidade da integral que  $P(A) \geq 0, \forall A \in \mathcal{A}$ . Ademais, temos  $P(\Omega) = \int_{\Omega} dx / \int_{\Omega} dx = 1$ . Por último, dados  $A_1, A_2, \dots \in \mathcal{A}$  dois a dois disjuntos, temos

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \int_{\bigsqcup A_i} dx / \int_{\Omega} dx = \sum_i \int_{A_i} dx / \int_{\Omega} dx = \sum_{i=1}^{\infty} P(A_i).$$

□

## 1.2 Condicionamento e independência

### Probabilidade Condicional

Para começarmos a discussão acerca da probabilidade condicional, começamos com um exemplo motivador.

**Exemplo.** Considere uma caixa com  $r$  bolas vermelhas numeradas de 1 a  $r$  e  $b$  bolas brancas numeradas de 1 a  $b$ . Seja  $\mathcal{E}$  o experimento “retira-se ao acaso uma bola da caixa”. Suponhamos resultados equiprováveis e sejam  $\Omega = \{V_1, \dots, V_r, B_1, \dots, B_b\}$  e  $\mathcal{A} = \mathcal{P}(\Omega)$ . Daí,  $\forall \omega \in \Omega$  temos  $P(\omega) = \frac{1}{r+b}$

e,  $\forall A \in \mathcal{A}, P(A) = \frac{|A|}{r+b}$ . Com isso, estamos interessados em saber

- (I) qual a probabilidade de retirar uma bola número 1?
- (II) qual a probabilidade de retirar uma bola número 1, sabendo que ela é vermelha?

**Solução.** Para a primeira pergunta, temos

$$P(B) = \frac{|B|}{r+b} = \frac{2}{r+b},$$

sendo  $B = \{\text{número da bola é 1}\} = \{V_1, B_1\}$ . Para a segunda pergunta, façamos  $A = \{\text{bola é vermelha}\} = \{V_1, \dots, V_r\}$ . Denotamos por  $B|A$  o evento “ocorrer  $B$  dado que  $A$  ocorreu”.

Queremos, então, calcular  $P(B|A)$ . Note que estamos restringindo o espaço amostral às  $r$  bolas vermelhas, i.e.,  $A$  é o “novo” espaço amostral. Nesse “novo espaço”, a probabilidade de retirada de cada bola é  $1/|A| = 1/r$ . Daí,

$$P(B|A) = \frac{1}{r} = \frac{|B \cap A|}{|A|} = \frac{|B \cap A|/|\Omega|}{|A|/|\Omega|} = \frac{P(B \cap A)}{P(A)}.$$

Isso motiva a

**Definição 1.10** (Probabilidade Condicional). Sejam  $(\Omega, \mathcal{A}, P)$  um espaço de probabilidade e  $A, B \in \mathcal{A}$  tais que  $P(A) > 0$ . A **probabilidade condicional de  $B$ , dado  $A$** , denotada por  $P(B|A)$ , é

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

**Observações.** (i) se  $P(A) = 0$ , então  $P(B|A)$  é indefinida. Alguns autores adotam  $P(B|A) = 0$  ou  $P(B|A) = P(B)$  neste caso;

(ii) sejam  $(\Omega, \mathcal{A}, P)$  espaço de probabilidade e  $A \in \mathcal{A}$  fixo com  $P(A) > 0$ . Então  $P(\cdot|A) : \mathcal{A} \rightarrow \mathbb{R}$  define uma medida de probabilidade sobre  $(\Omega, \mathcal{A})$ .

**Demonstração.** Note que  $P(\Omega|A) = \frac{P(\Omega \cap A)}{P(A)} = \frac{P(A)}{P(A)} = 1$  e que  $P(B|A) = \frac{P(B \cap A)}{P(A)} \geq 0, \forall B \in \mathcal{A}$  uma vez que  $P(B \cap A) \geq 0$  e  $P(A) > 0$ . Por fim, dados  $A_1, A_2, \dots \in \mathcal{A}$  disjuntos dois a dois, temos

$$P\left(\bigcup_{i=1}^{\infty} A_i | A\right) = \frac{P(\bigcup_{i=1}^{\infty} A_i \cap A)}{P(A)} = \sum_{i=1}^{\infty} \frac{P(A_i \cap A)}{P(A)} = \sum_{i=1}^{\infty} P(A_i | A),$$

onde a penúltima igualdade se deve ao fato de que  $A_i \cap A_j = \emptyset$  se  $i \neq j$  implica  $A \cap A_i \cap A_j = \emptyset$ .  $\square$

**Proposição 1.11.** Seja  $(\Omega, \mathcal{A}, P)$  espaço de probabilidade. Valem

(a) Regra da Multiplicação

- (i) se  $A, B \in \mathcal{A}$ , então  $P(A \cap B) = P(B|A)P(A) = P(A|B)P(B)$ ;  
(ii) se  $A_1, \dots, A_n \in \mathcal{A}$ , então

$$P\left(\bigcap_{i=1}^n A_i\right) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap \cdots \cap A_{n-1}).$$

(b) Regra da Probabilidade Total: se  $B_1, B_2, \dots \in \mathcal{A}$  são dois a dois disjuntos tais que  $\bigcup_{i=1}^{\infty} B_i = \Omega$  e  $P(B_i) > 0, \forall i$ , então

$$P(A) = \sum_{i=1}^{\infty} P(B_i)P(A|B_i), \forall A \in \mathcal{A}.$$

(c) Fórmula de Bayes: se  $B_1, B_2, \dots \in \mathcal{A}$  são dois a dois disjuntos tais que  $\bigcup_{i=1}^{\infty} B_i = \Omega, P(B_i) > 0, \forall i$  e  $P(A) > 0, A \in \mathcal{A}$ , então

$$P(B_j|A) = \frac{P(B_j)P(A|B_j)}{\sum_{i=1}^{\infty} P(B_i)P(A|B_i)}, \forall j \geq 1.$$

**Demonstração.** Demonstramos os resultados na ordem apresentada.

(i) Por definição, se  $P(A) > 0$  então  $P(A|B) = \frac{P(A \cap B)}{P(B)}$  e  $P(B|A) = \frac{P(B \cap A)}{P(A)}$ , i.e.,  $P(A \cap B) = P(A)P(A|B) = P(B)P(B|A)$ .

(ii) Para  $i = 1, 2$  vale a igualdade. Suponha que

$$P\left(\bigcap_{i=1}^k A_i\right) = P(A_1)P(A_2|A_1) \cdots P(A_k|A_1 \cap \cdots \cap A_{k-1}).$$

De (i), temos que

$$\begin{aligned} P\left(\bigcap_{i=1}^{k+1} A_i\right) &= P\left(A_{k+1} \mid \bigcap_{i=1}^k A_i\right) P\left(\bigcap_{i=1}^k A_i\right) \\ &= P(A_1)P(A_2|A_1) \cdots P(A_k|A_1 \cap \cdots \cap A_{k-1})P(A_{k+1}|A_1 \cap \cdots \cap A_k) \end{aligned}$$

e o resultado segue por indução.

(b) Note que como os  $B_i$ 's são dois a dois disjuntos, então  $(A \cap B_i) \cap (A \cap B_j) = \emptyset$  se  $i \neq j$ . Agora,

como  $A = A \cap \Omega = A \cap \left(\bigcup_{i=1}^{\infty} B_i\right)$ , temos

$$P(A) = P\left(A \cap \bigcup_{i=1}^{\infty} B_i\right) = P\left(\bigcup_{i=1}^{\infty} (A \cap B_i)\right) = \sum_{i=1}^{\infty} P(A \cap B_i) = \sum_{i=1}^{\infty} P(B_i)P(A|B_i), \forall A \in \mathcal{A}.$$

(c) Da definição de probabilidade condicional e de (b), temos

$$P(B_j|A) = \frac{P(B_j \cap A)}{P(A)} = \frac{P(B_j)P(A|B_j)}{\sum_{i=1}^{\infty} P(B_i)P(A|B_i)}, \forall j \geq 1.$$

□

## Independência

Como de praxe, comecemos com um exemplo motivador: seja  $\mathcal{E}$  o experimento “jogar um dado honesto duas vezes e observar os resultados”. Temos, então  $\Omega = \{(i, j) : 1 \leq i, j \leq 6\}$  e definimos  $\mathcal{A} = \mathcal{P}(\Omega)$ . Daí, temos

$$P(\{\omega\}) = P(\{(i, j)\}) = \frac{1}{36}, \forall \omega \in \Omega \quad \text{e} \quad P(A) = \frac{|A|}{36}, \forall A \in \mathcal{A}.$$

Sejam  $A = \{\text{soma dos resultados é } 5\} = \{(1, 4), (2, 3), (3, 2), (4, 1)\}$ ,  $B = \{1^\circ \text{ resultado é par}\} = \{(2, i), (4, i), (6, i) : i = 1, \dots, 6\}$ ,  $C = \{\text{pelo menos um dos resultados é maior que } 3\} = \Omega \setminus \{(i, j) : 1 \leq i, j \leq 3\}$  e  $D = \{\text{pelo menos um dos resultados é } 2\} = \{(2, i) : 1 \leq i \leq 6\} \cup \{(j, 2) : 1 \leq j \leq 6, j \neq 2\}$ . Temos

$$P(A) = \frac{4}{36} = \frac{1}{9}, P(B) = \frac{18}{36} = \frac{1}{2}, P(C) = \frac{27}{36} = \frac{3}{4}, P(D) = \frac{11}{36}.$$

Ademais, temos  $A \cap C = \{(1, 4), (4, 1)\}$ ,  $A \cap D = \{(2, 3), (3, 2)\}$ ,  $A \cap B = \{(2, 3), (4, 1)\}$ . Daí,

$$P(A|C) = \frac{2/36}{27/36} = \frac{2}{27} < \frac{1}{9} = P(A),$$

$$P(A|D) = \frac{2/36}{11/36} = \frac{2}{11} > \frac{1}{9} = P(A),$$

$$P(A|B) = \frac{2/36}{18/36} = \frac{2}{18} = \frac{1}{9} = P(A).$$

Note que o conhecimento da ocorrência de  $C$  **diminui** a chance de  $A$  ocorrer e o conhecimento da ocorrência de  $D$  **aumenta** a chance de  $A$  ocorrer. Por outro lado, o conhecimento da existência de  $B$  não altera a chance de ocorrência de  $A$ . Nesse caso, dizemos que  $A$  e  $B$  são **independentes**, conforme a definição abaixo.

**Definição 1.12** (Independência). Dois eventos  $A$  e  $B$  num espaço de probabilidade  $(\Omega, \mathcal{A}, P)$  são **independentes** se  $P(A \cap B) = P(A)P(B)$ .

Alguns comentários são necessários: primeiro, note que se  $P(A) = 0$ , então  $A$  é independente de qualquer outro evento  $B$ . Além disso, se  $P(A), P(B) > 0$ , então  $A$  e  $B$  são independentes se, e só se,  $P(A|B) = P(A)$  (ou  $P(B|A) = P(B)$ ).

**Demonstração.** Se  $A$  e  $B$  são independentes, então  $P(A \cap B) = P(A)P(B) = P(A)P(A|B) = P(B)P(B|A)$  e, portanto,  $P(B) = P(B|A)$  e  $P(A) = P(A|B)$ . Reciprocamente, se  $P(A) = P(A|B)(P(B) = P(B|A))$ , então  $P(A \cap B) = P(A)P(B)$ . □

**Observação.** Se  $A$  e  $B$  são independentes, então

- (i)  $A$  e  $B^C$  são independentes
- (ii)  $A^C$  e  $B$  são independentes
- (iii)  $A^C$  e  $B^C$  são independentes

**Demonstração.** Se  $P(A) = 0$  ou  $P(B) = 0$ , (i), (ii) e (iii) são imediatas. Suponha  $P(A), P(B) > 0$ .

- (i) Note que como  $A \cap B^C$  e  $A \cap B$  são disjuntos, então  $P(A \cap B^C) = P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)P(B^C)$ .
- (ii) Analogamente a acima,  $P(B \cap A^C) = P(B) - P(B \cap A) = P(B) - P(B)P(A) = P(B)P(A^C)$ .
- (iii) Note que  $P(A^C \cap B^C) = 1 - P(A \cup B) = 1 - P(A) - P(B) + P(A \cap B) = (1 - P(A))(1 - P(B)) = P(A^C)P(B^C)$ .

□

Nada mais natural do que generalizar a definição de independência da seguinte forma.

**Definição 1.13** (Independência 2 a 2 e Coletiva). Sejam  $(\Omega, \mathcal{A}, P)$  um espaço de probabilidade e  $A_1, \dots, A_n \in \mathcal{A}$ . Dizemos que

- (a)  $A_1, \dots, A_n$  são **2 a 2 independentes** ou **independentes aos pares** se  $P(A_i \cap A_j) = P(A_i)P(A_j), \forall 1 \leq i, j \leq n, i \neq j$ .
- (b)  $A_1, \dots, A_n$  são **independentes** ou **mutuamente independentes** ou **coletivamente independentes** se  $P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_m}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_m}), \forall 1 \leq i_1, i_2, \dots, i_m \leq n$  e  $\forall 2 \leq m \leq n$ .

Note que a independência coletiva é um pouco mais forte que a independência aos pares. Ao longo do texto, a não ser que seja dito o contrário, quando dissermos que vários eventos são independentes estamos nos referindo à independência coletiva.

**Exemplo** (Ensaio ou Provas de Bernoulli). Consideremos um experimento qualquer (por exemplo, lançar uma moeda) que possua duas propriedades:

- (i) há somente 2 resultados possíveis: **sucesso**, com probabilidade  $p$ , e **fracasso**, com probabilidade  $1 - p$ .
- (ii) ao repetirmos o experimento, os resultados são **independentes**.

Dado um experimento com as duas propriedades acima, os **ensaios** ou **provas de Bernoulli** consistem de  $n \geq 1$  ensaios, isto é, repetições, deste experimento, resultando em um experimento **composto** (por exemplo, lançar 3 vezes uma moeda). Nessas condições, vejamos como resolver os seguintes itens.

- (a) Construir um espaço de probabilidade adequado para o experimento composto
- (b) Calcular a probabilidade de que ocorram  $k \in \{0, \dots, n\}$  sucessos nas  $n$  repetições
- (c) Calcular a probabilidade de pelo menos um sucesso nas  $n$  repetições
- (d) Repetindo o experimento indefinidamente ( $n \rightarrow \infty$ ), calcular a probabilidade de todas as provas terem sucesso.

**Solução.**

- (a) Para  $n$  ensaios temos  $2^n$  resultados possíveis ao todo. Logo, definimos

$$\Omega = \{\tilde{\omega} = (\omega_1, \omega_2, \dots, \omega_n) : \omega_i \in \{0, 1\}, i = 1, \dots, n\},$$

sendo

$$\omega_i = \begin{cases} 1, & \text{se ocorre sucesso na } i\text{-ésima repetição} \\ 0, & \text{caso contrário.} \end{cases},$$

$\mathcal{A} = \mathcal{P}(\Omega)$  e, como  $\Omega$  é finito, para definirmos  $P$  basta definirmos  $P(\{\tilde{\omega}\}), \forall \tilde{\omega} \in \Omega$ . Seja  $\tilde{\omega} = (\underbrace{1, 1, \dots, 1}_k, \underbrace{0, 0, \dots, 0}_{n-k})$  (sucessos nas  $k$  primeiras repetições). Para cada  $i \in \{1, \dots, n\}$ ,

defina

$$A_i = \{\text{ocorrer sucesso na } i\text{-ésima repetição}\}.$$

Por hipótese, temos

$$P(A_i) = p, \forall i = 1, \dots, n \text{ e } A_1, \dots, A_n \text{ são independentes.}$$

Logo,

$$\begin{aligned} P(\{\tilde{\omega}\}) &= P(\{\underbrace{(1, 1, \dots, 1)}_k, \underbrace{(0, 0, \dots, 0)}_{n-k}\}) \\ &= P(A_1 \cap \dots \cap A_k \cap A_{k+1}^C \cap \dots \cap A_n^C) \\ &= P(A_1)P(A_2) \dots P(A_k)P(A_{k+1}^C) \dots P(A_n^C) \\ &= p^k(1-p)^{n-k}. \end{aligned}$$

Temos então, denotando por  $\tilde{\omega}^{(k)}$  um elemento que corresponde a  $k$  sucessos em qualquer ordem, que  $P(\{\tilde{\omega}\}) = p^k(1-p)^{n-k}$ . Está definido nosso espaço de probabilidade  $(\Omega, \mathcal{A}, P)$ .

(b) Seja  $B_k = \{\text{ocorrer exatamente } k \text{ sucessos nas } n \text{ repetições}\}$ . Note que  $|B_k| = \binom{n}{k}$ , de modo que

$$P(B_k) = \sum_{\tilde{\omega} \in B_k} P(\{\tilde{\omega}\}) = |B_k| \cdot p^k(1-p)^{n-k} = \binom{n}{k} p^k(1-p)^{n-k}, \quad k = 0, 1, \dots, n.$$

(c) Seja  $A = \{\text{pelo menos 1 sucesso}\}$ . Note que

$$A = \bigcup_{i=1}^n A_i \text{ e } A^C = \{\text{nenhum sucesso nas } n \text{ repetições}\} = \bigcap_{i=1}^n A_i^C,$$

sendo  $A_i = \{\text{sucesso na } i\text{-ésima repetição}\}$ . Temos, observando que os  $A_i$  são independentes, que

$$P(A^C) = P\left(\bigcap_{i=1}^n A_i^C\right) = \prod_{i=1}^n P(A_i^C) = \prod_{i=1}^n (1 - P(A_i)) = \prod_{i=1}^n (1-p) = (1-p)^n,$$

logo  $P(A) = 1 - P(A^C) = 1 - (1-p)^n$ .

(d) Seja  $C = \{\text{sucesso nas infinitas repetições}\}$ . Podemos escrever

$$C = \bigcap_{i=1}^{\infty} A_i = A_1 \cap (A_1 \cap A_2) \cap (A_1 \cap A_2 \cap A_3) \cap \dots = \bigcap_{n=1}^{\infty} C_n,$$

com

$$C_n = \bigcap_{i=1}^n A_i = \{\text{sucesso em todas as } n \text{ primeiras provas}\}.$$

Como  $C_n \supset C_{n+1} \forall n \geq 1$ , então

$$C = \bigcap_{n=1}^{\infty} C_n = \lim_{n \rightarrow \infty} C_n.$$

Pela continuidade da probabilidade e independência dos  $A_i$ , temos

$$P(C) = P\left(\lim_{n \rightarrow \infty} C_n\right) = \lim_{n \rightarrow \infty} P(C_n) = \lim_{n \rightarrow \infty} P\left(\bigcap_{i=1}^n A_i\right) = \lim_{n \rightarrow \infty} p^n = \begin{cases} 0, & 0 \leq p < 1, \\ 1, & p = 1 \end{cases}.$$

□

### 1.3 Exercícios

1. Seja  $(\Omega, \mathcal{A}, P)$  um espaço de probabilidade, onde  $\mathcal{A}$  é a  $\sigma$ -álgebra de todos os subconjuntos de  $\Omega$  e  $P$  é uma medida de probabilidade que associa a probabilidade  $p > 0$  a cada conjunto de um ponto de  $\Omega$ .

- Mostre que  $\Omega$  deve ter um número finito de pontos [*Sugestão*: mostre que  $\Omega$  não pode ter mais de  $p^{-1}$  pontos.]
- Mostre que se  $n$  é o número de pontos em  $\Omega$ , então  $p$  deve ser igual a  $n^{-1}$ .

**Solução.** Seja  $A_i$  o evento “escolher o ponto  $w_i$ ”.

a) Temos  $A_i \cap A_j = \emptyset$  para  $i \neq j$ . Logo, segue da  $\sigma$ -aditividade que

$$P\left(\bigcup_{i=1}^N A_i\right) = \sum_{i=1}^N P(A_i) = Np \leq 1 \Leftrightarrow N \leq 1/p.$$

Portanto,  $|\Omega| \leq 1/p < +\infty$ .

b) Temos que

$$1 = P(\Omega) = P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) = np \Leftrightarrow p = 1/n.$$

□

2. Pode-se construir um modelo para um *spinner* aleatório tomando um espaço uniforme de probabilidade sobre a circunferência de um círculo de raio 1, de modo que a probabilidade de que o ponteiro do *spinner* pare sobre um arco de comprimento  $s$  é  $s/2\pi$ . Suponha que o círculo esteja dividido em 37 zonas numeradas de 1 a 37. Determine a probabilidade de que o ponteiro pare sobre uma zona de número par.

**Solução.** Cada zona tem comprimento  $2\pi/37$ , uma vez que o espaço de probabilidade é uniforme. Seja  $Z_i$  o evento “parar sobre a zona de número  $i$ ”. Temos

$$P\left(\bigcup_{i \text{ par}} Z_i\right) = \sum_{i \text{ par}} P(Z_i) = 13 \cdot \frac{2\pi/37}{2\pi} = 13/37,$$

já que  $Z_i \cap Z_j = \emptyset$  para  $i \neq j$ .

□

3. Considere um ponto escolhido ao acaso sobre um quadrado unitário. Determine a probabilidade de que o ponto esteja no triângulo limitado por  $x = 0$ ,  $y = 0$  e  $x + y = 1$ .

**Solução.** Sejam  $\Omega = [0, 1] \times [0, 1]$  e  $\mathcal{A} = \mathcal{B}^2(\Omega)$ . Defina uma medidade de probabilidade  $P$  sobre  $(\Omega, \mathcal{A})$  por

$$P(A) = \iint_A f(x, y) dx dy$$

sendo

$$f(x, y) = \begin{cases} 1/\iint_{\Omega} dx dy, & (x, y) \in \Omega \\ 0, & (x, y) \notin \Omega \end{cases}, \forall A \in \mathcal{A}.$$

Assim, sendo  $A$  o evento “escolher um ponto no triângulo limitado por  $x = 0$ ,  $y = 0$  e  $x + y = 1$ ”, temos

$$P(A) = \iint_A \frac{1}{\iint_{\Omega} dx dy} dx dy = \iint_A 1 dx dy = 1/2.$$

□

4. Seja  $P$  um ponto escolhido ao acaso sobre um círculo unitário. Determine a probabilidade de que  $P$  esteja no setor angular de 0 a  $\pi/4$  radianos.

**Solução.** Seja  $(\Omega, \mathcal{A}, P)$  o espaço de probabilidade definido acima, mas com  $\Omega = \overline{D}(0, 1)$ . Seja  $A$  o evento “ $P$  no setor circular de 0 a  $\pi/4$ ”. Temos

$$P(A) = \iint_A \frac{1}{\iint_{\Omega} dx dy} dx dy = \int_0^{\pi/4} \int_0^1 \frac{r}{\pi} dr d\theta = \frac{1}{\pi} \cdot \frac{1}{2} \cdot \frac{\pi}{4} = 1/8.$$

□

5. Uma caixa contém 10 bolas numeradas de 1 a 10. Extraí-se ao acaso uma bola da caixa. Determine a probabilidade de que o número da bola seja 3, 4 ou 5.

**Solução.** Sejam  $\Omega = \{1, 2, \dots, 10\}$ ,  $\mathcal{P}(\Omega) = \mathcal{A}$  e  $P$  a medida de probabilidade sobre  $(\Omega, \mathcal{A})$  definida por  $P(A) = |A|/|\Omega|$ ,  $\forall A \in \mathcal{A}$ . Defina  $A$  como o evento “o número da bola é 3, 4 ou 5”. Temos

$$P(A) = |A|/|\Omega| = 3/10.$$

□

6. Suponha que se lance um par de dados e que os 36 resultados possíveis são igualmente prováveis.



Determine a probabilidade de que a soma dos números observados seja par.

**Solução.** Sejam  $\Omega = \{(i, j), 1 \leq i, j \leq 6\}$ ,  $\mathcal{A} = \mathcal{P}(\Omega)$  e  $P$  a medida de probabilidade sobre  $(\Omega, \mathcal{A})$  definida por  $P(A) = |A|/|\Omega|$ ,  $\forall A \in \mathcal{A}$ . Defina  $A$  como sendo o evento “a soma dos números é par”. Temos que  $|A| = 18$ , logo  $P(A) = 18/36 = 1/2$ .  $\square$

7. Suponha que  $A$  e  $B$  sejam eventos tais que  $P(A) = 2/5$ ,  $P(B) = 2/5$  e  $P(A \cup B) = 1/2$ . Determine  $P(A \cap B)$ .

**Solução.** Como  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ , temos que  $P(A \cap B) = 2/5 + 2/5 - 1/2 = 3/10$ .  $\square$

8. Se  $P(A) = 1/3$ ,  $P(A \cup B) = 1/2$  e  $P(A \cap B) = 1/4$ , determine  $P(B)$ .

**Solução.** Como no item anterior, temos  $P(B) = 1/2 - 1/3 + 1/4 = 5/12$ .  $\square$

9. Suponha que se escolha ao acaso um ponto sobre um quadrado unitário. Seja  $A$  o evento de que o ponto está no triângulo limitado por  $y = 0$ ,  $x = 1$  e  $x = y$ , e  $B$  o evento de que o ponto está no retângulo com vértices em  $(0, 0)$ ,  $(1, 0)$ ,  $(1, 1/2)$  e  $(0, 1/2)$ . Determine  $P(A \cup B)$  e  $P(A \cap B)$ .

**Solução.** Sejam  $\Omega = [0, 1] \times [0, 1]$ ,  $\mathcal{A} = \mathcal{B}^2(\Omega)$  e  $P$  a medida de probabilidade definida nos exercícios 3 e 4. Sendo  $A$  e  $B$  os eventos do enunciado, temos

$$P(A \cap B) = \iint_{A \cap B} \frac{1}{\iint_{\Omega} dx dy} dx dy = \iint_{A \cap B} dx dy = 1/8,$$

e também  $P(A) = 1/2 = P(B)$ . Logo,  $P(A \cup B) = 1 - 1/8 = 7/8$ .  $\square$

10. Suponha que temos quatro cofres, cada um com 2 gavetas. Os cofres 1 e 2, têm uma moeda de ouro em uma gaveta e uma de prata na outra. O cofre 3, tem duas moedas de ouro e, o cofre 4 tem duas de prata. Escolhe-se um cofre ao acaso, abre-se uma gaveta e encontra-se uma moeda de ouro. Determine a probabilidade de que a outra gaveta contenha:

- Uma moeda de prata;
- Uma moeda de ouro.

**Solução.**

- (a) Sejam  $Au_i$  e  $Ag_i$  os eventos “tirar uma moeda de ouro do cofre  $i$ ” e “tirar uma moeda de prata do cofre  $i$ ”, respectivamente. Queremos calcular  $P(Ag|Au)$ , sendo  $Ag$  e  $Au$  os eventos “tirar moeda de prata” e “tirar moeda de ouro”, respectivamente. Note que

$$P(Ag|Au) = \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2} = 2/3.$$

- (b) Analogamente, a probabilidade desejada é  $1/3 = (1/3)(1/2) + (1/3)(1/2)$ .  $\square$

11. Uma caixa contém 10 bolas das quais 6 são pretas e 4 são brancas. Remove-se três bolas sem observar suas cores. Determine a probabilidade de que uma quarta bola removida da caixa seja branca. Assuma que as 10 bolas são igualmente prováveis de serem removidas da caixa.

**Solução.** A probabilidade da quarta bola ser branca é

$$\frac{4 \cdot 3 \cdot 2 \cdot 1}{10 \cdot 9 \cdot 8 \cdot 7} + \frac{4 \cdot 3 \cdot 6 \cdot 2}{10 \cdot 9 \cdot 8 \cdot 7} + \frac{4 \cdot 6 \cdot 5 \cdot 3}{10 \cdot 9 \cdot 8 \cdot 7} + \frac{6 \cdot 5 \cdot 4 \cdot 4}{10 \cdot 9 \cdot 8 \cdot 7} = 1/5.$$

$\square$

12. Para uma caixa de mesma composição que a do exercício 11, determine a probabilidade de que todas as 3 primeiras bolas removidas sejam pretas, sabendo-se que pelo menos uma delas é preta.

**Solução.** Sejam  $PPP$  e  $P+$  os eventos “três pretas” e “pelo menos uma preta”, respectivamente. Seja ainda  $BBB$  o evento “três brancas”. Temos

$$P(PPP|P+) = \frac{P(PPP \cap P+)}{P(P+)} = \frac{P(PPP)}{1 - P(BBB)}.$$

Note que

$$P(PPP) = \frac{6 \cdot 5 \cdot 4}{10 \cdot 9 \cdot 8} = 1/6$$

e

$$1 - P(BBB) = 1 - \frac{4 \cdot 3 \cdot 2}{10 \cdot 9 \cdot 8} = 29/30.$$

Logo,  $P(PPP|P+) = 5/29$ . □

13. Suponha que uma fábrica tem duas máquinas  $A$  e  $B$ , responsáveis, respectivamente, por 60% e 40% da produção total. A máquina  $A$  produz 3% de itens defeituosos, enquanto a máquina  $B$  produz 5% de itens defeituosos. Determine a probabilidade de que um dado item defeituoso foi produzido pela máquina  $B$ .

**Solução.** Sejam  $D$ ,  $A$  e  $B$  os eventos “dar defeito”, “ser produzido por  $A$ ” e “ser produzido por  $B$ ”, respectivamente. A probabilidade pedida é

$$P(B|D) = \frac{P(D|B)P(B)}{P(D|B)P(B) + P(D|A)P(A)} = \frac{0,05 \cdot 0,4}{0,05 \cdot 0,4 + 0,03 \cdot 0,6} = 10/19. \quad \square$$

14. Um estudante se submete a um exame de múltipla escolha no qual cada questão tem 5 respostas possíveis das quais exatamente uma é correta. O estudante seleciona a resposta correta se ele sabe a resposta. Caso contrário, ele seleciona ao acaso uma resposta entre as 5 possíveis. Suponha que o estudante saiba a resposta de 70% das questões.

- Qual a probabilidade de que o estudante escolha a resposta correta para uma dada questão?
- Se o estudante escolhe a resposta correta para uma dada questão, qual a probabilidade de que ele sabia a resposta?

**Solução.** Sejam  $A$  e  $C$  os eventos “acertar a questão” e “chutar a resposta”, respectivamente. Assim, temos

a)  $P(A) = P(A|C)P(C) + P(A|C^c)P(C^c) = (1/5)(3/10) + 7/10 = 19/25$ .

b)  $P(C^c|A) = P(A|C^c)P(C^c)/P(A) = (7/10)(25/19) = 35/38$ . □

15. Suponha que se escolha ao acaso um ponto sobre um quadrado unitário. Sabendo-se que o ponto está no retângulo limitado por  $y = 0$ ,  $y = 1$  e  $x = 0$  e  $x = 1/2$ , qual é a probabilidade de que o ponto esteja no triângulo limitado por  $y = 0$ ,  $x = 1/2$  e  $x + y = 1$ ?

**Solução.** Seja  $(\Omega, \mathcal{A}, P)$  o espaço de probabilidade definido no exercício 9. Temos que a probabilidade  $p$  procurada é igual a 0, haja vista que o triângulo e o retângulo do enunciado não têm pontos em comum. □

16. Suponha que uma caixa contenha  $r$  bolas vermelhas e  $b$  bolas pretas. Extraí-se ao acaso uma bola da caixa e a seguir extraí-se, também ao acaso, uma segunda bola dentre as que ficaram na caixa. Determine a probabilidade de que:

- Ambas as bolas sejam vermelhas;
- A primeira bola seja vermelha e a segunda preta;
- A primeira bola seja preta e a segunda vermelha;
- Ambas as bolas sejam pretas.

**Solução.**

a)  $\frac{r(r-1)}{(r+b)(r+b-1)}$ .

b)  $\frac{rb}{(r+b)(r+b-1)}$ .

c)  $\frac{rb}{(r+b)(r+b-1)}$ .

$$d) \frac{b(b-1)}{(r+b)(r+b-1)}.$$

□

17. Uma caixa contém 10 bolas vermelhas e 5 pretas. Extraí-se uma bola da caixa. Se ela é vermelha, ela é recolocada na caixa. Se é preta, além de recolocá-la na caixa, adiciona-se duas bolas pretas à caixa. Determine a probabilidade de que uma segunda bola extraída da caixa seja
- vermelha;
  - preta.

**Solução.** Sejam  $V_i$  e  $P_i$  os eventos “vermelha na  $i$ -ésima retirada” e “preta na  $i$ -ésima retirada”. Temos

$$a) P(V_2) = P(V_2|V_1)P(V_1) + P(V_2|P_1)P(P_1) = (10/15)(10/15) + (10/17)(5/15) = 98/153.$$

$$b) P(P_2) = P(P_2|V_1)P(V_1) + P(P_2|P_1)P(P_1) = (5/15)(10/15) + (7/17)(5/15) = 55/153.$$

□

18. Extraí-se duas bolas, com reposição da primeira, de uma caixa contendo 3 bolas brancas e 2 bolas pretas.
- Construa um espaço amostral com pontos igualmente prováveis para este experimento
  - Determine a probabilidade de que as bolas extraídas sejam da mesma cor.
  - Determine a probabilidade de que, pelo menos, uma das bolas extraídas seja branca

**Solução.**

$$a) \Omega = \{(i, j), i, j = B, P\}$$

b) Sejam  $A$  e  $B$  os eventos “bolas da mesma cor” e “pelo menos uma bola branca”, respectivamente. Temos que

$$P(A) = \frac{3}{5} \cdot \frac{3}{5} + \frac{2}{5} \cdot \frac{2}{5} = 13/25.$$

$$c) P(B) = 1 - (2/5)(2/5) = 21/25.$$

□

19. Resolva o exercício 18, caso a primeira bola não seja repostada na caixa.

**Solução.**  $P(A) = (3/5)(2/4) + (2/5)(1/4) = 2/5$  e  $P(B) = 1 - (2/5)(1/4) = 9/10$ .

□

20. Uma caixa contém 3 bolas brancas e 2 bolas pretas. Seleccionam-se duas bolas sem reposição.
- Calcule a probabilidade de que a segunda bola seja preta, dado que a primeira é preta.
  - Calcule a probabilidade de que a segunda bola seja da mesma cor da primeira.
  - Calcule a probabilidade de que a primeira bola seja branca, dado que a segunda é branca.

**Solução.** Sejam  $B_i$  e  $P_i$  os eventos “branca na  $i$ -ésima retirada” e “preta na  $i$ -ésima retirada”, respectivamente.

$$a) P(P_2|P_1) = P(P_2 \cap P_1)/P(P_1) = 1/4.$$

$$b) P(B_1 \cap B_2) + P(P_1 \cap P_2) = (3/5)(1/2) + (2/5)(1/4) = 2/5.$$

$$c) P(B_1|B_2) = \frac{P(B_2|B_1)P(B_1)}{P(B_2|B_1)P(B_1) + P(B_2|P_1)P(P_1)} = 1/2.$$

□

21. Suponha que existisse um teste para câncer com probabilidade de que 90% das pessoas com câncer e 5% das pessoas sem câncer reagem positivamente. Admita que 1% dos pacientes de um hospital têm câncer. Qual a probabilidade de que um paciente escolhido ao acaso, que reage positivamente a esse teste, realmente tenha câncer?

**Solução.** Sejam  $D$  e  $P$  os eventos “ter câncer” e “testar positivo”, respectivamente.

$$a) P(D|P) = \frac{P(P|D)P(D)}{P(P|D)P(D) + P(P|D^c)P(D^c)} = \frac{0,9 \cdot 0,01}{0,9 \cdot 0,01 + 0,05 \cdot 0,99} = 2/13.$$

□

22. Suponha que se lança três moedas idênticas e perfeitamente equilibradas. Seja  $A_i$  o evento de observar cara na  $i$ -ésima moeda. Mostre que os eventos  $A_1$ ,  $A_2$  e  $A_3$  são mutuamente independentes.

**Solução.** Basta notar que

$$\begin{aligned} P(A_1 \cap A_2) &= 1/4 = P(A_1)P(A_2) \\ P(A_1 \cap A_3) &= 1/4 = P(A_1)P(A_3) \\ P(A_2 \cap A_3) &= 1/4 = P(A_2)P(A_3) \\ P(A_1 \cap A_2 \cap A_3) &= 1/8 = P(A_1)P(A_2)P(A_3). \end{aligned}$$

□

23. Suponha que as seis faces de um dado têm igual probabilidade de ocorrência e que sucessivos lançamentos do dado são independentes. Construa um espaço de probabilidade para o experimento composto de três lançamentos do dado.

**Solução.** Defina

$$\Omega = \{(i, j, k) : 1 \leq i, j, k \leq 6\}, \mathcal{A} = \mathcal{P}(\Omega)$$

e  $P : \mathcal{A} \rightarrow \mathbb{R}$  por  $P(A) = |A|/|\Omega|$ .

□

24. Suponha que  $A$ ,  $B$  e  $C$  são eventos mutuamente independentes e  $P(A \cap B) \neq 0$ . Mostre que  $P(C|A \cap B) = P(C)$ .

**Solução.** Temos

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

□

25. Experiência mostra que 20% das pessoas que fazem reservas de mesa num certo restaurante deixam de comparecer. Se o restaurante tem 50 mesas e aceita 52 reservas, qual a probabilidade de que seja capaz de acomodar todos os fregueses?

**Solução.** Temos

$$P(\text{acomodar}) = 1 - P(\text{só 1 não vai}) - P(\text{todos vão}) = 1 - \frac{1}{5} \cdot 52 \cdot \left(\frac{4}{5}\right)^{51} - \frac{4}{5} \left(\frac{4}{5}\right)^{51} = 1 - \left(\frac{4}{5}\right)^{51} \cdot \frac{56}{5}.$$

□

26. Um certo componente de um motor de foguete falha 5% das vezes quando o motor é acionado. Para obter maior confiabilidade no funcionamento do motor, usa-se  $n$  componentes em paralelo, de maneira que o motor falha somente se todos os componentes falharem. Suponha que as falhas dos componentes sejam independentes uma das outras. Qual é o menor valor de  $n$  que pode ser usado para garantir que o motor funcione 99% das vezes?

**Solução.** Queremos  $n$  tal que

$$1 - (0,05)^n \geq 0,99 \Leftrightarrow n(1 + \log 2) \geq 2 \Leftrightarrow n \geq 1,53.$$

Portanto,  $n = 2$ .

□

27. Existem 4 reis num baralho de 52 cartas. Extrai-se uma carta do baralho, registra-se o seu valor e a seguir repõe-se a carta extraída. Este procedimento é repetido 4 vezes. Determine a probabilidade de que existam exatamente 2 reis entre as cartas selecionadas, sabendo-se que entre elas existe pelo menos um rei.

**Solução.** Sejam  $RR$  e  $R+$  os eventos “exatamente dois reis” e “pelo menos um rei”. Temos

$$P(RR|R+) = \frac{P(RR \cap R+)}{P(R+)} = \frac{P(RR)}{1 - P(R^c)} = \frac{16 \cdot 48^2 \cdot 6/52^4}{52^4 - 48^4/52^4} = \frac{16 \cdot 48 \cdot 48 \cdot 6}{4^4} \cdot \frac{1}{13^4 - 12^4} = \frac{6 \cdot 12^2}{13^4 - 12^4}.$$

□

28. Mostre que se  $A$ ,  $B$  e  $C$  são eventos tais que  $P(A \cap B \cap C) \neq 0$  e  $P(C|A \cap B) = P(C|B)$ , então  $P(A|B \cap C) = P(A|B)$ .

**Solução.** Temos

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

Daí, segue que

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

□

29. Um homem dispara 12 tiros independentes num alvo. Qual a probabilidade de que ele atinja o alvo pelo menos uma vez, se tem probabilidade 9/10 de atingir o alvo em qualquer tiro?

**Solução.** A probabilidade desejada é  $1 - (1/10)^{12}$ .

□

30. No Exercício 29 qual a probabilidade de que o alvo seja atingido pelo menos duas vezes, sabendo-se que o mesmo foi atingido pelo menos uma vez?

**Solução.** Sejam  $A$ ,  $B$ ,  $C$  e  $D$  os eventos “acertar pelo menos uma vez”, “acertar pelo menos 2 vezes”, “acertar apenas uma vez” e “não acertar nada”, respectivamente. Temos

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

Ademais,

$$P(B) = 1 - P(C) - P(D) = 1 - 12 \frac{9}{10^{12}} - \frac{1}{10^{12}} = 1 - \frac{109}{10^{12}}$$

e, portanto,

$$P(B|A) = \frac{1 - \frac{109}{10^{12}}}{1 - \frac{1}{10^{12}}}.$$

□

## 2 Noções de Análise Combinatória

Dado um conjunto finito  $S$ , apresentaremos algumas maneiras de contar seus elementos, isto é, obter  $|S|$ . Desejamos fazer isto para calcular as probabilidades de eventos em espaços de probabilidade uniformes com espaço amostral finito já que, como vimos anteriormente, a probabilidade de um evento  $A$  neste espaço nada mais é que  $|A|/|\Omega|$ .

### 2.1 Amostras ordenadas

Sejam  $S = \{s_1, \dots, s_n\}$  e  $T = \{t_1, \dots, t_n\}$ . Temos  $|S| = m$  e  $|T| = n$ , de modo que  $\Omega = S \times T = \{(s, t) : s \in S, t \in T\}$  é tal que  $|\Omega| = mn$ . De maneira geral, se  $S_1, \dots, S_n$  são tais que  $|S_i| = m_i, 1 \leq i \leq n$ , então  $\Omega = S_1 \times S_2 \times \dots \times S_n = \{(s_1, \dots, s_n) : s_i \in S_i, 1 \leq i \leq n\}$  é tal que  $|\Omega| = m_1 \cdots m_n$ . Note que se  $m = m_1 = \dots = m_n$ , então  $|\Omega| = m^n$ .

### Amostragem com reposição

Considere uma caixa com  $m$  bolas numeradas de 1 a  $m$  em que retiramos uma bola, registramos seu número e a repomos na caixa. Repetimos esse procedimento  $n$  vezes. Temos, então

$$\Omega = \{(x_1, \dots, x_n) : 1 \leq x_i \leq m, 1 \leq i \leq n\},$$

sendo  $x_i$  o número da  $i$ -ésima bola retirada, de modo que

$$|\Omega| = m^n \text{ (total de amostras de tamanho } n \text{ com reposição).}$$

**Exemplo.** Seja  $\mathcal{E}$  o experimento que consiste em lançar  $n$  vezes uma moeda honesta e observar o resultado. Qual a probabilidade de se obter pelo menos uma cara entre  $n$  lançamentos? Ora, note que  $\mathcal{E} \leftrightarrow$  extrair uma amostra de tamanho  $n$ , com reposição, de uma população com 2 elementos,  $S = \{c, \hat{c}\}$ . Temos

$$\Omega = \{(x_1, \dots, x_n) : x_i = c, \hat{c}, i = 1, \dots, n\}.$$

Daí,  $|\Omega| = 2^n$  e tomamos  $\mathcal{A} = \mathcal{P}(\Omega)$  e  $P(\{\omega\}) = 1/2^n, \forall \omega \in \Omega$ , de modo que  $P(A) = |A|/2^n, \forall A \in \mathcal{A}$ . Se  $A = \{\text{obter pelo menos uma cara}\} \subset \Omega$ , então  $A^C =$

| {não obter nenhuma cara nos  $n$  lançamentos} =  $\{(\hat{c}, \dots, \hat{c})\}$ . Logo,  $P(A) = 1 - P(A^C) = 1 - 1/2^n$ .

## Amostragem sem reposição (arranjos e permutações)

Considere  $S$  um conjunto com  $m$  objetos distintos,  $S = \{s_1, \dots, s_m\}$ . Seleccionamos um objeto de  $S$  e não o repomos, repetindo esse processo  $n \leq m$  vezes. Temos, então,

$$\Omega = \{(s_1, \dots, s_n) : s_i \in S, 1 \leq i \leq n, s_i \neq s_j \forall i \neq j\}$$

e, portanto,

$$|\Omega| = m(m-1) \cdots (m-n+1) = \frac{m!}{(m-n)!} = A_{m,n}.$$

Note que se  $n = m$ , então  $A_{m,n} = A_{n,m} = P_m$  (permutação de  $m$  elementos).

**Exemplo.** Considere que temos 10 bolas numeradas de 1 a 10,  $S = \{1, \dots, 10\}$ , e 3 caixas distintas. Suponha que queiramos escolher ao acaso 3 bolas, sem reposição, e colocar cada uma em uma das caixas. Temos

$$\Omega = \{(x_1, x_2, x_3) : x_i \in S, i = 1, 2, 3, x_i \neq x_j, i \neq j\}.$$

Note que, aqui, estamos considerando  $(5, 2, 4) \neq (4, 2, 5)$  por exemplo, pois as caixas são distintas. Tomamos  $\mathcal{A} = \mathcal{P}(\Omega)$  e  $P(\{\omega\}) = 1/|\Omega|, \forall \omega \in \Omega$ , com  $|\Omega| = A_{10,3} = 10!/7! = 720$ . Se tivéssemos  $n$  bolas e  $n$  caixas, então

$$\Omega = \{(x_1, \dots, x_n) : x_i \in S, i = 1, \dots, n, x_i \neq x_j, i \neq j\},$$

sendo  $S = \{1, \dots, n\}$ . Daí, temos  $|\Omega| = n!$  e tomamos  $\mathcal{A} = \mathcal{P}(\Omega)$  e  $P(\{\omega\}) = 1/n!, \forall \omega \in \Omega$ . Seja  $A_{ij} = \{\text{bola } i \text{ na caixa } j\}$ , com  $i, j \in S$  fixos. Note que  $|A_{ij}| = (n-1)!$ . Logo,  $P(A_{ij}) = (n-1)!/n! = 1/n$  e, de modo geral, se  $A_k = \{k \text{ bolas específicas em } k \text{ caixas específicas}\}$ , então  $|A_k| = (n-k)!$  e  $P(A_k) = (n-k)!/n! = 1/A_{n,k}$ .

## 2.2 Amostras desordenadas e sem reposição (combinações)

Considere  $S$  com  $m$  objetos numerados distintos,  $S = \{1, \dots, m\}$ . Escolhemos, ao acaso,  $r \leq m$  elementos de  $S$ , sem reposição e sem considerar a ordem de escolha. Temos

$$\Omega = \{(x_1, \dots, x_r) : x_i \in S, i = 1, \dots, r, x_i < x_j, i < j\}.$$

Daí,  $|\Omega| = m!/(r!(m-r)!) = C_{m,r}$ . De fato, se  $|\Omega| \cdot P_r = A_{m,r}$ , donde segue que  $\Omega$  tem a forma acima. Denotamos também  $C_{m,r} = \binom{m}{r}$ .

| **Exemplo.** Sejam  $S = \{1, 2, 3, 4\}$ ,  $r = 3$ . O número de amostras é  $A_{4,3}/3! = 4 = C_{4,3}$ .

**Exemplo.** Suponha que há 75 professores no departamento de Matemática, sendo 25 titulares, 15 adjuntos e 35 assistentes. Queremos formar uma comissão de 6 membros, selecionados ao acaso. Seja  $A = \{\text{todos são assistentes}\}$ . Temos

$$\Omega = \{(x_1, \dots, x_6) : x_i \in \{1, \dots, 75\}, i = 1, \dots, 6 \text{ e } x_i < x_j, i < j\}.$$

Daí,  $|\Omega| = \binom{75}{6}$  e, ademais,  $A = \binom{35}{6}$ , de modo que  $P(A) = \binom{35}{6}/\binom{75}{6} \approx 1\%$ .

## 2.3 Permutações com elementos repetidos

(a). Considere um conjunto de  $n$  objetos distintos, agrupados em  $r$  espécies distintas:  $n_1$  da espécie 1,  $n_2$  da espécie 2, ...,  $n_r$  da espécie  $r$ , com  $n_1 + \dots + n_r = n$ . A quantidade de permutações dos  $n$  objetos é

$$\begin{aligned} P_n^{n_1, \dots, n_r} &= \underbrace{\binom{n}{n_1}}_{\text{lugares 1ª espécie}} \binom{n-n_1}{n_2} \cdots \binom{n-(n_1+n_2+\dots+n_{r-1})}{n_r} \\ &= \frac{n!}{n_1!(n-n_1)!} \cdot \frac{(n-n_1)!}{n_2!(n-(n_1+n_2))!} \cdots \frac{(n-(n_1+\dots+n_{r-1}))!}{n_r!(n-(n_1+\dots+n_r))!} \\ &= \frac{n!}{n_1!n_2! \cdots n_r!}. \end{aligned}$$

Denotamos também  $P_n^{n_1, \dots, n_r} = \binom{n}{n_1, n_2, \dots, n_r}$ . Note que se  $n_i = 1, i = 1, \dots, r$ , então  $P_n^{n_1, \dots, n_r} = P_n^{1, \dots, 1} = P_n = n!$  e, se  $r = 2$ , então

$$\binom{n}{n_1, n_2} = \frac{n!}{n_1! n_2!} = \begin{cases} \frac{n!}{n_1!(n-n_1)!} & = \binom{n}{n_1} \\ \frac{n!}{n_2!(n-n_2)!} & = \binom{n}{n_2} \end{cases}.$$

(b). Considere um conjunto com  $n$  item distintos que deve ser dividido em  $r$  grupos distintos de tamanhos  $n_1, \dots, n_r$ , com  $n_1 + \dots + n_r = n$ . O total de divisões possíveis é  $P_n^{n_1, \dots, n_r} = \binom{n}{n_1, n_2, \dots, n_r}$ . Note que tanto a expansão binomial quanto a multinomial usam desse raciocínio.

## 2.4 Partições — problema da urna

### Urna simples

Em uma urna, temos  $n$  bolas:  $n_A$  do tipo  $A$  e  $n_B$  do tipo  $B$ . Retira-se  $m \leq n$  bolas ao acaso. Queremos calcular a probabilidade de  $A_k = \{\text{a amostra tem exatamente } k \text{ bolas do tipo } A\}$ . Note que  $0 \leq k \leq \min\{n_A, m\}$  se não houver reposição e  $0 \leq k \leq m$  se houver reposição. Supondo resultados equiprováveis, temos  $P(A_k) = |A_k|/|\Omega|$ . Note que se  $x_i$  são as bolas do tipo  $A$  e  $y_j$  são as bolas do tipo  $B$ , então

$$\Omega = \{x_1, \dots, x_{n_A}, y_1, \dots, y_{n_B}\}.$$

Ademais,  $\Omega$  está em bijeção com  $\{1, \dots, n_A, n_A+1, \dots, n_A+n_B\}$ , de modo que podemos pensar também  $\Omega = \{1, \dots, n\}$ .

**Amostragem sem reposição.** Ao considerar  $A_k$ , queremos apenas a quantidade de bolas de cada tipo escolhidas, sem importar a ordem. Isso sugere que  $P(A_k)$  fica inalterada se considerarmos ou não a ordem, e veremos que esse, de fato, é o caso.

(a1) Considerando amostras não ordenadas, temos

$$\Omega = \{\omega = (\omega_1, \dots, \omega_m) : \omega_i \in S, i = 1, \dots, m, \omega_i < \omega_j, i < j\},$$

$\mathcal{A} = \mathcal{P}(\Omega)$  e  $P(A) = |A|/|\Omega|, \forall A \in \mathcal{A}$ . Note que  $|\Omega| = \binom{n}{m}$  e, para todo  $0 \leq k \leq \min\{n_A, m\}$ ,

temos  $|A_k| = \binom{n_A}{k} \binom{n_B}{m-k}$ . Daí, temos

$$P(A_k) = \frac{\binom{n_A}{k} \binom{n_B}{m-k}}{\binom{n}{m}} = \frac{\binom{n_A}{k} \binom{n-n_A}{m-k}}{\binom{n}{m}}, 0 \leq k \leq \min\{n_A, m\},$$

chamada de **probabilidade hipergeométrica**.

(a2) Considerando amostras ordenadas, temos

$$\Omega = \{\omega = (\omega_1, \dots, \omega_m) : \omega_i \in S_i, i = 1, \dots, m \text{ e } \omega_i \neq \omega_j, i \neq j\}.$$

A  $\sigma$ -álgebra e a medida de probabilidade são as mesmas do caso acima. Agora, temos  $|\Omega| = A_{n,m}$

e, para todo  $0 \leq k \leq \min\{n_A, m\}$ ,  $|A_k| = \binom{m}{k, m-k} A_{n_A, k} A_{n_B, m-k}$ . Daí,

$$P(A_k) = \frac{\binom{m}{k, m-k} \frac{n_A!}{(n_A-k)!} \cdot \frac{n_B!}{(n_B-m+k)!}}{\frac{n!}{(n-m)!}} = \frac{\binom{n_A}{k} \binom{n_B}{m-k}}{\binom{n}{m}}.$$

**Amostragem com reposição.** Neste caso, temos

$$\Omega = \{\omega = (\omega_1, \dots, \omega_m) : \omega_i \in S, i = 1, \dots, m\},$$

com  $m \leq n$  e  $\mathcal{A}$  e  $P$  como acima. Note que  $|\Omega| = n^m$  e, para todo  $k \in \{0, \dots, m\}$ ,  $|A_k| = \binom{m}{k, m-k} n_A^k n_B^{m-k}$  e  $P(A_k) = \binom{m}{k} \frac{n_A^k n_B^{m-k}}{n^m}$ ,  $k = 0, \dots, m$ , chamada de **probabilidade binomial**. Note também que se considerarmos a retirada de uma bola do tipo  $A$  como “sucesso”, então  $P(\text{sucesso}) = n_A/n := p$  e  $P(\text{fracasso}) = 1 - n_A/n = 1 - p$ . Devido à reposição, os resultados das retiradas são mutuamente independentes, de modo que  $P(A_k) = P(\text{exatamente } k \text{ sucessos}) = \binom{m}{k} p^k (1-p)^{m-k}$ ,  $k = 0, \dots, m$ .

## Urna geral

Agora, consideremos que em uma urna há  $n$  bolas, sendo  $n_i$  do tipo  $i$ , com  $i = 1, \dots, r$ . Note que  $n_1 + \dots + n_r = n$ . Retiramos  $m \leq n$  bolas ao acaso, e estamos interessados em calcular a probabilidade de  $A = A_{k_1, \dots, k_r} = \{\text{conter } k_i \text{ bolas do tipo } i, i = 1, \dots, r\}$ , com  $k_1 + \dots + k_r = m$ . Note que  $0 \leq k_i \leq \min\{n_i, m\}$ ,  $i = 1, \dots, r$  se não houver reposição e  $0 \leq k_i \leq m$ ,  $i = 1, \dots, r$  caso contrário. Sem perda de generalidade, podemos descrever o conjunto de bolas como  $S = \{1, \dots, n\}$ .

**Amostragem sem reposição.** Como no caso da urna simples, a ordem não importa e temos

$$\Omega = \{\omega = (\omega_1, \dots, \omega_m) : \omega_i \in S, i = 1, \dots, m, \omega_i < \omega_j, i < j\},$$

$$|\Omega| = \binom{n}{m} \text{ e } |A| = \binom{n_1}{k_1} \binom{n_2}{k_2} \dots \binom{n_r}{k_r}, \text{ logo}$$

$$P(A) = \frac{\binom{n_1}{k_1} \dots \binom{n_r}{k_r}}{\binom{n}{m}}.$$

**Amostragem com reposição.** Analogamente ao caso da urna simples, temos  $|\Omega| = n^m$ ,  $|A| = \binom{m}{k_1, \dots, k_r} n_1^{k_1} \dots n_r^{k_r}$ , logo

$$P(A) = \binom{m}{k_1, \dots, k_r} \frac{n_1^{k_1} \dots n_r^{k_r}}{n^m}, 0 \leq k_i \leq m, i = 1, \dots, r,$$

chamada **probabilidade multinomial**. Observe que, em cada retirada,  $P(\text{retirar bola do tipo } i) = n_i/n := p_i$ . A reposição implica independência entre os resultados de cada retirada, logo

$$P(A) = \binom{m}{k_1, \dots, k_r} p_1^{k_1} \dots p_r^{k_r}.$$

**Exemplo.** Em um lote de 100 peças, 20 são defeituosas e 80 estão em perfeitas condições. Escolhe-se, ao acaso e sem reposição, 10 peças. Estamos interessados em calcular a probabilidade de que exatamente metade das peças escolhidas seja defeituosa. Note que aqui temos um problema de urna simples, com  $n_1 = 20$ ,  $n_2 = 80$ ,  $m = 10$  e  $k = 5$ . Daí, sendo  $A_k = \{\text{a amostra contém 5 peças defeituosas e 5 peças perfeitas}\}$ , temos que

$$P(A_k) = \frac{\binom{20}{5} \binom{80}{5}}{\binom{100}{10}}.$$

## 3 Variáveis aleatórias

Começemos considerando o seguinte exemplo motivador: temos o experimento “lançar uma moeda 3 vezes”. A cada lançamento, podemos obter  $c$  (cara) com probabilidade  $p$  ou  $\hat{c}$  (coroa) com probabilidade  $1-p$ . Suponha que ganhamos R\$1,00 para cada cara obtida e perdemos o mesmo valor para cada coroa obtida. Assim, ao final do experimento, poderemos ter  $-3, -1, 1$  ou  $3$  reais (onde  $-x$  reais indica prejuízo de  $x$  reais). Estamos interessados em calcular a probabilidade de cada um desses resultados. Nosso espaço de probabilidade é  $(\Omega, \mathcal{A}, P)$  com

$$\Omega = \{(c, c, c), (c, c, \hat{c}), (c, \hat{c}, c), (\hat{c}, c, c), (c, \hat{c}, \hat{c}), (\hat{c}, c, \hat{c}), (\hat{c}, \hat{c}, c), (\hat{c}, \hat{c}, \hat{c})\},$$



$\mathcal{A} = \mathcal{P}(\Omega)$  e nossa medida de probabilidade é dada por

$$P(\{(c, c, c)\}) = p^3, P(\{(c, c, \hat{c})\}) = P(\{(c, \hat{c}, c)\}) = P(\{(\hat{c}, c, c)\}) = p^2(1-p),$$

$$P(\{(c, \hat{c}, \hat{c})\}) = P(\{(\hat{c}, c, \hat{c})\}) = P(\{(\hat{c}, \hat{c}, c)\}) = p(1-p)^2, P(\{(\hat{c}, \hat{c}, \hat{c})\}) = (1-p)^3.$$

Definamos  $X : \Omega \rightarrow \mathbb{R}$  tal que  $X(\omega) =$  quantia obtida pelo resultado  $\omega, \forall \omega \in \Omega$ . Assim, temos

$$X((c, c, c)) = 3, X((c, c, \hat{c})) = X((c, \hat{c}, c)) = X((\hat{c}, c, c)) = 1,$$

$$X((c, \hat{c}, \hat{c})) = X((\hat{c}, c, \hat{c})) = X((\hat{c}, \hat{c}, c)) = -1, X((\hat{c}, \hat{c}, \hat{c})) = -3,$$

ou seja,  $\text{Im}(X) = \{-3, -1, 1, 3\}$ . Observe que

$$\{X = 3\} := \{\omega \in \Omega : X(\omega) = 3\} = \{(c, c, c)\} \subset \Omega : \{X = 3\} \in \mathcal{A}$$

$$\{X = 1\} := \{\omega \in \Omega : X(\omega) = 1\} = \{(c, c, \hat{c}), (c, \hat{c}, c), (\hat{c}, c, c)\} \subset \Omega : \{X = 1\} \in \mathcal{A}$$

$$\{X = -1\} := \{\omega \in \Omega : X(\omega) = -1\} = \{(c, \hat{c}, \hat{c}), (\hat{c}, c, \hat{c}), (\hat{c}, \hat{c}, c)\} \subset \Omega : \{X = -1\} \in \mathcal{A}$$

$$\{X = -3\} := \{\omega \in \Omega : X(\omega) = -3\} = \{(\hat{c}, \hat{c}, \hat{c})\} \subset \Omega : \{X = -3\} \in \mathcal{A}$$

$$\{X = x\} := \{\omega \in \Omega : X(\omega) = x\} = \emptyset \in \mathcal{A}, \forall x \in \mathbb{R} \setminus \text{Im}(X).$$

Ademais,  $\{X = -3\} \cup \{X = -1\} \cup \{X = 1\} \cup \{X = 3\} = \Omega \in \mathcal{A}$  e note que

$$\{X \leq 1\} := \{\omega \in \Omega : X(\omega) \leq 1\} = \{X = -3\} \cup \{X = -1\} \cup \{X = 1\} \in \mathcal{A}$$

$$\{X \leq 0,5\} := \{\omega \in \Omega : X(\omega) \leq 0,5\} = \{X = -3\} \cup \{X = -1\} \in \mathcal{A}.$$

De maneira geral, como  $\text{Im}(X) \subset \mathbb{Z}$ , então temos

$$\{X \leq x\} = \{\omega \in \Omega : X(\omega) \leq x\} = \bigcup_{i: x_i \leq [x]} \{X = x_i\} \in \mathcal{A}, \quad x_i \text{ valor possível para } x,$$

ou seja,  $\{X \leq x\}$  é um evento para todo  $x$  real. Assim, podemos calcular

$$P(X = 3) = p^3, P(X = 1) = 3p^2(1-p), P(X = -1) = 3p(1-p)^2,$$

$$P(X = -3) = (1-p)^3 \text{ e } P(X = x) = 0, \forall x \in \mathbb{R} \setminus \text{Im}(X).$$

**Definição 3.1** (Variável aleatória). Uma variável aleatória  $X$  em um espaço de probabilidade  $(\Omega, \mathcal{A}, P)$  é uma função  $X : \Omega \rightarrow \mathbb{R}$  tal que  $\{X \leq x\} = \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{A}, \forall x \in \mathbb{R}$ .

**Observações.** (a) Note que

$$\{X \leq x\} = X^{-1}((-\infty, x]) = \{\omega \in \Omega : X(\omega) \in (-\infty, x]\}, \forall x \in \mathbb{R}.$$

(b) De modo geral,  $X : \Omega \rightarrow \mathbb{R}$  é uma variável aleatória (v.a.) em  $(\Omega, \mathcal{A}, P)$  se

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A}, \forall B \in \mathcal{B}(\mathbb{R}).$$

(c) Dado um espaço de probabilidade  $(\Omega, \mathcal{A}, P)$ , se

- $\Omega$  é enumerável (caso discreto), então  $\mathcal{A} = \mathcal{P}(\Omega)$  e toda função  $X : \Omega \rightarrow \mathbb{R}$  é uma v.a. (não será demonstrado).
- $\Omega \subset \mathbb{R}^n, n \geq 1$  é não enumerável, então  $\mathcal{A} = \mathcal{B}^n(\mathbb{R})$ . Nesse caso, nem toda função  $X : \Omega \rightarrow \mathbb{R}$  é uma v.a., mas as exceções são raras.

Todas as funções que definirmos no contexto de v.a.'s serão de fato v.a.'s, e portanto não será necessário verificar pela definição.

**Exemplos.** 1. No experimento da motivação,  $X$  é v.a.; de fato, vimos que  $\{X \leq x\} \in \mathcal{A}, \forall x \in \mathbb{R}$ .

2. Suponha que um ponto é escolhido ao acaso no círculo unitário centrado na origem. Seja  $Y$  a distância desse ponto à origem. Temos

$$\Omega = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}, \mathcal{A} = \mathcal{B}^2(\Omega)$$

e,  $\forall A \in \mathcal{A}, P(A) = \iint_A \frac{1}{\pi} dx dy$ . Daí, como  $Y : \Omega \rightarrow \mathbb{R}$  é dada por  $Y(\omega) = Y((x, y)) = \sqrt{x^2 + y^2}$ , temos que dado  $z \in \mathbb{R}$ ,

$$\{Y \leq z\} = \{(x, y) \in \Omega : \sqrt{x^2 + y^2} \leq z\} = \begin{cases} \emptyset, & z < 0, \\ \{(x, y) : \sqrt{x^2 + y^2} \leq z\}, & 0 \leq z < 1, \\ \Omega, & z \geq 1 \end{cases},$$

ou seja,  $\{Y \leq z\} \in \mathcal{A}, \forall z \in \mathbb{R}$ , de modo que  $Y$  é v.a. em  $(\Omega, \mathcal{A}, P)$ .

**Definição 3.2.** Seja  $X$  v.a. em  $(\Omega, \mathcal{A}, P)$ .

- (i)  $X$  é discreta se  $\text{Im}(X) \subset \mathbb{R}$  é finita ou enumerável;
- (ii)  $X$  é contínua se  $P(X = x) = 0, \forall x \in \mathbb{R}$ .

Note que se  $X$  é v.a. discreta com imagem enumerável, digamos  $\text{Im}(X) = \{x_1, x_2, \dots\}$ , então

- (a)  $P(X = x_i) > 0, i = 1, 2, \dots$
- (b)  $P(X = x_i) = 0, x \neq x_i, i = 1, 2, \dots$
- (c)  $\bigcup_{i=1}^{\infty} \{X = x_i\} = \Omega$
- (d)  $\sum_{i=1}^{\infty} P(X = x_i) = 1$

**Exemplo.** No Exemplo 1 anterior,  $X$  é v.a. discreta e, no Exemplo 2,  $Y$  é v.a. contínua pois  $P(Y = z) = \{(x, y) \in \Omega : x^2 + y^2 = z^2\} = 0, \forall z \in \mathbb{R}$ .

Concentraremos o nosso estudo primeiro nas v.a.'s discretas e, mais à frente, nas contínuas.

### 3.1 Variáveis aleatórias discretas

No estudo das v.a.'s, tanto discretas quanto contínuas, utilizamos de funções que determinam o “comportamento” da referida v.a.; no caso de v.a.'s discretas, temos a **função de probabilidade (f.p.)** e a **função de distribuição (f.d.)**. Começamos pela função de probabilidade.

**Definição 3.3.** Seja  $X$  v.a. discreta em  $(\Omega, \mathcal{A}, P)$ . A **função de probabilidade de  $X$**  é a função  $p_X : \mathbb{R} \rightarrow [0, 1]$  dada por

$$p_X(x) = P(X = x), \forall x \in \mathbb{R}.$$

A função de probabilidade associa, a cada número real, a probabilidade de  $X$  assumir tal valor.

**Observação.** Se  $x \in \mathbb{R}$  é tal que  $p_X(x) > 0$ , dizemos que  $x$  é um **valor possível** de  $X$ . Além disso, note que se  $X$  é discreta então existe  $\{x_1, x_2, \dots\} \subset \mathbb{R}$  finito ou enumerável tal que  $p_X(x_i) > 0, i = 1, 2, \dots$  e  $p_X(x) = 0, x \notin \{x_1, x_2, \dots\}$ . Como os eventos  $\{X = x_i\}, i = 1, 2, \dots$  são disjuntos e  $\bigcup_{i=1}^{\infty} \{X = x_i\} = \Omega$  (ou seja, esses eventos formam uma partição de  $\Omega$ ), segue que  $\sum_{i=1}^{\infty} p_X(x_i) = 1$ .

**Exemplo.** No exemplo motivador, temos  $P(\text{cara}) = p = 1/2$ . A função de probabilidade de  $X$  é dada por  $p_X(-3) = P(X = -3) = 1/8, p_X(-1) = p_X(1) = 3/8, p_X(3) = 1/8$  e  $p_X(x) = 0, \forall x \in \mathbb{R} \setminus \{-3, -1, 1, 3\}$ .

**Exemplo.** Dados  $(\Omega, \mathcal{A}, P)$  espaço de probabilidade e  $c \in \mathbb{R}$  constante, defina  $X : \Omega \rightarrow \mathbb{R}$  tal que  $X(\omega) = c, \forall \omega \in \Omega$ . Daí, dado  $x \in \mathbb{R}$ ,

$$\{X \leq x\} = \begin{cases} \emptyset, & x < c \\ \Omega, & x \geq c \end{cases} \in \mathcal{A},$$

ou seja,  $X$  é v.a. Como  $P(X = c) = P(\Omega) = 1$  e  $P(X = x) = 0, x \neq c$ , temos que  $X$  é v.a. discreta com apenas um valor possível,  $c$ .

**Exemplo.** Dados  $(\Omega, \mathcal{A}, P)$  espaço de probabilidade e  $A \in \mathcal{A}$ , defina  $X : \Omega \rightarrow \mathbb{R}$  tal que

$$X(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}.$$

Daí, dado  $x \in \mathbb{R}$  temos

$$\{X \leq x\} = \begin{cases} \emptyset, & x < 0 \\ A^C, & 0 \leq x < 1 \\ \Omega, & x \geq 1 \end{cases} \in \mathcal{A},$$

logo  $X$  é v.a. com dois valores possíveis, 0 e 1, chamada de **v.a. indicadora de  $A$** , usualmente denotada por  $1_A$  ou  $1_A$ . Se considerarmos  $P(A) = p$ , temos

$$p_X(x) = \begin{cases} p, & x = 1 \\ 1 - p, & x = 0 \\ 0, & x \in \mathbb{R} \setminus \{0, 1\} \end{cases} \quad (\text{função de probabilidade **Bernoulli** de parâmetro } p)$$

Reciprocamente, se  $X$  é uma v.a. discreta definida em  $(\Omega, \mathcal{A}, P)$  com conjunto de valores possíveis  $\{0, 1\}$ , então  $X$  é v.a. indicadora. De fato, denotando  $P(X = 1) = P(A) = p$ , a função de probabilidade de  $X$  é dada por

$$p_X(x) = \begin{cases} p, & x = 1 \\ 1 - p, & x = 0 \\ 0, & x \in \mathbb{R} \setminus \{0, 1\} \end{cases}.$$

Logo,  $X = 1_A$  sendo  $A = \{\omega \in \Omega : X(\omega) = 1\}$ .

**Proposição 3.4.** Seja  $X$  v.a. discreta em  $(\Omega, \mathcal{A}, P)$ . A função de probabilidade de  $X$ ,  $p_X$ , é tal que

(P1)  $p_X(x) \geq 0, \forall x \in \mathbb{R}$ ;

(P2)  $\{x \in \mathbb{R} : p_X(x) \neq 0\} = \{x_1, x_2, \dots\} \subset \mathbb{R}$  é finito ou infinito enumerável;

(P3)  $\sum_{i=1}^{\infty} p_X(x_i) = 1$ .

**Demonstração.** Os três itens seguem da definição de v.a. discreta e das propriedades da medida de probabilidade.  $\square$

**Definição 3.5** (Função de probabilidade). Dizemos que  $p : \mathbb{R} \rightarrow \mathbb{R}$  é uma **função de probabilidade** se satisfaz (P1), (P2) e (P3).

**Proposição 3.6.** Se  $p : \mathbb{R} \rightarrow \mathbb{R}$  é uma função de probabilidade, então existe um espaço de probabilidade  $(\Omega, \mathcal{A}, P)$  e uma v.a.  $X$  definida neste espaço tal que  $p = p_X$ .

**Demonstração.** Por hipótese,  $p$  é f.p., logo  $\{x \in \mathbb{R} : p_X(x) \neq 0\} = \{x \in \mathbb{R} : p_X(x) > 0\} = \{x_1, x_2, \dots\} \subset \mathbb{R}$  é finito ou enumerável. Tomemos  $\Omega = \{x_1, x_2, \dots\}$ ,  $\mathcal{A} = \mathcal{P}(\Omega)$  e  $P : \mathcal{A} \rightarrow \mathbb{R}$  dada por

$$P(\{x_i\}) = p(x_i), i = 1, 2, \dots$$

e, dado  $A \in \mathcal{A}$ ,

$$P(A) = \sum_{i: x_i \in A} p(x_i).$$

De (P1) e (P3), segue que  $P$  é medida de probabilidade em  $(\Omega, \mathcal{A})$ . Definamos  $X : \Omega \rightarrow \mathbb{R}$  tal que

$$X(\omega) = \omega \iff X(x_i) = x_i, i = 1, 2, \dots$$

Então  $X$  é v.a. discreta com valores possíveis  $x_1, x_2, \dots$  e f.p.

$$p_X(x_i) = P(X = x_i) = P(\{\omega \in \Omega : X(x_i) = x_i\}) = P(\{x_i\}) = p(x_i), i = 1, 2, \dots$$

$\square$

## 3.2 Exemplos clássicos de v.a.'s discretas

**Exemplo** (Uniforme discreta). Sejam  $\mathcal{E}$  um experimento com  $n$  resultados possíveis e equiprováveis,  $\Omega = \{\omega_1, \dots, \omega_n\}$ ,  $\mathcal{A} = \mathcal{P}(\Omega)$  e  $P : \mathcal{A} \rightarrow \mathbb{R}$  dada por  $P(A) = |A|/|\Omega|$ . Seja  $X : \Omega \rightarrow \mathbb{R}$  dada por

$X(\omega_i) = x_i, i = 1, 2, \dots, n$ .  $X$  é v.a. com valores possíveis  $\{x_1, \dots, x_n\}$ , e temos

$$p_X(x) = P(X = x) = \begin{cases} 1/n, x \in \{x_1, \dots, x_n\} \\ 0, \text{ c.c.} \end{cases}.$$

**Exemplo** (Binomial de parâmetros  $n$  e  $p$ ,  $X \sim B(n, p)$ ). Sejam  $\mathcal{E}$  um experimento,  $(\Omega, \mathcal{A}, P)$  espaço de probabilidade e  $A$  um evento associado a  $\mathcal{E}$  tal que  $P(A) = p$  (sucesso) e, conseqüentemente,  $P(A^C) = 1 - p$  (fracasso), com  $0 < p < 1$ . Suponha que são realizadas  $n \geq 1$  repetições **independentes** de  $\mathcal{E}$  (provas ou ensaios de Bernoulli), e seja  $X$  o número de ocorrências de  $A$  (sucessos) nas  $n$  repetições. Temos  $X$  v.a. com conjunto de valores possíveis  $\{0, 1, \dots, n\}$ , donde

$$p_X(k) = P(X = k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, k = 0, 1, \dots, n \\ 0, \text{ c.c.} \end{cases}.$$

Note que a função  $p := p_X$  define uma função de probabilidade.

**Demonstração.**

(P1)  $p_X(x) = \binom{n}{x} p^x (1-p)^{n-x} > 0$  para  $x \in \{0, 1, \dots, n\}$  e  $p_X(x) = 0$  para  $x \in \mathbb{R} \setminus \{0, 1, \dots, n\}$ .

(P2)  $\{x \in \mathbb{R} : p_X(x) \neq 0\} = \{0, 1, \dots, n\}$  é finito.

(P3)  $\sum_i p_X(x_i) = \sum_{k=0}^n p_X(k) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = (p + 1 - p)^n = 1$ .

□

Se  $n = 1$ ,  $X$  é v.a. **de Bernoulli** de parâmetro  $p$ :

$$p_X(k) = \begin{cases} p, k = 1 \\ 1 - p, k = 0 \\ 0, \text{ c.c.} \end{cases}$$

**Exemplo** (Geométrica de parâmetro  $p$ ,  $X \sim \text{Geom}(p)$ ). Sejam  $\mathcal{E}$  um experimento,  $(\Omega, \mathcal{A}, P)$  espaço de probabilidade e  $A$  um evento associado a  $\mathcal{E}$  tal que  $P(A) = p$  (sucesso) e  $P(A^C) = 1 - p$  (fracasso), com  $0 < p < 1$ . Repetimos  $\mathcal{E}$  até que  $A$  ocorra pela 1ª vez. Temos  $X$  v.a. com conjunto de valores possíveis  $\{1, 2, \dots\}$ . Note que  $\{X = k\} = \{A^C \text{ ocorre nas primeiras } k - 1 \text{ repetições e } A \text{ ocorre na } k\text{-ésima}\}$ . Daí,

$$p_X(k) = P(X = k) = \begin{cases} p(1-p)^{k-1}, k = 1, 2, \dots \\ 0, \text{ c.c.} \end{cases}$$

Note que a função  $p := p_X : \mathbb{R} \rightarrow \mathbb{R}$  define uma função de probabilidade.

**Demonstração.**

(P1)  $p_X(x) = p(1-p)^{x-1} > 0$  se  $x \in \{1, 2, \dots\}$  e  $p_X(x) = 0$  se  $x \in \mathbb{R} \setminus \{1, 2, \dots\}$ .

(P2)  $\{x \in \mathbb{R} : p_X(x) \neq 0\} = \{1, 2, \dots\}$  é enumerável.

(P3)  $\sum_i p_X(x_i) = \sum_{k=1}^{\infty} p(1-p)^{k-1} = p \frac{1}{p-1} = 1$ .

□

Podemos também representar a f.p. da v.a. geométrica de parâmetros  $n$  e  $p$  por (vide Exercício 2b, Lista 3)

$$p_X(k) = \begin{cases} p(1-p)^k, k = 0, 1, \dots \\ 0, \text{ c.c.} \end{cases}$$

**Exemplo** (Binomial negativa de parâmetros  $r$  e  $p$ ,  $X \sim BN(r, p)$ ). Estamos na mesma situação da v.a. geométrica, mas agora  $X$  é o número de repetições necessárias para que  $A$  ocorra exatamente  $r$  vezes.  $X$  é uma v.a. com conjunto de valores possíveis  $\{r, r + 1, \dots\}$ . Note que  $\{X = k\} =$

{A ocorre na  $k$ -ésima repetição e em exatamente  $r - 1$  vezes nas  $k - 1$  repetições anteriores}. Daí,

$$p_X(k) = P(X = k) = \begin{cases} \binom{k-1}{r-1} p^r (1-p)^{k-r}, & k = r, r+1, \dots \\ 0, & \text{c.c.} \end{cases}$$

Note que se  $r = 1$ ,  $X \sim \text{Geo}(p)$ . Ademais, se definirmos  $\binom{-r}{0} = 1$  e  $\binom{-r}{m} = (-1)^m \cdot \frac{(r+1) \cdots (r+m-1)}{m!}$ ,  $m \geq 1$ , temos  $\binom{k-1}{r-1} = (-1)^{k-r} \binom{-r}{k-r}$ ,  $k = r, r+1, \dots$

Daí, podemos reescrever  $p_X$  como

$$p_X(k) = \begin{cases} \binom{-r}{k-r} (-1)^{k-r} p^r (1-p)^{k-r}, & k = r, r+1, \dots \\ 0, & \text{c.c.} \end{cases}$$

Note também que  $p := p_X : \mathbb{R} \rightarrow \mathbb{R}$  dada acima (usaremos a primeira forma por simplicidade) é função de probabilidade.

### Demonstração.

(P1)  $p_X(x) = \binom{k-1}{r-1} p^r (1-p)^{k-r} > 0$  se  $x \in \{r, r+1, \dots\}$  e  $p_X(x) = 0$  se  $x \in \mathbb{R} \setminus \{r, r+1, \dots\}$ .

(P2)  $\{x \in \mathbb{R} : p_X(x) \neq 0\} = \{r, r+1, \dots\}$  é enumerável.

(P3) Temos

$$(1-t)^{-r} = \sum_{j=0}^{\infty} \underbrace{\binom{-r}{j} (-1)^j t^j}_{\binom{j+r-1}{r-1}}, |t| < 1.$$

Logo,

$$\begin{aligned} \sum_i p_X(x_i) &= \sum_{k=r}^{\infty} \binom{k-1}{r-1} p^r (1-p)^{k-r} = p^r \sum_{j=0}^{\infty} \binom{j+r-1}{r-1} (1-p)^j \\ &= p^r \sum_{j=0}^{\infty} \binom{-r}{j} (-1)^j (1-p)^j \\ &= p^r \frac{1}{p^r} = 1. \end{aligned}$$

□

Por fim, a f.p. da v.a. binomial negativa de parâmetros  $r$  e  $p$  pode ser reescrita como (vide Exercício 2c, Lista 3)

$$p_X(k) = \begin{cases} \binom{r+k-1}{r-1} p^r (1-p)^k, & k = 0, 1, \dots \\ 0, & \text{c.c.} \end{cases}$$

**Exemplo** (Hipergeométrica de parâmetros  $N, r$  e  $n$ ,  $X \sim \text{Hgeo}(N, r, n)$ ). Considere um conjunto com  $N$  objetos,  $r$  do tipo 1 e  $N - r$  do tipo 2. Uma amostra de tamanho  $n \leq r$  é selecionada, sem reposição. Seja  $X$  o número de objetos do tipo 1 na amostra. Temos  $X$  v.a. com conjunto de valores possíveis  $\{0, 1, \dots, n\}$ . Note que  $\{X = k\} = \{\text{a amostra tem exatamente } k \text{ objetos do tipo 1}\}$ . Daí,

$$p_X(k) = \begin{cases} \frac{\binom{r}{k} \binom{N-r}{n-k}}{\binom{N}{n}}, & k = 0, 1, \dots, n \\ 0, & \text{c.c.} \end{cases}$$

Note que  $p := p_X : \mathbb{R} \rightarrow \mathbb{R}$  é, de fato, uma função de probabilidade.

**Demonstração.**

(P1)  $p_X(x) = \binom{r}{x} \binom{N-r}{n-x} / \binom{N}{n} > 0$  se  $x \in \{0, 1, \dots, n\}$  e  $p_X(x) = 0$  se  $x \in \mathbb{R} \setminus \{0, 1, \dots, n\}$ .

(P2)  $\{x \in \mathbb{R} : p_X(x) \neq 0\} = \{0, 1, \dots, n\}$  é finito.

(P3)  $\sum_i p_X(x_i) = \sum_{k=0}^n \frac{\binom{r}{k} \binom{N-r}{n-k}}{\binom{N}{n}} \stackrel{*}{=} \binom{N}{n} / \binom{N}{n} = 1$ , onde em \* usamos a **identidade de Vandermonde**:

$$\binom{a+b}{n} = \sum_{k=0}^n \binom{a}{k} \binom{b}{n-k}, n \leq \min\{a, b\} \text{ e } a, b, \in \mathbb{N}.$$

□

**Exemplo** (Poisson com parâmetro  $\lambda > 0$ ,  $X \sim \text{Poisson}(\lambda)$ ). Considere  $n$  ensaios de Bernoulli com probabilidade de sucesso  $p_n = \lambda/n$ ,  $\lambda \in \mathbb{R}_{>0}$  fixo. Para cada  $n \geq 1$ , seja  $X_n$  o número de sucessos nas  $n$  provas. Daí, para cada  $n$ , temos  $X_n \sim B(n, p_n)$ , ou seja,

$$P(X_n = k) = \begin{cases} \binom{n}{k} p_n^k (1-p_n)^{n-k}, & k = 0, 1, \dots, n \\ 0, & \text{c.c.} \end{cases}.$$

Note que para cada  $k \in \{0, 1, \dots, n\}$ , temos

$$\begin{aligned} P(X_n = k) &= \frac{n(n-1)\cdots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \cdot \frac{n(n-1)\cdots(n-k+1)}{n^k} \left(1 - \frac{\lambda}{n}\right)^{-k} \left(1 - \frac{\lambda}{n}\right)^n \xrightarrow{n \rightarrow \infty} e^{-\lambda} \frac{\lambda^k}{k!}, \end{aligned}$$

de modo que

$$p_X(k) = \begin{cases} e^{-\lambda} \lambda^k / k!, & k = 0, 1, \dots \\ 0, & \text{c.c.} \end{cases}.$$

Temos que  $p := p_X : \mathbb{R} \rightarrow \mathbb{R}$  define uma função de probabilidade.

**Demonstração.**

(P1)  $p(x) = e^{-\lambda} \lambda^k / k! > 0$  se  $x \in \{0, 1, \dots\}$  e  $p(x) = 0$  se  $x \in \mathbb{R} \setminus \{0, 1, \dots\}$ .

(P2)  $\{x \in \mathbb{R} : p(x) \neq 0\} = \{0, 1, \dots\}$  é enumerável.

(P3)  $\sum_i p(x_i) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$ .

□

**Exemplo** (Zipf ou zeta). Colocamos este último exemplo de curiosidade. Uma v.a.  $X$  tem distribuição zeta (ou Zipf) se sua f.p. é dada por

$$p_X(k) = \begin{cases} \frac{C}{k^{\alpha+1}}, & k = 1, 2, \dots \\ 0, & \text{c.c.} \end{cases}, \alpha \in \mathbb{R}_{>0},$$

onde

$$C = \left[ \sum_{k=1}^{\infty} \left(\frac{1}{k}\right)^{\alpha+1} \right]^{-1}.$$

O nome de distribuição zeta vem do fato de que a função

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, s > 1$$

recebe o nome de **função zeta de Riemann**. Essa distribuição foi usada por um economista italiano, V. Pareto, para descrever a distribuição dos rendimentos das famílias de um dado país, mas foi G. K. Zipf(!) que aplicou a distribuição zeta em vários problemas de diferentes áreas, popularizando-a.

Tratada a f.p., passemos à função de distribuição. Agora, estamos interessados em calcular  $P(X \in A) = P(\omega \in \Omega : X(\omega) \in A)$ ,  $\forall A \subset \mathbb{R}$ . Consideraremos  $X$  uma v.a. discreta definida em  $(\Omega, \mathcal{A}, P)$  com

f.p.  $p_X$  e conjunto de valores possíveis  $\{x_1, x_2, \dots\}$  (finito ou enumerável). Note que dado  $\mathcal{B}(\mathbb{R}) \ni A \subset \mathbb{R}$ ,  $\{X \in A\}$  é um evento (pois  $X$  é v.a.), de modo que faz sentido calcular  $P(X \in A)$ . Temos que

$$\begin{aligned} P(X \in A) &= P\left(\bigcup_{i:x_i \in A} \{\omega \in \Omega : X(\omega) = x_i\}\right) = \sum_{i:x_i \in A} P(\{\omega \in \Omega : X(\omega) = x_i\}) \\ &= \sum_{i:x_i \in A} P(X = x_i) \\ &= \sum_{i:x_i \in A} p_X(x_i), \end{aligned}$$

ou seja,

$$P(X \in A) = \sum_{i:x_i \in A} p_X(x_i).$$

Portanto, vemos que a f.p. determina completamente a probabilidade  $P(X \in A), \forall A \in \mathcal{B}(\mathbb{R})$ .

**Definição 3.7** (Função de distribuição). Se  $X$  é uma v.a. definida em  $(\Omega, \mathcal{A}, P)$ , então  $F_X : \mathbb{R} \rightarrow \mathbb{R}$  dada por

$$F_X(x) = P(X \leq x) = P(X \in (-\infty, x]) = P(\omega \in \Omega : X(\omega) \leq x), \forall x \in \mathbb{R}$$

é chamada **função de distribuição** de  $X$ .

O exposto acima sugere que a f.p. determina por completo a f.d. de um v.a.  $X$  qualquer. De fato, esse é o caso, e há, na verdade, uma relação biunívoca entre ambas, como mostra a seguinte proposição.

**Proposição 3.8.** Seja  $X$  v.a. discreta definida em  $(\Omega, \mathcal{A}, P)$  com conjunto de valores possíveis  $\{x_1, x_2, \dots\}$ .

(a) Se  $X$  tem f.p.  $p_X$ , então  $F_X$  é dada por

$$F_X(x) = P(X \leq x) = \sum_{i:x_i \leq x} p_X(x_i), \forall x \in \mathbb{R}.$$

(b) Suponha  $x_1 \leq x_2 \leq \dots$ . Se  $X$  tem f.d.  $F_X$ , então sua f.p.  $p_X$  é dada por

$$p_X(x_i) = P(X = x_i) = F_X(x_i) - F_X(x_{i-1}), i = 1, 2, \dots$$

**Demonstração.**

(a) Dado  $x \in \mathbb{R}$ , temos

$$F_X(x) = P(X \leq x) = P(X \in [-\infty, x]) = \sum_{i:x_i \leq x} P(X = x_i) = \sum_{i:x_i \leq x} p_X(x_i).$$

(b) Note que  $\{X \leq x_i\} = \{X = x_i\} \sqcup \{X < x_i\}$ , donde

$$\begin{aligned} p_X(x_i) &= P(X = x_i) = P(X \leq x_i) - P(X < x_i) \\ &= P(X \leq x_i) - P(X \leq x_{i-1}) \\ &= F_X(x_i) - F_X(x_{i-1}), i = 1, 2, \dots \end{aligned}$$

em que a penúltima igualdade se deve ao fato de que  $x_1 \leq x_2 \leq \dots$ .

□

Como mencionamos, essa proposição nos diz que há uma correspondência  $F_X \leftrightarrow p_X$ . Por causa disso, chamamos de **distribuição de probabilidade** de  $X$  tanto  $F_X$  quanto  $p_X$ . Também segue da proposição que se  $X$  é v.a. discreta com conjunto de valores possíveis  $\{x_1, x_2, \dots\}$  (como na proposição, podemos supor  $x_1 \leq x_2 \leq \dots$  sem perda de generalidade), então a f.d.  $F_X$  é do tipo escada, com saltos em  $x_1, x_2, \dots$ ; o tamanho do salto em  $x_i$  é  $F_X(x_i) - F_X(x_{i-1}) = P(X = x_i)$ . A soma dos saltos é 1, isto é,  $\sum_{i \geq 1} \underbrace{F_X(x_i) - F_X(x_{i-1})}_{P(X=x_i)} = 1$ .

**Exemplo.** Considere dois lançamentos independentes de uma moeda honesta. Seja  $X$  o número de caras obtidas nos dois lançamentos. Temos 0, 1 e 2 como valores possíveis de  $X$ . A f.p. de  $X$ ,  $p_X$ , é

dada por

$$p_X(0) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}, \quad p_X(1) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}, \quad p_X(2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \quad \text{e} \quad p_X(x) = 0, x \in \mathbb{R} \setminus \{0, 1, 2\}.$$

A f.d. de  $X$ ,  $F_X$ , é dada por

$$F_X(x) = P(X \leq x) = P(\emptyset) = 0, x < 0$$

$$F_X(x) = P(X \leq x) = P(X = 0) = 1/4, 0 \leq x < 1$$

$$F_X(x) = P(X \leq x) = P(X = 0) + P(X = 1) = 3/4, 1 \leq x < 2$$

$$F_X(x) = P(X \leq x) = P(X = 0) + P(X = 1) + P(X = 2) = 1, x \geq 2,$$

ou seja,

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1/4, & 0 \leq x < 1 \\ 3/4, & 1 \leq x < 2 \\ 1, & x \geq 2 \end{cases}$$

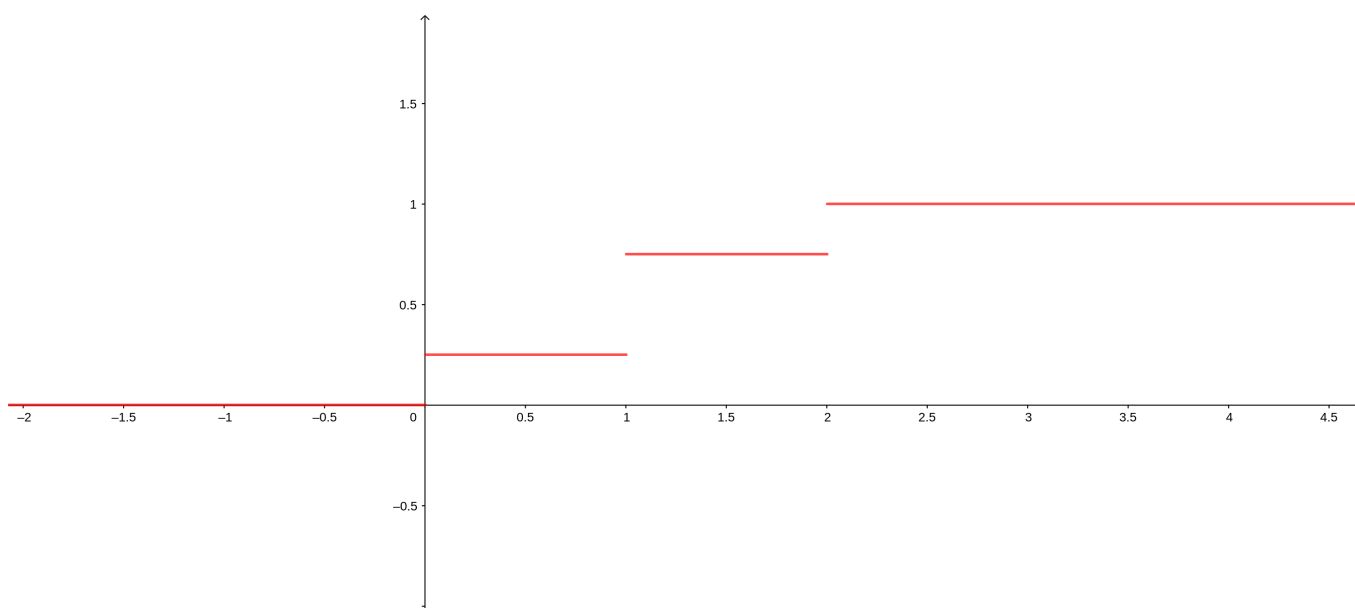


Figura 1: Gráfico da f.d. anterior.

**Exemplo.** Sejam  $(\Omega, \mathcal{A}, P)$  espaço de probabilidade,  $c \in \mathbb{R}$  constante e  $X$  v.a. em  $(\Omega, \mathcal{A}, P)$  tal que  $X(\omega) = c, \forall \omega \in \Omega$  (v.a. constante). Temos, então,

$$p_X(x) = \begin{cases} 1, & x = c \\ 0, & \text{c.c.} \end{cases}$$
$$F_X(x) = \begin{cases} 0, & x < c \\ 1, & x \geq c \end{cases}$$



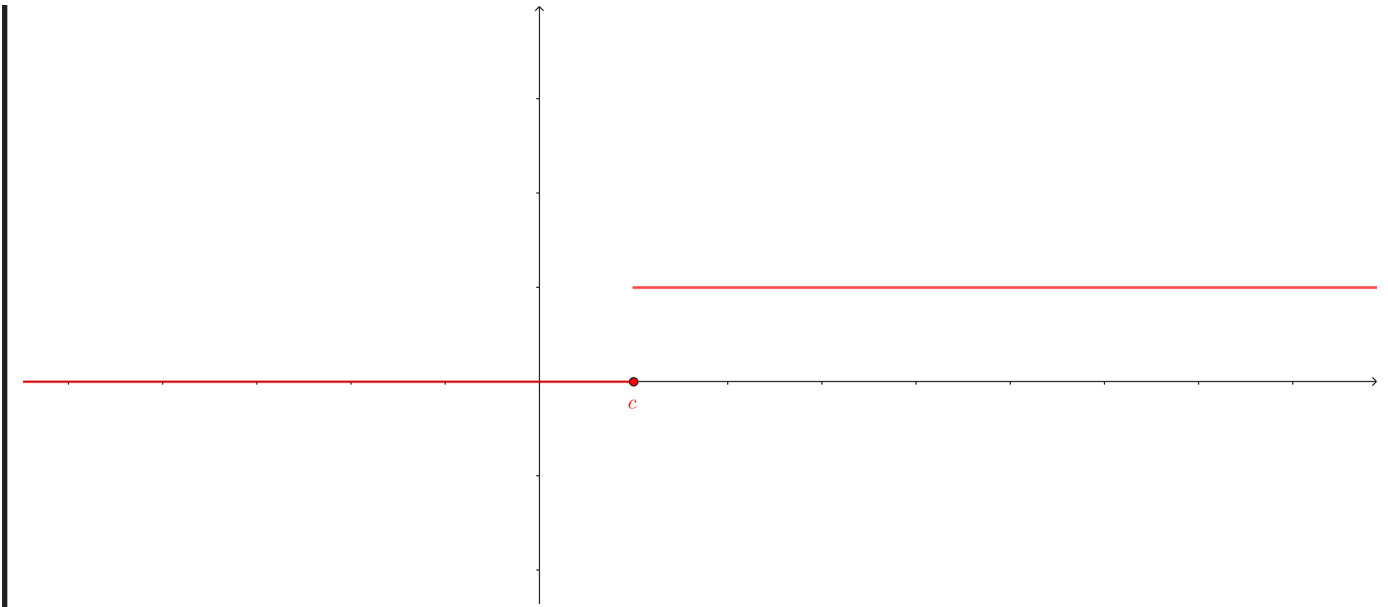


Figura 2: Gráfico da f.d. anterior.

**Exemplo.** Seja  $(\Omega, \mathcal{A}, P)$  espaço de probabilidade e  $X \sim \text{Geom}(p)$ ,  $0 < p < 1$ . Temos

$$p_X(k) = \begin{cases} p(1-p)^{k-1}, & k = 1, 2, \dots \\ 0, & \text{c.c.} \end{cases}.$$

Ademais,

$$\{X \leq x\} = \begin{cases} \emptyset, & x \leq 1 \\ \bigcup_{k=1}^{[x]} \{X = k\}, & x \geq 1 \end{cases}.$$

Logo, como

$$P\left(\bigcup_{k=1}^{[x]} \{X = k\}\right) = \sum_{k=1}^{[x]} P(X = k) = p \sum_{k=1}^{[x]} (1-p)^{k-1} = \frac{p}{1-p} \cdot (1-p) \frac{(1 - (1-p)^{[x]})}{p} = 1 - (1-p)^{[x]},$$

segue que

$$F_X(x) = \begin{cases} 0, & x < 1 \\ 1 - (1-p)^{[x]}, & x \geq 1. \end{cases}$$

**Definição 3.9** (Perda de memória). Uma v.a.  $X$  é dita ter a propriedade de **perda de memória** se

- (i)  $P(X > 0) > 0$ ;
- (ii)  $P(X > a + b | X > a) = P(X > b), \forall a, b \geq 0$ .

**Observação.** Note, em particular, que se  $X$  é v.a. discreta com valores possíveis  $1, 2, \dots$ , então  $X$  tem perda de memória se  $P(X > n + m | X > n) = P(X > m), \forall n, m \in \{1, 2, \dots\}$ . Ademais, se  $X$  é v.a. qualquer, então dados  $a, b \in \mathbb{R}_{>0}$  temos

$$P(X > a + b | X > a) = P(X > b) \Leftrightarrow P(X > a + b) = P(X > a)P(X > b).$$

**Demonstração.** De fato,

$$P(X > a + b | X > a) = P((X > a + b) \cap (X > a)) / P(X > a) = P(X > a + b) / P(X > a).$$

Logo,

$$P(X > a + b | X > a) = P(X > b) \Leftrightarrow P(X > a + b) = P(X > a)P(X > b).$$

□

**Proposição 3.10.** Seja  $X$  v.a. discreta em  $(\Omega, \mathcal{A}, P)$  assumindo valores em  $\{1, 2, \dots\}$ . Então  $X$  tem perda de memória se, e só se,  $X \sim \text{Geom}(p)$ , para algum  $0 < p < 1$ .

**Demonstração.**  $(\Leftrightarrow)$  Se  $X \sim \text{Geom}(p)$ ,  $0 < p < 1$ , então

$$P(X = k) = \begin{cases} p(1-p)^{k-1}, & k = 1, 2, \dots \\ 0, & \text{c.c.} \end{cases}.$$

Sejam  $n, m \in \{1, 2, \dots\}$  quaisquer. Temos

$$\begin{aligned} P(X > n+m | X > n) &= \frac{P(X > n+m)}{P(X > n)} = \frac{\sum_{k=0}^{\infty} p(1-p)^{n+m+k}}{\sum_{k=0}^{\infty} p(1-p)^{n+k}} \\ &= \frac{p(1-p)^{n+m}(1 - (1-p))^{-1}}{p(1-p)^n(1 - (1-p))^{-1}} \\ &= (1-p)^m \\ &= P(X > m). \end{aligned}$$

$(\Rightarrow)$  Suponha que  $X$  tenha perda de memória. Da proposição anterior, temos que

$$P(X > n+m) = P(X > n)P(X > m), \forall n, m \in \{0, 1, 2, \dots\}.$$

Se  $n = 0 = m$ , então

$$P(X > 0) = (P(X > 0))^2 \Rightarrow P(X > 0) = 0 \text{ ou } P(X > 0) = 1.$$

Como  $X$  tem perda de memória, então  $P(X > 0) = 1$ . Se  $P(X = 1) = p$ , então  $P(X > 1) = 1 - p$ . Daí, se  $m \geq 1$ , segue que

$$P(X > m+1) = P(X > m)(1-p).$$

Por indução, temos  $P(X > m) = (1-p)^m, \forall m \in \{0, 1, 2, \dots\}$ . Daí, dado  $n \in \{1, 2, \dots\}$ , temos

$$P(X = n) = P(X > n-1) - P(X > n) = (1-p)^{n-1} - (1-p)^n = p(1-p)^{n-1},$$

ou seja,  $X \sim \text{Geom}(p)$ . □

**Exemplo.** Se  $X \sim \text{Poisson}(\lambda)$ , então

$$p_X(k) = \begin{cases} e^{-\lambda} \frac{\lambda^k}{k!}, & k = 0, 1, 2, \dots \\ 0, & \text{c.c.} \end{cases}$$

e

$$F_X(x) = \begin{cases} 0, & x < 0 \\ e^{-\lambda} \sum_{k=0}^{[x]} \frac{\lambda^k}{k!}, & x \geq 0 \end{cases}.$$

**Exemplo.** Se  $X \sim B(n, p)$ , então

$$p_X(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & k = 0, 1, \dots, n \\ 0, & \text{c.c.} \end{cases}$$

e

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \sum_{k=0}^{[x]} \binom{n}{k} p^k (1-p)^{n-k}, & x \geq 0 \end{cases}.$$

**Proposição 3.11.** Temos que a f.d. satisfaz as seguintes propriedades:

- (i)  $0 \leq F_X(x) \leq 1, \forall x \in \mathbb{R}$ ;
- (ii) dados  $x, y \in \mathbb{R}, x < y$ , então  $F_X(x) \leq F_X(y)$ ;
- (iii)  $\forall a \in \mathbb{R}, F_X(a^+) := \lim_{x \rightarrow a^+} F_X(x) = F_X(a)$ ;
- (iv)  $F_X(-\infty) := \lim_{x \rightarrow -\infty} F_X(x) = 0$  e  $F_X(+\infty) := \lim_{x \rightarrow +\infty} F_X(x) = 1$ .

**Demonstração.**

- (i)  $F_X(x) = P(X \leq x) \Rightarrow 0 \leq F_X(x) \leq 1, \forall x \in \mathbb{R}$ .
- (ii) Dados  $x, y \in \mathbb{R}, x < y$ , temos  $\{X \leq x\} \subseteq \{X \leq y\}$  e, da monotonicidade da medida de probabilidade,  $P(X \leq x) \leq P(X \leq y)$ , i.e.,  $F_X(x) \leq F_X(y)$ .
- (iii) Sejam  $a \in \mathbb{R}$  e  $(x_n)_{n \geq 1}$  uma sequência decrescente com  $x_n \xrightarrow{n \rightarrow +\infty} a$ . Temos  $x_1 \geq x_2 \geq \dots \geq a$ ,

de modo que  $\{X \leq x_1\} \supset \{X \leq x_2\} \supset \dots$ , ou seja,  $(\{X \leq x_n\})_{n \geq 1}$  é uma sequência decrescente de eventos e

$$\lim_{n \rightarrow +\infty} \{X \leq x_n\} = \bigcap_{n=1}^{\infty} \{X \leq x_n\} = \{X \leq a\}.$$

Pela continuidade da probabilidade, temos

$$F_X(a) = P(X \leq a) = \lim_{n \rightarrow +\infty} P(X \leq x_n) = \lim_{n \rightarrow +\infty} F_X(x_n) = \lim_{x \rightarrow a^+} F_X(x) := F_X(a^+).$$

(iv) Seja  $(x_n)_{n \geq 1}$  uma sequência decrescente tal que  $x_n \xrightarrow{n \rightarrow +\infty} -\infty$ . De  $x_1 \geq x_2 \geq \dots$ , temos

$$\lim_{n \rightarrow +\infty} \{X \leq x_n\} = \bigcap_{n=1}^{\infty} \{X \leq x_n\} = \emptyset,$$

donde

$$F_X(-\infty) := \lim_{x \rightarrow -\infty} F_X(x) = \lim_{n \rightarrow +\infty} F_X(x_n) = \lim_{n \rightarrow +\infty} P(X \leq x_n) = \lim_{n \rightarrow +\infty} P(\emptyset) = 0.$$

Seja agora  $(x_n)_{n \geq 1}$  crescente tal que  $x_n \xrightarrow{n \rightarrow +\infty} +\infty$ . De  $x_1 \leq x_2 \leq \dots$ , temos  $\{X \leq x_1\} \subset \{X \leq x_2\} \subset \dots$ , ou seja,

$$\lim_{n \rightarrow +\infty} \{X \leq x_n\} = \bigcup_{n=1}^{\infty} \{X \leq x_n\} = \Omega.$$

Logo,

$$F_X(+\infty) := \lim_{x \rightarrow +\infty} F_X(x) = \lim_{x \rightarrow +\infty} P(X \leq x) = \lim_{n \rightarrow +\infty} P(X \leq x_n) = \lim_{n \rightarrow +\infty} P(\Omega) = 1. \quad \square$$

**Observação.** Seja  $X$  v.a. com f.d.  $F_X$ . A partir da f.d., podemos calcular  $P(X \in A)$ ,  $\forall A \in \mathcal{B}(\mathbb{R})$ :

(a) se  $A = (a, b]$ , então  $\{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\}$ , donde

$$P(X \leq b) = P(X \leq a) + P(a < X \leq b),$$

ou seja,

$$P(X \in A) = P(a < X \leq b) = F_X(b) - F_X(a).$$

(b) se  $A = (-\infty, b)$ , temos  $\{X \leq b\} = \bigcup_{n=1}^{\infty} \{X \leq b - 1/n\}$ , donde

$$\begin{aligned} P(X < b) &= P\left(\bigcup_{n=1}^{\infty} \{X \leq b - 1/n\}\right) = \lim_{n \rightarrow +\infty} P(X \leq b - 1/n) \\ &= \lim_{n \rightarrow +\infty} F_X(b - 1/n) \\ &= \lim_{x \rightarrow b^-} F_X(x) := F_X(b^-). \end{aligned}$$

Logo,  $P(X \in A) = P(X < b) = F_X(b^-) - \lim_{x \rightarrow b^-} F_X(x)$ .

(c) Se  $A = \{a\}$ , temos  $\{X \leq a\} = \{X < a\} \cup \{X = a\}$ , donde

$$P(X \leq a) = P(X < a) + P(X = a) \Leftrightarrow P(X = a) = F_X(a) - F_X(a^-).$$

Raciocínios análogos valem para outros subintervalos da reta.

Note também que como  $F_X$  é contínua à direita e não decrescente, seus pontos de descontinuidade são do tipo salto, com tamanho  $P(X = a) = F_X(a) - F_X(a^-)$  em  $a$ . Por fim, se  $X$  é v.a. contínua, então  $F_X$  é contínua.

**Demonstração.** De fato, dado  $x \in \mathbb{R}$ , temos

$$0 = P(X = x) = F_X(x) - F_X(x^-),$$

logo

$$F_X(x) = F_X(x^-) = F_X(x^+), \forall x \in \mathbb{R}. \quad \square$$

### 3.3 Exercícios - combinatória

1. O código Morse consiste de uma sequência de pontos e traços em que repetições são permitidas.

- a) Quantas letras podem ser codificadas usando exatamente  $n$  símbolos?  
 b) Qual é o número de letras que se pode codificar usando  $n$  ou menos símbolos?

**Solução.**

- (a) Como há duas opções para cada símbolo, temos  $2^n$  letras.  
 (b) Temos

$$\sum_{i=1}^n 2^i = \frac{2(2^n - 1)}{2 - 1} = 2(2^n - 1).$$

□

2. Um homem possui  $n$  chaves das quais, exatamente uma abre a fechadura. Ele experimenta as chaves uma de cada vez, escolhendo ao acaso em cada tentativa uma das chaves que não foram experimentadas. Determine a probabilidade de que ele escolha a chave correta na  $r$ -ésima tentativa?

**Solução.** A probabilidade é

$$\frac{n-1}{n} \cdot \frac{n-2}{n-1} \cdots \frac{n-(r-1)}{n-(r-2)} \cdot \frac{1}{n-(r-1)} = 1/n.$$

□

3. Um ônibus parte com 6 pessoas e para em 10 pontos diferentes. Supondo que os passageiros têm igual probabilidade de saltar em qualquer parada, determine a probabilidade de que dois passageiros não desembarquem na mesma parada.

**Solução.** Temos

$$\Omega = \{(x_1, x_2, \dots, x_6) : 1 \leq x_i \leq 10, i = 1, 2, \dots, 6\}, \forall \mathcal{A} = \mathcal{P}(\Omega)$$

e  $P : \mathcal{A} \rightarrow \mathbb{R}$  dada por  $P(A) = |A|/|\Omega|$ . Sendo  $A$  o evento “não desembarcar na mesma parada”, temos que

$$P(A) = |A|/|\Omega| = A_{10,6}/10^6.$$

□

4. Suponha que temos  $r$  caixas. Bolas são colocadas aleatoriamente nas caixas, uma de cada vez, até que alguma caixa contenha duas bolas pela primeira vez. Determine a probabilidade de que isto ocorra na  $n$ -ésima bola, com  $r \geq n - 1$ .

**Solução.** A probabilidade é  $\frac{(n-1)A_{r,n-1}}{r^n}$ .

□

5. Supondo que se distribui  $n$  bolas em  $n$  caixas.

- a) Qual a probabilidade de que exatamente uma caixa esteja vazia?  
 b) Dado que a caixa 1 está vazia, qual a probabilidade de que somente uma caixa esteja vazia?  
 c) Dado que somente uma caixa está vazia, qual a probabilidade de que a caixa 1 esteja vazia.

**Solução.**

a) A probabilidade é  $\binom{n}{2}n!/n^n$ .

b) Se a caixa 1 está vazia, as outras têm de estar preenchidas. Portanto, a probabilidade é  $\binom{n}{2}(n-1)!/(n-1)^n$ .

c) A probabilidade de uma caixa  $j$  qualquer estar vazia é  $1/n$ .

□

6. Se distribuimos aleatoriamente  $n$  bolas em  $r$  caixas, qual é a probabilidade de que a caixa 1 contenha  $j$  bolas, com  $0 \leq j \leq n$ ?

**Solução.** A probabilidade é  $\binom{n}{j}(r-1)^{n-j}/r^n$ .

□

7. Uma caixa contém  $b$  bolas pretas e  $r$  bolas vermelhas. Bolas são extraídas sem reposição, uma de cada vez. Determine a probabilidade de obter a primeira bola preta na  $n$ -ésima extração.

**Solução.** A probabilidade é

$$\frac{\binom{r}{n-1}}{\binom{r+b}{n-1}} \cdot \frac{b}{r+b-n+1}.$$

□

8. Considere um baralho com 52 cartas. Uma mão de pôquer consiste de 5 cartas extraídas do baralho sem reposição e sem consideração da ordem. Considera-se que constituem sequências as mãos dos seguintes tipos: A, 2, 3, 4, 5; 2, 3, 4, 5, 6; ... ; 10, J, Q, K, A. Determine a probabilidade de ocorrência de cada uma das seguintes mãos de pôquer:

- Royal flush ((10, J, Q, K, A) do mesmo naipe);
- Straight flush (cinco cartas do mesmo naipe em sequência);
- Four (valores da forma  $(x, x, x, x, y)$  onde  $x$  e  $y$  são distintos);
- Full House (valores da forma  $(x, x, x, y, y)$  onde  $x$  e  $y$  são distintos);
- Flush (cinco cartas do mesmo naipe);
- Straight (cinco cartas em sequência, sem consideração de naipes);
- Trinca (valores da forma  $(x, x, x, y, z)$  onde  $x, y$  e  $z$  são distintos);
- Dois pares (valores da forma  $(x, x, y, y, z)$  onde  $x, y$  e  $z$  são distintos);
- Um par (valores da forma  $(w, w, x, y, z)$  onde  $w, x, y$  e  $z$  são distintos).

**Solução.**

- $4/\binom{52}{5}$ .
- $4 \cdot 10/\binom{52}{5} = 4 \cdot 10 \cdot q$ .
- $13 \cdot 4 \cdot 12 \cdot q$ .
- $13 \cdot 4 \cdot 12 \cdot 6 \cdot q$ .
- $4 \cdot \binom{13}{5} \cdot q$ .
- $10 \cdot 4^5 \cdot q$ .
- $13 \cdot \binom{12}{2} \cdot 4^3 \cdot q$ .
- $\binom{13}{2} \cdot \binom{4}{2} \cdot \binom{4}{2} \cdot 44 \cdot q$ .
- $13 \cdot \binom{4}{2} \cdot \binom{12}{3} \cdot 4^3 \cdot q$ .

□

9. Uma caixa contém dez bolas numeradas de 1 a 10. Seleciona-se uma amostra aleatória de 3 elementos. Determine a probabilidade de que as bolas 1 e 6 estejam entre as bolas selecionadas.

**Solução.** Dado que 1 e 6 estão selecionadas, há 8 possíveis bolas que completam o conjunto. Logo, a probabilidade desejada é  $8/\binom{10}{3}$ . □

10. Suponha que se extrai, sem reposição, uma amostra de tamanho  $n$  de uma população de  $r$  elementos. Obtenha a probabilidade de que  $k$  objetos dados estejam incluídos na amostra.

**Solução.** Aqui, podemos imaginar que os  $k$  elementos já foram incluídos: resta então escolher  $n - k$  entre os  $r - k$  vizinhos restantes. A probabilidade procurada é, portanto,  $\binom{r-k}{n-k}/\binom{r}{n}$ . □

11. Qual a probabilidade de que 4 cartas extraídas de um baralho, 2 sejam pretas e 2 vermelhas?

**Solução.** A probabilidade é  $\binom{26}{2}\binom{26}{2}/\binom{52}{4}$ . □

12. Se você possui 3 bilhetes de uma loteria para a qual se vendeu  $n$  bilhetes e existem 5 prêmios, qual a probabilidade de você ganhar pelos menos um prêmio?

**Solução.**  $1 - \frac{\binom{5}{0}\binom{n-5}{3}}{\binom{n}{3}}$ . □

### 3.4 Exercícios - v.a.'s discretas

1. Qualquer ponto no intervalo  $[0, 1)$  pode ser representado por meio de sua expansão decimal  $0, x_1x_2 \dots$ ; suponha que se escolhe, aleatoriamente, um ponto do intervalo  $[0, 1)$ . Seja  $X$  o primeiro dígito da expansão decimal que representa o ponto. Determine a função de probabilidade de  $X$ .

**Solução.** Como  $X$  é uma v.a. uniforme, segue que

$$p_X(x) = \begin{cases} 1/10, & x \in \{0, 1, \dots, 9\} \\ 0, & \text{c.c.} \end{cases} .$$

□

2. a) Se  $X \sim B(n, p)$ , determine a função de probabilidade de  $Y = n - X$ .  
 b) Se  $X \sim \text{Geom}(p)$ , determine a função de probabilidade de  $Y = X - 1$ .  
 c) Se  $X$  tem uma função de probabilidade Binomial Negativa de parâmetros  $r$  e  $p$ , (sendo  $r$  inteiro e  $0 < p < 1$ ), determine a função de probabilidade de  $Y = X - r$ .

**Solução.**

- a) Temos

$$p_X(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & k = 0, 1, \dots, n \\ 0, & \text{c.c.} \end{cases} ,$$

logo

$$p_Y(k) = \begin{cases} \binom{n}{n-k} p^{n-k} (1-p)^k, & k = 0, 1, \dots, n \\ 0, & \text{c.c.} \end{cases} ,$$

ou seja,  $Y \sim B(n, 1-p)$ .

- b) Temos

$$p_Y(k) = \begin{cases} (1-p)^k p, & k = 0, 1, \dots \\ 0, & \text{c.c.} \end{cases} .$$

- c) Temos

$$p_Y(k) = \begin{cases} \binom{r+k-1}{r-1} p^r (1-p)^k, & k = 0, 1, \dots \\ 0, & \text{c.c.} \end{cases} .$$

□

3. Suponha que uma caixa contenha 6 bolas vermelhas e 4 pretas. Selecciona-se uma amostra aleatória de tamanho  $n$ . Seja  $X$  o número de bolas vermelhas na amostra. Determine a função de probabilidade de  $X$  para a amostragem:

- a) sem reposição;  
 b) com reposição.

**Solução.**

- a) Temos  $X \sim H_{\text{geo}}(10, 6, n)$ .  
 b) Temos  $X \sim B(n, 6/10)$ .

□

4. Seja  $N$  um número inteiro positivo e seja

$$p(x) = \begin{cases} c2^x, & x \in \{1, 2, \dots, N\} \\ 0, & \text{c.c.} \end{cases}$$

Determine o valor de  $c$  para o qual  $p$  é uma função de probabilidade.

**Solução.** Devemos ter  $\sum_i p(x_i) = 1$ , ou seja,

$$\sum_{i=1}^N c2^i = 1 \Leftrightarrow 2c(2^N - 1) = 1 \Leftrightarrow c = \frac{1}{2(2^N - 1)} .$$

□

5. Suponha que  $X$  tem uma distribuição geométrica com  $p = 0,8$ . Determine as probabilidades dos seguintes eventos:

- $P(X > 3)$
- $P(4 \leq X \leq 7 \text{ ou } X > 9)$ ;
- $P(3 \leq X \leq 5 \text{ ou } 7 \leq X \leq 10)$ ;

**Solução.**

(a) Temos

$$P(X > 3) = 1 - P(X \leq 3) = 1 - P(X = 1) - P(X = 2) - P(X = 3) = 0,008.$$

(b) Usando a f.d. de  $X$ , temos

$$P(4 \leq X \leq 7 \text{ ou } X > 9) = P(X \leq 7) - P(X < 4) + 1 - P(X \leq 9) = 0,2^3 - 0,2^7 + 0,2^9.$$

(c) Analogamente ao item b, temos que a probabilidade desejada é

$$P(X \leq 5) - P(X < 3) + P(X \leq 10) - P(X < 7) = 0,2^2 - 0,2^5 + 0,2^6 - 0,2^{10}.$$

□

6. Suponha que  $X$  tem uma distribuição uniforme sobre  $0, 1, \dots, 99$ . Determine:

- $P(X \geq 25)$ ;
- $P(2,6 < X < 12,2)$ ;
- $P(8 < X \leq 10 \text{ ou } 30 < X \leq 32)$ ;
- $P(25 \leq X \leq 30)$

**Solução.**

(a) Temos

$$P(X \geq 25) = 1 - P(X < 25) = 1 - P(X \leq 24) = 1 - 25/100 = 3/4.$$

(b) Temos

$$P(2,6 < X < 12,2) = P(3 \leq X \leq 12) = P(X \leq 12) - P(X \leq 2) = 1/10.$$

(c) Temos

$$P(8 < X \leq 10 \text{ ou } 30 < X \leq 32) = P(9 \leq X \leq 10) + P(31 \leq X \leq 32) = 1/25.$$

(d) Temos

$$P(25 \leq X \leq 30) = P(X \leq 30) - P(X \leq 24) = 3/50.$$

□

7. Suponha que o número de chegadas de clientes em um posto de informações turísticas seja uma variável aleatória com distribuição Poisson com taxa de 2 pessoas por hora ( $X \sim \text{Poisson}(2)$ ). Para uma hora qualquer, determine a probabilidade de ocorrer:

- pelo menos uma chegada;
- mais de duas chegadas, dado que chegaram menos de 5 pessoas.

**Solução.**

a) A probabilidade desejada é dada por

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X \leq 0) = 1 - P(X = 0) = 1 - e^{-2}.$$

b) A probabilidade desejada é

$$P(X > 2 | X < 5) = \frac{P(2 < X < 5)}{P(X < 5)} = \frac{P(3 \leq X \leq 4)}{P(X \leq 4)} = \frac{2e^{-2}}{7e^{-2}} = \frac{2}{7}.$$

□

8. Suponha que uma caixa contém 12 bolas numeradas de 1 a 12. Faz-se duas repetições independentes do experimento de selecionar aleatoriamente uma bola da caixa. Seja  $X$  o maior entre os dois números observados. Determine a função de probabilidade de  $X$ .

**Solução.** Temos

$$p_X(k) = \begin{cases} \frac{2k-1}{144}, k = 1, 2, \dots, 12 \\ 0, \text{c.c.} \end{cases} .$$

□

9. Considere a situação do Exercício 8 em que a seleção é feita sem reposição.

- Determine a função de probabilidade de  $X$ ;
- Determine a função de distribuição de  $X$ .

**Solução.**

a) Temos

$$p(k) = \begin{cases} \frac{k-1}{\binom{12}{2}}, k = 1, 2, \dots, 12 \\ 0, \text{c.c.} \end{cases} .$$

b) Temos

$$F_X(x) = \begin{cases} 0, x < 2 \\ \frac{\binom{\lfloor x \rfloor}{2}}{\binom{12}{2}}, 2 \leq x < 12 \\ 1, x \geq 12. \end{cases}$$

□

10. Suponha que uma caixa contenha  $r$  bolas numeradas de 1 a  $r$ . Selecciona-se sem reposição uma amostra aleatória de tamanho  $n$ . Seja  $Y$  o maior número observado na amostra e  $Z$  o menor. Determine:

- $P(Y \leq y), \forall y \in \mathbb{R}$ ;
- $P(Z \geq z), \forall y \in \mathbb{R}$ .

**Solução.**

a) Temos

$$P(Y \leq y) = \begin{cases} 0, y < n \\ \frac{\binom{\lfloor y \rfloor}{n}}{\binom{r}{n}}, n \leq y < r \\ 1, y \geq r. \end{cases}$$

b) Temos

$$P(Z \geq z) = \begin{cases} 1, z < 1 \\ \frac{\binom{r+1-z}{n}}{\binom{r}{n}}, z = 1, 2, \dots, r-n+1 \\ \frac{\binom{r-\lfloor z \rfloor}{n}}{\binom{r}{n}}, i < z < i+1, i = 1, 2, \dots, r-n \\ 0, z > r-n+1 \end{cases} .$$

□

11. Considere  $X$  uma variável aleatória assumindo valores em  $\{0, \pm 1, \pm 2\}$ . Suponha que  $P(X = -2) = P(X = -1)$  e  $P(X = 1) = P(X = 2)$  com a informação que  $P(X > 0) = P(X < 0) = P(X = 0)$ . Encontre a função de probabilidade e a função de distribuição de  $X$ .

**Solução.** Sejam  $P(X = -2) = x = P(X = -1)$ ,  $P(X = 2) = y = P(X = 1)$  e  $P(X = 0) = z$ . De  $P(X > 0) = P(X = 0) = P(X < 0)$ , temos  $2x = 2y = z$ . Daí, como  $2x + 2y + z = 1$ , temos  $x = y = 1/6$  e  $z = 1/3$ . Portanto,

$$p_X(k) = \begin{cases} 1/6, k = \pm 1, \pm 2 \\ 1/3, k = 0 \\ 0, \text{c.c.} \end{cases}$$



e

$$F_X(k) = \begin{cases} 0, & x < -2 \\ 1/6, & -2 \leq x < -1 \\ 1/3, & -1 \leq x < 0 \\ 2/3, & 0 \leq x < 1 \\ 5/6, & 1 \leq x < 2 \\ 1, & x \geq 2 \end{cases} .$$

□

12. Seja  $X$  uma v.a. com função de distribuição dada por:

$$F_X(x) = \begin{cases} 0, & x < 0 \\ x/2, & 0 \leq x < 1 \\ 2/3, & 1 \leq x < 2 \\ 11/12, & 2 \leq x < 3 \\ 1, & x \geq 3 \end{cases} .$$

Determine

- $P(X < 3)$ ;
- $P(X = 1)$ ;
- $P(X > 1/2)$ ;
- $P(2 < X \leq 4)$ ;
- $P(2 \leq X \leq 4)$ .

**Solução.**

- Temos  $P(X < 3) = F(3^-) = 11/12$ .
- Temos  $P(X = 1) = F(1) - F(1^-) = 1/6$ .
- Temos  $P(X > 1/2) = 1 - F(1/2) = 3/4$ .
- Temos  $P(2 < X \leq 4) = F(4) - F(2) = 1/12$ .
- Temos  $P(2 \leq X \leq 4) = F(4) - F(2^-) = 1/3$ .

□

### 3.5 Vetores aleatórios discretos (variáveis aleatórias multidimensionais)

Estamos agora interessados em estudar a relação entre duas ou mais variáveis aleatórias. A título de exemplo, em uma extração de uma amostra de tamanho  $n$  de bolas numeradas de 1 a  $m > n$ , poderíamos estar interessados, simultaneamente, no maior e no menor números retirados, digamos  $X$  e  $Y$ , respectivamente.

**Definição 3.12 (Vetor aleatório).** Sejam  $X_1, X_2, \dots, X_n$  v.a.'s definidas em  $(\Omega, \mathcal{A}, P)$ . O vetor  $\tilde{X} = (X_1, X_2, \dots, X_n)$  é chamado **vetor aleatório** ( $n$ -dimensional) definido em  $(\Omega, \mathcal{A}, P)$ . Note que  $\tilde{X}$  define uma função de  $\Omega \rightarrow \mathbb{R}^n$ , i.e., dado  $\omega \in \Omega$  temos  $\tilde{X}(\omega) = (X_1(\omega), X_2(\omega), \dots, X_n(\omega)) \in \mathbb{R}^n$ .

**Observação.** Se existir  $A = \{\tilde{x}_1, \tilde{x}_2, \dots\} \subset \mathbb{R}^n$  finito ou infinito enumerável tal que  $P(\tilde{X} \in A) = 1$ , ou seja, se  $\tilde{X}$  assume valores em um subconjunto finito ou infinito enumerável de  $\mathbb{R}^n$ , diremos que  $\tilde{X} = (X_1, X_2, \dots, X_n)$  é um **vetor aleatório discreto**.

De maneira geral, um vetor aleatório  $n$ -dimensional é uma função vetorial  $\tilde{X} : \Omega \rightarrow \mathbb{R}^n$  tal que  $\{\tilde{X} \in A\} \in \mathcal{A}, \forall A \in \mathcal{A}$ . Note que um vetor aleatório é uma variável aleatória com contradomínio  $\mathbb{R}^n$ .

**Observação.** Seja  $\tilde{X} = (X_1, X_2, \dots, X_n)$  vetor aleatório. Para  $\tilde{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ , temos

$$\begin{aligned} \{\tilde{X} = \tilde{x}\} &= \{(X_1, X_2, \dots, X_n) = (x_1, x_2, \dots, x_n)\} \\ &= \{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\} \\ &:= \{X_1 = x_1\} \cap \{X_2 = x_2\} \cap \dots \cap \{X_n = x_n\} \\ &= \bigcap_{i=1}^n \underbrace{\{X_i = x_i\}}_{\in \mathcal{A}} \in \mathcal{A}, \end{aligned}$$

ou seja,  $\{\tilde{X} = \tilde{x}\}$  é um evento, de modo que faz sentido calcular

$$P(\tilde{X} = \tilde{x}) = P(X_1 = x_1, \dots, X_n = x_n).$$

**Definição 3.13** (Função de probabilidade conjunta). Se  $\tilde{X} = (X_1, X_2, \dots, X_n)$  é um vetor aleatório discreto definido em  $(\Omega, \mathcal{A}, P)$ , então a função  $p_{\tilde{X}} : \mathbb{R}^n \rightarrow \mathbb{R}$  dada por

$$p_{\tilde{X}}(\tilde{x}) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n), \forall \tilde{x} \in \mathbb{R}^n$$

é chamada **função de probabilidade** de  $\tilde{X}$  ou **função de probabilidade conjunta** de  $X_1, X_2, \dots, X_n$ .

**Observação.** É comum usar a seguinte notação

$$\begin{aligned} p_{\tilde{X}}(\tilde{x}) &= P(\tilde{X} = \tilde{x}) = P((X_1, X_2, \dots, X_n) = (x_1, x_2, \dots, x_n)) \\ &= P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \\ &= p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) \\ &= p_{X_1, X_2, \dots, X_n}(\tilde{x}) \end{aligned}$$

e, se  $\tilde{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$  é tal que  $X_1, X_2, \dots, X_n(x_1, x_2, \dots, x_n) > 0$ , então  $\tilde{x}$  é um **valor possível** de  $\tilde{X} = (X_1, X_2, \dots, X_n)$ . Daí, se  $\tilde{X}$  é discreto, então existe  $\{\tilde{x}_1, \tilde{x}_2, \dots\} \subset \mathbb{R}^n$  finito ou infinito enumerável de valores possíveis de  $\tilde{X}$ . Nesse caso,

- (i)  $p_{\tilde{X}}(\tilde{x}_i) > 0, \forall \tilde{x}_i \in \{\tilde{x}_1, \tilde{x}_2, \dots\}$
- (ii)  $\sum_{i=1}^{\infty} p_{\tilde{X}}(\tilde{x}_i) = 1$  se  $\{\tilde{x}_1, \tilde{x}_2, \dots\}$  é infinito enumerável e  $\sum_{i=1}^n p_{\tilde{X}}(\tilde{x}_i) = 1$  caso contrário.
- (iii)  $P(\tilde{X} \in A) = \sum_{i: \tilde{x}_i \in A} p_{\tilde{X}}(\tilde{x}_i), \forall A \in \mathcal{B}(\mathbb{R}^n)$ .

Assim como no caso unidimensional, podemos tomar a seguinte definição.

**Definição 3.14** (Função de probabilidade  $n$ -dimensional). Toda função  $p : \mathbb{R}^n \rightarrow \mathbb{R}$  tal que

- (i)  $p(\tilde{x}) \geq 0, \forall \tilde{x} \in \mathbb{R}^n$ ;
- (ii)  $\{\tilde{x} \in \mathbb{R}^n : p(\tilde{x}) > 0\}$  é finito ou infinito enumerável;
- (iii)  $\sum_{i: p(\tilde{x}_i) > 0} p(\tilde{x}_i) = 1$

é chamada **função de probabilidade  $n$ -dimensional**.

Analogamente ao caso unidimensional, pode-se provar que dada uma função de probabilidade  $n$ -dimensional  $p$ , existe um espaço de probabilidade  $(\Omega, \mathcal{A}, P)$  e um vetor aleatório  $n$ -dimensional  $\tilde{X} = (X_1, X_2, \dots, X_n)$  tal que

$$p_{\tilde{X}}(\tilde{x}) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = p(x_1, x_2, \dots, x_n), \forall \tilde{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n, \text{ i.e., } p_{\tilde{X}} = p.$$

**Definição 3.15** (Função de distribuição conjunta). Dado um vetor aleatório  $\tilde{X} = (X_1, X_2, \dots, X_n)$  definido em  $(\Omega, \mathcal{A}, P)$ , a função  $F_{\tilde{X}} : \mathbb{R}^n \rightarrow \mathbb{R}$  dada por

$$F_{\tilde{X}}(\tilde{x}) = F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n), \tilde{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n,$$

é chamada **função de distribuição** de  $\tilde{X}$  ou **função de distribuição conjunta** de  $X_1, X_2, \dots, X_n$ .

**Observação.** Note, primeiro, que

$$P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) = P(\{X_1 \leq x_1\} \cap \dots \cap \{X_n \leq x_n\}).$$

Ademais, se  $\tilde{X} = (X_1, X_2, \dots, X_n)$  é um vetor aleatório discreto com função de probabilidade conjunta  $p_{X_1, X_2, \dots, X_n}$ , então

$$\begin{aligned} F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) &= P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) \\ &= \sum_{k_1 \leq x_1} \sum_{k_2 \leq x_2} \dots \sum_{k_n \leq x_n} P(X_1 = k_1, X_2 = k_2, \dots, X_n = k_n) \\ &= \sum_{k_1 \leq x_1} \sum_{k_2 \leq x_2} \dots \sum_{k_n \leq x_n} p_{X_1, X_2, \dots, X_n}(k_1, k_2, \dots, k_n). \end{aligned}$$

Como anteriormente, existe uma correspondência biunívoca entre a função de distribuição conjunta e a função de probabilidade conjunta. Por isso, chamamos ambas de **distribuição** de  $\tilde{X}$  ou **distribuição conjunta** de  $X_1, X_2, \dots, X_n$ .

**Exemplo.** Seja  $\alpha \in (0, 1)$  e considere  $p : \mathbb{R}^2 \rightarrow \mathbb{R}$  dada por

$$p(x, y) = \begin{cases} \alpha^2(1 - \alpha)^{x+y}, & x, y \in \{0, 1, 2, \dots\} \\ 0, & \text{c.c.} \end{cases}$$

Note que

- (i)  $p(x, y) \geq 0, \forall (x, y) \in \mathbb{R}^2$
- (ii)  $\{(x, y) \in \mathbb{R}^2 : p(x, y) > 0\} = \{(x, y) \in \mathbb{R}^2 : x, y \in \{0, 1, 2, \dots\}\}$  é enumerável
- (iii)  $\sum_{x=0}^{\infty} \sum_{y=0}^{\infty} p(x, y) = \alpha^2 \sum_{x=0}^{\infty} (1 - \alpha)^x \sum_{y=0}^{\infty} (1 - \alpha)^y = \alpha^2 \frac{1}{[1 - (1 - \alpha)]^2} = 1.$

Logo,  $p$  é uma função de probabilidade bidimensional. Seja  $(X, Y)$  um vetor aleatório com função de probabilidade  $p_{X, Y} = p$  acima. Note que  $(X, Y)$  assume valores em  $\{(x, y) \in \mathbb{R}^2 : x, y \in \{0, 1, 2, \dots\}\}$ . Daí, dado  $(x, y) \in \mathbb{R}^2$  com  $x, y \geq 0$ , temos

$$\begin{aligned} F_{X, Y} &= \sum_{i \leq x} \sum_{j \leq y} p(i, j) = \sum_{i=0}^{\lfloor x \rfloor} \sum_{j=0}^{\lfloor y \rfloor} \alpha^2(1 - \alpha)^{i+j} \\ &= \alpha^2 \cdot \frac{1 - (1 - \alpha)^{\lfloor x \rfloor + 1}}{1 - (1 - \alpha)} \cdot \frac{1 - (1 - \alpha)^{\lfloor y \rfloor + 1}}{1 - (1 - \alpha)} \\ &= [1 - (1 - \alpha)^{\lfloor x \rfloor + 1}][1 - (1 - \alpha)^{\lfloor y \rfloor + 1}]. \end{aligned}$$

Logo,

$$F_{X, Y}(x, y) = \begin{cases} [1 - (1 - \alpha)^{\lfloor x \rfloor + 1}][1 - (1 - \alpha)^{\lfloor y \rfloor + 1}], & x, y \geq 0 \\ 0, & \text{c.c.} \end{cases}.$$

Ademais, da distribuição conjunta de  $X$  e  $Y$ , podemos obter a distribuição de  $X$  e de  $Y$ : para  $k \in \{0, 1, 2, \dots\}$ , temos

$$\begin{aligned} P(X = k) &= P(X = k, Y \in \mathbb{R}) \\ &= P\left(X = k, \bigcup_{j=0}^{\infty} \{Y = j\}\right) \\ &= \sum_{j=0}^{\infty} P(X = k, Y = j) \\ &= \sum_{j=0}^{\infty} \alpha^2(1 - \alpha)^{k+j} \\ &= \alpha^2(1 - \alpha)^k \sum_{j=0}^{\infty} (1 - \alpha)^j \\ &= \alpha(1 - \alpha)^k, \end{aligned}$$

de modo que

$$p_X(k) = \begin{cases} \alpha(1-\alpha)^k, & k = 0, 1, 2, \dots \\ 0, & \text{c.c.} \end{cases},$$

ou seja,  $X \sim \text{Geom}(\alpha)$ . De maneira análoga, pode-se mostrar que  $Y \sim \text{Geom}(\alpha)$ .

## Distribuições marginais

As funções  $p_{X_i}$  e  $F_{X_i}$  são chamadas **função de probabilidade marginal** de  $X_i$  e **função de distribuição marginal** de  $X_i$ , respectivamente.

**Caso bidimensional.** Seja  $(X, Y)$  vetor aleatório discreto com função de probabilidade conjunta  $p_{X,Y}$  e função de distribuição conjunta  $F_{X,Y}$ . Assuma que os valores possíveis de  $X$  e  $Y$  são  $\{x_1, x_2, \dots\}$  e  $\{y_1, y_2, \dots\}$ , respectivamente. Vamos obter as distribuições marginais de  $X$  e  $Y$  a partir da distribuição conjunta.

(F.P.M.) Temos

$$\begin{aligned} p_X(x) &= P(X = x, Y \in \mathbb{R}) = P\left(X = x, \bigcup_{j=1}^{\infty} \{Y = y_j\}\right) \\ &= \sum_{j=1}^{\infty} P(X = x, Y = y_j) \\ &= \sum_{j=1}^{\infty} p_{X,Y}(x, y_j). \end{aligned}$$

Denotamos

$$p_X(x) = \sum_{j=1}^{\infty} p_{X,Y}(x, y_j) = \sum_y p_{X,Y}(x, y)$$

. De maneira análoga, temos também

$$p_Y(y) = \sum_x p_{X,Y}(x, y)$$

.

(F.D.M.) Temos

$$\begin{aligned} F_X(x) &= P(X \leq x, Y \in \mathbb{R}) = \sum_{j=1}^{\infty} P(X \leq x, Y = y_j) \\ &= \sum_{j=1}^{\infty} P\left(\bigcup_{i: x_i \leq x} \{X = x_i\}, Y = y_j\right) \\ &= \sum_{j=1}^{\infty} \sum_{i: x_i \leq x} P(X = x_i, Y = y_j) \\ &= \sum_{j=1}^{\infty} \sum_{i: x_i \leq x} p_{X,Y}(x_i, y_j) \\ &= \sum_{a \leq x} \sum_b p_{X,Y}(a, b). \end{aligned}$$

$$\text{Analogamente, } F_Y(y) = \sum_{i=1}^{\infty} \sum_{j: y_j \leq y} p_{X,Y}(x_i, y_j) = \sum_a \sum_{b \leq y} p_{X,Y}(a, b).$$

As funções de distribuição marginais  $F_X$  e  $F_Y$  podem também ser obtidas diretamente da função de distribuição conjunta,  $F_{X,Y}$ : considere  $(z_n)$  uma sequência crescente de números reais tal que  $z_n \xrightarrow{n \rightarrow +\infty} +\infty$ . Note que

$$\{X \leq x\} = \{X \leq x, Y \in \mathbb{R}\} = \bigcup_{n=1}^{\infty} \{X \leq x, Y \leq z_n\} = \lim_{n \rightarrow +\infty} \{X \leq x, Y \leq z_n\}.$$

Da continuidade da probabilidade, segue que

$$F_X(x) = \lim_{n \rightarrow +\infty} P(X \leq x, Y \leq z_n) = \lim_{n \rightarrow +\infty} F_{X,Y}(x, z_n) = \lim_{y \rightarrow +\infty} F_{X,Y}(x, y).$$

Denota-se  $F_X(x) = \lim_{y \rightarrow +\infty} F_{X,Y}(x,y) := F_{X,Y}(x, +\infty)$  e, analogamente,  $F_Y(y) = F_{X,Y}(+\infty, y)$ .

**Definição 3.16** (Distribuição marginal). Dado um vetor aleatório  $(X_1, X_2, \dots, X_n)$ , suas **distribuições marginais** são as distribuições de quaisquer  $(X_{i_1}, X_{i_2}, \dots, X_{i_k})$ , com  $i_1, i_2, \dots, i_k \in \{1, 2, \dots, n\}$  e  $k \leq n$ .

**Caso geral.** Seja  $X = (X_1, X_2, \dots, X_n)$  um vetor aleatório discreto com f.p. conjunta  $p_{X_1, X_2, \dots, X_n}$  f.d. conjunta  $F_{X_1, X_2, \dots, X_n}$ . Alguns exemplos de f.p.m. são

$$\begin{aligned} p_{X_1, X_2}(x_1, x_2) &= \sum_{x_3} \sum_{x_4} \cdots \sum_{x_n} p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) \\ p_{X_1}(x_1) &= \sum_{x_2} \sum_{x_3} \cdots \sum_{x_n} p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) \\ p_{X_2, X_3, X_5}(x_2, x_3, x_5) &= \sum_{x_1} \sum_{x_4} \sum_{x_6} \sum_{x_7} \cdots \sum_{x_n} p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n). \end{aligned}$$

Alguns exemplos de f.d.m. são

$$F_{X_1}(x_1) = F_{X_1, X_2, \dots, X_n}(x_1, +\infty, \dots, +\infty) = \sum_{a_1 \leq x_1} \sum_{x_2} \cdots \sum_{x_n} p_{X_1, X_2, \dots, X_n}(a_1, x_2, \dots, x_n)$$

$$F_{X_1, X_2}(x_1, x_2) = F_{X_1, X_2, \dots, X_n}(x_1, x_2, +\infty, \dots, +\infty) = \sum_{a_1 \leq x_1} \sum_{a_2 \leq x_2} \sum_{x_3} \cdots \sum_{x_n} p_{X_1, X_2, \dots, X_n}(a_1, a_2, x_3, \dots, x_n).$$

**Exemplo.** Considere uma caixa com 3 bolas numeradas de 1 a 3. Duas extrações são realizadas sem reposição. Assim, o espaço amostral é

$$\Omega = \{(1, 2), (1, 3), (2, 1), (2, 3), (3, 1), (3, 2)\}.$$

Sejam  $X$  e  $Y$  os números da primeira e da segunda bolas retiradas, respectivamente. Suponha que a função de probabilidade conjunta de  $X$  e  $Y$  seja dada por

$$\begin{aligned} p_{X,Y}(1, 2) &= \frac{3}{8}, p_{X,Y}(1, 3) = \frac{1}{8}, p_{X,Y}(2, 1) = \frac{1}{8}, \\ p_{X,Y}(2, 3) &= \frac{1}{24}, p_{X,Y}(3, 1) = \frac{5}{24}, p_{X,Y}(3, 2) = \frac{1}{8} \text{ e } p_{X,Y}(x, y) = 0, \text{ c.c.} \end{aligned}$$

Podemos apresentar a f.p. conjunta com a seguinte tabela.

$Y \setminus X$	1	2	3	$P(X = x)$
1	0	3/8	1/8	1/2
2	1/8	0	1/24	1/6
3	5/24	1/8	0	1/3
$P(Y = y)$	1/3	1/2	1/6	1

**Exemplo.** Uma urna contém 15 bolas: 5 vermelhas, 4 pretas e 6 brancas. São retiradas, aleatoriamente, 4 bolas. Considere as v.a.'s  $X_1, X_2$  e  $X_3$ , que representam, respectivamente, o número de bolas vermelhas, pretas e brancas retiradas.

**Sem reposição.** A f.p. conjunta de  $X_1, X_2$  e  $X_3$  é dada por

$$p_{X_1, X_2, X_3}(x_1, x_2, x_3) = \begin{cases} \frac{\binom{5}{x_1} \binom{4}{x_2} \binom{6}{x_3}}{\binom{15}{4}}, & 0 \leq x_i \leq 4, i = 1, 2, 3, x_1 + x_2 + x_3 = 4 \\ 0, & \text{c.c.} \end{cases}$$

Com ela, podemos calcular a probabilidade de se retirar exatamente 2 bolas vermelhas de uma maneira alternativa: enquanto o cálculo direto nos dá

$$P(X_1 = 2) = \frac{\binom{5}{2} \binom{10}{2}}{\binom{15}{4}},$$

a distribuição conjunta nos dá

$$\begin{aligned}
 P(X_1 = 2) &= P(X_1 = 2, X_2 \in \mathbb{R}, X_3 \in \mathbb{R}) = \sum_{0 \leq x_2, x_3 \leq 4, x_2 + x_3 = 2} P(X_1 = 2, X_2 = x_2, X_3 = x_3) \\
 &= \sum_{x_2=0}^2 P(X_1 = 2, X_2 = x_2, X_3 = 2 - x_2) \\
 &= \sum_{x_2=0}^2 \frac{\binom{5}{2} \binom{4}{x_2} \binom{6}{2-x_2}}{\binom{15}{4}} \\
 &= \frac{\binom{5}{2}}{\binom{15}{4}} \sum_{x_2=0}^2 \binom{4}{x_2} \binom{6}{2-x_2} \\
 &= \frac{\binom{5}{2} \binom{10}{2}}{\binom{15}{4}}.
 \end{aligned}$$

De maneira geral,

$$P(X_1 = x_1) = \frac{\binom{5}{x_1} \binom{10}{4-x_1}}{\binom{15}{4}}, x_1 \in \{0, 1, 2, 3, 4\},$$

ou seja,  $X \sim \text{Hgeo}(15, 5, 4)$ . Analogamente,  $X_2 \sim \text{Hgeo}(15, 4, 4)$  e  $X_3 \sim \text{Hgeo}(15, 6, 4)$ .

**Com reposição.** A f.p. conjunta de  $X_1, X_2$  e  $X_3$  é dada por

$$p_{X_1, X_2, X_3}(x_1, x_2, x_3) = \begin{cases} \binom{4}{x_1, x_2, x_3} \left(\frac{5}{15}\right)^{x_1} \left(\frac{4}{15}\right)^{x_2} \left(\frac{6}{15}\right)^{x_3}, & 0 \leq x_i \leq 4, i = 1, 2, 3, x_1 + x_2 + x_3 = 4 \\ 0, & \text{c.c.} \end{cases}$$

Fazendo  $p_1 = P(\text{tirar uma bola vermelha}) = 5/15$ ,  $p_2 = 4/15$  e  $p_3 = 6/15$ , temos

$$p_{X_1, X_2, X_3}(x_1, x_2, x_3) = \begin{cases} \binom{4}{x_1, x_2, x_3} p_1^{x_1} p_2^{x_2} p_3^{x_3}, & 0 \leq x_i \leq 4, i = 1, 2, 3, x_1 + x_2 + x_3 = 4 \\ 0, & \text{c.c.} \end{cases}$$

Para calcular a mesma probabilidade anterior, o cálculo direto nos fornece

$$P(X_1 = 2) = \binom{4}{2} \left(\frac{5}{15}\right)^2 \left(\frac{10}{15}\right)^2,$$

e, usando a distribuição conjunta, temos

$$\begin{aligned}
 P(X_1 = 2) &= P(X_1 = 2, X_2 \in \mathbb{R}, X_3 \in \mathbb{R}) \\
 &= \sum_{0 \leq x_2, x_3 \leq 4, x_2 + x_3 = 2} P(X_1 = 2, X_2 = x_2, X_3 = x_3) \\
 &= \sum_{x_2}^2 \binom{4}{2, x_2, 2-x_2} \left(\frac{5}{15}\right)^2 \left(\frac{4}{15}\right)^{x_2} \left(\frac{6}{15}\right)^{2-x_2} \\
 &= \frac{4!}{2!2!} \left(\frac{5}{15}\right)^2 \sum_{x_2=0}^2 \frac{2!}{x_2!(2-x_2)!} \left(\frac{4}{15}\right)^{x_2} \left(\frac{6}{15}\right)^{2-x_2} \\
 &= \binom{4}{2} \left(\frac{5}{15}\right)^2 \left(\frac{10}{15}\right)^2.
 \end{aligned}$$

De maneira geral,

$$P(X_1 = x_1) = \binom{4}{x_1} p_1^{x_1} (1 - p_1)^{1-x_1}, x_1 \in \{0, 1, 2, 3, 4\},$$

ou seja,  $X_1 \sim B(4, p_1)$ . De maneira inteiramente análoga,  $X_2 \sim B(4, p_2)$  e  $X_3 \sim B(4, p_3)$ .

**Exemplo (Distribuição multinomial).** Sejam  $\mathcal{E}$  um experimento aleatório com  $r$  valores possíveis e  $\{Y = i\} = \{\text{o experimento produz o } i\text{-ésimo resultado}\}$ ,  $i = 1, 2, \dots, r$ . Defina  $p_i = P(Y = i)$ . Considere  $n$  repetições do experimento e seja  $(Y_1, Y_2, \dots, Y_n)$  vetor aleatório tal que  $Y_j$  seja o resultado da  $j$ -ésima prova. Defina também  $X_i$  como o número de provas que produzem o resultado  $i$ . Temos  $\{X_i = x_i\} = \{\text{exatamente } x_i \text{ das } n \text{ v.a.'s } Y_j \text{ assume o valor } i\}$ . Note que  $(X_1, X_2, \dots, X_r)$

é um vetor aleatório assumindo valores em  $\mathbb{R}^n$  da forma  $(x_1, x_2, \dots, x_r)$  com  $x_i \in \{0, 1, \dots, n\}$  e  $x_1 + \dots + x_r = n$ . A f.p. conjunta de  $X_1, \dots, X_r$  é dada por

$$p_{X_1, \dots, X_r}(x_1, x_2, \dots, x_r) = \begin{cases} \binom{n}{x_1, x_2, \dots, x_r} p_1^{x_1} p_2^{x_2} \cdots p_r^{x_r}, & x_i \in \{0, 1, \dots, n\}, i = 1, 2, \dots, r, x_1 + \dots + x_r = n \\ 0, & \text{c.c.} \end{cases}$$

Dizemos que  $(X_1, \dots, X_r)$  tem **distribuição multinomial de parâmetros**  $n, p_1, \dots, p_r$  e denotamos  $(X_1, \dots, X_r) \sim \text{Multinomial}(n, p_1, \dots, p_r)$ .

De maneira mais concreta, se  $r = 3$  e  $p_1 + p_2 + p_3 = 1$ , então

$$\begin{aligned} P(X_1 = x_1) &= \sum_{0 \leq x_2, x_3 \leq n, x_1 + x_2 + x_3 = n} P(X_1 = x_1, X_2 = x_2, X_3 = x_3) \\ &= \sum_{x_2=0}^{n-x_1} \binom{n}{x_1, x_2, n-x_1-x_2} p_1^{x_1} p_2^{x_2} p_3^{n-x_1-x_2} \\ &= \binom{n}{x_1} p_1^{x_1} \sum_{x_2=0}^{n-x_1} \frac{(x_1-x_2)!}{x_2!(n-x_1-x_2)!} p_2^{x_2} p_3^{n-x_1-x_2} \\ &= \binom{n}{x_1} p_1^{x_1} (1-p_1)^{n-x_1}, \quad x_1 \in \{0, 1, \dots, n\}, \end{aligned}$$

ou seja,  $X_1 \sim B(n, p_1)$ . Analogamente,  $X_2 \sim B(n, p_2)$  e  $X_3 \sim B(n, p_3)$ . De maneira geral, se  $(X_1, X_2, \dots, X_n) \sim \text{Multinomial}(n, p_1, \dots, p_r)$ , então  $X_i \sim B(n, p_i), i = 1, 2, \dots, r$ .

### 3.6 Variáveis aleatórias independentes

**Definição 3.17** (V.a.'s independentes). Sejam  $X_1, X_2, \dots, X_n$  v.a.'s definidas em  $(\Omega, \mathcal{A}, P)$ . Dizemos que elas são **independentes** se

$$P(X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n) = \prod_{i=1}^n P(X_i \in A_i), \quad \forall A_1, A_2, \dots, A_n \in \mathcal{A}.$$

**Observação.** Segue da definição que  $X_1, X_2, \dots, X_n$  são independentes se, e somente se, os eventos  $\{X_1 \in A_1\}, \{X_2 \in A_2\}, \dots, \{X_n \in A_n\}$  o são.

**Proposição 3.18.** Sejam  $X_1, X_2, \dots, X_n$  v.a.'s discretas definidas em  $(\Omega, \mathcal{A}, P)$ . Então, temos

(a)  $X_1, X_2, \dots, X_n$  são independentes se, e só se,

$$p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p_{X_i}(x_i), \quad \forall x_1, x_2, \dots, x_n \in \mathbb{R},$$

(b)  $X_1, X_2, \dots, X_n$  são independentes se, e só se,

$$F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = \prod_{i=1}^n F_{X_i}(x_i), \quad \forall x_1, x_2, \dots, x_n \in \mathbb{R}.$$

**Demonstração.**

(a)  $(\Rightarrow)$  Se  $X_1, X_2, \dots, X_n$  são independentes, então

$$P(X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n) = \prod_{i=1}^n P(X_i \in A_i), \quad \forall A_1, A_2, \dots, A_n \in \mathcal{A}.$$

Em particular, se  $A_i = \{x_i\}, i = 1, 2, \dots, n$  então

$$\begin{aligned} p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) &= P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \\ &= P(X_1 = x_1) \cdots P(X_n = x_n) \\ &= p_{X_1}(x_1) \cdots p_{X_n}(x_n). \end{aligned}$$

( $\Leftarrow$ ) Dados  $A_1, A_2, \dots, A_n \in \mathcal{A}$ , temos

$$\begin{aligned} P(X_1 \in A_1, \dots, X_n \in A_n) &= P\left(\bigcup_{x_1 \in A_1} \{X_1 = x_1\}, \dots, \bigcup_{x_n \in A_n} \{X_n = x_n\}\right) \\ &= \sum_{x_1 \in A_1} \cdots \sum_{x_n \in A_n} P(X_1 = x_1, \dots, X_n = x_n) \\ &= \sum_{x_1 \in A_1} \cdots \sum_{x_n \in A_n} P(X_1 = x_1) \cdots P(X_n = x_n) \\ &= \sum_{x_1 \in A_1} P(X_1 = x_1) \cdots \sum_{x_n \in A_n} P(X_n = x_n) \\ &= P(X_1 \in A_1) \cdots P(X_n \in A_n). \end{aligned}$$

(b) ( $\Rightarrow$ ) Se  $X_1, \dots, X_n$  são independentes, então

$$P(X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n) = \prod_{i=1}^n P(X_i \in A_i), \forall A_1, A_2, \dots, A_n \in \mathcal{A}.$$

Em particular, se  $A_i = (-\infty, x_i), i = 1, 2, \dots, n$  temos

$$\begin{aligned} F_{X_1, \dots, X_n}(x_1, \dots, x_n) &= P(X_1 \in A_1, \dots, X_n \in A_n) \\ &= P(X_1 \leq x_1) \cdots P(X_n \leq x_n) \\ &= F_{X_1}(x_1) \cdots F_{X_n}(x_n). \end{aligned}$$

( $\Leftarrow$ ) A volta envolve técnicas de Teoria da Medida, e não será demonstrada. □

**Exemplo.** Sejam  $X, Y$  v.a.'s com f.p. conjunta

$$p_{X,Y}(x, y) = \begin{cases} p^2(1-p)^{x+y}, & x, y \in \{0, 1, \dots\} \\ 0, & \text{c.c.} \end{cases},$$

com  $0 < p < 1$ . Vimos que

$$\begin{aligned} p_X(x) &= \begin{cases} p(1-p)^x, & x \in \{0, 1, \dots\} \\ 0, & \text{c.c.} \end{cases} \\ p_Y(y) &= \begin{cases} p(1-p)^y, & y \in \{0, 1, \dots\} \\ 0, & \text{c.c.} \end{cases}, \end{aligned}$$

de modo que  $p_{X,Y}(x, y) = p_X(x)p_Y(y), \forall (x, y) \in \mathbb{R}^2$ , i.e.,  $X$  e  $Y$  são independentes.

**Exemplo.** Uma urna contém 15 bolas: 5 vermelhas, 4 pretas e 6 brancas. São retiradas aleatoriamente 4 bolas. Considere as v.a.'s  $X_1, X_2$  e  $X_3$ , que representam o número de bolas vermelhas, pretas e brancas retiradas, respectivamente.

**Sem reposição.** Vimos que

$$p_{X_1, X_2, X_3} = \begin{cases} \frac{\binom{5}{x_1} \binom{4}{x_2} \binom{6}{x_3}}{\binom{15}{4}}, & 0 \leq x_i \leq 4, i = 1, 2, 3, x_1 + x_2 + x_3 = 4 \\ 0, & \text{c.c.} \end{cases}$$

e que

$$p_{X_1}(x_1) = \frac{\binom{5}{x_1} \binom{10}{4-x_1}}{\binom{15}{4}}, p_{X_2}(x_2) = \frac{\binom{4}{x_2} \binom{11}{4-x_2}}{\binom{15}{4}}, p_{X_3}(x_3) = \frac{\binom{6}{x_3} \binom{9}{4-x_3}}{\binom{15}{4}}, x_i \in \{0, 1, 2, 3, 4\}, i = 1, 2, 3.$$

Note que existem  $x_1, x_2, x_3 \in \{0, 1, 2, 3, 4\}$  tais que

$$p_{X_1, X_2, X_3}(x_1, x_2, x_3) \neq p_{X_1}(x_1)p_{X_2}(x_2)p_{X_3}(x_3),$$

ou seja,  $X_1, X_2$  e  $X_3$  não são independentes.



**Com reposição.** Vimos que

$$p_{X_1, X_2, X_3}(x_1, x_2, x_3) = \begin{cases} \binom{4}{x_1, x_2, x_3} p_1^{x_1} p_2^{x_2} p_3^{x_3}, & 0 \leq x_i \leq 4, i = 1, 2, 3, x_1 + x_2 + x_3 = 4 \\ 0, & \text{c.c.} \end{cases},$$

com  $p_1 = 5/15, p_2 = 4/15$  e  $p_3 = 6/15$ . Além disso,

$$p_{X_1}(x_1) = \binom{4}{x_1} p_1^{x_1} (1 - p_1)^{1-x_1}, p_{X_2}(x_2) = \binom{4}{x_2} p_2^{x_2} (1 - p_2)^{1-x_2}, p_{X_3}(x_3) = \binom{4}{x_3} p_3^{x_3} (1 - p_3)^{1-x_3}.$$

Logo,  $X_1, X_2$  e  $X_3$  não são independentes.

### 3.7 Funções de variáveis aleatórias

Sejam  $X_1, X_2, \dots, X_n$  v.a.'s definidas em  $(\Omega, \mathcal{A}, P)$  e  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  uma função. É possível mostrar que  $Z = g(X_1, \dots, X_n)$  é v.a. em  $(\Omega, \mathcal{A}, P)$ .

**Exemplo.**  $Z = X^2$ , ou seja,  $Z = g(X)$  com  $X$  v.a. e  $g : \mathbb{R} \rightarrow \mathbb{R}$  tal que  $x \mapsto x^2$ .

**Exemplo.**  $W = X + Y + Z$ , ou seja,  $W = h(X, Y, Z)$  com  $X, Y, Z$  v.a.'s. e  $h : \mathbb{R}^3 \rightarrow \mathbb{R}$  tal que  $(x, y, z) \mapsto x + y + z$ .

**Exemplo.**  $Z = \max\{X, Y\}$ , ou seja,  $Z = l(X, Y)$  com  $X, Y$  v.a.'s. e  $l : \mathbb{R}^2 \rightarrow \mathbb{R}$  tal que  $(x, y) \mapsto \max\{x, y\}$ .

**Observação.** Quando as v.a.'s  $X_1, \dots, X_n$  são independentes e têm a mesma distribuição, dizemos que elas são **independentes e identicamente distribuídas** (i.i.d.).

**Exemplo.** Considere  $X_1, \dots, X_n$  v.a.'s i.i.d. com f.d.  $F$ .

(a) Seja  $Z = \max\{X_1, \dots, X_n\}$ . Note que dado  $z \in \mathbb{R}$ ,

$$\max\{x_1, \dots, x_n\} \leq z \Leftrightarrow x_i \leq z, x_i \in \mathbb{R}, i = 1, \dots, n.$$

Logo,

$$F_Z(z) = P(Z \leq z) = P(\max\{x_1, \dots, x_n\} \leq z) = P(X_1 \leq z) \cdots P(X_n \leq z) = [F(z)]^n$$

(b) Seja  $W = \min\{X_1, \dots, X_n\}$ . Note que dado  $w \in \mathbb{R}$ ,

$$\min\{x_1, \dots, x_n\} > w \Leftrightarrow x_i > w, x_i \in \mathbb{R}, i = 1, \dots, n.$$

Logo,

$$\begin{aligned} F_W(w) &= P(\min\{x_1, \dots, x_n\} \leq w) \\ &= 1 - P(X_1 > w) \cdots P(X_n > w) \\ &= 1 - (1 - P(X_1 \leq w)) \cdots (1 - P(X_n \leq w)) \\ &= 1 - [1 - F(w)]^n. \end{aligned}$$

**Exemplo.** Considere  $X, Y \sim \text{Geo}(p), 0 < p < 1$  independentes. Seja  $W = \min\{X, Y\}$ . A f.d. comum de  $X$  e  $Y$  é dada por

$$F(x) = \begin{cases} 0, & x < 1 \\ 1 - (1 - p)^{[x]}, & x \geq 1 \end{cases}.$$

Do exemplo anterior, temos

$$F_W(w) = \begin{cases} 0, & w < 1 \\ 1 - (1 - p)^{2[w]}, & w \geq 1 \end{cases},$$

ou seja  $W \sim \text{Geo}(1 - (1 - p)^2)$ . Ademais, temos também

$$\begin{aligned}
 P(\min\{X, Y\} = X) &= P(Y \geq X) = P\left(\bigcup_{k=1}^{\infty} \{X = k\}, Y \geq X\right) = \sum_{k=1}^{\infty} P(X = k, Y \geq X) \\
 &= \sum_{k=1}^{\infty} P(X = k)P(Y \geq k) \\
 &= \sum_{k=1}^{\infty} P(X = k)[1 - P(Y \leq k - 1)] \\
 &= \sum_{k=1}^{\infty} P(X = k)[1 - F(k - 1)] \\
 &= \sum_{k=1}^{\infty} p(1 - p)^{k-1}(1 - p)^{k-1} \\
 &= \frac{p}{(1 - p)^2} \sum_{k=1}^{\infty} (1 - p)^{2k} \\
 &= \frac{1}{2 - p}.
 \end{aligned}$$

### Soma de v.a.'s independentes

Sejam  $X, Y$  v.a.'s discretas independentes em  $(\Omega, \mathcal{A}, P)$ ,  $Z = X + Y$  e  $\{x_1, x_2, \dots\}$  o conjunto dos valores possíveis de  $X$ . Dado  $z \in \mathbb{R}$ , temos

$$\begin{aligned}
 p_Z(z) &= P\left(\bigcup_{i=1}^{\infty} \{X = x_i\}, X + Y = z\right) = \sum_{i=1}^{\infty} P(X = x_i, Y = z - x_i) \\
 &= \sum_{i=1}^{\infty} p_{X,Y}(x_i, z - x_i) \\
 &= \sum_{i=1}^{\infty} p_X(x_i)p_Y(z - x_i).
 \end{aligned}$$

Analogamente, se  $\{y_1, y_2, \dots\}$  é o conjunto de valores possíveis de  $Y$ , então

$$p_Z(z) = \sum_{i=1}^{\infty} p_X(z - y_i)p_Y(y_i).$$

Podemos denotar também

$$p_{X+Y}(z) = \sum_x p_{X,Y}(x, z - x) = \sum_x p_X(x)p_Y(z - x)$$

e

$$p_{X+Y}(z) = \sum_y p_{X,Y}(z - y, y) = \sum_y p_X(z - y)p_Y(y).$$

**Exemplo.** Sejam  $X, Y$  v.a.'s i.i.d. com  $X \sim \text{Geom}(p)$ ,  $0 < p < 1$ . Note que  $X + Y$  assume valores em  $\{2, 3, \dots\}$ . Logo, dado  $z$  neste conjunto, segue que

$$p_{X+Y}(z) = \sum_x p_{X,Y}(x, z - x) = \sum_{k=1}^{\infty} p_X(k) \underbrace{p_Y(z - k)}_{>0 \Leftrightarrow z - k \geq 1} = \sum_{k=1}^{z-1} p(1 - p)^{k-1} p(1 - p)^{z-k-1} = p^2(1 - p)^{z-2}(z - 1).$$

Logo,

$$p_{X+Y}(z) = \begin{cases} (z - 1)p^2(1 - p)^{z-2}, & z \in \{2, 3, \dots\} \\ 0, & \text{c.c.} \end{cases},$$

ou seja,  $X + Y \sim \text{BN}(2, p)$ .

**Exemplo.** Sejam  $X, Y$  v.a.'s independentes com  $X \sim \text{Poisson}(\lambda_1)$  e  $Y \sim \text{Poisson}(\lambda_2)$ . Então

$$p_X(x) = \begin{cases} e^{-\lambda_1} \frac{\lambda_1^x}{x!}, & x = 0, 1, \dots \\ 0, & \text{c.c.} \end{cases},$$

$$p_Y(y) = \begin{cases} e^{-\lambda_2} \frac{\lambda_2^y}{y!}, & y = 0, 1, \dots \\ 0, & \text{c.c.} \end{cases}.$$

Note que  $X + Y$  assume valores em  $\{0, 1, 2, \dots\}$ . Logo, dado  $z$  neste conjunto, temos

$$\begin{aligned} p_{X+Y}(z) &= \sum_x p_{X,Y}(x, z-x) \sum_{k=0}^{\infty} p_X k p_Y z - k = \sum_{k=0}^z e^{-\lambda_1} \frac{\lambda_1^k}{k!} e^{-\lambda_2} \frac{\lambda_2^{z-k}}{(z-k)!} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{z!} \sum_{k=0}^z \binom{z}{k} \lambda_1^k \lambda_2^{z-k} \\ &= e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1 + \lambda_2)^z}{z!}. \end{aligned}$$

Portanto,

$$p_{X+Y}(z) = \begin{cases} e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1+\lambda_2)^z}{z!}, & z = 0, 1, \dots \\ 0, & \text{c.c.} \end{cases},$$

ou seja,  $X + Y \sim \text{Poisson}(\lambda_1 + \lambda_2)$ .

### 3.8 Distribuição condicional de variáveis aleatórias discretas

**Proposição 3.19.** Sejam  $X, Y$  v.a.'s discretas em  $(\Omega, \mathcal{A}, P)$  e  $y \in \mathbb{R}$  tal que  $p_Y(y) > 0$ . A função  $p_{X|Y}(\cdot|y) : \mathbb{R} \rightarrow \mathbb{R}$  dada por

$$p_{X|Y}(x|y) = P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p_{X,Y}(x, y)}{p_Y(y)}, x \in \mathbb{R}$$

define uma função de probabilidade, chamada **f.p. condicional de  $X$  dado  $Y = y$** .

**Demonstração.** Seja  $\{x_1, x_2, \dots\}$  o conjunto de valores possíveis de  $X$ . Dado  $y \in \mathbb{R}$  tal que  $p_Y(y) > 0$ , temos

1.  $p_{X|Y}(x|y) \geq 0, \forall x \in \mathbb{R}$ ;
2.  $\{x : p_{X|Y}(x|y) > 0\} = \{x : p_{X,Y}(x, y) \neq 0\} \subset \{x : p_X(x) \neq 0\} = \{x_1, x_2, \dots\}$  é finito ou infinito enumerável;
3.  $\sum_{x_i} p_{X|Y}(x_i|y) = \sum_{x_i} \frac{p_{X,Y}(x_i, y)}{p_Y(y)} = \frac{\sum_{x_i} p_{X,Y}(x_i, y)}{p_Y(y)} = p_Y(y)/p_Y(y) = 1$ .

□

**Exemplo.** Sejam  $X, Y$  v.a.'s i.i.d. geométricas de parâmetro  $0 < p < 1$ . Vimos que

$$p_{X+Y}(z) = \begin{cases} (z-1)p^2(1-p)^{z-2}, & z \in \{2, 3, \dots\} \\ 0, & \text{c.c.} \end{cases}.$$

Logo, dado  $z \in \{2, 3, \dots\}$  temos

$$P_{X|X+Y}(x|z) = P(X = x|X + Y = z) = 0, x \notin \{1, 2, \dots, z-1\}$$

e

$$p_{X|X+Y}(x|z) = \frac{p(1-p)^{x-1}p(1-p)^{z-x-1}}{(z-1)p^2(1-p)^{z-2}} = \frac{1}{z-1}, x \in \{1, 2, \dots, z-1\}.$$

Logo,

$$p_{X|X+Y}(x|z) = \begin{cases} 1/z - 1, & x = 1, 2, \dots, z-1 \\ 0, & \text{c.c.} \end{cases}$$

para cada  $z \in \{2, 3, \dots\}$ , ou seja,  $p_{X|X+Y}$  é f.p. uniforme sobre  $1, 2, \dots, z-1$ .

### 3.9 Exercícios

1. A função de probabilidade conjunta de uma vetor aleatório  $(X, Y)$  é dada por

$$p_{X,Y}(x, y) = \begin{cases} k(2x + y), & x, y = 1, 2 \\ 0, & \text{c.c.} \end{cases}$$

sendo  $k$  uma constante real.

- Determine o valor de  $k$ .
- Determine as funções de probabilidade marginais de  $X$  e  $Y$ .
- São  $X$  e  $Y$  independentes?

#### Solução.

a) Devemos ter

$$\sum_{x,y} p_{X,Y}(x, y) = 1 \implies k \sum_{x=1}^2 2x \sum_{y=1}^2 y = 1 \implies k = 1/18.$$

b) Para  $x = 1, 2$ , temos

$$p_X(x) = \sum_y p_{X,Y}(x, y) = \frac{1}{18}(4x + 3)$$

e, caso contrário,  $p_X(x) = 0$ . Analogamente, para  $y = 1, 2$  temos

$$p_Y(y) = \sum_x p_{X,Y}(x, y) = \frac{1}{18}(2y + 6)$$

e, caso contrário,  $p_Y(y) = 0$ .

c) Não, pois  $p_{X,Y}(1, 1) = 1/6 \neq (7/18)(8/18) = p_X(1)p_Y(1)$ .

□

2. Considere um experimento de lançar três vezes duas moedas distintas  $A$  e  $B$ . Suponha que a moeda  $A$  é honesta, isto é,  $P(\text{cara}) = P(\text{coroa}) = 1/2$ , e a moeda  $B$  não é honesta, com  $P(\text{cara}) = 1/4$  e  $P(\text{coroa}) = 3/4$ . Seja  $X$  a v.a. que denota o número de caras resultantes da moeda  $A$  e  $Y$  a v.a. que denota o número de caras da moeda  $B$ .

- Determine os valores possíveis do vetor  $(X, Y)$ .
- Determine as funções de probabilidade marginais de  $X$  e  $Y$ .
- Determine a função de probabilidade conjunta de  $X$  e  $Y$ .
- Calcule  $P(X = Y)$ ,  $P(X > Y)$  e  $P(X + Y \leq 4)$ .

#### Solução.

a) O conjunto de valores possíveis é  $\{(i, j) : i, j = 0, 1, 2, 3\}$ .

b) Temos

$$p_X(x) = \begin{cases} 1/8, & x = 0, 3 \\ 3/8, & x = 1, 2 \\ 0, & \text{c.c.} \end{cases}, \quad p_Y(y) = \begin{cases} 27/64, & y = 0, 1 \\ 9/64, & y = 2 \\ 1/64, & y = 3 \\ 0, & \text{c.c.} \end{cases}.$$

c)  $p_{X,Y}(x, y) = p_X(x)p_Y(y)$ .

d) Temos

$$P(X = Y) = \sum_{k=1}^3 P(X = k)P(Y = k) = \frac{27 + 81 + 27 + 1}{512} = 136/512,$$

$$P(X > Y) = \frac{81 + 162 + 63}{512} = \frac{306}{512}$$

e

$$P(X + Y \leq 4) = \frac{1}{8} + \frac{3}{8} + \frac{189 + 54}{512} = \frac{499}{512}.$$

□

3. Seja  $X$  uma variável aleatória geometricamente distribuída com parâmetro  $p$  e seja  $M \in \mathbb{N}$  uma

constante. Determine a função de probabilidade de  $Y = \min(X, M)$ .

4. Considere 10 lançamentos independentes de um dado honesto e seja  $X_i$  o número de ocorrências da face  $i, i = 1, \dots, 6$ .

**Solução.** Temos dois casos

(a)  $k < M: P(Y = k) = P(X = k) = p(1 - p)^{k-1}$

(b)  $k = M: P(Y = M) = P(X \geq M) = (1 - p)^{M-1}$ .

Portanto,

$$p_Y(k) = \begin{cases} p(1 - p)^{k-1}, & k \in \{2, \dots, M - 1\} \\ (1 - p)^{M-1}, & k = M \\ 0, & \text{c.c.} \end{cases}$$

□

- a) Determine a função de probabilidade conjunta de  $X_1, \dots, X_6$ .  
 b) Determine as funções de probabilidade marginais de  $X_i$ , para  $i = 1, \dots, 6$   
 c) São  $X_1, \dots, X_6$  independentes?

**Solução.**

- a) Temos  $(X_1, X_2, \dots, X_6) \sim \text{Multinomial}(10, 1/6, 1/6, 1/6, 1/6, 1/6, 1/6)$ .  
 b) Temos  $X_i \sim B(n, 1/6), i = 1, 2, \dots, 6$ .  
 c) Não, porque  $p_{X_1, \dots, X_6}(x_1, \dots, x_6) \neq p_{X_1}(x_1) \cdots p_{X_6}(x_6)$  em geral.

□

5. Suponha que se distribui aleatoriamente  $2r$  bolas em  $r$  caixas. Seja  $X_i$  o número de bolas na caixa  $i$ .

- a) Obtenha a função de probabilidade conjunta de  $X_1, \dots, X_r$ .  
 b) Obtenha a probabilidade de que cada caixa contenha exatamente 2 bolas.

**Solução.**

- a) Temos  $(X_1, \dots, X_r) \sim \text{Multinomial}(2r, p_1, \dots, p_r)$ , com  $p_i = 1/r, i = 1, 2, \dots, r$ .  
 b) Temos

$$P(X_1 = 2, \dots, X_r = 2) = \binom{2r}{2, 2, \dots, 2} \cdot \frac{1}{r^2} \cdots \frac{1}{r^2} = \frac{(2r)!}{2^r \cdot r^{2r}}$$

□

6. Sejam  $X$  e  $Y$  duas variáveis aleatórias independentes que se distribuem uniformemente sobre  $\{0, \dots, N\}$ . Determine:

- a)  $P(X \geq Y)$ .  
 b)  $P(X = Y)$ .  
 c) a função de probabilidade de  $Z = \min(X, Y)$ .  
 d) a função de probabilidade de  $W = \max(X, Y)$ .  
 e) a função de probabilidade de  $U = |Y - X|$ .

**Solução.**

- a) Temos

$$\begin{aligned} P(X \geq Y) &= P(X = 0)P(Y \leq 0) + \cdots + P(X = N)P(Y \leq N) \\ &= \frac{1}{N+1} \left( \frac{1}{N+1} + \frac{2}{N+1} + \cdots + \frac{N}{N+1} + 1 \right) \\ &= \frac{N+2}{2(N+1)}. \end{aligned}$$

b) Temos

$$P(X = Y) = \sum_{i=0}^N P(X = i)P(Y = i) = \frac{1}{N+1}.$$

c) Para  $k = 0, 1, \dots, N$  temos

$$\begin{aligned} P(Z = k) &= P(X = k)P(Y = k) + P(X > k)P(Y = k) + P(X = k)P(Y > k) \\ &= \frac{1}{(N+1)^2} + 2\frac{N-k}{(N+1)^2} \\ &= \frac{2(N-k) + 1}{(N+1)^2} \end{aligned}$$

e, caso contrário,  $P(Z = k) = 0$ .

d) Para  $k = 0, 1, \dots, N$ , temos

$$\begin{aligned} P(W = k) &= P(X = k)P(Y = k) + P(X = k)P(Y < k) + P(X < k)P(Y = k) \\ &= \frac{1}{(N+1)^2} + 2\frac{k}{(N+1)^2} \\ &= \frac{2k+1}{(N+1)^2} \end{aligned}$$

e, caso contrário,  $P(W = k) = 0$ .

e) Para  $k = 0$ , temos

$$P(U = k) = P(X = Y) = \frac{1}{N+1}.$$

Para  $k = 1, \dots, N$ , temos

$$\begin{aligned} P(U = k) &= 2P(X = Y + k) \\ &= 2[P(X = k)P(Y = 0) + P(X = k+1)P(Y = 1) + \dots + P(X = N)P(Y = N-k)] \\ &= 2\frac{N-k+1}{(N+1)^2} \end{aligned}$$

e, para  $k \notin \{0, 1, \dots, N\}$ , temos  $P(U = k) = 0$ .

□

7. Sejam  $X$  e  $Y$  duas variáveis aleatórias independentes com funções de probabilidade geométricas de parâmetros  $p_1$  e  $p_2$ , respectivamente. Obtenha:

- $P(X \geq Y)$
- $P(X = Y)$
- a função de probabilidade de  $Z = \min(X, Y)$
- a função de probabilidade de  $W = X + Y$ .

### Solução.

a) Temos

$$P(X \geq Y) = \sum_{i=1}^{\infty} P(X = i)P(Y \leq i).$$

Após algumas simplificações, chegamos em

$$P(X \geq Y) = \frac{p_2}{p_1 + p_2 - p_1 p_2}.$$

b) Temos

$$\begin{aligned} P(X = Y) &= \sum_{i=1}^{\infty} P(X = i)P(Y = i) \\ &= \frac{p_1 p_2}{p_1 + p_2 - p_1 p_2} \end{aligned}$$

após algumas simplificações.

c) Para  $k = 1, 2, \dots$ , temos que

$$P(Z > k) = P(X > k)P(Y > k) = [(1 - p_1)(1 - p_2)]^k.$$

Daí,

$$F_Z(k) = \begin{cases} 0, & k < 1 \\ 1 - [(1 - p_1)(1 - p_2)]^{[k]}, & k \geq 1. \end{cases}$$

Logo,  $Z \sim \text{Geo}(1 - (1 - p_1)(1 - p_2)) = \text{Geo}(p_1 + p_2 - p_1 p_2)$ .

□

8. Sejam  $X$  e  $Y$  duas variáveis aleatórias independentes com a mesma função de probabilidade geométrica de parâmetro  $p$ . Sejam  $Z = Y - X$  e  $W = \min(X, Y)$ .

a) Mostre que para  $w \geq 1$  e  $z$  inteiros, temos

$$P(W = w, Z = z) = \begin{cases} P(X = w - z)P(Y = w), & z < 0 \\ P(X = w)P(Y = w + z), & z \geq 0 \end{cases}$$

b) Conclua do item anterior que dados  $w \geq 1$  e  $z$  inteiros, temos

$$P(W = w, Z = z) = p^2(1 - p)^{2(w-1)}(1 - p)^{|z|}$$

c) Use o item anterior e o exercício 7c) para mostrar que  $W$  e  $Z$  são independentes.

### Solução.

a) Se  $z < 0$ , então  $X > Y$ , ou seja,  $Y = w$  e  $X = w - z$ . Se  $z \geq 0$ , então  $Y \geq X$ , ou seja,  $X = w$  e  $Y = w + z$ . Portanto,

$$P(W = w, Z = z) = \begin{cases} P(X = w - z)P(Y = w), & z < 0 \\ P(X = w)P(Y = w + z), & z \geq 0. \end{cases}$$

b) Do item anterior,

$$P(W = w, Z = z) = \begin{cases} p^2(1 - p)^{2(w-1)}(1 - p)^{-z}, & z < 0 \\ p^2(1 - p)^{2(w-1)}(1 - p)^z, & z \geq 0. \end{cases} = p^2(1 - p)^{2(w-1)}(1 - p)^{|z|}.$$

c) Note que

$$\begin{aligned} p_Z(z) &= p^2(1 - p)^{|z|-2} \sum_{w=1}^{\infty} (1 - p)^{2w} \\ &= p^2(1 - p)^{|z|-2} \frac{(1 - p)^2}{1 - (1 - p)^2}. \end{aligned}$$

Logo,

$$\begin{aligned} p_Z(z)p_W(w) &= p^2(1 - p)^{|z|-2} \frac{(1 - p)^2}{1 - (1 - p)^2} [1 - (1 - p)^2](1 - p)^{2(w-1)} \\ &= p_{Z,W}(z, w) \end{aligned}$$

e, portanto,  $Z$  e  $W$  são independentes.

□

9. Sejam  $X$  e  $Y$  v.a.'s independentes. Determine a função de probabilidade de  $Z = X + Y$  seguintes casos:

a)  $X \sim \text{Poisson}(\lambda_1)$  e  $Y \sim \text{Poisson}(\lambda_2)$ .

b)  $X$  e  $Y$  uniformemente distribuídas sobre  $\{1, 2, \dots, N\}$ .

### Solução.

(a) Foi feito no texto.

(b) Para  $z = 2, \dots, N$ , temos

$$p_Z(z) = P(X + Y = z) = \sum_{x=1}^{z-1} P(X = x)P(Y = z - x) = \frac{z - 1}{N^2}.$$

Para  $z = N + 1, \dots, 2N$ , temos

$$p_Z(z) = P(X + Y = z) = \sum_{x=1}^{2N-z+1} P(X = x)P(Y = z - x) = \frac{2N - z + 1}{N^2}.$$

Para  $z \notin \{2, 3, \dots, 2N\}$ ,  $p_Z(z) = 0$ .

□

10. Sejam  $X_1, \dots, X_l$  v.a.'s independentes, tais que  $X_j \sim \text{Poisson}(\lambda_j), j = 1, \dots, l$ . Determine a função de probabilidade de  $Z = X_1 + \dots + X_l$ .

**Solução.** Do exercício anterior, temos que

$$\sum_{i=1}^l X_i \sim \text{Poisson} \left( \sum_{i=1}^l \lambda_i \right).$$

□

11. Considere um experimento com três resultados possíveis que ocorrem com probabilidades  $p_1, p_2$  e  $p_3$ , respectivamente. Suponha que se realiza  $n$  repetições independentes do experimento e seja  $X_i$  o número de vezes que ocorre o resultado  $i, i = 1, 2, 3$ .

- a) Determine a probabilidade de  $X_1 + X_2$ .  
 b) Para cada  $z$ , determine  $P(X_2 = y | X_1 + X_2 = z), y \in \mathbb{R}$ .

**Demonstração.**

- a) Para  $z = 0, 1, \dots, n$ , temos

$$\begin{aligned} p_{X_1+X_2}(z) &= \sum_{x_1} p_{X_1, X_2}(x_1, z - x_1) \\ &= \sum_{x_1=0}^z \frac{n!}{(n-z)!x_1!(z-x_1)!} p_1^{x_1} p_2^{z-x_1} p_3^{n-z} \\ &= \binom{n}{z} p_3^{n-z} \sum_{x_1=0}^z \binom{z}{x_1} p_1^{x_1} p_2^{z-x_1} \\ &= \binom{n}{z} [1 - (p_1 + p_2)]^{n-z} (p_1 + p_2)^z. \end{aligned}$$

Logo,  $X_1 + X_2 \sim B(n, p_1 + p_2)$ .

- b) Dado  $z \in \{0, 1, \dots, n\}$ , temos

$$\begin{aligned} P(X_2 = y | X_1 + X_2 = z) &= \frac{P(X_2 = y, X_1 = z - y)}{P(X_1 + X_2 = z)} \\ &= \frac{P(X_2 = y, X_1 = z - y, X_3 = n - z)}{P(X_1 + X_2 = z)} \\ &= \frac{n!}{(z-y)!y!(n-z)!} \cdot \frac{p_1^{z-y} p_2^y p_3^{n-z}}{(p_1 + p_2)^z p_3^{n-z}} \cdot \frac{z!(n-z)!}{n!} \\ &= \binom{z}{y} \left( \frac{p_1}{p_1 + p_2} \right)^{z-y} \left( \frac{p_2}{p_1 + p_2} \right)^y \end{aligned}$$

para  $y = 0, 1, \dots, z$  e 0 caso contrário.

□

12. Use a aproximação de Poisson para calcular a probabilidade de:

- a) que no máximo 2 dentre 50 motoristas tenham carteiras de habilitação inválida se normalmente 5% dos motoristas o tem;  
 b) que uma caixa com 100 fusíveis contenha no máximo 2 fusíveis defeituosos se 3% dos fusíveis fabricados são defeituosos.

**Solução.**

- a) Temos  $X \sim B(50, 0, 05) \approx \text{Poisson}(5/2)$ . Daí,  $P(X \leq 2) = e^{-5/2} 5^3/8$ .  
 b) Temos  $X \sim B(100, 0, 03) \approx \text{Poisson}(3)$ . Logo,  $P(X \leq 2) = e^{-3} 17/2$ .

□

13. Lança-se um dado até observar o número 6. Considere  $X$  o número de lançamentos até observar 6 pela primeira vez. Responda:

- a) qual é a probabilidade de que sejam necessários seis lançamentos no máximo?



- b) quantos lançamentos são necessários para que a probabilidade de obter 6 seja no mínimo  $1/2$ ?

**Solução.**

- a) Temos  $P(X \leq 6) = 1 - (5/6)^6$ .  
 b) Queremos  $k$  inteiro tal que  $(5/6)^k \leq 1/2$ . Para isso, devemos ter

$$k \geq \frac{\ln(1/2)}{\ln(5/6)} = 3,8.$$

Logo,  $k = 4$ . □

14. Sejam  $X$  e  $Y$  v.a.'s independentes com distribuição de Poisson de parâmetros  $\lambda_1$  e  $\lambda_2$ , respectivamente. Para cada  $z \in \{0, 1, \dots\}$  determine  $P(X = x | X + Y = z), x \in \mathbb{R}$ .

**Solução.** Para  $x = 0, 1, \dots, z$  temos

$$\begin{aligned} P(X = x | X + Y = z) &= \frac{P(X = x)P(Y = z - x)}{P(X + Y = z)} \\ &= \binom{z}{x} \frac{\lambda_1^x \lambda_2^{z-x}}{(\lambda_1 + \lambda_2)^z}. \end{aligned}$$

Caso contrário, a probabilidade é 0. □

15. Sejam  $X, Y$  e  $Z$  v.a.'s independentes com distribuições de Poisson de parâmetros  $\lambda_1, \lambda_2$  e  $\lambda_3$ , respectivamente. Para cada  $m = 0, 1, \dots$ , determine  $P(X = x, Y = y, Z = z | X + Y + Z = m)$ , para números inteiros  $x, y$  e  $z$ .

**Solução.** Para  $x, y, z \in \{0, 1, 2, \dots\}$  tais que  $x + y + z = m$ , temos

$$\begin{aligned} P(X = x, Y = y, Z = z | X + Y + Z = m) &= \frac{P(X = x)P(Y = y)P(Z = z)}{P(X + Y + Z = m)} \\ &= \frac{m!}{x!y!z!} \cdot \frac{\lambda_1^x \lambda_2^y \lambda_3^z}{(\lambda_1 + \lambda_2 + \lambda_3)^m}. \end{aligned}$$

Do contrário, a probabilidade é nula. □

## 4 Esperança de variáveis aleatórias discretas

Começamos com uma motivação para depois introduzirmos a definição de esperança: dada uma amostra  $\{a_1, a_2, \dots, a_n\} \subset \mathbb{R}$ , a média amostral é dada por  $\frac{a_1 + a_2 + \dots + a_n}{n}$ . Podemos representar esses dados numa tabela:

Valores distintos na amostra	Frequência absoluta
$x_1$	$N_1$
$x_2$	$N_2$
$\vdots$	$\vdots$
$x_m$	$N_m$

Note que  $N_i$  é o número de vezes que  $x_i$  aparece na amostra,  $i = 1, 2, \dots, m$ . Além disso,  $N_1 + N_2 + \dots + N_m = n$  e  $x_1 N_1 + x_2 N_2 + \dots + x_m N_m = a_1 + a_2 + \dots + a_n$ . Com isso, podemos escrever a média dos  $a_i$  como  $\frac{x_1 N_1 + x_2 N_2 + \dots + x_m N_m}{N_1 + N_2 + \dots + N_m}$  ou, ainda,  $x_1 \frac{N_1}{n} + \dots + x_m \frac{N_m}{n}$ , em que  $N_i/n$  é a frequência relativa de  $x_i$ ,  $i = 1, 2, \dots, m$ . Nesse caso, a f.p.  $p: \mathbb{R} \rightarrow \mathbb{R}$  é dada por

$$p(x) = \begin{cases} N_i/n, & x = x_i \text{ para algum } i = 1, 2, \dots, m \\ 0, & \text{c.c.} \end{cases}$$

Agora, suponhamos que  $x_1, x_2, \dots, x_m$  sejam valores possíveis de uma v.a.  $X$  e que  $a_1, a_2, \dots, a_n$  são valores (independentes) observados de  $X$ . Então, de acordo com a interpretação da probabilidade como

frequência relativa (Lei Forte dos Grandes Números), temos que para  $n$  grande a frequência relativa se aproxima da probabilidade real, i.e.,

$$\lim_{n \rightarrow \infty} \frac{N_i}{n} = p(x_i) = P(X = x_i).$$

Assim, o **valor esperado** de  $X$ , representado por  $EX$  ou  $E[X]$ , é  $EX = \sum_{i=1}^m x_i p(x_i)$ , sendo  $p$  a f.p. de  $X$ .

**Definição 4.1 (Esperança).** Seja  $X$  uma v.a. discreta em  $(\Omega, \mathcal{A}, P)$  com valores possíveis  $\{x_1, x_2, \dots\}$ . Se  $\sum_i |x_i| p(x_i) < \infty$ , dizemos que  $X$  tem **esperança finita** e definimos sua esperança ou **esperança matemática** ou **valor esperado** ou **média** como

$$EX = \sum_{i=1}^{\infty} x_i p(x_i) = \sum_x x p(x).$$

Note, em particular, que se o conjunto dos valores possíveis de  $X$  é finito, então  $X$  tem esperança finita.

**Exemplo.** Seja  $X \sim B(n, p)$ . Temos

$$p_X(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & k = 0, 1, \dots, n \\ 0, & \text{c.c.} \end{cases}.$$

Como  $X$  tem uma quantidade finita de valores possíveis, então  $X$  tem esperança finita dada por

$$\begin{aligned} EX &= \sum_x x p(x) = \sum_{k=0}^n k P(X = k) = \sum_{k=1}^n k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^n \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{n-1-(k-1)} \\ &= np. \end{aligned}$$

Em particular, se  $n = 1$  então  $EX = p$ .

**Exemplo.** Seja  $X \sim \text{Poisson}(\lambda)$ . Temos

$$p_X(k) = \begin{cases} e^{-\lambda} \frac{\lambda^k}{k!}, & k = 0, 1, \dots \\ 0, & \text{c.c.} \end{cases}.$$

Note que como  $X$  tem infinitos valores possíveis, não sabemos a priori se  $EX < \infty$ . Temos

$$\sum_x |x| p_X(x) = \sum_{k=0}^{\infty} |k| e^{-\lambda} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda < \infty.$$

Logo,  $X$  tem esperança finita e, como seus valores possíveis são não negativos, segue que  $EX = \lambda$ .

**Exemplo.** Seja  $X \sim \text{Geom}(p)$ . Temos

$$p_X(k) = \begin{cases} p(1-p)^{k-1}, & k = 1, 2, \dots \\ 0, & \text{c.c.} \end{cases}, \quad 0 < p < 1.$$

Note que

$$\begin{aligned} \sum_x |x| p_X(x) &= \sum_{k=1}^{\infty} |k| p(1-p)^{k-1} = p \sum_{k=1}^{\infty} k(1-p)^{k-1} = -p \sum_{k=1}^{\infty} \frac{d}{dp} [(1-p)^k] \\ &= -p \frac{d}{dp} \left[ \sum_{k=1}^{\infty} (1-p)^k \right] \\ &= -p(-1/p^2) \\ &= 1/p, \end{aligned}$$

em que a série converge pelo teste da integral. Logo,  $X$  tem esperança finita e, como seus valores possíveis são não negativos, temos  $EX = 1/p$ .

**Exemplo.** Considere  $p : \mathbb{R} \rightarrow \mathbb{R}$  dada por

$$p(x) = \begin{cases} \frac{1}{x(x+1)}, & x = 1, 2, \dots \\ 0, & \text{c.c.} \end{cases}$$

Note que  $p(x) \geq 0, \forall x \in \mathbb{R}$ ,  $\{x \in \mathbb{R} : p(x) > 0\} = \mathbb{N}$  é enumerável e  $\sum_x p(x) = \sum_{x=1}^{\infty} \frac{1}{x(x+1)} = \lim_{N \rightarrow \infty} \sum_{x=1}^N \frac{1}{x} - \frac{1}{x+1} = \lim_{N \rightarrow \infty} 1 - \frac{1}{N+1} = 1$ . Logo,  $p$  é uma f.p. e, se  $X$  é v.a. com f.p.  $p$ , temos

$$\sum_x |x|p(x) = \sum_{x=1}^{\infty} \frac{1}{x+1}$$

que diverge pelo teste da integral. Logo,  $X$  não tem esperança finita e escrevemos  $EX = +\infty$ .

## 4.1 Esperança de função de variável aleatória

Nosso interesse aqui é determinar a esperança de uma v.a. (discreta) que é função de uma ou mais v.a.'s:  $Z = g(X)$ , com  $X = (X_1, \dots, X_n)$  e  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  função. Note que o conjunto de valores possíveis de  $X$  é  $\{x_1, x_2, \dots\} \subset \mathbb{R}^n$  (enumerável), ou seja,  $P(X = x_j) > 0, j = 1, 2, \dots$  e  $P(X \in \{x_1, x_2, \dots\}) = 1$ . Note que  $x_j = (x_{j1}, \dots, x_{jn}) \in \mathbb{R}^n, j = 1, 2, \dots$  e  $\sum_x g(x)p_X(x) = \sum_j g(x_j)p_X(x_j)$ .

**Teorema 4.2.** Sejam  $X = (X_1, \dots, X_n)$  um vetor aleatório discreto com f.p.  $p_X$  e  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  função. A v.a.  $Z = g(X)$  tem esperança finita se, e só se,  $\sum_x |g(x)|p_X(x) < \infty$ . Nesse caso,

$$EZ = \sum_x g(x)p_X(x) = \sum_{x_1} \cdots \sum_{x_n} g(x_1, \dots, x_n)P(X_1 = x_1, \dots, X_n = x_n).$$

**Demonstração.** Vamos mostrar o caso  $n = 1$ ; o caso  $n > 1$  é análogo. Temos  $g : \mathbb{R} \rightarrow \mathbb{R}$  função e, se  $\{x_1, x_2, \dots\}$  é o conjunto de valores possíveis de  $X$ , então  $\{g(x_1), g(x_2), \dots, g(x_n)\}$  é o conjunto de valores possíveis de  $Z = g(X)$ . Note que pode haver repetição, i.e., podemos ter  $g(x_i) = g(x_k)$  com  $x_i \neq x_k$ . Seja  $A_j = \{x : x = x_i \text{ e } g(x) = z_j\}, j = 1, 2, \dots$ , sendo  $\{z_1, z_2, \dots\}$  o conjunto de valores possíveis de  $Z$ . Note que  $\{X \in A_j\} = \{Z = z_j\}$  e, para cada  $j = 1, 2, \dots, A_j \subset \{x_1, x_2, \dots\}$  e os  $A_j$ 's são dois a dois disjuntos já que  $g$  função. Além disso,

$$\bigcup_{j=1}^{\infty} A_j = \{x_1, x_2, \dots\}.$$

Logo,

$$\begin{aligned} \sum_j |z_j|p_Z(z_j) &= \sum_j |z_j|P(X \in A_j) = \sum_j |z_j| \sum_{i:x_i \in A_j} P(X = x_i) \\ &= \sum_j \sum_{i:x_i \in A_j} |z_j|P(X = x_i) \\ &= \sum_j \sum_{i:x_i \in A_j} |g(x_i)|P(X = x_i) \\ &= \sum_k |g(x_k)|P(X = x_k). \end{aligned}$$

Logo,  $EZ < \infty \Leftrightarrow \sum_x |g(x)|p_X(x) < \infty$ . Agora, se  $EZ < \infty$ , então analogamente ao acima,  $EZ = \sum_x g(x)p_X(x)$ .  $\square$

**Observação.** Dada uma v.a. discreta  $X$ , vimos que (i)  $EX < \infty \Leftrightarrow \sum_x |x|P(X = x) < \infty$  (por definição) e (ii)  $E|X|$ , quando existe, é  $\sum_x |x|P(X = x)$  pelo teorema acima. Logo,  $EX < \infty \Leftrightarrow E|X| < \infty$ .

**Exemplo.** Sejam  $X, Y$  v.a.'s independentes com  $X \sim \text{Poisson}(\lambda_1)$  e  $Y \sim \text{Poisson}(\lambda_2)$ . Vamos calcular  $E[X^2], E[XY]$  e  $E[X + Y]$ .

(i) Pelo teorema,

$$\begin{aligned}
\sum_x |g(x)|p_X(x) &= \sum_{k=0}^{\infty} k^2 e^{-\lambda_1} \frac{\lambda_1^k}{k!} \\
&= \sum_{k=1}^{\infty} k e^{-\lambda_1} \frac{\lambda_1^k}{(k-1)!} \\
&= \lambda_1 \left[ \sum_{k=1}^{\infty} (k-1) e^{-\lambda_1} \frac{\lambda_1^{k-1}}{(k-1)!} + \sum_{k=1}^{\infty} e^{-\lambda_1} \frac{\lambda_1^{k-1}}{(k-1)!} \right] \\
&= \lambda_1 [EX + e^{\lambda_1} e^{-\lambda_1}] \\
&= \lambda_1(\lambda_1 + 1) \in \mathbb{R}.
\end{aligned}$$

Logo,  $X^2$  tem esperança finita e, como seus valores possíveis são não negativos,  $E[X^2] = \lambda_1(\lambda_1 + 1)$ .

(ii) Pelo teorema,

$$\begin{aligned}
\sum_x \sum_y |g(x, y)|p_{X,Y}(x, y) &= \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} kj P(X = k, Y = j) \\
&= \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} kje^{-\lambda_1} \frac{\lambda_1^k}{k!} e^{-\lambda_2} \frac{\lambda_2^j}{j!} \\
&= \sum_{k=1}^{\infty} \left[ ke^{-\lambda_1} \frac{\lambda_1^k}{k!} \sum_{j=1}^{\infty} je^{-\lambda_2} \frac{\lambda_2^j}{j!} \right] \\
&= \lambda_2 \sum_{k=1}^{\infty} ke^{-\lambda_1} \frac{\lambda_1^k}{k!} \\
&= \lambda_1 \lambda_2 \in \mathbb{R},
\end{aligned}$$

logo  $XY$  tem esperança finita e  $E[XY] = \lambda_1 \lambda_2$  pois os valores possíveis de  $XY$  são não negativos.

(iii) Temos, pelo teorema, que

$$\begin{aligned}
\sum_x \sum_y |g(x, y)|p_{X,Y}(x, y) &= \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} (k+j)P(X = k, Y = j) \\
&= \sum_{k=0}^{\infty} k \sum_{j=0}^{\infty} P(X = k, Y = j) + \sum_{j=0}^{\infty} j \sum_{k=0}^{\infty} P(X = k, Y = j) \\
&= \sum_{k=0}^{\infty} kp_X(k) + \sum_{j=0}^{\infty} jp_Y(j) \\
&= EX + EY \\
&= \lambda_1 + \lambda_2 \in \mathbb{R},
\end{aligned}$$

logo  $X + Y$  tem esperança finita e  $E[X + Y] = EX + EY = \lambda_1 + \lambda_2$  pois os valores possíveis de  $X + Y$  são não negativos.

**Teorema 4.3** (Propriedades da esperança). Sejam  $X$  e  $Y$  v.a.'s definidas em  $(\Omega, \mathcal{A}, P)$  com  $EX, EY < \infty$ . Temos

(a)  $P(X = c) = 1 \implies EX = c, \forall c \in \mathbb{R}$

(b)  $E[cX] < \infty$  e  $E[cX] = cEX, \forall c \in \mathbb{R}$

(c)  $E[X + Y] < \infty$  e  $E[X + Y] = EX + EY$ . De maneira geral, dados  $c_1, \dots, c_n \in \mathbb{R}$  e  $EX_1, \dots, EX_n < \infty$ , então

$$E[c_1X_1 + \dots + c_nX_n] = \sum_i c_i EX_i < \infty$$

(d) Se  $P(X \geq Y) = 1$ , então  $EX \geq EY$  e  $EX = EY \iff P(X = Y) = 1$ . Em particular,  $P(X \geq 0) = 1 \implies EX \geq 0$

$$(e) |EX| \leq E|X|.$$

**Demonstração.**

(a) Temos  $\sum_x |x|p(x) = |c| < \infty \therefore EX = c < \infty$ .

(b) Temos  $\sum_x |cx|p(x) = |c| \sum_x |x|p(x) < \infty \therefore E[cX] = cEX < \infty$ .

(c) Temos

$$\begin{aligned} & \sum_{x_1} \cdots \sum_{x_n} |c_1x_1 + \cdots + c_nx_n|P(X_1 = x_1, \dots, X_n = x_n) \\ & \leq \sum_{x_1} |c_1x_1| \sum_{x_2} \cdots \sum_{x_n} P(X_1 = x_1, \dots, X_n = x_n) + \cdots + \sum_{x_n} |c_nx_n| \sum_{x_1} \cdots \sum_{x_{n-1}} P(X_1 = x_1, \dots, X_n = x_n) \\ & = |c_1| \sum_{x_1} |x_1|P(X_1 = x_1) + \cdots + |c_n| \sum_{x_n} |x_n|P(X_n = x_n) < \infty. \end{aligned}$$

Logo,  $E \left[ \sum_i c_i X_i \right] = \sum_i c_i EX_i < \infty$ .

(d) Se  $P(X \geq Y) = 1$ , então  $x_i \geq y_j$  para todo par de valores possíveis  $(x_i, y_j)$ . Daí,

$$EX = \sum_x xP(X = x) \geq \sum_y yP(X = x) \geq \sum_y yP(Y = y) = EY.$$

Se  $P(X = Y) = 1$ , então  $X$  e  $Y$  têm o mesmo conjunto de valores possíveis e  $P(X = x) = P(Y = y)$ . Logo,  $EX = \sum_x xP(X = x) = \sum_y yP(Y = y) = EY$ . Se  $EX = EY$ , devemos ter  $X$  e  $Y$  com o mesmo conjunto de valores possíveis e  $P(X = x) = P(Y = y)$ .

(e)  $|EX| = \left| \sum_x xp(x) \right| \leq \sum_x |x|p(x) = E|X|.$

□

**Teorema 4.4.** Seja  $X$  v.a. definida em  $(\Omega, \mathcal{A}, P)$  tal que  $P(|X| \leq M) = 1, M \in \mathbb{R}$ . Então  $EX < \infty$  e  $|EX| \leq M$ .

**Demonstração.** Seja  $x$  valor possível de  $X$ . Se  $|x| > M$ , então  $P(|X| > M) \geq P(X = x) > 0$ . Mas  $P(|X| > M) = 1 - P(|X| \leq M) = 0$ . Logo,  $|x| \leq M$  e, assim,

$$\sum_x |x|p_X(x) \leq M \sum_x p_X(x) = M < \infty,$$

ou seja,  $EX < \infty$ . Por fim,

$$|EX| \leq E|X| \leq EM = M.$$

□

**Exemplo.** Se  $X \sim B(n, p)$ , então  $EX = np$ , como vimos. Outra maneira de calcular  $EX$  seria tomar  $X_i \sim B(1, p), i = 1, \dots, n$  e  $S_n = X_1 + \cdots + X_n \sim B(n, p)$ . Então  $ES_n = E[X_1 + \cdots + X_n] = \sum_i EX_i = np$ .

**Exemplo.** Seja  $X \sim \text{Hgeo}(N, m, n)$ . Temos

$$p_{S_n}(k) = \begin{cases} \frac{\binom{m}{k} \binom{N-m}{n-k}}{\binom{N}{n}}, & k = 0, 1, \dots, n \\ 0, & \text{c.c.} \end{cases}.$$

Poderíamos calcular  $ES_n$  pela definição, mas isso seria trabalhoso e complicado; vamos proceder de outro modo. Sejam  $X_i$  v.a.'s indicadoras da presença do objeto de tipo  $i$ . Temos, então,

$$EX_i = \frac{\binom{1}{1} \binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}, i = 1, 2, \dots, n.$$

Logo,  $ES_n = \sum_i EX_i = mn/N$ .

**Observação.** Vimos que  $EX, EY < \infty$  implica  $E[X + Y] = EX + EY$ . Contudo, o mesmo não vale para  $E[XY]$ . De fato, sejam  $X, Y$  v.a.'s com  $p_X(-1) = 1/2 = p_X(1)$  e  $Y = X$ . A f.p. conjunta de

$X$  e  $Y$  pode ser representada como

$X \setminus Y$	-1	1	$p_X(x)$
-1	1/2	0	1/2
1	0	1/2	1/2
$p_Y(y)$	1/2	1/2	1

Temos  $EX = -1/2 + 1/2 = 0 = EY$  mas  $E[XY] = 1/2 + 1/2 = 1 \neq EXEY$ . É importante notar que  $X$  e  $Y$  **não são independentes!**

**Teorema 4.5.** Sejam  $X$  e  $Y$  v.a.'s definidas em  $(\Omega, \mathcal{A}, P)$  com  $EX, EY < \infty$ . Se  $X$  e  $Y$  são independentes, então  $E[XY] < \infty$  e  $E[XY] = EXEY$ .

**Demonstração.** Temos

$$\sum_x \sum_y |xy| p_{X,Y}(x, y) = \sum_x |x| p_X(x) \sum_y |y| p_Y(y) < \infty$$

pois  $EX, EY < \infty$ . Daí, segue que

$$E[XY] = \left( \sum_x x p_X(x) \right) \left( \sum_y y p_Y(y) \right) = EXEY.$$

□

**Observação.** Note que a recíproca deste teorema **é falsa!** O exemplo anterior é um contraexemplo disso.

**Teorema 4.6.** Seja  $X$  v.a. inteira não negativa, i.e., com conjunto de valores possíveis  $\{0, 1, 2, \dots\}$ . Então  $EX < \infty \Leftrightarrow \sum_{x=1}^{\infty} P(X \geq x) < \infty$  e, nesse caso,  $EX = \sum_{x=1}^{\infty} P(X \geq x) = \sum_{x=0}^{\infty} 1 - F_X(x)$ .

**Demonstração.** Temos que  $EX < \infty$  se, e só se,

$$\sum_x |x| p_X(x) = \sum_{x=1}^{\infty} x p_X(x) = \sum_{x=1}^{\infty} p_X(x) + \sum_{x=2}^{\infty} p_X(x) + \dots = \sum_{k=1}^{\infty} \sum_{x=k}^{\infty} p_X(x) = \sum_{k=1}^{\infty} P(X \geq k) < \infty,$$

em que a segunda igualdade vale pois a convergência é absoluta. Daí, quando existe,  $EX$  é dada por

$$\sum_{k=1}^{\infty} P(X \geq k) = \sum_{k=0}^{\infty} 1 - F_X(k).$$

□

**Exemplo.**  $X \sim \text{Geo}(p)$  é inteira não negativa, logo

$$EX = \sum_{k=1}^{\infty} P(X \geq k) = \sum_{k=0}^{\infty} 1 - F_X(k) = \sum_{k=0}^{\infty} (1-p)^k = 1/p.$$

**Exemplo.** Sejam  $X, Y \sim \text{Geom}(p)$  e  $W = \min\{X, Y\}$ . Temos  $W$  inteira não negativa e

$$F_W(w) = \begin{cases} 0, & w < 1 \\ 1 - (1-p)^{2[w]}, & w \geq 1 \end{cases}.$$

Logo,

$$EW = \sum_{k=0}^{\infty} 1 - F_W(k) = \sum_{k=0}^{\infty} (1-p)^{2k} = \frac{1}{1 - (1-p)^2} = \frac{1}{2p - p^2} = \frac{1}{p(2-p)}.$$

## 4.2 Momentos

**Definição 4.7** (Momentos e variância). Sejam  $X$  v.a. e  $k \in \mathbb{N}$ . Se  $E[X^k] < \infty$ , dizemos que  $X$  tem momento de ordem  $k$  e chamamos  $E[X^k]$  de **momento de ordem  $k$**  ou  $k$ -ésimo momento de  $X$ .

Quando  $X$  tem momento de ordem  $k$ , chamamos  $E[(X - b)^k]$  de **momento de ordem  $k$**  em torno de  $b, \forall b \in \mathbb{R}$ . Em particular, se  $k = 2$  e  $b = EX$ , temos

$$E[(X - EX)^2] := \text{Var}(X) = \sigma_X^2 \quad \text{e} \quad \sqrt{\text{Var}(X)} := \sigma_X$$

a **variância** e o **desvio-padrão** de  $X$ , respectivamente. É comum denotar também  $EX = \mu$ .

**Proposição 4.8** (Propriedades dos momentos). Os momentos de uma v.a. satisfazem as seguintes propriedades.

- (P1) Se  $X$  é v.a. com momento de ordem  $r$ , então  $X$  tem momento de ordem  $k$  para todo  $k \leq r$ .
- (P2) Se  $X$  e  $Y$  são v.a.'s com momento de ordem  $r$ , então  $X + Y$  também tem momento de ordem  $r$  e, de maneira geral, se  $X_1, \dots, X_n$  são v.a.'s com momento de ordem  $r$ , então  $X_1 + \dots + X_n$  também tem momento de ordem  $r$  e, se  $E[X^2] < \infty$ , então  $\text{Var}(X) < \infty$ .
- (P3) Seja  $X$  v.a. com segundo momento finito. Então
- (i)  $\text{Var}(X) \geq 0$
  - (ii)  $\text{Var}(X) = E[X^2] - EX^2$
  - (iii)  $P(X = c) = 1 \Leftrightarrow \text{Var}(X) = 0$
  - (iv)  $\text{Var}(X + c) = \text{Var}(X), \forall c \in \mathbb{R}$
  - (v)  $\text{Var}(cX) = c^2 \text{Var}(X), \forall c \in \mathbb{R}$
  - (vi) Se  $Y$  é v.a. com segundo momento finito, então  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2E[(X - EX)(Y - EY)] < \infty$ .

### Demonstração.

(P1) Da definição de esperança, temos  $E[X^r] = \sum_x x^r p_X(x) < \infty$  e essa série converge absolutamente. Logo, podemos derivá-la, obtendo os demais momentos.

(P2) Note que  $E[Y^r] < \infty \implies E[(-Y)^r] = (-1)^r E[Y^r] < \infty$ , logo  $-Y$  tem momento de ordem  $k$ . Note que

$$E[X^2] < \infty \implies EX < \infty \implies E[(EX)^2] = EX^2,$$

logo  $EX$  tem segundo momento finito; como  $X$  e  $EX$  têm segundo momento finito, então  $X - EX$  também tem e  $\text{Var}(X) < \infty$ .

(P3) (i) Temos

$$\text{Var}(X) = E[(X - EX)^2] \geq 0$$

pois  $(X - EX)^2$  é não negativa.

(ii) Temos

$$\begin{aligned} \text{Var}(X) &= E[(X - EX)^2] = E[X^2 - 2XEX + EX^2] \\ &= E[X^2] - 2EXEX + EX^2 \\ &= E[X^2] - EX^2. \end{aligned}$$

(iii) Note que

$$P(X = c) = 1 \implies P(X^2 = c^2) = 1 \implies E[X^2] = c^2 \therefore \text{Var}(X) = c^2 - c^2 = 0.$$

Reciprocamente,

$$\text{Var}(X) = 0 \implies E[(X - EX)^2] = 0 \implies \sum_i \underbrace{(x_i - EX)^2}_{\geq 0} \underbrace{P(X = x_i)}_{> 0} = 0$$

$$\implies x_i = EX, i = 1, 2, \dots$$

$$\implies EX = c \text{ é o único valor possível de } X$$

$$\therefore P(X = c) = 1.$$

(iv) Temos

$$\begin{aligned}\text{Var}(X + c) &= E[(X + c - E[X + c])^2] = E[(X + c - EX - c)^2] \\ &= E[(X - EX)^2] \\ &= \text{Var}(X), \forall c \in \mathbb{R}.\end{aligned}$$

(v) Temos

$$\text{Var}(cX) = E[(cX - E[cX])^2] = E[c^2(X - EX)^2] = c^2\text{Var}(X), \forall c \in \mathbb{R}.$$

(vi) Temos

$$\begin{aligned}\text{Var}(X + Y) &= E[(X + Y - E[X + Y])^2] \\ &= E[(X + Y - EX - EY)^2] \\ &= E[(X - EX)^2 + (Y - EY)^2 + 2(X - EX)(Y - EY)] \\ &= \text{Var}(X) + \text{Var}(Y) + 2E[(X - EX)(Y - EY)].\end{aligned}$$

□

**Exemplo.** Seja  $X \sim B(n, p)$ . Temos

$$\begin{aligned}E[X^2] &= \sum_{k=0}^n k^2 \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^n [k(k-1) + k] \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^n k(k-1) \binom{n}{k} p^k (1-p)^{n-k} + \sum_{k=1}^n k \binom{n}{k} p^k (1-p)^{n-k} \\ &= np + n(n-1)p^2 \sum_{k=2}^n \binom{n-2}{k-2} p^{k-2} (1-p)^{n-2-(k-2)} \\ &= np + n(n-1)p^2 \\ &= np(1-p) + n^2p^2 \therefore \text{Var}(X) = n^2p^2 + np(1-p) - n^2p^2 = np(1-p).\end{aligned}$$

**Exemplo.** Seja  $X \sim \text{Geom}(p)$ . Temos

$$\begin{aligned}E[X^2] &= \sum_{k=1}^{\infty} k^2 p(1-p)^{k-1} = p \sum_{k=1}^{\infty} [(k-1) + 1]^2 (1-p)^{k-1} \\ &= \sum_{k=1}^{\infty} (k-1)^2 p(1-p)^{k-1} + \sum_{k=1}^{\infty} 2(k-1)p(1-p)^{k-1} + \sum_{k=1}^{\infty} p(1-p)^{k-1} \\ &= (1-p)E[X^2] + 2(1-p)EX + 1.\end{aligned}$$

Daí,

$$E[X^2] = \frac{2-p}{p^2} \therefore \text{Var}(X) = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

**Definição 4.9** (Covariância). Sejam  $X, Y$  v.a.'s em  $(\Omega, \mathcal{A}, P)$  com  $E[X^2], E[Y^2] < \infty$ . A **covariância** de  $X$  e  $Y$  é definida por

$$\text{Cov}(X, Y) = E[(X - EX)(Y - EY)] = \sigma_{X,Y}.$$

**Proposição 4.10** (Propriedades da covariância). Sejam  $X, Y, X_1, \dots, X_n, Y_1, \dots, Y_m$  v.a.'s com segundo momento finito. Temos

- (i)  $\text{Cov}(X, Y) = E[XY] - EXEY$
- (ii)  $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- (iii)  $\text{Cov}(X, X) = \text{Var}(X)$
- (iv)  $\text{Cov}(cX, Y) = c\text{Cov}(X, Y) = \text{Cov}(X, cY), \forall c \in \mathbb{R}$



$$(v) \operatorname{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m \operatorname{Cov}(X_i, Y_j)$$

**Demonstração.**

(i) Temos

$$\operatorname{Cov}(X, Y) = E[(X - EX)(Y - EY)] = E[XY - XEY - YEX + EXEY] = E[XY] - EXEY.$$

(ii) Note que

$$\operatorname{Cov}(X, Y) = E[XY] - EXEY = E[YX] - EYEX = \operatorname{Cov}(Y, X).$$

(iii) Temos

$$\operatorname{Cov}(X, X) = E[X^2] - EX^2 = \operatorname{Var}(X).$$

(iv) Basta notar que

$$\operatorname{Cov}(cX, Y) = E[cXY] - E[cX]EY = c\operatorname{Cov}(X, Y) = \operatorname{Cov}(X, cY).$$

(v) Por fim, temos

$$\begin{aligned} \operatorname{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) &= E\left[\sum_{i=1}^n \sum_{j=1}^m X_i Y_j\right] - E\left[\sum_{i=1}^n X_i\right] E\left[\sum_{j=1}^m Y_j\right] \\ &= \sum_{i=1}^n \sum_{j=1}^m E[X_i Y_j] - EX_i EY_j \\ &= \sum_{i=1}^n \sum_{j=1}^m \operatorname{Cov}(X_i, Y_j). \end{aligned}$$

□

**Proposição 4.11.** Sejam  $X_1, \dots, X_n$  v.a.'s com esperança finita. Então

$$\operatorname{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \operatorname{Var}(X_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \operatorname{Cov}(X_i, X_j).$$

**Demonstração.** Usando a proposição acima, temos

$$\operatorname{Var}\left(\sum_{i=1}^n X_i\right) = \operatorname{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^n X_j\right) = \sum_{i=1}^n \sum_{j=1}^n \operatorname{Cov}(X_i, X_j) = \sum_{i=1}^n \operatorname{Var}(X_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \operatorname{Cov}(X_i, X_j).$$

□

**Observação.** Note que se  $X$  e  $Y$  são independentes, então  $\operatorname{Cov}(X, Y) = E[XY] - EXEY = 0$  e, se  $X_1, \dots, X_n$  são independentes, então  $\operatorname{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \operatorname{Var}(X_i)$ .

**Definição 4.12 (Correlação).** Se  $\operatorname{Cov}(X, Y) = 0$ , então  $X$  e  $Y$  são ditas **não-correlacionadas**.

Note, em particular, que se  $X$  e  $Y$  são **não-correlacionadas**, então  $\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y)$ . Ademais, se  $X$  e  $Y$  são **independentes** então elas são não-correlacionadas, mas a recíproca é **falsa!** Intuitivamente,  $X$  e  $Y$  serem independentes significa que não existe **nenhuma relação** entre as duas v.a.'s; por outro lado, como veremos mais à frente, as duas v.a.'s serem não-correlacionadas significa que não existe **relação linear** entre ambas.

**Exemplo.** Seja  $S_n \sim \text{Hgeo}(N, m, n)$ . Vimos que  $S_n$  é a soma das variáveis indicadoras  $X_i$ , com  $EX_i = n/N$  e  $ES_n = mn/N$ . Como  $X_i^2 = X_i$ , então  $E[X_i^2] = n/N$  e  $\operatorname{Var}(X_i) = n/N(1 - n/N)$ . Ademais, dados  $1 \leq i < j \leq m$ , então

$$E[X_i X_j] = P(X_i = 1, X_j = 1) = \frac{n}{N} \cdot \frac{n-1}{N-1}.$$

Logo,

$$\text{Cov}(X_i, X_j) = \frac{n}{N} \cdot \frac{n-1}{N-1} - \frac{n^2}{N^2} = \frac{n}{N} \cdot \frac{n-N}{N(N-1)}$$

e, daí,

$$\begin{aligned} \text{Var}(S_n) &= \sum_{i=1}^m \text{Var}(X_i) + 2 \sum_{i=1}^{m-1} \sum_{j=i+1}^m \text{Cov}(X_i, X_j) \\ &= \frac{mn}{N} \left(1 - \frac{n}{N}\right) + 2 \sum_{i=1}^{m-1} (m-i) \frac{n}{N} \cdot \frac{n-N}{N(N-1)} \\ &= \frac{mn}{N} \left(1 - \frac{n}{N}\right) + 2 \left[ \frac{mn}{N} \cdot \frac{n-N}{N(N-1)} (m-1) + \frac{n}{N} \cdot \frac{n-N}{N(N-1)} \cdot \frac{m(m-1)}{2} \right] \\ &= \frac{mn}{N} \cdot \frac{(N-n)(N-m)}{N(N-1)}. \end{aligned}$$

**Definição 4.13** (Coeficiente de correlação). Sejam  $X, Y$  v.a.'s em  $(\Omega, \mathcal{A}, P)$  com variância finita e positiva. O **coeficiente de correlação** entre  $X$  e  $Y$  é definido por

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$

**Observação.** Note que  $X, Y$  são não-correlacionadas se, e só se,  $\rho(X, Y) = 0$ . Além disso, se  $X, Y$  são independentes, então  $\rho(X, Y) = 0$ .

**Teorema 4.14** (Desigualdade de Cauchy-Schwarz). Sejam  $X, Y$  v.a.'s com segundo momento finito. Então

$$E[|XY|] \leq \sqrt{E[X^2]E[Y^2]},$$

com igualdade se, e só se,  $P(Y = aX) = 1$ , para algum  $a \in \mathbb{R}$ .

**Demonstração.** Se  $E[X^2] = 0$  ou  $E[Y^2] = 0$ , então  $P(X = 0) = 1$  ou  $P(Y = 0) = 1$ , de modo que  $P(XY = 0) = 1$  e  $E[XY] = 0 = E[X^2]E[Y^2]$ . Se  $E[X^2], E[Y^2] > 0$ , temos

$$\begin{aligned} E \left[ \left( \frac{|X|}{\sqrt{E[X^2]}} - \frac{|Y|}{\sqrt{E[Y^2]}} \right)^2 \right] &\geq 0 \\ \Leftrightarrow 2 - 2 \frac{E[|XY|]}{\sqrt{E[X^2]E[Y^2]}} &\geq 0 \\ \Leftrightarrow E[|XY|] &\leq \sqrt{E[X^2]E[Y^2]}. \end{aligned}$$

A igualdade ocorre se, e só se,

$$\begin{aligned} P \left( \frac{|X|}{\sqrt{E[X^2]}} - \frac{|Y|}{\sqrt{E[Y^2]}} = 0 \right) = 1 &\Leftrightarrow P \left( |Y| = \frac{\sqrt{E[Y^2]}}{\sqrt{E[X^2]}} |X| \right) = 1 \\ &\Leftrightarrow P(Y = aX) = 1, a = \pm \sqrt{\frac{E[Y^2]}{E[X^2}}} \in \mathbb{R}. \end{aligned}$$

□

**Proposição 4.15.** Sejam  $X, Y$  v.a.'s com variância finita e positiva. Então

- (i)  $|\rho(X, Y)| \leq 1$
- (ii)  $|\rho(X, Y)| = 1 \Leftrightarrow P(Y = aX + b) = 1$ , para algum par  $a, b \in \mathbb{R}$ . Nesse caso,  $\rho(X, Y) = 1$  se  $a > 0$  e  $\rho(X, Y) = -1$  se  $a < 0$ .

### Demonstração.

(i) Temos

$$|E[(X - \mu_X)(Y - \mu_Y)]| \leq E[|(X - \mu_X)(Y - \mu_Y)|] \leq \sqrt{E[(X - \mu_X)^2]E[(Y - \mu_Y)^2]},$$

ou seja,

$$|\text{Cov}(X, Y)| \leq \sigma_X \sigma_Y \Leftrightarrow |\rho(X, Y)| \leq 1.$$

(ii) Se  $\rho(X, Y) = \pm 1$ , então

$$\text{Cov}(X, Y) = \pm \sigma_X \sigma_Y$$

$$\Rightarrow E[(X - \mu_X)(Y - \mu_Y)] = \pm \sqrt{E[(X - \mu_X)^2]E[(Y - \mu_Y)^2]}$$

$$\Rightarrow P(Y - \mu_Y = a(X - \mu_X)) = 1,$$

ou seja,

$$P(Y = aX + b) = 1, a = \pm \sqrt{\frac{E[(Y - \mu_Y)^2]}{E[(X - \mu_X)^2]}}, b = \mu_Y - a\mu_X.$$

Se  $P(Y = aX + b) = 1$ , então  $\mu_Y = a\mu_X + b$  e  $\sigma_Y^2 = a^2\sigma_X^2$ . Logo,

$$\text{Cov}(X, Y) = E[XY] - \mu_X\mu_Y = E[X(aX + b) - \mu_X(a\mu_X + b)] = a\sigma_X^2,$$

de modo que

$$\rho(X, Y) = \frac{a\sigma_X^2}{\sigma_X|a|\sigma_X} = \frac{a}{|a|} = \begin{cases} 1, & a > 0 \\ -1, & a < 0 \end{cases}.$$

□

**Observação.** Como mencionamos, note que  $\rho(X, Y)$  é uma medida do grau de dependência **linear** entre  $X$  e  $Y$ :  $\rho(X, Y) \approx \pm 1$  indica alta linearidade entre  $X$  e  $Y$ , com  $\rho(X, Y) > 0$  indicando que  $Y$  cresce quando  $X$  cresce e  $\rho(X, Y) < 0$  indicando que  $Y$  decresce quando  $X$  cresce;  $\rho(X, Y) \approx 0$  indica ausência/fraca de linearidade entre  $X$  e  $Y$ . Se  $X$  e  $Y$  são independentes, então elas são não-correlacionadas pois não existe nenhuma relação (em particular, nenhuma relação linear) entre  $X$  e  $Y$ . Entretanto, a recíproca é **falsa!** Por exemplo,

$$P(Y = X^2) = 1 \Rightarrow \begin{cases} \rho(X, Y) = 0 \\ X, Y \text{ não independentes} \end{cases}.$$

Observe que, nesse caso, a dependência entre  $X$  e  $Y$  não é linear.

## 4.3 Exercícios - esperança e funções de v.a.'s discretas

1. Seja  $X$  uma v.a. com função de distribuição dada por:

$$F_X(x) = \begin{cases} 0, & x < -2 \\ 1/8, & -2 \leq x < 1 \\ 5/8, & 1 \leq x < 2 \\ 7/8, & 2 \leq x < 4 \\ 1, & x \geq 4 \end{cases}.$$

Determine:

- a função de probabilidade de  $X$ ;
- $EX$ .

### Solução.

a) Temos

$$p_X(x) = \begin{cases} 1/8, & x = -2, 4 \\ 1/2, & x = 1 \\ 1/4, & x = 2 \end{cases}$$

b) Temos

$$EX = -2/8 + 4/8 + 1/2 + 2/4 = 5/4.$$

□

2. Seja  $X$  uma v.a. com função de probabilidade dada por:

$$p_X(x) = \begin{cases} \frac{1}{2^{|x|}(|x|+1)}, & x = \pm 1, \pm 2, \dots \\ 0, & \text{c.c.} \end{cases}$$

a) Calcule  $\sum_{x \in \mathbb{Z}} |x|p_X(x)$ .

b) Mostre que  $EX$  não existe.

**Solução.**

a) Temos

$$\sum_{x \in \mathbb{Z}} |x|p_X(x) = \frac{1}{2} \sum_{x \in \mathbb{Z}} \frac{1}{|x|+1} = \infty.$$

b) Como a série acima diverge, não existe  $EX$ .

□

3. Seja  $N$  um número inteiro positivo e seja  $p$  a função definida por

$$p = \begin{cases} \frac{2x}{N(N+1)}, & x \in \{1, 2, \dots, N\} \\ 0, & \text{c.c.} \end{cases}$$

a) Mostre que  $p$  é uma função de probabilidade.

b) Seja  $X$  v.a. com f.p.  $p$ ; calcule  $EX$ .

**Solução.**

(a) Temos  $p(x) \geq 0$  para todo  $x \in \mathbb{R}$ , o conjunto  $\{1, 2, \dots, N\}$  é finito e

$$\sum_{x=1}^N \frac{2x}{N(N+1)} = 1.$$

(b) Temos

$$\sum_x |x|p(x) = \sum_{x=1}^N \frac{2x^2}{N(N+1)} = \frac{2N+1}{3}.$$

□

4. Suponha que  $X$  se distribui uniformemente em  $\{1, \dots, N\}$ . Determine  $EX$  e  $E[X^2]$ .

**Solução.** Temos

$$EX = \sum_{x=1}^N \frac{x}{N} = \frac{N+1}{2}$$

e

$$EX^2 = \sum_{x=1}^N \frac{x^2}{N} = \frac{(N+1)(2N+1)}{6}.$$

□

5. Suponha que  $X$  tem distribuição binomial de parâmetros  $n = 4$  e  $p$ . Obtenha  $E[\sin(\pi X/2)]$ .

**Solução.** Temos

$$E[\sin(\pi X/2)] = \sum_{x=0}^3 \sin(\pi x/2) \binom{4}{x} p^x (1-p)^{4-x} = 4p(1-p)(1-2p).$$

□

6. Suponha que  $X$  tem distribuição de Poisson de parâmetro  $\lambda$ . Determine  $E\left[\frac{1}{1+X}\right]$ .

**Solução.** Temos

$$E[1/(1+X)] = \sum_{x=0}^{\infty} \frac{1}{1+x} e^{-\lambda} \frac{\lambda^x}{x!} = \frac{e^{-\lambda}}{\lambda} (e^{\lambda} - 1) = \frac{1 - e^{-\lambda}}{\lambda}.$$

□

7. Seja  $X$  uma variável aleatória com distribuição geométrica de parâmetro  $p$  e seja  $M > 0$  um

número inteiro positivo. Determine a esperança das seguintes v.a.'s:

- $Z = \min(X, M)$
- $W = \max(X, M)$ .

**Solução.**

a) A esperança é

$$\sum_{x=1}^{\infty} |\min(x, M)| p(1-p)^{x-1} = \sum_{x=1}^{\infty} \min(x, M) p(1-p)^{x-1} = \frac{1 - (1-p)^M}{p}$$

após algumas simplificações.

b) A esperança é

$$\sum_{x=1}^{\infty} |\max(x, M)| p(1-p)^{x-1} = \sum_{x=1}^{\infty} \max(x, M) p(1-p)^{x-1} = M + \frac{(1-p)^M}{p}$$

após algumas simplificações. □

8. Em ensaios de Bernoulli independentes, com probabilidade  $p$  de sucesso, sejam  $X$  o número de ensaios até a ocorrência do  $r$ -ésimo sucesso e  $Y$  o número de fracassos anteriores ao  $r$ -ésimo sucesso. Determine  $EX$  e  $EY$ .

**Solução.** Note que  $X = \sum_{i=1}^r T_i$ , sendo  $T_i \sim \text{Geo}(p)$ ,  $i = 1, 2, \dots, r$ . Daí,  $EX = r/p$  e, além disso,  $Y = X - r$ , ou seja,  $Y = \sum_{i=1}^r T_i - r$ . Logo,  $EY = r/p - r = r(1-p)/p$ . □

9. Seja  $(X, Y)$  um vetor aleatório com função de probabilidade conjunta dada por:

$$p_{X,Y}(x, y) \begin{cases} p, & x = \pm 1, y = 0 \\ 1 - 2p, & x = 0, y = 1 \\ 0, & \text{c.c.} \end{cases},$$

onde  $0 < p < 1/2$ . Verifique que  $E[XY] = EXEY$ , mas  $X$  e  $Y$  não são independentes.

**Solução.** Temos

$$E(XY) = \sum_{x=-1}^1 \sum_{y=0}^1 xyp_{X,Y}(x, y) = 0.$$

Ademais,  $EXEY = 0(1-2p) = 0 = E(XY)$ . Ora, mas  $P(X = 0, Y = 1) = 1 - 2p \neq (1 - 2p)^2 = P(X = 0)P(Y = 1)$ , logo  $X$  e  $Y$  não são independentes. □

## 4.4 Exercícios - variância e covariância

1. Suponha que  $X$  se distribui uniformemente em  $\{1, \dots, N\}$ . Determine  $\text{Var}(X)$ .

**Solução.** Já sabemos que  $EX = (N + 1)/2$  e  $EX^2 = (N + 1)(2N + 1)/6$ , logo  $\text{Var}(X) = (N^2 - 1)/12$ . □

2. Considere a seguinte função:

$$p(x) = \begin{cases} x^{-(r+2)}/c, & x \in \mathbb{N} \\ 0, & \text{c.c.} \end{cases},$$

onde  $c$  é um número real positivo e  $r$  um número inteiro positivo.

- a) Mostre que  $\sum_{x=1}^{\infty} x^{-(r+2)}$  converge. Conclua que  $p$  é uma função de probabilidade com  $c =$

$$\sum_{x=1}^{\infty} x^{-(r+2)}.$$

- b) Seja  $X$  uma v.a. com função de probabilidade  $p$ . Mostre que  $E[X^r]$  é finito, mas  $X$  não tem nenhum momento de ordem maior do que  $r$ .

**Solução.**

a) A convergência dessa série é garantida pelo teste da integral. Portanto, como o produto

de  $1/c$  por essa série deve ser igual a 1, segue que  $c$  é igual a essa série.

b) Temos

$$E(X^r) = \frac{1}{c} \sum_{x=1}^{\infty} \frac{1}{x^2} < \infty,$$

logo  $X$  tem momento de ordem  $r$ . Note que se  $k > r$ , então a série que aparece em  $E(X^k)$  diverge pois é uma  $p$ -série com  $p = -k + r_2 \leq 1$ .

□

3. Em ensaios de Bernoulli independentes, com probabilidade  $p$  de sucesso, sejam  $X$  o número de ensaios até a ocorrência do  $r$ -ésimo sucesso e  $Y$  o número de fracassos anteriores ao  $r$ -ésimo sucesso. Determine  $\text{Var}(X)$  e  $\text{Var}(Y)$ .

**Solução.** Temos

$$\begin{aligned} E(X^2) &= \sum_{n=r}^{\infty} n^2 \binom{n-1}{r-1} p^r (1-p)^{n-r} \\ &= \frac{r}{p} E(Z-1) \\ &= \frac{r}{p} \left( \frac{r+1}{p} - 1 \right), \end{aligned}$$

sendo  $Z \sim BN(r+1, p)$ . Daí,

$$\text{Var}(X) = \frac{r(1-p)}{p^2}.$$

Como  $Y = X - r$ , temos  $\text{Var}(Y) = \text{Var}(X)$ .

□

4. Suponha que  $X$  e  $Y$  são duas v.a.'s independentes tais que  $E[X^4] = 2$ ,  $E[X^2] = 1$ ,  $E[Y^2] = 1$  e  $EY = 0$ . Determine  $\text{Var}(X^2Y)$ .

**Solução.** Como  $X$  e  $Y$  são independentes, temos  $\text{Var}(X^2Y) = E(X^4)E(Y^2) - E(X^2)^2E(Y)^2 = 2$ .

□

5. Sejam  $X_1, \dots, X_n$  v.a.'s i.i.d. com média  $\mu$  e variância  $\sigma^2$  e seja  $\bar{X} = S_n/n$ , onde  $S_n = X_1 + \dots + X_n$ . (Se  $X_1, \dots, X_n$  têm função de distribuição  $F$ , dizemos que eles são uma amostra aleatória de tamanho  $n$  da v.a.  $X$ , cuja função de distribuição é  $F$ , e  $\bar{X}$  é chamada média amostral.) Mostre que:

- $E[\bar{X}] = \mu$
- $\text{Var}(\bar{X}) = \sigma^2/n$
- $E\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] = (n-1)\sigma^2$ .

**Solução.**

a) Temos

$$E(\bar{X}) = E(S_n/n) = n\mu/n = \mu.$$

b) Temos

$$\text{Var}(\bar{X}) = \text{Var}(S_n/n) = \text{Var}(S_n)/n^2 = \sigma^2/n.$$

c) Temos

$$\begin{aligned} E\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] &= \sum_{i=1}^n E[(X_i - \bar{X})^2] \\ &= n\sigma^2 + 2n\mu^2 + \sigma^2 - \frac{2}{n} \sum_{i=1}^n E[X_i(X_1 + \dots + X_n)] \\ &= (n-1)\sigma^2. \end{aligned}$$

□

6. Suponha que tenhamos dois baralhos de  $n$  cartas, cada um com as cartas numeradas de 1 a  $n$ . Utilizando-se estas cartas forma-se  $n$  pares, de tal forma que cada par contendo uma carta de cada baralho. Dizemos que ocorre um encontro na posição  $i$  se o par  $i$  é constituído de cartas de

mesmo número. Seja  $S_n$  o número de encontros. Determine:

- $E[S_n]$
- $\text{Var}(S_n)$

[Sugestões: para cada  $i = 1, 2, \dots, n$ , considere a v.a.  $X_i$  definida por  $X_i = 1$  se ocorre um encontro na  $i$ -ésima posição e  $X_i = 0$  caso contrário. Assim,  $S_n = X_1 + \dots + X_n$ ; use os seguintes resultados: para cada  $i, j = 1, 2, \dots, n$ , tem-se que  $P(X_i = 1) = 1/n$  e  $P(X_i = 1, X_j = 1) = 1/n(n-1)$  se  $i \neq j$ .]

**Solução.**

- Segue da linearidade da esperança que  $E(S_n) = 1$ .
- Usando que  $E(X_i X_j) = 1/n(n-1)$  para todo  $i \neq j$ , temos que

$$\text{Var}(S_n) = 2 \cdot \frac{n(n-1)}{2} \cdot \frac{1}{n(n-1)} = 1.$$

□

- Sejam  $X_1, X_2$  e  $X_3$  variáveis aleatórias independentes tendo variâncias finitas e positivas  $\sigma_1^2, \sigma_2^2$  e  $\sigma_3^2$ , respectivamente. Obtenha a correlação entre  $X_1 - X_2$  e  $X_2 + X_3$ .

**Solução.** Temos

$\text{Cov}(X_1 - X_2, X_2 + X_3) = E[(X_1 - X_2 - E(X_1 - X_2))(X_2 + X_3 - E(X_2 + X_3))] = -\sigma_2^2$  após várias simplificações. Daí, como  $\text{Var}(X_1 - X_2) = \sigma_1^2 + \sigma_2^2$  e  $\text{Var}(X_2 + X_3) = \sigma_2^2 + \sigma_3^2$ , obtemos

$$\rho(X_1 - X_2, X_2 + X_3) = -\frac{\sigma_2^2}{\sqrt{(\sigma_1^2 + \sigma_2^2)(\sigma_2^2 + \sigma_3^2)}}.$$

□

- Suponha que  $X$  e  $Y$  são duas v.a.'s tais que  $\rho(X, Y) = 1/2$ ,  $\text{Var}(X) = 1$  e  $\text{Var}(Y) = 2$ . Obtenha  $\text{Var}(X - 2Y)$ .

**Solução.** Temos  $1/2 = \text{Cov}(X, Y)/\sqrt{2}$ , logo  $\text{Cov}(X, Y) = \sqrt{2}/2$ . Daí, segue que

$$\text{Var}(X - 2Y) = E[(X - 2Y)^2] = \text{Var}(X) + 4\text{Var}(Y) - 4\text{Cov}(X, Y) = 9 - 2\sqrt{2}.$$

□

- Uma caixa contém 3 bolas vermelhas e 2 pretas. Extraí-se uma amostra sem reposição de tamanho dois. Sejam  $U$  e  $V$  os números de bolas vermelhas e pretas, respectivamente, na amostra. Determine  $\rho(U, V)$ .

**Solução.** Temos  $U \sim \text{Hgeo}(5, 3, 2)$  e  $V \sim \text{Hgeo}(5, 2, 2)$ . Daí,  $\text{Var}(U) = 9/25$  e  $\text{Var}(V) = 9/25$ . Ademais,  $\text{Cov}(U, V) = -9/25$  e, portanto,  $\rho(U, V) = -1$ .

□

- Suponha que uma caixa contém 3 bolas numeradas de 1 a 3. Selecciona-se, ao acaso e sem reposição, duas bolas da caixa. Sejam  $X$  o número da primeira bola e  $Y$  o número da segunda bola. Determine  $\text{Cov}(X, Y)$  e  $\rho(X, Y)$ .

**Solução.** Temos  $E(XY) = 11/3$  e  $E(X)E(Y) = 4$ . Logo,  $\text{Cov}(X, Y) = -1/3$ . Como  $\text{Var}(X) = 2/3 = \text{Var}(Y)$ , segue que  $\rho(X, Y) = -1/2$ .

□

## 5 Variáveis aleatórias contínuas

Vimos situações em que as v.a.'s representavam o número de “objetos” ou “coisas”. Entretanto, há muitas situações (tanto teóricas quanto práticas) em que a v.a. natural a se considerar é “contínua” num certo sentido, e.g. o tempo que decorre até a recuperação completa de um paciente com determinada doença.

**Definição 5.1** (V.a. contínua). Uma v.a.  $X$  em  $(\Omega, \mathcal{A}, P)$  é dita **contínua** se  $P(X = x) = 0, \forall x \in \mathbb{R}$ .

Recordando das propriedades da f.d. de uma v.a., temos o seguinte fato.

**Proposição 5.2.**  $X$  é v.a. contínua se, e só se,  $F_X$  é contínua.

**Demonstração.** Dado  $x \in \mathbb{R}$ ,

$$0 = P(X = x) = F_X(x) - F_X(x^-) \Leftrightarrow F_X(x^+) = F_X(x) = F_X(x^-).$$

□

No caso de v.a.'s contínuas, podemos trocar os sinais  $<$  e  $\leq$  à vontade nos cálculos de probabilidades.

**Exemplo.** Considere o experimento de escolher um ponto ao acaso no círculo de centro na origem e raio  $R > 0$  (jogar um dardo). Vimos que, nesse caso, um espaço de probabilidade adequado é  $(\Omega, \mathcal{A}, P)$  com

$$\Omega = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq R^2\}, \mathcal{A} = \mathcal{B}^2(\Omega)$$

e  $P : \mathcal{A} \rightarrow \mathbb{R}$  tal que

$$P(A) = \iint_A \frac{1}{\pi R^2} dx dy, \forall A \in \mathcal{A}.$$

Podemos definir a v.a.  $X : \Omega \rightarrow \mathbb{R}$  por  $X(\omega) = X((x, y)) = \sqrt{x^2 + y^2}$ . Note que, dado  $z \in \mathbb{R}$ , temos  $\{X = z\} = \emptyset$  se  $z < 0$  ou  $z > R$  e  $\{X = z\} = \{(x, y) \in \Omega : x^2 + y^2 = z^2\}$  se  $0 \leq z \leq R$ . Logo,  $P(X = z) = 0, \forall z \in \mathbb{R}$ , ou seja,  $X$  é v.a. contínua. Por outro lado,  $\{X \leq z\} = \emptyset$  se  $z < 0$ ,  $\{X \leq z\} = \{(x, y) \in \Omega : x^2 + y^2 \leq z^2\}$  se  $0 \leq z < R$  e  $\{X \leq z\} = \Omega$  se  $z \geq R$ . Logo,

$$F_X(z) = \begin{cases} 0, & z < 0 \\ z^2/R^2, & 0 \leq z < R \\ 1, & z \geq R \end{cases}.$$

Em particular, se  $0 \leq a < b \leq R$ , então

$$P(a < X < b) = F_X(b) - F_X(a) = \frac{b^2 - a^2}{R^2} > 0.$$

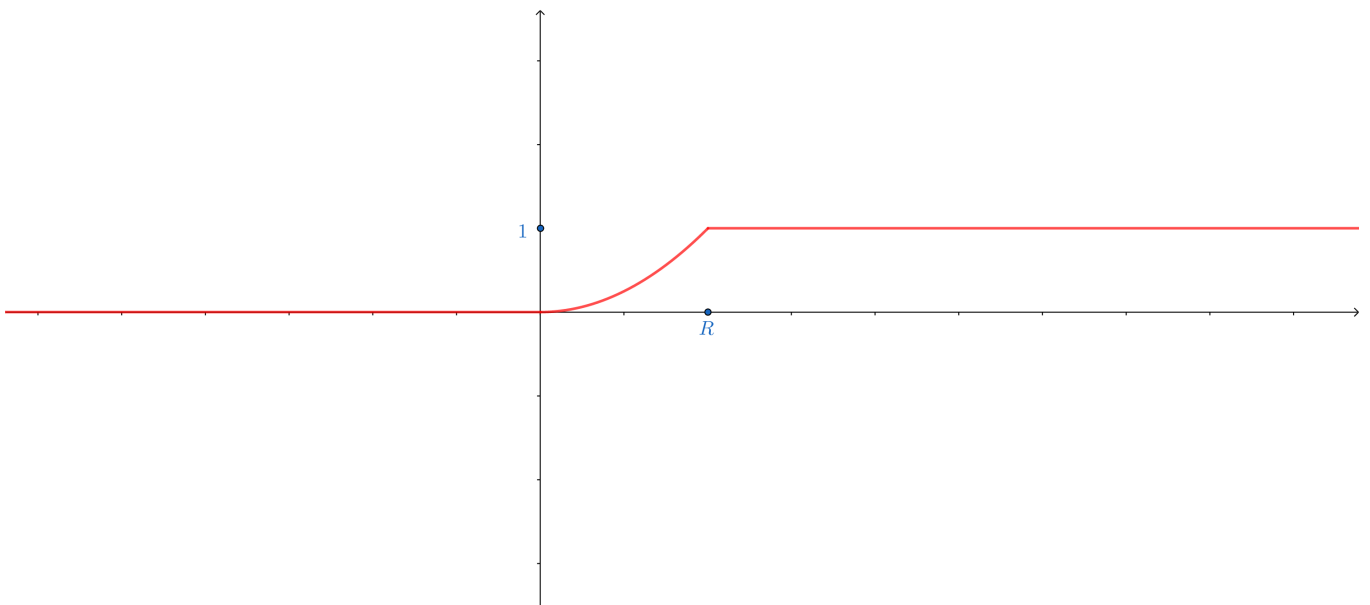


Figura 3: Gráfico da f.d. anterior

**Definição 5.3** (Função de densidade). Uma função  $f : \mathbb{R} \rightarrow \mathbb{R}$  tal que

(i)  $f(x) \geq 0, \forall x \in \mathbb{R}$

(ii)  $\int_{\mathbb{R}} f(x) dx = 1$

é dita **função de densidade de probabilidade** ou apenas **densidade**.



**Definição 5.4** (Função de distribuição). Uma função  $F : \mathbb{R} \rightarrow \mathbb{R}$  tal que

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

para alguma densidade  $f$  é dita **função de distribuição (absolutamente) contínua**. Dizemos ainda que  $f$  é a densidade de  $F$ .

**Observação.** É possível, mas complicado, construir exemplos de funções  $F$  que sejam contínuas mas não tenham densidade. As que têm densidade são chamadas de **absolutamente contínuas**. Aqui não faremos distinção, pois os casos não absolutamente contínuos são raros; sempre que nos referirmos a uma f.d., estará implícito que ela é absolutamente contínua.

Outro ponto importante é que dada  $F$ , a densidade  $f$  não é única, já que  $F$  pode não ser derivável. Contudo, os pontos onde  $F$  não é derivável (ou onde  $f$  não é contínua) formam um conjunto enumerável, de maneira que a integral não se altera. Em geral é comum tomar  $f$  como

$$f(x) = \begin{cases} F'(x), \forall x \in \mathbb{R} : \exists F'(x) \\ 0, \text{ c.c.} \end{cases}$$

**Definição 5.5.** Uma v.a.  $X$  definida em  $(\Omega, \mathcal{A}, P)$  é **(absolutamente) contínua** se sua f.d.  $F_X$  é **(absolutamente) contínua**, i.e., se  $F_X(x) = \int_{-\infty}^x f(t)dt, \forall x \in \mathbb{R}$  para alguma função de densidade  $f := f_X$ , chamada **densidade** de  $X$ .

**Observação.** Como no caso discreto, podemos nos referir tanto a  $F_X$  quanto a  $f_X$  quando dizemos “distribuição”, devido à relação biunívoca entre ambas as funções. Além disso, se  $X$  é v.a. contínua com densidade  $f_X$ , então dados  $a, b \in \mathbb{R}$  quaisquer com  $a \leq b$ , temos

$$P(a < X < b) = F_X(b) - F_X(a) = \int_a^b f_X(x)dx,$$

ou seja,  $P(a < X < b)$  é a área da região  $A$ .

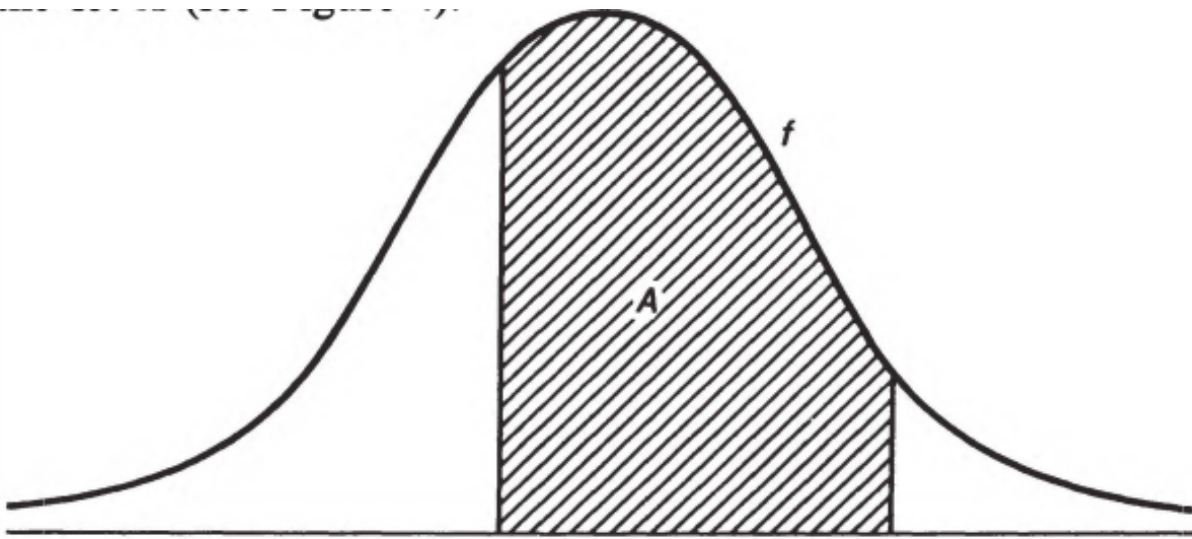


Figura 4: Gráfico obtido de [2].

**Exemplo.** No exemplo do dardo acima, vimos que

$$F_X(x) = \begin{cases} 0, x < 0 \\ x^2/R^2, 0 \leq x < R \\ 1, x \geq R \end{cases} .$$

Logo, a densidade  $f$  de  $X$  é

$$f_X(x) = \begin{cases} 0, & x < 0 \\ 2x/R^2, & 0 \leq x < R \\ 0, & x \geq R \end{cases} .$$

Note que não existe  $F'_X(R)$ , pois  $F'_{X-}(R) = 2/R \neq 0 = F'_{X+}(R)$ , ou seja,  $f$  não é contínua em  $R$ . Assim, sem perda de generalidade, tomamos  $f(R) = 0$ , de modo que

$$f_X(x) = \begin{cases} 2x/R^2, & 0 \leq x < R \\ 0, & \text{c.c.} \end{cases}$$

e ainda vale  $F_X(x) = \int_{-\infty}^x f_X(t)dt, \forall x \in \mathbb{R}$ .

**Observação.** É importante notar que há v.a.'s que não são nem contínuas nem discretas, chamadas **mistas**. Por exemplo, a v.a.  $X$  com f.d.  $F_X$  dada pelo gráfico abaixo não é contínua, pois  $F_X$  é descontínua em  $a$ ; entretanto,  $X$  tampouco é discreta, pois  $F_X$  não é do tipo escada.

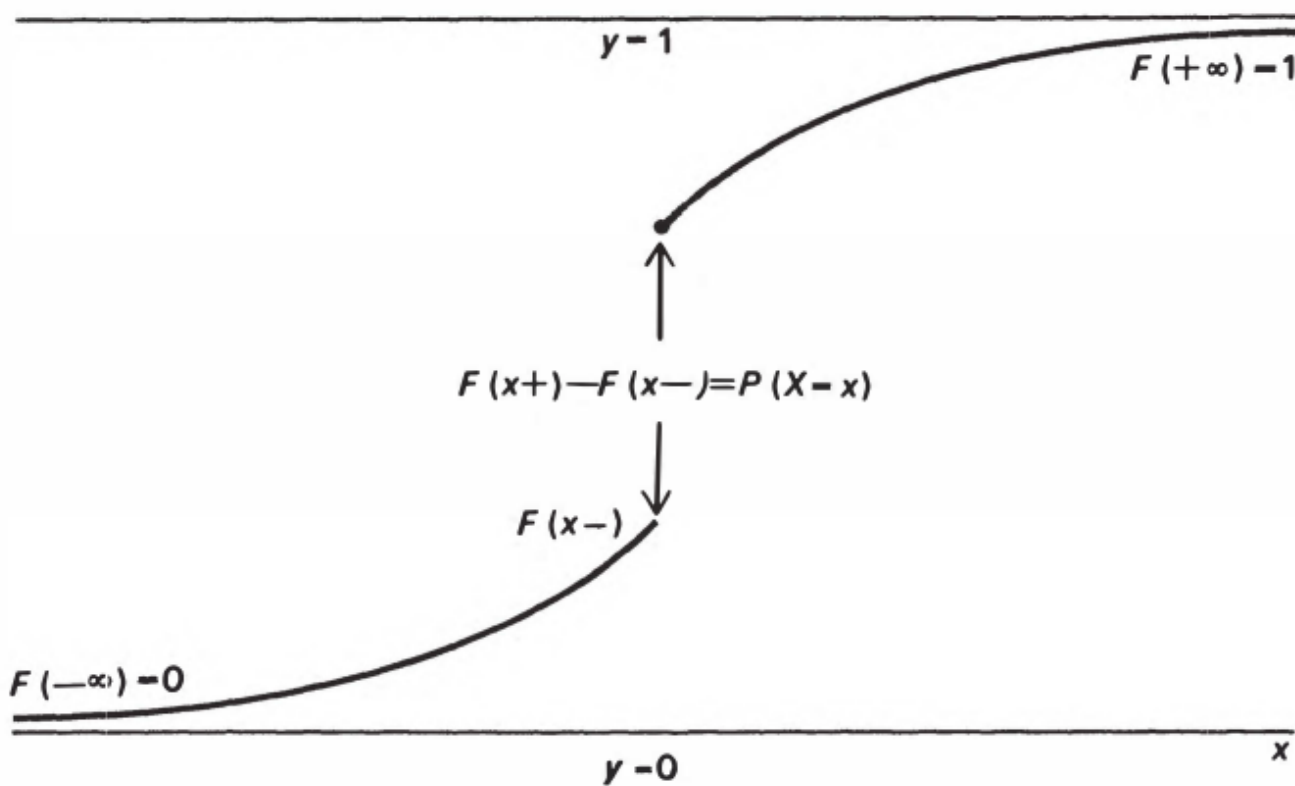


Figura 5: Gráfico obtido de [2].

**Definição 5.6.** Seja  $X$  uma v.a.; dizemos que

- (i)  $X$  é simétrica em torno de 0 se  $P(X \geq x) = P(X \leq -x), \forall x \in \mathbb{R}$ , i.e.,  $X$  e  $-X$  têm a mesma distribuição;
- (ii)  $X$  é simétrica em torno de  $\mu$  se existe  $\mu \in \mathbb{R}$  tal que  $P(X \geq \mu + x) = P(X \leq \mu - x), \forall x \in \mathbb{R}$ .

**Teorema 5.7.** Seja  $X$  v.a. contínua com densidade  $f$ . Então  $X$  é simétrica em torno de 0 se, e só se,  $f$  é par; nesse caso,  $f$  é **densidade simétrica**. De modo geral,  $X$  é simétrica em torno de  $\mu \in \mathbb{R}$  se, e só se,  $f(\mu + x) = f(\mu - x)$ . Nesse caso,  $f$  é **densidade simétrica em torno de  $\mu$** .

**Demonstração.** Provamos em torno de 0. Se  $f$  é par, então

$$P(X \geq x) = \int_x^{\infty} f(t)dt = \int_{-x}^{\infty} f(-y)dy = \int_{-x}^{\infty} f(y)dy = P(X \leq -x),$$

logo  $X$  é simétrica. Reciprocamente, se  $P(X \geq x) = P(X \geq -x)$ , defina

$$g(x) = \frac{f(x) + f(-x)}{2}.$$

Note que  $g$  é simétrica, logo

$$\begin{aligned} \int_{-\infty}^x g(y)dy &= \frac{1}{2} \int_{-\infty}^x f(y)dy + \frac{1}{2} \int_{-\infty}^x f(-y)dy \\ &= \frac{1}{2} \int_{-\infty}^x f(y)dy + \frac{1}{2} \int_{-x}^{\infty} f(y)dy \\ &= \frac{1}{2}P(X \leq x) + \frac{1}{2}P(-X \geq -x) \\ &= P(X \leq x), \end{aligned}$$

ou seja,  $X$  tem densidade  $g$  simétrica. A demonstração do caso geral é análoga, substituindo  $x$  por  $x + \mu$ .  $\square$

**Observação.** É possível mostrar que o resultado acima vale também para v.a.'s discretas; além disso, note que se  $X$  é simétrica em torno da origem então  $F(0) = 1/2$  e, se  $X$  é simétrica em torno de  $\mu$ , então  $F(\mu) = 1/2$ . De maneira geral, se  $X$  é simétrica em torno da origem então

$$F(-x) = \int_{-\infty}^{-x} f(y)dy = \int_x^{\infty} f(-y)dy = \int_x^{\infty} f(y)dy = \int_{-\infty}^{\infty} f(y)dy - \int_{-\infty}^x f(y)dy,$$

ou seja,  $F(-x) = 1 - F_X(x), \forall x \in \mathbb{R}$ .

## 5.1 Exemplos clássicos de distribuições contínuas

**Exemplo** (Uniforme contínua,  $X \sim U(a, b)$ ). Dizemos que  $X$  é uma v.a. contínua com distribuição uniforme no intervalo  $(a, b)$  se tem densidade dada por

$$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{c.c.} \end{cases}.$$

Note que  $f(x) \geq 0, \forall x \in \mathbb{R}$  e  $\int_{\mathbb{R}} f(x)dx = 1$ . A f.d.  $F$  associada a  $f$  é calculada como segue:

$$\begin{aligned} F(x) &= \int_{-\infty}^x f(t)dt = 0, x < a \\ F(x) &= \int_{-\infty}^x f(t)dt = \int_a^x \frac{1}{b-a} dt = \frac{x-a}{b-a}, a \leq x < b \\ F(x) &= \int_{-\infty}^x f(t)dt = \int_a^b \frac{1}{b-a} dt = 1, x \geq b, \end{aligned}$$

ou seja,

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x < b \\ 1, & x \geq b \end{cases}.$$

Note também que  $f(x) = F'(x), \forall x \in \mathbb{R} \setminus \{a, b\}$ .

**Exemplo** (Exponencial,  $X \sim \text{Exp}(\lambda)$ ). Dizemos que  $X$  é uma v.a. contínua com distribuição exponencial de parâmetro  $\lambda > 0$  se tem densidade dada por

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

ou, equivalentemente, f.d. dada por

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}.$$

Note que, de fato,  $f(x) \geq 0, \forall x \in \mathbb{R}$  e que

$$\int_{\mathbb{R}} f(t)dt = \int_0^{+\infty} \lambda e^{-\lambda t} dt = 1.$$

Ademais,  $f(x) = F'(x), \forall x \in \mathbb{R} \setminus \{0\}$ .

**Observação.** A distribuição exponencial é utilizada muitas vezes quando a v.a. em questão é um tempo de espera, e.g. o tempo até que um componente eletrônico apresente falhas. Além disso, se  $X \sim \text{Exp}(\lambda)$ , então  $X$  tem perda de memória, i.e.,  $P(X > a + b) = P(X > a)P(X > b), a, b \geq 0$  ou, equivalentemente,  $P(X > a + b | X > a) = P(X > b), a, b \geq 0$ .

**Demonstração.** De fato, se  $X \sim \text{Exp}(\lambda)$ , então

$$\begin{aligned} P(X > a + b | X > a) &= \frac{P(X > a + b, X > a)}{P(X > a)} \\ &= \frac{P(X > a + b)}{P(X > a)} \\ &= \frac{e^{-\lambda(a+b)}}{e^{-\lambda a}} \\ &= e^{-\lambda b} \\ &= P(X > b), \forall a, b \geq 0. \end{aligned}$$

□

Na verdade, a v.a. exponencial é a única v.a. contínua não negativa com perda de memória.

**Teorema 5.8.**  $X$  é v.a. contínua com perda de memória se, e só se,  $X \sim \text{Exp}(\lambda)$  para algum  $\lambda > 0$  ou  $P(X > 0) = 0$ .

**Demonstração.** ( $\Leftarrow$ ) Já vimos o caso que  $X$  tem distribuição exponencial; se  $P(X > 0) = 0$ , então  $P(X > a + b) = 0 = P(X > a)P(X > b), \forall a, b \geq 0$ .

( $\Rightarrow$ ) Se  $X$  tem perda de memória e  $P(X > 0) \neq 0$ , então tomando  $a = 0 = b$  segue que

$$P(X > 0) = [P(X > 0)]^2 \implies P(X > 0) = 1,$$

ou seja,  $X$  é v.a. positiva. Seja  $F$  a f.d. de  $X$  e defina  $G(x) = 1 - F(x)$ . Temos  $G$  não crescente, contínua à direita,  $G(0) = 1, G(+\infty) = 0$  e  $G(a + b) = G(a)G(b), \forall a, b > 0$ . Daí, se  $c \in \mathbb{R}_+^*$  e  $m, n \in \mathbb{R}$ , temos

$$G(c) = G(c - c/m)G(c/m) = G(c - 2c/m)[G(c/m)]^2 = \dots = G(0)[G(c/m)]^m = [G(c/m)]^m,$$

donde segue que

$$G(nc) = [G(c)]^n.$$

Além disso, temos  $0 < G(1) < 1$ . De fato, se  $G(1) = 1$  então teríamos

$$G(n) = [G(1)]^n \implies G(+\infty) = 1,$$

absurdo. Se  $G(1) = 0$ , então teríamos

$$G(1/m) = 0 \implies G(0) = G(0^+) = \lim_{m \rightarrow +\infty} G(1/m) = \lim_{m \rightarrow +\infty} [G(1)]^{1/m} = 0,$$

absurdo. Logo, como  $0 < G(1) < 1$ , podemos tomar  $G(1) = e^{-\lambda}$ , para algum  $\lambda > 0$ . Tomando  $c = 1$ , segue que  $G(1/m) = e^{-\lambda/m}, m \in \mathbb{N}$  e, fazendo  $c = 1/m$ , temos  $G(n/m) = [G(1/m)]^n = e^{-\lambda n/m}, \forall n, m \in \mathbb{N}$ . Logo,  $G(y) = e^{-\lambda y}, \forall y \in \mathbb{Q}$ . Da continuidade à direita, temos

$$G(x) = \lim_{y \rightarrow x^+, y \in \mathbb{Q}} G(y) = e^{-\lambda x}, \forall x \in \mathbb{R}_+^*,$$

pois  $\mathbb{Q}$  é denso em  $\mathbb{R}$ . Daí, segue que  $F(x) = 1 - e^{-\lambda x}, \forall x > 0$ , ou seja,  $X \sim \text{Exp}(\lambda)$ . □

**Observação.** Além de tempo de falha, variáveis exponenciais são úteis para estudar o tempo de decaimento de uma partícula radioativa e, também, no estudo de processos de Poisson e cadeias de Markov. Ademais, pensando na variável exponencial como o tempo de falha, a propriedade de perda de memória diz que dado que não houve falha até o tempo  $a$ , a probabilidade de que não haja falha nas próximas  $b$  unidades de tempo é igual à probabilidade incondicional de que não haja falha nas primeiras  $b$  unidades de tempo. Isso implica que o desgaste de uma peça de equipamento não aumenta nem diminui a probabilidade de falha em um dado intervalo de tempo.

**Exemplo (Normal/gaussiana).** Dizemos que  $X$  é v.a. com distribuição normal **padrão**,  $X \sim N(0, 1)$ , se tem densidade dada por

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, x \in \mathbb{R}.$$

**Demonstração.** Vamos mostrar que  $f$  de fato é densidade. Primeiro,  $f(x) > 0, \forall x \in \mathbb{R}$ . Ademais, vamos verificar que  $\int_{\mathbb{R}} f(x) dx = 1$ . Seja  $g(x) = e^{-x^2/2}, x \in \mathbb{R}$ . Note que  $g$  é par, contínua e não negativa. Além disso, se  $x \geq 1$  então  $0 < g(x) < e^{-x/2}$  e, então,

$$\int_1^{\infty} e^{-x^2/2} dx \leq \int_1^{\infty} e^{-x/2} dx = \lim_{a \rightarrow +\infty} \int_1^a e^{-x/2} dx = \lim_{a \rightarrow +\infty} -2e^{-x/2} \Big|_1^a = 2e^{-1/2} \in \mathbb{R},$$

logo  $\int_1^{\infty} e^{-x^2/2} dx \in \mathbb{R}$ . Como  $g$  é par, temos  $\int_1^{\infty} e^{-x^2/2} dx = \int_{-\infty}^{-1} e^{-x^2/2} dx \in \mathbb{R}$  e, como  $g$  é contínua em  $\mathbb{R}$ , temos também  $\int_{-1}^1 e^{-x^2/2} dx \in \mathbb{R}$ . Logo,  $\int_{-\infty}^{\infty} e^{-x^2/2} dx \in \mathbb{R}$ , digamos  $c$ . Segue então que

$$c^2 = \int_{-\infty}^{\infty} e^{-x^2/2} dx \int_{-\infty}^{\infty} e^{-y^2/2} dy = \iint_{\mathbb{R}^2} e^{-\frac{x^2+y^2}{2}} dx dy = \int_0^{\infty} \int_0^{2\pi} r e^{-r^2/2} dr d\theta = 2\pi \int_0^{\infty} r e^{-r^2/2} dr = 2\pi,$$

ou seja,  $c = \sqrt{2\pi}$ , pois  $c > 0$ . Logo,  $\int_{\mathbb{R}} f(x) dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} g(x) dx = 1$ . Note que  $f$  é simétrica em torno de 0 e, além disso, a f.d.  $F$  associada a  $f$  é dada por

$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt, x \in \mathbb{R}.$$

Verifica-se que  $F$  não tem uma forma fechada, mas podemos aproximar seus valores numericamente. Como mencionado antes, temos  $F(0) = 1/2 = 1 - F(0)$  e  $F(x) = 1 - F(x), \forall x \in \mathbb{R}$ . É comum também denotar  $f$  por  $\varphi$  e  $F$  por  $\Phi$  no caso da densidade e da f.d. de uma v.a. normal padrão, respectivamente.  $\square$

**Observação.** Tabelas de valores para a distribuição normal geralmente fornecem as probabilidades do tipo  $P(0 < X < a)$ . De fato, basta apenas esta probabilidade, pois  $P(X \leq a) = 1/2 + P(0 < X < a), a > 0$  e analogamente para os demais casos.

De maneira geral, dizemos que  $X$  é v.a. contínua com distribuição **normal de parâmetros**  $\mu$  e  $\sigma^2$ ,  $X \sim N(\mu, \sigma^2)$  (veremos o que esses parâmetros significam mais à frente), se tem densidade dada por

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R},$$

com  $\mu, \sigma \in \mathbb{R}$  e  $\sigma > 0$ . A verificação de que  $f$  de fato é densidade é análoga ao que fizemos acima, bastando apenas efetuar a mudança de variável  $y = \frac{x - \mu}{\sigma}$ . Note também que  $f$  é simétrica em torno de  $\mu$ .

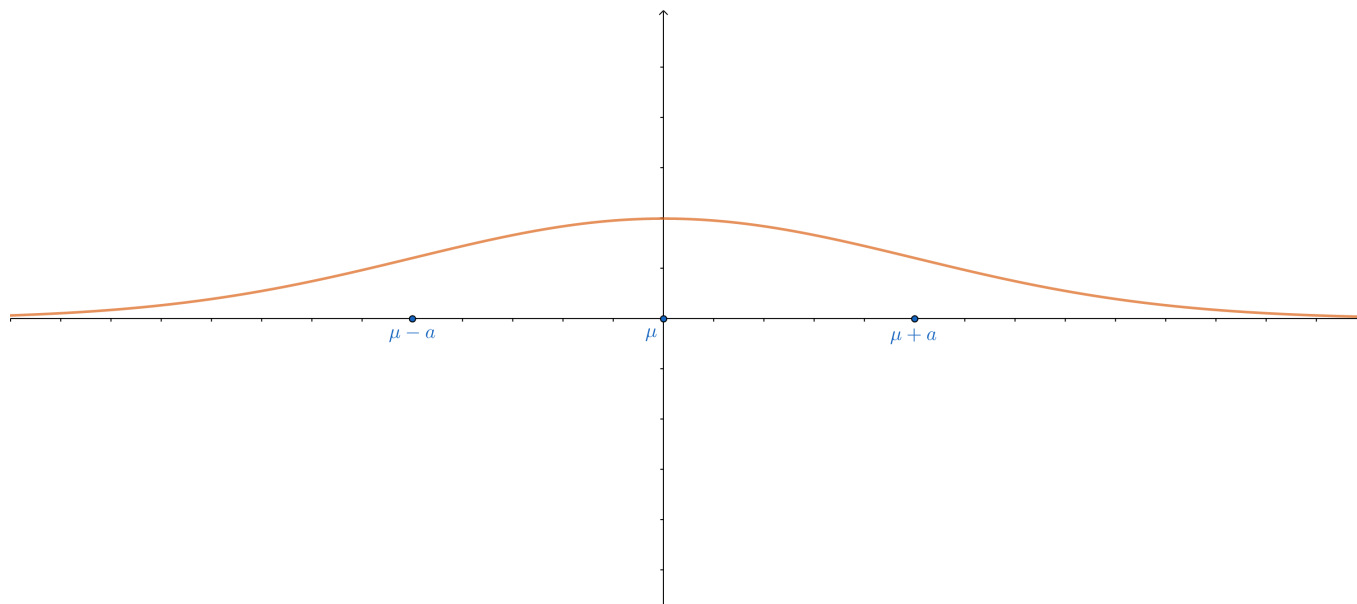


Figura 6: Gráfico obtido de [2].

As variáveis aleatórias com distribuição normal ocorrem frequentemente em aplicações práticas. A Lei de Maxwell da Física afirma que, sob condições adequadas, as componentes da velocidade de uma molécula de gás estarão aleatoriamente distribuídas seguindo uma distribuição normal  $N(0, \sigma^2)$ , onde  $\sigma^2$  depende de certas quantidades físicas. Entretanto, na maioria das aplicações, as v.a.'s de interesse têm distribuições que é **aproximadamente** normal. Por exemplo, erros instrumentais em experimentos físicos e variabilidade biológica (e.g., altura e massa) foram verificados, empiricamente, como possuindo distribuições aproximadamente normais. Veremos esse tipo de comportamento mais adiante.

**Exemplo (Gama,  $X \sim \Gamma(\alpha, \lambda)$ ).** Dizemos que  $X$  é v.a. contínua com distribuição **gama** de parâmetros  $\alpha$  e  $\lambda$  (positivos) se tem densidade dada por

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

sendo

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx, \alpha > 0$$

a **função gama**. Em particular, pode-se mostrar que  $\Gamma(\alpha) \in \mathbb{R}_+^*$  para todo  $\alpha$  real positivo, apesar de não haver uma forma fechada para a integral. Verificamos que  $f$  é densidade.

**Demonstração.** É imediato que  $f(x) \geq 0, \forall x \in \mathbb{R}$ . Ademais, temos

$$\int_0^\infty x^{\alpha-1} e^{-\lambda x} dx = \int_0^\infty y^{\alpha-1} \lambda^{1-\alpha} e^{-y} \frac{1}{\lambda} dy = \frac{1}{\lambda^\alpha} \int_0^\infty y^{\alpha-1} e^{-y} dy = \frac{\Gamma(\alpha)}{\lambda^\alpha},$$

logo

$$\int_{\mathbb{R}} f(x) dx = \int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{\lambda^\alpha} = 1.$$

□

A função gama possui algumas propriedades, muito úteis nos cálculos:

- (i)  $\Gamma(1) = \int_0^\infty e^{-x} dx = 1$
- (ii)  $\Gamma(1/2) = \int_0^\infty \frac{e^{-x}}{\sqrt{x}} dx = \sqrt{2} \int_0^\infty e^{-t^2/2} dt = \sqrt{2} \frac{1}{2} \sqrt{2\pi} = \sqrt{\pi}$
- (iii)  $\Gamma(\alpha + 1) = \int_0^\infty x^\alpha e^{-x} dx = -x^\alpha e^{-x} \Big|_0^\infty + \int_0^\infty \alpha x^{\alpha-1} e^{-x} dx = \alpha \Gamma(\alpha)$
- (iv)  $\Gamma(1) = 1$  e  $\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$  implicam  $\Gamma(n) = (n - 1)!, \forall n \in \mathbb{N}$

(v)  $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$  e  $\Gamma(1/2) = \sqrt{\pi}$  implicam

$$\Gamma(n/2) = \frac{\sqrt{\pi}}{2^{\frac{n-1}{2}}} \prod_{j=1}^{\frac{n-1}{2}} (n-2j) = \frac{\sqrt{\pi}}{2^{\frac{n-1}{2}}} \frac{1}{2^{\frac{n-1}{2}}} \frac{(n-1)!}{\left(\frac{n-1}{2}\right)!} = \frac{\sqrt{\pi}}{2^{n-1}} \frac{(n-1)!}{\left(\frac{n-1}{2}\right)!}, \forall n \in \mathbb{N} \text{ ímpar.}$$

(vi)  $X \sim \text{Exp}(\lambda) \Leftrightarrow X \sim \Gamma(1, \lambda)$ , ou seja, a distribuição exponencial é um caso particular da distribuição gama

(vii) se  $\alpha = m \in \mathbb{N}$  com  $m \geq 2$ , então a f.d.  $F$  de  $X \sim \Gamma(m, \lambda)$  tem uma forma fechada:

$$\begin{aligned} \int_0^x \frac{\lambda^m y^{m-1} e^{-\lambda y}}{(m-1)!} dy &= \frac{-(\lambda y)^{m-1} e^{-\lambda y}}{(m-1)!} \Big|_0^x + \int_0^x \frac{\lambda^{m-1} y^{m-2} e^{-\lambda y}}{(m-2)!} dy \\ &= \int_0^x \frac{\lambda^{m-1} y^{m-2} e^{-\lambda y}}{(m-2)!} dy - \frac{(\lambda x)^{m-1} e^{-\lambda x}}{(m-1)!} \\ &\vdots \\ &= 1 - \sum_{k=0}^{m-1} \frac{(\lambda x)^k}{k!} e^{-\lambda x}, x > 0, \end{aligned}$$

em que integramos por partes  $m$  vezes. Isso sugere uma conexão com  $Y \sim \text{Poisson}(\lambda x)$ . De fato, segue do fato acima que  $P(X \leq x) = P(Y \geq m)$ , e essa conexão tem aplicações em processos de Poisson.

**Observação** (Construção de funções de densidade). Seja  $g : \mathbb{R} \rightarrow \mathbb{R}$  uma função tal que  $g(x) \geq 0, \forall x \in \mathbb{R}$  e  $\int_{\mathbb{R}} g(x) dx = c \in \mathbb{R}_+^*$ . Logo, se  $f(x) := g(x)/c, \forall x \in \mathbb{R}$ , então  $f$  é densidade.

**Exemplo.** Seja  $g(x) = \frac{1}{1+x^2}, x \in \mathbb{R}$ . Temos  $g(x) \geq 0, \forall x \in \mathbb{R}$  e

$$\int_{\mathbb{R}} g(x) dx = \lim_{a \rightarrow +\infty} \left[ \arctan x \Big|_{-a}^a \right] = \pi \in \mathbb{R}_+^*.$$

Logo,  $f(x) = \frac{1}{\pi(1+x^2)}$  é densidade.

**Exemplo** (Cauchy padrão,  $X \sim \text{Cauchy}(0, 1)$ ). Dizemos que  $X$  é v.a. contínua com distribuição Cauchy de parâmetros 0 e 1 se tem densidade dada por

$$f(x) = \frac{1}{\pi(1+x^2)}, \forall x \in \mathbb{R}.$$

A f.d.  $F$  associada a  $f$  é dada por

$$F(x) = \int_{-\infty}^x f(t) dt = \frac{1}{\pi} \arctan t \Big|_{-\infty}^x = \frac{1}{2} + \frac{1}{\pi} \arctan x, \forall x \in \mathbb{R}.$$

Podemos considerar também  $X \sim \text{Cauchy}(\mu, \beta)$ , com  $\mu$  real e  $\beta$  real positivo, chamada **distribuição Cauchy** de parâmetros  $\mu$  e  $\beta$ . Essa v.a. tem densidade dada por

$$f(x) = \frac{\beta}{\pi(\beta^2 + (x - \mu)^2)}, \forall x \in \mathbb{R}.$$

Verificamos que  $f$  é densidade.

**Demonstração.** Note que  $f(x) \geq 0, \forall x \in \mathbb{R}$  e que

$$\int_{\mathbb{R}} f(x) dx = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1}{\beta} \cdot \frac{1}{1 + \left(\frac{x-\mu}{\beta}\right)^2} dx = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1}{1+y^2} dy = 1.$$

□

## 5.2 Funções de variáveis aleatórias contínuas

Sendo  $X$  v.a. contínua e  $g$  função em  $\mathbb{R}$ , estamos interessados em determinar a densidade  $f_Y$  de  $Y = g(X)$ .

**Exemplo.** Seja  $Y = X^2$ , i.e.,  $Y = g(X)$  com  $g : \mathbb{R} \rightarrow \mathbb{R}$  tal que  $g(x) = x^2$ . Note que se  $y \leq 0$ , então  $F_Y(y) = P(Y \leq y) = 0$  e, se  $y > 0$ , então  $F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$ . Logo, segue que para  $y > 0$  temos

$$f_Y(y) = F'_X(\sqrt{y}) \frac{d}{dy}(\sqrt{y}) - F'_X(-\sqrt{y}) \frac{d}{dy}(-\sqrt{y}) = \frac{1}{2\sqrt{y}} [f_X(\sqrt{y}) + f_X(-\sqrt{y})],$$

ou seja,

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} [f_X(\sqrt{y}) + f_X(-\sqrt{y})], & y > 0 \\ 0, & y \leq 0 \end{cases}.$$

Essa fórmula vale para todo  $y$  onde  $F'_X(\sqrt{y})$  existe.

**Observação.** Se  $X$  é v.a. contínua com densidade  $f_X$  e  $g$  é uma função discreta, i.e., com imagem finita ou enumerável, então  $Y = g(X)$  é v.a. discreta.

**Exemplo.** Se  $X \sim \text{Exp}(1)$  e  $Y = I_{X \leq 3}$ , i.e.,

$$Y = g(X) = \begin{cases} 1, & X \leq 3 \\ 0, & X > 3 \end{cases},$$

então os valores possíveis de  $Y$  são 0 e 1. Temos  $P(Y = 1) = P(X \leq 3) = 1 - e^{-3}$ ,  $P(Y = 0) = P(X > 3) = e^{-3}$  e  $P(Y = y) = 0, \forall y \in \mathbb{R} \setminus \{0, 1\}$ . Logo,

$$p_Y(y) = \begin{cases} e^{-3}, & y = 0 \\ 1 - e^{-3}, & y = 1 \\ 0, & \text{c.c.} \end{cases}$$

é a f.p. de  $Y$  (discreta!).

**Teorema 5.9 (Mudança de variável).** Seja  $X$  v.a. contínua com f.d.  $f_X$  tal que  $f_X(x) = 0$  para todo  $x \in I$ , sendo  $I$  um aberto da reta. Suponha que  $g : \mathbb{R} \rightarrow \mathbb{R}$  seja uma função estritamente monótona e derivável em  $I$ . Então  $Y = g(X)$  tem densidade dada por

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|, & y \in g(I) \\ 0, & y \notin g(I) \end{cases}.$$

Podemos simplificar a notação tomando  $x = g^{-1}(y)$ .

**Demonstração.** Note que  $g$  é derivável e inversível em  $I$ . Suponhamos  $g$  estritamente crescente em  $I$ . Então  $g^{-1}$  é estritamente crescente em  $g(I)$ . Daí, dado  $y \in g(I)$ , temos

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y)).$$

Logo,

$$f_Y(y) = \frac{d}{dy} [F_X(g^{-1}(y))] = f_X(g^{-1}(y)) \underbrace{\frac{d}{dy}(g^{-1}(y))}_{>0},$$

ou seja, temos

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|.$$

Se  $g$  é estritamente decrescente em  $I$ , então  $g^{-1}$  também o é e, dado  $y \in g(I)$ , temos

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y)).$$

Logo,

$$f_Y(y) = \frac{d}{dy} [1 - F_X(g^{-1}(y))] = f_X(g^{-1}(y)) \underbrace{\frac{d}{dy}(-g^{-1}(y))}_{>0},$$



ou seja, temos

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|.$$

□

**Exemplo.** Sejam  $X \sim \text{Exp}(\lambda)$  e  $Y = X^{1/\beta}$ ,  $\beta \in \mathbb{R}^*$ . Temos

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}.$$

Ademais,  $Y = g(X)$  com  $g$  função da reta dada por  $g(x) = x^{1/\beta}$ . Note que  $g$  é crescente e derivável em  $(0, +\infty)$  e também

$$y = x^{1/\beta} \Leftrightarrow x = y^\beta \implies \frac{dx}{dy} = \beta y^{\beta-1}.$$

Daí, se  $y > 0$ , temos

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right| = \lambda e^{-\lambda y^\beta} |\beta y^{\beta-1}|$$

e, portanto,

$$f_Y(y) = \begin{cases} \lambda |\beta| y^{\beta-1} e^{-\lambda y^\beta}, & y > 0 \\ 0, & y \leq 0. \end{cases}$$

**Exemplo.** Sejam  $X$  v.a. contínua com densidade  $f_X$  e  $a, b \in \mathbb{R}$  com  $b \neq 0$ . Se  $Y = a + bX$ , então

$$y = a + bx \Leftrightarrow x = \frac{y-a}{b} \implies \frac{dx}{dy} = \frac{1}{b},$$

com  $g$  função da reta dada por  $g(x) = a + bx$  estritamente monótona em  $\mathbb{R}$  (que é um aberto). Logo, dado  $y \in \mathbb{R}$ , temos

$$f_Y(y) = f_X\left(\frac{y-a}{b}\right) \left| \frac{1}{b} \right|.$$

**Proposição 5.10.** Sejam  $X$  v.a. contínua e  $\mu, \sigma \in \mathbb{R}$  com  $\sigma > 0$ . Temos

- (i) se  $Y = \frac{X - \mu}{\sigma}$ , então  $X \sim N(\mu, \sigma^2) \Leftrightarrow Y \sim N(0, 1)$
- (ii) se  $X \sim N(\mu, \sigma^2)$  e  $Y = a + bX$  com  $a, b \in \mathbb{R}$  e  $b \neq 0$ , então  $Y \sim N(a + b\mu, b^2\sigma^2)$
- (iii) se  $X \sim N(0, 1)$  e  $Y = -X$ , então  $Y \sim N(0, 1)$

**Demonstração.**

- (i) Note que  $g : \mathbb{R} \rightarrow \mathbb{R}$  dada por  $g(x) = \frac{x - \mu}{\sigma}$  é estritamente crescente e derivável na reta. Daí, temos

$$x = \sigma y + \mu \implies \frac{dx}{dy} = \sigma.$$

Logo,

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \sigma = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}, y \in \mathbb{R},$$

ou seja,  $Y \sim N(0, 1)$ . Reciprocamente, se  $Y \sim N(0, 1)$  então

$$f_X(x) = f_Y(g(x)) \left| \frac{dy}{dx} \right| = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

ou seja,  $X \sim N(\mu, \sigma^2)$ .

- (ii) Utilizando o exemplo anterior, temos

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dx}{dy} \right| = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}\left(\frac{y-a}{b} - \mu\right)^2\right] \left| \frac{1}{b} \right| = \frac{1}{\sqrt{2\pi}\sigma|b|} \exp\left[-\frac{1}{2} \frac{[y - (a + b\mu)]^2}{(\sigma b)^2}\right],$$

ou seja,  $Y \sim N(a + b\mu, b^2\sigma^2)$ .

- (iii) Note que  $g : \mathbb{R} \rightarrow \mathbb{R}$  dada por  $g(x) = -x$  é estritamente decrescente e derivável em  $\mathbb{R}$ . Ademais,

temos  $x = -y$  e  $dx/dy = -1$ , de modo que

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dx}{dy} \right| = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(-y)^2\right] |-1| = \frac{1}{\sqrt{2\pi}} e^{-y^2/2},$$

ou seja,  $Y \sim N(0, 1)$ . □

## 6 Vetores aleatórios contínuos

Começamos com as definições gerais, generalizando de maneira natural as definições vistas na seção anterior.

### 6.1 Definições gerais

**Definição 6.1** (Vetor aleatório). Sejam  $X_1, X_2, \dots, X_n$  v.a.'s absolutamente contínuas, definidas em  $(\Omega, \mathcal{A}, P)$ . Chamamos  $\bar{X} = (X_1, \dots, X_n)$  de **vetor aleatório contínuo** ( $n$ -dimensional).

Dito de outro modo,  $\bar{X}$  é vetor aleatório contínuo se sua função de distribuição  $F_{X_1, \dots, X_n}$  é dada por

$$F_{\bar{X}}(\bar{x}) = P(X_1 \leq x_1, \dots, X_n \leq x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(u_1, \dots, u_n) du_n \cdots du_1, \forall \bar{x} \in \mathbb{R}^n,$$

para alguma função  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  função de densidade  $n$ -dimensional, isto é,  $f(\bar{x}) \geq 0 \forall \bar{x} \in \mathbb{R}^n$  e  $\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(x_1, \dots, x_n) dx_1 \cdots dx_n = 1$ . Nesse caso, dizemos que  $f$  é **função de densidade** de  $\bar{X}$  e denotamos  $f := f_{\bar{X}}$ .

**Definição 6.2.** Seja  $F_{X_1, \dots, X_n}$  a f.d. de  $(X_1, \dots, X_n)$ . Para cada  $k \in \{1, \dots, n\}$ , definimos a **função de distribuição marginal** de  $X_k$  por

$$F_{X_k}(x_k) = P(X_k \leq x_k) = \lim_{x_i \rightarrow +\infty, i \neq k} F_{X_1, \dots, X_n}(x_1, \dots, x_n), x_k \in \mathbb{R}.$$

Podemos definir a f.d. marginal de pares, triplas e  $r$ -uplas de maneira inteiramente análoga.

### 6.2 Distribuições marginais e independência

**Caso bidimensional.** Seja  $(X, Y)$  um vetor aleatório contínuo com densidade conjunta  $f = f_{X, Y}$  e f.d. conjunta  $F = F_{X, Y}$ . Temos

$$F(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) dv du, \forall (x, y) \in \mathbb{R}^2$$

e também

$$P(a < X \leq b, c < Y \leq d) = \int_a^b \int_c^d f(x, y) dy dx.$$

De maneira geral,

$$P((X, Y) \in A) = \iint_A f(x, y) dx dy, \forall A \in \mathcal{B}^2(\mathbb{R}).$$

Podemos, ainda, obter as densidades marginais de  $X$  e  $Y$ . Note que

$$F_X(x) = P(X \leq x) = P(X \leq x, Y \in \mathbb{R}) = \int_{-\infty}^x \int_{-\infty}^{\infty} f(u, y) dy du.$$

Por outro lado, sabemos que

$$F_X(x) = \int_{-\infty}^x f_X(u) du.$$

Logo, como ambas as igualdades valem para todo  $x \in \mathbb{R}$ , devemos ter

$$f_X(x) = \int_{-\infty}^{\infty} f_{X, Y}(x, y) dy, \forall x \in \mathbb{R}.$$

De maneira análoga, temos

$$F_Y(y) = \int_{-\infty}^y \int_{-\infty}^{\infty} f(x, v) dx dv$$

e

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx, \forall y \in \mathbb{R}.$$

Sob algumas fracas condições de regularidades e para  $(x,y)$  ponto de continuidade de  $f$ , temos

$$\frac{\partial}{\partial y} F(x,y) = \int_{-\infty}^x \left( \frac{\partial}{\partial y} \int_{-\infty}^y f(u,v) dv \right) du = \int_{-\infty}^x f(u,y) du,$$

de modo que

$$\frac{\partial^2}{\partial x \partial y} F(x,y) = f(x,y)$$

e também, de maneira análoga,

$$\frac{\partial^2}{\partial y \partial x} F(x,y) = f(x,y).$$

**Exemplo.** Sejam  $X, Y$  v.a.'s contínuas com f.d. conjunta  $F$  dada por

$$F(x,y) = \begin{cases} 0, & x < 0 \text{ e } y < 0 \\ \frac{x}{5}(1 - e^{-y}), & 0 \leq x \leq 5 \text{ e } y \geq 0 \\ 1 - e^{-y}, & x \geq 5 \text{ e } y \geq 0 \end{cases} .$$

Note que

$$F_X(x) = \lim_{y \rightarrow \infty} F(x,y) = \begin{cases} 0, & x < 0 \\ x/5, & 0 \leq x < 5 \\ 1, & x \geq 5 \end{cases}$$

e também

$$F_Y(y) = \lim_{x \rightarrow \infty} F(x,y) = \begin{cases} 0, & y < 0 \\ 1 - e^{-y}, & y \geq 0 \end{cases} .$$

Ademais, para todo  $(x,y) \in \mathbb{R}^2$  com  $x \notin \{0,5\}$  e  $y \neq 0$ , temos

$$\frac{\partial}{\partial y} F(x,y) = \begin{cases} 0, & x < 0 \text{ e } y < 0 \\ \frac{x}{5}e^{-y}, & 0 < x < 5 \text{ e } y > 0 \\ e^{-y}, & x > 5 \text{ e } y > 0 \end{cases}$$

e também

$$\frac{\partial^2}{\partial x \partial y} F(x,y) = \begin{cases} 0, & x < 0 \text{ e } y < 0 \\ \frac{e^{-y}}{5}, & 0 < x < 5 \text{ e } y > 0 \\ 0, & x > 5 \text{ e } y > 0 \end{cases} .$$

Portanto,

$$f_{X,Y}(x,y) = \begin{cases} e^{-y}/5, & 0 < x < 5 \text{ e } y > 0 \\ 0, & \text{c.c.} \end{cases}$$

e obtemos

$$f_X(x) = \frac{d}{dx} F_X(x) = \begin{cases} 1/5, & 0 < x < 5 \\ 0, & \text{c.c.} \end{cases}$$

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \begin{cases} e^{-y}, & y > 0 \\ 0, & \text{c.c.} \end{cases} ,$$

ou seja,  $X \sim U(0,5)$  e  $Y \sim \text{Exp}(1)$ .

**Exemplo.** Seja  $D = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq R^2\}$  e considere o experimento de escolher um ponto em  $D$ . Defina  $X$  e  $Y$  como as coordenadas do ponto escolhido. Supondo uniformidade, temos que

a densidade conjunta de  $X$  e  $Y$  é dada por

$$f(x, y) = \begin{cases} \frac{1}{\pi R^2}, & (x, y) \in D \\ 0, & \text{c.c.} \end{cases}.$$

De fato, note que dado  $A \subset D$  temos

$$P((X, Y) \in A) = \iint_A f(x, y) dx dy.$$

Da hipótese de uniformidade, temos

$$P((X, Y) \in A) = \frac{\iint_A dx dy}{\pi R^2} = \iint_A \frac{1}{\pi R^2} dx dy,$$

de modo que  $f$  tem a forma acima. Para as densidades marginais, temos que dado  $-R < x < R$ ,

$$f_X(x) = \int_{\mathbb{R}} f(x, y) dy = \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \frac{1}{\pi R^2} dy = 2 \frac{\sqrt{R^2-x^2}}{\pi R^2}.$$

Logo,

$$f_X(x) = \begin{cases} 2 \frac{\sqrt{R^2-x^2}}{\pi R^2}, & -R < x < R \\ 0, & \text{c.c.} \end{cases}$$

e, de maneira análoga,

$$f_Y(y) = \begin{cases} 2 \frac{\sqrt{R^2-y^2}}{\pi R^2}, & -R < y < R \\ 0, & \text{c.c.} \end{cases}$$

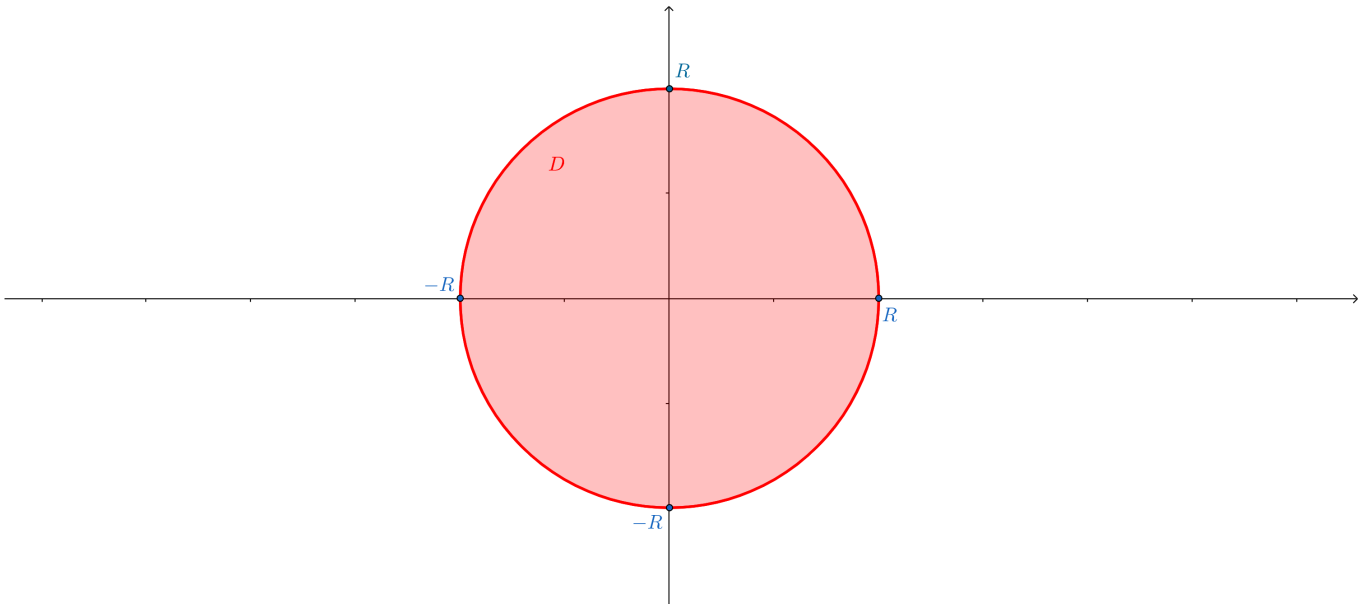


Figura 7: Gráfico da região  $D$ .

Queremos agora discutir sobre a independência de duas v.a.'s contínuas. Recorde que duas v.a.'s  $X$  e  $Y$  são ditas independentes se estão definidas no mesmo espaço  $(\Omega, \mathcal{A}, P)$  e, dados quaisquer  $A_1, A_2 \in \mathcal{B}(\mathbb{R})$ , vale  $P(X \in A_1, Y \in A_2) = P(X \in A_1)P(Y \in A_2)$ . Vimos anteriormente que  $X$  e  $Y$  são v.a.'s independentes se, e só se,  $F(x, y) = F_X(x)F_Y(y), \forall (x, y) \in \mathbb{R}^2$ . Um resultado análogo vale para a densidade.

**Proposição 6.3.** Duas v.a.'s contínuas  $X$  e  $Y$  definidas em  $(\Omega, \mathcal{A}, P)$  são independentes se, e só se,  $f_{X,Y}(x, y) = f_X(x)f_Y(y), \forall (x, y) \in \mathbb{R}^2$ .

**Demonstração.** Dado  $(x, y) \in \mathbb{R}^2$ , temos  $X$  e  $Y$  independentes se, e só se,

$$\begin{aligned} \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u, v) dv du &= F_{X,Y}(x, y) \\ &= F_X(x)F_Y(y) \\ &= \left( \int_{-\infty}^x f_X(u) du \right) \left( \int_{-\infty}^y f_Y(v) dv \right) \\ &= \int_{-\infty}^x \int_{-\infty}^y f_X(u) f_Y(v) dv du. \end{aligned}$$

Como  $(x, y)$  é um ponto arbitrário do plano, devemos ter

$$f_{X,Y}(x, y) = f_X(x)f_Y(y), \forall (x, y) \in \mathbb{R}^2.$$

□

**Exemplo.** No dois exemplos anteriores temos  $X$  e  $Y$  independentes no primeiro mas não no segundo.

**Exemplo (Construção de densidade conjunta).** Suponha  $X$  e  $Y$  v.a.'s i.i.d. com  $X \sim N(0, 1)$ . Então temos

$$f(x, y) = f_X(x)f_Y(y) = \frac{1}{2\pi} \exp\left[-\frac{x^2 + y^2}{2}\right], \forall (x, y) \in \mathbb{R}^2.$$

Pode-se verificar, de maneira análoga à densidade da distribuição normal, que esta  $f$  define uma densidade; ela é chamada **densidade normal bidimensional padrão**.

**Exemplo.** Sejam  $X$  e  $Y$  v.a.'s com densidade conjunta dada por  $f(x, y) = c \exp\left[-\frac{x^2 - xy + y^2}{2}\right]$ ,  $\forall (x, y) \in \mathbb{R}^2$ , com  $c$  real positivo tal que  $\iint_{\mathbb{R}^2} f(x, y) dx dy = 1$ . Note que esta última integral é complicada de calcular diretamente. Portanto, tomamos uma rota alternativa para encontrar  $c$ . Note que

$$f(x, y) = c \exp\left[-\frac{1}{2} \left( \left(y - \frac{x}{2}\right)^2 + \frac{3}{4}x^2 \right)\right].$$

Daí, segue que

$$f_X(x) = \int_{\mathbb{R}} f(x, y) dy = ce^{-3x^2/8} \int_{\mathbb{R}} \exp\left[-\frac{1}{2} \left(y - \frac{x}{2}\right)^2\right] dy = ce^{-3x^2/8} \int_{\mathbb{R}} e^{-u^2/2} du = c\sqrt{2\pi}e^{-3x^2/8}, \forall x \in \mathbb{R}.$$

Ora, mas  $c$  deve ser tal que  $\int_{\mathbb{R}} f_X(x) dx = 1$ , logo

$$c\sqrt{2\pi} \int_{\mathbb{R}} e^{-3x^2/8} dx = 1 \implies c\sqrt{2\pi} \frac{2}{\sqrt{3}} \int_{\mathbb{R}} e^{-u^2/2} du = 1 \implies c = \frac{\sqrt{3}}{4\pi}.$$

Logo,

$$f(x, y) = \frac{\sqrt{3}}{4\pi} \exp\left[-\frac{x^2 - xy + y^2}{2}\right], \forall (x, y) \in \mathbb{R}^2$$

e  $X \sim N(0, 4/3)$ . De maneira análoga,  $Y \sim N(0, 4/3)$  e observamos que  $X$  e  $Y$  não são independentes, pois

$$\frac{3}{8\pi} = f_X(0)f_Y(0) \neq f(0, 0) = \frac{\sqrt{3}}{4\pi}.$$

**Caso  $n$ -dimensional.** Considere  $(X_1, \dots, X_n)$  um vetor aleatório contínuo com densidade  $f = f_{X_1, \dots, X_n}$  e f.d.  $F = F_{X_1, \dots, X_n}$ . As distribuições marginais são dadas por

$$f_{X_k}(x_k) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_1 \cdots dx_{k-1} dx_{k+1} \cdots dx_n$$

e analogamente para os demais casos. Ademais,

$$F_{X_k}(x_k) = \int_{-\infty}^{x_k} f_{X_k}(u) du = \int_{-\infty}^{x_k} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_{k-1}, u, x_{k+1}, \dots, x_n) dx_1 \cdots dx_{k-1} dx_{k+1} \cdots dx_n du$$

e analogamente para os demais casos. De maneira geral, se  $A \in \mathcal{B}^n(\mathbb{R})$  então

$$P((X_1, \dots, X_n) \in A) = \int \cdots \int_A f(x_1, \dots, x_n) dx_1 \cdots dx_n.$$

Novamente sob algumas condições fracas de regularidade, podemos escrever

$$f(x_1, \dots, x_n) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F(x_1, \dots, x_n), \forall (x_1, \dots, x_n) \in \mathbb{R}^n \text{ tal que } f \text{ é contínua.}$$

Naturalmente,  $X_1, \dots, X_n$  são independentes se, e só se,  $F(x_1, \dots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n), \forall (x_1, \dots, x_n) \in \mathbb{R}^n$  ou, equivalentemente, se, e só se,  $f(x_1, \dots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n), \forall (x_1, \dots, x_n) \in \mathbb{R}^n$ .

### 6.3 Função de distribuição $n$ -dimensional

Dada  $F : \mathbb{R}^n \rightarrow \mathbb{R}$  uma função, estamos interessados em saber as condições que  $F$  deve satisfazer para ser função de distribuição em  $\mathbb{R}^n$ . Pensando na f.d. unidimensional, as três condições a seguir são naturais de serem exigidas.

(F1)  $F(x_1, \dots, x_n)$  é não decrescente em cada variável, ou seja, se  $x < y$  então

$$F(x_1, \dots, x_{k-1}, x, x_{k+1}, \dots, x_n) \leq F(x_1, \dots, x_{k-1}, y, x_{k+1}, \dots, x_n), \forall k \in \{1, \dots, n\}.$$

(F2)  $F(x_1, \dots, x_n)$  é contínua à direita em cada variável, ou seja, se  $y_m \xrightarrow{m \rightarrow \infty} x_k$  então

$$F(x_1, \dots, x_{k-1}, y_m, x_{k+1}, \dots, x_n) \xrightarrow{m \rightarrow \infty} F(x_1, \dots, x_{k-1}, x_k, x_{k+1}, \dots, x_n), \forall k \in \{1, \dots, n\}.$$

(F3) Vale

$$\lim_{x_i \rightarrow -\infty} F(x_1, \dots, x_i, \dots, x_n) := F(x_1, \dots, -\infty, \dots, x_n) = 0, \forall i = 1, 2, \dots, n$$

e também

$$\lim_{x_1 \rightarrow \infty} \cdots \lim_{x_n \rightarrow \infty} F(x_1, \dots, x_n) := F(+\infty, \dots, +\infty) = 1.$$

No entanto, apenas essas três condições não são suficientes. De fato, tome  $F : \mathbb{R}^2 \rightarrow \mathbb{R}$  tal que

$$F(x, y) = \begin{cases} 1, & x \geq 0, y \geq 0 \text{ e } x + y \geq 1 \\ 0, & \text{c.c.} \end{cases}.$$

Note que  $F$  satisfaz as três condições acima, mas supondo que  $F$  fosse f.d. de  $(X, Y)$  então teríamos

$$\begin{aligned} 0 \leq P(0 < X \leq 1, 0 < Y \leq 1) &= P(X \leq 1, 0 < Y \leq 1) - P(X \leq 0, 0 < Y \leq 1) \\ &= F(1, 1) - F(1, 0) - F(0, 1) + F(0, 0) \\ &= 1 - 1 - 1 + 0 \\ &= -1, \end{aligned}$$

absurdo. Assim, queremos que dados  $a_1 < b_1$  e  $a_2 < b_2$  reais,  $F$  seja tal que

$$0 \leq P(a_1 < X \leq b_1, a_2 < Y \leq b_2) = F(b_1, b_2) - F(b_1, a_2) - [F(a_1, b_2) - F(a_1, a_2)].$$

De maneira geral, temos a condição

(F4) Se  $I_k = (a_k, b_k], k = 1, 2, \dots, n$  e  $\Delta_{k, I_k} F(x_1, \dots, x_n)$  for definido como

$$F(x_1, \dots, x_{k-1}, b_k, x_{k+1}, \dots, x_n) - F(x_1, \dots, x_{k-1}, a_k, x_{k+1}, \dots, x_n),$$

então

$$\Delta_{1, I_1} \cdots \Delta_{n, I_n} F(x_1, \dots, x_n) \geq 0, \forall I_k = (a_k, b_k], k = 1, \dots, n.$$

Com isso, definimos

**Definição 6.4.** Uma função  $F : \mathbb{R}^n \rightarrow \mathbb{R}$  que satisfaz (F1), (F2), (F3) e (F4) é chamada **função de distribuição  $n$ -dimensional**.

**Proposição 6.5.** Dado um vetor aleatório  $(X_1, \dots, X_n)$ , a função  $F : \mathbb{R}^n \rightarrow \mathbb{R}$  dada por

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n), \forall (x_1, \dots, x_n) \in \mathbb{R}^n$$

define uma f.d. em  $\mathbb{R}^n$ .

**Demonstração.** Basta verificar (F1)-(F4) usando as propriedades da medida de probabilidade.  $\square$

### 6.4 Funções de vetores aleatórios

Sejam  $(X_1, \dots, X_n)$  vetor aleatório com densidade  $f_{X_1, \dots, X_n} = f$ , função de distribuição  $F_{X_1, \dots, X_n} = F$  e  $Z = g(X_1, \dots, X_n)$ , com  $g : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  tal que  $D \supset \text{Im}(X_1, \dots, X_n)$ . Estamos interessados na

distribuição de  $Z$ . Note que dado  $z$  real qualquer, temos

$$\{Z \leq z\} = \{(X_1, \dots, X_n) \in A_z\},$$

com

$$A_z = \{(x_1, \dots, x_n) \in \mathbb{R}^n : g(x_1, \dots, x_n) \leq z\}.$$

Logo,

$$F_Z(z) = P(Z \leq z) = P((X_1, \dots, X_n) \in A_z) = \int \cdots \int_{A_z} f(x_1, \dots, x_n) dx_1 \cdots dx_n = \int_{-\infty}^z \varphi(v) dv,$$

em que queremos verificar quando ocorre a última igualdade; no caso de ser válida,  $\varphi = f_Z$  é a densidade de  $Z$ . Nos restringiremos aqui ao caso  $n = 2$ .

**Distribuição da soma.** Sejam  $X$  e  $Y$  v.a.'s contínuas com densidade conjunta  $f$  e distribuição conjunta  $F$ . Seja  $Z = \varphi(X, Y)$  com  $\varphi : D \subset \mathbb{R}^2 \rightarrow \mathbb{R}$  tal que  $D$  contém a imagem de  $(X, Y)$ . Dado  $z \in \mathbb{R}$  fixo, temos

$$\{Z \leq z\} = \{(X, Y) \in A_z\},$$

com

$$A_z = \{(x, y) \in \mathbb{R}^2 : \varphi(x, y) \leq z\}.$$

Então

$$F_Z(z) = P(Z \leq z) = P((X, Y) \in A_z) = \iint_{A_z} f(x, y) dx dy.$$

Considere  $Z = X + Y$ , de modo que

$$A_z = \{(x, y) \in \mathbb{R}^2 : x + y \leq z\}.$$

Daí,

$$\begin{aligned} F_Z(z) &= \iint_{A_z} f(x, y) dx dy = \int_{\mathbb{R}} \int_{-\infty}^{z-x} f(x, y) dy dx \\ &= \int_{\mathbb{R}} \int_{-\infty}^z f(x, v-x) dv dx \\ &= \int_{-\infty}^z \underbrace{\int_{\mathbb{R}} f(x, v-x) dx}_{f_Z(v)} dv, \end{aligned}$$

ou seja,

$$f_{X+Y}(z) = \int_{\mathbb{R}} f(x, z-x) dx$$

ou, analogamente,

$$f_{X+Y}(z) = \int_{\mathbb{R}} f(z-y, y) dy.$$

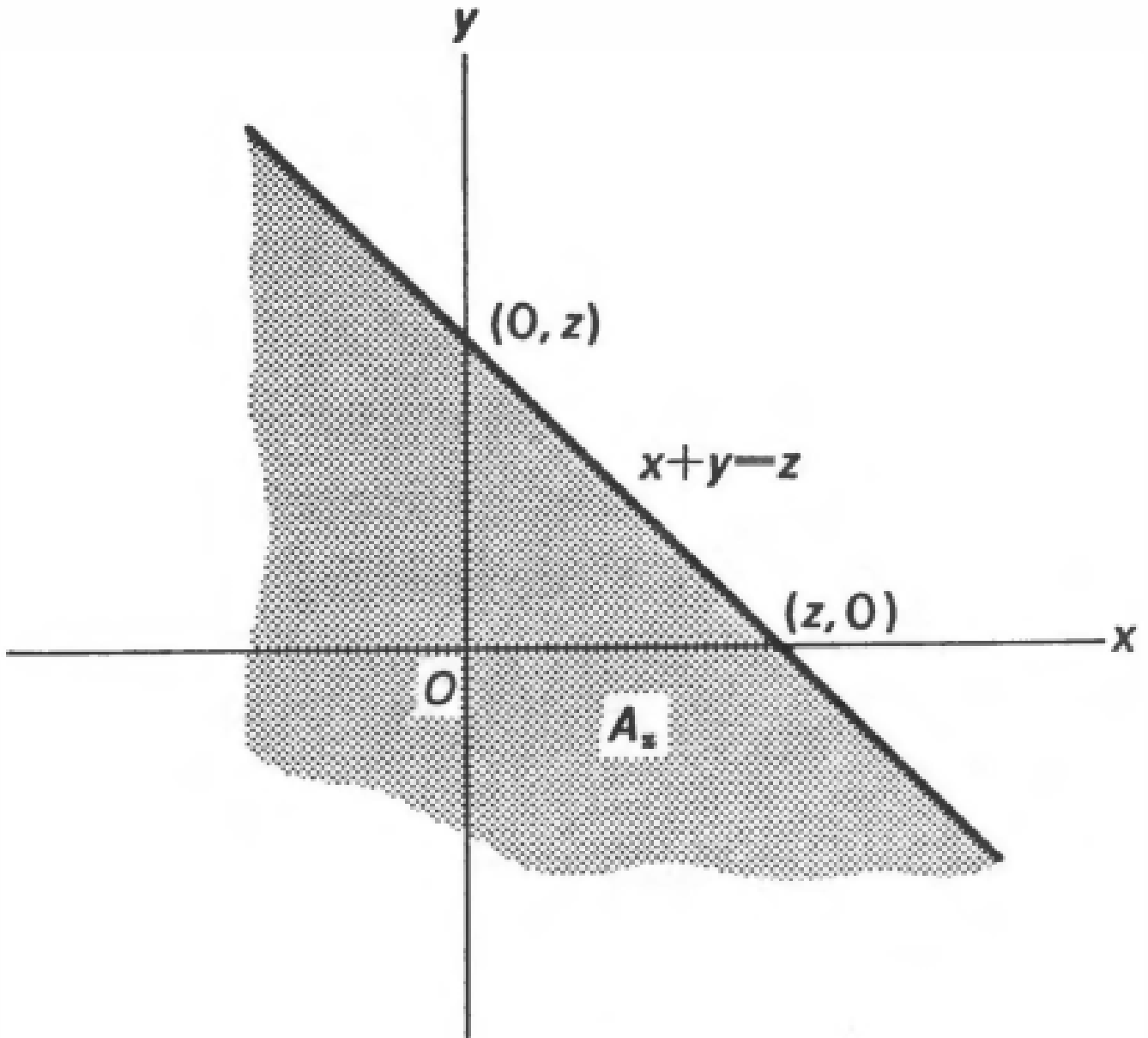


Figura 8: Imagem obtida de [2].

Em particular, se  $X$  e  $Y$  são duas v.a.'s independentes e não negativas, então

$$f_{X+Y}(z) = \begin{cases} \int_0^z f_X(x)f_Y(z-x)dx, z > 0 \\ 0, \text{ c.c.} \end{cases}.$$

Nesse caso,  $f_{X+Y}$  é a **convolução** de  $f_X$  e  $f_Y$ , denotada por  $f_X * f_Y$ .

**Exemplo.** Sejam  $X, Y$  v.a.'s i.i.d. com distribuição comum  $\text{Exp}(\lambda)$ . Note que  $Z = X + Y$  é v.a. não negativa, de modo que  $f_Z(z) = 0$  se  $z \leq 0$ . Para  $z > 0$ , temos

$$f_Z(z) = \int_0^z f_X(x)f_Y(z-x)dx = \int_0^z \lambda e^{-\lambda x} \lambda e^{-\lambda(z-x)} dx = \lambda^2 e^{-\lambda z} \int_0^z dx = \lambda^2 z e^{-\lambda z}.$$

Logo,

$$f_Z(z) = \begin{cases} \lambda^2 z e^{-\lambda z}, z > 0 \\ 0, \text{ c.c.} \end{cases},$$

ou seja,  $Z \sim \Gamma(2, \lambda)$ .

Esse exemplo tem uma generalização importante, enunciada como segue.



**Teorema 6.6.** Sejam  $X$  e  $Y$  v.a.'s independentes tais que  $X \sim \Gamma(\alpha_1, \lambda)$  e  $Y \sim \Gamma(\alpha_2, \lambda)$ . Então  $X + Y \sim \Gamma(\alpha_1 + \alpha_2, \lambda)$ .

**Demonstração.** Como  $X$  e  $Y$  são v.a.'s não negativas, segue que  $f_{X+Y}(z) = 0, z \leq 0$ . Para  $z > 0$ , temos

$$\begin{aligned} f_{X+Y}(z) &= \frac{\lambda^{\alpha_1+\alpha_2} e^{-\lambda z}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^z x^{\alpha_1-1} (z-x)^{\alpha_2-1} dx \\ &= \frac{\lambda^{\alpha_1+\alpha_2} z^{\alpha_1+\alpha_2-1} e^{-\lambda z}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^1 u^{\alpha_1-1} (1-u)^{\alpha_2-1} du \\ &= \frac{\lambda^{\alpha_1+\alpha_2} z^{\alpha_1+\alpha_2-1} e^{-\lambda z}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)} \\ &= \frac{\lambda^{\alpha_1+\alpha_2} z^{\alpha_1+\alpha_2-1} e^{-\lambda z}}{\Gamma(\alpha_1 + \alpha_2)}, \end{aligned}$$

em que usamos que

$$\int_0^1 u^{\alpha_1-1} (1-u)^{\alpha_2-1} du = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)}.$$

Logo,  $X + Y \sim \Gamma(\alpha_1 + \alpha_2, \lambda)$ . □

**Exemplo (Beta,  $X \sim \text{Beta}(\alpha_1, \alpha_2)$ ).** A integral que usamos no argumento acima nos permite definir uma nova família de densidades, a dois parâmetros, chamadas **densidades betas**. Dizemos que  $X$  tem densidade beta de parâmetros  $\alpha_1$  e  $\alpha_2$  (positivos) se sua densidade é dada por

$$f_X(x) = \begin{cases} \frac{\Gamma(\alpha_1+\alpha_2)x^{\alpha_1-1}(1-x)^{\alpha_2-1}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}, & 0 < x < 1 \\ 0, & \text{c.c.} \end{cases}$$

A nomenclatura “beta” vem do fato de que a função

$$B(\alpha_1, \alpha_2) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)}, \quad 0 < \alpha_1, \alpha_2 < \infty$$

é chamada **função beta**.

**Exemplo.** Sejam  $X, Y$  i.i.d. com distribuição uniforme em  $(0, 1)$  e  $Z = X + Y$ . Note que  $P((X, Y) \in (0, 2)) = 1$ , de modo que  $f_{X+Y}(z) = 0$  se  $z < 0$  ou  $z > 2$ . Se  $0 \leq z \leq 1$ , então

$$f_{X+Y}(z) = \int_0^z f_X(x)f_Y(z-x)dx = \int_0^z dx = z.$$

Se  $1 < z \leq 2$ , então

$$f_{X+Y}(z) = \int_0^z f_X(x)f_Y(z-x)dx = \int_{z-1}^1 dx = 2 - z.$$

Logo,

$$f_{X+Y}(z) = \begin{cases} z, & 0 \leq z \leq 1 \\ 2 - z, & 1 < z \leq 2 \\ 0, & \text{c.c.} \end{cases}.$$

Antes de passarmos à distribuição de quocientes, enunciamos um teorema importante acerca da soma de distribuições normais.

**Teorema 6.7.** Sejam  $X$  e  $Y$  v.a.'s independentes com  $X \sim N(\mu_1, \sigma_1^2)$  e  $Y \sim N(\mu_2, \sigma_2^2)$ . Então  $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ .

**Demonstração.** Assumamos  $\mu_1 = \mu_2 = 0$ . Então

$$\begin{aligned} f_{X+Y}(z) &= \frac{1}{2\pi\sigma_1\sigma_2} \int_{\mathbb{R}} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_1^2} + \frac{(z-x)^2}{\sigma_2^2}\right)\right] dx \\ &= \frac{1}{2\pi\sqrt{\sigma_1^2 + \sigma_2^2}} \int_{\mathbb{R}} \exp\left[-\frac{1}{2}\left(u^2 - \frac{2uz\sigma_1}{\sigma_2\sqrt{\sigma_1^2 + \sigma_2^2}} + \frac{z^2}{\sigma_2^2}\right)\right] du \\ &= \frac{e^{-z^2/2(\sigma_1^2 + \sigma_2^2)}}{\sqrt{2\pi}\sqrt{\sigma_1^2 + \sigma_2^2}} \int_{\mathbb{R}} \frac{e^{-v^2/2}}{\sqrt{2\pi}} dv \\ &= \frac{e^{-z^2/2(\sigma_1^2 + \sigma_2^2)}}{\sqrt{2\pi}\sqrt{\sigma_1^2 + \sigma_2^2}}, \end{aligned}$$

em que fizemos a troca de variável

$$u = \frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{\sigma_1\sigma_2}x,$$

completamos quadrados

$$u^2 - \frac{2uz\sigma_1}{\sigma_2\sqrt{\sigma_1^2 + \sigma_2^2}} + \frac{z^2}{\sigma_2^2} = \left(u - \frac{z\sigma_1}{\sigma_2\sqrt{\sigma_1^2 + \sigma_2^2}}\right)^2 + \frac{z^2}{\sigma_1^2 + \sigma_2^2}$$

e fizemos a segunda mudança de variável

$$v = u - \frac{z\sigma_1}{\sigma_2\sqrt{\sigma_1^2 + \sigma_2^2}}.$$

Notamos, então, que  $X + Y \sim N(0, \sigma_1^2 + \sigma_2^2)$ . Daí, de modo geral,  $X - \mu_1$  e  $Y - \mu_2$  também são independentes com densidades normais de parâmetros 0 e  $\sigma_1^2$  e 0 e  $\sigma_2^2$ , respectivamente. Daí,  $X + Y - (\mu_1 + \mu_2)$  tem densidade normal de parâmetros  $\mu_1 + \mu_2$  e  $\sigma_1^2 + \sigma_2^2$ , como desejado.  $\square$

**Distribuição do quociente.** Sejam  $X$  e  $Y$  v.a.'s com densidade conjunta  $f$  e distribuição conjunta  $F$ . Estamos agora interessados em encontrar a distribuição de  $Z = X/Y$ . Note que dado  $z \in \mathbb{R}$  fixo, temos

$$\{Z \leq z\} = \{(X, Y) \in A_z\},$$

com

$$A_z = \{(x, y) \in \mathbb{R}^2 : y/x \leq z\} = \{(x, y) \in \mathbb{R}^2 : x < 0, y \geq xz\} \cup \{(x, y) \in \mathbb{R}^2 : x > 0, y \leq xz\}.$$

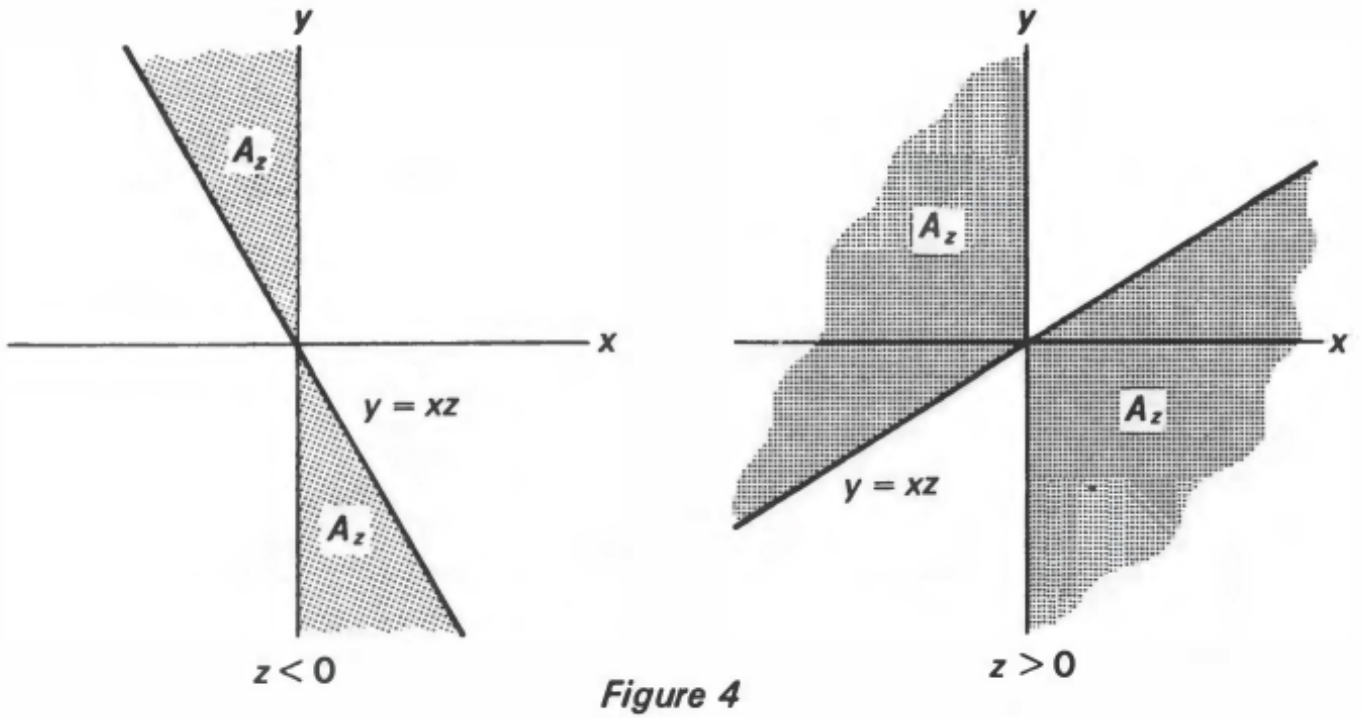


Figure 9: Imagem obtida de [2].

Daí, segue que

$$\begin{aligned}
 F_Z(z) = P(Z \leq z) &= P((X, Y) \in A_z) = \iint_{A_z} f(x, y) dx dy \\
 &= \int_{-\infty}^0 \int_{xz}^{\infty} f(x, y) dy dx + \int_0^{\infty} \int_{-\infty}^{xz} f(x, y) dy dx \\
 &= \int_{-\infty}^0 \int_z^{\infty} x f(x, xu) du dx + \int_0^{\infty} \int_{-\infty}^z x f(x, xu) du dx \\
 &= \int_{-\infty}^0 \int_{-\infty}^z (-x) f(x, xu) du dx + \int_0^{\infty} \int_{-\infty}^z x f(x, xu) du dx \\
 &= \int_{\mathbb{R}} \int_{-\infty}^z |x| f(x, xu) du dx,
 \end{aligned}$$

donde segue que

$$F_{Y/X}(z) = \int_{-\infty}^z \underbrace{\int_{\mathbb{R}} |x| f(x, xu) dx}_{f_{Y/X}(u)} du \implies f_{Y/X}(z) = \int_{\mathbb{R}} |x| f_{X,Y}(x, xz) dx, \forall z \in \mathbb{R}.$$

Em particular, se  $X$  e  $Y$  são independentes e não negativas, então

$$f_{Y/X}(z) = \begin{cases} \int_0^{\infty} x f_X(x) f_Y(xz) dx, & z > 0 \\ 0, & z \leq 0 \end{cases}.$$

**Exemplo.** Sejam  $X$  e  $Y$  v.a.'s independentes tais que  $X \sim \Gamma(\alpha_1, \lambda)$  e  $Y \sim \Gamma(\alpha_2, \lambda)$ . Note que  $X$  e  $Y$  são não negativas, de modo que se  $z \leq 0$  temos  $f_{Y/X}(z) = 0$ . Se  $z > 0$ , então

$$\begin{aligned}
 f_{Y/X}(z) &= \int_0^{\infty} x f_X(x) f_Y(xz) dx = \int_0^{\infty} x \frac{\lambda^{\alpha_1} x^{\alpha_1-1} e^{-\lambda x}}{\Gamma(\alpha_1)} \frac{\lambda^{\alpha_2} (xz)^{\alpha_2-1} e^{-\lambda xz}}{\Gamma(\alpha_2)} dx \\
 &= \frac{\lambda^{\alpha_1+\alpha_2} z^{\alpha_2-1}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^{\infty} x^{\alpha_1+\alpha_2-1} e^{-x\lambda(z+1)} dx \\
 &= \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \frac{z^{\alpha_2-1}}{(z+1)^{\alpha_1+\alpha_2}},
 \end{aligned}$$

em que usamos que

$$\int_0^{\infty} x^{\alpha_1 + \alpha_2 - 1} e^{-x\lambda(z+1)} dx = \frac{\Gamma(\alpha_1 + \alpha_2)}{[\lambda(z+1)]^{\alpha_1 + \alpha_2}}.$$

**Exemplo.** Sejam  $X, Y$  v.a.'s i.i.d. com distribuição  $N(0, \sigma^2)$ . Da Lista de Exercícios 7, sabemos que  $X^2, Y^2 \sim \Gamma(1/2, 1/2\sigma^2)$ . Daí, do exemplo anterior, temos  $f_{Y^2/X^2}(z) = 0$  se  $z \leq 0$  e, para  $z > 0$ , temos

$$f_{Y^2/X^2}(z) = \frac{\Gamma(1)}{\Gamma(1/2)\Gamma(1/2)} \frac{z^{-1/2}}{z+1} = \frac{1}{\pi(z+1)\sqrt{z}}.$$

Note também que dado  $z \in \mathbb{R}$  qualquer, temos

$$\begin{aligned} f_{Y/X}(z) &= 2 \int_0^{\infty} x f_X(x) f_Y(xz) dx \\ &= 2 \frac{1}{2\pi\sigma^2} \int_0^{\infty} x \exp\left[-\frac{x^2(1+z^2)}{2\sigma^2}\right] dx \\ &= \frac{1}{\pi\sigma^2} \frac{\sigma^2}{1+z^2} \int_0^{\infty} e^{-u} du \\ &= \frac{1}{\pi(1+z^2)}, \end{aligned}$$

ou seja,  $Y/X \sim \text{Cauchy}(0, 1)$ . É possível mostrar também que  $Y/|X| \sim \text{Cauchy}(0, 1)$ .

**Observação.** Podemos também falar de densidades condicionais e da Regra da Bayes no caso de v.a.'s contínuas; entretanto, optamos por não fazê-lo e focar nas distribuições clássicas de v.a.'s contínuas. Ao leitor interessado, consulte [2] pp. 153-157.

Podemos ainda generalizar os exemplos da soma de v.a.'s gama e normal com os seguintes teoremas.

**Teorema 6.8.** Sejam  $X_1, \dots, X_n$  v.a.'s independentes. Sejam  $Y$  uma v.a. definida em termos de  $X_1, \dots, X_m$  e  $Z$  uma v.a. definida em termos de  $X_{m+1}, \dots, X_n$ , com  $1 \leq m < n$ . Então  $Y$  e  $Z$  são independentes.

A prova deste teorema envolve argumentos de Teoria da Medida e, portanto, não será apresentada; entretanto, por indução e usando este teorema pode-se demonstrar os dois teoremas a seguir.

**Teorema 6.9.** Sejam  $X_1, \dots, X_n$  v.a.'s independentes tais que  $X_m \sim \Gamma(\alpha_m, \lambda)$ , com  $m = 1, \dots, n$ . Então  $X_1 + \dots + X_n \sim \Gamma(\alpha_1 + \dots + \alpha_n, \lambda)$ .

**Corolário 6.10.** Se  $X_1, \dots, X_n$  são v.a.'s independentes com distribuição exponencial de parâmetro  $\lambda$ , então  $X_1 + \dots + X_n \sim \Gamma(n, \lambda)$ .

**Teorema 6.11.** Sejam  $X_1, \dots, X_n$  v.a.'s independentes tais que  $X_m \sim N(\mu_m, \sigma_m^2)$ , com  $m = 1, \dots, n$ . Então  $X_1 + \dots + X_n \sim N(\mu_1 + \dots + \mu_n, \sigma_1^2 + \dots + \sigma_n^2)$ .

**Distribuição de vetores aleatórios bidimensionais.** Sejam  $X$  e  $Y$  v.a.'s contínuas com densidade conjunta  $f_{X,Y}$ . Considere  $g_1 : \mathbb{R}^2 \rightarrow \mathbb{R}$  e  $g_2 : \mathbb{R}^2 \rightarrow \mathbb{R}$  e tome

$$\begin{cases} U = g_1(X, Y) \\ V = g_2(X, Y) \end{cases}.$$

Queremos determinar a distribuição conjunta de  $U$  e  $V$ . Por definição,

$$F_{U,V}(u, v) = P(U \leq u, V \leq v) = \iint_{B_{u,v}} f_{X,Y}(x, y) dx dy,$$

com  $B_{u,v} = \{(x, y) \in \mathbb{R}^2 : g_1(x, y) \leq u, g_2(x, y) \leq v\}$ . Em geral,

$$f_{U,V}(u, v) = \frac{\partial^2}{\partial u \partial v} F_{U,V}(u, v).$$

Em particular, podemos considerar

$$\begin{cases} u = g_1(x, y) \\ v = g_2(x, y) \end{cases}.$$

Supondo que  $g_1$  e  $g_2$  têm derivadas parciais contínuas em todos os pontos e que

$$J(x, y) = \begin{vmatrix} \frac{\partial}{\partial x}g_1 & \frac{\partial}{\partial y}g_1 \\ \frac{\partial}{\partial x}g_2 & \frac{\partial}{\partial y}g_2 \end{vmatrix} \neq 0,$$

então pelo teorema da função inversa podemos resolver (localmente) o sistema acima, ou seja, existem  $h_1, h_2 : \mathbb{R}^2 \rightarrow \mathbb{R}$  tais que

$$\begin{cases} x = h_1(u, v) \\ y = h_2(u, v) \end{cases}.$$

Logo,  $U$  e  $V$  são v.a.'s contínuas com densidade conjunta

$$f_{U,V}(u, v) = \frac{1}{|J(x, y)|} f_{X,Y}(x, y),$$

com  $x, y$  dados como acima. Ademais,

$$F_{U,V}(u, v) = \int_{-\infty}^u \int_{-\infty}^v f_{U,V}(z, w) dw dz.$$

**Exemplo.** Sejam  $X, Y$  i.i.d. com distribuição exponencial de parâmetro  $\lambda$  e

$$\begin{cases} U = X + Y \\ V = X - Y \end{cases}.$$

Note que  $P(U > 0) = 1$  e  $P(V \in \mathbb{R}) = 1$ . Daí,

$$\begin{cases} u = x + y \\ v = x - y \end{cases} \implies \begin{cases} x = \frac{u+v}{2} \\ y = \frac{u-v}{2} \end{cases}, u > 0 \text{ e } -u < v < u.$$

Ademais,  $J = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2$ . Logo, para  $u > 0$  e  $-u < v < u$ , temos

$$f_{U,V}(u, v) = \frac{1}{|J|} f_{X,Y} \left( \frac{u+v}{2}, \frac{u-v}{2} \right) = \frac{1}{|J|} f_X \left( \frac{u+v}{2} \right) f_Y \left( \frac{u-v}{2} \right) = \frac{\lambda^2}{2} e^{-\lambda u},$$

ou seja,

$$f_{U,V}(u, v) = \begin{cases} \frac{\lambda^2}{2} e^{-\lambda u}, u > 0 \text{ e } -u < v < u \\ 0, \text{ c.c.} \end{cases}.$$

Daí, dado  $u > 0$ , temos

$$f_U(u) = \int_{\mathbb{R}} f_{U,V}(u, v) dv = \int_{-u}^u \frac{\lambda^2}{2} e^{-\lambda u} dv = \lambda^2 u e^{-\lambda u},$$

ou seja,

$$f_U(u) = \begin{cases} \lambda^2 u e^{-\lambda u}, u > 0 \\ 0, \text{ c.c.} \end{cases} \quad \therefore U \sim \Gamma(2, \lambda).$$

Ademais, dado  $v \in \mathbb{R}$ ,

$$f_V(v) = \int_{\mathbb{R}} f_{U,V}(u, v) du = \begin{cases} \int_v^{\infty} \frac{\lambda^2}{2} e^{-\lambda u} du = \frac{\lambda}{2} e^{-\lambda v}, v > 0 \\ \int_{-v}^{\infty} \frac{\lambda^2}{2} e^{-\lambda u} du = \frac{\lambda}{2} e^{-\lambda(-v)}, v < 0 \end{cases},$$

ou seja,

$$f_V(v) = \frac{\lambda}{2} e^{-\lambda|v|}, \forall v \in \mathbb{R}.$$

**Exemplo.** Sejam  $X$  e  $Y$  v.a.'s i.i.d. exponenciais de parâmetro  $\lambda$  e  $Z = X + 3Y$ . Considere  $U = X$ . Daí,

$$\begin{cases} u = x \\ z = x + 3y \end{cases} \implies \begin{cases} x = u \\ y = \frac{z-u}{3} \end{cases}, z > 0 \text{ e } 0 < u < z.$$

Ademais,  $J = 3$ . Logo, dados  $z > 0$  e  $0 < u < z$ , temos

$$f_{U,Z}(u, z) = \frac{\lambda^2}{3} e^{-2\lambda/3} e^{-\lambda z/3}$$

e, se  $z \leq 0$  ou  $u \leq 0$  ou  $u \geq z$ ,  $f_{U,Z}(u, z) = 0$ . Daí, dado  $z > 0$ , temos

$$f_Z(z) = \int_0^z \frac{\lambda^2}{3} e^{-2\lambda/3} e^{-\lambda z/3} du = \frac{\lambda}{2} (e^{-\lambda z/3} - e^{-\lambda z})$$

e  $f_Z(z) = 0$  se  $z \leq 0$ .

**Observação.** Podemos ainda generalizar esse método do jacobiano para  $n$  variáveis, de maneira natural:

$$f_{Y_1, \dots, Y_n}(y_1, \dots, y_n) = \frac{1}{|J(x_1, \dots, x_n)|} f(x_1, \dots, x_n),$$

com os  $x_i$ 's definidos implicitamente pelos  $y_j$ 's por

$$y_i = g_i(x_1, \dots, x_n), i = 1, \dots, n,$$

e o jacobiano é a matriz  $n \times n$  generalizada naturalmente do jacobiano  $2 \times 2$ .

## 7 Esperança de variáveis aleatórias contínuas

**Definição 7.1.** Seja  $X$  v.a. contínua com densidade  $f_X$ . Definimos a **esperança** de  $X$  por

$$EX = \int_{\mathbb{R}} x f_X(x) dx,$$

desde que a integral exista, isto é, seja bem definida.

**Observação.** Dizemos que  $EX$  é bem definida quanto podemos escrever

$$EX = \int_{\mathbb{R}} x f_X(x) dx = \int_0^{\infty} x f_X(x) dx - \int_{-\infty}^0 -x f_X(x) dx$$

e quando não ocorre a indefinição  $\infty - \infty$ . Note que  $EX < \infty$  sempre que  $\int_0^{\infty} x f_X(x) dx$  e  $\int_{-\infty}^0 x f_X(x) dx$  forem finitas. Ademais, como veremos mais à frente,  $EX < \infty \Leftrightarrow E|X| < \infty$ .

**Exemplo.** Seja  $X \sim U(a, b)$ . Temos

$$\int_{\mathbb{R}} x f_X(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2} \in \mathbb{R} \therefore EX = \frac{a+b}{2}.$$

**Exemplo.** Seja  $X \sim \text{Exp}(\lambda)$ . Temos

$$\begin{aligned} \int_{\mathbb{R}} x f_X(x) dx &= \int_0^{\infty} x \lambda e^{-\lambda x} dx \\ &= \lambda \left[ -\frac{x e^{-\lambda x}}{\lambda} \Big|_0^{\infty} + \frac{1}{\lambda} \int_0^{\infty} e^{-\lambda x} dx \right] \\ &= \frac{1}{\lambda} \in \mathbb{R} \therefore EX = \frac{1}{\lambda}. \end{aligned}$$

**Exemplo.** Seja  $X \sim \Gamma(\alpha, \lambda)$ . Temos

$$\begin{aligned} \int_{\mathbb{R}} x f_X(x) dx &= \int_0^{\alpha} x \frac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx \\ &= \frac{\lambda^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha} e^{-\lambda x} dx \\ &= \frac{\lambda^{\alpha}}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}} \\ &= \frac{\alpha}{\lambda} \in \mathbb{R} \therefore EX = \frac{\alpha}{\lambda}. \end{aligned}$$

**Exemplo.** Seja  $X \sim \text{Cauchy}(0, 1)$ . Temos

$$\int_{\mathbb{R}} x f_X(x) dx = \frac{1}{\pi} \int_0^{\infty} \frac{x}{1+x^2} dx + \frac{1}{\pi} \int_{-\infty}^0 \frac{x}{1+x^2} dx = \frac{1}{2\pi} \left[ \underbrace{\int_0^{\infty} \frac{1}{u} du - \int_0^{\infty} \frac{1}{u} du}_{\infty - \infty} \right],$$

logo não existe  $EX$ .

**Observação.** Com a noção de densidade condicional, é possível também definir uma **esperança condicional**. Como não tratamos de densidades condicionais, também não trataremos de esperanças condicionais.

## 7.1 Esperança de função de variável aleatória contínua

Sejam  $X$  v.a. contínua com densidade  $f_X$  e  $Y = \varphi(X)$ , sendo  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  função. Por definição,  $Y$  tem esperança finita se, e só se,

$$\int_{\mathbb{R}} |y| f_Y(y) dy < \infty$$

e, nesse caso,

$$EY = \int_{\mathbb{R}} y f_Y(y) dy < \infty.$$

No caso de  $\varphi$  ser discreta, temos  $Y$  v.a. discreta e podemos aplicar os resultados vistos anteriormente sobre esperanças de v.a.'s discretas para calcular  $EY$ . No caso de  $\varphi$  ser tal que  $Y$  é v.a. contínua, temos que  $Y$  tem esperança finita se, e só se,

$$\int_{\mathbb{R}} |\varphi(x)| f_X(x) dx < \infty$$

e, nesse caso,

$$EY = \int_{\mathbb{R}} \varphi(x) f_X(x) dx < \infty.$$

De maneira geral, dadas  $X_1, \dots, X_n$  v.a.'s contínuas com densidade conjunta  $f_{X_1, \dots, X_n}$  e  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  função real de modo que  $Y = \varphi(X_1, \dots, X_n)$  tem densidade, então

$$EY < \infty \Leftrightarrow \int_{\mathbb{R}} \dots \int_{\mathbb{R}} |\varphi(x_1, \dots, x_n)| f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \dots dx_n < \infty$$

e, nesse caso,

$$EY = \int_{\mathbb{R}} \dots \int_{\mathbb{R}} \varphi(x_1, \dots, x_n) f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \dots dx_n < \infty.$$

**Exemplo.** Sejam  $X \sim U(0, 1)$  e  $Y = X^2$ . Temos

$$EY = E[X^2] = \int_{\mathbb{R}} x^2 f_X(x) dx = \int_0^1 x^2 dx = 1/3.$$

**Exemplo.** Sejam  $X$  e  $Y$  v.a.'s contínuas com densidade conjunta  $f_{X,Y}$  e suponha que  $EX, EY < \infty$ . Seja  $Z = X + Y$ . Temos

$$\int_{\mathbb{R}} \int_{\mathbb{R}} x f_{X,Y}(x, y) dx dy = \int_{\mathbb{R}} x \int_{\mathbb{R}} f_{X,Y}(x, y) dy dx = \int_{\mathbb{R}} x f_X(x) dx = EX < \infty.$$

Analogamente,

$$\int_{\mathbb{R}} \int_{\mathbb{R}} y f_{X,Y}(x, y) dx dy = EY < \infty.$$

Logo,

$$EZ = E[X + Y] = \int_{\mathbb{R}} \int_{\mathbb{R}} (x + y) f_{X,Y}(x, y) dx dy = \int_{\mathbb{R}} \int_{\mathbb{R}} x f_{X,Y}(x, y) dx dy + \int_{\mathbb{R}} \int_{\mathbb{R}} y f_{X,Y}(x, y) dx dy = EX + EY$$

**Proposição 7.2** (Propriedades da esperança). Sejam  $X$  e  $Y$  v.a.'s contínuas com  $E|X|, E|Y| < \infty$ . Temos

- (i)  $P(X = c) = 1 \implies EX = c, c \in \mathbb{R}$
- (ii)  $E[cX] = cEX, \forall c \in \mathbb{R}$
- (iii)  $E[X + Y] = EX + EY$
- (iv)  $|EX| \leq E|X|$

- (v) se  $X, Y$  são independentes, então  $E[XY] = EXEY$
- (vi) se  $P(X \geq Y) = 1$ , então  $EX \geq EY$ ; em particular,  $P(X = Y) = 1 \Leftrightarrow EX = EY$
- (vii) se  $P(|X| \leq M) = 1$  para algum  $M$  real positivo, então  $|EX| \leq M$ .

**Demonstração.** As demonstrações são análogas ao caso discreto. □

## 7.2 Momentos de uma variável aleatória contínua

Seja  $X$  v.a. contínua com densidade  $f_X$ .

- Definição 7.3.**
- (i) Dado  $k \in \mathbb{N}$ ,  $E[X^k]$  é o **momento de ordem  $k$**  de  $X$  ou  **$k$ -ésimo momento** de  $X$ .
  - (ii) Dado  $k \in \mathbb{N}$  e sendo  $EX < \infty$ ,  $E[(X - EX)^k]$  é o **momento central de ordem  $k$**  de  $X$  ou  **$k$ -ésimo momento central** de  $X$ . Em particular,  $\text{Var}(X) = E[(X - EX)^2]$  é a **variância** de  $X$  e  $\sigma = \sqrt{\text{Var}(X)}$  é o **desvio-padrão** de  $X$ .
  - (iii) Dado  $k \in \mathbb{N}$ ,  $E|X|^k$  é o **momento absoluto de ordem  $k$**  de  $X$  ou  **$k$ -ésimo momento absoluto** de  $X$ .

**Observação.** Se  $EX = \mu \in \mathbb{R}$ , então

$$\text{Var}(X) = \sigma^2 = \int_{\mathbb{R}} (x - \mu)^2 f_X(x) dx.$$

Daí, mostra-se que

- $\text{Var}(X) = E[X^2] - EX^2 = E[X^2] - \mu^2$
- $\sigma^2 = 0 \Leftrightarrow P(X = \mu) = 1$ , logo se  $X$  é v.a. contínua então  $\text{Var}(X) > 0$
- $E[|X|^k] < \infty \implies E[|X|^j] < \infty, \forall j \leq k$ .

**Exemplo.** Seja  $X \sim \Gamma(\alpha, \lambda)$ . Vimos que  $EX = \alpha/\lambda$ . Ademais, dado  $k \in \mathbb{N}$ , temos

$$E[X^k] = \int_0^{\infty} x^k \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^{\infty} x^{k+\alpha-1} e^{-\lambda x} dx = \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)} \frac{1}{\lambda^k} = \frac{\alpha(\alpha+1)\cdots(\alpha+k-1)}{\lambda^k}.$$

Em particular, segue daí que  $\text{Var}(X) = \alpha/\lambda^2$ . Além disso, se  $X \sim \chi^2(n)$ , isto é,  $X \sim \Gamma(n/2, 1/2)$ , então  $EX = n$  e  $\text{Var}(X) = 2n$ . Por fim, se  $\alpha = 1$ , então  $X \sim \text{Exp}(\lambda)$  e temos  $E[X^k] = k!/\lambda^k$  e  $\text{Var}(X) = 1/\lambda^2$ .

**Exemplo.** Seja  $X$  v.a. contínua com densidade simétrica  $f_X$ . Suponha que  $E[X^k] < \infty$ . Então

$$\int_0^{\infty} x^k f_X(x) dx < \infty \text{ e } \int_{-\infty}^0 x^k f_X(x) dx < \infty. \text{ Logo,}$$

$$\int_{-\infty}^0 x^k f_X(x) dx = \int_0^{\infty} (-u)^k f_X(-u) du = \int_0^{\infty} (-u)^k f_X(u) du,$$

ou seja,

$$E[X^k] = \int_0^{\infty} x^k f_X(x) dx + \int_0^{\infty} (-x)^k f_X(x) dx.$$

Portanto,

$$E[X^k] = \begin{cases} 2 \int_0^{\infty} x f_X(x) dx, & k \text{ par} \\ 0, & k \text{ ímpar} \end{cases}.$$

**Exemplo.** a) Seja  $X \sim N(0, \sigma^2)$ . Note que  $f_X$  é simétrica, de modo que  $E[X^k] = 0$  se  $k$  ímpar;



para  $k$  par, como  $X^2 \sim \Gamma(1/2, 1/2\sigma^2)$ , então

$$\begin{aligned} E[X^k] &= E[(X^2)^m] = \frac{\Gamma(m + 1/2)}{\Gamma(1/2)} \frac{1}{(1/2\sigma^2)^m} \\ &= \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma(1/2)} \frac{1}{(1/2\sigma^2)^{k/2}} \\ &= \frac{1}{\sqrt{\pi}} \frac{\sqrt{\pi} k!}{2^k (k/2)!} \frac{1}{2^{k/2} \sigma^k} \\ &= \frac{k!}{2^{k/2} (k/2)!} \sigma^k. \end{aligned}$$

b) Seja  $X \sim N(\mu, \sigma^2)$ , de modo que  $X - \mu \sim N(0, \sigma^2)$ . Se  $k$  é ímpar, então  $E[(X - \mu)^k] = 0 \implies E[X - \mu] = 0 \implies EX = \mu$ . Se  $k$  é par,

$$E[(X - \mu)^k] = \frac{k! \sigma^k}{2^{k/2} (k/2)!}.$$

Daí,  $\text{Var}(X - \mu) = E[(X - \mu)^2] = \sigma^2$ .

**Definição 7.4** (Covariância e correlação). Sejam  $X$  e  $Y$  v.a.'s tais que  $EX, EY < \infty$ . A **covariância** de  $X$  e  $Y$  é definida por

$$\text{Cov}(X, Y) := E[(X - EX)(Y - EY)] = E[XY] - EXEY.$$

Se  $X$  e  $Y$  são v.a.'s contínuas com densidade conjunta  $f_{X,Y}$ , então

$$\text{Cov}(X, Y) = \iint_{\mathbb{R}^2} (x - EX)(y - EY) f_{X,Y}(x, y) dx dy.$$

Se além de  $EX, EY < \infty$  tivermos  $0 < E[X^2], E[Y^2] < \infty$ , definimos a **correlação** entre  $X$  e  $Y$  por

$$\rho(X, Y) := \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)} \sqrt{\text{Var}(Y)}}.$$

**Observação.** As mesmas propriedades da covariância e da correlação para v.a.'s discretas valem para v.a.'s contínuas:

- (a) se  $X$  e  $Y$  são v.a.'s contínuas independentes, então  $\text{Cov}(X, Y) = 0$  e, portanto,  $\rho(X, Y) = 0$ , mas a recíproca é falsa pelo mesmo raciocínio do caso discreto;
- (b) se  $X_1, \dots, X_n$  são v.a.'s contínuas independentes tais que  $0 < E[X_i^2] < \infty, i = 1, \dots, n$ , então

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{Cov}(X_i, X_j).$$

Se as v.a.'s são independentes, então

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i)$$

- (c) se  $X$  e  $Y$  são v.a.'s contínuas com segundo momento finito e positivo, então vale a desigualdade de Cauchy-Schwarz,

$$E[XY]^2 \leq E[X^2]E[Y^2]$$

e, daí,

$$\text{Cov}(X, Y)^2 \leq \text{Var}(X)\text{Var}(Y) \implies |\rho(X, Y)| \leq 1.$$

**Exemplo.** Sejam  $X$  e  $Y$  v.a.'s com densidade conjunta

$$f_{X,Y}(x, y) = \frac{\sqrt{3}}{4\pi} \exp\left[-\frac{x^2 - xy + y^2}{2}\right], \forall (x, y) \in \mathbb{R}^2.$$

Vimos que  $X, Y \sim N(0, 4/3)$ , mas não eram independentes. Logo,  $EX = 0 = EY$  e  $\text{Var}(X) =$

$4/3 = \text{Var}(Y)$ . Temos também

$$\begin{aligned} E[XY] &= \iint_{\mathbb{R}^2} xy f_{X,Y}(x,y) dx dy = \sqrt{2\pi} \frac{\sqrt{3}}{4\pi} \int_{\mathbb{R}} x e^{-3x^2/8} \left( \frac{1}{\sqrt{2\pi}} y \exp \left[ -\frac{1}{2} \left( y - \frac{x}{2} \right)^2 \right] dy \right) dx \\ &= \frac{\sqrt{6}}{8\sqrt{\pi}} \int_{\mathbb{R}} x^2 e^{-3x^2/8} dx \\ &= \frac{\sqrt{6}}{8\sqrt{\pi}} \sqrt{2\pi} \sqrt{\frac{4}{3}} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi} \sqrt{4/3}} x^2 \exp \left[ -\frac{1}{2} \frac{x^2}{4/3} \right] dx \\ &= \frac{\sqrt{6}}{8\sqrt{\pi}} \sqrt{2\pi} \sqrt{\frac{4}{3}} \frac{4}{3} \\ &= \frac{2}{3}. \end{aligned}$$

Logo,  $\text{Cov}(X, Y) = E[XY] - EXEY = 2/3$  e

$$\rho(X, Y) = \frac{2/3}{\sqrt{4/3}\sqrt{4/3}} = \frac{1}{2}.$$

### 7.3 Exercícios - v.a.'s contínuas

1. Seja  $X$  uma variável aleatória tal que  $P(|X - 1| = 2) = 0$ . Expresse  $P(|X - 1| \geq 2)$  em termos da função de distribuição  $F_X$ .

**Solução.** Como  $P(|X - 1| = 2) = 0$ , temos  $P(X = 3) = 0 = P(X = 1)$ . Daí,  $P(|X + 1| \geq 2) = P(X \leq -1) + P(X \geq 3) = F_X(-1) + 1 - F_X(3)$ .  $\square$

2. Considere um ponto escolhido uniformemente no intervalo  $[0, a]$ . Seja  $X$  a distância da origem ao ponto escolhido. Obtenha a função de distribuição de  $X$ .

**Solução.** Se  $x < 0$ ,  $\{X \leq x\} = \emptyset$ ; se  $0 \leq x < a$ ,  $\{X \leq x\} = [0, x]$ ; se  $x \geq a$ ,  $\{X \leq x\} = [0, a]$ . Logo,

$$F_X(x) = \begin{cases} 0, & x < 0 \\ x/a, & 0 \leq x < a \\ 1, & x \geq a \end{cases}.$$

$\square$

3. Seja o ponto  $(u, v)$  escolhido uniformemente no quadrado  $[0, 1] \times [0, 1]$ . Seja  $X$  a v.a. que associa o número  $u + v$  ao ponto  $(u, v)$ . Obtenha a função de distribuição de  $X$ .

**Solução.** Se  $x < 0$ , então  $\{X \leq x\} = \emptyset$ ; se  $0 \leq x < 1$ , então  $\{X \leq x\} = \{(u, v) \in \mathbb{R}^2 | 0 \leq u + v < x\}$  (triângulo); se  $1 \leq x < 2$ , então  $\{X \leq x\} = \{(u, v) \in \mathbb{R}^2 | 1 \leq u + v < x\}$ ; se  $x \geq 2$ , então  $\{X \leq x\} = [0, 1] \times [0, 1]$ . Logo

$$F_X(x) = \begin{cases} 0, & x < 0 \\ x^2/2, & 0 \leq x < 1 \\ -1 + 2x - x^2/2, & 1 \leq x < 2 \\ 1, & x \geq 2 \end{cases}.$$

$\square$

4. Obtenha a função de densidade para cada uma das variáveis aleatórias dos exercícios 2 e 3.

**Solução.** Para o exercício 2, temos

$$f_X(x) = \begin{cases} 0, & x \leq 0 \text{ ou } x \geq a \\ 1/a, & 0 < x < a \end{cases}.$$

Para o exercício 3, temos

$$f_X(x) = \begin{cases} x, & 0 < x < 1 \\ 2 - x, & 1 < x < 2 \\ 0, & \text{c.c.} \end{cases} .$$

□

5. Seja  $F$  a função de distribuição exponencial de parâmetro  $\lambda$ . Obtenha um número  $m$  tal que  $F(m) = 1/2$  ( $m$  é chamado de mediana de  $F$ ).

**Solução.** Temos  $F(m) = 1/2 \Leftrightarrow e^{-\lambda m} = 1/2 \Leftrightarrow m = \ln(2)/\lambda$ .

□

6. Seja  $X$  uma variável aleatória contínua com densidade  $f$  dada por:

$$f(x) = \frac{1}{2}e^{-|x|}, x \in \mathbb{R}.$$

Obtenha  $P(1 \leq |X| \leq 2)$ .

**Solução.** Temos

$$P(1 \leq |X| \leq 2) = 2P(1 \leq X \leq 2) = 2(F_X(2) - F_X(1)) = e^{-1} - e^{-2}.$$

□

7. Seja  $F$  a função de distribuição definida por:

$$F(x) = \frac{1}{2} + \frac{x}{2(|x| + 1)}, x \in \mathbb{R}.$$

Obtenha uma densidade  $f$  para  $F$ . Para que valores de  $x$  teremos  $F'(x) = f(x)$ ?

**Solução.** Se  $x \geq 0$ , então

$$F'(x) = \frac{1}{2(|x| + 1)^2}.$$

Se  $x < 0$ , então

$$F'(x) = \frac{1}{2(|x| + 1)^2}.$$

Logo, podemos tomar  $f(x) = 1/[2(|x| + 1)^2]$  para todo  $x \in \mathbb{R}$  e vale  $f(x) = F'(x)$  para todo  $x \in \mathbb{R}$ .

□

## 7.4 Exercícios - funções de v.a.'s contínuas

1. Considere um ponto escolhido uniformemente no intervalo  $[0, a]$ . Seja  $X$  a distância da origem ao ponto escolhido. Obtenha a função de distribuição de  $Y = \min(X, a/2)$ .

**Solução.** Se  $y < 0$ , então  $P(Y \leq y) = 0$  pois  $\min(X, a/2) \geq 0$ . Se  $0 \leq y < a/2$ , então  $F_Y(y) = F_X(y) = y/a$ ; se  $y \geq a/2$ , então  $F_Y(y) = 1$  pois  $Y \leq a/2$ .

□

2. Escolhe-se aleatoriamente um ponto em  $(-10, 10)$ . Seja  $X$  uma variável aleatória definida de tal forma que  $X$  represente a coordenada do ponto se o mesmo estiver em  $[-5, 5]$ ,  $X = -5$  se o ponto estiver em  $(-10, -5)$  e  $X = 5$  se o ponto estiver em  $(5, 10)$ . Obtenha a função de distribuição de  $X$ .

**Solução.** Se  $x < -5$ , então  $F_X(x) = 0$  pois  $X$  assume valores em  $[-5, 5]$ ; se  $-5 \leq x < 5$ , então  $F_X(x) = P(X \in (-10, x)) = (x + 10)/20$ ; se  $x \geq 5$ , então  $P(X \leq x) = 1$ .

□

3. Verifique que:

a)  $X \sim N(0, 1)$  se, e somente se,  $\sigma X + \mu \sim N(\mu, \sigma^2)$ , onde  $\mu \in \mathbb{R}$  e  $\sigma > 0$ .

b)  $X \sim \text{Cauchy}(0, 1)$  se, e somente se,  $bX + a \sim \text{Cauchy}(a, b)$ , onde  $a \in \mathbb{R}$  e  $b > 0$ .

**Solução.**

a) Foi feito no texto.

b) Temos

$$f_{a+bX}(x) = \frac{1}{b} \cdot \frac{1}{\pi(1 + (\frac{x-a}{b})^2)} = \frac{b}{\pi(b^2 + (x-a)^2)},$$

logo  $a + bX \sim \text{Cauchy}(a, b)$ . A recíproca segue de maneira análoga. □

4. Suponha que  $X \sim \text{Exp}(\lambda)$ ,  $\lambda > 0$ . Obtenha a densidade de  $Y = cX$ , onde  $c > 0$ .

**Solução.** Temos  $Y = g(X)$  sendo  $g(x) = cx$  para todo  $x$  real. Note que  $g$  é crescente e derivável em  $I = (0, +\infty)$  e que, nesse intervalo, temos  $dx/dy = 1/c$  e  $x = y/c$ . Logo,

$$f_Y(y) = \frac{1}{c} \lambda e^{-\lambda y/c}, y > 0$$

e  $f_Y(y) = 0$  caso contrário, ou seja,  $Y \sim \text{Exp}(\lambda/c)$ . □

5. Suponha que  $X \sim U(0, 1)$ . Obtenha a densidade de  $Y = X^{1/\beta}$ , onde  $\beta \neq 0$ .

**Solução.** Temos  $Y = g(X)$  com  $g(x) = x^{1/\beta}$ . Temos  $g$  estritamente monótona (crescente se  $\beta > 0$  e decrescente caso contrário) e derivável em  $(0, +\infty)$ . Logo,  $x = y^\beta$  e  $dx/dy = \beta y^{\beta-1}$ . Para  $\beta > 0$ , temos  $f_Y(y) = \beta y^{\beta-1}$  para  $0 < y < 1$  e  $f_Y(y) = 0$  caso contrário; se  $\beta < 0$ , então  $f_Y(y) = -\beta y^{\beta-1}$  para  $y > 1$  e  $f_Y(y) = 0$  caso contrário. □

6. Seja  $X$  uma variável aleatória contínua com densidade  $f$ .

a) Obtenha uma fórmula para a densidade de  $Y = |X|$  em termos de  $f$ .

b) Obtenha uma fórmula para a densidade de  $Y = X^2$  em termos de  $f$ .

**Solução.**

a) Se  $y > 0$ , então  $Y = \pm X$  e  $f_Y(y) = f(y) + f(-y)$ . Se  $y \leq 0$ , então  $f_Y(y) = 0$ .

b) Foi feito no texto. □

7. Seja  $X \sim N(0, \sigma^2)$ . Obtenha a densidade de:

a)  $Y = |X|$ ;

b)  $Y = X^2$ .

**Solução.**

a) Do exercício anterior,

$$f_Y(y) = \begin{cases} \frac{2}{\sqrt{2\pi}\sigma} e^{-y^2/2\sigma^2}, y > 0 \\ 0, \text{c.c.} \end{cases}$$

b) Temos

$$f_Y(y) = \begin{cases} \frac{(1/2\sigma^2)^{1/2}}{\sqrt{\pi}} y^{-1/2} e^{-y/2\sigma^2}, y > 0 \\ 0, \text{c.c.} \end{cases},$$

ou seja,  $Y \sim \text{Gama}(1/2, 1/2\sigma^2)$ . □

8. Seja  $X \sim N(\mu, \sigma^2)$ . Obtenha a densidade de  $Y = e^X$ . Essa densidade chama-se densidade lognormal. □

**Solução.** Temos  $Y = g(X)$  com  $g(x) = e^x$ . Temos  $g$  crescente e derivável em  $\mathbb{R}$  e também  $x = \ln y$ , donde segue que  $dx/dy = 1/y$ . Portanto,

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma} e^{-(\ln y - \mu)^2/2\sigma^2} \frac{1}{y}, y > 0 \\ 0, \text{c.c.} \end{cases}$$

□

9. Seja  $X$  uma v.a. contínua com densidade simétrica  $f$  e tal que  $X^2 \sim \text{Exp}(\lambda)$ ,  $\lambda > 0$ . Obtenha  $f$ .

**Solução.** Temos

$$f_{X^2}(y) = \begin{cases} \lambda e^{-\lambda y}, y > 0 \\ 0, \text{c.c.} \end{cases},$$

logo  $\lambda e^{-\lambda y} = f(\sqrt{y})/\sqrt{y}$ , ou seja,  $f(|x|) = \lambda|x|e^{-\lambda x^2} = f(x)$ ,  $x \in \mathbb{R}$ . □

10. Seja  $\Theta \sim U[-\pi/2, \pi/2]$ . Determine a função de distribuição e a densidade de:

- $X = \tan(\Theta)$
- $Y = \sin(\Theta)$ .

**Solução.**

- Como a função tangente é estritamente crescente e derivável em  $\mathbb{R}$ , temos  $\Theta = \arctan x$  e  $d\Theta/dx = 1/(1+x^2)$  e, daí,  $f_X(x) = (1/\pi)1/(1+x^2)$  para todo  $x \in \mathbb{R}$ , ou seja,  $X \sim \text{Cauchy}(0, 1)$ .
- Como a função seno é estritamente crescente e derivável em  $(-\pi/2, \pi/2)$ , temos  $\Theta = \arcsin y$  e  $d\Theta/dy = 1/\sqrt{1-y^2}$  e, daí,

$$f_Y(y) = \begin{cases} \frac{1}{\pi} \cdot \frac{1}{\sqrt{1-y^2}}, & -1 < y < 1 \\ 0, & \text{c.c.} \end{cases}$$

□

11. Seja  $X \sim \Gamma(\alpha, \lambda)$ . Determine a densidade de

- $Y = cX, c > 0$ ;
- $Y = \sqrt{X}$ .

**Solução.**

- Raciocinando como no exercício 4, temos

$$f_Y(y) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \frac{1}{c^\alpha} y^{\alpha-1} e^{-\lambda y/c}, & y > 0 \\ 0, & \text{c.c.} \end{cases},$$

ou seja,  $Y \sim \text{Gama}(\alpha, \lambda/c)$ .

- Raciocinando como no exercício 5, temos

$$f_Y(y) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} 2y^{2\alpha-1} e^{-\lambda y^2}, & y > 0 \\ 0, & \text{c.c.} \end{cases}.$$

□

## 7.5 Exercícios - soma e quociente de v.a.'s contínuas

1. Sejam  $X$  e  $Y$  variáveis aleatórias independentes, cada uma com distribuição uniforme em  $(0, 1)$ . Obtenha:

- $P(|X - Y| \leq 1/2)$
- $P(|X/Y - 1| \leq 1/2)$

**Solução.**

- Usando desenhos, temos que  $P(|X - Y| \leq 1/2)$  é a área de um hexágono, a saber  $3/4$ . Também podemos calcular como  $P(X - Y \leq 1/2) - P(X - Y \leq -1/2) = 7/8 - 1/8 = 3/4$ .
- Analogamente, podemos calcular a probabilidade através de uma área ou simplesmente como  $P(X/Y \leq 3/2) - P(X/Y \leq 1/2) = 2/3 - 1/4 = 5/12$ .

□

2. Suponha que os tempos que dois estudantes levam para resolver um problema são independentes e se distribuem exponencialmente com parâmetro  $\lambda$ . Determine a probabilidade de que o primeiro estudante necessite pelo menos do dobro do tempo gasto pelo segundo para resolver o problema.

**Solução.** Sejam  $X$  e  $Y$  os tempos do primeiro e segundo estudantes, respectivamente. Temos que

$$P(X \geq 2Y) = \int_0^\infty P(X \geq 2t, Y = t) dt = \int_0^\infty P(X \geq 2t)P(Y = t) dt = \lambda \int_0^\infty e^{-3\lambda t} dt = 1/3.$$

□

3. Seja  $f(x, y) = ce^{-(x^2 - xy + 4y^2)/2}, (x, y) \in \mathbb{R}^2$ .

- Determine o valor de  $c$  para que  $f$  seja uma densidade.

- b) Se  $X$  e  $Y$  têm densidade conjunta  $f$ , determine as densidades marginais de  $X$  e  $Y$  e verifique se são independentes.

**Solução.**

- a) Podemos reescrever

$$f(x, y) = c \exp \left[ -\frac{1}{2} \left( x - \frac{y}{2} \right)^2 \right] \exp \left[ -\frac{15y^2}{8} \right].$$

Daí, temos

$$\begin{aligned} f_Y(y) &= \int_{\mathbb{R}} f(x, y) dx \\ &= c \exp \left[ -\frac{15y^2}{8} \right] \int_{\mathbb{R}} \exp \left[ -\frac{1}{2} \left( x - \frac{y}{2} \right)^2 \right] dx \\ &= c\sqrt{2\pi} e^{-15y^2/8}, \forall y \in \mathbb{R}. \end{aligned}$$

Devemos ter  $c$  tal que a integral de  $f_Y(y)$  em  $y$  sobre a reta seja igual a 1, ou seja,

$$c\sqrt{2\pi} \int_{\mathbb{R}} f_Y(y) dy = 1 \implies c = \frac{\sqrt{15}}{4\pi}.$$

- b) Do item anterior, temos que  $Y \sim N(0, 4/15)$ . Realizando um procedimento análogo, concluímos também que  $X \sim N(0, 16/15)$ . Observe que como  $f(x, y) \neq f_X(x)f_Y(y)$  em geral, então  $X$  e  $Y$  não são independentes. □

4. Suponha que  $X$  e  $Y$  têm densidade conjunta uniforme no interior do triângulo com vértices em  $(0, 0)$ ,  $(2, 0)$  e  $(1, 2)$ . Obtenha  $P(X \leq 1, Y \leq 1)$ .

**Solução.** Podemos interpretar essa probabilidade como a área de uma região de um triângulo, de modo que  $P(X \leq 1, Y \leq 1) = 3/8$ . □

5. Sejam  $X$  e  $Y$  v.a.'s contínuas independentes tendo as densidades marginais especificadas abaixo. Obtenha, em cada item, a densidade de  $Z = X + Y$ .

- a)  $X \sim \text{Exp}(\lambda_1)$  e  $Y \sim \text{Exp}(\lambda_2)$   
 b)  $X \sim \Gamma(\alpha_1, \lambda)$  e  $Y \sim \Gamma(\alpha_2, \lambda)$   
 c)  $X \sim N(\mu_1, \sigma_1^2)$  e  $Y \sim N(\mu_2, \sigma_2^2)$

**Solução.**

- (a) Para  $z > 0$ , temos

$$\begin{aligned} f_{X+Y}(z) &= \int_{\mathbb{R}} f_{X,Y}(x, z-x) dx \\ &= \lambda_1 \lambda_2 e^{-\lambda_2 z} \int_0^z e^{(\lambda_2 - \lambda_1)x} dx \\ &= \frac{\lambda_1 \lambda_2}{\lambda_2 - \lambda_1} (e^{-\lambda_1 z} - e^{-\lambda_2 z}). \end{aligned}$$

Do contrário,  $f_{X+Y}(z) = 0$ .

- (b) Para  $z > 0$ , temos

$$\begin{aligned} f_{X+Y}(z) &= \frac{\lambda_1^{\alpha_1 + \alpha_2} e^{-\lambda z}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^z x^{\alpha_1 - 1} (z-x)^{\alpha_2 - 1} dx \\ &= \frac{\lambda^{\alpha_1 + \alpha_2} z^{\alpha_1 + \alpha_2 - 1}}{\Gamma(\alpha_1 + \alpha_2)} e^{-\lambda z}. \end{aligned}$$

Do contrário,  $f_{X+Y}(z) = 0$ . Logo,  $X + Y \sim \text{Gama}(\alpha_1 + \alpha_2, \lambda)$ .

- (c) Por simplicidade, supomos  $\mu_1 = 0 = \mu_2$ . Realizando contas análogas às acima, obtemos que  $X + Y \sim N(0, \sigma_1^2 + \sigma_2^2)$ . Daí, basta somar  $\mu_1 + \mu_2$  para obter que  $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ . □

6. Verifique que se  $X_1, X_2, \dots, X_n$  são v.a.'s i.i.d. com distribuição comum  $N(0, 1)$ , então  $X_1^2 + X_2^2 + \dots + X_n^2$  tem distribuição  $\Gamma(n/2, 1/2)$  (esta distribuição é conhecida como qui-quadrado com  $n$  graus de liberdade e denotada por  $\chi^2(n)$ ). [Sugestão: use os exercícios (7b) da lista 8 e (5b) anterior.]

**Solução.** Do exercício 7b, temos  $X_i^2 \sim \text{Gama}(1/2, 1/2)$  para  $i = 1, \dots, n$ . Do exercício 5b, temos  $\sum_i X_i^2 \sim \text{Gama}(n/2, 1/2)$ .  $\square$

7. Suponha que se escolhe aleatoriamente um ponto no plano de tal forma que suas coordenadas  $X$  e  $Y$  se distribuem independentemente segundo a densidade normal  $N(0, \sigma^2)$ . Obtenha a função densidade da v.a.  $R$  que representa a distância do ponto escolhido à origem. (Esta densidade é conhecida como densidade de Rayleigh).

**Solução.** Temos  $Z = X^2 + Y^2 \sim \Gamma(1/2, 1/2\sigma^2)$ . Temos então  $R = g(Z)$  sendo  $g$  a função raiz quadrada. Como ela é estritamente crescente e derivável em  $(0, +\infty)$ , segue que para  $r > 0$  temos

$$f_R(r) = \frac{2r}{2\sigma^2} e^{-r^2/2\sigma^2} = \frac{r}{\sigma^2} e^{-r^2/2\sigma^2}$$

e, para  $r \leq 0$ ,  $f_R(r) = 0$ .  $\square$

8. Sejam  $X$  e  $Y$  v.a.'s i.i.d. com distribuição comum  $\text{Exp}(\lambda)$ . Obtenha a densidade de  $Z = Y/X$ .

**Solução.** Para  $z > 0$ , temos

$$\begin{aligned} f_{Y/X}(z) &= \int_{\mathbb{R}} |x| f_{X,Y}(x, xz) dx \\ &= \int_0^{\infty} x f_X(x) f_Y(xz) dx \\ &= \lambda^2 \int_0^{\infty} x e^{-\lambda x(z+1)} dx \\ &= \lambda^2 \cdot \frac{\Gamma(2)}{[\lambda(z+1)]^2} \\ &= \frac{1}{(z+1)^2}. \end{aligned}$$

Do contrário,  $f_{Y/X}(z) = 0$ .  $\square$

9. Sejam  $X$  e  $Y$  v.a.'s independentes tais que  $X \sim \Gamma(\alpha_1, \lambda)$  e  $Y \sim \Gamma(\alpha_2, \lambda)$ . Obtenha a densidade de  $Z = X/(X+Y)$ . [Sugestão: expresse  $Z$  em função de  $Y/X$ ].

**Solução.** Observe que  $Z = (1+W)^{-1}$  sendo  $W = Y/X$ . Como  $W$  é uma v.a. positiva, os valores possíveis de  $Z$  formam o intervalo  $(0, 1)$ . Ademais, a função de  $W$  que define  $Z$  é estritamente decrescente e derivável nesse intervalo, de modo que podemos usar a mudança de variáveis para obter

$$f_Z(z) = \frac{1}{z^2} \cdot \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \cdot \frac{(1-z)^{\alpha_2-1} z^{1-\alpha_2}}{z^{-\alpha_1-\alpha_2}} = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} z^{\alpha_1-1} (1-z)^{\alpha_2-1}$$

para  $0 < z < 1$  e  $f_Z(z) = 0$  caso contrário.  $\square$

10. Sejam  $U$  e  $V$  v.a.'s i.i.d. com distribuição comum  $N(0, 1)$ . Seja  $Z = \rho U + \sqrt{1-\rho^2}V$ , onde  $-1 < \rho < 1$ .

a) Obtenha a densidade de  $Z$ .

b) Obtenha a densidade conjunta de  $U$  e  $Z$ .

c) Obtenha a densidade conjunta de  $X = \mu_1 + \sigma_1 U$  e  $Y = \mu_2 + \sigma_2 Z$ , onde  $\sigma_1, \sigma_2 > 0$ . Esta densidade é conhecida como densidade normal bidimensional.

**Solução.**

(a) Temos  $\rho U \sim N(0, \rho^2)$  e  $\sqrt{1-\rho^2}V \sim N(0, 1-\rho^2)$ . Logo,  $Z \sim N(0, 1)$ .

(b) Usando a fórmula da mudança de variáveis com o jacobiano, temos que

$$J(u, v) = \begin{vmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{vmatrix} = \sqrt{1-\rho^2}.$$

Daí, temos

$$\begin{aligned} f_{U,Z}(u, z) &= \frac{1}{|J(u, v)|} f_{U,V} \left( u, \frac{z - \rho u}{\sqrt{1-\rho^2}} \right) \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left[ -\frac{u^2 - 2\rho uz + z^2}{2(1-\rho^2)} \right], \end{aligned}$$

para todo  $(u, z) \in \mathbb{R}^2$ .

(c) Utilizando a mesma fórmula com o jacobiano, temos que

$$\begin{aligned} f_{X,Y}(x, y) &= \frac{1}{|J(u, z)|} f_{U,Z} \left( \frac{x - \mu_1}{\sigma_1}, \frac{y - \mu_2}{\sigma_2} \right) \\ &= \frac{1}{\sigma_1\sigma_2} \cdot \frac{1}{2\pi\sqrt{1-\rho^2}} \\ &\exp \left[ -\frac{1}{2(1-\rho^2)} \left( \left( \frac{x - \mu_1}{\sigma_1} \right)^2 - 2\rho \left( \frac{x - \mu_1}{\sigma_1} \right) \left( \frac{y - \mu_2}{\sigma_2} \right) + \left( \frac{y - \mu_2}{\sigma_2} \right)^2 \right) \right], \end{aligned}$$

para todo  $(x, y) \in \mathbb{R}^2$ .

□

11. Sejam  $R$  e  $\Theta$  v.a.'s independentes de modo que  $R$  tem densidade de Rayleigh:

$$f_R(r) = \begin{cases} \sigma^{-2} r e^{-r^2/2\sigma^2}, & r \geq 0 \\ 0, & r < 0 \end{cases}.$$

e  $\Theta$  se distribui uniformemente em  $U(-\pi, \pi)$ . Mostre que se  $X = R \cos \Theta$  e  $Y = R \sin \Theta$  são v.a.'s independentes e que cada uma tem densidade normal  $N(0, \sigma^2)$ .

**Solução.** Usando novamente a fórmula da mudança de variáveis com o jacobiano, temos

$$\begin{aligned} f_{X,Y}(x, y) &= \frac{1}{|J(r, \theta)|} f_{R,\Theta}(r, \theta) \\ &= \frac{1}{r} f_R(r) f_\Theta(\theta) \\ &= \frac{1}{\sqrt{x^2 + y^2}} \cdot \frac{\sqrt{x^2 + y^2}}{\sigma^2} \exp \left[ -\frac{x^2 + y^2}{2\sigma^2} \right] \cdot \frac{1}{2\pi} \\ &= \frac{1}{2\pi\sigma^2} \exp \left[ -\frac{x^2 + y^2}{2\sigma^2} \right]. \end{aligned}$$

Daí,

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y) dy = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}$$

e também

$$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x, y) dx = \frac{1}{\sqrt{2\pi}\sigma} e^{-y^2/2\sigma^2}$$

para todo  $(x, y) \in \mathbb{R}^2$ , ou seja,  $X, Y \sim N(0, \sigma^2)$ . Como  $f_X(x)f_Y(y) = f_{X,Y}(x, y)$  para todo  $(x, y) \in \mathbb{R}^2$ , temos  $X$  e  $Y$  independentes. □



# 8 Função Geradora de Momentos, Lei dos Grandes Números e o Teorema do Limite Central

## 8.1 Função Geradora de Momentos

**Definição 8.1.** Seja  $X$  uma v.a. definida em  $(\Omega, \mathcal{A}, P)$ . A função geradora de momentos de  $X$ ,  $M_X : \mathbb{R} \rightarrow \mathbb{R}$ , é dada por

$$M_X(t) = E[e^{tX}],$$

para todo  $t \in \mathbb{R}$  tal que  $e^{tX}$  tem esperança finita. Note que se  $X$  é discreta, então

$$M_X(t) = \sum_x e^{tx} p_X(x)$$

e, se  $X$  é contínua, então

$$M_X(t) = \int_{\mathbb{R}} e^{tx} f_X(x) dx.$$

**Exemplo.** Seja  $X \sim B(n, p)$ . Temos

$$\begin{aligned} \sum_{k=0}^n e^{tk} \binom{n}{k} p^k (1-p)^{n-k} &= \sum_{k=0}^n \binom{n}{k} (pe^t)^k (1-p)^{n-k} \\ &= [pe^t + 1 - p]^n \in \mathbb{R}, \forall t \in \mathbb{R} \therefore M_X(t) = [pe^t + 1 - p]^n, t \in \mathbb{R}. \end{aligned}$$

**Exemplo.** Seja  $X \sim \Gamma(\alpha, \lambda)$ . Temos

$$\begin{aligned} \int_{\mathbb{R}} e^{tx} f_X(x) dx &= \int_0^{\infty} e^{tx} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \underbrace{\int_0^{\infty} x^{\alpha-1} e^{-(\lambda-t)x} dx}_{\frac{\Gamma(\alpha)}{(\lambda-t)^\alpha}, \text{ se } t < \lambda} \\ &= \left( \frac{\lambda}{\lambda-t} \right)^\alpha, t < \lambda \therefore M_X(t) = \left( \frac{\lambda}{\lambda-t} \right)^\alpha, t < \lambda. \end{aligned}$$

**Exemplo.** Seja  $X \sim N(\mu, \sigma^2)$ . Temos

$$\begin{aligned} \int_{\mathbb{R}} e^{tx} f_X(x) dx &= \int_{\mathbb{R}} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx \\ &= \int_{\mathbb{R}} e^{t(y+\mu)} \frac{1}{\sigma\sqrt{2\pi}} e^{-y^2/2\sigma^2} dy \\ &= e^{\mu t} \int_{\mathbb{R}} \frac{1}{\sigma\sqrt{2\pi}} e^{ty-y^2/2\sigma^2} dy \\ &= e^{\mu t} e^{\sigma^2 t^2/2} \int_{\mathbb{R}} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(y-\sigma^2 t)^2}{2\sigma^2}\right] dy \\ &= e^{\mu t} e^{\sigma^2 t^2/2}, t \in \mathbb{R} \therefore M_X(t) = e^{\mu t} e^{\sigma^2 t^2/2}, t \in \mathbb{R}. \end{aligned}$$

O nome da f.g.m. vem do fato de que  $E[X^n] = M_X(0)^{(n)}$ . De fato, pode-se mostrar que para  $t \in (-t_0, t_0)$  tal que  $M_X(t)$  está definida, temos

$$M_X(t) = E[e^{tX}] = E\left[\sum_{n=0}^{\infty} \frac{t^n X^n}{n!}\right] = \sum_{n=0}^{\infty} E\left[\frac{t^n X^n}{n!}\right] = \sum_{n=0}^{\infty} E[X^n] \frac{t^n}{n!}.$$

Em particular, se  $M_X$  está definida para todo  $t$ , então as igualdades acima valem para todo  $t$ . Note que a expansão em série de Taylor de  $M_X$  é

$$\sum_{n=0}^{\infty} \frac{t^n}{n!} \frac{d^n}{dt^n} M_X(t) \Big|_{t=0}.$$

Comparando os coeficientes, temos o resultado afirmado.

**Observação.** Note que se  $X$  e  $Y$  são v.a.'s independentes, então  $e^{tX}$  e  $e^{tY}$  também o são e, daí,

$$M_{X+Y}(t) = E[e^{t(X+Y)}] = E[e^{tX} e^{tY}] = E[e^{tX}]E[e^{tY}] = M_X(t)M_Y(t).$$

Segue daí, por indução, que se  $X_1, \dots, X_n$  são i.i.d. então

$$M_{X_1+\dots+X_n}(t) = M_{X_1}(t)^n.$$

**Exemplo.** Para  $X \sim B(n, p)$ , temos  $M_X(t) = (pe^t + 1 - p)^n, \forall t \in \mathbb{R}$ . Daí,

$$M_X(t)' = n(pe^t + 1 - p)^{n-1}pe^t \implies M_X(0)' = EX = np$$

e também

$$M_X(t)'' = n(n-1)(pe^t + 1 - p)^{n-2}p^2e^{2t} + n(pe^t + 1 - p)^{n-1}pe^t \implies M_X(0)'' = E[X^2] = n(n-1)p^2 + np,$$

de modo que  $\text{Var}(X) = E[X^2] - EX^2 = np(1-p)$ .

**Exemplo.** Se  $X \sim N(0, \sigma^2)$ , vimos que

$$M_X(t) = e^{\sigma^2 t^2/2} = \sum_{n=0}^{\infty} \left( \frac{\sigma^2 t^2}{2} \right)^n \frac{1}{n!} = \sum_{n=0}^{\infty} \frac{\sigma^{2n}}{2^n n!} t^{2n}.$$

Daí, os momentos de ordem ímpar de  $X$  são nulos e os de ordem par são tais que

$$\frac{E[X^{2n}]}{(2n)!} = \frac{\sigma^{2n}}{2^n n!} \Leftrightarrow E[X^{2n}] = \frac{\sigma^{2n}(2n)!}{2^n n!}.$$

## 8.2 Desigualdades de Chebyshev-Markov e Lei dos Grandes Números

### Desigualdades de Chebyshev e Markov

**Proposição 8.2** (Desigualdade de Chebyshev). Seja  $X$  v.a. tal que  $E[X^2] < \infty$ . Dado  $\varepsilon > 0$ , considere  $\mu = EX$ . Então

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2}.$$

**Demonstração.** Dado  $\varepsilon > 0$ , considere a v.a. discreta

$$Y = \begin{cases} \varepsilon^2, & (X - \mu)^2 \geq \varepsilon^2 \\ 0, & (X - \mu)^2 < \varepsilon^2 \end{cases}.$$

Note que  $P(Y = 0) = P((X - \mu)^2 < \varepsilon^2)$  e  $P(Y = \varepsilon^2) = P((X - \mu)^2 \geq \varepsilon^2)$ . Daí,  $EY = \varepsilon^2 P((X - \mu)^2 \geq \varepsilon^2)$ , ou seja,  $P((X - \mu)^2 \geq \varepsilon^2) = EY/\varepsilon^2$ . Ademais,  $(X - \mu)^2 \geq Y$ , pois  $Y = 0 \leq (X - \mu)^2$  ou  $Y = \varepsilon^2 \leq (X - \mu)^2$ . Da monotonicidade da esperança, temos  $E[(X - \mu)^2] \geq EY$ , de modo que

$$P(|X - \mu| \geq \varepsilon) = P((X - \mu)^2 \geq \varepsilon^2) \leq \frac{E[(X - \mu)^2]}{\varepsilon^2} = \frac{\text{Var}(X)}{\varepsilon^2}.$$

□

**Observação.** Considere  $X$  v.a. não negativa com  $EX < \infty$  e seja

$$Y = \begin{cases} \varepsilon, & X \geq \varepsilon \\ 0, & X < \varepsilon \end{cases},$$

para  $\varepsilon > 0$  dado. Raciocinando como acima, temos  $EX \geq EY = \varepsilon P(X \geq \varepsilon)$ , isto é,  $P(X \geq \varepsilon) \leq EX/\varepsilon$ , chamada **desigualdade básica de Chebyshev**. A desigualdade de Chebyshev nos dá uma cota superior, em termos de  $\text{Var}(X)$  e  $t$ , para a probabilidade de que  $X$  desvie de sua média em mais de  $t$  unidades. Sua força está na sua grande generalidade: nenhuma hipótese é feita acerca da distribuição de  $X$  além de que tenha variância finita. Essa desigualdade é o ponto inicial de vários desenvolvimentos teóricos. Para a maioria das distribuições que surgem na prática, há limites muito mais precisos para  $P(|X - \mu| \geq t)$  do que a cota fornecida pela desigualdade de Chebyshev; entretanto, exemplos mostram que, em geral, a cota dada pela desigualdade não pode ser melhorada.

**Proposição 8.3** (Desigualdade de Markov). Seja  $X$  v.a. qualquer com variância finita. Dados  $\varepsilon, k > 0$

quaisquer, temos

$$P(|X| \geq \varepsilon) \leq \frac{E[|X|^k]}{\varepsilon^k}.$$

**Demonstração.** Pela desigualdade básica de Chebyshev,

$$P(|X| \geq \varepsilon) = P(|X|^k \geq \varepsilon^k) \leq \frac{E[|X|^k]}{\varepsilon^k}.$$

□

**Exemplo.** Uma fábrica produz, em média, 50 peças por semana. Considere  $X$  a v.a. que denota o número de peças produzidas semanalmente. Nesse contexto, (a) o que é possível dizer sobre a probabilidade da produção semanal exceder 75 peças? e, supondo que a produção semanal tenha variância 25, (b) como estimar a probabilidade de que a produção semanal exceda 75 peças e (c) o que podemos dizer sobre a probabilidade da produção semanal estar entre 40 e 60?

(a) Note que  $EX = 50$  e, pela desigualdade de Markov,

$$P(X > 75) \leq \frac{50}{75} = \frac{2}{3}.$$

(b) Temos  $E[X^2] = \text{Var}(X) + EX^2 = 2525$ . Pela desigualdade de Markov,

$$P(X > 75) \leq \frac{E[X^2]}{75^2} = \frac{2525}{5625} = \frac{101}{225}.$$

(c) Temos, usando a desigualdade de Chebyshev, que

$$\begin{aligned} P(40 < X < 60) &= P(-10 < X - 50 < 10) \\ &= P(|X - 50| < 10) \\ &= 1 - P(|X - 50| \geq 10) \\ &\geq 1 - \frac{25}{100} \\ &= \frac{3}{4}. \end{aligned}$$

## Lei dos Grandes Números (L.G.N.)

Seja  $X_1, X_2, \dots$  uma sequência de v.a.'s i.i.d. com  $E[X_i] = \mu \in \mathbb{R}, \forall i \in \mathbb{N}$ . Queremos analisar a média amostral  $S_n/n = (X_1 + \dots + X_n)/n$  e verificar que, em algum sentido,  $S_n/n \xrightarrow{b \rightarrow \infty} E[X_1] = \mu$ . Dito de outro modo, queremos verificar que a média amostral se “aproxima” da média real.

Para tanto, considere  $F$  uma função de distribuição cuja média é finita e desconhecida. Queremos estimar  $\mu$  a partir de uma amostra de  $F$ , digamos  $X_1, \dots, X_n$  i.i.d. com distribuição  $F$ . Considerando  $S_n = X_1 + \dots + X_n$ , queremos saber se  $S_n/n \rightarrow \mu$  ou, equivalentemente, se  $(S_n - E[S_n])/n \rightarrow 0$  de alguma maneira. O que nos fornece essa resposta são as leis dos grandes números.

**Proposição 8.4** (Lei Fraca dos Grandes Números (LfGN)). Sejam  $X_1, X_2, \dots$  v.a.'s i.i.d. com  $E[X_1] = \mu < \infty$  e  $S_n = X_1 + \dots + X_n$ . Daí, para todo  $\varepsilon > 0$ ,

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) = 0.$$

Note que  $\varepsilon$  representa a precisão desejada na aproximação de  $\mu$  pela média amostral. Assim, por menor que seja  $\varepsilon$ , a probabilidade de  $S_n/n$  se aproximar de  $\mu$  com precisão  $\varepsilon$  se aproxima de 1 à medida que  $n$  cresce. Nesse caso, a LfGN nos diz que  $S_n/n$  converge **em probabilidade** para  $\mu$ , denotando-se  $S_n/n \xrightarrow{p} \mu$ . A demonstração da LfGN (e também do TLC, que será visto mais à frente, usa funções características, e será dada ao final da seção.

**Proposição 8.5** (Lei Forte dos Grandes Números (LFGN)). Sejam  $X_1, X_2, \dots$  v.a.'s i.i.d. com  $E[X_1] =$

$\mu < \infty$  e  $S_n = X_1 + \dots + X_n$ . Daí, para todo  $\varepsilon > 0$ ,

$$P\left(\lim_{n \rightarrow \infty} \left| \frac{S_n}{n} - \mu \right| > \varepsilon\right) = 0.$$

A LFGN nos diz que  $S_n/n \rightarrow \mu$  com probabilidade 1. Nesse caso, dizemos que  $S_n/n$  converge **quase certamente** para  $\mu$ , e denotamos  $S_n/n \xrightarrow{\text{q.c.}} \mu$ . Dizemos também, nesse caso, que  $T_n = S_n/n$  é um **estimador fortemente consistente** para  $\mu$ . A demonstração da LFGN envolve argumentos de Teoria da Medida, e não será apresentada.

**Exemplo.** Um candidato a prefeito gostaria de ter uma ideia de quantos votos receberá nas próximas eleições. Para isso, foi feita uma pesquisa com a população da cidade, onde  $p$  representa a proporção de votos a favor do candidato em questão. Quantas pessoas devem ser entrevistadas para que a pesquisa tenha 95% de confiança de que o erro seja inferior a 5%, supondo que a escolha dos entrevistados é feita de maneira aleatória e independente?

Seja  $n$  o número de pessoas entrevistadas e defina

$$X_k = \begin{cases} 1, & k\text{-ésima pessoa entrevistada a favor do candidato} \\ 0, & \text{c.c.} \end{cases}, \quad k = 1, \dots, n.$$

Temos  $X_1, \dots, X_n$  v.a.'s i.i.d. Bernoulli, de modo que  $P(X_k = 1) = p$  e  $P(X_k = 0) = 1 - p$ ,  $k = 1, \dots, n$ . Sendo  $S_n = X_1 + \dots + X_n$ , temos  $E[S_n] = np$  e  $\text{Var}(S_n) = np(1 - p)$ , com  $S_n/n$  fornecendo uma aproximação de  $p$ , que é desconhecida. Queremos  $n$  tal que

$$P\left(\left|\frac{S_n}{n} - p\right| < 0,05\right) > 0,95 \Leftrightarrow P\left(\left|\frac{S_n}{n} - p\right| \geq 0,05\right) \leq 0,05.$$

Por Chebyshev, temos

$$P\left(\left|\frac{S_n}{n} - p\right| > 0,05\right) \leq \frac{p(1-p)}{n(0,05)^2} \leq \frac{1}{4n(0,05)^2} \leq 0,05 \implies n \geq 2000.$$

### 8.3 Teorema do Limite Central

**Teorema 8.6 (Limite Central).** Dada uma sequência  $X_1, X_2, \dots$  de v.a.'s i.i.d. com  $E[X_1] = \mu < \infty$  e  $0 \neq \text{Var}(X_1) = \sigma^2 < \infty$ , vale

$$\lim_{n \rightarrow \infty} P\left(\frac{S_n - E[S_n]}{\sqrt{\text{Var}(S_n)}} \leq x\right) = \Phi(x), \quad \forall x \in \mathbb{R},$$

sendo  $S_n = X_1 + \dots + X_n$  e  $\Phi$  a f.d. de  $N(0, 1)$ .

Note que  $\frac{S_n - E[S_n]}{\sqrt{\text{Var}(S_n)}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$ , de modo que o TLC nos diz que  $\frac{S_n - n\mu}{\sigma\sqrt{n}}$  converge **em distribuição** para  $N(0, 1)$ . Podemos ainda enxergar o TLC como uma aproximação através da normal: para  $n$  suficientemente grande,

$$\Phi(x) \approx P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = P(S_n \leq x\sigma\sqrt{n} + n\mu),$$

ou seja,

$$P(S_n \leq x) = P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq \frac{x - n\mu}{\sigma\sqrt{n}}\right) \approx \Phi\left(\frac{x - n\mu}{\sigma\sqrt{n}}\right), \quad \forall x \in \mathbb{R}.$$

Note também que como  $\Phi$  é simétrica, temos

$$\begin{aligned}
 P\left(\left|\frac{S_n}{n} - \mu\right| \leq \varepsilon\right) &= P\left(\frac{S_n}{n} \leq \mu + \varepsilon\right) - P\left(\frac{S_n}{n} < \mu - \varepsilon\right) \\
 &= P(S_n \leq n\mu + n\varepsilon) - P(S_n < n\mu - n\varepsilon) \\
 &= P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq \frac{n\varepsilon}{\sigma\sqrt{n}}\right) - P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} < -\frac{n\varepsilon}{\sigma\sqrt{n}}\right) \\
 &= \Phi\left(\frac{n\varepsilon}{\sigma\sqrt{n}}\right) - \Phi\left(-\frac{n\varepsilon}{\sigma\sqrt{n}}\right) \\
 &= 2\Phi\left(\frac{n\varepsilon}{\sigma\sqrt{n}}\right) - 1.
 \end{aligned}$$

**Exemplo.** Suponha que a duração de um certo componente eletrônico distribui-se exponencialmente, com duração média de 2 meses. Quando o componente queima, instala-se outro do mesmo tipo em seu lugar. Estamos interessados na probabilidade de que o 30º componente não queime antes de 3 anos.

Seja  $X_i$  o tempo de duração, em meses, de um componente  $i$ . Temos  $X_i \sim \text{Exp}(\lambda)$ ,  $\lambda = 1/2$ ,  $E[X_i] = 2$  e  $\text{Var}(X_i) = 4$ ,  $i = 1, 2, \dots$ ; temos  $X_1, X_2, \dots$  i.i.d. e  $S_n = X_1 + \dots + X_n$  é o tempo necessário para que o  $n$ -ésimo componente queime. Temos  $E[S_n] = 2n$  e  $\text{Var}(S_n) = 4n$ , e queremos estimar  $P(S_{30} > 36)$ . Uma primeira maneira de calcular é notar que  $S_n \sim \Gamma(n, \lambda)$ , donde

$$P(S_n > x) = \sum_{k=0}^{n-1} e^{-\lambda x} \frac{(\lambda x)^k}{k!} \implies \sum_{k=0}^{29} e^{-18} \frac{18^k}{k!} \approx 0,9941 = P(S_{30} > 36).$$

Uma segunda maneira seria usar o TLC:

$$\begin{aligned}
 P(S_n > x) &= P\left(\frac{S_n - n/\lambda}{\sqrt{n}/\lambda} > \frac{x - n/\lambda}{\sqrt{n}/\lambda}\right) \approx 1 - \Phi\left(\frac{x - n/\lambda}{\sqrt{n}/\lambda}\right) \\
 \implies P(S_{30} > 36) &\approx 1 - \Phi\left(\frac{36 - 2 \cdot 30}{2\sqrt{30}}\right) \approx 1 - \Phi(-2, 19) = \Phi(2, 19) \approx 0,9857,
 \end{aligned}$$

em que o valor de  $\Phi(2, 19)$  foi retirado da tabela ao final da subseção. Note que o erro entre as duas formas é de aproximadamente 0,8%.

**Exemplo.** Numa determinada rota doméstica são utilizados aviões com 90 assentos. Verificou-se que cerca de 20% dos passageiros com reserva marcada não comparecem para o voo. Por isso, a companhia adotou a estratégia de confirmar a reserva de 100 passageiros em cada voo. Pergunta-se (a) qual a probabilidade aproximada de no máximo 90 passageiros que confirmaram reserva compareçam para o voo? e (b) qual a probabilidade aproximada de que o número de passageiros que confirmaram reserva e compareceram ao voo esteja entre 85 e 90?

(a) Defina

$$X_i = \begin{cases} 1, & i - \text{ésimo passageiro com reserva comparece} \\ 0, & \text{c.c.} \end{cases}, i = 1, \dots, 100.$$

Logo,  $X_1, \dots, X_{100}$  são i.i.d. com  $P(X_i = 1) = 0,8$  e  $P(X_i = 0) = 0,2$ . Seja  $S_n = X_1 + \dots + X_n$ , que representa o número de passageiros, entre os  $n$ , que fizeram reserva e compareceram para o voo. Temos  $E[X_i] = 0,8 = \mu$  e  $E[X_i^2] = 0,8$ , donde  $\text{Var}(X_i) = 0,16 = \sigma^2$ . Ademais,  $E[S_n] = n\mu$  e  $\text{Var}(S_n) = n\sigma^2$ . Por Chebyshev,

$$\begin{aligned}
 P(S_{100} \leq 90) &= P(S_{100} - 80 \leq 10) \geq P(|S_{100} - 80| \leq 10) \\
 &= 1 - P(|S_{100} - 80| > 10) \\
 &\geq 1 - \frac{100 \cdot 0,16}{100} \\
 &= 0,84.
 \end{aligned}$$

Usando o TLC, temos

$$\begin{aligned} P(S_{100} \leq 90) &= P\left(\frac{S_{100} - 80}{\sqrt{16}} \leq \frac{90 - 80}{\sqrt{16}}\right) \\ &= P\left(\frac{S_{100} - 80}{4} \leq 2,5\right) \\ &\approx \Phi(2,5) \\ &\approx 0,9938. \end{aligned}$$

(b) Pelo TLC

$$\begin{aligned} P(85 < S_{100} < 90) &= P(S_{100} < 90) - P(S_{100} < 85) \\ &= P\left(\frac{S_{100} - 80}{4} < 2,5\right) - P\left(\frac{S_{100} - 80}{4} < 1,25\right) \\ &\approx \Phi(2,5) - \Phi(1,25) \\ &\approx 0,9938 - 0,8944 \\ &= 0,0994 \\ &\approx 10\%. \end{aligned}$$

**Exemplo.** Decide-se tomar uma amostra de tamanho  $n$  para determinar a porcentagem de eleitores que têm intenção de votar num determinado candidato. Considere

$$X_k = \begin{cases} 1, & k - \text{ésimo eleitor vota no candidato} \\ 0, & \text{c.c.} \end{cases}, \quad k = 1, \dots, n.$$

Suponha  $X_1, \dots, X_n$  i.i.d. com  $P(X_k = 1) = p$  e  $P(X_k = 0) = 1 - p$ ,  $k = 1, \dots, n$ . Daí,  $\mu = E[X_1] = p$  e  $\sigma^2 = \text{Var}(X_1) = p(1 - p)$ . Note que  $\sigma^2$  é máximo para  $p = 1/2$ , valendo  $1/4$ . Consideraremos este valor para  $p$ . A v.a.  $S_n/n$  representa a proporção (ou frequência relativa) de eleitores que têm a intenção de votar no candidato, e pode ser usada para estimar a proporção verdadeira  $p$ .

(a) Suponha  $n = 900$ . Temos

$$\begin{aligned} P\left(\left|\frac{S_n}{n} - p\right| \geq 0,025\right) &= 1 - P\left(\left|\frac{S_n}{n} - p\right| < 0,025\right) \\ &\approx 1 - \left[2\Phi\left(\frac{0,025n}{\sigma\sqrt{n}}\right) - 1\right] \\ &= 2\left[1 - \Phi\left(\frac{900 \cdot 0,025}{1/2\sqrt{900}}\right)\right] \\ &= 2(1 - \Phi(1,5)) \\ &\approx 2(1 - 0,9332) \\ &= 0,134 \therefore P\left(\left|\frac{S_n}{n} - p\right| < 0,025\right) \approx 0,87. \end{aligned}$$

(b) Suponha  $n = 900$ . Queremos determinar  $\varepsilon$  tal que

$$P\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) = 0,01$$

ou seja, tal que

$$\Phi\left(\frac{\sqrt{900}\varepsilon}{0,5}\right) = 0,995,$$

donde devemos ter

$$\frac{\sqrt{900}\varepsilon}{0,5} \approx 2,58 \implies \varepsilon \approx 0,043,$$

ou seja,

$$P\left(\left|\frac{S_n}{n} - p\right| < 0,043\right) \approx 0,99.$$

Dito de outro modo, tem-se aproximadamente 99% de chance de garantir um erro menor que

0,043 quando estima-se  $p$  por  $S_n/n$  com  $n = 900$ .

(c) Queremos determinar  $n$  tal que

$$P\left(\left|\frac{S_n}{n} - p\right| \geq 0,025\right) = 0,01,$$

ou seja, tal que

$$\Phi\left(\frac{0,025\sqrt{n}}{0,5}\right) = 0,995,$$

isto é,

$$\frac{0,025\sqrt{n}}{0,5} \approx 2,58 \implies n \approx 2663.$$

$z$	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998

Tabela 1:  $\Phi(z)$  para  $0,00 \leq z \leq 3,59$ .

## 8.4 Funções Características

**Definição 8.7** (Função característica). A **função característica** de uma v.a.  $X$  é definida por

$$\varphi_X(t) = E[e^{itX}], t \in \mathbb{R}.$$

Funções características são um pouco mais complicadas que funções geradoras de momentos por envolverem números complexos; entretanto, elas têm duas vantagens importantes sobre f.g.m.'s: primeiro,  $\varphi_X(t)$  é finita para toda v.a.  $X$  e todo  $t \in \mathbb{R}$ ; segundo, a f.d. de  $X$  (e geralmente a densidade, se existir) pode ser obtida a partir da função característica através de uma “fórmula de inversão”. Usando propriedades das funções características poderemos provar a LfGN e o TLC. Assumimos, daqui em diante, conhecimentos sobre números complexos e o básico de derivadas complexas.

Uma v.a. complexa  $Z$  pode ser escrita como  $X+iY$ , sendo  $X$  e  $Y$  v.a.'s reais. As mesmas propriedades da esperança continuam válidas para  $Z$ , notando que  $EZ = EX + iEY$ . Reservaremos  $X$  e  $Y$  para denotar v.a.'s reais. Suponha que  $X$  é uma v.a. e  $t \in \mathbb{R}$  é constante (reservamos  $t$  para denotar constantes reais). Então  $|e^{itX}| = 1$ , de modo que  $e^{itX}$  tem esperança finita e a função característica de  $X$  é bem definida. Note que  $\varphi_X(0) = E[1] = 1$  e, dado  $t \in \mathbb{R}$ ,

$$|\varphi_X(t)| = |E[e^{itX}]| \leq E[|e^{itX}|] = E[1] = 1.$$

A explicação para que as funções características sejam finitas para todo  $t \in \mathbb{R}$  enquanto as f.g.m.'s não são finitas em geral é que  $e^{it}$  é limitado e  $e^t$  não.

**Exemplo.** Seja  $X$  uma v.a. que toma o valor  $a$  com probabilidade 1. Então

$$\varphi_X(t) = E[e^{itX}] = e^{ita}, t \in \mathbb{R}.$$

Em particular, se  $X$  toma o valor 0 com probabilidade 1, então sua função característica é idênticamente igual a 1.

Se  $X$  é v.a. e  $a, b$  são reais quaisquer, então

$$\varphi_{a+bX}(t) = E[e^{it(a+bX)}] = E[e^{ita}e^{ibtX}] = e^{ita}E[e^{ibtX}],$$

logo

$$\varphi_{a+bX}(t) = e^{ita}\varphi_X(bt), t \in \mathbb{R}.$$

**Exemplo.** Seja  $U$  uniformemente distribuída em  $(-1, 1)$ . Então dado  $t \neq 0$ ,

$$\begin{aligned} \varphi_U(t) &= \int_{-1}^1 e^{itu} \frac{1}{2} du \\ &= \frac{1}{2} \left( \frac{e^{it} - e^{-it}}{it} \right) \\ &= \frac{\sin t}{t}. \end{aligned}$$

Dados  $a < b$ , seja

$$X = \frac{a+b}{2} + \frac{b-a}{2}U.$$

Temos  $X$  uniformemente distribuída em  $(a, b)$  e, para  $t \neq 0$ ,

$$\varphi_X(t) = e^{it(a+b)/2} \frac{\sin((b-a)t/2)}{(b-a)t/2}.$$

Por outro lado,

$$\begin{aligned} \varphi_X(t) &= \int_a^b e^{itx} \frac{1}{b-a} dx \\ &= \frac{e^{ibt} - e^{iat}}{it(b-a)}. \end{aligned}$$

Não é muito complicado verificar que os dois valores concordam.



**Exemplo.** Seja  $X$  uma v.a. exponencial de parâmetro  $\lambda$ . Então

$$\begin{aligned}\varphi_X(t) &= \int_0^\infty e^{itx} \lambda e^{-\lambda x} dx \\ &= \lambda \int_0^\infty e^{-(\lambda-it)x} dx \\ &= \frac{\lambda}{\lambda-it} e^{-(\lambda-it)x} \Big|_0^\infty.\end{aligned}$$

Como  $\lim_{x \rightarrow \infty} e^{-\lambda x} = 0$  e  $e^{itx}$  é limitado em  $x$ , segue que

$$\lim_{x \rightarrow \infty} e^{-(\lambda-it)x} = \lim_{x \rightarrow \infty} e^{-\lambda x} e^{itx} = 0.$$

Logo,

$$\varphi_X(t) = \frac{\lambda}{\lambda-it}.$$

Suponha que  $X$  e  $Y$  são v.a.'s independentes. Então  $e^{itX}$  e  $e^{itY}$  também são independentes, de modo que

$$\varphi_{X+Y}(t) = E[e^{it(X+Y)}] = E[e^{itX} e^{itY}] = E[e^{itX}]E[e^{itY}],$$

ou seja,

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t), t \in \mathbb{R}.$$

Essa fórmula se estende facilmente para nos dar o fato de que a função característica de uma soma finita de v.a.'s independentes é o produto das funções características individuais.

Podemos mostrar também que  $\varphi_X(t)$  é contínua em  $t$ ; ademais, se  $X$  tem  $n$ -ésimo momento finito, então  $\varphi_X^{(n)}(t)$  existe, é contínua em  $t$  e pode ser calculada por

$$\varphi_X^{(n)}(t) = \frac{d^n}{dt^n} E[e^{itX}] = E\left[\frac{d^n}{dt^n} e^{itX}\right] = E[(iX)^n e^{itX}].$$

Em particular,

$$\varphi_X^{(n)}(0) = i^n E[X^n].$$

Expandindo  $\varphi_X(t)$  em série de potências, temos

$$\varphi_X(t) = E[e^{itX}] = E\left[\sum_{n=0}^{\infty} \frac{(itX)^n}{n!}\right] = \sum_{n=0}^{\infty} \frac{i^n E[X^n]}{n!} t^n.$$

**Exemplo.** Seja  $X$  normalmente distribuída com média 0 e variância  $\sigma^2$ . Vimos que  $E[X^n] = 0$  se  $n$  é ímpar e, se  $n$  é par,

$$E[X^n] = E[X^{2k}] = \frac{\sigma^{2k} (2k)!}{2^k k!}.$$

Logo,

$$\varphi_X(t) = \sum_{k=0}^{\infty} \frac{i^{2k} E[X^{2k}]}{(2k)!} t^{2k} = \sum_{k=0}^{\infty} \frac{(-\sigma^2 t^2/2)^k}{k!} = e^{-\sigma^2 t^2/2}.$$

De forma geral, se  $X \sim N(\mu, \sigma^2)$ , então  $X = \mu + Y$  com  $Y \sim N(0, \sigma^2)$ . Daí,

$$\varphi_X(t) = e^{it\mu} e^{-\sigma^2 t^2/2}, t \in \mathbb{R}.$$

Seja  $X$  v.a. com f.g.m.  $M_X(t)$  finita em  $(-t_0, t_0)$  para algum  $t_0 > 0$ . Como

$$M_X(t) = E[e^{tX}]$$

e

$$\varphi_X(t) = E[e^{itX}],$$

é razoável esperar que

$$\varphi_X(t) = M_X(it).$$

Dito de outro modo, é razoável esperar que trocando  $t$  por  $it$  na f.g.m. obtemos a fórmula da função característica. Esse de fato é o caso, mas um entendimento adequado da teoria envolve o conceito de continuação analítica da teoria de funções complexas. Para exemplificar, seja  $X \sim N(\mu, \sigma^2)$ . Vimos

que

$$M_X(t) = e^{\mu t} e^{\sigma^2 t^2 / 2},$$

de modo que

$$M_X(it) = e^{\mu(it)} e^{\sigma^2(it)^2 / 2} = e^{i\mu t} e^{-\sigma^2 t^2 / 2},$$

como vimos.

## 8.5 Fórmulas de inversão e o Teorema da Continuidade

Seja  $X$  uma v.a. inteira. Sua função característica é dada por

$$\varphi_X(t) = \sum_{j \in \mathbb{Z}} e^{ijt} f_X(j).$$

Uma das propriedades mais úteis de  $\varphi_X(t)$  é que ela pode ser usada para calcular  $f_X(k)$ . Mais especificamente, temos a “fórmula de inversão”

$$f_X(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ikt} \varphi_X(t) dt.$$

Para verificá-la, escrevemos o lado direito como

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ikt} \left[ \sum_{j \in \mathbb{Z}} e^{ijt} f_X(j) \right] dt.$$

Pode-se mostrar que é permitido trocar a integral com a série, de maneira que temos

$$\sum_{j \in \mathbb{Z}} f_X(j) \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(j-k)t} dt.$$

Para completar a demonstração, precisamos mostrar que essa última expressão é igual a  $f_X(k)$ . Para isso, é suficiente mostrar que

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(j-k)t} dt = \begin{cases} 1, & j = k \\ 0, & j \neq k \end{cases}.$$

É claro que para  $j = k$  vale o exposto acima. Se  $j \neq k$ , então

$$\begin{aligned} \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(j-k)t} dt &= \frac{e^{i(j-k)\pi} - e^{-i(j-k)\pi}}{2\pi i(j-k)} \\ &= \frac{\sin((j-k)\pi)}{\pi(j-k)} \\ &= 0, \end{aligned}$$

como queríamos.

**Exemplo.** Sejam  $X_1, \dots, X_n$  v.a.’s inteiras i.i.d. e  $S_n = X_1 + \dots + X_n$ . Então  $\varphi_{S_n}(t) = (\varphi_{X_1}(t))^n$  e, pela fórmula de inversão,

$$f_{S_n}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ikt} (\varphi_{X_1}(t))^n dt.$$

Essa igualdade é a base de quase todos os métodos para a análise do comportamento de  $f_{S_n}(k)$  para  $n$  grande e, em particular, é a base da prova do TLC “local”.

Existe, ainda, um análogo para v.a.’s contínuas da fórmula de inversão. Seja  $X$  v.a. cuja função característica é integrável, ou seja,

$$\int_{\mathbb{R}} |\varphi_X(t)| dt < \infty.$$

Pode-se mostrar que nesse caso  $X$  é v.a. contínua com densidade  $f_X$  dada por

$$f_X(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-ixt} \varphi_X(t) dt.$$

**Exemplo.** Seja  $X \sim N(0, \sigma^2)$ . Vamos mostrar que a fórmula de inversão acima é válida para  $X$ .

Sabemos que  $\varphi_X(t) = e^{-\sigma^2 t^2/2}$ . Logo, por definição,

$$e^{-\sigma^2 t^2/2} = \int_{\mathbb{R}} e^{itx} \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} dx.$$

Se substituirmos  $t$  por  $-t$  e  $\sigma$  por  $1/\sigma$ , obtemos

$$e^{-t^2/2\sigma^2} = \int_{\mathbb{R}} e^{-itx} \frac{\sigma}{\sqrt{2\pi}} e^{-\sigma^2 x^2/2} dx$$

ou, equivalentemente,

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-t^2/2\sigma^2} = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} e^{-\sigma^2 x^2/2} dx.$$

Finalmente, trocando  $x$  por  $t$  na última igualdade, obtemos

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} e^{-\sigma^2 t^2/2} dt,$$

que nada mais é que a fórmula de inversão no caso particular de  $X \sim N(0, \sigma^2)$ .

Seja  $X$  v.a. qualquer e seja  $Y$  v.a. independente de  $X$  com distribuição normal padrão. Seja ainda  $c > 0$  uma constante positiva. Então  $X + cY$  tem função característica

$$\varphi_{X+cY}(t) = e^{-c^2 t^2/2}.$$

Como  $\varphi_X(t)$  é limitada em módulo por 1 e  $e^{-c^2 t^2/2}$  é integrável, segue que  $X + cY$  tem uma função característica integrável. Logo, podemos aplicar a fórmula de inversão e  $X + cY$  é v.a. contínua com densidade dada por

$$\varphi_{X+cY}(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \varphi_X(t) e^{-c^2 t^2/2} dt.$$

Integrando ambos os lados em  $[a, b]$  e trocando a ordem de integração, concluímos que

$$\begin{aligned} P(a \leq X + cY \leq b) &= \frac{1}{2\pi} \int_a^b \left( \int_{\mathbb{R}} e^{-itx} \varphi_X(t) e^{-c^2 t^2/2} dt \right) dx \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \left( \int_a^b e^{-itx} dx \right) \varphi_X(t) e^{-c^2 t^2/2} dt \end{aligned}$$

ou

$$P(a \leq X + cY \leq b) = \frac{1}{2\pi} \int_{\mathbb{R}} \left( \frac{e^{-ibt} - e^{-iat}}{-it} \right) \varphi_X(t) e^{-c^2 t^2/2} dt.$$

A importância desta última equação é que ela é válida para qualquer v.a.  $X$ . O lado direito dela depende apenas de  $X$  através de  $\varphi_X(t)$ . Usando esse fato e fazendo  $c \rightarrow 0$ , pode-se mostrar que a f.d. de  $X$  é determinada por sua função característica. Esse resultado é conhecido como **teorema da unicidade**, e pode ser enunciado como segue.

**Teorema 8.8 (Unicidade).** Se duas v.a.'s têm a mesma função característica, elas têm a mesma função de distribuição.

**Exemplo.** Vamos usar o teorema da unicidade para mostrar que a soma de duas v.a.'s normais independentes é também normal. Sejam  $X$  e  $Y$  independentes com distribuições  $N(\mu_1, \sigma_1^2)$  e  $N(\mu_2, \sigma_2^2)$ , respectivamente. Então

$$\varphi_X(t) = e^{i\mu_1 t} e^{-\sigma_1^2 t^2/2}$$

e

$$\varphi_Y(t) = e^{i\mu_2 t} e^{-\sigma_2^2 t^2/2},$$

de modo que

$$\varphi_{X+Y}(t) = \varphi_X(t) \varphi_Y(t) = e^{i(\mu_1 + \mu_2)t} e^{-(\sigma_1^2 + \sigma_2^2)t^2/2}.$$

Logo, a função característica de  $X + Y$  é a mesma que a de uma v.a. com distribuição normal de média  $\mu_1 + \mu_2$  e variância  $\sigma_1^2 + \sigma_2^2$ . Pelo teorema da unicidade,  $X + Y$  deve ter essa distribuição normal.

A aplicação mais importante da fórmula de inversão “contínua” é que ela pode ser usada para derivar o resultado a seguir, que é a base da prova da LfGN e do TLC.

**Teorema 8.9** (Continuidade). Sejam  $X_n, n \geq 1$  e  $X$  v.a.'s tais que

$$\lim_{n \rightarrow \infty} \varphi_{X_n}(t) = \varphi_X(t), t \in \mathbb{R}.$$

Então

$$\lim_{n \rightarrow \infty} F_{X_n}(t) = F_X(x)$$

em todos os pontos onde  $F_X$  é contínua.

Esse teorema diz que a convergência de funções características implica a convergência das funções de distribuição correspondentes ou, dito de outro modo, que as funções de distribuição “dependem continuamente” de suas funções características. Por isso, esse teorema é conhecido por **teorema da continuidade**. Sua demonstração é trabalhosa e não será apresentada aqui.

## 8.6 Demonstração da Lei Fraca dos Grandes Números e do Teorema do Limite Central

Nessa seção, usaremos o teorema da continuidade para provar a LfGN e o TLC. Para tal, precisamos discutir primeiro o comportamento assintótico de  $\log \varphi_X(t)$  próximo de  $t = 0$ .

Seja  $z \in \mathbb{C}$  tal que  $|z - 1| < 1$ . Podemos definir  $\log z$  pela série de potências (para  $|z - 1| \geq 1$ , outras definições de  $\log z$  são necessárias)

$$\log z = (z - 1) - \frac{(z - 1)^2}{2} + \frac{(z - 1)^3}{3} - \dots$$

Com essa definição, temos as propriedades usuais que  $\log 1 = 0$ ,  $e^{\log z} = z$ ,  $|z - 1| < 1$ , e se  $h(t)$ ,  $a < t < b$  é uma função holomorfa tal que  $|h(t) - 1| < 1$ , então

$$\frac{d}{dt} \log h(t) = \frac{h'(t)}{h(t)}.$$

Seja  $X$  v.a. com função característica  $\varphi_X$ . Então  $\varphi_X$  é contínua e  $\varphi_X(0) = 1$ . Logo,  $\log \varphi_X(t)$  é bem definida para  $t$  próximo de 0 e  $\log \varphi_X(0) = 0$ .

Suponha agora que  $X$  tem média finita  $\mu$ . Então  $\varphi_X(t)$  é diferenciável e  $\varphi'_X(0) = i\mu$ . Logo,

$$\begin{aligned} \lim_{t \rightarrow 0} \frac{\log \varphi_X(t)}{t} &= \lim_{t \rightarrow 0} \frac{\log \varphi_X(t) - \log \varphi_X(0)}{t - 0} \\ &= \frac{\varphi'_X(0)}{\varphi_X(0)} \\ &= i\mu. \end{aligned}$$

Consequentemente,

$$\lim_{t \rightarrow 0} \frac{\log \varphi_X(t) - i\mu t}{t} = 0.$$

Suponha que  $X$  tem variância finita  $\sigma^2$ . Então  $\varphi_X(t)$  é duas vezes diferenciável e

$$\varphi''_X(0) = -E[X^2] = -(\mu^2 + \sigma^2).$$

Podemos aplicar a regra de l'Hôpital para obter

$$\begin{aligned} \lim_{t \rightarrow 0} \frac{\log \varphi_X(t) - i\mu t}{t^2} &= \lim_{t \rightarrow 0} \frac{\varphi'_X(t)/\varphi_X(t) - i\mu}{2t} \\ &= \lim_{t \rightarrow 0} \frac{\varphi'_X(t) - i\mu\varphi_X(t)}{2t\varphi_X(t)} \\ &= \lim_{t \rightarrow 0} \frac{\varphi'_X(t) - i\mu\varphi_X(t)}{2t}. \end{aligned}$$

Aplicando a regra de l'Hôpital novamente, temos

$$\lim_{t \rightarrow 0} \frac{\log \varphi_X(t) - i\mu t}{t^2} = \frac{\varphi_X''(0) - i\mu\varphi_X'(0)}{2} \quad (2)$$

$$= \frac{-(\mu^2 + \sigma^2) - (i\mu)^2}{2} \quad (3)$$

$$= \frac{-\sigma^2}{2}. \quad (4)$$

**Teorema 8.10** (LfGN). Sejam  $X_1, \dots, X_n$  v.a.'s i.i.d. com média finita  $\mu$  e  $S_n = X_1 + \dots + X_n$ . Então para todo  $\varepsilon < 0$ ,

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) = 0.$$

**Demonstração.** A função característica de  $S_n/n - \mu$  é

$$e^{-i\mu t} (\varphi_{X_1}(t/n))^n. \quad (5)$$

Fixe  $t$ . Então para  $n$  suficientemente grande,  $t/n$  está suficientemente próximo de 0 de modo que  $\log \varphi_{X_1}(t/n)$  é bem definida e

$$e^{-i\mu t} (\varphi_{X_1}(t/n))^n = \exp[n(\log \varphi_{X_1}(t/n) - i\mu(t/n))].$$

Afirmamos que

$$\lim_{n \rightarrow \infty} n(\log \varphi_{X_1}(t/n) - i\mu(t/n)) = 0.$$

Essa equação é imediata para  $t = 0$  já que  $\log \varphi_{X_1}(0) = \log 1 = 0$ . Se  $t \neq 0$ , podemos escrever o lado esquerdo como

$$t \lim_{n \rightarrow \infty} \frac{\log \varphi_X(t/n) - i\mu(t/n)}{t/n}.$$

Mas  $t/n \xrightarrow{n \rightarrow \infty} 0$ , de modo que o último limite é 0, e está mostrado o limite desejado. Segue então que a função característica de  $S_n/n - \mu$  se aproxima de 1 quando  $n \rightarrow \infty$ . Agora, 1 é a função característica de uma v.a.  $X$  tal que  $P(X = 0) = 1$ . A f.d. de  $X$  é dada por

$$F_X(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}.$$

A f.d. é contínua em todo ponto menos em  $x = 0$ . Tome  $\varepsilon > 0$ . Pelo teorema da continuidade,

$$\lim_{n \rightarrow \infty} P(S_n/n - \mu \leq -\varepsilon) = F_X(-\varepsilon) = 0$$

e

$$\lim_{n \rightarrow \infty} P(S_n/n - \mu \leq \varepsilon) = F_X(\varepsilon) = 1.$$

Esse último resultado implica que

$$\lim_{n \rightarrow \infty} P\left(\frac{S_n}{n} - \mu > \varepsilon\right) = 0$$

que, junto com o penúltimo limite, implica a LfGN.  $\square$

**Teorema 8.11** (Limite Central). Sejam  $X_1, X_2, \dots$  v.a.'s i.i.d. com média finita  $\mu$  e variância não nula  $\sigma^2$ . Então

$$\lim_{n \rightarrow \infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x), x \in \mathbb{R}.$$

**Demonstração.** Faça

$$S_n^* = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

Então, para  $t$  fixo e  $n$  suficientemente grande,

$$\varphi_{S_n^*}(t) = e^{-in\mu t/\sigma\sqrt{n}} \varphi_{S_n}(t/\sigma\sqrt{n}) = e^{-in\mu t/\sigma\sqrt{n}} (\varphi_{X_1}(t/\sigma\sqrt{n}))^n,$$

ou

$$\varphi_{S_n^*}(t) = \exp\left[n(\log \varphi_{X_1}(t/\sigma\sqrt{n}) - i\mu(t/\sigma\sqrt{n}))\right].$$

Afirmamos que

$$\lim_{n \rightarrow \infty} n(\log \varphi_{X_1}(t/\sigma\sqrt{n}) - i\mu(t/\sigma\sqrt{n})) = -t^2/2.$$

Se  $t = 0$ , então ambos os lados são 0 e a equação vale. Se  $t \neq 0$ , podemos escrever o lado esquerdo como

$$\frac{t^2}{\sigma^2} \lim_{n \rightarrow \infty} \frac{\log \varphi_{X_1}(t/\sigma\sqrt{n}) - i\mu(t/\sigma\sqrt{n})}{(t/\sigma\sqrt{n})^2},$$

que é igual a

$$\frac{t^2}{\sigma^2} \left(-\frac{\sigma^2}{2}\right) = -\frac{t^2}{2}.$$

Logo a equação vale para todo  $t$ . Segue daí que

$$\lim_{n \rightarrow \infty} \varphi_{S_n^*}(t) = e^{-t^2/2}, t \in \mathbb{R}.$$

Ora, mas essa é a função característica de uma v.a.  $X$  com distribuição normal padrão e f.d.  $\Phi(x)$ . Logo, pelo teorema da continuidade,

$$\lim_{n \rightarrow \infty} P(S_n^* \leq x) = \Phi(x), x \in \mathbb{R},$$

como queríamos mostrar. □

## 8.7 Exercícios - esperança e momentos de v.a.'s contínuas

- Sejam  $X$  e  $Y$  v.a.'s independentes tais que  $X \sim \Gamma(\alpha_1, \lambda)$  e  $Y \sim \Gamma(\alpha_2, \lambda)$ . Considere  $Z = Y/X$ .
  - Determine para quais valores de  $\alpha_1$  e  $\alpha_2$  teremos  $EZ$  finita e calcule  $EZ$  neste caso. Determine para quais valores de  $\alpha_1$  e  $\alpha_2$  teremos  $E[Z^2]$  finita e calcule  $\text{Var}(Z)$  neste caso.

**Solução.**

a) Temos

$$\begin{aligned} EZ &= \int_0^\infty \int_0^\infty \frac{y}{x} \cdot \frac{\lambda^{\alpha_2} y^{\alpha_2-1}}{\Gamma(\alpha_2)} e^{-\lambda y} \cdot \frac{\lambda^{\alpha_1} x^{\alpha_1-1}}{\Gamma(\alpha_1)} e^{-\lambda x} dx dy \\ &= \frac{\lambda^{\alpha_1}}{\Gamma(\alpha_1)} \int_0^\infty x^{\alpha_1-1-1} e^{-\lambda x} \frac{\alpha_2}{\lambda} dx \\ &= \frac{\lambda^{\alpha_1-1}}{\Gamma(\alpha_1)} \cdots \frac{\Gamma(\alpha_1-1)}{\lambda^{\alpha_1-1}} \alpha_2 \\ &= \frac{\alpha_2}{\alpha_1-1}, \end{aligned}$$

onde impomos as condições  $\alpha_2 > 0$  e  $\alpha_1 > 1$  para que as integrais existam.

b) Para  $\alpha_1 > 2$ , temos

$$\begin{aligned} EZ^2 &= \int_0^\infty \int_0^\infty \frac{y^2}{x^2} \cdot \frac{\lambda^{\alpha_2} y^{\alpha_2-1}}{\Gamma(\alpha_2)} e^{-\lambda y} \frac{\lambda^{\alpha_1} x^{\alpha_1-1}}{\Gamma(\alpha_1)} e^{-\lambda x} dx dy \\ &= \frac{\lambda^{\alpha_1}}{\Gamma(\alpha_1)} \cdot \frac{\alpha_2(\alpha_2+1)}{\lambda^2} \int_0^\infty x^{\alpha_1-2-1} e^{-\lambda x} dx \\ &= \frac{\lambda^{\alpha_1-1}}{\Gamma(\alpha_1)} \alpha_2(\alpha_2+1) \cdot \frac{\Gamma(\alpha_1-2)}{\lambda^{\alpha_1-2}} \\ &= \frac{\alpha_2(\alpha_2+1)}{(\alpha_1-1)(\alpha_1-2)}, \end{aligned}$$

onde impomos a condição  $\alpha_1 > 2$  para que a integral exista. Daí, temos

$$\text{Var}(Z) = \frac{\alpha_2(\alpha_2+1)}{(\alpha_1-1)(\alpha_1-2)} - \frac{\alpha_2}{(\alpha_1-1)^2} = \frac{\alpha_2(\alpha_1+\alpha_2-1)}{(\alpha_1-1)^2(\alpha_1-2)}.$$

□

- Seja  $X \sim \chi^2(n)$ , ou seja,  $X \sim \Gamma(n/2, 1/2)$ . Calcule a esperança de  $Y = \sqrt{X}$ .

**Solução.** Temos

$$\begin{aligned} EY &= \int_{\mathbb{R}} \sqrt{x} f_X(x) dx \\ &= \int_0^{\infty} \frac{x^{1/2} (1/2)^{n/2}}{\Gamma(n/2)} x^{n/2-1} e^{-x/2} dx \\ &= \frac{(1/2)^{n/2}}{\Gamma(n/2)} \int_0^{\infty} x^{(n+1)/2-1} e^{-x/2} dx \\ &= \sqrt{2} \frac{\Gamma((n+1)/2)}{\Gamma(n/2)}. \end{aligned}$$

□

3. Sejam  $U_1$  e  $U_2$  v.a.'s i.i.d. com distribuição comum  $\text{Exp}(\lambda)$  e seja  $Y = \max(U_1, U_2)$ . Obtenha a esperança e a variância de  $Y$ .

**Solução.** Da independência de  $U_1$  e  $U_2$ , temos que

$$F_Y(y) = F_U(y)F_V(y) = (1 - e^{-\lambda y})^2,$$

para todo  $y \geq 0$  e  $F_Y(y)$  caso contrário. Portanto,

$$f_Y(y) = \begin{cases} 2\lambda e^{-\lambda y}(1 - e^{-\lambda y}), & y \geq 0 \\ 0, & \text{c.c.} \end{cases}$$

Daí, segue que

$$\begin{aligned} EY &= \int_{\mathbb{R}} y f_Y(y) dy \\ &= 2 \left( \int_0^{\infty} y \lambda e^{-\lambda y} dy - \int_0^{\infty} y \lambda e^{-2\lambda y} dy \right) \\ &= 2 \left( \frac{1}{\lambda} - \frac{1}{4\lambda} \right) \\ &= \frac{3}{2\lambda}. \end{aligned}$$

Temos também

$$\begin{aligned} EY^2 &= 2 \left( \int_0^{\infty} y^2 \lambda e^{-\lambda y} dy - \int_0^{\infty} y^2 \lambda e^{-2\lambda y} dy \right) \\ &= 2 \left( \frac{2}{\lambda^2} - \frac{1}{4\lambda^2} \right) \\ &= \frac{7}{2\lambda^2}. \end{aligned}$$

Portanto,

$$\text{Var}(Y) = \frac{5}{4\lambda^2}.$$

□

4. Seja  $X = \sin \Theta$ , em que  $\Theta \sim U(-2, 2)$ . Determine  $EX$  e  $\text{Var}(X)$ .

**Solução.** Temos

$$\begin{aligned} EX &= \int_{\mathbb{R}} \sin \theta f_{\Theta}(\theta) d\theta \\ &= \int_{-\pi/2}^{\pi/2} \sin \theta \cdot \frac{1}{\pi} d\theta \\ &= 0 \end{aligned}$$

e

$$\begin{aligned} EX^2 &= \int_{\mathbb{R}} \sin^2 \theta f_{\Theta}(\theta) d\theta \\ &= \int_{-\pi/2}^{\pi/2} \sin^2 \theta \cdot \frac{1}{\pi} d\theta \\ &= \frac{1}{2\pi} \int_{-\pi/2}^{\pi/2} 1 - \cos(2\theta) d\theta \\ &= 1/2. \end{aligned}$$

Portanto,  $\text{Var}(X) = 1/2$ . □

5. Seja  $X \sim N(0, \sigma^2)$ . Determine a esperança e a variância das seguintes v.a.'s:

- a)  $|X|$ ;
- b)  $X^2$ .

**Solução.**

a) Temos

$$\begin{aligned} E|X| &= 2 \int_0^{\infty} x f_X(x) dx \\ &= \sqrt{\frac{2}{\pi}} \sigma \int_0^{\infty} e^{-u} du \\ &= \sqrt{\frac{2}{\pi}} \sigma. \end{aligned}$$

Ademais,

$$\text{Var}(X) = E(X^2) - E|X|^2 = \sigma^2 \left(1 - \frac{2}{\pi}\right).$$

b) Do texto, sabemos que

$$EX^k = \begin{cases} \frac{k! \sigma^k}{(k/2)! 2^{k/2}}, & k \text{ par} \\ 0, & \text{c.c.} \end{cases}.$$

Portanto,  $E(X^2) = \sigma^2$  e  $\text{Var}(X^2) = E(X^4) - E^2(X^2) = 2\sigma^4$ . □

6. Sejam  $X$  e  $Y$  v.a.'s com densidade conjunta

$$f_{X,Y}(x,y) = \frac{\sqrt{15}}{4\pi} \exp\left[-\frac{x^2 - xy + 4y^2}{2}\right], (x,y) \in \mathbb{R}^2.$$

Determine o coeficiente de correlação entre  $X$  e  $Y$ .

**Solução.** Como  $EX = 0 = EY$ , temos que

$$\rho(X,Y) = \frac{E(XY)}{8/15}.$$

Como

$$\begin{aligned} E(XY) &= \iint_{\mathbb{R}^2} xy f_{X,Y}(x,y) dx dy \\ &= \frac{\sqrt{15}}{4\pi} \frac{\sqrt{2\pi}}{2} \sqrt{2\pi} \sqrt{\frac{4}{15}} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \cdot \frac{y^2}{\sqrt{4/15}} \exp(-15y^2/8) dy \\ &= \frac{2}{15}. \end{aligned}$$

Portanto,

$$\rho(X,Y) = \frac{1}{4}. \quad \square$$

7. Sejam  $X$  e  $Y$  v.a.'s independentes tais que  $X \sim N(\mu, \sigma^2)$  e  $Y \sim \Gamma(\alpha, \lambda)$ . Obtenha a esperança e a variância de  $Z = XY$ .



**Solução.** Como  $X$  e  $Y$  são independentes, temos  $E(Z) = EXEY = \mu\alpha/\lambda$ . Daí,

$$\text{Var}(Z) = (\sigma^2 + \mu^2) \frac{\alpha(\alpha + 1)}{\lambda^2} - \frac{\mu^2 \alpha^2}{\lambda^2} = \frac{\alpha}{\lambda} (\sigma^2 \alpha + \sigma^2 + \mu^2).$$

□

8. Sejam  $X$  e  $Y$  v.a.'s tais que  $EX = EY = 0$ ,  $\text{Var}(X) = \text{Var}(Y) = 1$  e  $\rho(X, Y) = \rho$ . Mostre que  $X - \rho Y$  e  $Y$  são não-correlacionadas,  $E[X - \rho Y] = 0$  e  $\text{Var}(X - \rho Y) = 1 - \rho^2$ .

**Solução.** Temos

$$\text{Cov}(X - \rho Y, Y) = E[(X - \rho Y)Y] - E[X - \rho Y]E[Y] = E[XY] - \rho E[Y^2].$$

Como  $EY = 0$  e  $\text{Var}(Y) = 1$ , temos  $E[Y^2] = 1$ . Ademais, como  $\rho(X, Y) = \text{Cov}(X, Y) = E[XY]$ , temos  $\text{Cov}(X - \rho Y, Y) = \rho - \rho = 0$ . Por fim,  $E[X - \rho Y] = EX - \rho EY = 0$  e  $\text{Var}(X - \rho Y) = E[X^2] - 2\rho E[XY] + \rho^2 E[Y^2] = 1 - \rho^2$ . □

9. Seja  $X \sim U(a, b)$ . Obtenha  $M_X(t)$ , a função geradora de momentos de  $X$ .

**Solução.** Temos

$$M_X(t) = \int_{\mathbb{R}} e^{tx} f_X(x) dx = \int_a^b e^{tx} \frac{1}{b-a} dx.$$

Se  $t = 0$ ,  $M_X(0) = 1$ . Se  $t \neq 0$ , então

$$M_X(t) = \frac{e^{bt} - e^{at}}{(b-a)t}.$$

□

10. Use a função geradora de momentos para obter a esperança e a variância de  $X$ , nos seguintes casos:

- $X \sim B(n, p)$
- $X \sim \text{Poisson}(\lambda)$

**Solução.**

- Feito no texto.
- Para todo  $t \in \mathbb{R}$ , temos

$$M_X(t) = \sum_{x=0}^{\infty} e^{tx} e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} = e^{-\lambda} e^{\lambda e^t}.$$

Portanto,

$$M'_X(t) = e^{-\lambda} e^{\lambda e^t} \lambda e^t \implies M'_X(0) = EX = \lambda.$$

Ademais,

$$M''_X(t) = \lambda e^{-\lambda} (e^{\lambda e^t} \lambda e^{2t} + e^{\lambda e^t} e^t),$$

donde segue que  $M''_X(0) = \lambda^2 + \lambda = E[X^2]$ . Portanto,  $\text{Var}(X) = \lambda$ . □

11. Use a função geradora de momentos para obter os momentos de todas as ordens de  $X$ , nos seguintes casos:

- $X \sim N(0, \sigma^2)$
- $X$  tem densidade  $f_X(x) = e^{-|x|}$ ,  $x \in \mathbb{R}$ .

**Demonstração.**

- Feito no texto.

b) Temos, para  $t \in (-1, 1)$ ,

$$\begin{aligned} M_X(t) &= \frac{1}{2} \left( \int_0^\infty e^{x(t-1)} dx + \int_{-\infty}^0 e^{x(t+1)} dx \right) \\ &= -\frac{1}{2} \cdot \frac{1}{t-1} + \frac{1}{2} \cdot \frac{1}{t+1} \\ &= \frac{1}{1-t^2} \\ &= \sum_{k=0}^{\infty} t^{2k} \\ &= \sum_{k=0}^{\infty} (2k)! \frac{t^{2k}}{(2k)!}. \end{aligned}$$

Portanto,

$$EX^m = \begin{cases} 0, & m \text{ impar} \\ m!, & m \text{ par} \end{cases}.$$

□

12. Sejam  $X_1, X_2, \dots, X_n$  v.a.'s independentes. Use a função geradora de momentos para obter a distribuição de  $S_n = X_1 + X_2 + \dots + X_n$ , nos seguintes casos:

- $X_i \sim \Gamma(\alpha_i, \lambda)$
- $X_i \sim \text{Exp}(\lambda)$
- $X_i \sim B(n_i, p)$
- $X_i \sim \text{Poisson}(\lambda_i)$ , para  $i = 1, \dots, n$ .

### Solução.

a) Temos

$$M_{S_n}(t) = \prod_{i=1}^n M_{X_i}(t) = \prod_{i=1}^n \left( \frac{\lambda}{\lambda - t} \right)^{\alpha_i} = \left( \frac{\lambda}{\lambda - t} \right)^{\sum_i \alpha_i}, \quad t < \lambda.$$

Portanto,  $S_n \sim \Gamma(\sum_i \alpha_i, \lambda)$ .

b) Temos

$$M_{S_n}(t) = \prod_{i=1}^n M_{X_i}(t) = \prod_{i=1}^n \frac{1}{1 - t/\lambda} = \frac{1}{(1 - t/\lambda)^n} = \left( \frac{\lambda}{\lambda - t} \right)^n, \quad t < \lambda.$$

Portanto,  $S_n \sim \Gamma(n, \lambda)$ .

c) Temos

$$M_{S_n}(t) = \prod_{i=1}^n M_{X_i}(t) = \prod_{i=1}^n (pe^t + 1 - p)^{n_i} = (pe^t + 1 - p)^{\sum_i n_i}, \quad \forall t \in \mathbb{R}.$$

Portanto,  $S_n \sim B(\sum_i n_i, p)$ .

d) Temos

$$M_{S_n}(t) = \prod_{i=1}^n M_{X_i}(t) = \prod_{i=1}^n \exp(\lambda_i(e^t - 1)) = \exp\left[ (e^t - 1) \sum_i \lambda_i \right], \quad \forall t \in \mathbb{R}.$$

Portanto,  $S_n \sim \text{Poisson}(\sum_i \lambda_i)$ .

□

13. Sejam  $X_1, X_2, \dots, X_n$  v.a.'s i.i.d. tendo média  $\mu$  e variância  $\sigma^2$  e seja  $\bar{X} = S_n/n$ , onde  $S_n = X_1 + \dots + X_n$ .

- Mostre que  $EX = \mu$  e  $\text{Var}(\bar{X}) = \sigma^2/n$ .
- Qual o tamanho da amostra que devemos considerar de tal forma que  $P(|\bar{X} - \mu| \leq \sigma/10) \geq 0,95$ ?

### Solução.

a) Feito em exercícios anteriores.

b) Temos que

$$\begin{aligned}P(|\bar{X} - \mu| \leq \sigma/10) &= P(-\sigma/10 \leq \bar{X} - \mu \leq \sigma/10) \\&= P(\bar{X} \leq \mu + \sigma/10) - P(\bar{X} \leq \mu - \sigma/10) \\&= P(S_n \leq n\mu + n\sigma/10) - P(S_n \leq n\mu - n\sigma/10).\end{aligned}$$

Usando o TLC, temos que essa probabilidade pode ser aproximada por

$$\Phi(\sqrt{n}/10) - \Phi(-\sqrt{n}/10) = 2\Phi(\sqrt{n}/10) - 1.$$

Portanto, para que a probabilidade desejada seja pelo menos 0,95, devemos ter  $n$  tal que

$$\Phi(\sqrt{n}/10) \geq 0,975 \Leftrightarrow \frac{\sqrt{n}}{10} \geq 1,96 \Leftrightarrow n \geq 384,16.$$

Como  $n \in \mathbb{N}$ , devemos ter uma amostra de tamanho 385, pelo menos.

□

14. Da experiência passada, um professor sabe que a pontuação de um estudante no seu exame final é uma v.a. com média 75.

- Dê um limite superior para a probabilidade de que a pontuação do estudante excederá 85.
- Se, além disso, o professor também saiba que a variância da pontuação do estudante é 25, o que pode ser dito sobre a probabilidade de que o estudante terá uma pontuação entre 65 e 85?

### Solução.

a) Pela desigualdade de Markov, temos

$$P(X \geq 85) \leq \frac{75}{85} \approx 0,88.$$

b) Usando o TLC, temos

$$\begin{aligned}P(65 \leq X \leq 85) &= P(X \leq 85) - P(X \leq 65) \\&= P\left(\frac{X - 75}{5} \leq 2\right) - P\left(\frac{X - 75}{5} \leq -2\right) \\&\approx \Phi(2) - \Phi(-2) \\&= 2\Phi(2) - 1 \\&= 2 \cdot 0,9772 - 1 \\&= 0,9544.\end{aligned}$$

Portanto, a probabilidade procurada é aproximadamente 0,9544.

□

15. Um corredor procura controlar seus passos em uma corrida de 100 metros. De sua experiência, ele sabe que o tamanho de seu passo na corrida é uma v.a. com média 0,97 metro e desvio-padrão 0,1 metro. Determine a probabilidade de que 100 passos difiram de 100 metros por não mais de 5 metros.

**Solução.** Se  $X$  é a v.a. que modela o tamanho do passo na corrida, queremos estimar  $P(|100X - 100| \leq 5)$ . Temos, usando o TLC, que essa probabilidade é igual a

$$\begin{aligned}P(|X - 1| \leq 0,05) &= P(0,95 \leq X \leq 1,05) \\&= P\left(\frac{X - 0,97}{0,1} \leq 0,8\right) - P\left(\frac{X - 0,97}{0,1} \leq -0,2\right) \\&\approx \Phi(0,8) - \Phi(-0,2) \\&\approx \Phi(0,8) + \Phi(0,2) - 1 \\&= 0,7881 + 0,5793 - 1 \\&= 0,3674.\end{aligned}$$

□

# Referências

- [1] Sheldon Ross, *Probabilidade: um curso moderno com aplicações*, 8<sup>a</sup> ed., 2010.
- [2] Paul Hoel; Sidney Port; Charles Stone, *Introduction to Probability Theory*, 1971.