INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DA BAHIA
CAMPUS SALVADOR

Manuel Aniceto Pereira Neto

A MATEMÁTICA DO AZAR: EXPLORANDO A RUÍNA DO JOGADOR POR MEIO DE CADEIAS DE MARKOV

Salvador-BA
2026

Manuel Aniceto Pereira Neto

A MATEMÁTICA DO AZAR: EXPLORANDO A RUÍNA DO JOGADOR POR MEIO DE CADEIAS DE MARKOV

Trabalho de Conclusão de Curso apresentado como requisito parcial para obtenção do grau de Licenciado em Matemática, pelo Instituto Federal de Educação, Ciência e Tecnologia Bahia – Campus Salvador.

Área de concentração: Probabilidade.

Orientadora: Profa. Ma. Renata de Moura Issa Vianna

Coorientador: Prof. Me. Ruan dos Santos Vieira Miranda

Salvador-BA
2026

Manuel Aniceto Pereira Neto

A MATEMÁTICA DO AZAR: EXPLORANDO A RUÍNA DO JOGADOR POR MEIO DE CADEIAS DE MARKOV

A banca examinadora, abaixo listada, aprova o Trabalho de Conclusão de Curso "A MATEMÁTICA DO AZAR: EXPLORANDO A RUÍNA DO JOGADOR POR MEIO DE CADEIAS DE MARKOV" elaborado por "Manuel Aniceto Pereira Neto" como requisito parcial para obtenção do grau de Licenciado(a) em Matemática, pelo Instituto Federal de Educação, Ciência e Tecnologia da Bahia.

Salvador-BA, 05 / 02 / 2026

Banca Examinadora

Documento assinado digitalmente
gov.br
RENATA DE MOURA ISSA VIANNA
Data: 25/03/2026 10:29:07-0300
Verifique em https://validar.iti.gov.br

Profa. Ma. Renata de Moura Issa Vianna
IFBA
(Orientadora)

Documento assinado digitalmente
gov.br
RUAN DOS SANTOS VIEIRA MIRANDA
Data: 25/03/2026 10:43:08-0300
Verifique em https://validar.iti.gov.br

Prof. Me. Ruan dos Santos Vieira Miranda
UFBA
(Coorientador)

Documento assinado digitalmente
gov.br
FELLIPE ANTONIO DOS SANTOS CARDOSO LEIT
Data: 22/05/2026 15:41:12-0300
Verifique em https://validar.iti.gov.br

Prof. Me. Fellipe Antônio dos Santos Cardoso Leite
IFBA

Documento assinado digitalmente
gov.br
YURE CARNEIRO DE OLIVEIRA
Data: 25/03/2026 17:03:59-0300
Verifique em https://validar.iti.gov.br

Prof. Me. Yure Carneiro de Oliveira
UFBA

FICHA CATALOGRÁFICA ELABORADA PELO SISTEMA DE BIBLIOTECAS DO IFBA, COM OS DADOS FORNECIDOS PELO(A) AUTOR(A)

P968m Pereira Neto, Manuel Aniceto

A matemática do azar: explorando a ruína do jogador por meio da cadeia de Markov / Manuel Aniceto Pereira Neto; orientadora Renata de Moura Issa Vianna; coorientador Ruan dos Santos Vieira Miranda -- Salvador, 2026.

62 p.

Trabalho de Conclusão de Curso (Matemática) -- Instituto Federal da Bahia, 2026.

1. Probabilidade. 2. Processos estocásticos. 3. Cadeias de Markov. I. Vianna, Renata de Moura Issa, orient. II. Miranda, Ruan dos Santos Vieira, coorient. III. TÍTULO.

CDU 519.2

Agradecimentos

Gostaria de agradecer a minha companheira Maria Amália Veloso da Silva, por ter me apoiado tanto em toda essa caminhada e nos momentos de sofrimento nos quais eu acreditava que nunca passaria. Gostaria de agradecer também aos meus professores, especialmente a professora Ma. Renata de Moura Issa Vianna e ao professor Me. Ruan dos Santos Vieira Miranda, por terem me orientado e me dado aulas durante nossos estudos. Gostaria de agradecer também aos professores do IFBA que mais acreditaram em mim e esse agradecimento vai aos professores Me. Fellipe Antônio Cardoso Leite, Dra. Isabel Cristina Costa Leite, Dra. Anete Otilia Cardoso de Santana Cruz, Dra. Tânia Jussara Silva Santana, Dra. Daniela Santa Inês Cunha e a Dra. Jamille Vilas Bôas de Souza por terem me apoiado tanto em momentos muito difíceis. Alguns dos meus colegas de curso também merecem um agradecimento especial por terem me apoiado tanto, gostaria então de agradecer aos colegas Alana Lima Brandão, Rebeca Camille Brito de Oliveira, Henrique Neves dos Santos e Ana Ellen Sena de Oliveira Santos. Gostaria de agradecer também a uma pessoa muito especial, Marligia Nascimento Araújo que tanto nos ajudou durante a caminhada da graduação.

"A matemática, vista de forma correta, não se trata apenas de verdades, mas beleza suprema."

(Betrand Russell)

A MATEMÁTICA DO AZAR: EXPLORANDO A RUÍNA DO JOGADOR POR MEIO DE CADEIAS DE MARKOV

Resumo

Este Trabalho de Conclusão de Curso apresenta uma análise teórica e introdutória sobre as Cadeias de Markov, fundamentando-as dentro do campo dos processos estocásticos. O estudo inicia-se com uma revisão básica da Teoria da Probabilidade, partindo das abordagens clássicas até a consolidação do método axiomático proposto por Kolmogorov. São discutidos conceitos fundamentais de variáveis aleatórias e distribuições de probabilidade, essenciais para a compreensão da dinâmica de sistemas que evoluem com o tempo. O foco central reside na definição e propriedade das Cadeias de Markov em tempo discreto, explorando a característica de "ausência de memória"e as matrizes de transição. Através de exemplos clássicos, como a Ruína do Jogador, demonstra-se a aplicabilidade do modelo em fenômenos aleatórios. Por fim, o trabalho discute brevemente as perspectivas futuras de expansão para o campo da computação quântica por meio dos Passeios Aleatórios Quânticos, ressaltando a relevância da natureza probabilística na ciência moderna.

Palavras-chave: Probabilidade. Processos Estocásticos. Cadeias de Markov.

Abstract

This undergraduate work presents a theoretical and introductory analysis of Markov chains, situating them within the field of stochastic processes. The study begins with a fundamental review of Probability Theory, moving from classical approaches to the consolidation of the axiomatic method proposed by Kolmogorov. Fundamental concepts of random variables and probability distributions are discussed as essentials for understanding the dynamics of systems that evolve over time. The central focus lies on the definition and properties of discrete-time Markov chains, exploring the 'memorylessness' characteristic and transition matrices. Through classical examples, such as the Gambler's Ruin, the model's applicability to random phenomena is demonstrated. Finally, the work briefly discusses future perspectives for expansion into the field of quantum computing via Quantum Random Walks, highlighting the relevance of probabilistic nature in modern science.

Keywords: Probability. Stochastic Processes. Markov Chains.

Lista de figuras

Figura 1 – Espaço amostral Ω particionado 24
Figura 2 – Representação Limsup A_n 28
Figura 3 – Representação Liminf A_n 29
Figura 4 – Representação do problema da moeda binomial 36
Figura 5 – Jogador de basquete 38
Figura 6 – Distribuição de Poisson 41
Figura 7 – Distribuição exponencial 42
Figura 8 – Roleta do cassino 44
Figura 9 – Diagrama das vitórias-régias 54
Figura 10 – Passeio aleatório simples 56
Figura 11 – Passeio aleatório em duas dimensões 57
Figura 12 – Diagrama da ruína do jogador 58
Figura 13 – Diagrama de classes do problema da ruína do jogador 58

Sumário

1 – Introdução 15

2 – Introdução à probabilidade 17

2.1 Probabilidade clássica e probabilidade moderna 17
2.2 Espaço amostral 18
2.3 Espaço mensurável 19
2.4 Probabilidade 20
2.5 Probabilidade condicional 23
2.6 Teorema da probabilidade total 25
2.7 Independência 25
2.8 Sigma-álgebra de Borel e conjuntos borelianos 26
2.9 Lema de Borel-Cantelli 27

3 – Processos estocásticos 31

3.1 Variáveis aleatórias 31
3.2 Função indicadora 32
3.3 Variáveis aleatórias discretas 33
3.4 Variáveis aleatórias contínuas 34
3.5 Variáveis aleatórias independentes 34
3.5.1 Distribuição binomial 34
3.5.2 Distribuição geométrica 36
3.5.3 Distribuição de Poisson 38
3.5.4 Distribuição exponencial 41
3.6 Função de distribuição 42
3.7 Esperança 43
3.8 Propriedades da esperança 46
3.9 Variância 47
3.10 Desigualdades clássicas 50
3.11 Processo estocástico 51
3.12 Espaço de estados 51
3.13 Matriz estocástica 51

4 – Cadeias de Markov 53

4.1 Cadeias de Markov 53
4.2 A ruína do jogador 58

5 – Conclusões e Perspectivas 64

Referências **65**

1 Introdução

Este trabalho traz o problema da ruína do jogador que ao se deparar com estados extremos, sem dinheiro pra apostar, ou quando já tem a quantia limite permitido pela banca, o mesmo não pode mais continuar jogando. O objetivo é apresentar as ferramentas básicas da probabilidade, aprofundar em probabilidade, trazendo os principais aspectos dos processos estocásticos e falar sobre as cadeias de Markov, apresentando exemplos de cadeias de Markov finitas.

Os processos estocásticos são processos que ditam o comportamento da natureza, seja na previsão do tempo ou na variação de preço de ações. Esses fenômenos são incríveis, pois regem o universo a nossa volta, no entanto, podemos estudá-los matematicamente e criar modelos de aplicação. É muito intrigante saber que a vida é regida por processos estocásticos, não importa o planejamento de cada ação, algo imprevisível sempre acontece. É fascinante perceber o quanto a aleatoriedade afeta a vida das pessoas.

Para a elaboração deste trabalho, foram utilizados livros textos considerados clássicos, que são os trabalhos do James (2023) que é um livro de nível intermediário, muito utilizado no mestrado em matemática e Durrett (2010) que é amplamente utilizado nos programas de doutorado pelo mundo. Foi usado também o Rolla e Lima (2025) que é uma referência mais nova e foi elaborado a partir de notas de aula. Um outro clássico que não posso deixar de citar é o trabalho do Levin e Peres (2017), mas na maior parte do tempo foi utilizado o livro do Franco (2020), pois é uma mistura de intermediário com avançado. Um outro material, além dos clássicos que foi explorado, principalmente para os exemplos da ruína do jogador e do sapo foi o relatório de iniciação científica do Cardoso (2020).

O livro do Tertuliano Franco não só apresenta conceitos de probabilidade mais avançados, como também traz uma ótima explicação/tradução desses conceitos e os exemplos são bem elucidativos. Outros livros foram usados também em uma porcentagem menor quando por exemplo na indicação de onde esta uma determinada demonstração, essa referência foi o Lima (2022). Foi utilizado também alguns trabalhos para trazer uma contextualização da probabilidade clássica e uma definição de processo estocástico, ambos são, respectivamente os livros do Carvalho et al. (2020) e do Bertsekas e Tsitsiklis (2008). As aulas no youtube para incrementar o entendimento, também foram bastante uteis para a escrita deste trabalho, mais especificamente as aulas do programa de mestrado do IMPA - Instituto de Matemática Pura e Aplicada (2025) e outros vídeos como o do professor Timpanaro (2021).

Então, baseado nesses autores, neste trabalho será abordado os conceitos de probabilidade moderna como por exemplo a σ-álgebra. Sem essa ferramenta a probabilidade

como conhecemos hoje não existiria, pois é essa ferramenta que garante a mensurabilidade de conjuntos infinitos. Ainda no primeiro capitulo é abordado o conceito de probabilidade condicional, que falando de uma forma mais coloquial, a probabilidade restringe a resposta dado uma condição. No segundo capítulo ressalto a importância das variáveis aleatórias, da esperança e por fim, no terceiro capítulo é discutido a ideia do que é uma cadeia de Markov, o mesmo é relacionado com o processo estocástico e essa ideia é aplicada no problema da ruína do jogador, onde toda essa teoria toma forma, pois é nele que vemos a utilidade da probabilidade condicional, teorema da probabilidade total e etc.

Para finalizar o trabalho, é mostrado as perspectivas futuras de pesquisa em probabilidade, mais especificamente as cadeias de Markov aplicadas a computação quântica já que esse é um tema muito atual. O que justifica o presente trabalho é o fato desse material ter sido elaborado com o intuito de traduzir os conceitos de probabilidade para uma linguagem mais acessível, sem perder o rigor matemático, pois a linguagem que os livros clássicos como do James (2023) é uma linguagem que os estudantes sentem dificuldade. Pensando nesse fato, este trabalho foi escrito utilizando uma linguagem mais próxima do trabalho do Franco (2020), que embora traga conceitos mais avançados de probabilidade, ele faz uma ótima explicação desses conceitos e utiliza muitos exemplos.

2 Introdução à probabilidade

A probabilidade clássica é conhecida como Probabilidade de Laplace, na qual a probabilidade de um evento A ocorrer é definida pela razão entre o número de casos favoráveis e o número de casos possíveis. Segundo Carvalho et al. (2020), tal definição é denotada por:

$$\mathbb{P}(A) = \frac{\text{Número de casos favoráveis}}{\text{Número de casos possíveis}}.$$

Conforme a probabilidade foi tomando uma característica mais axiomática, graças a Andrei Nikolaevich Kolmogorov, um dos pioneiros na formulação da probabilidade como a conhecemos hoje, é possível nos dias atuais mensurar coisas que só com a definição de probabilidade clássica era impossível, como por exemplo calcular a probabilidade de sair um determinado número em um dado viciado.

Este primeiro capítulo trata-se de uma introdução à probabilidade moderna, trazendo as ferramentas para o desenvolvimento do seu conceito. São introduzidas as ideias de espaço amostral, sigma-álgebra, espaço de medida, espaço de probabilidade e outras ferramentas essenciais para apresentar as cadeias de Markov.

2.1 Probabilidade clássica e probabilidade moderna

Considere um experimento aleatório. Chamamos de espaço amostral o conjunto de todos os resultados possíveis deste experimento. A definição clássica de probabilidade fundamenta-se em espaços amostrais equiprováveis, nos quais todos os elementos do conjunto de possibilidades possuem a mesma chance de ocorrência, ou seja, a probabilidade clássica define sistemas não viciados, conforme aponta Franco (2020).

Definição 1 (Probabilidade clássica)**.** A probabilidade de um subconjunto qualquer $A \subset \Omega$, denominado evento, é dada por:

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

onde:

i. $n(\Omega)$ é o número de casos possíveis ou a cardinalidade do espaço amostral;

ii. $n(A)$ é o número de casos favoráveis ou a cardinalidade do conjunto A.

Precisamos, agora, estabelecer uma definição de probabilidade para incorporar os sistemas viciados, fazendo uma abordagem axiomática moderna, a ser detalhada adiante.

Porém, antes disso, torna-se necessário definir objetos matemáticos que viabilizem a formalização dessa teoria.

2.2 Espaço amostral

Definição 2 (Espaço Amostral)**.** O espaço amostral é conjunto que denota todos os casos possíveis de um experimento aleatório, e sua notação é dada por Ω.

Exemplo 2.1. Suponha que desejamos listar todos os casos possíveis de um lançamento de um dado. Esse conjunto será descrito como $\{1, 2, 3, 4, 5, 6\}$. A rigor matemático denotamos um espaço amostral da seguinte forma:

$$\begin{aligned}\Omega &= \{(a_1, ..., a_n) \mid a_i \in \{1, 2, 3, 4, 5, 6\}\} \\ &= \{1, 2, 3, 4, 5, 6\} \cdot ... \cdot \{1, 2, 3, 4, 5, 6\} \ n \text{ vezes} \\ &= \{1, 2, 3, 4, 5, 6\}^n.\end{aligned}$$

Exemplo 2.2. No lançamento de uma moeda, o espaço amostral é $\Omega = \{C, K\}$, C para representar cara e K para representar coroa.

Exemplo 2.3. No lançamento de 2 dados, o espaço amostral na prática pode ser escrito como $\{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), ..., (6,6)\}$

Como escrever espaços amostrais muito grandes é impraticável da forma anterior, podemos escrever matematicamente como:

$$\Omega = \{(x, y) \mid x \in \{1, 2, 3, 4, 5, 6\} \text{ e } y \in \{1, 2, 3, 4, 5, 6\}\}.$$

Definição 3 (σ-álgebra)**.** Dizemos que uma coleção de subconjuntos do espaço amostral é uma σ-álgebra, denotado por $\mathcal{F}$, se satisfizer as seguintes propriedades:

i. $\Omega \in \mathcal{F}$;

ii. Se A $\in \mathcal{F}$, então $A^c \in \mathcal{F}$;

iii. Se $A_1, A_2, A_3, ... \in \mathcal{F}$, então $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$. (Não necessariamente a união disjunta).

Exemplo 2.4 (σ-Álgebra constituída pelo conjunto das partes)**.** Seja $\Omega = \{C, K\}$ o espaço amostral. Uma σ-Álgebra desse espaço amostral é:

$$\mathcal{P}(\Omega) = \{\emptyset, \{C\}, \{K\}, \{C, K\}\}.$$

i. $\Omega \in \mathcal{F}$;

ii. O complementar dos elementos estão em $\mathcal{F}$, pois $\{C\}^c = \{K\}, \{K\}^c = \{C\}$, $\{C, K\}^c = \emptyset$ e $\emptyset^c = \{C, K\}$;

iii. Veja que $\{C\} \cup \{K\} \cup \{C, K\} \cup \emptyset$ também está na σ-Álgebra.

Vale a pena observar que uma σ-Álgebra não se restringe só ao conjunto das partes, veja o exemplo a seguir.

Exemplo 2.5 (σ-Álgebra que não é o conjunto das partes)**.** Vamos analisar o lançamento de um dado, o seu espaço amostral é $\Omega = \{1, 2, 3, 4, 5, 6\}$. Uma σ-álgebra desse espaço amostral pode ser escrita da seguinte forma:

$$\mathcal{F} = \{\emptyset, \{1\}, \{2, 3, 4, 5, 6\}, \Omega\}.$$

Essa é uma σ-álgebra, pois

i. $\Omega \in \mathcal{F}$;

ii. Seja $A = \{1\} \in \mathcal{F}$. O complementar $A^c = \{2, 3, 4, 5, 6\}$ também está na σ-álgebra. Da mesma forma, como Ω pertence a σ-álgebra, o seu complementar $\Omega^c = \emptyset$ também pertence à σ-álgebra;

iii. Sejam $A_1 = \{1\}, A_2 = \{2, 3, 4, 5, 6\}$. Então, $\bigcup_{i=1}^{2} A_i = \{1\} \cup \{2, 3, 4, 5, 6\} = \{1, 2, 3, 4, 5, 6\}$. E assim para qualquer união de elementos de $\mathcal{F}$.

Logo, $\mathcal{F}$ é uma σ-Álgebra.

Exemplo 2.6 (σ-álgebra trivial)**.** A σ-Álgebra trivial é formada pelo espaço amostral e pelo conjunto vazio, denotado como $\{\emptyset, \Omega\}$.

É possível, então, observar que a σ-álgebra pode ser diferente do conjunto das partes. Construir uma σ-álgebra do tipo do exemplo 2.5, na qual não faz sentido calcular a probabilidade de sair 2, é dizer que a σ-álgebra não enxerga esse elemento do conjunto Ω. Portanto, não faz sentido calcular a probabilidade de 2, pois esta σ-álgebra só pode medir a probabilidade de sair 1 ou não 1.

2.3 Espaço mensurável

Definição 4 (Espaço mensurável)**.** Seja $(\Omega, \mathcal{F})$ um par, onde Ω é o espaço amostral e $\mathcal{F}$ é uma σ-álgebra formada por subconjuntos de Ω. Essa dupla é chamada de espaço mensurável.

Agora que temos as ferramentas necessárias, podemos enunciar de forma mais precisa a probabilidade.

2.4 Probabilidade

Definição 5 (Probabilidade)**.** Seja a dupla $(\Omega, \mathcal{F})$ um espaço mensurável, uma probabilidade, denotada por $\mathbb{P}$, nesse espaço é dada por uma função $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ que satisfaz as seguintes propriedades:

i. $\mathbb{P}(A) \geq 0, \forall A \in \mathcal{F}$;

ii. (Sigma-aditividade) Se $A_1, A_2, ... \in \mathcal{F}$ são disjuntos dois a dois, então:

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

iii. $\mathbb{P}(\Omega) = 1$.

Então, a uma terna $(\Omega, \mathcal{F}, \mathbb{P})$ chamamos de um espaço de probabilidade, sendo as propriedades apresentadas acima chamadas de axiomas de Kolmogorov.

Proposição 2.4.1 (Propriedades de um espaço de probabilidade)**.** Seja a tripla $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade e sejam $A, B, A_1, A_2, ... \in \mathcal{F}$, elementos da σ-álgebra. Então valem as seguintes propriedades:

i. $\mathbb{P}(\emptyset) = 0$;

ii. (Aditividade finita) Se $A_1, A_2, ..., A_n \in \mathcal{F}$ são disjuntos dois a dois, então:

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

iii. Se $A \subset B$, então $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \backslash A)$;

iv. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$;

v. $\mathbb{P}(A) \leq 1, \forall A \in \mathcal{F}$;

vi. Se $A \subset B$, então, $\mathbb{P}(A) \leq \mathbb{P}(B)$;

vii. Se $A_n \nearrow A$, então $\mathbb{P}(A_n) \nearrow \mathbb{P}(A)$;

viii. Se $A_n \searrow A$, então $\mathbb{P}(A_n) \searrow \mathbb{P}(A)$.

Demonstração. Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade, então:

i. Note que o conjunto vazio pode ser escrito como a união enumerável de si mesmo:

$$\emptyset = \emptyset \cup ... \cup \emptyset \cup ... = \bigcup_{n=1}^{\infty} \emptyset.$$

Como é uma união disjunta, pela definição de probabilidade, propriedade *ii*. Definição 5:

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

Pela propriedade *iii* da definição 5, temos que $\mathbb{P}(\Omega) = 1$, o que implica que $\mathbb{P}(\emptyset)$ deve ser um número real finito, especificamente, $\mathbb{P}(\emptyset) \leq 1$. Se tivéssemos $\mathbb{P}(\emptyset) > 0$, a soma infinita divergiria para o infinito, o que contradiz o fato de $\mathbb{P}(\emptyset)$ ser um valor finito. Portanto, a única solução possível é $\mathbb{P}(\emptyset) = 0$.

ii. Se $A_1, A_2, ..., A_n, ... \in \mathcal{F}$ e $A_{n+k} = \emptyset, \forall k \geq 1$ temos

$$\bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{+\infty} A_i.$$

Logo:

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

mas

$$\sum_{i=n+1}^{+\infty} \mathbb{P}\left(A_i\right) = \sum_{i=n+1}^{+\infty} \mathbb{P}(\emptyset) = \sum_{i=n+1}^{+\infty} 0 = 0.$$

Portanto,

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

iii. Sejam $A, B \in \mathcal{F}$ com $A \subset B$, note que $A \cap (B \backslash A) = \emptyset$. Além disso, observe que podemos escrever:

$$B = A \cap (B \backslash A).$$

Como $\mathcal{F}$ é uma σ-álgebra, temos que $A^c \in \mathcal{F}$. Logo, $(B \backslash A) = B \cap A^c \in \mathcal{F}$, pois $B, A^c \in \mathcal{F}$. Portanto, A e $B \backslash A$ são eventos disjuntos em $\mathcal{F}$. Pela aditividade:

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

Assim, $\mathbb{P}(A) + \mathbb{P}(B \backslash A) = \mathbb{P}(B)$.

iv. Para demonstrar essa propriedade, basta fazer $B = \Omega$ e pela propriedade *iii* da proposição 2.4.1 temos que $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B - A)$. Segue que $\mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(\Omega - A)$. Como $\Omega \backslash A = A^c$ e $\mathbb{P}(\Omega) = 1$, temos que $1 = \mathbb{P}(A) + \mathbb{P}(A^c)$. Logo, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

v. Pelo axioma de Kolmogorov *i*, Definição 2.5, $\mathbb{P}(A) \geq 0, \forall A \in \Omega$. Além disso, pela propriedade antrior *iv*, $\mathbb{P}(A) + \mathbb{P}(A^c) = 1$. Assim, como $\mathbb{P}(A^c) \geq 0$, temos que $\mathbb{P}(A) \leq 1, \forall A \in \mathcal{F}$.

vi. Sejam $A, B \in \mathcal{F}$, $A \subset B$, então $B = A \cup (B \backslash A)$ e $A \cap (B \backslash A) = \emptyset$. Assim, $\mathbb{P}(B) = \mathbb{P}(A \cup (B \backslash A))$. Pela propriedade *ii*, (aditividade finita), temos que

$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \backslash A).$$

Pelo axioma de Kolmogorov *i*, $\mathbb{P}(B \backslash A) \geq 0$. Logo, $\mathbb{P}(B) \geq \mathbb{P}(A)$.

vii. Suponha que A_n cresce para A. Denotado como $A_n \nearrow A$, isto é, $A_n \subset A_{n+1}$ e $A = \bigcup_{n=1}^{+\infty} A_n$. Defina $B_1 = A_1$ e $B_n = B_n \backslash A_{n-1}, n \geq 2$. Então, os B_n são disjuntos, $A = \bigcup_{n=1}^{+\infty} B_n$ e $A_n = \bigcup_{k=1}^{n} B_k$. Aplicando o axioma de Kolmogorov e a propriedade *ii*, temos que

$$\mathbb{P}(A_n) = \sum_{k=1}^{n} \mathbb{P}(B_k)$$

e

$$\mathbb{P}(A) = \sum_{k=1}^{+\infty} \mathbb{P}(B_k).$$

Como a soma parcial cresce para a soma total,

$$\lim_{n \to \infty} \mathbb{P}(A_n) = \sum_{k=1}^{\infty} \mathbb{P}(B_k) = \mathbb{P}(A).$$

Portanto, dizemos que a probabilidade de A_n cresce para a probabilidade de A. Denotado como $\mathbb{P}(A_n) \nearrow \mathbb{P}(A)$.

viii. Suponha que A_n decresce para A. Denotado como $A_n \searrow A$, isto é, $A_{n+1} \subset A_n$ e $A = \bigcap_{n=1}^{+\infty} A_n$. Passando para os complementares, temos que:

$$A_n^c \subset A_{n+1}^c.$$

Logo, A_n^c é crescente e $\bigcup_{n=1}^{+\infty} A_n^c = A^c$. Pela propriedade anterior, $\mathbb{P}(A_n^c) \nearrow \mathbb{P}(A^c)$. Usando a propriedade *iv*, $\mathbb{P}(A_n^c) \nearrow \mathbb{P}(A^c)$, segue que

$$\lim_{n \to \infty} \mathbb{P}(A_n) = 1 - \lim_{n \to \infty} \mathbb{P}(A_n^c) = 1 - \mathbb{P}(A^c) = \mathbb{P}(A).$$

Portanto, a probabilidade de A_n decresce para a probabilidade de A. Denotado como $\mathbb{P}(A_n) \searrow \mathbb{P}(A)$.

□

Proposição 2.4.2 (Subaditividade)**.** Seja a tripla $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade. Se $A_1, A_2, ... \in \mathcal{F}$ são eventos quaisquer, não necessariamente disjuntos, então

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

A demonstração será omitida aqui, mas pode ser encontrada no livro do Franco (2020), página 202.

2.5 Probabilidade condicional

A probabilidade condicional é de suma importância para um bom entendimento sobre as cadeias de Markov. Usamos muito a probabilidade condicional de forma empírica. Por exemplo, quando tentamos adivinhar se irá chover ou não, dependendo de como esteja o céu, límpido ou carregado de nuvem, nós conseguimos restringir nossa opção de resposta, que muda dada uma condição.

Definição 6. Sejam a terna $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade e os eventos $A, B \in \mathcal{F}$ com $\mathbb{P}(B) \neq 0$. A probabilidade condicional de A dado B é definida como:

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

Uma observação muito interessante é que a probabilidade condicional restringe o espaço amostral a condição dada. No caso do exemplo onde tentamos adivinhar se irá chover ou não, note que restringimos o espaço amostral depois de saber se o céu está carregado de nuvens ou não.

Exemplo 2.7. Suponha um lançamento de um dado. Qual a probabilidade de sair o número 2, dado que o resultado obtido é um número par?

Solução: O espaço amostral é $\Omega = \{1, 2, 3, 4, 5, 6\}$. Os números pares são $B = \{2, 4, 6\}$ e o evento $A = \{2\}$. Então:

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

onde

$$\begin{aligned} \mathbb{P}(A \cap B) &= \mathbb{P}(\{2\} \cap \{2, 4, 6\}) \\ &= \mathbb{P}(\{2\}) \\ &= \frac{n(A \cap B)}{n(\Omega)} = \frac{1}{6}. \end{aligned}$$

Além disso,

$$\mathbb{P}(B) = \frac{n(B)}{n(\Omega)} = \frac{3}{6} = \frac{1}{2}.$$

Logo,

$$\mathbb{P}(A \mid B) = \frac{1}{6} \cdot \frac{3}{1} = \frac{1}{3}.$$

Para demonstrar que a probabilidade condicional é uma probabilidade, basta usar os axiomas de Kolmogorov. Isso pode ser verificado no livro do Franco (2020).

Teorema 2.1 (Teorema do produto)**.** Sejam $A, B \in \mathcal{F}$ eventos tais que $\mathbb{P}(B) > 0$. Então,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B) \cdot \mathbb{P}(B).$$

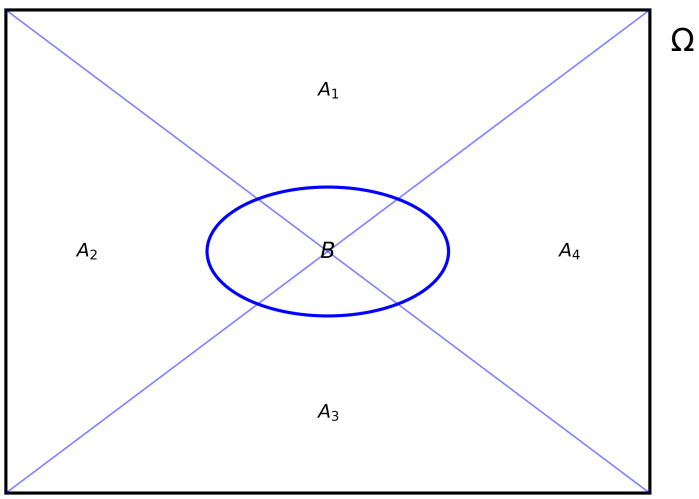
Mais geralmente, se $A_1, ..., A_n \in \mathcal{F}$ com $\mathbb{P}(A_1 \cap ... \cap A_k) > 0, \forall k \geq 2$, então:

$$\mathbb{P}\left(\bigcap_{k=1}^{n} A_k\right) = \mathbb{P}(A_1) \cdot \prod_{k=2}^{n} \mathbb{P}\left(A_k \mid \bigcap_{j=1}^{k-1} A_j\right).$$

A demonstração pode ser encontrada em Durrett (2010), sendo consequência direta da definição de probabilidade condicional.

Seja agora, Ω um espaço amostral particionado pelos eventos A_1, A_2, A_3 e $A_4 \in \mathcal{F}$, todos mutuamente exclusivos, ou seja, a interseção entre eles é vazia, como mostra na figura a seguir.

Figura 1 – Espaço amostral Ω particionado



Fonte: Próprio autor (2025).

Consequentemente, essa partição induz outra em B. Então, temos as interseções $(A_1 \cap B), (A_2 \cap B), (A_3 \cap B)$ e $(A_4 \cap B)$ que também são mutuamente exclusivos, conforme a figura. Daí, podemos dizer que $B = (A_1 \cap B) \cup (A_2 \cap B) \cup (A_3 \cap B) \cup (A_4 \cap B)$. Então, pelo item *ii* da definição 5, podemos escrever:

$$\mathbb{P}(B) = \mathbb{P}(A_1 \cap B) + \mathbb{P}(A_2 \cap B) + \mathbb{P}(A_3 \cap B) + \mathbb{P}(A_4 \cap B).$$

É possível aplicar agora o teorema do produto:

$$\mathbb{P}(B) = \mathbb{P}(B \mid A_1) \cdot \mathbb{P}(A_1) + \mathbb{P}(B \mid A_2) \cdot \mathbb{P}(A_2) + \mathbb{P}(B \mid A_3) \cdot \mathbb{P}(A_3) + \mathbb{P}(B \mid A_4) \cdot \mathbb{P}(A_4).$$

Essa ideia pode ser generalizada para uma partição enumerável do espaço amostral, na forma de um teorema, como apresentaremos a seguir.

2.6 Teorema da probabilidade total

Teorema 2.2 (Teorema da probabilidade total)**.** Se a sequência (finita ou enumerável) de eventos aleatórios $A_1, A_2, ...$ formar uma partição de Ω, então:

$$\mathbb{P}(B) = \sum_i \mathbb{P}(B|A_i) \cdot \mathbb{P}(A_i).$$

E essa é a fórmula do teorema da probabilidade total. A demonstração pode ser encontrada em Ross (2019), sendo consequência direta da definição de probabilidade condicional.

Teorema 2.3 (Teorema de Bayes)**.** Sejam $B_1, B_2, ... \in \mathcal{F}$ eventos, todos de probabilidade positiva, que particionem o espaço amostral Ω, ou seja, são disjuntos e sua união é igual a Ω. Dado um evento A, suponha que $\mathbb{P}(A) > 0$. Então, para qualquer índice $i \in \mathbb{N}$:

$$\mathbb{P}(B_i \mid A) = \frac{\mathbb{P}(A \mid B_i) \cdot \mathbb{P}(B_i)}{\sum_{k=1}^{\infty} \mathbb{P}(A \mid B_k) \cdot \mathbb{P}(B_k)}.$$

A demonstração pode ser encontrada em Franco (2020).

2.7 Independência

O conceito de independência também é bem intuitivo, pois dois eventos são independentes quando o conhecimento da ocorrência de um dos eventos não altera a probabilidade de ocorrência do outro.

Definição 7. Seja a terna $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade, os eventos $A, B \in \mathcal{F}$ são independentes se:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

Exemplo 2.8. Em um cassino, em Las Vegas, num jogo de dados, um dado honesto foi lançado 4 vezes. No primeiro lançamento saiu 1, no segundo saiu 4, no terceiro saiu 5 e no quarto saiu 6. Qual a probabilidade de todos os eventos ocorrerem?

Solução: Chamaremos o primeiro lançamento de A, o segundo de B, o terceiro de C e o quarto de D. Como cada lançamento é feito um após o outro, podemos afirmar que o resultado de um lançamento não afeta a probabilidade de ocorrência dos outros. Dado que os dados são honestos, podemos calcular a probabilidade de cada lançamento.

$$\mathbb{P}(A) = \frac{1}{6}, \mathbb{P}(B) = \frac{1}{6}, \mathbb{P}(C) = \frac{1}{6}, \mathbb{P}(D) = \frac{1}{6}.$$

Podemos agora calcular a probabilidade de todos eles ocorrerem, ou seja, iremos calcular a probabilidade da intersecção $(A \cap B \cap C \cap D)$. Como esses eventos são independentes, vale a definição 7:

$$\begin{aligned} \mathbb{P}(A \cap B \cap C \cap D) &= \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C) \cdot \mathbb{P}(D) \\ &= \left(\frac{1}{6}\right) \cdot \left(\frac{1}{6}\right) \cdot \left(\frac{1}{6}\right) \cdot \left(\frac{1}{6}\right) \\ &= \frac{1}{1296} \\ &\approx 0,000772 \\ &\approx 0,0772\%. \end{aligned}$$

2.8 Sigma-álgebra de Borel e conjuntos borelianos

Muito do que fazemos em probabilidade tem a ver com a σ-álgebra de Borel ou σ-álgebra gerada por intervalos abertos. Por exemplo, quando falamos de variáveis aleatórias contínuas (que serão apresentadas adiante). Esta classe de conjuntos borelianos será a base para o desenvolvimento de nosso trabalho.

Teorema 2.4 (σ-álgebra gerada por uma classe de subconjuntos)**.** Seja Ω um espaço amostral e ε uma classe de subconjuntos de Ω. Então, existe uma única σ-álgebra, chamada σ-álgebra gerada por ε, denotada por $\sigma(\varepsilon)$, que satisfaz às seguintes propriedades:

i. $\sigma(\varepsilon)$ é a menor σ-álgebra que contém a classe ε;

ii. $\sigma(\varepsilon) \supseteq \varepsilon$;

iii. Se $\mathcal{F} \supseteq \varepsilon$ é uma σ-álgebra, então $\mathcal{F} \supseteq \sigma(\varepsilon)$.

Definição 8 (σ-álgebra de Borel)**.** Seja $\Omega = \mathbb{R}$ e considere $\varepsilon = \{(a, b) \subseteq \mathbb{R} : -\infty < a < b < \infty\}$ uma familia de subconjuntos, ou classe de subconjuntos. Dizemos que $\sigma(\varepsilon)$ é a menor σ-álgebra que contém ε, ou seja, $\sigma(\varepsilon)$ é a σ-álgebra que contém os intervalos abertos de $\mathbb{R}$ e essa é a σ-álgebra de Borel. Denotamos ela por $\mathcal{B}(\mathbb{R})$.

Exemplo 2.9. Seja $\Omega = \mathbb{R}$ e seja a $\{\varepsilon_1\} = \{\emptyset\}$, a σ-álgebra que contém $\{\varepsilon_1\}$ é $\sigma(\varepsilon_1) = \{\emptyset, \Omega\} = \{\emptyset, \mathbb{R}\}$. Note que é preciso adicionar elementos para que $\sigma(c_1)$ seja a menor σ-álgebra que contém $\{\varepsilon_1\}$.

Exemplo 2.10. Seja $\Omega = \mathbb{R}$ e seja a $\{\varepsilon_2\} = \{A\}$, então $\sigma(\varepsilon_2) = \{\emptyset, \mathbb{R}, A, A^c\}$. Note que mais uma vez, foi preciso adicionar elementos para que $\sigma(\varepsilon_2)$ seja uma σ-álgebra.

Definição 9 (Conjuntos borelianos na reta)**.** Para o espaço $\Omega = \mathbb{R}$, a σ-álgebra de Borel $\mathcal{B}(\mathbb{R})$ é a σ-álgebra gerada por conjuntos abertos em $\mathbb{R}$. Quando estiver claro no contexto que $\Omega = \mathbb{R}$, podemos escrever $\mathcal{B}$ ao invés de $\mathcal{B}(\mathbb{R})$. Os conjuntos em $\mathcal{B}$ são chamados borelianos.

Exemplo 2.11. Um conjunto pontual $\{x\}$ pode ser obtido como interseção de intervalos abertos $\left(x - \frac{1}{n}, x + \frac{1}{n}\right)$.

É possível observar que é muito difícil encontrar conjuntos que não são borelianos, pois fazendo interseções desse intervalo aberto, o conjunto vai diminuindo ate chegar muito próximo de x. Isso pode ser extrapolado para a maioria dos conjuntos em $\mathbb{R}$.

2.9 Lema de Borel-Cantelli

De acordo com Franco (2020), o lema de Borel-Cantelli, um recurso essencial na área de probabilidade, é empregado para determinar a possibilidade de um evento ocorrer de forma infinita.

Definição 10 (Limite superior e limite inferior)**.** Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade e sejam $A_1, A_2, ..., A_n, ...$ uma sequência de eventos, o limite superior e o limite inferior de A_n é definido por:

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{+\infty} \bigcup_{k=n}^{+\infty} A_k.$$

O limite inferior é dado por:

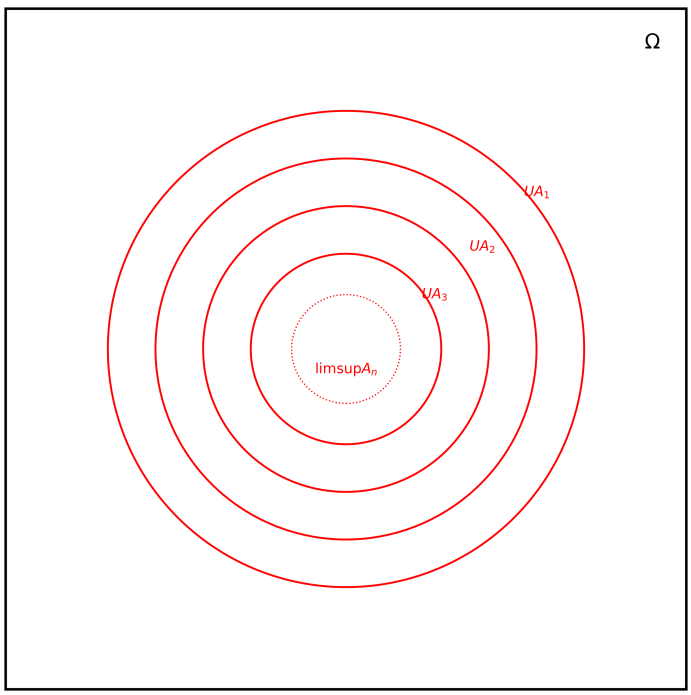
$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{+\infty} \bigcap_{k=n}^{+\infty} A_k.$$

A interpretação para o $\limsup A_n$ é o evento A_n ocorrer infinitas vezes, mesmo que de forma descontínua. Por exemplo, no lançamento infinito de uma moeda, podemos concordar que não vai sair só cara, mas podemos concordar que sairá cara infinitas vezes. A notação frequentemente utilizada é $\limsup A_n = [A_n \text{ Infinitas vezes}]$ ou só $\limsup A_n = [A_n \text{ i.v}]$.

O $\liminf_{n\to\infty} A_n$ é uma condição muito mais forte. Ela indica que existe um ponto crítico na sequência a partir do qual o evento passa a ocorrer para todo k subsequente. Em termos probabilísticos, dizemos que o evento ocorre para quase todo n suficientemente grande, frequentemente denotado segundo James (2023) como $\liminf A_n$ =[Ocorrência de $A_n, \forall n$ suficientemente grande].

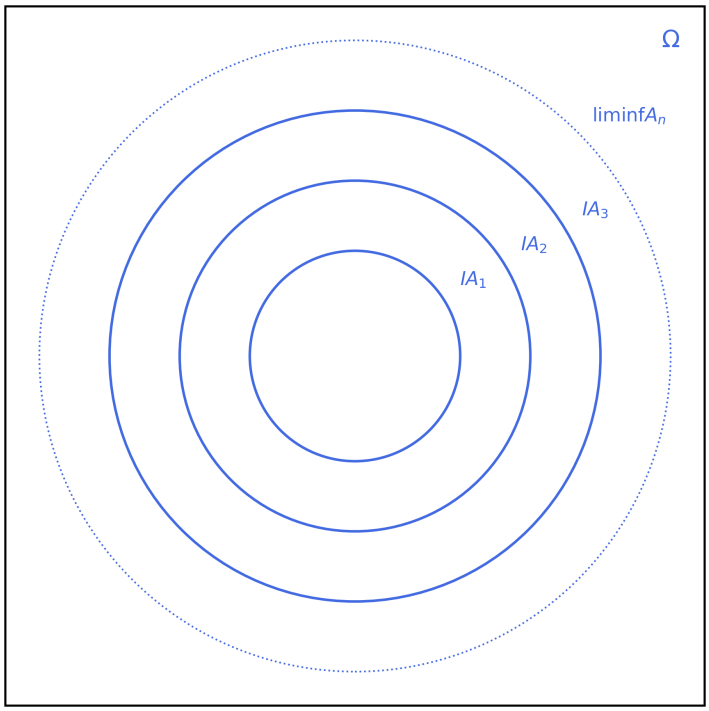
Usando a mesma ideia da moeda, em um lançamento infinito, o $\liminf A_n$ fala que a partir do momento em que sair cara, então só irá sair cara no resto dos lançamentos. Dessa forma, é possível perceber que o $\liminf A_n$ é mais raro de ocorrer do que o $\limsup A_n$. Então, é possível observar que $\liminf \subset \limsup$, pois um evento acontecer infinitas vezes, mesmo que de forma descontínua, é menos raro do que a partir de um ponto só ocorrer esse evento para um número muito grande de vezes. Veja a representação do $\liminf A_n$ e o $\limsup A_n$ abaixo.

Figura 2 – Representação Limsup A_n



Fonte: Próprio autor (2025).

Figura 3 – Representação Liminf A_n



Fonte: Próprio autor (2025).

Teorema 2.5 (Lema de Borel–Cantelli)**.** Seja $(A_n)_{n\geq 1}$ uma sequência de eventos em $(\Omega, \mathcal{F}, \mathbb{P})$.

i. Se $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, então:

$$\mathbb{P}(\limsup A_n) = 0.$$

ii. Se os eventos $(A_n)_{n\geq 1}$ são independentes e $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, então:

$$\mathbb{P}(\limsup A_n) = 1.$$

Corolário 2.6. Considere uma sequência de ensaios de Bernoulli independentes, em que $X_n = 1$ indica sucesso no n-ésimo ensaio e $X_n = 0$ indica fracasso, com $\mathbb{P}(X_n = 1) = p_n$.

i. Se $\sum_{n=1}^{\infty} p_n = \infty$, então

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

ii. Se $\sum_{n=1}^{\infty} p_n < \infty$, então

$$\mathbb{P}\left(\sum_{n=1}^{\infty} X_n < \infty\right) = 1.$$

Traduzindo em palavras, esses dois itens querem dizer que:

i. $\sum P_n = +\infty \Rightarrow$ Um número infinito de sucessos, quase certamente;

ii. $\sum P_n < +\infty \Rightarrow$ Um número finito de sucessos, quase certamente.

A demonstração desse teorema e do corolário serão omitidas aqui, mas é possível encontrá-las no livro do Franco (2020) e do James (2023). O James (2023) traz um exemplo clássico e muito interessante que veremos a seguir.

Exemplo 2.12 (O macaco infinito)**.** Considere o seguinte experimento clássico. Um macaco é colocado diante de uma máquina de escrever contendo um conjunto finito de caracteres (letras, espaços e pontuação). Em cada ensaio, o macaco pressiona teclas de forma completamente aleatória e independente, produzindo uma sequência finita de caracteres.

Fixemos um texto específico, por exemplo, uma obra completa de Shakespeare, e definamos um ensaio como *sucesso* se o macaco reproduzir exatamente esse texto, sem qualquer erro. Caso contrário, o ensaio é considerado um *fracasso*. Após cada ensaio, o processo é reiniciado de forma independente, garantindo que os ensaios sucessivos sejam independentes e identicamente distribuídos.

Denote por A_n o evento "o macaco reproduz corretamente o texto no n-ésimo ensaio". Como o alfabeto é finito e o texto possui comprimento finito, a probabilidade de sucesso em cada ensaio é estritamente positiva, embora extremamente pequena. Assim,

$$\mathbb{P}(A_n) = p > 0, \quad \forall n \in \mathbb{N}.$$

Consequentemente, a série:

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \sum_{n=1}^{\infty} p = +\infty$$

diverge. Além disso, como os eventos $(A_n)_{n\geq 1}$ são independentes, as hipóteses da segunda parte do Lema de Borel–Cantelli são satisfeitas. Logo,

$$\mathbb{P}(\limsup_{n\to\infty} A_n) = 1.$$

Em termos probabilísticos, isso significa que, com probabilidade um, o macaco reproduzirá corretamente o texto infinitas vezes ao longo de uma sequência infinita de ensaios.

3 Processos estocásticos

Neste capítulo, vamos entender o que são variáveis aleatórias discretas e contínuas. A princípio, uma variável aleatória é uma função que associa a cada resultado possível de um experimento aleatório um número real. Iremos entender melhor esses conceitos a partir da definição formal de variáveis aleatórias. Veremos também a definição formal de função de distribuição, função indicadora, desigualdades clássicas, processos estocásticos etc. Essas são ferramentas básicas para um bom entendimento sobre cadeias de Markov.

3.1 Variáveis aleatórias

Uma variável aleatória, conforme definida por Franco (2020), é um valor numérico, real, atribuído a cada resultado possível dentro de um experimento aleatório.

Definição 11 (Variáveis aleatórias)**.** Uma variável aleatória é uma função $X : \Omega \to \mathbb{R}$, que vai do espaço amostral até o conjunto dos números reais, ou de uma forma mais precisa, seja a terna $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade, dizemos que a função $X : \Omega \to \mathbb{R}$ é uma variável aleatória se, para todo intervalo $I \subset \mathbb{R}$, vale:

$$X^{-1}(I) = \{w \in \Omega : X(w) \in I\} \in \mathcal{F}.$$

O evento associado à variável aleatória pode ser denotado por $\{X \in I\}$, que é uma forma mais simples de notação. Outra observação é que podemos nos referir a uma variável aleatória apenas por $v.a.$

Alguns autores usam a pré-imagem, denotado por $X^{-1}(B) = \{w \in \Omega : X(w) \in B\} \in \mathcal{F}$, para todo conjunto boreliano $B \subset \mathbb{R}$, para falar de variáveis aleatórias, ou usam a notação $\{X \leq x\}$ para pré-imagens de intervalos do tipo $(-\infty, x]$.

Vale a pena observar também que existem alguns tipos de variáveis aleatórias e podemos separá-las em categorias. A primeira categoria é a que podemos classificar as variáveis aleatórias em qualitativas e quantitativas. Os nomes já sugerem que uma variável quantitativa está relacionada a uma quantidade e a qualitativa está relacionada a uma qualidade ou característica. Essa classificação é mais comum em Estatística, mas ajuda na intuição inicial.

3.2 Função indicadora

A função indicadora nos permite observar se determinados conjuntos são mensuráveis, e não só isso, este conceito é muito utilizado para quantificar coisas que normalmente são características, como contar quantas caras saíram em um jogo de cara ou coroa. A função indicadora também pode ser usada como ferramenta para demonstrar desigualdades importantes, como a desigualdade de Tchebyshev item 3.10.1.

Definição 12. Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade. Dado $A \subset \Omega$, a função indicadora, denotado por $\mathbb{1}_A(x)$, é definida como:

$$\mathbb{1}_A(x) = \begin{cases} 1, & \text{se } x \in A \\ 0, & \text{se } x \notin A. \end{cases}$$

A função indicadora permite verificar se determinados subconjuntos do espaço amostral são mensuráveis. Isso ocorre porque a função indicadora associada a um conjunto mensurável define uma variável aleatória. Assim, a função indicadora é uma ferramenta importante na identificação e construção de variáveis aleatórias. Veja o exemplo a seguir.

Exemplo 3.1. Seja o espaço amostral $\Omega = \{1, 2, 3, 4\}$ e $\mathcal{F} = \{\emptyset, \Omega, \{1, 2\}, \{3, 4\}\}$, com $A = \{1, 2\}$ e $B = \{1, 3\}$. Verifique se $\mathbb{1}_B$ é uma variável aleatória.

Já temos experiencia suficiente para de forma visual, perceber que o conjunto B não pertence a σ-álgebra, pois a σ-álgebra não enxerga esse conjunto. Entretanto, podemos verificar isso usando a função indicadora. Em palavras, nos perguntamos se B é mensurável. Note que

$$\mathbb{1}_B^{-1}((-\infty, a]) = \{\omega \in \Omega : \mathbb{1}_B(\omega) \leq a\}.$$

Observamos que:

$$\mathbb{1}_B^{-1}((-\infty, a]) = \begin{cases} \emptyset, & a < 0, \\ B^c, & 0 \leq a < 1, \\ \Omega, & a \geq 1. \end{cases}$$

Como $B^c \notin \mathcal{F}$, segue que $\mathbb{1}_B$ não é mensurável, e portanto não é uma variável aleatória.

Exemplo 3.2 (Variável aleatória qualitativa)**.** A cor dos olhos de uma amostra de uma determinada população.

Exemplo 3.3 (Variável aleatória quantitativa)**.** O número de coroas que saíram num jogo de cara ou coroa.

Existem subcategorias para as variáveis aleatórias quantitativas e qualitativas. Uma variável aleatória qualitativa pode ser classificada em ordinal ou nominal. Já as variáveis aleatórias quantitativas podem ser classificadas em variáveis aleatórias quantitativas discretas ou contínuas.

Exemplo 3.4 (Variável aleatória qualitativa ordinal)**.** Estágios de uma doença (Leve, Moderado, Grave) é um exemplo de variável aleatória qualitativa ordinal, pois é necessário uma ordem ou hierarquia.

Exemplo 3.5 (Variável aleatória qualitativa nominal)**.** A cor dos olhos de uma amostra de uma determinada população é um exemplo de variável aleatória qualitativa nominal.

Exemplo 3.6 (Variável aleatória quantitativa discreta)**.** O número de coroas que saíram num jogo de cara ou coroa é um exemplo de variável aleatória quantitativa discreta.

Exemplo 3.7 (Variável aleatória quantitativa contínua)**.** A altura de uma amostra de uma população é um exemplo de variável aleatória quantitativa contínua.

Os tipos de variáveis aleatórias que nos interessam, de fato, são as variáveis aleatórias quantitativas discretas e contínuas.

3.3 Variáveis aleatórias discretas

Definição 13 (Variáveis aleatórias discretas)**.** Uma variável aleatória é dita discreta se existe um conjunto enumerável $B = \{x_1, x_2, ...\} \subset \mathbb{R}$ tal que:

$$\mathbb{P}(X \in B) = 1.$$

A função de probabilidade associada a X é a função:

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

tal que:

$$p_X(x) \geq 0 \quad \text{e} \quad \sum_x p_X(x) = 1.$$

Em outras palavras, uma variável aleatória discreta é uma função $X : \Omega \to \mathbb{R}$ cujo conjunto de valores possíveis é finito ou enumerável, ou seja, a variável aleatória assume valores que possam ser contados. Irei omitir a explicação do que é um conjunto enumerável mas essa ideia junto com a demonstração pode ser verificada no livro do Elon Lima (2022).

Exemplo 3.8. Um exemplo de uma variável aleatória quantitativa discreta é o resultado do lançamento de um dado, cujo resultado é um número de 1 a 6.

3.4 Variáveis aleatórias contínuas

Uma variável aleatória é contínua, ou absolutamente contínua, quando sua distribuição admite uma função densidade de probabilidade, a qual permite calcular probabilidades por meio de integrais.

Definição 14 (Variáveis aleatórias contínuas)**.** A variável aleatória X é absolutamente contínua se existe uma função $f : \mathbb{R} \to \mathbb{R}_+$, digamos contínua por partes, tal que:

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)\, dx.$$

Neste caso, dizemos que f é a função densidade de probabilidade de X ou simplesmente densidade de X. Note que $\int_{-\infty}^{+\infty} f(x)\, dx = 1$.

Intuitivamente, dizemos que uma variável aleatória é contínua quando as probabilidades não são atribuídas a valores isolados, mas a intervalos de valores.

Exemplo 3.9. O tempo que uma lampada dura até parar de funcionar é um exemplo de variável aleatória contínua.

3.5 Variáveis aleatórias independentes

Definição 15 (Variáveis aleatórias independentes)**.** As variáveis aleatórias $X_1, ..., X_n$ são coletivamente independentes se:

$$\mathbb{P}(X_1 \in B_1, X_2 \in B_2, ..., X_n \in B_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in B_i),$$

$\forall B_i \in \mathcal{F}, i = 1, ..., n.$

Podemos traduzir isso como:

$$\mathbb{P}(\{X_1 \in B_1\} \cap \{X_2 \in B_2\} \cap ... \cap \{X_n \in B_n\}) = \mathbb{P}(X_1 \in B_1) \cdot \mathbb{P}(X_2 \in B_2) \cdot ... \cdot \mathbb{P}(X_n \in B_n).$$

Segue a mesma lógica da definição 7, só é preciso lembrar que estamos falando de variáveis aleatórias agora. Essa ideia será de extrema importância em cadeias de Markov. Vejamos alguns exemplos de variáveis aleatórias.

3.5.1 Distribuição binomial

A distribuição binomial é utilizada para modelar o número de sucessos em n ensaios independentes, nos quais cada ensaio possui apenas dois resultados possíveis: sucesso ou fracasso, com probabilidade de sucesso p.

Seja X a variável aleatória que representa o número de sucessos nesses n ensaios. Dizemos que X possui distribuição binomial com parâmetros n e p se sua função de probabilidade é dada por:

$$\mathbb{P}(X = x) = p^x \cdot (1-p)^{n-x} \cdot \binom{n}{x}.$$

É comum ver nos livros a seguinte notação para distribuição binomial: $X \sim \text{Binomial}(n, p)$.

Exemplo 3.10. Imagine o lançamento de uma moeda honesta 10 vezes. Qual é a probabilidade de obter exatamente 7 caras?

Seja X a variável aleatória que representa o número de caras obtidas nos 10 lançamentos. Considere:

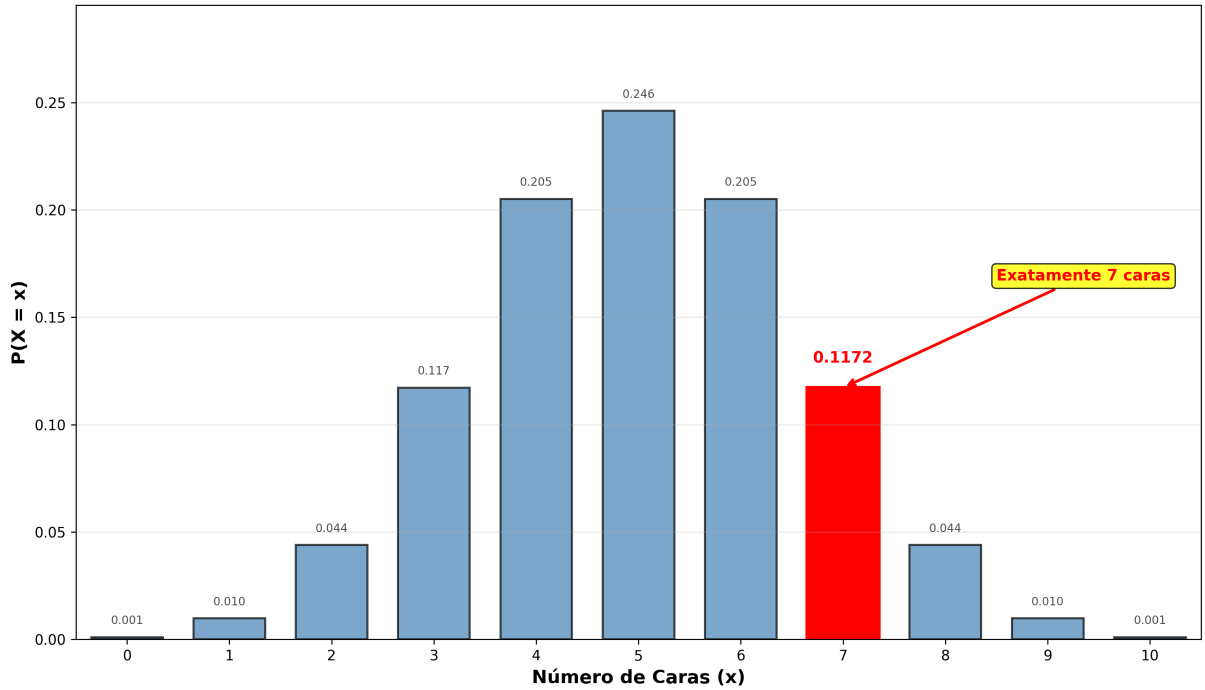
- $n = 10$ é o número de lançamentos;
- $x = 7$ é o numero de sucessos (Caras);
- $p = 0,5$ é a probabilidade de sair cara;
- $1 - p$ é o complementar de p, ou seja, é a probabilidade de sair coroa.

Logo,

$$\mathbb{P}(X = 7) = (0,5)^7 \cdot (0,5)^{10-7} \cdot \binom{10}{7} \approx 11,72\%.$$

O grafico abaixo ilustra a solução do problema:

Figura 4 – Representação do problema da moeda binomial



Fonte: Próprio autor (2025).

3.5.2 Distribuição geométrica

A distribuição geométrica é utilizada quando realizamos o experimento e queremos obter o número de tentativas até ocorrer o primeiro estado chamado sucesso. A distribuição geométrica tem parâmetro p, sendo seu complementar $q = (1 - p)$. Então, a distribuição geométrica tem função de probabilidade dada por:

$$\mathbb{P}(X = x) = p \cdot (1 - p)^{x-1}.$$

Uma observação é que é muito comum ver nos livros a seguinte notação para distribuição geométrica: $X \sim \text{Geom}(p)$.

Uma curiosidade é que a distribuição geométrica possui o que é chamado de perda de memória. Esta característica é:

$$\mathbb{P}(X \geq m + n | X \geq m) = \mathbb{P}(X > n).$$

Demonstração. Usando a definição de probabilidade condicional, segue que:

$$\mathbb{P}(X \geq m + n | X \geq m) = \frac{\mathbb{P}(\{X \geq m + n\} \cap \{X \geq m\})}{\mathbb{P}(\{X \geq m\})}.$$

Como $\{X \geq m + n\} \subset \{X \geq m\}$, temos que:

$$\mathbb{P}(X \geq m+n | X \geq m) = \frac{\mathbb{P}(\{X \geq m+n\}}{\mathbb{P}(\{X \geq m\})}.$$

Agora, quem é $\mathbb{P}(X \geq m)$?

$$\begin{aligned}\mathbb{P}(X \geq m) &= \mathbb{P}(\{X = m\} \cup \{X = m+1\} \cup \{X = m+2\} \cup \dots) \\ &= \mathbb{P}\left(\bigcup_{k=m}^{+\infty} \{X = k\}\right).\end{aligned}$$

Como essa união é disjunta, então, pelas definições de probabilidade e da função de probabilidade da distribuição geométrica, obtemos:

$$\begin{aligned}\mathbb{P}(X \geq m) &= \sum_{k=m}^{+\infty} \mathbb{P}(X = k) \\ &= \sum_{k=m}^{+\infty} p \cdot (1-p)^{k-1} \\ &= p \cdot \sum_{k=m}^{+\infty} (1-p)^{k-1} = p \cdot \sum_{k=m}^{+\infty} \frac{(1-p)^k}{(1-p)} \\ &= \frac{p}{(1-p)} \cdot \sum_{k=m}^{+\infty} (1-p)^k.\end{aligned}$$

Analisando $\sum_{k=m}^{+\infty} (1-p)^k = (1-p)^m + (1-p)^{m+1} + (1-p)^{m+2} + ...$, vemos que é a soma de uma PG infinita com primeiro termo $(1-p)^m$ e razão $(1-p)$, cuja soma é $\frac{(1-p)^m}{1-(1-p)}$. Logo,

$$\mathbb{P}(X \geq m) = \frac{p}{1-p} \cdot \left(\frac{(1-p)^m}{1-(1-p)}\right) = \frac{(1-p)^m}{(1-p)} = (1-p)^{m-1}.$$

Voltando ao problema, obtemos então:

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

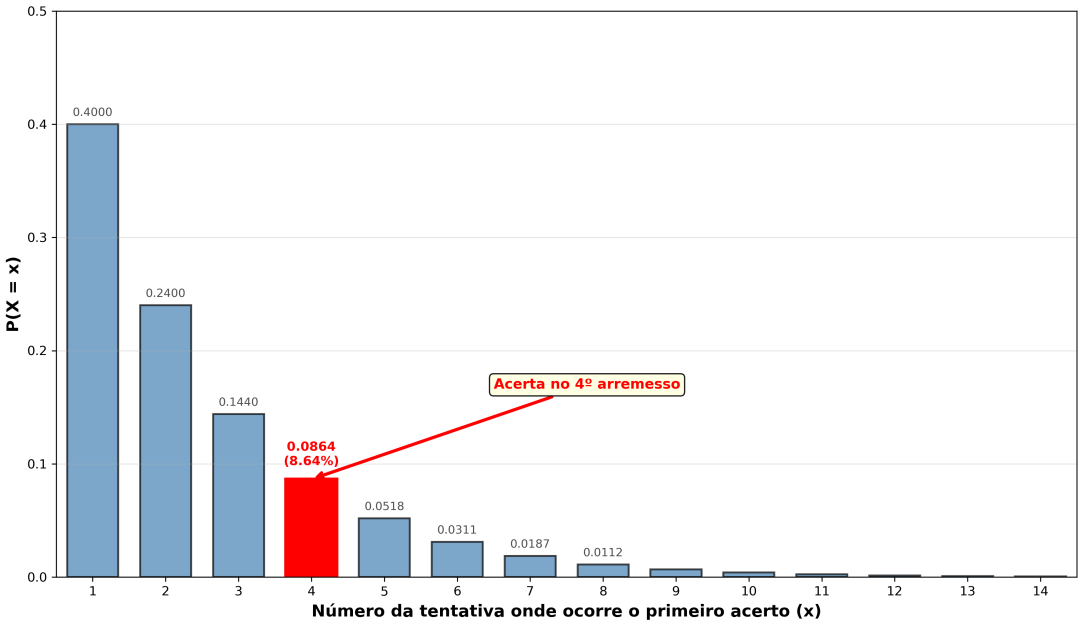
□

Exemplo 3.11. Um jogador de basquete tem uma probabilidade de sucesso de $p = 0,4$ ou 40% de acertar um arremesso livre. Qual a probabilidade de ele acertar a sua primeira cesta no 4º arremesso?

Seja X o número de arremessos até o primeiro acerto. Temos que $p = 0,4$; $1-p = 0,6$ e $x = 4$ é o número de tentativas onde ocorre o primeiro acerto. Usando a fórmula $\mathbb{P}(X = x) = p \cdot (1-p)^{x-1}$ podemos substituir x, p e $1-p$, obtendo:

$$\mathbb{P}(X = 4) = (0,4) \cdot (0,6)^3 \approx 0,0864 \approx 8,64\%.$$

Figura 5 – Jogador de basquete



Fonte: Próprio autor (2025).

3.5.3 Distribuição de Poisson

A distribuição de Poisson é utilizada para modelar o número de ocorrências de um evento em um intervalo fixo de tempo ou espaço, sob a hipótese de que tais ocorrências acontecem de forma independente e com taxa média constante. Trata-se de uma distribuição de variável aleatória discreta.

Dizemos que uma variável aleatória X possui distribuição de Poisson com parâmetro $\lambda > 0$ se:

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

Nesse caso, escrevemos $X \sim \text{Poisson}(\lambda)$.

A distribuição de Poisson pode ser obtida como limite da distribuição binomial quando o número de ensaios tende ao infinito e a probabilidade de sucesso tende a zero, de modo que $np = \lambda$ permanece constante.

Escrevendo a fórmula da distribuição binomial, obtemos:

$$\mathbb{P}(X = x) = p^x \cdot (1-p)^{n-x} \binom{n}{x}.$$

Agora, chamamos λ de esperança ou valor esperado. Então, podemos fazer $\lambda = n \cdot p$, sendo n o número de eventos e p a probabilidade de ocorrer um evento. Logo, $\lambda = n \cdot p \Rightarrow p = \frac{\lambda}{n}$. Substituindo no lugar de p, obtemos:

$$\mathbb{P}(X = x) = \binom{\lambda}{n}^{x} \left(1 - \frac{\lambda}{n}\right)^{n-x} \binom{n}{x}.$$

Como n tende ao infinito, podemos escrever como limite, então:

$$\mathbb{P}(X = x) = \lim_{n\to\infty} \left[\binom{\lambda}{n}^{x} \left(1 - \frac{\lambda}{n}\right)^{n-x} \binom{n}{x} \right],$$

desenvolvendo, obtemos:

$$\begin{aligned}
\mathbb{P}(X = x) &= \lim_{n\to\infty} \left[\frac{\lambda^x}{n^x} \cdot \frac{(1 - \frac{\lambda}{n})^n}{(1 - \frac{\lambda}{n})^x} \cdot \frac{n!}{(n-x)! \cdot x!} \right] \\
&= \frac{\lambda^x}{x!} \cdot \lim_{n\to\infty} \frac{n!}{n^x (n-x)!} \cdot \lim_{n\to\infty} \left[\left(1 - \frac{\lambda}{n}\right)^n \right] \cdot \left(\frac{1}{\lim\limits_{n\to\infty} (1 - \frac{\lambda}{n})^x} \right)
\end{aligned}$$

Note que $\lim\limits_{n\to\infty} \frac{1}{(1-\frac{\lambda}{n})^x} = 1$ e $\lim\limits_{n\to\infty} \left[(1 - \frac{\lambda}{n})^n\right] = e^{-\lambda}$. Além disso,

$$\begin{aligned}
\lim_{n\to\infty} \frac{n!}{n^x (n-x)!} &= \lim_{n\to\infty} \left[\frac{1}{n^x} \cdot \frac{n(n-1)(n-2) \cdot \cdots \cdot (n-x+1)(n-x)!}{(n-x)!} \right] \\
&= \lim_{n\to\infty} \left[\frac{n}{n} \cdot \frac{n-1}{n} \cdot \frac{n-2}{n} \cdot \cdots \cdot \frac{n-x+1}{n} \right] \\
&= \lim_{n\to\infty} \left(1 - \frac{1}{n}\right) \cdot \lim_{n\to\infty} \left(1 - \frac{2}{n}\right) \cdot \cdots \cdot \lim_{n\to\infty} \left(1 - \frac{x+1}{n}\right) = \cdot 1 \cdot 1 \cdot \ldots \cdot 1 = 1.
\end{aligned}$$

Assim, segue que:

$$\mathbb{P}(X = x) = \frac{\lambda^x}{x!} \cdot e^{-\lambda},$$

logo,

$$\mathbb{P}(X = x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}.$$

Exemplo 3.12. Um call center de suporte técnico recebe em média 8 chamadas por hora durante o horário comercial. A gerência deseja analisar a demanda de atendimento para planejar melhor a equipe. Considerando que o número de chamadas segue uma distribuição de Poisson, calcule:

i. Qual a probabilidade de receber exatamente 10 chamadas em uma hora?

ii. Qual a probabilidade de receber no máximo 5 chamadas em uma hora?

iii. Qual a probabilidade de receber mais de 12 chamadas em uma hora?

Solução:

i. $\lambda = 8$ chamadas por hora (taxa média) e $X =$ número de chamadas por hora. Basta substituir:

$$\mathbb{P}(X = 10) = \frac{e^{-8} \cdot \lambda^{10}}{10!} \approx 9,93\%.$$

ii. Receber no máximo 5 chamadas é receber 0 chamadas, 1 chamada, 2 chamadas, 3 chamadas, 4 chamadas até 5 chamadas, denotado por:

$$\mathbb{P}(X \leq 5) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) + \mathbb{P}(X = 3) + \mathbb{P}(X = 4) + \mathbb{P}(X = 5).$$

Desenvolvendo, obtemos:

$$\begin{aligned}\mathbb{P}(X \leq 5) &= \frac{e^{-8} \cdot 8^0}{0!} + \frac{e^{-8} \cdot 8^1}{1!} + \frac{e^{-8} \cdot 8^2}{2!} + \frac{e^{-8} \cdot 8^3}{3!} + \frac{e^{-8} \cdot 8^4}{4!} + \frac{e^{-8} \cdot 8^5}{5!} \\ &= 0,0003 + 0,0027 + 0,0107 + 0,0286 + 0,0573 + 0,0916 \\ &\approx 0,1912 \\ &\approx 19,12\%.\end{aligned}$$

iii. A probabilidade de receber mais de 12 chamadas pode ser denotada por:

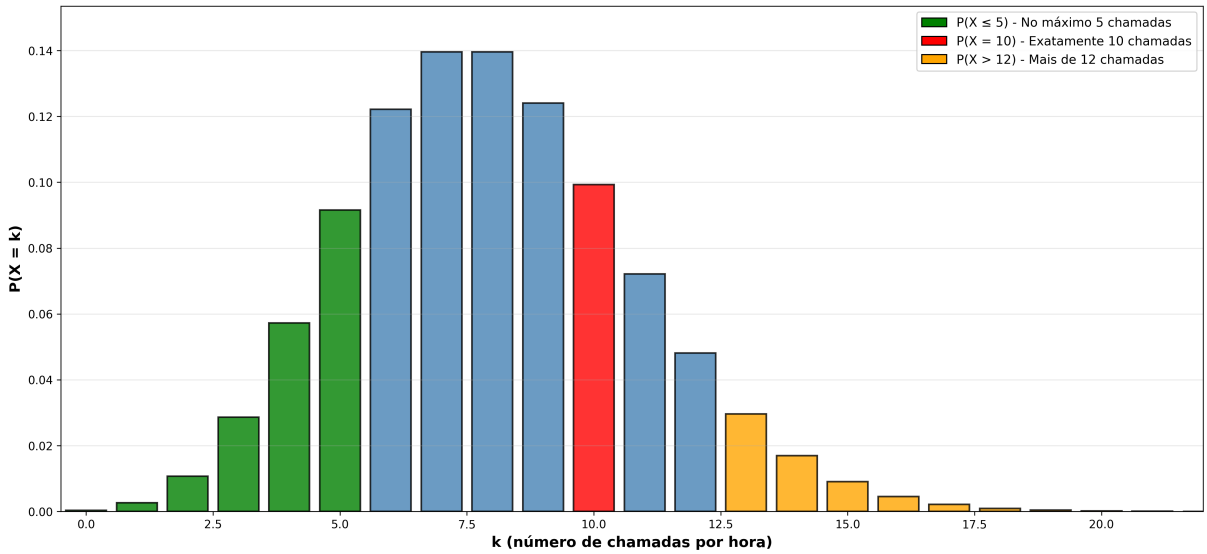
$$\mathbb{P}(X > 12) = 1 - \mathbb{P}(X \leq 12).$$

Logo,

$$\begin{aligned}\mathbb{P}(X > 12) &= 1 - (\mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) + ... + \mathbb{P}(X = 12)) \\ \mathbb{P}(X > 12) &= 1 - \left[\frac{e^{-8} \cdot 8^0}{0!} + \frac{e^{-8} \cdot 8^1}{1!} + ... + \frac{e^{-8} \cdot 8^{12}}{12!}\right] \\ &= 1 - 0,9362 \\ &\approx 0,0638 \\ &\approx 6,38\%.\end{aligned}$$

O gráfico abaixo ilustra esses resultados.

Figura 6 – Distribuição de Poisson



Fonte: Próprio autor (2025).

3.5.4 Distribuição exponencial

A distribuição exponencial é uma distribuição de variável aleatória contínua, cuja função densidade de probabilidade é dada por:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

O parâmetro $\lambda > 0$ representa a taxa média de ocorrência do evento por unidade de tempo. Valores maiores de λ indicam maior frequência de ocorrências e, consequentemente, menor tempo médio até o evento.

A distribuição exponencial é amplamente utilizada em estudos de confiabilidade, especialmente para modelar o tempo até a falha de componentes cuja taxa de falha é aproximadamente constante ao longo do tempo, como lâmpadas, dispositivos eletrônicos e sistemas simples.

Ademais, a distribuição exponencial também possui a propriedade de perda de memória, sendo a única distribuição contínua que possui esta propriedade. Assim, o tempo restante até a ocorrência do evento é independente do tempo já decorrido.

Além disso, a distribuição exponencial está intimamente relacionada ao processo de Poisson. Em um processo de Poisson com taxa λ, os tempos entre ocorrências consecutivas são variáveis aleatórias independentes com distribuição exponencial de parâmetro λ.

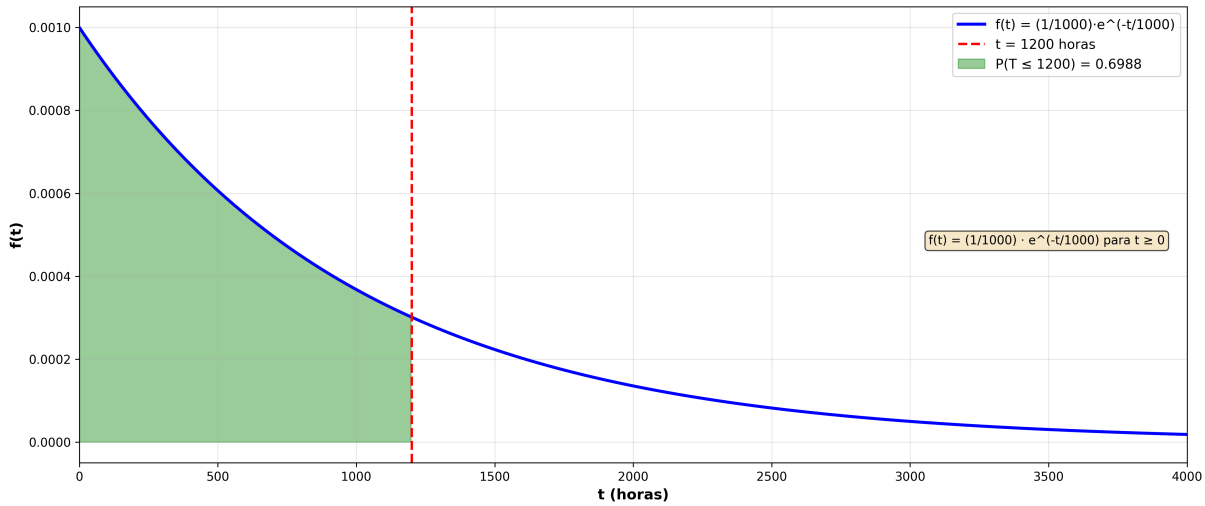
Exemplo 3.13 (TRT 7° R - Adaptada)**.** A duração de uma lâmpada é uma variável aleatória T com distribuição exponencial de parâmetro $\lambda = \frac{1}{1000}$. Qual é a probabilidade de a lâmpada durar menos de 1200 horas?

$$f(x) = \begin{cases} \frac{1}{1000} \cdot e^{\frac{-t}{1000}} & \text{se } t \geq 0 \\ 0 & \text{se } t < 0. \end{cases}$$

O tempo t estará entre zero e 1200. Assim, segue que:

$$\mathbb{P}(T \leq 1200) = \int_0^{1200} \frac{1}{1000} e^{-t/1000}\, dt = 1 - e^{-1,2} \approx 0,70.$$

Figura 7 – Distribuição exponencial



Fonte: Próprio autor (2025).

Este exemplo foi tirado de uma prova de concurso BRASIL. Tribunal Regional do Trabalho da 7a Região (Ceará) (2009).

3.6 Função de distribuição

Definição 16. A função de distribuição da variável aleatória X é definida por $F_X : \mathbb{R} \to [0, 1]$ com:

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

Proposição 3.6.1 (Propriedades da função de distribuição)**.** Se X é uma variável aleatória, sua função de distribuição F_X satisfaz as seguintes propriedades:

i. Se $x \leq y$, então $F_X(x) \leq F_X(y)$, isto é, F_X é não decrescente;

ii. Se $x_n \searrow y$, então $F_X(x_n) \searrow F_X(y)$, isto é, F_X é contínua à direita;

iii. Se $x_n \searrow -\infty$, então $F_X(x_n) \searrow 0$, e se $x_n \nearrow +\infty$, então $F_X(x_n) \nearrow 1$.

As distribuições discretas, como a binomial, a geométrica e a de Poisson, possuem funções de distribuição com saltos, enquanto distribuições contínuas, como a exponencial e a normal, apresentam funções de distribuição contínuas, suaves, como é possível observar no exemplo 3.13.

Demonstração. Seja a tripla $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade, então:

i. Sejam os valores x e y, sendo $x \leq y$. Note que $\{X \leq x\} \subseteq \{X \leq y\}$. Pela definição de função de distribuição, temos que $F_X(x) = \mathbb{P}(X \leq x) \leq \mathbb{P}(X \leq y) = F_X(y)$.

ii. Se $x_n \searrow y$, então $\{X \leq x_n\}$ é uma sequência decrescente de eventos aleatórios. Logo, $\{X \leq x_n\} \searrow \{X \leq y\}$ e, pela continuidade da probabilidade, $\mathbb{P}(\{X \leq x_n\}) \searrow \mathbb{P}(\{X \leq y\})$. Portanto, $F_X(x_n) \searrow F_X(y)$.

iii. A terceira propriedade, é possível demonstrar da mesma forma que a propriedade *ii*. Se $x_n \searrow -\infty$, então $\{X \leq x_n\}$ diminuem para o conjunto vazio, ou seja, $\{X \leq x_n\} \searrow \emptyset$. Logo, pela continuidade da probabilidade, temos que $\mathbb{P}(\{X \leq x_n\}) \searrow 0$. Assim, pela definição temos que $F_X(x_n) \searrow 0$ e, da mesma forma, seguindo os mesmos passos, só que agora $x_n \nearrow +\infty$, pela continuidade também chegamos a conclusão de que $F_X(x_n) \nearrow 1$.

□

Variáveis aleatórias independentes e identicamente distribuídas (i.i.d), significa que as variáveis aleatórias são independentes (definição 15) e elas possuem a mesma distribuição.

3.7 Esperança

A esperança, ou média ponderada de uma variável aleatória, é justamente o que a palavra diz, é o valor esperado. Esse valor esperado para a variável aleatória, nem sempre estará no espaço amostral.

Definição 17. Seja X uma variável aleatória. Se X é discreta assumindo os valores $\{x_1, x_2, ...\}$, então definimos a esperança de X como:

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} x_k \cdot \mathbb{P}(X = x_k).$$

Definição 18. Seja X uma variável aleatória. Se X é absolutamente contínua com densidade $f : \mathbb{R} \to \mathbb{R}_+$, então definimos a esperança de X como:

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

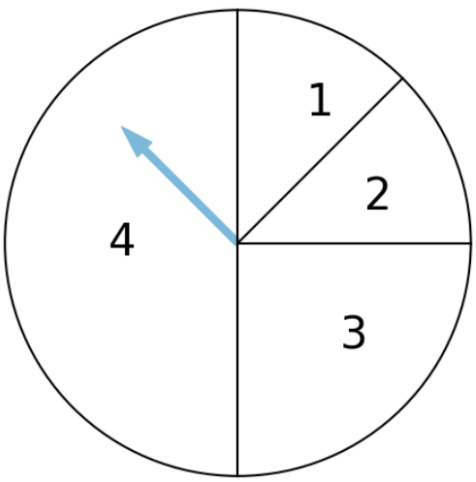
Exemplo 3.14. Em Las Vegas, num cassino, um apostador entra em um jogo de roleta. Essa roleta é diferente, pois é dividida em setores com áreas diferentes, conforme a Figura 8. Surgem duas perguntas: qual a esperança da variável aleatória X? Qual setor da roleta a seta tem uma maior chance de parar??

Considere o espaço amostral Ω como o disco unitário de raio $r = 1$ no plano cartesiano, centrado na origem $(0, 0)$, cuja fronteira é definida pela equação $x^2 + y^2 = 1$. Assumimos que a posição final da seta é distribuída uniformemente em relação à área do disco. Assim, a probabilidade de a seta parar em uma região é proporcional à sua área. O espaço amostral Ω possui área total $A(\Omega) = \pi r^2 = \pi$. A variável aleatória X é discreta e assume os valores $\{1, 2, 3, 4\}$, dependendo do setor em que a seta parar.

Definimos a probabilidade de a seta cair em um setor S_k como a razão entre a área do setor e a área total do círculo denotado por:

$$\mathbb{P}(X = k) = \frac{A(S_k)}{A(\Omega)}.$$

Figura 8 – Roleta do cassino



Fonte: Possani (2022).

Os setores S_k e suas respectivas probabilidades são dados:
Setor 4: Ocupa metade do círculo. $\mathbb{P}(X = 4) = \frac{1}{2}$.
Setor 3: Ocupa um quarto do círculo. $\mathbb{P}(X = 3) = \frac{1}{4}$.
Setor 2: Ocupa um oitavo do círculo. $\mathbb{P}(X = 2) = \frac{1}{8}$.
Setor 1: Ocupa um oitavo do círculo. $\mathbb{P}(X = 1) = \frac{1}{8}$.

A esperança (ou média) da variável X é calculada conforme a definição 17:

$$\mathbb{E}[X] = \sum_{k=1}^{4} x_k \cdot \mathbb{P}(X = x_k).$$

Seguindo isso, obtemos

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=1}^{4} k \cdot \mathbb{P}(X = k) \\
&= 4 \cdot \mathbb{P}(X = 4) + 3 \cdot \mathbb{P}(X = 3) + 2 \cdot \mathbb{P}(X = 2) + 1 \cdot \mathbb{P}(X = 1) \\
&= 4\left(\frac{1}{2}\right) + 3\left(\frac{1}{4}\right) + 2\left(\frac{1}{8}\right) + 1\left(\frac{1}{8}\right) \\
&= \frac{16}{8} + \frac{6}{8} + \frac{2}{8} + \frac{1}{8} \\
&= \frac{25}{8}.
\end{aligned}$$

Portanto, a esperança da variável sob análise é $\mathbb{E}[X] = 3,125$.

A esperança, ou média dessa variável X não é um valor que seja possível da seta interceptar e a jogada de maior vantagem certamente é apostar que a seta ira cair no setor de número 4, pois a probabilidade da seta parar nele é maior.

Exemplo 3.15 (Adaptado)**.** Uma seguradora disponibiliza uma apólice de seguros para um estudante do IFBA que irá fazer estágio externo. O valor dessa apólice é de $R\$50.000{,}00$ caso o estudante sofra algum acidente e o estudo feito mostrou que existe uma chance de 2% que um acidente ocorra. Qual o valor esperado pela empresa de seguros ao disponibilizar esse produto, admitindo como variável aleatória a ocorrência ou não de acidente?

Solução:

Primeiro temos dois estados, o primeiro é o acidente ocorrer, que tem uma probabilidade de 2%, e o segundo é a probabilidade de não ocorrer que equivale ao complementar de 2%, que é 98%. Sendo assim, o valor para caso o acidente ocorra é de $R\$50.000,00$ e o valor para o caso do acidente não ocorrer é $R\$0,00$. Definimos assim a variável aleatória X como o valor, em reais, pago pela seguradora:

$$X = \begin{cases} 50.000, & \text{com probabilidade 0,02,} \\ 0, & \text{com probabilidade 0,98.} \end{cases}$$

Logo, a esperança é dada por:

$$\mathbb{E}[X] = v_1 \cdot \mathbb{P}(X = 50.000) + v_2 \cdot \mathbb{P}(X = 0).$$

Sendo $v_1 = 50.000,00, v_2 = 0,00; \mathbb{P}(X = 0) = 2\%$ e $\mathbb{P}(X = 50.000) = 98\%$, substituindo, obtemos

$$\begin{aligned}\mathbb{E}[X] &= 50.000 \cdot 0,02 + 0 \cdot 0,98 \\ &= 1.000.\end{aligned}$$

Portanto, o valor esperado do pagamento da seguradora é de R$ 1.000,00.

Exemplo 3.16 (Exemplo de esperança de variável aleatória contínua)**.** No exemplo 3.13, a função densidade de probabilidade é dada por:

$$f(t) = \begin{cases} \frac{1}{1000} \cdot e^{\frac{-t}{1000}} & \text{se } t \geq 0 \\ 0 & \text{se } t < 0. \end{cases}$$

Iremos agora calcular o valor esperado para esse problema, sendo $\lambda = \frac{1}{1000}$ e $t \geq 0$. Então usando usando a formula da esperança de variável aleatória contínua obtemos:

$$\mathbb{E}[T] = \int_{-\infty}^{+\infty} t \cdot f(t)\, dt.$$

Substituindo, temos:

$$\mathbb{E}[T] = \int_{0}^{\infty} t \cdot \frac{1}{1000} e^{-\frac{t}{1000}} dt.$$

Resolvendo essa integral por partes, obtemos:

$$\mathbb{E}[T] = \frac{1}{\frac{1}{1000}} = \frac{1}{1} \cdot \frac{1000}{1} = 1000.$$

Curiosamente, a esperança para uma distribuição exponencial é $\frac{1}{\lambda}$. A demonstração para essa e as demais distribuições, são encontradas no livro do Franco (2020).

3.8 Propriedades da esperança

Teorema 3.1 (Propriedades da esperança)**.** A esperança matemática possui as seguintes propriedades:

i. Se $X = c$, então $\mathbb{E}[X] = c$;

ii. Se $X \leq Y$ então $\mathbb{E}[X] \leq \mathbb{E}[Y]$ (Monotonicidade);

iii. Se $Z = aX + Y$ então, $\mathbb{E}[Z] = a\mathbb{E}[X] + \mathbb{E}[Y]$ (Linearidade);

iv. Se X, Y são variáveis aleatórias independentes, então $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$ (Esperança do Produto).

v. Dados $h : \mathbb{R} \to \mathbb{R}$ e X variável aleatória, então

$$\mathbb{E}[h(X)] = \sum_{k=1}^{\infty} h(x_k) \cdot \mathbb{P}(X = x_k),$$

se X é discreta, ou

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

se x for absolutamente contínua. (Teorema da mudança de variável)

A demonstração dessas propriedades serão omitidas pois algumas usam teoria da medida, mas podem ser consultadas no livro do James (2023).

3.9 Variância

A variância é uma das medidas de dispersão de uma variável aleatória que mede o quanto a variável se distancia da média, ou valor esperado.

Definição 19 (Variância)**.** Seja a tripla $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade. A variância de uma variável aleatória X é definida por

$$V_{ar}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

Proposição 3.9.1 (Propriedades da variância)**.** As propriedades da variância são

i. $V_{ar}(X + c) = V_{ar}(X), \forall c \in \mathbb{R}$;

ii. $V_{ar}(a \cdot X) = a^2 \cdot V_{ar}(X), \forall a \in \mathbb{R}$;

iii. $V_{ar}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$;

iv. Se X e Y são independentes, então $V_{ar}(X + Y) = V_{ar}(X) + V_{ar}(Y)$;

v. Se $X_1, X_2, ..., X_n$ são independentes, então

$$V_{ar}\left(\sum_{k=1}^{n} X_i\right) = \sum_{k=1}^{n} V_{ar}(X_n).$$

Demonstração. Seja a tripla $(\Omega, \mathcal{F}, \mathbb{P})$, um espaço de probabilidade, temos:

i. Pela definição de variância, temos:

$$V_{ar}(X + c) = \mathbb{E}\left[((X + c) - \mathbb{E}[X + c])^2\right].$$

Pela linearidade da esperança, sabe-se que $\mathbb{E}[X + c] = \mathbb{E}[X] + c$. Substituindo este resultado na expressão anterior, obtemos:

$$\begin{aligned} V_{ar}(X + c) &= \mathbb{E}\left[(X + c - (\mathbb{E}[X] + c))^2\right] \\ &= \mathbb{E}\left[(X + \not{c} - \mathbb{E}[X] - \not{c})^2\right] \\ &= \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \\ &= V_{ar}(X). \end{aligned}$$

Portanto, a adição de uma constante não altera a dispersão dos dados em relação à média.

ii. Pela definição de variância, temos:

$$V_{ar}(aX) = \mathbb{E}\left[(aX - \mathbb{E}[aX])^2\right].$$

Pela propriedade de linearidade da esperança, temos que $\mathbb{E}[aX] = a\mathbb{E}[X]$. Substituindo esta relação na expressão, obtemos:

$$\begin{aligned} V_{ar}(aX) &= \mathbb{E}\left[(aX - a\mathbb{E}[X])^2\right] \\ &= \mathbb{E}\left[(a(X - \mathbb{E}[X]))^2\right] \\ &= \mathbb{E}\left[a^2(X - \mathbb{E}[X])^2\right] \\ &= a^2\mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \\ &= a^2 V_{ar}(X). \end{aligned}$$

Onde a penúltima igualdade decorre da linearidade da esperança para constantes. Conclui-se que o fator de escala é extraído da variância elevado ao quadrado.

iii. Partindo da definição canônica de variância e desenvolvendo o binômio de grau dois no interior do operador de esperança, temos:

$$\begin{aligned} V_{ar}(X) &= \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \\ &= \mathbb{E}\left[X^2 - 2X\mathbb{E}[X] + (\mathbb{E}[X])^2\right] \end{aligned}$$

Pela propriedade de linearidade do operador de esperança, e considerando que $\mathbb{E}[X]$ é um valor constante, podemos distribuir o operador nos termos da soma:

$$\begin{aligned} V_{ar}(X) &= \mathbb{E}[X^2] - \mathbb{E}[2X\mathbb{E}[X]] + \mathbb{E}[(\mathbb{E}[X])^2] \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + (\mathbb{E}[X])^2 \\ &= \mathbb{E}[X^2] - 2(\mathbb{E}[X])^2 + (\mathbb{E}[X])^2 \\ &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \end{aligned}$$

Portanto, a identidade está demonstrada, revelando uma forma alternativa e frequentemente mais simples para o cálculo da dispersão.

iv. Pela definição de variância aplicada à soma, temos:

$$V_{ar}(X+Y) = \mathbb{E}\left[((X+Y) - \mathbb{E}[X+Y])^2\right].$$

Considerando a linearidade da esperança e reagrupando os termos:

$$\begin{aligned} V_{ar}(X+Y) &= \mathbb{E}\left[((X-\mathbb{E}[X]) + (Y-\mathbb{E}[Y]))^2\right] \\ &= \mathbb{E}\left[(X-\mathbb{E}[X])^2 + 2(X-\mathbb{E}[X])(Y-\mathbb{E}[Y]) + (Y-\mathbb{E}[Y])^2\right]. \end{aligned}$$

Distribuindo o operador de esperança pelo princípio da linearidade:

$$V_{ar}(X+Y) = V_{ar}(X) + 2\mathbb{E}\left[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\right] + V_{ar}(Y).$$

Como X e Y são independentes, as funções de X e Y também o são. Assim, a esperança do produto dos desvios é igual ao produto das esperanças:

$$\mathbb{E}\left[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\right] = \mathbb{E}[X-\mathbb{E}[X]] \cdot \mathbb{E}[Y-\mathbb{E}[Y]].$$

Visto que a esperança do desvio em relação à média é sempre nula, ou seja, $\mathbb{E}[X - \mathbb{E}[X]] = 0$, o termo central da expressão desaparece. Portanto:

$$V_{ar}(X+Y) = V_{ar}(X) + V_{ar}(Y).$$

v. Utilizaremos o princípio da indução finita sobre o número de variáveis n.

Base da Indução: Para $n = 1$, a igualdade é trivial, pois $V_{ar}(X_1) = V_{ar}(X_1)$. Para $n = 2$, a propriedade foi demonstrada no item anterior, onde a independência entre X_1 e X_2 implica que $V_{ar}(X_1 + X_2) = V_{ar}(X_1) + V_{ar}(X_2)$.

Hipótese de Indução: Suponhamos que a propriedade seja válida para um natural $k \geq 2$, ou seja:

$$V_{ar}\left(\sum_{i=1}^{k} X_i\right) = \sum_{i=1}^{k} V_{ar}(X_i).$$

Passo Indutivo: Devemos mostrar que a validade para k implica a validade para $k+1$. Seja $S_k = \sum_{i=1}^{k} X_i$. Note que S_k e X_{k+1} são independentes, pois X_{k+1} é independente de cada X_i individualmente. Assim:

$$\begin{aligned} V_{ar}\left(\sum_{i=1}^{k+1} X_i\right) &= V_{ar}(S_k + X_{k+1}) \\ &= V_{ar}(S_k) + V_{ar}(X_{k+1}) \quad \text{(pelo caso } n=2\text{)} \\ &= \sum_{i=1}^{k} V_{ar}(X_i) + V_{ar}(X_{k+1}) \quad \text{(pela Hipótese de Indução)} \\ &= \sum_{i=1}^{k+1} V_{ar}(X_i). \end{aligned}$$

Portanto, pelo princípio da indução finita, a propriedade é válida para todo $n \in \mathbb{N}$.

□

3.10 Desigualdades clássicas

Proposição 3.10.1 (Desigualdade de Markov)**.** Seja X uma variável aleatória, com $X \geq 0$ e $\lambda \in \mathbb{R}$, com $\lambda > 0$. Então:

$$\mathbb{P}(X \geq \lambda) \leq \frac{\mathbb{E}[X]}{\lambda}.$$

Demonstração. Seja a tripla $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade, temos:

Para demonstrar essa desigualdade, podemos usar a função indicadora

$$\mathbb{1}_{\{X \geq a\}}(w) = \begin{cases} 1, & \text{se } X(w) \geq a \\ 0, & \text{se } X(w) < a. \end{cases}$$

Então, temos que $X = X \cdot \mathbb{1}_{\{X \geq a\}} + X \cdot \mathbb{1}_{\{X < a\}}$. Logo, podemos escrever a esperança de X como

$$\mathbb{E}[X] = \mathbb{E}[X \cdot \mathbb{1}_{\{X \geq a\}} + X \cdot \mathbb{1}_{\{X < a\}}] = \mathbb{E}[X \cdot \mathbb{1}_{\{X \geq a\}}] + \mathbb{E}[X \cdot \mathbb{1}_{\{X < a\}}].$$

Note que $\mathbb{E}[X \cdot \mathbb{1}_{\{X < a\}}] \geq 0$, logo:

$$\mathbb{E}[X] \geq \mathbb{E}[a \cdot \mathbb{1}_{\{X \geq a\}}].$$

Pela propriedade da linearidade da esperança, propriedade *ii*, teorema 3.1, podemos escrever:

$$\mathbb{E}[X] \geq a \cdot \mathbb{E}[\mathbb{1}_{\{X \geq a\}}] \Rightarrow \mathbb{E}[X] \geq a \cdot \mathbb{P}(X \geq a).$$

E, portanto:

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a},$$

com $a > 0$. □

Proposição 3.10.2 (Desigualdade de Tchebyshev)**.** Seja X uma variável aleatória com $\mathbb{E}[X]$ e $\mathrm{V}_{\mathrm{ar}}(X) < \infty$. Então, para todo $\lambda > 0$,

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \lambda) \leq \frac{\mathrm{V}_{\mathrm{ar}}(X)}{\lambda^2}.$$

Demonstração. Aplicando a desigualdade de Markov à variável não-negativa $(X - \mathbb{E}(X))^2$, obtemos:

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \lambda) = \mathbb{P}\big((X - \mathbb{E}(X))^2 \geq \lambda^2\big) \leq \frac{\mathbb{E}[(X - \mathbb{E}(X)^2]}{\lambda^2} = \frac{\mathrm{V}_{\mathrm{ar}}(X)}{\lambda^2}.$$

□

Proposição 3.10.3 (Desigualdade de Jensen)**.** Seja $\phi : \mathbb{R} \to \mathbb{R}$ função convexa e, suponha que $\mathbb{E}[|X|] < +\infty$. Então,

$$\mathbb{E}[\phi(X)] \geq \phi(\mathbb{E}[X])$$

Omitiremos a demonstração, que pode ser encontrada no livro de Franco (2020).

3.11 Processo estocástico

Um processo estocástico pode ser entendido como uma família de variáveis aleatórias indexadas por um conjunto que representa o tempo. Esses processos são utilizados para modelar fenômenos aleatórios que evoluem ao longo do tempo, como sistemas físicos, econômicos ou naturais.

Definição 20 (Processo estocástico)**.** Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade e seja T um conjunto de índices (tempos). Uma família de variáveis aleatórias $\{X(t) : t \in T\}$ é chamada de processo estocástico se, para cada $t \in T$, $X(t)$ é uma variável aleatória definida em Ω.

Segundo Bertsekas e Tsitsiklis (2008), um processo estocástico consiste em uma coleção de variáveis aleatórias definidas sobre um mesmo espaço de probabilidade, permitindo a análise probabilística de sua evolução temporal. Uma cadeia de Markov é um exemplo de um processo estocástico.

3.12 Espaço de estados

Definição 21 (Espaço de estados)**.** O espaço de estados S é o conjunto de todos os valores possíveis que a variável aleatória $X(t)$ pode assumir. No caso das cadeias de Markov estudadas neste trabalho, considera-se S um conjunto finito ou enumerável.

3.13 Matriz estocástica

Definição 22 (Matriz estocástica)**.** Seja I um conjunto finito ou enumerável. Uma matriz $P = (P_{ij})_{i,j \in I}$ é chamada de matriz estocástica (ou matriz de transição) se satisfaz:

i. $P_{ij} \in [0, 1]$, para todo $i, j \in I$;

ii. $\sum_{j \in I} P_{ij} = 1$, para todo $i \in I$.

Então, P é chamada também de matriz de transição.

Exemplo 3.17. Suponha que o tempo possa estar em 3 estados:

i. Ensolarado (S);

ii. Nublado (N);

iii. Chuvoso (C).

A evolução do clima segue a seguinte matriz estocástica ou matriz de transição:

$$P = \begin{pmatrix} 0,8 & 0,15 & 0,05 \\ 0,2 & 0,6 & 0,2 \\ 0,4 & 0,25 & 0,35 \end{pmatrix}.$$

Cada linha da matriz representa um estado atual e cada coluna representa um estado futuro. Repare que a soma das entradas das linhas é igual a 1, é o que a propriedade *ii* da definição 22 está dizendo. Por exemplo, para saber qual a probabilidade do tempo ficar nublado dado que esta ensolarado, denotado por $\mathbb{P}(X_1 = N|X_0 = S)$, basta buscar pelo i e j da matriz que corresponde a esses estados, então:

$$\mathbb{P}(X_1 = N|X_0 = S) = 0,15.$$

Para ficar mais didático, observe a matriz a seguir:

$$P = \begin{pmatrix} P_{SS} & P_{SN} & P_{SC} \\ P_{NS} & P_{NN} & P_{NC} \\ P_{CS} & P_{CN} & P_{CC} \end{pmatrix}.$$

Então, para encontrar a probabilidade do tempo ficar nublado dado que está ensolarado, basta buscar a entrada da matriz que representa $\mathbb{P}(X_1 = N|X_0 = S) = P_{SN} = 0,15$.

Cada linha da matriz estocástica é chamado de vetor de probabilidade e a soma das suas entradas sempre é igual a 1. A matriz estocástica é de grande importância para as cadeias de Markov pois, por meio delas, é possível analisar a transição de um estado para o outro. Um fato interessante é que as entradas da matriz estocástica representam as probabilidades de transição de um estado para o outro. Essa matriz é chamada de matriz de transição P.

4 Cadeias de Markov

Uma cadeia de Markov é um processo estocástico que é conhecido pela sua propriedade fundamental, propriedade de Markov, que nos diz que o estado seguinte só precisa do estado anterior e não das condições iniciais. Em outras palavras, o que caracteriza uma cadeia de Markov é a perda de memória.

4.1 Cadeias de Markov

Definição 23 (Cadeias de Markov)**.** Seja a tripla $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade. Uma sequência de variáveis aleatórias $(X_0, X_1, X_2, ...)$ é uma cadeia de Markov com o espaço de estados Ω e matriz de transição P se, $\forall x, y \in \Omega, t \geq 1$, e todos os eventos $H_{t-1} = \bigcap_{s=0}^{t-1} \{X_s = x_s\}$ satisfazendo $\mathbb{P}(H_{t-1} \cap \{X_t = x\}) > 0$, temos:

$$\mathbb{P}(X_{t+1} = y \mid H_{t-1} \cap \{X_t = x\}) = \mathbb{P}(X_{t+1} = y \mid X_t = x).$$

Isto é, os estados anteriores não importam, só o estado presente, o que é conhecido como propriedade de Markov.

Imagine uma partícula se movendo no espaço. Esse movimento é um fenômeno de origem probabilística, chamado de passeio aleatório. Um passeio aleatório é um exemplo de uma cadeia de Markov e será muito bem explorado no exemplo 4.4.

Exemplo 4.1 (Exemplo do sapo e das duas vitórias-régias)**.** Considere um sistema composto por um sapo e duas vitórias-régias, denotadas pelos estados Leste (**l**) e Oeste (**o**), que compõem o espaço de estados finito $\Omega = \{\mathbf{l}, \mathbf{o}\}$ e $(X_0, X_1, ...)$ a sequência de vitórias-régias que o sapo utiliza ao longo da semana. A dinâmica de transição do sistema é regida por um mecanismo probabilístico, muito parecido com o lançamento de moedas. A cada passo de tempo discreto n, a evolução do sistema ocorre sob as seguintes condições:

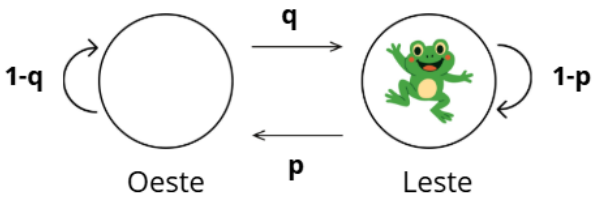
A partir do estado e:

A probabilidade do sapo saltar para o estado w no tempo $n + 1$ é p, enquanto a probabilidade de permanecer em e é $1 - p$.

A partir do estado w:

A probabilidade de transição para o estado e é q, e a de permanência no estado w é $1 - q$. Sendo X_n a variável aleatória que indica a posição do sapo no tempo n, observamos que a probabilidade de X_{n+1} depende exclusivamente do valor de X_n, satisfazendo a propriedade de Markov.

Figura 9 – Diagrama das vitórias-régias



Fonte: Próprio autor (2025).

Essas probabilidades podem ser organizadas na matriz de transição estocástica P:

$$P = \begin{pmatrix} P_{\mathbf{l},\mathbf{l}} & P_{\mathbf{l},\mathbf{o}} \\ P_{\mathbf{o},\mathbf{l}} & P_{\mathbf{o},\mathbf{o}} \end{pmatrix} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix},$$

onde cada entrada P_{ij} representa $\mathbb{P}(X_{n+1} = j \mid X_n = i)$, com $i, j \in \{e, w\}$. Note que a soma de cada linha da matriz é unitária, uma característica fundamental de matrizes estocásticas.

Suponha que na segunda-feira, estando na vitória-régia leste, o sapo jogue a moeda para decidir se muda de vitória-régia ou não. Com probabilidade p ele vai para o oeste e com probabilidade $1-p$ continua no leste. Temos que:

$$\mathbb{P}(X_1 = \mathbf{o} \mid X_0 = \mathbf{l}) = p \quad \text{e} \quad \mathbb{P}(X_1 = \mathbf{l} \mid X_0 = \mathbf{l}) = 1 - p.$$

O que pode acontecer na terça-feira? Ao considerar as duas possibilidades de X_1, obtemos:

$$\mathbb{P}(X_2 = \mathbf{l} \mid X_0 = \mathbf{l}) = (1-p) \cdot (1-p) + p \cdot q$$

e

$$\mathbb{P}(X_2 = \mathbf{o} \mid X_0 = \mathbf{l}) = (1-p) \cdot p + p \cdot (1-q).$$

Isso pode continuar indefinidamente, mas podemos escrever a distribuição como um vetor:

$$\mu_t := (\mathbb{P}(X_t = \mathbf{l} \mid X_0 = \mathbf{o}), \mathbb{P}(X_t = \mathbf{o} \mid X_0 = \mathbf{l})).$$

Podemos pensar o seguinte. Digamos que o sapo no domingo está na vitória-régia do leste. Essa hipótese pode ser escrita como:

$$\mu_0 = (1, 0).$$

Então, a posição do sapo na segunda-feira é dada por:

$$\mu_1 = \mu_0 \cdot P, \quad \text{sendo} \quad P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}.$$

Com essa multiplicação, é possível encontrar a distribuição para o passo μ_t seguinte. Então,

$$\mu_t = \mu_{t-1} \cdot P, \forall t \geq 1.$$

Finalmente, para qualquer distribuição μ_0, obtemos:

$$\mu_t = \mu_0 \cdot P^t, \forall t \geq 0.$$

Um objetivo central no estudo das Cadeias de Markov é determinar o comportamento do sistema após um longo período de tempo. Buscamos um vetor de probabilidade $\pi = (\pi_1, \pi_2)$ que seja invariante pelo sistema, ou seja, que satisfaça a equação de equilíbrio:

$$\pi P = \pi.$$

Para este modelo de dois estados, o sistema de equações lineares resultante

$$(\pi_1(1-p) + \pi_2 q = \pi_1 \text{ e } \pi_1 p + \pi_2(1-q) = \pi_2),$$

sob a condição de normalização $\pi_1 + \pi_2 = 1$, fornece a seguinte distribuição estacionária:

$$\pi_1 = \frac{q}{p+q}, \quad \pi_2 = \frac{p}{p+q}.$$

Uma abordagem mais detalhada deste exemplo é encontrada em Levin e Peres (2017), ou no relatório de iniciação científica de Cardoso (2020). Este resultado demonstra que, no longo prazo, a proporção de tempo que o agente passa em cada estado é inversamente proporcional à probabilidade de saída daquele estado. Este exemplo simples serve como base para a análise de cadeias mais complexas.

Exemplo 4.2 (Passeio aleatório simples)**.** Sejam $X_1, X_2, ...$ variáveis aleatórias independentes e identicamente distribuídas (i.i.d), tais que:

$$\begin{cases} \mathbb{P}(X_i = 1) = \frac{1}{2} \\ \mathbb{P}(X_i = -1) = \frac{1}{2}. \end{cases}$$

O passeio aleatório simples em $\mathbb{Z}$ é definido como:

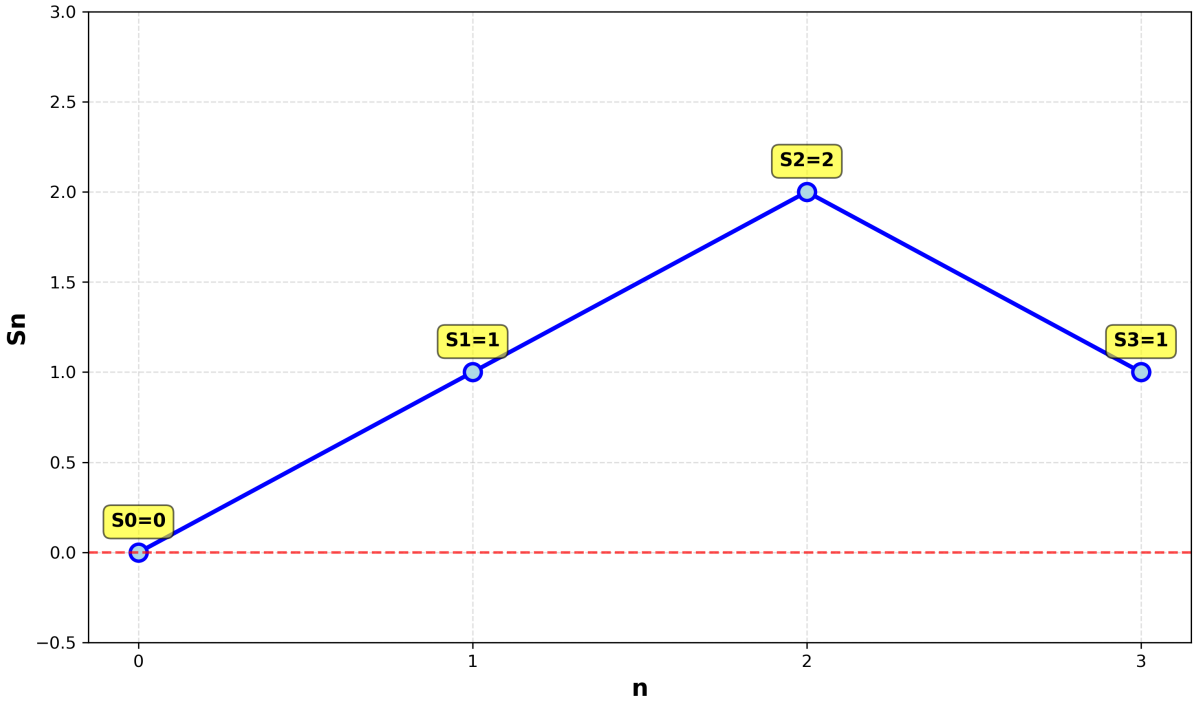
$$S_n = \sum_{k=1}^{n} X_k, n \geq 1, \quad \text{com} \quad S_0 = 0.$$

Fixando os valores $X_1 = 1$, $X_2 = 1$, $X_3 = -1$, obtemos os seguintes valores para o passeio, representados na Figura 10:

$$S_1 = X_1 = 1, \quad S_2 = \sum_{k=1}^{2} X_k = 1 + 1 = 2, \quad S_3 = \sum_{k=1}^{3} X_k = 1 + 1 - 1 = 1.$$

O passeio aleatório simples $\{S_j\}_{j\geq 0}$, definido por $S_j = \sum_{i=1}^{j} X_i$, onde $\{X_i\}_{i\in\mathbb{N}}$ são v.a. i.i.d. com valores em $\{-1, +1\}$, é uma cadeia de Markov.

Figura 10 – Passeio aleatório simples



Fonte: Próprio autor (2025).

Demonstração. Para demonstrar que $\{S_j, j \geq 0\}$ é uma cadeia de Markov, devemos verificar se:

$$P(S_{j+1} = s_{j+1} \mid S_j = s_j, S_{j-1} = s_{j-1}, \ldots, S_0 = s_0) = P(S_{j+1} = s_{j+1} \mid S_j = s_j).$$

Considerando a definição do processo, podemos escrever $S_{j+1} = S_j + X_{j+1}$. Substituindo na probabilidade condicional do lado esquerdo:

$$P(S_j + X_{j+1} = s_{j+1} \mid S_j = s_j, S_{j-1} = s_{j-1}, \ldots, S_0 = s_0).$$

Dado que o evento condicionado estabelece que $S_j = s_j$, a expressão torna-se:

$$P(X_{j+1} = s_{j+1} - s_j \mid S_j = s_j, S_{j-1} = s_{j-1}, \ldots, S_0 = s_0).$$

Como X_{j+1} é independente de $X_1, X_2, \ldots, X_j$, ele é consequentemente independente de $S_0, S_1, \ldots, S_j$ (pois estes são funções de X_i para $i \leq j$). Assim, a probabilidade condicional simplifica-se para a probabilidade marginal:

$$P(X_{j+1} = s_{j+1} - s_j).$$

Por outro lado, calculando o termo da direita da propriedade de Markov:

$$P(S_{j+1} = s_{j+1} \mid S_j = s_j) = P(S_j + X_{j+1} = s_{j+1} \mid S_j = s_j) = P(X_{j+1} = s_{j+1} - s_j).$$

Como ambos os lados resultam em $P(X_{j+1} = s_{j+1} - s_j)$, a igualdade é satisfeita e o processo é, por definição, uma cadeia de Markov. □

Portanto, é possível observar que o passeio aleatório é uma cadeia de Markov, pois é um processo estocástico e preserva a propriedade de Markov: a perda de memória.

Exemplo 4.3 (Passeios aleatórios em duas dimensões)**.** Sejam $X_1, X_2, ...$ vetores aleatórios independentes e identicamente distribuídos (i.i.d), para todo $i \in \mathbb{N}$:

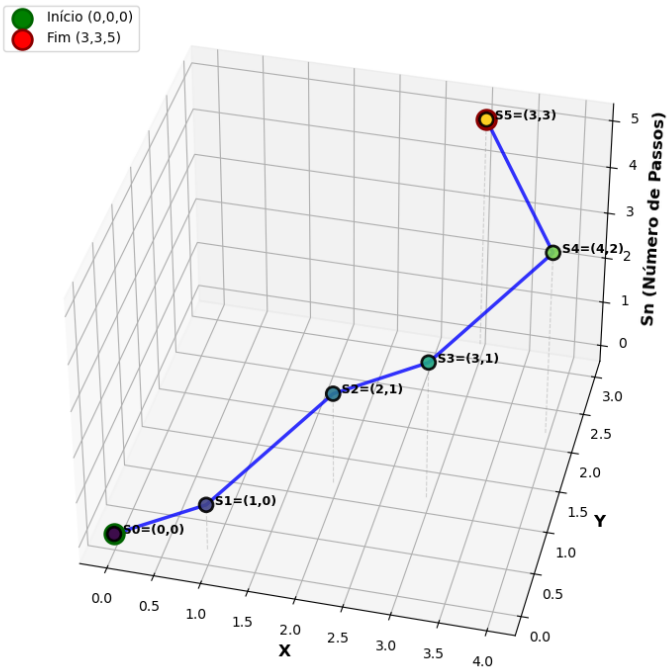
$$\begin{cases} \mathbb{P}(X_i = (1,0)) = \frac{1}{4} \\ \mathbb{P}(X_i = (-1,0)) = \frac{1}{4} \\ \mathbb{P}(X_i = (0,1)) = \frac{1}{4} \\ \mathbb{P}(X_i = (0,-1) = \frac{1}{4} \end{cases}.$$

Em palavras, cada X_i é um par ordenado (x, y), com $(x, y) \in \mathbb{Z}^2$. Então, o passeio aleatório em duas dimensões é o passeio aleatório no plano, definido como:

$$S_n = \sum_{i=1}^{n} X_i, \quad \text{com} \quad S_0 = (0,0).$$

Fixando $X_1 = (1,0)$, $X_2 = (1,1)$, $X_3 = (1,0)$, $X_4 = (1,1)$ e $X_5 = (-1,1)$, obtemos o gráfico 11 a seguir:

Figura 11 – Passeio aleatório em duas dimensões



Fonte: Próprio autor (2025).

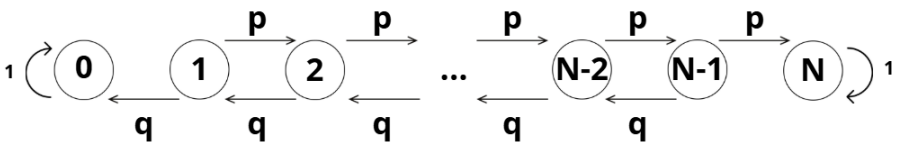
Os passeios aleatórios constituem um modelo eficaz para descrever o movimento de partículas ao longo do tempo, com aplicações relevantes na física.

4.2 A ruína do jogador

Exemplo 4.4. Considere um jogador apostando baseado no resultado de uma sequência de lançamentos independentes de uma moeda justa. Se o resultado do lançamento for cara, ele ganha um real; se for coroa, ele perde um real. Se ele conseguir N reais ou ficar sem dinheiro, ele para de jogar.

A situação do apostador pode ser modelada por um passeio aleatório num caminho de vértices $\{0, 1, 2, ..., N\}$. Em todos esses vértices, o passeio é igualmente provável de subir ou descer uma unidade (moeda justa: $p = q = 1/2$), como no diagrama abaixo.

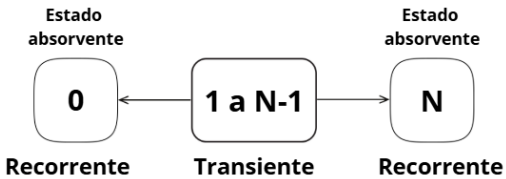
Figura 12 – Diagrama da ruína do jogador



Fonte: Timpanaro (2021).

Nesse exemplo, o processo possui dois estados absorventes: o estado 0, no qual o jogador não pode mais apostar, e o estado N, que representa o término do jogo ao se atingir o capital máximo. Os estados intermediários, de 1 a $N - 1$, são transientes, pois o jogador pode transitar entre eles livremente. O diagrama de estados correspondente é mostrado na Figura 13.

Figura 13 – Diagrama de classes do problema da ruína do jogador



Fonte: Timpanaro (2021).

A partir disso, surgem duas questões: Quanto tempo irá demorar para o apostador alcançar um dos seus dois destinos? Quais são as suas probabilidades? As respostas encontram-se no seguinte enunciado:

Proposição 4.2.1. Suponha que um jogador que faz apostas justas em lançamentos de moedas, abandonará o jogo quando sua fortuna cair para 0 ou subir para N. Sejam X_t a fortuna do apostador no tempo t e τ o tempo necessário para alcançar o estado 0 ou N. Suponha que $X_0 = j$, onde $0 \leq j \leq N$. Então

$$\mathbb{P}(X_\tau = N | X_0 = j) = \mathbb{P}_j(X_\tau = N) = \frac{j}{N}$$

e

$$\mathbb{E}_j(\tau) = j(N - j).$$

Demonstração. Sejam j, $0 \leq j \leq N$, a fortuna inicial do jogador e τ o tempo necessário para alcançar o estado 0 ou N, ou seja:

$$\tau := \min\{t \geq 0 : X_t = N \quad \text{ou} \quad X_t = 0\}.$$

Além disso, seja p_j a probabilidade do apostador alcançar a fortuna N antes de falir, dado que ele começou com j reais, $j \in \{0, 1, \ldots, N\}$. Claramente, $p_0 = 0$ e $p_N = 1$, enquanto que:

$$p_j = \mathbb{P}(X_\tau = N \mid X_0 = j). \tag{1}$$

Condicionando (1) às possibilidades de X_1 e usando o teorema da Probabilidade Total 2.2, obtemos:

$$\begin{aligned} p_j =& \mathbb{P}(X_\tau = N \mid X_0 = j, X_1 = j - 1) \cdot \mathbb{P}(X_1 = j - 1 \mid X_0 = j) \\ &+ \mathbb{P}(X_\tau = N \mid X_0 = j, X_1 = j + 1) \cdot \mathbb{P}(X_1 = j + 1 \mid X_0 = j). \end{aligned}$$

Sejam $p := \mathbb{P}(X_1 = j + 1 \mid X_0 = j)$ e $q := \mathbb{P}(X_1 = j - 1 \mid X_0 = j)$. Como já temos nosso estado atual atualizado para $X_1 = j - 1$, não precisamos da informação $X_0 = j$ anterior; logo, podemos descartar essa informação na equação (propriedade de Markov), restando apenas:

$$\mathbb{P}(X_\tau = N \mid X_1 = j - 1) \cdot \mathbb{P}(X_1 = j - 1 \mid X_0 = j) = p_{j-1} \cdot q.$$

Da mesma forma, para $X_1 = j + 1$:

$$\mathbb{P}(X_\tau = N \mid X_1 = j + 1) \cdot \mathbb{P}(X_1 = j + 1, X_0 = j) = p_{j+1} \cdot p.$$

E, assim, obtemos:

$$\begin{cases} p_j = p_{j-1} \cdot q + p_{j+1} \cdot p, \quad \text{para todo } j \in \{1, \ldots, N - 1\} \\ p_0 = 0 \\ p_N = 1. \end{cases} \tag{2}$$

Agora, veja que p_j pode ser escrito como $p \cdot p_j + q \cdot p_j$, sendo $p + q = 1$. Logo, podemos reescrever $p_j = p \cdot p_{j+1} + q \cdot p_{j-1}$ como $p \cdot p_j + q \cdot p_j = p \cdot p_{j+1} + q \cdot p_{j-1}$. Organizando, obtemos:

$$\begin{aligned} p \cdot p_{j+1} - p \cdot p_j = q \cdot p_j - q \cdot p_{j-1} &\Rightarrow p(p_{j+1} - p_j) = q(p_j - p_{j-1}) \\ &\Rightarrow p_{j+1} - p_j = \frac{q}{p}(p_j - p_{j-1}) \\ &\Rightarrow \Delta_{j+1} = \frac{q}{p}\Delta_j. \end{aligned}$$

Obtemos, portanto, um sistema de recorrência. Para $j = 1$, temos:

$$\Delta_2 = \frac{q}{p}\Delta_1.$$

Para $j = 2$:

$$\Delta_3 = \frac{q}{p}\Delta_2.$$

Usando que $\Delta_2 = \frac{q}{p}\Delta_1$:

$$\Delta_3 = \frac{q}{p} \cdot \frac{q}{p}\Delta_1 = \left(\frac{q}{p}\right)^2 \Delta_1,$$

e assim por diante. Logo, por indução, obtemos que:

$$\Delta_{n+1} = \left(\frac{q}{p}\right)^n \Delta_1 \quad \text{para todo } n \in \mathbb{N}.$$

Agora, como $p_0 = 0$ e $p_N = 1$, temos:

$$\Delta_1 = p_1 - p_0 \Rightarrow p_1 = \Delta_1.$$

$$\Delta_2 = p_2 - p_1 \Rightarrow p_2 = \Delta_1 + \Delta_2.$$

$$\Delta_3 = p_3 - p_2 \Rightarrow p_3 = p_2 + \Delta_3 \Rightarrow= p_3 = \Delta_1 + \Delta_2 + \Delta_3.$$

Logo,

$$p_j = \sum_{k=1}^{j} \Delta_k \quad \text{para todo } j \in \mathbb{N}.$$

Por fim, ao substituir Δ_k na soma, temos:

$$P_j = \sum_{k=1}^{j} \left[\left(\frac{q}{p}\right)^{k-1} \Delta_1 \right].$$

Como Δ_1 é uma constante em relação ao índice k, podemos colocá-lo para fora do somatório:

$$p_j = \Delta_1 \sum_{k=1}^{j} \left(\frac{q}{p}\right)^{k-1}.$$

Para facilitar a visualização da progressão geométrica (PG), podemos fazer uma mudança de índice. Seja $m = k - 1$: quando $k = 1$, $m = 0$; quando $k = j$, $m = j - 1$. Assim:

$$p_j = \Delta_1 \sum_{m=0}^{j-1} \left(\frac{q}{p}\right)^m,$$

que é uma série geométrica de razão $r = \frac{q}{p}$. A fórmula para a soma de uma PG finita com n termos é dada por:

$$S_n = \Delta_1 \cdot \frac{1 - r^n}{1 - r}.$$

Em nosso problema, temos j termos, que vai de 0 até $j - 1$. Portanto:

$$\sum_{m=0}^{j-1} \left(\frac{q}{p}\right)^m = \frac{1 - (\frac{q}{p})^j}{1 - (\frac{q}{p})},$$

substituindo a soma da PG de volta na equação de p_j, concluimos que:

$$p_j = \Delta_1 \cdot \frac{1 - (\frac{q}{p})^j}{1 - (\frac{q}{p})}. \tag{3}$$

Usando a condição de contorno $p_N = 1$:

$$p_N = 1 = \Delta_1 \cdot \frac{1 - (\frac{q}{p})^N}{1 - (\frac{q}{p})} \Rightarrow \Delta_1 = \frac{1 - (\frac{q}{p})}{1 - (\frac{q}{p})^N}.$$

substituindo em (3), obtemos:

$$p_j = \frac{1 - (\frac{q}{p})^j}{1 - (\frac{q}{p})^N}.$$

Seja $x := \frac{q}{p}$, tomando o limite $x \to 1$ (considerando a moeda honesta) e aplicando L´Hospital, obtemos finalmente

$$\lim_{x \to 1} \frac{1 - x^j}{1 - x^N} = \frac{j}{N}.$$

Ou seja:

$$p_j = \begin{cases} \dfrac{1 - \left(\frac{q}{p}\right)^j}{1 - \left(\frac{q}{p}\right)^N} & \text{se } p \neq q \neq 1/2, \\ \frac{j}{N} & \text{se } p = q = 1/2, \end{cases} \tag{4}$$

com $p + q = 1$. E, assim, obtemos a probabilidade de chegar a N saindo de um j, antes de falir.

Agora veremos quanto tempo demora para isso acontecer. O tempo de chegada é definido como:

$$\tau \coloneqq \min\{t \geq 0 : X_t = N \quad \text{ou} \quad X_t = 0\}.$$

E o seu valor esperado:

$$E_j = \mathbb{E}[X_\tau = N \mid X_0 = j].$$

Novamente, condicionando nas possibilidades para X_1 e utilizando a propriedade de Markov, obtemos:

$$\begin{cases} \mathbb{E}[X_\tau = N \mid X_1 = j+1] = E_{j+1} + 1 & \text{com probabilidade } p \\ \mathbb{E}[X_\tau = N \mid X_1 = j-1] = E_{j-1} + 1 & \text{com probabilidade } q \end{cases}.$$

Por quê? Quando o primeiro passo aumenta a fortuna do apostador (com probabilidade p de avançar), então a esperança condicional para τ é 1 (para o passo inicial), mais o tempo esperado (esperança) adicional. A esperança adicional esperada é E_{j+1}, porque agora o apostador encontra-se no estado $j+1$. Raciocínio equivalente para quando a fortuna diminui.

Agora precisamos avaliar as condições de contorno. A esperança de chegar em zero ou N dado que ele já esta em zero será $E_0 = E_N = 0$. A solução será obtida de forma semelhante ao caso da probabilidade. Temos que:

$$E_j = p \cdot (E_{j+1} + 1) + q \cdot (E_{j-1} + 1) \Rightarrow E_j = p \cdot E_{j+1} + q \cdot E_{j-1} + 1.$$

Agora precisamos resolver a relação de recorrência. Seja $\Delta_j := E_j - E_{j-1},\ 1 \leq j \leq N-1$. Temos que, para $p = q = 1/2$:

$$\frac{E_j}{2} + \frac{E_j}{2} = \frac{E_{j+1}}{2} + \frac{E_{j-1}}{2} + 1 \Rightarrow$$

$$\frac{\Delta_j}{2} = \frac{\Delta_{j+1}}{2} + 1 \Rightarrow$$

$$\Delta_{j+1} = \Delta_j - 2.$$

Podemos avaliar agora os valores para $j = 1, 2, 3, ...$ então: Para $j = 1$. Temos:

$$\Delta_2 = \Delta_1 - 2.$$

Para $j = 2$:

$$\Delta_3 = \Delta_2 - 2 \Rightarrow \Delta_3 = \Delta_1 - 2 - 2 = \Delta_1 - 4.$$

Para $j = 3$:

$$\Delta_4 = \Delta_3 - 2 \Rightarrow \Delta_4 = \Delta_1 - 4 - 2 = \Delta_1 - 6.$$

E assim sucessivamente. Portanto, $\Delta_{j+1} = \Delta_1 - 2j$. Novamente, substituindo as condições de contorno, obtemos:

$$\Delta_1 = E_1 - E_0 \Rightarrow E_1 = \Delta_1$$

$$\Delta_2 = E_2 - E_1 \Rightarrow E_2 = E_1 + \Delta_2 = \Delta_1 + \Delta_2$$

$$\Delta_3 = E_3 - E_2 \Rightarrow E_3 = E_2 + \Delta_3 = \Delta_1 + \Delta_2 + \Delta_3$$

Logo,

$$E_j = \sum_{k=1}^{j} \Delta_k = \sum_{k=1}^{j} \Delta_1 - (2k-2) = j \cdot \Delta_1 - 2 \cdot \sum_{k=0}^{j-1} k.$$

Essa é a soma dos termos de uma PA. Então:

$$j \cdot \Delta_1 - 2 \cdot \sum_{k=0}^{j-1} k \Rightarrow E_j = j \cdot \Delta_1 - j(j-1)$$

para a condição $E_n = 0$ temos:

$$0 = E_N = N \cdot \Delta_1 - N(N-1) \Rightarrow$$

$$N \cdot \Delta_1 - N(N-1) = 0 \Rightarrow$$

$$\Delta_1 = \frac{N(N-1)}{N} \Rightarrow \Delta_1 = N - 1$$

substituindo Δ_1 em $E_j = j \cdot \Delta_1 - j(j-1)$ temos

$$E_j = j \cdot (N-1) - j(j-1) \Rightarrow$$

$$E_j = j \cdot N - j - j^2 + j = j \cdot N - j^2$$

logo

$$E_j = j(N-j)$$

e essa equação representa o tempo esperado para sair de j até N. □

Exemplo 4.5. Qual a probabilidade do jogador chegar em $N = 20$, saindo de uma posição inicial $j = 4$?

$$p_j = \frac{j}{N}$$

$$p_4 = \frac{4}{20} = \frac{1}{5} = 0,2$$

Ou seja, $p_4 = 20\%$.

Exemplo 4.6. Qual o tempo esperado para sair do estado $j = 4$ para o estado $N = 20$? Temos que:

$$E_j = j(N-j),$$

para $j = 4$, obtemos $E_4 = 4 \cdot (20-4) = 64$.

5 Conclusões e Perspectivas

O trabalho se propôs a investigar e apresentar as bases da probabilidade até ter material suficiente para apresentar as cadeias de Markov, culminando no problema da ruína do jogador. Um outro objetivo foi desenvolver um material utilizando uma linguagem acessível, sem perder o rigor matemático, para que o futuro estudante de probabilidade tenha acesso a um material de qualidade. Dados esses objetivos, é possível observar pela estrutura lógica e rigor apresentados que o trabalho concluiu com sucesso esses objetivos.

As limitações do trabalho estão no fato de não ter entrado mais a fundo na teoria da medida, pois até o momento da publicação do mesmo, o autor não cursou essa componente curricular.

Uma possível expansão deste trabalho seria unir as teorias de passeios aleatórios com mecânica quântica e investigar os passeios aleatórios quânticos, que também é possível modelar utilizando cadeias de Markov.

Esta investigação é justificada pelo fato da tecnologia de computação cientifica estar caminhando para os computadores quânticos, pois os mesmos conseguem simular problemas da física, da química, da medicina, da farmacologia entre outras áreas importantes para a manutenção da vida humana.

Os computadores quânticos conseguem simular, por exemplo, as interações de um novo fármaco com moléculas biológicas de forma mais eficiente e em um tempo hábil, coisa que os computadores clássicos demorariam muitas vezes a idade do universo para concluir as simulações.

Referências

BERTSEKAS, Dimitri; TSITSIKLIS, John N. **Introduction to Probability**. Nashua: Athena Scientific, 2008. v. 1.

BRASIL. Tribunal Regional do Trabalho da 7a Região (Ceará). **Concurso Público para provimento de cargos de Analista Judiciário - Área Apoio Especializado**: Especialidade estatística. Fortaleza: TRT 7ª Região, 2009. Caderno de Prova 'M', Tipo 001. Organizado pela Fundação Carlos Chagas. Referência: Caderno de Prova 'M', Tipo 001 [cite: 3, 10].

CARDOSO, Ruan. **Cadeias de Markov e mixing times**: Relatório final de projeto de iniciação científica (CNPq/UFBA). Salvador, 2020. 19 p.

CARVALHO, Paulo Cezar Pinto et al. **Análise Combinatória e Probabilidade**. 11. ed. Rio de Janeiro: SBM, 2020.

DURRETT, Rick. **Probability**: Theory and examples. Cambridge: Cambridge University Press, 2010.

FRANCO, Tertuliano. **Princípios de Combinatória e Probabilidade**. Rio de Janeiro: IMPA, 2020.

IMPA - Instituto de Matemática Pura e Aplicada. **Programa de Mestrado**: Cadeia de markov (2025). IMPA, 2025. Playlist de vídeos. Disponível em: http://www.youtube.com/playlist?list=PLo4jXE-LdDTQQ64mRYr_Sop4E--Jnodae.

JAMES, Barry R. **Probabilidade**: Um curso em nível intermediário. Rio de Janeiro: IMPA, 2023.

LEVIN, David A.; PERES, Yuval. **Markov Chains and Mixing Times**. 2nd edition. ed. Providence, Rhode Island: American Mathematical Society, 2017. v. 107.

LIMA, Elon Lages. **Curso de Análise**. 15. ed. Rio de Janeiro: IMPA, 2022. v. 1.

POSSANI, Claudio. **Valor Esperado**. [S.l.]: Professor Possani, 2022. YouTube. Canal: Professor Possani. Disponível em: <https://www.youtube.com/watch?v=bm5PJBlY9Ns>. Acesso em: 1 fev. 2026.

ROLLA, Leonardo T.; LIMA, Bernardo N. B. de. **Probabilidade**. [S. l.]: Edição dos autores, 2025. Versão de 18 de março de 2025. Disponível em: https://www.ime.usp.br/~leorolla/probabilidade/. Acesso em: 1 fev. 2026.

ROSS, Sheldon M. **A First Course in Probability**. 10th edition. ed. Upper Saddle River: Pearson, 2019.

TIMPANARO, André. **Aula 7.1 - O modelo da ruína**. [S. l.]: [s. n.], 2021. 1 vídeo (32 min). Publicado pelo canal André Timpanaro Aulas. Disponível em: http://www.youtube.com/watch?v=l7HIr6NmItI. Acesso em: 13 jan. 2026.