UNIVERSIDADE DE BRASÍLIA
Instituto de Ciências Exatas
Departamento de Matemática
Mestrado Profissional em Matemática em Rede Nacional

Matrizes de Markov: o Teorema de Perron-Frobenius; PageRank e outras aplicações

por
Lázaro Sousa Pereira

Brasília-DF
2019

UNIVERSIDADE DE BRASÍLIA
Instituto de Ciências Exatas
Departamento de Matemática
Mestrado Profissional em Matemática em Rede Nacional

Lázaro Sousa Pereira

Matrizes de Markov: o Teorema de Perron-Frobenius; PageRank e outras aplicações

Dissertação de Mestrado apresentada ao Departamento de Matemática da Universidade Brasília como requisito parcial do Programa de Mestrado Profissional em Matemática em Rede Nacional - PROFMAT, para obtenção do título de Mestre em Matemática.

Orientador: Prof. Dr. Theo Allan Darn Zapata

Brasília-DF
2019

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

SSO725m Sousa Pereira, Lázaro
Matrizes de Markov: o Teorema de Perron-Frobenius; PageRank e outras aplicações. / Lázaro Sousa Pereira; orientador Theo Allan Darn Zapata. -- Brasília, 2019.
68 p.

Dissertação (Mestrado - Mestrado Profissional em Matemática) -- Universidade de Brasília, 2019.

1. Matrizes de Markov. 2. Matrizes Irredutíveis. 3. Teorema de Perron-Frobenius. 4. PageRank. 5. Modelo de Difusão de Ehrenfest. I. Allan Darn Zapata, Theo, orient. II. Título.

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Matemática

Matrizes de Markov: o teorema de Perron-Frobenius; PageRank e outras aplicações.

por

LÁZARO SOUSA PEREIRA

Dissertação apresentada ao Departamento de Matemática da Universidade de Brasília, como parte dos requisitos "Programa" de Mestrado Profissional em Matemática em Rede Nacional – PROFMAT, para obtenção do grau de

MESTRE EM MATEMÁTICA

Brasília, 19 de julho de 2019.

Comissão Examinadora:

Prof. Theo Allan Darn Zapata (Orientador)

Prof. Mauro Ribeiro de Oliveira Júnior - CEF/Unieuro

Prof. Ary Vasconcelos Medino - MAT/UnB

Com amor, à minha esposa e aos meus filhos; aos meus pais, irmão e familiares.
Com respeito, a cada colega professor e professora da educação pública do nosso país; em especial, à minha colega e irmã D. Marllene (Profª Allene Martins Rezende).
Em memória, aos meus avós: João e Regino; ao colega e amigo Prof. Romeu(Antonio Vidal).
Até algum dia!

Agradecimentos

As minhas experiências de vida revelam que devo ser grato até mesmo, por ter pelo que agradecer. A realização desse curso e desse trabalho tornou-se possível graças a Deus, a todo plano espiritual e, a tolerância e a cumplicidade direta ou indireta de algumas pessoas. Registro então, meus sinceros agradecimentos a todas, ainda que eu não as mencione nessas linhas:

A todos os professores do MAT/UnB que empreenderam esforços no PROFMAT. Pelas aulas, pelo tempo dedicado a cada conversa e a cada atendimento além das aulas. Em especial aos membros da banca examinadora, o Prof. Dr. Ary Vasconcelos Medino pela apreciação crítica e valiosas sugestões e o Prof. Dr. Nilton Moura Barroso Neto pela leitura atenta e considerações. Ao Prof. Dr. Mauro Ribeiro de Oliveira Junior pela leitura crítica e sugestões. Suas contribuições foram inspiradoras e determinantes para esta importante e conclusiva etapa de minha formação.

Em particular, agradeço ao meu orientador Prof. Dr. Theo Allan Darn Zapata; ter sido seu aluno e tê-lo como orientador foi uma experiência muito significativa, pelo que aprendi, pela consciência da dimensão do que preciso aprender, e pelo sentido e valorização das buscas por entendimento, compreensão e aprendizagem significativa. Muito obrigado pela generosidade! O tenho também como inspiração e referência a partir de então.

A todos os colegas e amigos que encontrei durante os dois anos dessa jornada. Nossas discussões virtuais e presenciais foram muito importantes. Em particular, deixo meus agradecimentos aos colegas de grupo em várias aulas - Cesa Sabino, Marcelo Carvalho, Edeilson Cavalcante e Roosevelt Bessoni - e em muitas horas de estudos, especialmente nos fins de semana e feriados - Rubens Cardoso, Ricardo Pinto e Michel. Foi relevante compartilhar com vocês essa experiência.

A agência de fomento à pesquisa CAPES, pela bolsa concedida.

♡ Aos meus pais Antonia e Joel, pelo amor, pelo exemplo e pelas escolhas que fizeram em minha educação; ao meu irmão Leandro, pela torcida de sempre. Mesmo distantes física e geograficamente, os tenho sempre aqui!

♡ À minha esposa e companheira Emília Grazielle, pelo amor, pelo sacrifício e pela reiterada resiliência; aos meus filhos João Miguel e Bernardo pela paciência e companhia. Obrigado por jamais terem duvidado ou desacreditado, mesmo quando eu tive dúvidas.

O correr da vida embrulha tudo, a vida é assim: esquenta e esfria, aperta e daí afrouxa,
sossega e depois desinquieta. O que ela quer da gente é coragem.
João Guimarães Rosa, Grande Sertão: Veredas, (p.448), Nova Aguilar, 1ª ed. 1994

Resumo

Nesta dissertação de mestrado, discutiremos de forma básica a Teoria da Probabilidade e o cálculo de probabilidades além de algumas de suas propriedades; mostraremos e discutiremos as propriedades das Matrizes de Markov e apresentaremos uma demonstração do Teorema de Perron-Frobenius. Além disso daremos exemplos com aplicações no contexto da Teoria da Probabilidade.

Palavras-chave: Probabilidade, matrizes de Markov, matrizes irredutíveis, Perron-Frobenius, PageRank, modelo de Ehrenfest.

Abstract

In this master dissertation, we shall basically discuss Probability Theory and the calculation of probabilities in addition to some of its properties; we will show and discuss the properties of the Markov Matrices and present a demonstration of the Perron-Frobenius Theorem. At the end, we shall show you four examples with applications in the context of Probability Theory.

Key words: Probability, Markov matrices, irreducible matrices, Perron-Frobenius, PageRank, Ehrenfest model.

Sumário

Introdução

Estudar matemática em nosso país, ainda que não seja o desejo inicial de muitos dos estudantes de escolas públicas Brasil afora, vem sendo uma atividade cada vez mais incentivada pelas universidades em seus cursos de graduação e por instituições como o IMPA - Instituto Nacional de Matemática Pura e Aplicada. Como um professor de escola pública, a necessidade de atualização e aperfeiçoamento é companheira constante. Ainda que nossos incentivos e motivações sejam quase sempre pessoais, estudar matemática, compreender sua dimensão e relevância para tantos outros ramos das ciências, nos habilita continuamente a trabalhar em nossas salas de aula de forma mais segura e honesta. Na última etapa da educação básica - no ensino médio - onde todos os conceitos matemáticos anteriormente estudados precisam ser consolidados, ampliados e aprofundados, faz-se necessário uma abordagem contextualizada com a realidade e para isso, é natural trabalharmos com aplicações e modelos que deem sentido à teoria. Esse é um dos principais motivos dessa jornada: estudar e aprender.

Nesta dissertação é apresentada e discutida uma parte interessante, porém, pouco estudada em cursos de Álgebra Linear e Probabilidade em graduações de licenciatura Matemática: as matrizes de Markov. Esse tipo de matriz dá suporte para ramificações de estudos diversos da matemática, como na teoria da probabilidade e na teoria e representação de grupos, nas teorias de automorfismos de grupos livres [BH12], no estudo de sistemas dinâmicos e ergódicos [PY98] e de categorias tensoriais [EGNO15]. Sua principal característica é o fato de ser quadrada e ter cada vetor-coluna com entradas ≥ 0, e soma igual a 1. Esse tipo de vetor é também chamado de vetor de probabilidade, ou simplesmente, vetor estocástico. Essas matrizes resultam dos estudos do matemático russo Andrey Andreyevich Markov, que descreveu do ponto de vista do produto matricial, o cálculo de probabilidades de ocorrência de certos eventos que dependem apenas do estado em que o fenômeno se encontra para que seja calculada a probabilidade de estar num estado seguinte. Essa característica de não estar associada a resultados de uma memória mais extensa de repetições, é o que define uma Cadeia de Markov.

Utilizaremos ao longo de todo o trabalho a denominação *matrizes de Markov* para nos referirmos às tradicionalmente conhecidas, *matrizes de transição de probabilidades*, porém com uma interpretação distinta da tradicional, visto que, consideraremos os vetores-coluna como os vetores estocásticos. Com alguns recursos de Álgebra Linear referentes à propriedades e operações entre matrizes, obtemos o mesmo efeito do produto tradicionalmente conhecido.

Desde o início do programa de mestrado PROFMAT em 2011, foram publicadas até o momento mais de 4600 dissertações. A cadeia de Markov foi objeto de estudo e discussão em menos de duas dezenas de dissertações, sendo 10 delas, com foco numa discussão que compreendesse o contexto do ensino médio. Sendo um incentivo e um desafio a professores e estudantes desta etapa pois, a maioria dos recursos matemáticos necessários encontram-se nas obras e nos cursos de graduação e de pós-graduação. Destas dissertações, destaco: Silva, C. E. V. da. Aplicações da Álgebra Linear nas Cadeias de Markov, UFG - Universidade Federal de Goiás, 2013, que explana e propõe exemplos explorando a resolução de sistemas lineares; Oliveira, J. C. F. de, Noções de grafos dirigidos, cadeias de Markov e as buscas do Google, Universidade Federal de Sergipe, 2014, que propôs uma aplicação utilizando um exemplo com o PageRank com estudantes do ensino médio; Ribeiro, T. S. G. Processos de Markov discretos: exemplos voltados para o ensino médio, UNESP - Universidade Estadual Paulista Júlio de Mesquita Filho, 2017, que discorre sobre matrizes regulares e utiliza o Teorema de Perron-Frobenius numa versão que não utiliza recursos considerados avançados de Álgebra Linear, demonstrando-o para matrizes de ordem 2. Em todos esses trabalhos encontramos material pertinente acerca da cadeia de Markov e aplicações. Todo o acervo de dissertações pode ser consultado livremente no endereço http://www.profmat-sbm.org.br/dissertacoes/. Com o firme propósito de apresentar aspectos teóricos e aplicações das matrizes de Markov, o centro do nosso trabalho está no Teorema de Perron-frobenius com ênfase na ação das matrizes irredutíveis, e suas respectivas consequências. Para tanto, lançamos mão de recursos como autovalores e autovetores em sua demonstração e nas resoluções das aplicações que propomos.

No capítulo 1, apresentamos e discutimos de forma básica o cálculo de probabilidade, da probabilidade condicional, além de propriedades que tornam práticas as iterações na resolução de certos problemas, como o problema dos n mentirosos [Fel67]. Apresentamos ainda definições úteis no contexto das variáveis aleatórias e dos processos estocásticos; apresentamos e discutimos também, tópicos sobre matrizes, operações entre matrizes, autovalores e autovetores.

No capítulo 2 está o cerne da discussão acerca das matrizes de Markov. Nele apresentamos uma matriz de Markov e alguns resultados relevantes como por exemplo, a existência de um autovalor maximal igual a 1. Em seguida, definimos uma matriz irredutível, exibimos resultados e exemplos que a caracterizam e apresentamos uma demonstração do Teorema de Perron-Frobenius sugerida por Wielandt em sua obra sobre matrizes irredutíveis.

No capítulo 3, apresentamos e discutimos a resolução de quatro problemas onde aplicamos esse suporte teórico. Iniciamos com um problema que foi proposto na 23ª Olimpíada Brasileira de Matemática, onde deseja-se calcular a probabilidade de um ratinho estar numa certa gaiola, sob certas condições, após 23 sinais sonoros. Em seguida, discutimos a resolução do problema dos n mentirosos, onde desejamos calcular a probabilidade de, num grupo com n pessoas, uma determinada pessoa estar dizendo a verdade diante de um conjunto de afirmações feitas pelas demais. Na edição de janeiro / fevereiro de 2000 da Computing in Science and Engineering, Jack Dongarra e Francis Sullivan selecionaram 10 algoritmos com a maior influência em ciência, análise numérica e engenharia no século XX. Em março de 2016, Nick Higham (presidente do SIAM, 2017-2018) apresentou uma lista ligeiramente revisada. Em nenhuma ordem particular, a lista é: (1) métodos Newton e quasi-Newton; (2) fatorações matriciais (LU, Cholesky, QR); (3) decomposição do valor singular, algoritmos QR e QZ; (4) métodos de Monte-Carlo; (5) Transformada rápida de Fourier; (6) métodos do subespaço de Krylov; (7) JPEG; (8) PageRank; (9) método simplex; e (10) filtro de Kalman. No terceiro problema deste capítulo, calcularemos o PageRank para uma rede formada por sítios da web. Desde 1998, quando da publicação do resultado obtido por Brin e Page, o PageRank é assunto que vem sendo discutido e estudado em diversos institutos de pesquisa mundo afora. Destacamos o *Google Matrix: Fundamentals Applications and Beyond*, workshop organizado em outubro de 2018 pelo IHES - Institut des Hautes Études Scientifiques (Instituto de Altos Estudos Científicos), que é uma das principais referências - francesa e mundial - em estudos avançados de matemática, física teórica e ciências afins. Ao final, apresentamos e discutimos o modelo de difusão de Ehrenfest, onde deseja-se calcular a probabilidade de retorno de cada molécula para um recipiente ocupado por elas inicialmente. Este modelo foi estudado pelos físicos Tatyana e Paul Ehrenfest no início do século XX, e nessa versão básica, descreve a origem da irreversibilidade em sistemas físicos.

1 Preliminares

1.1 Noções Básicas de Probabilidade

Historicamente, a Teoria da Probabilidade surgiu a partir de problemas sobre a distribuição das apostas em jogos de azar. Seu desenvolvimento subsequente trouxe não só para a Matemática, mas, para a Estatística, as ciências da natureza e também sociais, uma poderosa ferramenta de investigação e análise que subsidiaram importantes resultados desde então. Nesta seção, vamos definir e apresentar elementos básicos e propriedades elementares dessa rica teoria. Ao leitor que desejar verificar uma exposição mais detalhada e com bastante exemplos ilustrativos, recomendamos a leitura dos capítulos 2, 3 e 5 de [MCCF06].

Definição 1.1.1. *Um experimento que não apresenta exatamente o mesmo resultado, ainda que repetido sob condições fixas, é denominado* ***experimento aleatório****. Do contrário, é denominado* ***experimento determinístico****.*

Definição 1.1.2. *Ao conjunto que coleciona todos os resultados possíveis de um experimento dá-se o nome de* ***espaço amostral****.*

Definição 1.1.3. *Denomina-se* ***evento****, todo resultado ou subconjunto de resultados de um experimento; quando o subconjunto é unitário, o evento é chamado de* ***evento simples****.*

Além disso, um espaço amostral é chamado de discreto se contiver um número finito de pontos ou se seus infinitos pontos podem ser postos em correspondência biunívoca com o conjunto dos números naturais. Nesse contexto, vamos denotar o espaço amostral por $\mathcal{S}$ e um evento qualquer por $\mathcal{E}$. Para uma sequência de eventos $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \ldots$, de $\mathcal{S}$, conforme seja finita ou infinita, temos

$$\bigcup_{i=1}^{n} \mathcal{E}_i = \mathcal{S} \quad \text{ou} \quad \bigcup_{i=1}^{\infty} \mathcal{E}_i = \mathcal{S}.$$

Convém ainda registrar que:

(i) a um ponto de $\mathcal{S}$ corresponde um, e tão somente um, resultado possível; e

(ii) cada resultado distinto corresponde a pontos distintos em $\mathcal{S}$.

Neste trabalho, consideraremos apenas espaços amostrais discretos.

Exemplo 1.1.1. *Uma linha de produção de uma industria metalúrgica, produz peças para automóveis que são sempre inspecionadas e classificadas como B (boa), ou D (defeituosa). Após uma inspeção, foram retiradas ao acaso, três dessas peças.*

Note que o espaço amostral associado a esse experimento é o conjunto

$$\mathcal{S} = \{BBB, DDD, DBB, BDB, BBD, DDB, DBD, BDD\}.$$

O evento $\mathcal{E}_1 = \{DBB, BDB, BBD\}$, por exemplo, reúne as possíveis amostras onde exatamente uma peça é classificada defeituosa; enquanto cada sequência, como por exemplo, $\{BDB\}$, é um evento simples.

Quando não mencionado de outra forma, tomaremos por referência um experimento com um número finito de elementos, e que cada um desses elementos têm a mesma chance de ocorrência, isto é, são equiprováveis. Temos, dessa forma, elementos suficientes para uma definição de probabilidade que atende nossas pretensões neste trabalho. Para uma explanação mais aprofundada ver [Jam15], e o clássico [Fel67].

Definição 1.1.4. *Dado um espaço amostral discreto $\mathcal{S}$ de N elementos, seja $\mathcal{E}$ um subconjunto de $\mathcal{S}$ composto por n elementos. Então, a probabilidade de $\mathcal{E}$, que denota-se por $P(\mathcal{E})$ é o número real não-negativo obtido a partir da razão:*

$$P(\mathcal{E}) = \frac{n}{N}.$$

Tradicionalmente, interpretamos esse número como *a divisão do número de casos favoráveis à ocorrência do evento $\mathcal{E}$ pelo número total de casos possíveis N*. Observamos também que se trata de uma função definida para uma classe dos eventos (subconjuntos) de $\mathcal{S}$.

De forma axiomática, para uma classe $\mathcal{C}$ de eventos de $\mathcal{S}$, a probabilidade satisfaz:

I. $P(A) \geq 0$ para todo $A \in \mathcal{C}$;

II. Se $A_1, \ldots, A_m$ é uma sequência de eventos disjuntos dois a dois, de $\mathcal{C}$ então,

$$P\left(\bigcup_{k=1}^{m} A_k\right) = \sum_{k=1}^{m} P(A_k);$$

III. $P(\mathcal{S}) = 1$.

A partir da definição e dos axiomas, verificamos que a probabilidade possui as propriedades:

(i) Para todo $\mathcal{E} \subset \mathcal{S}$, tem-se $P(\mathcal{E}) \geq 0$;

(ii) Sendo $\mathcal{E}_1$ e $\mathcal{E}_2$ eventos distintos de $\mathcal{S}$, tais que, $\mathcal{E}_1 \cap \mathcal{E}_2 = \emptyset$ então:

$$P(\mathcal{E}_1 \cup \mathcal{E}_2) = P(\mathcal{E}_1) + P(\mathcal{E}_2).$$

De modo geral, para $\mathcal{E}_1$ e $\mathcal{E}_2$ arbitrários, segue que,

$$P(\mathcal{E}_1 \cup \mathcal{E}_2) \leq P(\mathcal{E}_1) + P(\mathcal{E}_2);$$

(iii) Para $\mathcal{E}^c = \mathcal{S} - \mathcal{E}$, segue que $P(\mathcal{E}^c) = 1 - P(\mathcal{E})$.

De fato, para (i) basta verificarmos que como $N > 0$ e $n \geq 0$ temos o suficiente para $P(\mathcal{E}) \geq 0$. Supondo que $\mathcal{E}_1$ e $\mathcal{E}_2$ têm, respectivamente, n_1 e n_2 eventos simples, e sabendo que eles não têm eventos simples comuns, o número de eventos simples de $\mathcal{E}_1 \cup \mathcal{E}_2$ é, então, igual a $n_1 + n_2$. Logo, pela definição de probabilidade,

$$P(\mathcal{E}_1 \cup \mathcal{E}_2) = \frac{n_1 + n_2}{N} = \frac{n_1}{N} + \frac{n_2}{N} = P(\mathcal{E}_1) + P(\mathcal{E}_2).$$

Para $\mathcal{E}_1$ e $\mathcal{E}_2$ arbitrários, é suficiente considerarmos a possibilidade de terem algum evento simples comum. Denotemos esse evento comum por n_3. Então $\mathcal{E}_1 \cup \mathcal{E}_2$ tem $n_1 + n_2 - n_3$ eventos, e da definição de probabilidade, chegamos a

$$P(\mathcal{E}_1 \cup \mathcal{E}_2) = \frac{n_1 + n_2 - n_3}{N} = \frac{n_1}{N} + \frac{n_2}{N} - \frac{n_3}{N} = P(\mathcal{E}_1) + P(\mathcal{E}_2) - P(\mathcal{E}_1 \cap \mathcal{E}_2) \leq P(\mathcal{E}_1) + P(\mathcal{E}_2),$$

e assim obtemos (ii).
Por fim, como $P(\mathcal{S}) = 1$, para (iii), temos

$$1 = P(\mathcal{S}) = P(\mathcal{E} \cup \mathcal{E}^c) = P(\mathcal{E}) + P(\mathcal{E}^c),$$

pois, $\mathcal{E}$ e $\mathcal{E}^c$ são disjuntos. Consequentemente,

$$P(\mathcal{E}^c) = 1 - P(\mathcal{E}).$$

Exemplo 1.1.2. ***Problema dos Aniversários [Fel67, p. 33]:*** *- Em um grupo formado por r pessoas, qual é a probabilidade de pelo menos duas delas fazerem aniversário no mesmo dia?*

Esse problema tem intrigado e surpreendido estudantes, pois, uma primeira observação natural a fazer é que, em função do valor de r, a probabilidade pode ser muito alta. Sabemos que nem todos os anos têm a mesma duração (anos bissextos têm 366 dias). Com isso, consideraremos um ano com 365 dias. Além disso, cada uma das r pessoas pode ter nascido em qualquer uma das 365 datas disponíveis, isto é, todos os dias são equiprováveis, o que nos dá um espaço amostral $\mathcal{S}$ com 365^r eventos simples (que são sequências de tamanho r formadas por datas). Assim, vamos supor que $r < 365$ uma vez que, se $r \geq 365$ a probabilidade desejada seria 1.

Seja o evento $\mathcal{E} = \{$*ao menos 2 pessoas aniversariam no mesmo dia*$\}$; então, seu complementar é $\mathcal{E}^c = \{$*ninguém faz aniversário num mesmo dia*$\}$. Note que, para calcular o número de casos favoráveis à $\mathcal{E}^c$ precisamos contar o número de sequências distintas com r elementos (datas) tomados de um total de 365 datas disponíveis, ou seja, basta calcularmos um arranjo de 365 datas tomadas de r em r. Com isso, torna-se mais viável calcularmos $P(\mathcal{E}^c)$ e em seguida, obtemos $P(\mathcal{E}) = 1 - P(\mathcal{E}^c)$. Assim,

$$
\begin{aligned}
P(\mathcal{E}) = 1 - \frac{\left[\frac{365!}{(365-r)!}\right]}{365^r} &= 1 - \left[\frac{365 \cdot 364 \cdot \ldots \cdot (365-r+1)}{365^r}\right] \\
&= 1 - \left[1.\left(1 - \frac{1}{365}\right) \cdot \ldots \cdot \left(1 - \frac{r-1}{365}\right)\right].
\end{aligned}
$$

A seguinte tabela mostra que não é necessário um grupo muito grande de pessoas para que tenhamos uma possibilidade real de ao menos duas, aniversariarem numa mesma data.

Número de pessoas	$P(\mathcal{E})$
10	0,11695
20	0,41144
22	0,47569
23	0,50729
40	0,89123
60	0,99412
80	0,99991

Tabela 1 – *Probabilidades para um grupo com r pessoas*

Em particular, para $r = 23$ pessoas, a chance de pelo menos duas terem aniversário em comum excede 50%.

Isso mostra que não é necessário uma grupo muito grande de pessoas para que tenhamos a possibilidade real de ao menos duas, aniversariarem numa mesma data.

1.1.1 Probabilidade Condicional

Os conceitos de probabilidade condicional e de independência de eventos são fundamentais na Teoria da Probabilidade e lastreiam parte significativa de seu desenvolvimento. Para tanto, consideremos um espaço amostral com resultados equiprováveis.

Definição 1.1.5. *Sejam A e B dois eventos de um espaço amostral $\mathcal{S}$. Se $P(B) > 0$, a probabilidade condicional de A dado que ocorreu B, é dada por:*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}; \tag{1.1}$$

caso $P(B) = 0$, para alguns autores, é conveniente definir $P(A|B) = P(A)$.

Comparando com a definição de probabilidade, notamos que, uma vez "certo que B ocorreu", esse evento passa a figurar como referência para um novo espaço amostral S'. Como ilustração para essa definição, vamos considerar a seguinte situação especial: um experimento consiste em lançar um dado honesto duas vezes sobre uma mesa perfeitamente plana, e observar o número de pontos na face superior em cada um dos lançamentos. Suponha que não houve a observação dos lançamentos, porém, foi informado que em cada um dos lançamentos, o número de pontos observados é menor do que ou igual a três. Nessas condições, *qual é a probabilidade de que a soma dos pontos observados nos dois lançamentos seja igual a um número primo?* Para organizar as ideias, denotemos por B, o evento: *em cada um dos lançamentos, o número de pontos observados é menor do que ou igual a três*; e por A o evento: *a soma dos pontos observados nos dois lançamentos seja igual a um número primo*. Ora, claramente, estamos interessados em saber qual é a probabilidade de ocorrer o evento A, dado como certo, que ocorreu o evento B. Para o espaço amostral associado a esse experimento, temos os seguintes pares (n_1, n_2) das possibilidades para o 1º e 2º lançamentos, respectivamente:

$(1,1)$	$(1,2)$	$(1,3)$	$(1,4)$	$(1,5)$	$(1,6)$
$(2,1)$	$(2,2)$	$(2,3)$	$(2,4)$	$(2,5)$	$(2,6)$
$(3,1)$	$(3,2)$	$(3,3)$	$(3,4)$	$(3,5)$	$(3,6)$
$(4,1)$	$(4,2)$	$(4,3)$	$(4,4)$	$(4,5)$	$(4,6)$
$(5,1)$	$(5,2)$	$(5,3)$	$(5,4)$	$(5,5)$	$(5,6)$
$(6,1)$	$(6,2)$	$(6,3)$	$(6,4)$	$(6,5)$	$(6,6)$

Tabela 2 – *Resultados possíveis para dois lançamentos*

Consideremos $B = \{n_k \text{ é inteiro positivo} \leq 3\}$ com $1 \leq k \leq 2$ e, $A = \{n_1 + n_2 \text{ é primo}\}$ e a tabela acima, temos resumidamente:

$$B = \{(1,1);(1,2);(1,3);(2,1);(2,2);(2,3);(3,1);(3,2);(3,3)\} \quad \text{e}$$

$$A = \{(1,1);(1,2);(1,4);(1,6);(2,1);(2,3);(2,5);(3,2);(3,4);(4,1);(4,3);(5,2);(5,6);(6,1);(6,5)\}.$$

Entretanto, os únicos pontos de A que interessam são os que têm a soma de suas coordenadas como valores primos ≤ 5. Nessas condições, afirmar que o evento B ocorreu significa dizer que, agora, devemos considerar apenas os pontos do espaço amostral que pertençam a B. Assim, fica fácil concluir que a probabilidade procurada é $\frac{5}{9}$; pois dos 9 pontos de B, apenas 5 deles pertencem a A. Em outros termos, como $A \cap B = \{(1,1);(1,2);(2,1);(2,3);(3,2)\}$, segue $P(A \cap B) = \dfrac{5}{36}$ e $P(B) = \dfrac{9}{36}$, e então,

$$P(A|B) = \frac{\left(\dfrac{5}{36}\right)}{\left(\dfrac{9}{36}\right)} = \frac{5}{9}.$$

Além de servirem como base para modelagens diversas em situações práticas, as probabilidades condicionais ainda podem ser usadas para calcularmos probabilidades em geral, embora nem sempre tenhamos facilidade em caracterizar eventos. Para os casos eventualmente mais espinhosos, convém construirmos um condicionamento menos complicado. Da fórmula (1.1), obtemos $P(A \cap B) = P(B)P(A|B)$. Para um número finito n de interseções, a ocorrência de um evento numa etapa $k < n$ é influenciada pelas ocorrências das $k-1$ etapas anteriores. Essa generalização exprime a probabilidade da interseção de n eventos por meio das probabilidades condicionais sucessivas, como veremos nos lemas a seguir.

Lema 1.1.1 (**Regra do Produto de Probabilidades**). *Seja n um inteiro ≥ 2 e sejam A_1, A_2, ..., A_n eventos do espaço amostral $\mathcal{S}$, para o qual está definida a probabilidade P e com $P\left(\bigcap_{i=1}^{n} A_i\right) > 0$, tem-se então:*

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)\cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}). \quad (1.2)$$

Demonstração: Por indução sobre n. Para $n = 2$, segue da definição de probabilidade condicional que,

$$P(A_1 \cap A_2) = P(A_1)P(A_2|A_1),$$

pois $P(A_1) > 0$. Suponha que a igualdade (1.2) seja válida para $n = k$. Temos,

$$\begin{aligned} P(A_1 \cap A_2 \cap \cdots \cap A_k \cap A_{k+1}) &= P((A_1 \cap A_2 \cap \cdots \cap A_k) \cap A_{k+1}) \\ &= P(A_1 \cap A_2 \cap \cdots \cap A_k)P(A_{k+1}|A_1 \cap A_2 \cap \cdots \cap A_k) \\ &= (P(A_1)P(A_2|A_1)\cdots P(A_k|A_1 \cap A_2 \cap \cdots \cap A_{k-1})) \end{aligned}$$

$$P(A_{k+1}|A_1 \cap A_2 \cap \cdots \cap A_k).$$

Logo, pelo Princípio da Indução Matemática, o lema está provado. □

O próximo lema nos dá as ferramentas para calcular a probabilidade de um evento A conhecidas as probabilidades dos eventos que compõem uma partição de $\mathcal{S}$ e as respectivas probabilidades condicionais de A com cada um deles.

Lema 1.1.2 (**Fórmula da Probabilidade Total**). *Seja C_1, C_2, ..., C_n uma partição do espaço amostral $\mathcal{S}$, isto é, tem-se $\bigcup_{i=1}^{n} C_i = \mathcal{S}$ e $C_i \cap C_j = \emptyset$ sempre que $i \neq j$. Seja A um evento e P uma probabilidade > 0 definida nos eventos de $\mathcal{S}$, então:*

$$P(A) = \sum_{i=1}^{n} P(C_i)P(A|C_i).$$

Demonstração: Temos por hipótese, $\mathcal{S} = \bigcup_{i=1}^{n} C_i$ e também $A \in \mathcal{S}$. Como $A \cap \mathcal{S} = A$, segue que,

$$A = A \cap \mathcal{S} = A \cap \left(\bigcup_{i=1}^{n} C_i \right) = \bigcup_{i=1}^{n} (A \cap C_i).$$

Calculando a probabilidade de A, utilizando o fato dos eventos serem disjuntos, o lema anterior e a definição de probabilidade condicional, obtemos:
$P(A) = P\left(\bigcup_{i=1}^{n} (A \cap C_i)\right) = \sum_{i=1}^{n} P(A \cap C_i) = \sum_{i=1}^{n} P(C_i)P(A|C_i)$ □

Com o propósito de ilustrar esses dois últimos resultados, vamos discutir um exemplo.

Exemplo 1.1.3. *Tem-se 33 bolas que se distinguem apenas pela cor, distribuídas em cinco urnas distintas. Dessas bolas, 13 são brancas e 20 são pretas. Dessas urnas, sabe-se que: uma delas tem 5 bolas brancas e 10 pretas; duas delas têm 2 bolas brancas e 3 pretas e, as outras duas, têm 2 bolas brancas e 2 pretas.*

Uma dessas urnas é aleatoriamente escolhida, e desta, uma bola é aleatoriamente retirada. Qual é a probabilidade de que a bola retirada seja preta?

Resolução: Vamos denotar por $\mathcal{P}$, o evento: a bola retirada é preta. Denotemos ainda os eventos: uma delas tem 5 bolas brancas e 10 bolas pretas por A_1; duas delas têm 2 bolas brancas e 3 bolas pretas por A_2; e, as outras duas, têm 2 bolas brancas e 2 pretas por A_3. Sabendo que uma bola somente pode ser retirada de uma urna de composição A_1 , A_2 ou A_3, temos que $\mathcal{P} = A_1 \cap \mathcal{P} + A_2 \cap \mathcal{P} + A_3 \cap \mathcal{P}$. Pela definição de probabilidade condicional e pela fórmula de probabilidade total,

$$P(\mathcal{P}) = P(A_1)P(\mathcal{P}|A_1) + P(A_2)P(\mathcal{P}|A_2) + P(A_3)P(\mathcal{P}|A_3).$$

Por outro lado, temos as probabilidades:
$P(A_1) = \frac{1}{5}$; $P(A_2) = \frac{2}{5}$; $P(A_3) = \frac{2}{5}$; $P(\mathcal{P}|A_1) = \frac{2}{3}$; $P(\mathcal{P}|A_2) = \frac{3}{5}$ e $P(\mathcal{P}|A_3) = \frac{1}{2}$.
Com isso, concluímos que $P(\mathcal{P}) = \frac{1}{5} \cdot \frac{2}{3} + \frac{2}{5} \cdot \frac{3}{5} + \frac{2}{5} \cdot \frac{1}{2} = \frac{43}{75}$. □

1.1.2 Independência de Eventos

Apresentamos inicialmente uma noção intuitiva da independência de dois eventos, pois, mostra que a probabilidade de um não é alterada pela informação de que o outro ocorreu.

Definição 1.1.6. *Sejam A e B dois eventos de um espaço amostral $\mathcal{S}$, e suponha que $P(B) > 0$. O evento A é dito independente do evento B se: $P(A|B) = P(A)$.*

Se o evento A é independente do evento B, então,

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

Ademais, parece natural que o evento B seja independente do evento A. De fato, pois, para $P(A) > 0$,

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

Com isso, podemos afirmar que os eventos A e B são independentes se, e somente se,

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

A partir dessas condições, e da última igualdade, podemos generalizar a definição de independência de eventos para mais de dois eventos.

Definição 1.1.7. *Sejam os eventos* $A_1, A_2, ..., A_n$ *de um espaço amostral* $\mathcal{S}$*. Esses eventos são ditos independentes se:* $P(A_{i_1} \cap A_{i_2} \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k})$ *para todo* $k = 2, 3, \ldots, n$ *e todo* $\{i_1, i_2, \ldots, i_k\} \subset \{1, 2, \ldots, n\}$.

Um problema clássico no estudo da Teoria das Probabilidades e cuja solução é atribuída a Pierre de Fermat no século XVII, é o de observar que, para ensaios do tipo sucesso-fracasso que são realizados de forma independente, com probabilidade de sucesso p, qual seria a probabilidade de ocorrerem N sucessos antes de M fracassos?

Na solução é inicialmente apontado que o evento: *N sucessos antes de M fracassos* é equivalente à: *ocorrerem, ao menos, N sucessos nos primeiros M+N-1 ensaios*. Para obtermos a probabilidade de k sucessos em $N + M - 1$ ensaios, como a ordem de ocorrência dos ensaios bem sucedidos não importa, então, pela independência entre os ensaios, temos,

$$P(\{k \text{ sucessos em } (N + M - 1) \text{ ensaios}\}) = \binom{N + M - 1}{k} \cdot p^k \cdot (1 - p)^{(N+M-1)-k}$$

e, então,

$$P(\{N \text{ sucessos antes de } M \text{ fracassos}\}) = \sum_{k=N}^{N+M-1} \binom{N + M - 1}{k} \cdot p^k \cdot (1 - p)^{(N+M-1)-k}.$$

1.1.3 Variáveis Aleatórias e Processos Estocásticos

Nesta seção, apresentaremos definições de variável aleatória e de processos estocásticos que serão úteis na resolução de aplicações do capítulo 3. Havendo curiosidade e interesse por parte do leitor em uma exposição mais detalhada e aprofundada, vale a leitura de [Jam15] e [Doo53].

Podemos definir, informalmente, uma *variável aleatória* como uma função capaz de associar a cada ponto de um espaço amostral um número real. Neste trabalho, estamos interessados em um tipo específico que é a *variável aleatória discreta*.

Definição 1.1.8. *Uma variável aleatória* X *em um espaço de probabilidade é uma função real* $X : S \to \mathbb{R}$ *se o conjunto* $\{X \leq x\} = \{r \in S, \ talque, \ X(r) \leq x\}$ *é um evento aleatório para todo* $x \in \mathbb{R}$.

Definição 1.1.9. *Seja* X *uma variável aleatória. Se* X *assume valores em um conjunto enumerável diz-se que* X *é uma variável aleatória discreta.*

Em outros termos, se X é uma variável com valores $x_1, x_2, \ldots$, então, para $i = 1, 2, \ldots$ tem-se $p(x_i) = P(X = x_i)$.

Exemplo 1.1.4. *Consideremos a situação descrita no exemplo 1.1.1. Precisamos determinar a probabilidade de, ao final da terceira retirada ao acaso, exatamente duas dessas peças sejam* *boas. Para isso, definamos X a variável aleatória que nos dá o número de peças boas, e, considere ainda que nessa linha de produção 4% das peças produzidas são defeituosas e que as retiradas são independentes.*

Estamos interessados em calcular $P(X = 2)$. Observando os pontos do espaço amostral, temos favoravelmente BBD, BDB e DBB. Assim e sendo óbvio que 96% das peças produzidas são boas, segue que:

$$P(X = 2) = P(BBD) + P(BDB) + P(DBB) = 3(0,96)^2(0,04) = 11,06\%.$$

Definição 1.1.10. *Um processo estocástico é um modelo matemático utilizado no estudo de fenômenos aleatórios cujos resultados são funções.*

Em outros termos, consideremos um conjunto K não-vazio de reais ≥ 0. Fixando um k, podemos pensar num processo estocástico como uma família de variáveis aleatórias $\{X_k\}_{k\in K}$ onde $X_k : S \to E$ para cada $k \in K$ e um conjunto enumerável E que possui uma sequência de resultados possíveis $E_1, E_2, \ldots$. Um processo estocástico onde, dado um número finito de variáveis, o valor condicionalmente associado a uma delas depende do valor imediatamente anterior é chamado *processo de Markov*. Resumidamente,

$$P(X_k = x_k \mid X_1 = x_1, \ldots, X_{k-2} = x_{k-2},\ X_{k-1} = x_{k-1}) = P(X_k = x_k \mid X_{k-1} = x_{k-1}) = p_{k-1k};$$

e um processo cuja probabilidade condicional é satisfeita nesses termos é dito ser uma *cadeia de Markov*.

1.2 Matrizes

Nesta seção, definiremos uma matriz e, resumidamente, um espaço vetorial. Após, apresentaremos alguns dos principais tipos de matrizes. Em seguida, apresentaremos de forma sucinta conceitos considerados básicos, como as operações entre matrizes, determinantes, sistemas e transformações lineares, autovalores e autovetores. Para uma apreciação mais aprofundada do que trataremos nesta seção, é necessário algumas leituras da jovem, porém, vasta literatura da Álgebra Linear. Em particular, de autores com os quais servir-me, como [Ser02], [Bol78] e [Lim04].

Definição 1.2.1. *Dados m e n naturais, definimos uma* ***matriz*** *real A de ordem m–por–n, ou simplesmente, $A_{m\times n}$ - como uma tabela formada por números reais a_{ij} com $1 \leq i \leq m$ e $1 \leq j \leq n$, dispostos em m linhas e n colunas. Os números reais a_{ij} são chamados de entradas da matriz A e localizam-se na intersecção da linha i com a coluna j.*

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

Deste ponto em diante, cabe registrarmos para fins práticos que, sempre que nos referirmos a uma matriz o faremos denominando-a por uma letra maiúscula do nosso alfabeto e, quando necessário, indicaremos sua ordem; também a representaremos por uma tabela entre parênteses conforme fizemos na definição acima. O símbolo $\mathcal{M}(m,n)$ denota o conjunto de todas as matrizes m−por−n. Em relação às suas entradas, para uma referência escrita também de ordem prática, utilizaremos a notação a_{ij}, e também, A_{ij} sempre que não houver dúvidas ou ambiguidades. Uma matriz m−por−n cujas entradas são todas nulas, dá-se o nome de ***matriz nula***; na literatura encontra-se simplesmente expressa por $0_{m\times n}$.

Embora o termo *vetor* evoque intuitivamente aos estudantes, a noção de uma grandeza física com direção, sentido e magnitude, aqui entretanto, o tomaremos com uma noção e um contexto mais amplos. Em outros termos, o tomaremos como um elemento de um conjunto algebricamente estruturado, no qual estão bem definidas a adição entre esses elementos e a multiplicação por um número real. Estas operações satisfazem a comutatividade, a associatividade e a distributividade. Este conjunto, que é o terreno onde se desenvolve a Álgebra Linear é denominado ***espaço vetorial***. Como um exemplo, para todo natural n, temos o conjunto $\mathbb{R}^n$, que representa o ***espaço vetorial euclidiano*** **n**-***dimensional***. Os elementos de $\mathbb{R}^n$ são as listas ordenadas de números reais denominadas como $\vec{u} = (u_1, u_2, \ldots, u_n)$. Para o espaço vetorial $\mathbb{R}^\infty$, seus elementos são as listas infinitas como $\vec{a} = (a_1, a_2, \ldots, a_n \ldots)$.

Um subconjunto F de um espaço vetorial E, para o qual sejam definidas as mesmas operações, e satisfaça as mesmas condições, é um ***subespaço vetorial*** do espaço E. Dizemos que F é um conjunto *linearmente independente* (abreviadamente, *l.i.*) quando nenhum vetor $\vec{v} \in F$ é escrito como combinação linear de outros vetores de F, isto é, não existem α_i reais $\neq 0$, para $1 \leq i \leq n$ tais que, $\vec{v} = \alpha_1\vec{v_1} + \cdots + \alpha_n\vec{v_n}$, com $\vec{v_i} \in F$. Para o caso em que $F = \{\vec{v}\}$, se $\vec{v} \neq \vec{0}$ então, por definição, F é *l.i.* Quando F é *l.i.*, diz-se também que seus elementos são todos $\neq \vec{0}$ e são *vetores linearmente independentes*, pois, naturalmente, o vetor nulo é combinação linear de quaisquer outros. Do contrário, um conjunto $X \in E$ diz-se *linearmente dependente* (abreviadamente *l.d.*) quando não é *l.i.* Isto significa que algum dos vetores $\vec{u} \in X$ é combinação linear de outros elementos de X, ou então, que $X = \left\{\vec{0}\right\}$.

Uma ***base*** de um espaço vetorial E é um conjunto $\mathcal{B} \subset E$ linearmente independente que gera E. Isto significa que todo vetor $\vec{v} \in E$ é expresso de modo único como combinação linear de elementos de $\mathcal{B}$. Os vetores $e_1 = (1,0,0,\ldots,0)$, $e_2 = (0,1,0,\ldots,0), \cdots, e_n = (0,0,\ldots,1)$ formam uma base de $\mathbb{R}^n$ que se chama ***base canônica***. Se um espaço vetorial E admite uma base com k elementos, então, todas as bases de E tem exatamente k elementos, e esse número k é chamado de ***dimensão*** do espaço vetorial E.

No contexto das matrizes, por exemplo, uma matriz A de ordem n−por−m, podemos representar a i-ésima linha pelo vetor-linha $(a_{i1}\ a_{i2}\ \cdots\ a_{in})$, enquanto a j-ésima coluna pelo vetor-coluna

$$\begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}$$

que são, respectivamente, elementos dos espaços vetoriais $\mathbb{R}^n$ e $\mathbb{R}^m$.

Utilizaremos a partir desse ponto, exceto quando mencionado de forma distinta, a representação dos vetores como

$$\vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}.$$

A ***transposta*** de uma matriz $A \in \mathcal{M}(m,n)$ é a matriz A^T cujas entradas são a_{ji}, isto é, as colunas de A são as linhas de A^T. Quando $m = n$, então A é dita ser uma ***matriz quadrada*** (ou simplesmente, de ordem n). Essa matriz, é ainda chamada de ***simétrica*** se, e somente se $A^T = A$, isto é, $a_{ij} = a_{ji}$ para todo i e todo j. Ilustrando; a matriz

$$A = \begin{pmatrix} 1 & -7 & 6 \\ -7 & 1 & 0 \\ 6 & 0 & 5 \end{pmatrix}$$

é uma matriz simétrica.

As entradas a_{ii} com $1 \leq i \leq n$, de uma matriz quadrada A formam a sua ***diagonal principal***. Uma ***matriz diagonal*** de ordem n é uma matriz quadrada de ordem n em que todos os elementos que não pertençam à sua diagonal principal são nulos, isto é, se A é uma matriz diagonal, então, $a_{ij} = 0$ sempre que $i \neq j$:

$$A = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}.$$

Uma matriz diagonal de ordem n para a qual, $m_{ij} = \begin{cases} 0, \text{ se } i \neq j \\ 1 \text{ se } i = j \end{cases}$ é chamada de ***matriz identidade*** de ordem n e denotada por I_n. Em outras palavras,

$$I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

A matriz identidade é um caso especial de uma ***matriz de permutação***. Esta, é uma matriz quadrada com exatamente uma entrada diferente de zero em cada linha e cada coluna, sendo essa entrada um 1. Trataremos de forma detalhada das matrizes de permutação na seção 2.2.

A uma matriz quadrada A tal que, $a_{ij} = 0$ sempre que $i > j$, dá-se o nome de ***matriz triangular superior*** de ordem n:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix};$$

por outro lado, quando $a_{ij} = 0$ sempre que $i < j$, é uma ***matriz triangular inferior*** de ordem n:

$$A = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$

1.2.1 Operações com Matrizes

Dizemos que duas matrizes A e B de mesma ordem m−por−n são iguais, e escrevemos $A = B$, quando, para todo $1 \leq i \leq m$ e $1 \leq j \leq n$ tem-se $a_{ij} = b_{ij}$.

Se A e B são duas matrizes de mesma ordem $m \times n$ e, α e β são constantes reais quaisquer, então, a ***soma*** de A e B denotada por $A + B$, é uma matriz C com ordem m−por−n tal que, $c_{ij} = a_{ij} + b_{ij}$ para todo $1 \leq i \leq m$, e todo $1 \leq j \leq n$; a ***multiplicação por escalar*** α de uma matriz A é uma matriz $\alpha A = \alpha a_{ij}$ pata todo i e todo j.

$$\alpha A = \alpha \cdot \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} = \begin{pmatrix} \alpha a_{11} & \alpha a_{12} & \cdots & \alpha a_{1n} \\ \alpha a_{21} & \alpha a_{22} & \cdots & \alpha a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha a_{m1} & \alpha a_{m2} & \cdots & \alpha a_{mn} \end{pmatrix}.$$

A adição entre matrizes tem propriedades semelhantes à adição entre números reais e, à adição entre elementos de um espaço vetorial:

(i) $A + B = B + A$ (*Comutatividade*);

(ii) $(A + B) + C = A + (B + C)$ (*Associatividade*);

(iii) $A + (-A) = \mathbf{0}$ *com* $\mathbf{0}$ *a matriz nula com mesma ordem de* A;

(iv) $A + \mathbf{0} = A$ (*Elemento Neutro*).

Verificamos ainda que, para quaisquer A e B em $\mathcal{M}(m,n)$, e α, β em $\mathbb{R}$ vale:

(i) $\alpha(A+B) = \alpha A + \alpha B$ (*Distributividade*);

(ii) $(\alpha+\beta)A = \alpha A + \beta A$;

(iii) $\alpha(\beta A) = (\alpha\beta)A$ e $1A = A$.

O conceito de *multiplicação de matrizes*, cuja ideia foi expandida, formalizada e apresentada por Arthur Cayley em 1858, em seu trabalho intitulado *A Memoir on the Theory of Matrices*, publicado na revista *Philosophical Transactions of the Royal Society of London*; embora não seja de trivial execução, é de amplo conhecimento por estudantes do Ensino Médio. O *produto matricial AB* entre as matrizes $A_{m\times n}$ e $B_{n\times p}$ é definido como a matriz $D_{m\times p}$, tal que, $D = AB$, isto é,

$$d_{ij} = \sum_{1\leq k\leq n} a_{ik}b_{kj} = a_{i1}b_{1j} + \cdots + a_{in}b_{nj}.$$

Observando a definição assim apresentada, verificamos que, para que seja possível o produto AB é necessário que o número de colunas de A seja igual ao número de linhas de B. Uma outra forma de entender esse produto é por meio de uma *combinação linear* entre os vetores-coluna de A e as entradas (ou escalares) de cada vetor-coluna de B. Resumidamente, consideremos a matriz $A_{m\times n}$ e o vetor $\vec{v}$ em $\mathbb{R}^n$. Então, da definição acima temos:

$$A\vec{v} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} a_{11}v_1 + a_{12}v_2 + \cdots + a_{1n}v_n \\ a_{21}v_1 + a_{22}v_2 + \cdots + a_{2n}v_n \\ \vdots \\ a_{m1}v_1 + a_{m2}v_2 + \cdots + a_{mn}v_n \end{pmatrix}.$$

Vale ressaltar ainda que, em geral, o produto matricial não e comutativo, isto é, $AB \neq BA$. Algumas razões justificam essa particularidade: para $A_{m\times n}$ e $B_{n\times p}$, segue que AB está bem definida e BA, por sua vez, não se $m \neq p$; supondo ambas bem definidas e com mesmo tamanho, podemos ter elementos de AB distintos dos elementos de BA, por contra-exemplo, se $A = \begin{pmatrix} 0 & 1 \\ 2 & 3 \end{pmatrix}$ e $B = \begin{pmatrix} 2 & 1 \\ 5 & 6 \end{pmatrix}$, temos que, $AB = \begin{pmatrix} 5 & 6 \\ 19 & 20 \end{pmatrix}$ enquanto $BA = \begin{pmatrix} 2 & 5 \\ 12 & 23 \end{pmatrix}$. Dentre as propriedades do produto matricial, destacamos:

(i) $A(BC) = (AB)C$ Associatividade;

(ii) $A(B+C) = AB + AC$ Distributividade à esquerda;

(iii) $(A+B)C = AC+BC$ Distributividade à direita.

Convém destacarmos que para uma matriz quadrada A de ordem n−por−n, define-se $A^0 = I_n$. Ademais, $A^1 = A$; $A^2 = A \cdot A$; e de modo geral $A^k = A \cdots A$ com k fatores iguais a A.

Definição 1.2.2. *Seja A uma matriz quadrada de ordem n. Se existir uma matriz quadrada B com ordem n, tal que, $AB = BA = I_n$, então, diremos que A é **invertível** e que B é sua **inversa** denotada por $B = A^{-1}$. Se não existir uma matriz B, então a matriz A não é invertível (equivalentemente, é singular).*

Da definição acima, verificamos que vale também a afirmação de que B é invertível e que A é sua inversa. Verificamos também que se B e C são inversas de uma matriz A então $B = C$. Com efeito, pois, dado que B é inversa de A, temos $BA = I$. Com isso, multiplicando à direita ambos os membros, pela matriz C , temos $(BA)C = IC = C$. Por outro lado, é fato que $(BA)C = B(AC) = BI = B$ dado que C também é inversa de A. Então, concluímos que $B = C$. Há ainda um rico contexto onde as operações entre matrizes, em particular, a multiplicação, tem papel de destaque. Trata-se das ***transformações lineares***. Por ser distinto das ferramentas que precisaremos, não trataremos desse contexto.

Outro importante conceito acerca das matrizes quadradas e da aplicabilidade para modelar e resolver problemas diversos, é o de ***determinante***. Inicialmente sendo utilizado na resolução de ***sistemas de equações lineares*** ou simplesmente, ***sistemas lineares*** com n equações e n incógnitas, que encontramos na literatura inclusive do ensino médio.

Resumidamente apresentado de forma matricial como $A\vec{x} = \vec{b}$, é expresso como a igualdade a seguir:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}.$$

Posteriormente, o determinante foi também identificado como a área de um paralelogramo, como o volume de um paralelepípedo e, como uma função multilinear alternada que diferentemente das funções reais, associam números reais a variáveis que são matrizes quadradas. Denotaremos o determinante de uma matriz A por $\det A$. A função determinante tem propriedades úteis que tornam práticas provas e demonstrações de resultados importantes. Destacamos algumas:

Dada uma matriz quadrada A de ordem n,

(i) $\det A = 0$ se, e somente se A é singular. Outro aspecto dessa singularidade são as matrizes que possuem linhas(ou colunas) nulas, possuem duas linhas idênticas, ou possuem duas linhas proporcionais ou duas colunas proporcionais;

(ii) $\det A = \det A^T$;

(iii) $\det(k.A) = k^n.\det A$ para $k \in \mathbb{R}$;

(iv) $\det(A.B) = \det A.\det B$ se B for quadrada com mesma ordem de A;

Ademais, se $\det A \neq 0$, então A é invertível, e segue diretamente que $\det A^{-1} = \dfrac{1}{\det A}$.

1.3 Autovalor e Autovetor

Sejam A uma matriz quadrada de ordem n e $\vec{u}$ um vetor não-nulo em $\mathbb{R}^n$. O vetor $\vec{u}$ é chamado de **autovetor** de A se $A\vec{u}$ for um múltiplo escalar de $\vec{u}$, isto é, $A\vec{u} = \lambda\vec{u}$ para algum escalar real (ou complexo) λ, que por sua vez, é denominado **autovalor** de A. Dizemos ainda, mais precisamente que $\vec{u}$ é um autovetor associado ao autovalor λ.

Em linhas gerais, quando $\vec{u}$ for um autovetor de A, a multiplicação por A preservará sua direção. Em virtude do sinal e da magnitude do autovalor λ associado a $\vec{u}$, a multiplicação $A\vec{u} = \lambda\vec{u}$ comprime ou expande o vetor $\vec{u}$, invertendo seu sentido no caso em que $\lambda < 0$. A equação $A\vec{u} = \lambda\vec{u}$ pode ser equivalentemente escrita como $A\vec{u} = \lambda I_n\vec{u}$, e daí, $(\lambda I_n - A)\vec{u} = \vec{0}$. Dessa forma, para que λ seja um autovalor de A esta ***equação característica*** deve possuir alguma solução $\vec{u}$ não-nula (não-trivial), e isso ocorre se, e somente se a matriz $(\lambda I_n - A)$ é singular, ou seja, se $\det(\lambda I_n - A) = 0$. Expandindo $\det(\lambda I_n - A)$ encontramos um polinômio de grau n chamado ***polinômio característico*** de A em λ da forma, $P(\lambda) = \lambda^n + \alpha_1\lambda^{n-1} + \cdots + \alpha_n$ onde o coeficiente líder é 1.

Pelo Teorema Fundamental de Álgebra, ele terá exatamente n raízes complexas (contadas suas multiplicidades) que são os autovalores de A. A ***multiplicidade algébrica*** é o número de vezes que λ zera o polinômio e, a ***multiplicidade geométrica*** é a dimensão do subespaço de autovetores associados a λ. Como consequências das definições de autovalores e de autovetores, e das características acima, seguem as proposições [1.3.1] e [1.3.2].

Proposição 1.3.1. *Autovetores de uma matriz quadrada A de ordem n associados a autovalores distintos são linearmente independentes.*

Proposição 1.3.2. *Se A for uma matriz quadrada de ordem n triangular (superior, inferior ou diagonal) então seus autovalores são as entradas de sua diagonal principal.*

Por outro lado, observamos ainda que, sendo $\vec{u}$ e $\vec{v}$ dois autovetores associados ao mesmo autovalor λ_0, isto é, $A\vec{u} = \lambda_0\vec{u}$ e $A\vec{v} = \lambda_0\vec{v}$, então, para quaisquer c_1 e c_2 complexos (e obviamente, reais) um vetor $\vec{w} = c_1\vec{u} + c_2\vec{v}$ é tal que, $A\vec{w} = \lambda_0\vec{w}$.

Com efeito, pois,

$$A\vec{w} = A(c_1\vec{u} + c_2\vec{v}) = c_1A\vec{u} + c_2A\vec{v} = c_1\lambda_0\vec{u} + c_2\lambda_0\vec{v} = \lambda_0(c_1\vec{u} + c_2\vec{v}) = \lambda_0\vec{w}.$$

Como os autovetores associados a um autovalor λ de uma matriz A são os vetores não nulos que satisfazem a equação $(\lambda I_n - A)\vec{u} = \vec{0}$, podemos afirmar que esses autovetores são os vetores não nulos do ***espaço nulo*** da matriz $(\lambda I_n - A)$, ou seja, do conjunto que coleciona as soluções não-nulas do sistema $(\lambda I_n - A)\vec{u} = \vec{0}$. Esse conjunto é denominado ***auto-espaço*** de A associado λ.

Proposição 1.3.3. *Se k for inteiro positivo, λ um autovalor de uma matriz quadrada A de ordem n e $\vec{u}$ um autovetor associado, então λ^k é um autovalor de A^k e $\vec{u}$ é um autovetor associado.*

Proposição 1.3.4. *Se A é uma matriz quadrada de ordem n, então são equivalentes as seguintes afirmações:*

(a) $det(A) \neq 0$, isto é, A é invertível(não singular);

(b) Os vetores-coluna de A são linearmente independentes e geram $\mathbb{R}^n$;

(c) Os vetores-linha de A são linearmente independentes e geram $\mathbb{R}^n$;

(d) Os vetores-coluna(e também, linha) de A formam uma base de $\mathbb{R}^n$.

Como consequência desses resultados e para tornar prática a resolução de problemas onde há necessidade de calcular potências de matrizes, a decomposição de uma matriz A na forma $A = SDS^{-1}$ onde S é uma matriz invertível e D uma matriz diagonal, é uma chave importante. Essa decomposição é chamada de ***decomposição espectral*** da matriz A que por sua vez, é dita ser ***diagonalizável***. Os vetores-coluna da matriz S são os autovetores de A e as entradas da diagonal principal de D são os autovalores de A associados aos respectivos autovetores.

Como exemplo dessa decomposição, consideremos a matriz

$$A = \begin{pmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{pmatrix}.$$

Seu polinômio característico é $P_A(\lambda) = (\lambda-2)^2(\lambda-1)$ e disso obtemos os autovalores $\lambda_1 = 2$ com autovetores associados $\vec{v} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$ e $\vec{u} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, e, para $\lambda = 1$, o autovetor

$\vec{w} = \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}$. Assim, fazendo $S = \begin{pmatrix} -1 & 0 & -2 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$, temos $S^{-1} = \begin{pmatrix} 1 & 0 & 2 \\ 1 & 1 & 1 \\ -1 & 0 & -1 \end{pmatrix}$, e dessa forma,

$$S^{-1}AS = \begin{pmatrix} 1 & 0 & 2 \\ 1 & 1 & 1 \\ -1 & 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{pmatrix} \cdot \begin{pmatrix} -1 & 0 & -2 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} = D$$

donde concluímos que,

$$A = SDS^{-1} = \begin{pmatrix} -1 & 0 & -2 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 2 \\ 1 & 1 & 1 \\ -1 & 0 & -1 \end{pmatrix}.$$

2 Matrizes de Markov

Neste capítulo mostraremos uma importante ferramenta no estudo e na solução de problemas que surgem a partir da análise do comportamento de certos sistemas físicos que evolvem constantemente, sob a influência de um conjunto finito de variáveis cujos valores mudam com o passar do tempo. Esses sistemas são chamados de Sistemas Dinâmicos e seu estudo compreende campos de pesquisa importantes da matemática como a Teoria das Probabilidades, Álgebra Linear, Física Matemática e Equações Diferenciais; tendo aplicações relevantes nas engenharias, biologia, economia, dentre outras ciências sociais. Definiremos as ***Matrizes de Markov***, fruto de estudos do matemático russo Andrei Andreyevich Markov que descreveu matematicamente, sob o ponto de vista do produto matricial, o cálculo de probabilidades de ocorrência de certos eventos que dependem apenas do estado em que o fenômeno se encontra para que seja calculada a probabilidade de estar num estado seguinte.

2.1 Definição e exemplos iniciais

Definição 2.1.1. *Uma matriz quadrada ≥ 0 diremos ser uma* ***Matriz de Markov*** *se a soma das entradas de cada vetor-coluna é sempre igual a 1.*

Na Teoria da Probabilidade essas matrizes são tradicionalmente chamadas de matrizes de transição de probabilidades (ou simplesmente matriz de transição) e, cada vetor-linha tem como soma de suas entradas o valor 1. Em nosso trabalho consideraremos a soma das entradas de cada vetor-coluna sempre igual a 1.

Como exemplo trivial, a matriz $A = \begin{pmatrix} \frac{2}{3} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} \end{pmatrix}$ é uma matriz de Markov; matrizes identidade são matrizes de Markov.

As matrizes de Markov também são conhecidas como ***matrizes estocásticas***. Cada coluna de uma matriz de Markov é um vetor cujas entradas são ≥ 0 e com soma 1. Cada um desses vetores são chamados ***vetor de probabilidade*** (ou simplesmente, ***vetor estocástico***). Vejamos alguns resultados acerca desse tipo especial de matriz.

Proposição 2.1.1. *Se $\vec{p}$ é um vetor estocástico e A é uma matriz de Markov, então, $\vec{Ap}$ é um vetor estocástico.*

Demonstração: Suponha que $\vec{a}_1$, $\vec{a}_2, \ldots, \vec{a}_n$ sejam os vetores coluna de A e $p_1, p_2, \ldots, p_n$ as entradas não negativas do vetor estocástico $\vec{p}$. O produto $A\vec{p}$ pode ser expresso como uma combinação linear dos vetores-coluna de A, com os coeficientes sendo as entradas de $\vec{p}$, isto é, denotando por $\overline{p}_i$ com $1 \leq i \leq n$, as entradas de $A\vec{p}$, segue que

$$A\vec{p} = \begin{pmatrix} \overline{p}_1 \\ \overline{p}_2 \\ \vdots \\ \overline{p}_n \end{pmatrix} \text{ para } \overline{p}_1 = \sum_{1 \leq k \leq n} p_k a_{1k};\ \overline{p}_2 = \sum_{1 \leq k \leq n} p_k a_{2k}; \cdots;\ \overline{p}_n = \sum_{1 \leq k \leq n} p_k a_{nk}.$$

Ora, se somarmos $\overline{p}_1 + \overline{p}_2 + \cdots + \overline{p}_n$ obtemos $p_1 \cdot 1 + p_2 \cdot 1 + \cdots + p_n \cdot 1 = 1$, donde concluímos que $A\vec{p}$ é um vetor estocástico. □

Proposição 2.1.2. *O produto de duas matrizes de Markov de uma mesma ordem é uma matriz de Markov.*

Demonstração: Sejam A e B duas matrizes de Markov de uma mesma ordem, e $\vec{b}_i$ com $1 \leq i \leq n$ os vetores-coluna de B. A matriz AB é tal que, suas colunas são os vetores $A\vec{b}_i$ com $1 \leq i \leq n$. Note que cada $A\vec{b}_i$ é um vetor estocástico, conforme provamos acima, para todo i. Concluímos que AB é uma matriz de Markov. □

Verificamos com facilidade que a soma de duas matrizes de Markov não é uma matriz de Markov, pois, considerando as matrizes A e B, como da proposição acima, a soma das entradas de cada vetor-coluna da matriz $A + B$ será igual a 2. A inversa de uma matriz de Markov não é, necessariamente, uma matriz de Markov. Com efeito, pois, a matriz $A = \begin{pmatrix} \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{2}{3} \end{pmatrix}$ tem como inversa a matriz $A^{-1} = \begin{pmatrix} 4 & -2 \\ -3 & 3 \end{pmatrix}$ que, embora tenha soma das entradas dos vetores-coluna iguais a 1, possui entradas negativas.

No que diz respeito a autovalores e autovetores de uma matriz de Markov, temos o seguinte resultado.

Proposição 2.1.3. *Uma matriz de Markov tem sempre um autovalor igual a 1. Qualquer outro é, em valor absoluto, menor do que ou igual a 1.*

Demonstração: Seja A uma matriz de Markov de ordem n. A soma das entradas de cada vetor-linha de sua transposta A^T é igual a 1. Dessa forma, a matriz A^T tem, portanto, o autovetor

$$\vec{v} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Como matrizes transpostas têm o mesmo determinante, como vimos no capítulo 1, as matrizes $(\lambda I_n - A)$ e $(\lambda I_n - A^T)$ têm o mesmo determinante, de maneira que os autovalores de A e de A^T são os mesmos. Assim, como A^T tem autovalor 1, a matriz A também possui autovalor 1.

Suponha agora que $\vec{u}$ é um autovetor associado a um autovalor com $|\lambda| > 1$. Então, $A^k\vec{u} = |\lambda|^k\vec{u}$ tem entradas que crescem exponencialmente para k suficientemente grande. Isso implica que existe $(A^k)_{ij}>1$. Mas, A^k é uma matriz de Markov e tem todas as suas entradas ≤ 1. E assim, a suposição de um autovalor maior do que 1 não pode ser válida, e então concluímos que todos os outros autovalores são, em valor absoluto, menor do que ou igual a 1. □

Vamos ilustrar a definição e os resultados acima.

Exemplo 2.1.1. *A matriz de Markov* $\begin{pmatrix} \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \end{pmatrix}$ *tem a soma das entradas de cada vetor-linha igual a 1. Então, pela proposição acima, tem um autovalor* $\lambda_1 = 1$ *e um autovetor associado* $\vec{u} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$.

Ademais, essa matriz tem polinômio característico $P(\lambda) = \lambda^4(\lambda - 1)$, *donde tiramos seu outro autovalor* $\lambda_2 = 0$ *com multiplicidade 4. Os autovetores associados a esse autovalor são:*

$$\vec{v_1} = \begin{pmatrix} -1 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad \vec{v_2} = \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \vec{v_3} = \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \vec{v_4} = \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

2.2 Matrizes Irredutíveis

Nesta seção, trataremos de conceitos que são menos estudados e trabalhados nos cursos de Álgebra Linear, como os de matriz ≥ 0, positiva, redutível e irredutível. Nosso objetivo é conhecer os elementos e propriedades relevantes que utilizaremos na demonstração do Teorema de Perron-Frobenius. Faremos então, uma breve exposição destes conceitos e de suas propriedades.

Definição 2.2.1. *Uma matriz real* A *de tamanho* n*−por−*n *é uma matriz não-negativa, a qual denota-se por* $A \geq 0$, *se* $A_{ij} \geq 0$ *para cada* i *e* j *em* $\{1, 2, ..., n\}$. *Por outro lado, se* A *é* *uma matriz positiva então ela tem* $A_{ij} > 0$ *para cada* i *e* j *em* $\{1, 2, \ldots, n\}$*, e por conseguinte, a denotamos por* $A > 0$.

Notamos da definição acima que, se $A \geq 0$ então, para todo $k \geq 1$, verifica-se que $A^k \geq 0$, e ainda, que $A^k > 0$ sempre que a matriz A for positiva ou tiver a maior parte das entradas de cada linha e de cada coluna não-nula. Como exemplos de matrizes ≥ 0 que utilizaremos neste trabalho, destacamos as *matrizes de permutação*, que são matrizes quadradas na qual cada linha e cada coluna tem uma única entrada igual a 1 e, todas as outras, nulas. Outra maneira de concebermos as matrizes de permutação é verificarmos que elas resultam de um reordenamento das linhas(ou colunas) das matrizes identidade. Nesse contexto, utilizando os vetores da base canônica do espaço $\mathbb{R}^n$ com $n \geq 2$, temos que, para $n = 2$ existem duas matrizes de permutações possíveis, a saber: (e_1, e_2) e (e_2, e_1). De modo análogo, verificamos que para $n = 3$ existem seis matrizes de permutações possíveis, para $n = 4$ existem vinte e quatro, etc. Nessas condições, existem $n!$ matrizes de permutações possíveis a partir da identidade de ordem n. As matrizes

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

são exemplos de algumas matrizes de permutações de ordens 2, 3 e 4, respectivamente.

Pela definição das matrizes de permutação, podemos verificar facilmente que:

(i) O produto matricial entre matrizes de permutação é uma matriz de permutação;

(ii) Elas são matrizes *ortogonais*, isto é, sua inversa é igual à sua transposta ($P^{-1} = P^T$). Em particular, o determinante de uma matriz de permutação é sempre igual a ± 1.

De fato, para (i) a conclusão é imediata, pois, é suficiente notarmos que, se trata da permutação das linhas da identidade. Para (ii), note que $P^T P = I$ (onde I é a identidade de ordem correspondente). Ademais, $\det(P^T P) = \det(P^T)\det(P) = (\det(P))^2 = 1$, donde concluímos que $\det(P) = \pm 1$.

Dada uma matriz A, o produto PA troca as linhas da matriz A, enquanto o produto AP troca as colunas. Já o produto $P^{-1}AP$ troca as linhas e as colunas de A do mesmo modo.

Definição 2.2.2. *Uma matriz quadrada* A *é dita redutível se existe uma matriz de permutação* P*, tal que*

$$P^{-1}AP = \begin{pmatrix} X & Y \\ 0 & Z \end{pmatrix},$$

onde X e Z são matrizes quadradas; a matriz 0 *é nula e Y qualquer. Se uma matriz quadrada A não é redutível, então ela é irredutível, o que significa dizer que a matriz A não é similar, por uma matriz de permutação, a uma matriz bloco-triangular-superior.*

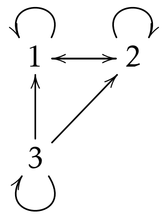
Nesse contexto, é natural então, questionarmos: Há alguma característica ou propriedade comum às matrizes redutíveis? E às irredutíveis? Caso exista, é possível ilustrarmos tais características ou propriedades? Mencionamos que matrizes irredutíveis são fundamentais nos estudos de grupos lineares e representações de grupos [Za93].

Para toda matriz $A \geq 0$ de tamanho n–por–n podemos associar um *grafo orientado* $\Gamma(A)$, onde seus vértices são os números $1, 2, ..., n$ e para cada par (i, j) de vértices existe uma aresta orientada de i para j se, e somente se $A_{ij} > 0$. Em um grafo como este, um *caminho orientado* define uma sequência de vértices conectados por arestas orientadas. A partir desta definição e do efeito do produto $P^{-1}AP$ sobre A, temos que $\Gamma(P^{-1}AP) = \Gamma(A)$ a menos da ordem dos vértices.

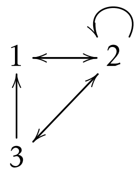
Vamos ilustrar alguns exemplos de matrizes $A \geq 0$ e seus respectivos grafos associados que nos ajudarão a identificar características ou propriedades comuns e que sirvam também como conveniente critério de redutibilidade e de irredutibilidade.

(1) Matrizes com algumas entradas nulas

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 4 & 5 & 0 \\ 7 & 8 & 9 \end{pmatrix}, \qquad \Gamma(A):$$



$$B = \begin{pmatrix} 0 & 2 & 0 \\ 4 & 5 & 6 \\ 7 & 8 & 0 \end{pmatrix}, \qquad \Gamma(B):$$



Verificamos com os exemplos acima, a existência de *laços* nos vértices 1, 2 e 3 da matriz A e no vértice 2 da matriz B, que ilustram uma entrada positiva na diagonal principal dessas matrizes. Verificamos também facilmente que, em relação à matriz A, não há caminhos orientados dos vértices 1 e 2 para o vértice 3, o que ilustra um grafo não fortemente conexo.

O resultado seguinte garante que a transposição de matrizes preserva a irredutibilidade.

Proposição 2.2.1. *Se A é uma matriz irredutível, então A^T também é irredutível.*

Demonstração: Se A^T fosse redutível, então, existiria uma matriz de permutação Q tal que,

$$Q^{-1}A^TQ = \begin{bmatrix} X & Y \\ 0 & Z \end{bmatrix},$$

para X e Z quadradas. Com isso, e do fato de Q ser ortogonal, seguiria que,

$$Q^{-1}A^TQ = [Q^{-1}AQ]^T$$

e disso teríamos, $Q^{-1}AQ = \begin{bmatrix} X^T & 0^T \\ Y^T & Z^T \end{bmatrix}$.

Considere a matriz de permutação P cujas entradas são todas nulas salvo aquelas na diagonal não-principal que são iguais a 1. Em outros termos, se P possui ordem n-por-n então $p_{ij} = 1$ se $i + j = n + 1$, $p_{ij} = 0$ sempre que $i + j \neq n + 1$. Notemos que se A possui ordem n-por-n então AP é obtida de A escrevendo-se as colunas de A na ordem inversa, e PA por sua vez, obtida de A escrevendo-se as linhas de A na ordem inversa. Em particular, $P^{-1} = P$. Assim

$$P^{-1}(Q^{-1}AQ)P = \begin{bmatrix} Z' & Y' \\ 0' & X' \end{bmatrix}$$

com submatrizes Z', X' quadradas. Entretanto, isso contradiz a hipótese de A ser irredutível. Portanto, se A é irredutível, então A^T também é irredutível. □

Podemos obter exemplos mais evidentes acerca das matrizes redutíveis e irredutíveis através do resultado seguinte.

Proposição 2.2.2. *Seja A uma matriz ≥ 0 de tamanho n–por–n. Então, as seguintes afirmações são equivalentes:*

(1) A é irredutível;

(2) O conjunto $\{1, 2, ..., n\}$ não pode ser dividido em dois subconjuntos não-vazios I e J com a propriedade: $A_{ij} = 0$ se $i \in I$ e $j \in J$;

(3) O grafo $\Gamma(A)$ é fortemente conexo, isto é, para quaisquer dois vértices distintos i e j de $\Gamma(A)$, existe um caminho orientado de i para j em $\Gamma(A)$.

Demonstração: (1) $\Rightarrow$ (2) : Suponhamos que o conjunto $\{1, 2, ..., n\}$ seja particionado em dois subconjuntos não-vazios I e J, tal que, $A_{ij} = 0$ sempre que $(i, j) \in I \times J$. Então a matriz A tem em alguma linha (ou coluna), no mínimo, $(n - 1)$ entradas nulas abaixo (ou acima) da sua diagonal principal conforme $i > j$ (ou $i < j$, respectivamente). Dessa forma, e pela proposição anterior, conseguimos uma matriz de permutação P de tal sorte que $P^{-1}AP$ é similar, por P a uma matriz bloco-triangular-superior, o que contradiz a irredutibilidade de A.

(2) $\Rightarrow$ (3) Por outro lado, não sendo possível uma partição do conjunto $\{1, 2, ..., n\}$ em dois subconjuntos I e J com $A_{ij} = 0$, então devemos ter $A_{ij} > 0$ para cada par (i, j), ou ao menos, para os pares consecutivos dois a dois, tomados numa sequência como $i_1 < i_2 < \cdots < i_k < j$. Pela definição de grafo, temos associado à matriz A um grafo $\Gamma(A)$ fortemente conexo.

(3) $\Rightarrow$ (1) Se $\Gamma(A)$ é fortemente conexo então todos os seus vértices estão dois a dois, conectados direta ou indiretamente. Em outros termos, se ocorrer $A_{ij_\alpha} = 0$, então ocorrerá $A_{ij_\beta} > 0$ e $A_{j_\beta j_\alpha} > 0$. Nessas condições, o conjunto $\{1, 2, \ldots, n\}$ não pode ser particionado em dois subconjuntos não vazios I e J com $A_{ij} = 0$ sempre que $(i, j) \in I \times J$, e consequentemente a matriz A é irredutível. $\square$

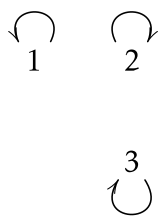
A partir do exposto até aqui, podemos elencar algumas características e propriedades referentes às matrizes redutíveis e às irredutíveis por meio de alguns exemplos.

2.2.1 Exemplos de matrizes redutíveis

- ***Matriz Identidade***. De fato, pois, é suficiente observarmos que todas as entradas nulas estão distribuídas acima e abaixo da diagonal principal, e nesses casos, seus grafos não são fortemente conexos e o conjunto $\{1, 2, \ldots, n\}$ pode ser facilmente dividido em subconjuntos I e J com $A_{ij} = 0$ sempre que $i \in I$ e $j \in J$.

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \Gamma(I_2):$$

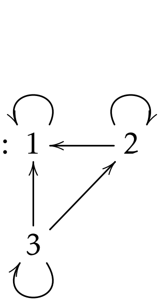
$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \Gamma(I_3):$$



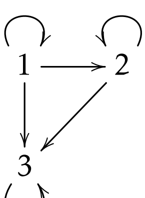
O que estende aos demais casos de dimensão superior a 3.

- ***Matrizes Triangulares (inferiores e superiores)***. É um caso análogo ao das matrizes identidade, porém, aqui com entradas nulas inferiores (ou superiores) à diagonal principal, o que nos possibilita verificar.

$$T_i = \begin{pmatrix} \alpha_1 & 0 & 0 \\ \alpha_2 & \alpha_3 & 0 \\ \alpha_4 & \alpha_5 & \alpha_6 \end{pmatrix} \qquad \Gamma(T_i):$$



$$T_s = \begin{pmatrix} \beta_1 & \beta_2 & \beta_3 \\ 0 & \beta_4 & \beta_5 \\ 0 & 0 & \beta_6 \end{pmatrix} \qquad \Gamma(T_s):$$



Em ambos os casos, verificamos que os grafos não são fortemente conexos, pois, em $\Gamma(T_i)$ não há caminho orientado do vértice 1 para o vértice 2, por exemplo. Assim como ocorre em $\Gamma(T_s)$ que não há caminho orientado do vértice 2 para o vértice 1. Nestes casos, o mesmo se aplica às matrizes de ordens superiores.

- ***Matrizes com filas nulas (linha ou coluna)***. Notamos com facilidade que um vetor-linha (ou vetor-coluna) nulo em uma matriz $A_{n\times n}$ nos dá uma divisão do conjunto $\{1,2,...,n\}$ em dois subconjunto I e J com a propriedade: $A_{ij} = 0$ se $i \in I$ e $j \in J$, pois, fixando uma linha k para $1 \leq k \leq n$, isso nos dá $A_{kj} = 0$ com j percorrendo os valores $\{1,2,...,n\}$, e então, tomamos os subconjuntos $I = \{k\}$ e $J = \{1,2,...,n\} \setminus \{k\}$. Ilustrando

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ A_{k1} = 0 & A_{k2} = 0 & \cdots & A_{kn} = 0 \\ \vdots & \vdots & \cdots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{pmatrix}$$

Concluímos de forma análoga tomando nula uma coluna arbitrária de A.

2.2.2 Exemplos de matrizes irredutíveis

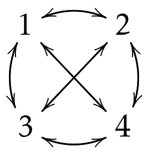
- ***Matrizes Positivas***. Com efeito, pois, dada uma matriz $A_{n\times n}$ positiva, a matriz $P^{-1}AP$ não será bloco-triangular-superior, visto que não possui entradas nulas. Temos ainda, associado a essa matriz um grafo fortemente conexo. Vejamos um caso que ilustra essa característica.

$$A = \begin{pmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{pmatrix}, \qquad \Gamma(A): \begin{tikzcd} 1 \arrow[loop above] \arrow[r, leftrightarrow] \arrow[d, leftrightarrow] \arrow[dr, leftrightarrow] & 2 \arrow[loop above] \arrow[d, leftrightarrow] \arrow[dl, leftrightarrow] \\ 3 \arrow[loop below] \arrow[r, leftrightarrow] & 4 \arrow[loop below] \end{tikzcd}$$

Analogamente estende-se para matrizes de ordem menor e maior do que 4.

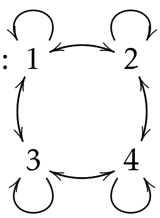
- ***Matrizes com diagonal principal nula e demais entradas positivas***. Dada uma matriz A com essa característica, é imediata a sua irredutibilidade uma vez que $A_{ij} = 0$ se, e somente se $i = j$. Nessas condições, ilustramos.

$$A = \begin{pmatrix} 0 & A_{12} & A_{13} & A_{14} \\ A_{21} & 0 & A_{23} & A_{24} \\ A_{31} & A_{32} & 0 & A_{34} \\ A_{41} & A_{42} & A_{43} & 0 \end{pmatrix}$$

$\Gamma(A):$ 

- ***Matrizes com diagonal não-principal nula e demais entradas positivas***. Dada uma matriz $A_{n\times n}$ com essa característica, basta verificarmos que $A_{ij} = 0$ se, e somente se $i + j = (n+1)$. Esse fato nos diz que, para cada i tem um único j que satisfaz essa igualdade, e portanto, o conjunto $\{1, 2, ..., n\}$ não pode ser dividido em dois subconjuntos I e J. Vale também salientar, que o grafo associado a uma matriz nessas condições, é fortemente conexo. Vejamos um exemplo.

$$A = \begin{pmatrix} A_{11} & A_{12} & A_{13} & 0 \\ A_{21} & A_{22} & 0 & A_{24} \\ A_{31} & 0 & A_{33} & A_{34} \\ 0 & A_{42} & A_{43} & A_{44} \end{pmatrix}$$

$\Gamma(A):$ 

2.3 Perron-Frobenius

O teorema de Perron-Frobenius emerge em diversos ramos da matemática, por exemplo nas teorias de automorfismos de grupos livres [BH12], de sistemas dinâmicos e ergódicos [PY98], Sec. 3.2; e de categorias tensoriais [EGNO15], Sec. 3.2. Nesta seção, utilizaremos os conceitos de matriz ≥ 0, positiva, de irredutibilidade e propriedades dessas matrizes, que tratamos na seção anterior, para a demonstração desse teorema que é o principal resultado desse capítulo.

Definição 2.3.1. *Seja A uma matriz irredutível ≥ 0, não-nula. O autovalor denotado por $PF(A)$ é chamado de* ***Autovalor Perron-Frobenius*** *de A. Se A é irredutível, nula de ordem 1, definimos $PF(A) = 0$. O vetor-coluna positivo $\vec{v}$ com $A\vec{v} = PF(A)\vec{v}$ é chamado de* ***Autovetor Direito Perron-Frobenius*** *de A.*

Teorema 2.3.1. ***(Perron-Frobenius)*** *Seja A uma matriz real ≥ 0 não-nula, irredutível. Então,*

(1) A tem um único autovetor > 0, real, a menos da multiplicação por um número real positivo; o autovalor $PF(A)$ associado é positivo;

(2) $PF(A) = PF(A^T)$;

(3) Para cada autovalor λ *de* A, *vale* $|\lambda| \leq PF(A)$;

(4) Para o autovalor $PF(A)$ *corresponde um único autovetor real à direita, a menos da multiplicação por um número real;*

Além disso, sendo $\vec{\omega}$ *um vetor-coluna real, não-nulo , não-negativo, e* α *um número real não-negativo, segue que*

(5) Se $A\vec{\omega} \leq PF(A)\vec{\omega}$, *então* $A\vec{\omega} = PF(A)\vec{\omega}$;

(6) Se $A\vec{\omega} \leq \alpha\vec{\omega}$, *então* $PF(A) \leq \alpha$.

Sugerimos a quem não queira ler a demonstração dele que prove-o, à guisa de exercício, no caso 2-por-2. Comecemos por demonstrar o lema a seguir.

Lema 2.3.2. *Seja* A *uma matriz irredutível* ≥ 0, *não-nula de tamanho* n*-por-*n. *Então a matriz* $B = \sum_{i=0}^{n-1} A^i$ *é positiva. Em particular, se* $x \in \mathbb{R}^n$, *com* $0 \neq x \geq 0$, *então* $Bx > 0$.

Demonstração: Note que cada entrada da diagonal principal de B é positiva, isto é, $B_{ii} > 0$. Note ainda que a matriz B pode ser expressa como

$$B = I_n + \sum_{i=1}^{n-1} A^i.$$

Devemos provar que $B_{ij} > 0$ para $i \neq j$. Como por hipótese, A é irredutível, não-negativa e não-nula, o grafo $\Gamma(A)$ de A é fortemente conexo, isto é, conforme provamos na Proposição 2.2.2, para quaisquer dois vértices distintos i, j existe um caminho orientado de i para j formado por arestas orientadas entre dois vértices consecutivos. Se todos esses vértices forem distintos, segue ainda da definição do grafo $\Gamma(A)$ de A que $A_{i_0 i_1} A_{i_1 i_2} ... A_{i_{k-1} i_k} > 0$. Portanto, $(A^k)_{ij} > 0$ e consequentemente, $B_{ij} > 0$ como queríamos provar. □

Definição 2.3.2. *Sejam* N *e* M *matrizes reais. Escrevemos* $N \leq M$ *para significar que* N *e* M *têm o mesmo tamanho e que* $N_{ij} \leq M_{ij}$ *para todo par de índices* (i, j). *Nós também dizemos que a matriz* N *é dominada pela matriz* M *se* $N \leq A$, *onde* A *é uma submatriz da matriz* M.

Definição 2.3.3. *Para* $\vec{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ *definimos* $|\vec{x}| = \sum_{i=1}^{n} |x_i|$ *que é denominada norma da soma. Definimos também* $\Delta = \{\vec{x} \in \mathbb{R}^n \mid \vec{x} \geq 0; |\vec{x}| = 1\}$.

Temos uma representação geométrica desse conjunto quando associamos seus elementos a pontos do plano ou do espaço, como um segmento de extremos $(1, 0)$ e $(0, 1)$ em $\mathbb{R}^2$, ou uma região triangular com vértices $(1, 0, 0)$, $(0, 1, 0)$ e $(0, 0, 1)$ em $\mathbb{R}^3$.

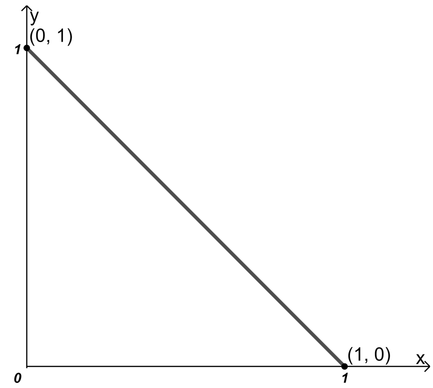


Figura 1 – *Representação geométrica de* Δ *em* $\mathbb{R}^2$

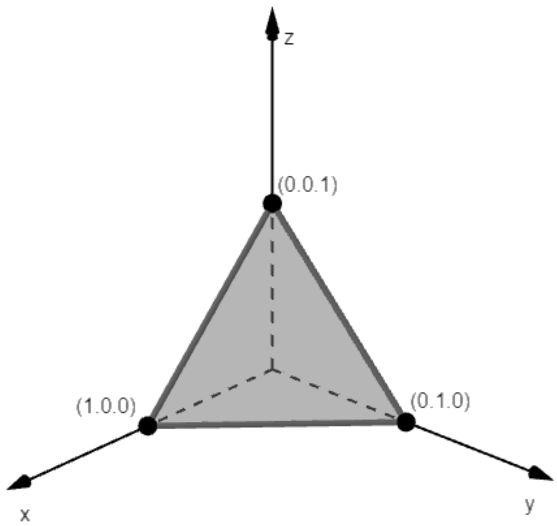


Figura 2 – *Representação geométrica de* Δ *em* $\mathbb{R}^3$

Deste ponto em diante, apresentamos uma demonstração do Teorema de Perron-Frobenius que segue uma ideia sugerida pelo matemático alemão Helmut Wielandt, em *Unzerlegbare nicht negative Matrizen*, cuja tradução livre é *"Matrizes não-negativas irredutíveis"*.

Demonstração Perron-Frobenius: (1) e (2) Primeiro provamos que existe um vetor-coluna $\vec{y} > 0$ e um número $\lambda > 0$, tal que, $A\vec{y} = \lambda\vec{y}$. Seja A uma matriz quadrada de ordem n. Seja $\vec{u}$ um vetor-linha de tamanho n com cada entrada igual a 1. Considerando um vetor $\vec{x} \in \Delta$, afirmamos que $\sup\{\rho \mid \textit{existe } \vec{x} \in \Delta, \textit{ tal que, } A\vec{x} \geq \rho\vec{x}\}$ é um número real. De fato, pois, se $A\vec{x} \geq \rho\vec{x}$ para algum $\vec{x} \in \Delta$, então, claramente teremos

$$\rho = \rho\vec{u}\vec{x} \leq \vec{u}A\vec{x} \leq \vec{u}A(\vec{u})^T.$$

Denotando o supremo acima por λ, segue do fato de Δ ser compacto que existe um vetor $\vec{y} \in \Delta$, tal que, $A\vec{y} \geq \lambda\vec{y}$. Suponhamos por contradição, que $A\vec{y} \neq \lambda\vec{y}$. Então,

$BA\vec{y} > \lambda B\vec{y}$, onde $B = I_n + \sum_{i=1}^{n-1} A^i$ é positiva conforme provamos no Lema 2.3.2. Assim, sendo fato que $AB = BA$, conseguimos obter $A\vec{x} > \lambda\vec{x}$ para $\vec{x} = \dfrac{B\vec{y}}{|B\vec{y}|} \in \Delta$, entretanto, isso contradiz a maximalidade de λ. Consequentemente, $B\vec{y} = \sum_{i=0}^{n-1} \lambda^i \vec{y}$, e sendo $B\vec{y} > 0$ obtemos também do lema anterior que $\vec{y} > 0$. Novamente, de $A\vec{y} = \lambda\vec{y}$ segue que $\lambda > 0$. Por simetria, existe um vetor-linha $\vec{z} > 0$ e um número real $\mu > 0$, tais que, $\vec{z}A = \mu\vec{z}$. Então, $\mu\vec{z}\vec{y} = \vec{z}A\vec{y} = \lambda\vec{z}\vec{y}$ e, $\vec{z}\vec{y} > 0$, donde segue que $\mu = \lambda$.

Agora, seja $\vec{y'}$ um autovetor positivo arbitrário de A, e seja λ' o autovalor associado. Como anteriormente, obteremos $\mu = \lambda'$, e portanto, $\lambda = \lambda'$. Suponhamos que $\vec{y'}$ não seja um múltiplo escalar de $\vec{y}$. Então os pontos $\dfrac{\vec{y'}}{|\vec{y'}|}$ e $\dfrac{\vec{y}}{|\vec{y}|}$ pertencem a Δ, são distintos e a reta que os contém intercepta a borda de Δ em algum ponto $\vec{v}$. Como $\vec{v}$ é autovetor de A, e portanto, de B, concluímos que $B\vec{v}$ é um múltiplo escalar de $\vec{v}$. Assim, alguma coordenada do vetor $\vec{Bv}$ é nula, o que contradiz o Lema 2.3.2. Novamente, obtemos então que $\vec{y'}$ é um múltiplo escalar e assim, fica provado (1). Da identidade acima $\mu = \lambda$, obtemos a prova da afirmação (2). Em particular, como A e A^T têm o mesmo polinômio característico, possuem os mesmos autovalores.

(3) Seja $A\vec{u} = \tau\vec{u}$ com $\tau \in \mathbb{C}$ e $\vec{u} \in \mathbb{C}^n$, para $\vec{u} \neq 0$. Defina $\vec{u'} = (|\, u_1 \,|, \ldots, |\, u_n \,|)^T$. Então $A\vec{u'} \geq |\, \tau \,|\, \vec{u'}$. Seja $\vec{a} > 0$ o vetor-linha, tal que, $\vec{a}A = \lambda\vec{a}$ onde $\lambda = PF(A)$. Então $\lambda\vec{a}\vec{u'} = \vec{a}A\vec{u'} \geq |\, \tau \,|\, \vec{a}\vec{u'}$ e $\vec{a}\vec{u'} > 0$, de modo que $\lambda \geq |\, \tau \,|$.

(4) Seja $\vec{b}$ um autovetor positivo e $\vec{c}$ um autovetor arbitrário associado a $PF(A)$. Então para um número suficientemente grande $r > 0$ o vetor $\vec{c} + r\vec{b}$ também é um autovetor positivo associado ao autovalor $PF(A)$. Pela afirmação (1) este vetor, e portanto o vetor $\vec{c}$ é um múltiplo escalar do vetor $\vec{b}$.

(5) Suponha que $A\vec{\omega} \leq \lambda\vec{\omega}$ e $A\vec{\omega} \neq \lambda\vec{\omega}$, onde $\lambda = PF(A)$. Seja $\vec{z} > 0$ um vetor-linha tal que, $\vec{z}A = \lambda\vec{z}$. Então $\lambda\vec{z}\vec{\omega} = \vec{z}A\vec{\omega} < \lambda\vec{z}\vec{\omega}$ é uma contradição. Portanto, se $A\vec{\omega} \leq \lambda\vec{\omega}$ então $A\vec{\omega} = \lambda\vec{\omega}$.

(6) Suponhamos que $A\vec{\omega} \leq \alpha\vec{\omega}$ e $\lambda > \alpha$, onde $\lambda = PF(A)$. Tomando o vetor-linha $\vec{z} > 0$ tal que $\vec{z}A = \lambda\vec{z}$, segue que $\lambda\vec{z}\vec{\omega} = \vec{z}A\vec{\omega} \leq \alpha\vec{z}\vec{\omega}$, o que é um absurdo. Logo, se $A\vec{\omega} \leq \alpha\vec{\omega}$, então $PF(A) \leq \alpha$, como queríamos demonstrar. □

Teorema 2.3.3. *Sejam M e M_1 matrizes quadradas reais, ≥ 0, e seja M_1 irredutível e dominada por M. Suponha que $M\vec{w} \leq \lambda\vec{w}$ para algum número $\lambda > 0$ e vetor $\vec{w} > 0$. Então ou $PF(M_1) < \lambda$ ou $PF(M_1) = \lambda$ e, a menos da conjugação por uma matriz de permutação,*

$$M = \begin{pmatrix} M_1 & 0 \\ C & D \end{pmatrix}.$$

Demonstração: Sem perda de generalidade, vamos supor que

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix},$$

e $M_1 \leq A$. Pondo $\vec{w} = \begin{pmatrix} \vec{u} \\ \vec{v} \end{pmatrix}$, onde o tamanho de $\vec{u}$ corresponde ao tamanho da submatriz A. Então,

$$\lambda \begin{pmatrix} \vec{u} \\ \vec{v} \end{pmatrix} \geq M \begin{pmatrix} \vec{u} \\ \vec{v} \end{pmatrix} = \begin{pmatrix} A\vec{u} + B\vec{v} \\ C\vec{u} + D\vec{v} \end{pmatrix},$$

Daí,

$$M_1\vec{u} \leq A\vec{u} \leq A\vec{u} + B\vec{v} \leq \lambda\vec{u}.$$

Como $M_1\vec{u} \leq \lambda\vec{u}$, então segue da afirmação (6) do Teorema 2.3.1 que $PF(M_1) \leq \lambda$. Caso $PF(M_1) = \lambda$, então pela afirmação (5), $M_1\vec{u} = \lambda\vec{u}$. Portanto, $B\vec{v} = 0$. Sendo $\vec{v} > 0$, como $B \geq 0$ então $B = 0$. □

Corolário 2.3.4. *Sejam M e M_1 matrizes quadradas reais, irredutíveis, ≥ 0, e sendo M_1 dominada por M e $M \neq M_1$. Então $PF(M_1) < PF(M)$.*

Demonstração: Pelo Teorema 2.3.3 anterior, segue diretamente que $M_1\vec{u} \leq \lambda\vec{u}$. Suponha que $PF(M_1) = \lambda$. Então teríamos $M = \begin{pmatrix} A & 0 \\ C & D \end{pmatrix}$, e isso contraria a hipótese de M ser irredutível. Portanto, $PF(M_1) < PF(M)$ como queríamos mostrar. □

Teorema 2.3.5. *Seja A uma matriz ≥ 0, não-nula, irredutível, com entradas inteiras, e autovalor Perron-Frobenius λ. Então, são válidos*

(a) $\lambda \geq 1$;

(b) Se $\lambda = 1$, então A é uma matriz de permutação;

(c) $A_{ij} \leq \lambda^n$ para todo i, j em $\{1, \ldots, n\}$.

Demonstração: Pelo Teorema de Perron-Frobenius, existe um vetor-coluna $\vec{v} > 0$ tal que, $A\vec{v} = \lambda\vec{v}$. Seja $\vec{u}$ o vetor-linha de tamanho n e cada entrada igual a 1. Assim, para provarmos *(a)* e *(b)* calculamos diretamente

$$\vec{u}A\vec{v} = \lambda\vec{u}\vec{v} = \left(\sum_{i=1}^{n} A_{i1}\right) v_1 + \cdots + \left(\sum_{i=1}^{n} A_{in}\right) v_n = \lambda v_1 + \cdots + \lambda v_n,$$

onde notamos que para cada j a soma $\sum_{i=1}^{n} A_{ij}$ é pelo menos 1, pois, a matriz A é não-negativa, não-nula, irredutível e tem entradas inteiras. Portanto, $\lambda \geq 1$. Se $\lambda = 1$, então cada uma dessas somas será igual a 1. Dessa forma, cada coluna da A é constituída inteiramente de zeros, com exceção de uma única entrada igual a 1. Com isso, A não contém linha nula e é uma matriz de permutação.

Para *(c)*, fixamos $i, j \in \{1, \ldots, n\}$. Na demonstração do Lema 2.3.2 mostramos que existe um número natural k tal que $(A^k)_{ij} > 0$; além disso, $0 \leq k \leq n-1$. Visto que A tem suas entradas inteiras, segue que $(A^k)_{ij} \geq 1$. De $A^k\vec{v} = \lambda^k\vec{v}$ temos

$$v_j \leq (A^k)_{ij} v_j \leq \lambda^k v_j.$$

Ademais, de $A\vec{v} = \lambda\vec{v}$ deduzimos que

$$A_{ij} v_j \leq \lambda v_j \leq \lambda^{k+1} v_j \leq \lambda^n v_j,$$

o que prova *(c)*. $\square$

Corolário 2.3.6. *Seja r um número real e n um número natural. Então existe apenas um número finito de matrizes quadradas ≥ 0, irredutíveis, com entradas inteiras e autovalores Perron-Frobenius não excedendo r.*

Demonstração: De fato, pelo teorema anterior como cada entrada de uma tal matriz A é inteira, ela está no intervalo $[0, r^n]$. $\square$

3 Aplicações

Neste capítulo, apresentaremos 4 exemplos com problemas onde podemos verificar aplicações das matrizes de Markov e do Teorema de Perron-Frobenius. São eles: **Problema 4 da 23º OBM**; **Problema dos n mentirosos**; **O PageRank** e **O Modelo de Difusão de Ehrenfest**.

3.1 XXIII OBM - Olimpíada Brasileira de Matemática

Um ratinho ocupa inicialmente a gaiola A e é treinado para mudar de gaiola atravessando um túnel sempre que soa um alarme. Cada vez que soa o alarme o ratinho escolhe qualquer um dos túneis incidentes a sua gaiola com igual probabilidade e sem ser afetado por escolhas anteriores. Qual é a probabilidade de que após o alarme soar 23 vezes, o ratinho ocupe a gaiola B?

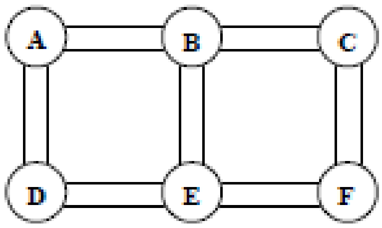


Figura 3 – *Um esquema das gaiolas*

Este é o problema 4 proposto na 23ª Olimpíada Brasileira de Matemática, [Eur02].

Resolução: Diferentemente do que é apresentado nas resoluções já publicadas anteriormente, utilizaremos a matemática subjacente a partir do instrumental referente à Cadeia de Markov [Dia09] e diagonalização de matrizes. Para tanto, começamos por considerar o conjunto $K = \{1,2,3,4,5,6\}$, que representa os estados de transições ordenadas das gaiolas de A até F, respectivamente. De acordo com as informações do problema, "cada vez que o alarme soa o ratinho escolhe qualquer um dos túneis incidentes a sua gaiola com igual probabilidade e sem ser afetado por escolhas anteriores", isso então nos diz que as transições satisfazem a uma Cadeia de Markov. A partir das disposições das gaiolas, verificamos as probabilidades condicionais $P(X_{n+1} = j | X_n = i)$ associadas às transições, e disso, os valores $p = \frac{1}{2}$ quando estiver saindo das gaiolas A, C, F e D, e, $q = \frac{1}{3}$ quando estiver saindo das gaiolas B e E.
Com isso, podemos construir a seguinte matriz M de transição:

$$M = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}$$

onde M^Té uma Matriz de Markov. Considerando $\vec{M_0} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$ o vetor-linha estocástico inicial de distribuição de probabilidade (posto que o ratinho está inicialmente na gaiola Ⓐ) precisamos calcular $M_{12}^{(23)}$, isto é, a probabilidade do ratinho ocupar a gaiola B após o alarme soar 23 vezes, e que por sua vez, é o mesmo que a segunda coordenada do vetor $\vec{M_0}M^{23}$. Para isso, devemos calcular a potência M^{23} o que pode parecer intimidador, computacionalmente. Entretanto, se pudermos lançar mão da decomposição da matriz M na forma $M = SDS^{-1}$, onde a matriz S é invertível e a matriz D é diagonal, obteremos então, com facilidade

$$M^{23} = (SDS^{-1})^{23} = SD^{23}S^{-1} = S. \begin{pmatrix} D_{11}^{(23)} & 0 & 0 & 0 & 0 & 0 \\ 0 & D_{22}^{(23)} & 0 & 0 & 0 & 0 \\ 0 & 0 & D_{33}^{(23)} & 0 & 0 & 0 \\ 0 & 0 & 0 & D_{44}^{(23)} & 0 & 0 \\ 0 & 0 & 0 & 0 & D_{55}^{(23)} & 0 \\ 0 & 0 & 0 & 0 & 0 & D_{66}^{(23)} \end{pmatrix} .S^{-1}$$

uma vez que, para elevar uma matriz diagonal a uma potência n basta elevar cada entrada de sua diagonal principal a n. Assim, a matriz M será diagonalizável e, as entradas da diagonal principal da matriz D serão os autovalores de M. Uma pergunta surge naturalmente: A matriz M é diagonalizável? De acordo com o que mostramos em Seção 1.3 a resposta é sim, pois, M tem seis autovalores distintos. Considerando os vetores-linha de M, notamos que para cada um deles a soma de suas entradas é igual 1. Esse fato, de imediato nos dá que $\lambda_1 = 1$ é um autovalor de M que tem como autovetor direito associado $\vec{v_1} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$. Do fato do traço da matriz M ser nulo, isso nos garante que, ou -1 é um dos outros cinco autovalores, ou, é a soma de ao menos, dois outros. Buscando os demais autovalores encontramos o polinômio característico

$$det(\lambda I_6 - M) = \begin{vmatrix} \lambda & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \lambda & \frac{1}{3} & 0 & \frac{1}{3} & 0 \\ 0 & \frac{1}{2} & \lambda & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \lambda & \frac{1}{2} & 0 \\ 0 & \frac{1}{3} & 0 & \frac{1}{3} & \lambda & \frac{1}{3} \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & \lambda \end{vmatrix} = \lambda^6 - \frac{23}{18}\lambda^4 + \frac{41}{144}\lambda^2 - \frac{1}{144}$$

o qual notamos facilmente a partir dos expoentes pares da variável λ, que tem raízes com valores simétricos. Podemos, então, afirmar que $\lambda_2 = -1$ é outro autovalor de M. Calculando as demais raízes do polinômio equivalente $144\lambda^6 - 184\lambda^4 + 41\lambda^2 - 1 = 0$, estas são $\lambda = \pm\frac{1}{2}$ e $\lambda = \pm\frac{1}{6}$. Com os demais autovetores, obtemos:

- $\lambda_1 = 1$ associado a $\vec{v_1} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$;
- $\lambda_2 = -1$ associado a $\vec{v_2} = \begin{pmatrix} -1 & 1 & -1 & 1 & -1 & 1 \end{pmatrix}$;
- $\lambda_3 = \dfrac{1}{2}$ associado a $\vec{v_3} = \begin{pmatrix} -1 & 0 & 1 & -1 & 0 & 1 \end{pmatrix}$;
- $\lambda_4 = -\dfrac{1}{2}$ associado a $\vec{v_4} = \begin{pmatrix} 1 & 0 & -1 & -1 & 0 & 1 \end{pmatrix}$;
- $\lambda_5 = -\dfrac{1}{6}$ associado a $\vec{v_5} = \begin{pmatrix} 1 & -\frac{4}{3} & 1 & 1 & -\frac{4}{3} & 1 \end{pmatrix}$;
- $\lambda_6 = \dfrac{1}{6}$ associado a $\vec{v_6} = \begin{pmatrix} -1 & -\frac{4}{3} & -1 & 1 & \frac{4}{3} & 1 \end{pmatrix}$.

A partir desses resultados, e do fato de que todos os autovalores são distintos e com multiplicidade algébrica 1, e que portanto, seus autovetores são linearmente independentes, a matriz M é diagonalizável. Formando uma matriz S com os seis autovetores de M sendo seus vetores-coluna, temos de imediato que S é invertível. Com isso, encontramos a decomposição espectral de $M = SDS^{-1}$ como segue,

$$\begin{pmatrix} -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 0 & -\frac{4}{3} & -\frac{4}{3} & 0 & 1 \\ -1 & -1 & 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & 1 & -1 & 1 \\ -1 & 0 & -\frac{4}{3} & \frac{4}{3} & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{6} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{6} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} -\frac{1}{7} & \frac{3}{14} & -\frac{1}{7} & \frac{1}{7} & -\frac{3}{14} & \frac{1}{7} \\ \frac{1}{4} & 0 & -\frac{1}{4} & -\frac{1}{4} & 0 & \frac{1}{4} \\ \frac{3}{28} & -\frac{3}{14} & \frac{3}{28} & \frac{3}{28} & -\frac{3}{14} & \frac{3}{28} \\ -\frac{3}{28} & -\frac{3}{14} & -\frac{3}{28} & \frac{3}{28} & \frac{3}{14} & \frac{3}{28} \\ -\frac{1}{4} & 0 & \frac{1}{4} & -\frac{1}{4} & 0 & \frac{1}{4} \\ \frac{1}{7} & \frac{3}{14} & \frac{1}{7} & \frac{1}{7} & \frac{3}{14} & \frac{1}{7} \end{pmatrix}.$$

Finalmente, calculamos

$$\vec{M_0}.M^{23} = \left(0 \quad \frac{3}{7} \cdot \left(\frac{6^{23}+1}{6^{23}}\right) \quad 0 \quad \frac{2}{7} + \frac{1}{2^{24}} - \frac{3}{14} \cdot \frac{1}{6^{23}} \quad 0 \quad \frac{2}{7} + \frac{1}{2^{24}} - \frac{3}{14} \cdot \frac{1}{6^{23}}\right)$$

donde tiramos que $M_{12}^{(23)} = \frac{3}{7} \cdot \left(\frac{6^{23}+1}{6^{23}} \right) \cong 0,4285$, isto é, uma probabilidade aproximada de $42,85\%$. □

3.2 O Problema dos n mentirosos

O Problema dos n mentirosos e a Cadeia de Markov [Fel67] *"Se cada uma das pessoas A, B, C, D dizem a verdade uma vez em três vezes (independentemente), e A afirma que B nega que C declara que D é um mentiroso, qual a probabilidade de que D estava dizendo a verdade?"*

Este problema foi primeiramente tratado por A.S. Eddington e teve sua solução publicada em Monthly [4288, Vol.57(1950) pp. 43-45].

Resolução: Note inicialmente que, existem apenas oito declarações distintas que podem ser feitas pela pessoa A. Na formulação original, as pessoas C, B e A declaram sucessivamente três afirmações que podem ser verdadeiras ou falsas, e que podem se contradizer. Tem-se ainda que, apenas D e C conhecem o estado inicial. No entanto, tomado pelo valor nominal, cada afirmação implicaria ou que D está dizendo a verdade ou que ele mente. Assim, por exemplo, se B negar que C afirma que D é um mentiroso, ele implica que D diz a verdade. Se, por outro lado, A nega que B negou, etc., então isso implica que D é um mentiroso. Utilizaremos uma terminologia neutra, e falaremos de um processo casual com dois estados possíveis onde, eventualmente, o estado observado é 1 ou 2 conforme a última afirmação de que D é honesto ou mentiroso. Inicialmente, ou seja, no instante 0, o estado observado é 1 ou 2 de acordo com o que D diz verdade ou mentira, respectivamente, e, cada sequência de amostra possível do processo acaba sendo representada por uma sucessão dos dígitos 1 e 2.

Fundamentalmente, o processo de Markov caracteriza-se pelo fato de que cada pessoa conhece apenas a afirmação da última falante. Ainda, no instante n o estado observado muda ou permanece o mesmo de acordo com a n-ésima falante diz a verdade ou mente. Nessas condições, obtemos um modelo que representa uma Cadeia de Markov mais simples:

- Temos dois estados possíveis 1 e 2 nos quais, inicialmente, as suas respectivas probabilidades são α e β, naturalmente com $\alpha + \beta = 1$. Independentemente do que ocorra até o tempo n, temos probabilidade p de que exatamente no tempo n, o estado observado não sofra alteração, e $q = (1-p)$ de que haja alteração. Buscamos as probabilidades condicionais x_n e y_n que o processo realmente começou a partir do estado 1 dado que, no tempo n o estado observado é, respectivamente, 1 ou 2. Na versão original, tem-se $\alpha = p = \frac{1}{3}$, $n = 3$ e somente x_n é requerida;

- Em cada etapa temos quatro transições possíveis $1 \mapsto 1$, $1 \mapsto 2$, $2 \mapsto 1$, $2 \mapsto 2$ e as correspondentes probabilidades de transição, por suposição, são $p_{11} = p_{22} = p$ e $p_{12} = p_{21} = (1-p)$. Assim, supondo que em um determinado momento o sistema está num estado j e sendo $p_{jk}^{(n)}$ a probabilidade de que n etapas depois, o estado observado seja k, temos que $p_{jk}^{(n)}$ é chamada a probabilidade de transição $j \mapsto k$ da n-ésima etapa, onde claramente $p_{jk}^{(1)} = p_{jk}$ e, a partir das possibilidades de transição do estado j para o estado k em duas etapas e de suas respectivas probabilidades,
$$p_{jk}^{(2)} = p_{j1}p_{1k} + p_{j2}p_{2k}. \tag{3.1}$$
O que de modo geral, a partir de igualdades recursivas, pode ser expresso por,
$$p_{jk}^{(n+1)} = p_{j1}^{(n)}p_{1k} + p_{j2}^{(n)}p_{2k}. \tag{3.2}$$
As igualdades (3.1) e (3.2) acima podem ser obtidas de forma direta por meio das potências da matriz
$$P = \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix} = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}. \tag{3.3}$$

- Sendo α e β as probabilidades iniciais respectivas aos estados 1 e 2, então, denotando por $e_k^{(n)}$ a probabilidade de observar no instante n o estado k, segue que,
$$e_k^{(n)} = \alpha p_{1k}^{(n)} + \beta p_{2k}^{(n)}. \tag{3.4}$$
Encontramos, portanto, para as probabilidades condicionais x_n e y_n, que
$$x_n = \frac{\alpha p_{11}^{(n)}}{e_1^{(n)}}, \quad y_n = \frac{\alpha p_{12}^{(n)}}{e_2^{(n)}}. \tag{3.5}$$

- Retomando à matriz P em (3.3). Precisamos calcular P^n, e para isso, conforme utilizamos na resolução do problema (3.1), escreveremos a sua decomposição espectral. Como P é uma matriz de Markov, e seus vetores-linha tem a soma de suas coordenadas iguais a 1, temos que $\lambda_1 = 1$ é um autovalor associado ao autovetor $\vec{v_1} = \begin{pmatrix} 1 & 1 \end{pmatrix}$. Buscando o outro autovalor e autovetor associado encontramos, $\lambda_2 = (2p-1)$ e $\vec{v_2} = \begin{pmatrix} 1 & -1 \end{pmatrix}$. Assim, escrevemos a decomposição
$$P = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 2p-1 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Consequentemente,

$$P^n = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} . \begin{pmatrix} (2p-1)^n & 0 \\ 0 & 1 \end{pmatrix} . \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \dfrac{(2p-1)^n+1}{2} & \dfrac{1-(2p-1)^n}{2} \\ \dfrac{1-(2p-1)^n}{2} & \dfrac{(2p-1)^n+1}{2} \end{pmatrix}. \tag{3.6}$$

Finalmente, para explicitarmos as probabilidades condicionais conforme exibidas em (3.5) basta substituirmos as expressões (3.6) e (3.4) nesta igualdade, o que nos dá,

$$x_n = \frac{\alpha.[(2p-1)^n+1]}{1+(2p-1)^n.(\alpha-\beta)}, \quad y_n = \frac{\alpha.[1-(2p-1)^n]}{1-[(2p-1)^n.(\alpha-\beta)]}. \tag{3.7}$$

Como na versão original, estamos interessados em determinar x_n para $n = 3$ e $\alpha = p = \dfrac{1}{3}$. Substituindo os valores em questão, concluímos que a probabilidade procurada é $x_3 = \dfrac{13}{41} \cong 31,71\%$. Além disso, temos $y_n = \dfrac{7}{20}$. Convém ainda verificarmos que, como $0 < p < 1$, então claramente $x_n \to \alpha$ conforme $n \to \infty$, isto é, a probabilidade condicional de que o processo começou a partir do estado 1, dado que para um número suficientemente grande de etapas o estado observado é novamente 1, converge para a probabilidade inicial α.

Mentiroso Preferencial. Até aqui, as chances de uma pessoa dizer a verdade não dependem da afirmação que ele deve transmitir. Suponhamos então, que cada pessoa tem a preferência de afirmar que D diz a verdade. Assim, uma transição $1 \mapsto 1$ é mais provável que $2 \mapsto 2$, enquanto $1 \mapsto 2$ é menos provável que $2 \mapsto 1$. Podemos tratar este caso geral da mesma forma que tratamos o problema específico anterior. Para isso, façamos $p_{11} = p$, $p_{12} = 1-p$, $p_{21} = 1-p'$ e $p_{22} = p'$. Naturalmente, todas as igualdades acima são aplicadas, exceto a igualdade (3.6) que deve ser substituída por

$$P^n = \begin{pmatrix} p & 1-p \\ 1-p' & p' \end{pmatrix}^n = \begin{pmatrix} \dfrac{(p+p'-1)^n(p-1)+p'-1}{p+p'-2} & \dfrac{(p+p'-1)^n(1-p)+p-1}{p+p'-2} \\ \dfrac{(p+p'-1)^n(1-p')+p'-1}{p+p'-2} & \dfrac{(p+p'-1)^n(p'-1)+p-1}{p+p'-2} \end{pmatrix}.$$

Nessas condições, o resultado final é agora

$$x_n = \frac{\alpha \cdot [(1-p')+(1-p)\cdot(p+p'-1)^n]}{(1-p')+[\alpha(1-p)-\beta(1-p')]\cdot(2p'-1)^n}$$

□

3.3 O PageRank

Um conjunto de nós com conexões é um grafo. Qualquer rede pode ser descrita por um grafo. A estrutura de links da web forma um grafo, onde os sites individuais são os nós e existe uma seta do site a_i para o site a_j se a_i se liga a a_j. A matriz de adjacência A deste grafo é chamada de grafo da web. Se houver n sites, então a matriz de adjacência é uma matriz $n \times n$ com entradas $A_{ij} = 1$ se existir um link de a_i para a_j. Se dividirmos cada coluna pela quantidade de números 1 dessa coluna, obtemos uma matriz de Markov A que é chamada de matriz da web normalizada. Defina a matriz E que satisfaz $E_{ij} = \dfrac{1}{n}$ para todo i, j. Os estudantes de pós-graduação e posteriores empresários Sergey Brin e Lawrence Page tiveram em 1996 a seguinte ideia de um bilhão de dólares:

A matriz do Google é a matriz $G = dA + (1-d)E$, onde $0 < d < 1$ é um parâmetro chamado fator de amortecimento e A é a matriz de Markov obtida da matriz de adjacência escalonando as linhas para se tornar matriz estocástica. Esta é uma matriz de Markov $n \times n$ com autovalor 1. Seu autovetor Perron-Frobenius $\vec{v}$ escalado de forma que o maior valor seja 10 é chamado de **PageRank** do fator de amortecimento d. A equação do PageRank é $[dA + (1-d)E]\vec{v} = \vec{v}$.

O fator de amortecimento pode parecer um pouco misterioso. Brin e Page escreveram: "O PageRank pode ser também considerado um modelo de comportamento do usuário. Assumimos que há um 'surfista aleatório' que recebe uma página da Web aleatoriamente e continua clicando nos links, nunca atingindo 'de volta', mas eventualmente fica entediado e começa em outra página aleatória. A probabilidade de que o surfista aleatório visite uma página é seu PageRank. O fator de amortecimento é a probabilidade em cada página de o 'surfista aleatório' ficar entediado e solicitar outra página aleatória. Uma variação importante é adicionar apenas o fator de amortecimento d a uma única página ou a um grupo de páginas. Isso permite personalização e pode tornar quase impossível enganar deliberadamente o sistema para obter uma classificação mais alta. Temos várias outras extensões para o PageRank."

Vamos resolver um caso simples que ilustra essas condições.

Considere uma rede com quatro sites S_1, S_2, S_3 e S_4 conectados por meio de links conforme ilustrado no grafo a seguir.

$$\begin{tikzcd}
S_1 \arrow[r, leftrightarrow] \arrow[d] \arrow[dr, leftrightarrow] & S_3 \\
S_2 \arrow[r] \arrow[ur] & S_4 \arrow[u]
\end{tikzcd}$$

Vamos encontrar o pagerank para o fator $d = 0{,}1$.

Resolução: A matriz de adjacência do grafo acima e sua matriz da web normalizada são respectivamente,

$$M = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix}, \quad e \quad A = \begin{pmatrix} 0 & 1 & \frac{1}{3} & \frac{1}{2} \\ 0 & 0 & \frac{1}{3} & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{3} & 0 \end{pmatrix}.$$

Com isso, e tendo ainda a matriz

$$E=\begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4}\\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4}\\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4}\\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix},$$

obtemos a matriz do Google

$$G=\frac{1}{10}\cdot\begin{pmatrix} 0 & 1 & \frac{1}{3} & \frac{1}{2}\\ 0 & 0 & \frac{1}{3} & \frac{1}{2}\\ \frac{1}{2} & 0 & 0 & 0\\ \frac{1}{2} & 0 & \frac{1}{3} & 0 \end{pmatrix}+\frac{9}{10}\cdot\begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4}\\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4}\\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4}\\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}=\begin{pmatrix} \frac{9}{40} & \frac{13}{40} & \frac{31}{120} & \frac{11}{40}\\ \frac{9}{40} & \frac{9}{40} & \frac{31}{120} & \frac{11}{40}\\ \frac{11}{40} & \frac{9}{40} & \frac{9}{40} & \frac{9}{40}\\ \frac{11}{40} & \frac{9}{40} & \frac{31}{120} & \frac{9}{40} \end{pmatrix}.$$

O Pagerank do fator de amortecimento d será o autovetor Perron-Frobenius $\vec{v}$, tal que, $G\vec{v}=\vec{v}$ e sua maior entrada será 10. Assim, pondo $\vec{v}=\begin{pmatrix} x\\ y\\ z\\ w \end{pmatrix}$ obtemos o sistema

$$\begin{cases} \frac{9}{40}x+\frac{13}{40}y+\frac{31}{120}z+\frac{11}{40}w=x\\ \frac{9}{40}x+\frac{9}{40}y+\frac{31}{120}z+\frac{11}{40}w=y\\ \frac{11}{40}x+\frac{9}{40}y+\frac{9}{40}z+\frac{9}{40}w=z\\ \frac{11}{40}x+\frac{9}{40}y+\frac{31}{120}z+\frac{9}{40}w=w \end{cases}$$

que possui infinitas soluções, sendo que, a sua solução relevante para o fator d é $\vec{v}=\begin{pmatrix} 10\\ \frac{443100}{48741}\\ \frac{63300}{7161}\\ \frac{2110}{231} \end{pmatrix}$ que é o seu Pagerank, e nos diz que o site S_1 é o mais acessado diretamente ou, por meio de links nos sites S_2, S_3 e S_4 que levam o usuário até S_1.

Esse mesmo caso visto sob outra interessante abordagem nos dá a distribuição das probabilidades de acesso aos sites S_1, S_2, S_3 e S_4. Para este fim, consideremos a matriz de transição P de uma cadeia de Markov, associada às probabilidades do usuário estando no site i acessar o site j,

$$P=\begin{pmatrix} 0 & 0 & 1 & \frac{1}{2}\\ \frac{1}{3} & 0 & 0 & 0\\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2}\\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{pmatrix}.$$

Seu autovetor Perron-Frobenius é a sua distribuição de equilíbrio estável. Nesse contexto, as entradas v_k desse autovetor $\vec{v}$, tais que, $\sum_{k=1}^{4} v_k = 1$ são as probabilidades que buscamos e revelam qual site tem o maior índice de acessos. Assim, o sistema

$$\begin{pmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix}$$

tem infinitas soluções caracterizadas por $t \cdot \begin{pmatrix} 2 \\ \frac{2}{3} \\ \frac{3}{2} \\ 1 \end{pmatrix}$, obviamente para $t \in \mathbb{R}$.

Escolhendo $t = \dfrac{6}{31}$ de modo conveniente a satisfazer a condição dada, obtemos

$$\vec{v} = \begin{pmatrix} \frac{12}{31} \\ \frac{4}{31} \\ \frac{9}{31} \\ \frac{6}{31} \end{pmatrix}.$$

Daí, temos que as probabilidades de acesso aos respectivos sites são $S_1 = \frac{12}{31} \cong 38,71\%$, $S_2 = \frac{4}{31} \cong 12,9\%$, $S_3 = \frac{9}{31} \cong 29,03\%$ e $S_4 = \frac{6}{31} \cong 19,35\%$, donde concluímos que quanto mais a rede é acessada, o site S_1 tem maior probabilidade de ser visitado com $38,71\%$. □

3.4 O Modelo de Difusão de Ehrenfest

Este modelo foi estudado e apresentado pelos físicos teóricos Tatyana e Paul Ehrenfest ainda na primeira década do século XX. Trata-se da sua versão considerada básica e fornece explicações relevantes fundamentadas na mecânica estatística - que tem como objetivo estudar e explicar o comportamento macroscópico a partir das leis microscópicas - para um problema inquietante, o da irreversibilidade. As leis da física baseiam-se em mecânica quântica (que é probabilística), portanto, um tratamento estocástico faz-se necessário nesse contexto. O casal Ehrenfest conseguiu 20 anos antes do desenvolvimento da mecânica quântica moderna, introduzir um modelo probabilístico para descrever a origem da irreversibilidade em sistemas físicos. Este modelo vem sendo usado extensivamente para estudar aplicações do Teorema H de Boltzmann (sobre a entropia termodinâmica e o caos molecular) e entender como a simetria do tempo subjacente nas leis do movimento geram a assimetria de tempo dos processos de difusão.

Considere um sistema formado por dois recipientes idênticos A e B ligados por uma pequena abertura, e que o recipiente A contém um gás constituído por um número finito de moléculas indistinguíveis. Tais moléculas transitam de A para B regidas pelas leis da mecânica e da termodinâmica até que num dado momento, os recipientes encontram-se em equilíbrio macroscópico (quando suas pressões internas são iguais). Considere também que não há interferências externas ao sistema e que, as velocidades e as posições das moléculas no interior dos recipientes são irrelevantes. Qual a probabilidade de mesmo num instante de tempo arbitrariamente remoto, cada molécula regressar ao recipiente A? Caso esse regresso ocorra, é possível determinar o tempo necessário?.

***Resolução**:* Como neste trabalho não há a pretensão de apresentar e nem discutir os fundamentos da mecânica estatística, apresentaremos de forma introdutória um modelo de evolução estocástica que responde satisfatoriamente ao modelo em questão.

Seja m o número total de moléculas. Estamos interessados em observar do ponto de vista probabilístico o trânsito das moléculas, então, consideraremos que apenas uma molécula passa de um recipiente para o outro a cada instante de tempo, e que todas as moléculas têm a mesma probabilidade de passar de um recipiente para o outro. Temos um número fixo de moléculas, por isso, deduzimos o número de moléculas de um recipiente a partir do número de moléculas do outro. O estado desse sistema será determinado por

$$X_n := \text{número de moléculas em } B \text{ no instante } n,$$

onde, como o recipiente B pode conter 0, 1, 2, etc. até no máximo m moléculas, então, para $E_m := \{0, 1, 2, \ldots, m\}$, que é o espaço dos estados, temos $X_n \in E_m$. Suporemos a condição inicial $X_0 = 0$, isto é, todas as moléculas estão no recipiente A. Precisamos determinar a evolução do número de moléculas em B em função do tempo n, ou seja, X_0, X_1, X_2, $\ldots$, a partir do fato de que apenas uma molécula passa de um recipiente para o outro. Nessas condições, temos $X_{n+1} = X_n \pm 1$, e assim, quando $X_n = k$ temos k moléculas em B e consequentemente, $m - k$ em A. Dessa forma, no tempo $n + 1$ a molécula estará ou em $k - 1$: uma molécula passou de B para A; ou em $k + 1$: uma molécula passou de A para B.

Até aqui, temos uma Cadeia de Markov com $m + 1$ estados $E_{(0)} = 0$, $E_{(1)} = 1$, $\cdots$, $E_{(m)} = m$ e transições possíveis apenas para os estados, imediatamente, anterior e posterior, pois, em cada instante uma molécula move-se aleatoriamente de seu recipiente para o outro enquanto o estado do sistema é determinado pelo número de moléculas em B. Para X_1, o primeiro estado imediatamente após a condição inicial, segue que $X_1 = 1$, pois, dado que em $n = 0$ o recipiente B estava vazio, em $n = 1$ devemos ter um molécula em B, o que probabilisticamente, representamos por $p(0 \mapsto 1) = 1$, isto é, estando B vazio, uma molécula passa de A para B com probabilidade 1. Supondo então $X_n = k$, a probabilidade da molécula a ser movida estar em B será

$$p(k \mapsto k-1) = \frac{k}{m},$$

consequentemente, se for uma molécula que esteja em A, a probabilidade será

$$p(k \mapsto k+1) = \frac{m-k}{m}.$$

Essas probabilidades correspondentes acima, são chamadas de *probabilidades de transição* e podem ser escritas como entradas da matriz quadrada M com ordem $(m+1)$, tais que, $M_{ij} := p(i \mapsto j)$. A matriz M construída como segue, é chamada de *matriz de transição*, onde $M_{ij} = \frac{j}{m}$ ou $M_{ij} = 1 - \frac{j}{m}$, conforme respectivamente, $j = i-1$ ou $j = i+1$,

$$M = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ \frac{1}{m} & 0 & 1-\frac{1}{m} & 0 & \cdots & 0 & 0 \\ 0 & \frac{2}{m} & 0 & 1-\frac{2}{m} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & \frac{1}{m} \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

Temos claramente para todo i, j, que $M_{ij} \geq 0$, e que $\sum_{k=i\pm1} M_{ik} = 1$ para todo i. Essas propriedades fazem com que M^T seja uma matriz de Markov.

Note que temos o vetor-coluna $\overrightarrow{u^{(0)}} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$ com $m+1$ entradas, representando a condição inicial $X_0 = 0$, e significando que no tempo $n = 0$ suas entradas são,

$$u_k^{(0)} := p(X_0 = k) = \begin{cases} 1, \text{se } k{=}1, \\ 0, \text{caso contrário.} \end{cases}$$

Dado que, a probabilidade contida em $\overrightarrow{u^{(0)}}$ concentrada em 0 no tempo $n = 0$, espalha-se pelo sistema, segue para $n = 1$, que

$$\vec{u}^{(1)} = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ 0 \end{pmatrix} = M^T \vec{u}^{(0)},$$

e

$$\vec{u}^{(2)} = \begin{pmatrix} \frac{1}{m} \\ 0 \\ \frac{m-1}{m} \\ 0 \\ \vdots \\ 0 \end{pmatrix} = M^T \vec{u}^{(1)},$$

e de modo generalizado, notamos que

$$\vec{u}^{(n)} = \begin{pmatrix} u_0^{(n)} \\ \vdots \\ u_m^{(n)} \end{pmatrix} = M^T \vec{u}^{(n-1)}$$

é o vetor de distribuição de probabilidade no tempo n, onde as entradas $u_k^{(n)}$ são não-negativas e $\sum u_k^{(n)} = 1$ para k percorrendo o conjunto $\{0, 1, 2, \ldots, m\}$ com a interpretação de que $u_k^{(n)} = P(\textit{a molécula estar em k no instante n})$. De fato, estando a molécula em k no instante n, isto significa que no instante $n-1$ devia estar em $k-1$ ou mesmo, em $k+1$. Disso, tiramos que

$$u_k^{(n)} = P(k-1 \mapsto k).u_{k-1}^{(n-1)} + P(k+1 \mapsto k).u_{k+1}^{(n-1)},$$

o que equivalentemente representamos de forma vetorial como $\vec{u}^{(n)} = M^T \vec{u}^{(n-1)}$.

Para a condição inicial dada, a evolução do sistema converge para um estado de equilíbrio em termos de número de moléculas em cada recipiente. Estamos interessados em provar esse comportamento a partir da matriz M^T, mais precisamente, verificando o que acontece com $(M^T)^n$ quando $n \to \infty$. Tal verificação será mais interessante se a matriz M^T satisfizer as condições do Teorema de Perron-Frobenius, entretanto, isso não ocorre. Consideremos, então, uma matriz M_ε, para $0 < \varepsilon < \dfrac{1}{m}$ sendo uma pequena variação associada à fração de tempo em que a molécula permanece num estado, conforme definimos a seguir:

$$M_\varepsilon = \begin{pmatrix} 2\varepsilon & \frac{1}{m} - \varepsilon & 0 & \cdots & 0 & 0 \\ 1 - 2\varepsilon & 2\varepsilon & \frac{2}{m} - \varepsilon & \cdots & 0 & 0 \\ 0 & 1 - \frac{1}{m} - \varepsilon & 2\varepsilon & \frac{3}{m} - \varepsilon & 0 & \vdots \\ \vdots & 0 & 1 - \frac{2}{m} - \varepsilon & 2\varepsilon & \cdots & 0 \\ \vdots & 0 & 0 & 1 - \frac{3}{m} - \varepsilon & \ldots & 1 - 2\varepsilon \\ 0 & 0 & \cdots & 0 & \frac{1}{m} - \varepsilon & 2\varepsilon \end{pmatrix}.$$

Com isso, dadas as matrizes quadradas M^T e M_ε de ordem $m+1$, definimos a distância entre M^T e M_ε por

$$d(M^T, M_\varepsilon) := \max_{1 \leq i,j \leq m+1} |\,(M^T)_{ij} - (M_\varepsilon)_{ij}\,|,$$

donde verificamos que $|\,(M^T)_{ij} - (M_\varepsilon)_{ij}\,| \leq 2\varepsilon$ para todo i,j. Nessas condições, a distância $d(M^T, M_\varepsilon)$ entre as matrizes M^T e M_ε torna-se tão pequena quanto se queira, isto é,

$$\lim_{\varepsilon \to 0} M_\varepsilon = M^T.$$

Além disso, temos associado à matriz M_ε um grafo $\Gamma(M_\varepsilon)$ tal que, todas as entradas de M_ε cujos índices são números consecutivos, são positivas. Consequentemente, para quaisquer dois vértices distintos i,j de $\Gamma(M_\varepsilon)$ deve existir uma aresta orientada de i para j, e assim, o grafo $\Gamma(M_\varepsilon)$ é fortemente conexo e a matriz M_ε é irredutível, de sobremaneira que $(M_\varepsilon^m)_{ij} > 0$ e a matriz satisfaz as condições do Teorema de Perron-Frobenius.

Concluímos daí que, como $\vec{u}^{(n)} = M^T \vec{u}^{(n-1)}$, isto é, a distribuição para um tempo n é calculada recursivamente a partir de $n-1$; existe no equilíbrio, uma distribuição $\vec{v}$ com entradas

$$v_k := \lim_{n \to \infty} u_k^{(n)} = \lim_{n \to \infty} P(X_n = k)$$

qualquer que seja a distribuição inicial $\vec{u}$.

Dessa forma, obtemos $\vec{v}$ fazendo

$$(M^T v)_k = (\lim_{n \to \infty} M^T u^{(n)})_k = (\lim_{n \to \infty} u^{(n+1)})_k = (v)_k = v_k$$

e assim, $M^T \vec{v} = \vec{v}$, o que mostra que $\vec{v}$ é uma distribuição invariante em relação M^T. Com outras palavras, $\vec{v}$ é o único autovetor associado ao autovalor de Perron-Frobenius $PF(M^T) = 1$. Portanto, há uma probabilidade de regresso das moléculas para o recipiente A, ainda que para um tempo humanamente impossível de ser vivido. □

Referências

[BH12] M. Bestvina, and M. Handel, ***Train tracks and automorphisms of free groups***, Annals of Mathematics (2), vol. 135 (1992), no. 1, pp. 1–51.

[Bog08] Bogopolski, O. ***Introduction to Group Theory***, EMS: European Mathematical Society, English edition; Dortmund, Germany 2008.

[Bol78] Boldrini, J. L.; Costa, S. I. R.; Ribeiro, V. L. F. F.; Wetzler, H. G. ***Álgebra Linear*** - Harper & Row do Brasil, São Pulo, 1ªed. 1978.

[Dia09] Diaconis, P. ***The Markov chain Monte Carlo revolution***, Bulletin of the AMS, Vol. 46, 2009, p. 179–205.

[Doo53] Doob, J.L., ***Stochastic Processes***. New York, John Wiley, 1953.

[EGNO15] Etingof, P.; Gelaki, Sh.; Nikshych, D.; Ostrik, V. ***Tensor categories, Mathematical Surveys and Monographs***, Vol. 205, American Mathematical Society, 2015.

[Eur02] *Revista Eureka!* ***nº13***, *p. 41-44*, Sociedade Brasileira de Matemática, 2002.

[Fel67] Feller, W. ***An Introduction to Probability Theory and Its Applications***. JOHN WILEY & SONS, New York, 3nd Ed.,1967.

[GM18] ***Google Matrix: fundamentals, applications and beyond***; Workshop, 15-18 October 2018, https://indico.math.cnrs.fr/event/3475/overview.

[Jam15] James, B. R. ***Probabilidade: Um Curso em Nível Intermediário***, Projeto Euclides, IMPA, Rio de Janeiro, 2015.

[Lim04] Lima, E. L. ***Álgebra Linear***, Coleção Matemática Universitária, IMPA, 7ª ed., Rio de Janeiro, 2004.

[Mag15] Magalhães, M. N. ***Probabilidade e Variáveis Aleatórias***. EDUSP: Editora da Universidade de São Paulo, São Paulo, 3ª Edição, 2ª reimpressão, 2015.

[MCCF06] Morgado, A. C.; de Carvalho, J. B. P.; Carvalho, P. C. P.; e Fernandez, P. ***Análise Combinatória e Probabilidade***, SBM; Coleção do Professor de Matemática, 9ª ed., Rio de Janeiro, 2006.

[PY98] Pollicott, M.; Yuri, M. ***Dynamical Systems and Ergodic Theory***, London Mathematical Society Student Texts 40, Cambridge University Press.

[Ser02] Serre, D. ***Matrices : Theory and Applications***, Springer-Verlag New York, 2002.

[Za93] Zalesskii, A.E. ***Linear Groups***, Em Algebra IV, Encyclopaedia of Mathematical Sciences, Vol. 37, Springer-Verlag, 1993.