Instituto Federal de Educação, Ciência e
Tecnologia da Bahia - Campus Valença
Licenciatura em Matemática

Carolina Silva Fraga

Cadeias de Markov: um estudo espectral

Valença-BA
2023

Carolina Silva Fraga

Cadeias de Markov: um estudo espectral

Trabalho de Conclusão de Curso apresentado ao Instituto Federal de Educação, Ciência e Tecnologia da Bahia - *Campus* Valença como parte dos requisitos para a obtenção do título de Licenciada em Matemática.

Orientador: Prof. Dr. Diogo Soares D. da Silva

Coorientador: Prof. Dr. Diego Daltro Conceição

Valença-BA
2023

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

F811c Fraga, Carolina Silva

Cadeias de Markov: um estudo espectral / Carolina Silva Fraga; orientador Diogo Soares D. da Silva; coorientador Diego Daltro Conceição -- Valença : IFBA, 2023.

40f.

Trabalho de Conclusão de Curso (Licenciada em Matemática.) -- Instituto Federal da Bahia, 2023.

1. Probabilidade. 2. processo markoviano. 3. convergência. 4. Teoria Espectral. I. Silva, Dr. Diogo Soares D. da, orient. II. Conceição, Prof. Dr. Diego Daltro, coorient. III. TÍTULO.

CDD. 512.5

Compreender as coisas que nos rodeiam é a melhor preparação para compreender o que há mais além.

Hipátia de Alexandria (370 EC - 415 EC)

Agradecimentos

Agradeço ao meu pai, Henrique Fraga, por todo o apoio e incentivo que sempre me proporcionou durante toda a minha trajetória e também por todos os ensinamentos e valores repassados.

Agradeço ao professor Dr. Diogo Soares Dórea da Silva pela orientação, suporte e por todos os aprendizados que agregaram na minha formação acadêmica, me tornando uma profissional mais competente.

Ao meu coorientador Dr. Diego Daltro Conceição agradeço sobretudo, por suas contribuições valiosas para este trabalho e por sua parceria e gentileza inestimáveis.

Agradeço também aos meus amigos pelos momentos de diversão e reflexão nos quais tornaram a minha formação mais leve e empolgante.

Aos membros da banca, Me. Raphael Dantas Pinho e Me. Roque da Silva Lyrio, agradeço pelo pelas orientações e sugestões realizadas, que foram muito úteis para a composição deste trabalho.

Agradeço ao meu amigo Yuval Peres pela atenção e disponibilidade que foram importantes no desenvolvimento de meus conhecimentos matemáticos. Sou grata também, por sua gentileza, bom humor e palavras de incentivo que tornaram essa etapa acadêmica ainda mais especial.

Resumo

Este trabalho tem como objetivo apresentar um estudo sobre Cadeias de Markov em tempo discreto e espaço de estados finito. O tema será abordado de forma gradativa, de tal modo que, iniciaremos com elementos básicos e essenciais para uma compreensão geral. Em seguida falaremos sobre classificações de estados e cadeias, assim como proposições relacionadas a este tópico que serão importantes para verificação de resultados posteriores como o Teorema de Convergência em Variação Total. Neste teorema, veremos o modo de como a matriz de transição irá convergir e as condições necessárias que garantem a convergência. Para tanto, abordaremos sobre um tipo de distribuição que chamaremos de estacionária e estabeleceremos uma métrica de distância entre duas distribuições de probabilidade. Em seguida, iremos tratar sobre Tempo de Mistura, onde será introduzido um conceito que define o tempo necessário para que a distância entre duas distribuições de probabilidade seja suficientemente pequena. Todos estes conceitos abordados serão importantes para a construção da relação entre Cadeia de Markov e Teoria Espectral, pois a partir dos resultados construídos junto com o estudo de autovalores e autovetores de uma matriz, iremos encontrar limitantes para distâncias entre distribuições de probabilidade para determinados tipos de cadeias em função dos autovalores da matriz de transição.

Palavras-chave: Probabilidade; processo markoviano; convergência; Teoria Espectral.

Abstract

The aim of this work is to present a study on Markov Chains in discrete time and finite state space. The theme will be approached gradually, in such a way that we will start with basic and essential elements for a general understanding. Next, we will talk about classifications of states and chains, as well as propositions related to this topic that will be important for verification of later results such as the Convergence Theorem in Total Variation. In this theorem, we will see how the transition matrix will converge and the necessary conditions that guarantee convergence. To do so, we will approach a type of distribution that we will call stationary and we will establish a metric of distance between two probability distributions. Finally, we will deal with Mixing Time, where we will introduce a concept that defines the time required for the distance between two probability distributions be sufficiently small. All these concepts discussed will be important for building the relationship between Markov Chain and Spectral Theory in which, based on the results constructed with the study of eigenvalues and eigenvectors of a matrix, we will find limits for distances between probability distributions for certain types of chains in function of the eigenvalues of the transition matrix.

Keywords: Probability; Markovian Processes; convergence; Spectral Theory.

Sumário

Capítulo 0

Introdução

Em 1906, o matemático russo Andrei Andreyevich Markov iniciou um estudo de um processo em que apenas o resultado de uma experiência atual pode interferir no resultado da experiência seguinte. Tal processo ficou conhecido como Cadeia de Markov e suas aplicações podem ser encontradas em diversas áreas, como Economia, Física e Biologia.

Cadeia de Markov é um caso particular de processo estocástico que pode ser subdividido em dois tipos com relação ao tempo, o tipo contínuo e o tipo discreto. Para esta pesquisa, iremos apenas trabalhar com Cadeias de Markov em tempo discreto e todas as formulações incluídas ao longo do texto como definições, proposições e teoremas, foram direcionas apenas para os casos discretos. Além disso, todo o estudo desenvolvido delimita-se para Cadeia de Markov do tipo homogênea com espaço de estados finito.

A pesquisa foi desenvolvida através da revisão bibliográfica em que a principal obra utilizada foi (LEVIN; PERES; WILMER, 2009). Tal pesquisa é constituída de três capítulos partindo do capítulo um. Todo o trabalho desenvolvido foi feito de forma gradativa, desde conceitos fundamentais de Probabilidade até recursos mais avançados que relacionam as distâncias entre distribuições de probabilidades com a existência de limitantes superiores e inferiores.

O primeiro capítulo é iniciado com conceitos preliminares básicos de Probabilidade que serão necessários para a compreensão do tema. Então, são apresentadas as primeiras definições, resultados e exemplos. Neste capítulo, também trataremos sobre classificações e proposições que serão essenciais para a compreensão dos capítulos posteriores.

No capítulo dois, será feita a demonstração do Teorema de Convergência em Variação Total, um dos principais teoremas do texto. Nele iremos analisar como ocorre a convergência e as condições necessárias que a garantem. Após isso, trataremos sobre tempo de mistura, além de caracterizações de distância de probabilidade, assim como resultados importantes derivados das definições apresentadas.

No terceiro capítulo, iremos construir uma relação entre Cadeias de Markov e Teoria Espectral, principal objetivo da pesquisa. Iniciaremos com o estudo de autovalores, autovetores e produto interno, e então abordaremos tempo de relaxamento, onde a partir dos resultados apresentados em função de alguns entes relativos à Teoria Espectral, iremos encontrar um limitante superior e um limitante inferior para o tempo de mistura. Em seguida, concluiremos o capítulo, encontrando limitantes superiores entre distância de distribuições de probabilidade para determinados tipos de cadeias, desta vez, utilizando resultados conhecidos do espaço l^p.

Capítulo 1

Cadeias de Markov: principais conceitos e definições

Neste capítulo serão apresentados alguns aspectos introdutórios sobre Cadeias de Markov, como definições, classificações de estados e cadeias e exemplos de aplicações. Os resultados aqui abordados serão essenciais para a construção dos próximos capítulos.

1.1 Conceitos Preliminares

Definição 1.1. *Dizemos que um experimento é **aleatório** se, repetido diversas vezes sob as mesmas condições, os resultados obtidos podem ser diferentes. O conjunto formado por todos os resultados possíveis em um determinado experimento aleatório é chamado **espaço amostral**. Representaremos esse conjunto por Ω. Por fim, um **evento** é qualquer subconjunto de um espaço amostral.*

Exemplo 1.1. *Em um lançamento de um dado convencional não viciado, temos que o resultado do número da face representa um exemplo de experimento aleatório, visto que, repetido diversas vezes em condições idênticas, os resultados produzidos não podem ser previsto com exatidão. Contudo, sabemos que o espaço amostral Ω, formado por todos os resultados possíveis, é dado por:*

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

Além disso, considerando A o evento de ocorrer um número ímpar no lançamento do dado, então

$$A = \{1, 3, 5\}.$$

Ou seja, $A \subset \Omega$.

Definição 1.2. *Uma variável aleatória $X : \Omega \to \mathbb{R}$ é uma função real que mapeia os resultados de um experimento aleatório em valores numéricos. Ademais, quando estes* *resultados numéricos são finitos ou enumeráveis, então dizemos que a variável aleatória é discreta.*

Exemplo 1.2. *No lançamento simultâneo de dois dados honestos, observamos a quantidade de vezes em que o número quatro ocorre. Sendo X a função que mapeia os resultados deste experimento, então temos que*

$$X : \Omega \to \{0, 1, 2\}, \text{ em que } \Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}.$$

Em outras palavras, um lançamento simultâneo de dois dados pode resultar em nenhum dado com o número quatro, pode ocorrer apenas um dado com o número quatro, e por fim, ambos os dados podem obter o resultado observado.

Definição 1.3. *Sejam A e B dois eventos, com $\mathbb{P}(B) > 0$. A probabilidade condicional $\mathbb{P}(A|B)$ do evento A, dado que ocorreu um evento B, é dada por*

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

Definição 1.4. *Dizemos que $A_1, A_2, \ldots, A_n$ são eventos independentes dois-a-dois quando a probabilidade de ocorrência de um deles não interfere na probabilidade de qualquer outro evento, isto é,*

$$\mathbb{P}(A_i|A_j) = \mathbb{P}(A_i), \; \forall \, i, j \in \{1, 2, \ldots, n\}.$$

Além disso, dizemos que tais eventos são coletivamente independentes, se para qualquer coleção finita de índices $1 \leq i_1 < i_2 \cdots i_l$, vale

$$\mathbb{P}\left(\bigcap_{k=1}^{l} A_{i_k}\right) = \prod_{k=1}^{l} \mathbb{P}(A_{i_k}).$$

Exemplo 1.3. *Um dado é lançado cinco vezes. Qual a probabilidade de observarmos a face com o número dois em cada lançamento?*

Sejam os eventos:

A_1*: Ocorre o número dois no* 1° *lançamento;*

A_2*: Ocorre o número dois no* 2° *lançamento;*

$\vdots$

A_5*: Ocorre o número dois no* 5° *lançamento.*

Observe que o resultado de cada lançamento não interfere no resultado dos demais, e como para cada lançamento, a probabilidade de ocorrer o número dois é dada por $\frac{1}{6}$, então

$$\mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5) = \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{7776}.$$

Definição 1.5. *Uma distribuição discreta de probabilidade é uma função que toma valores de uma variável aleatória e associa cada valor à sua correspondente probabilidade.*

Exemplo 1.4. *Se uma variável aleatória X toma valores em $x_1, x_2, \ldots, x_n$, com respectivas probabilidades $\mathbb{P}(x_1), \mathbb{P}(x_2), \ldots, \mathbb{P}(x_n)$, então, a distribuição de probabilidade X é dada por*

Tabela 1.1: Distribuição de probabilidade da variável aleatória X

x_1	$\mathbb{P}(x_1)$
x_2	$\mathbb{P}(x_2)$
$\vdots$	$\vdots$
x_n	$\mathbb{P}(x_n)$

Fonte: Autoria própria (2023).

Definição 1.6. *Dizemos que uma distribuição de probabilidade é uniforme quando todos os valores, $x_1, x_2, \ldots, x_n$ de sua variável aleatória X possuem a mesma probabilidade de ocorrer, ou seja,*

$$\mathbb{P}(x_i) = \frac{1}{n} \qquad \forall i \in \{1, 2, \ldots, n\}.$$

Definição 1.7. *Seja X uma variável aleatória discreta de valores $x_1, x_2, \ldots x_n$, e probabilidades $\mathbb{P}(x_1), \mathbb{P}(x_2), \ldots, \mathbb{P}(x_n)$. A esperança de X, ou valor médio de X, é dada por*

$$\begin{aligned}\mathbb{E}(X) &= x_1\mathbb{P}(x_1) + x_2\mathbb{P}(x_2) + \cdots + x_n\mathbb{P}(x_n)\\ &= \sum_{i=1}^{n} x_i\mathbb{P}(x_i).\end{aligned}$$

Exemplo 1.5. *Um supermercado realizou uma promoção em que através do lançamento de um dado, o cliente poderia ganhar um desconto no valor final da compra. Se o lançamento do dado resultar na face 6, então o desconto recebido é de 30%, caso resulte na face 5 o desconto correspondente é de 20%. Se ocorrer face 4, recebe 10% e caso resulte nas faces 1, 2 ou 3 o desconto será de 5%. Qual o desconto médio concedido pelo supermercado?*

Inicialmente, note que, a variável aleatória X é a função que transforma os resultados de Ω em valores numéricos, neste caso, X : n° da face $\rightarrow$ desconto concedido. Logo, para obter o desconto médio, basta calcular $\mathbb{E}(X)$. Para isto, devemos encontrar as probabilidades dos resultados das faces correspondentes aos quatro valores dos descontos, como mostra a tabela.

Tabela 1.2: Probabilidades correspondentes aos tipos de descontos possíveis

X	$\mathbb{P}(X)$
5%	$\mathbb{P}$(face 1, 2 ou 3) = 3/6
10%	$\mathbb{P}$(face 4) = 1/6
20%	$\mathbb{P}$(face 5) = 1/6
30%	$\mathbb{P}$(face 6) = 1/6

Fonte: Autoria própria (2023).

Deste modo, temos que o desconto médio, $\mathbb{E}(X)$, é igual a

$$\begin{aligned}\mathbb{E}(X) &= 0,05 \cdot \frac{3}{6} + 0,10 \cdot \frac{1}{6} + 0,20 \cdot \frac{1}{6} + 0,30 \cdot \frac{1}{6} \\ &= 12,5\%.\end{aligned}$$

1.2 Elementos Básicos de uma Cadeia de Markov

Definição 1.8. *Uma Cadeia de Markov, com espaço de estados finito dado por Ω, é uma sequência de variáveis aleatórias $(X_0, X_1, \ldots)$, em que a cada tempo $t \in \mathbb{N}$, o estado seguinte dado por X_{t+1}, depende apenas do estado atual X_t. Para cada estado, a passagem para o próximo estado é calculada seguindo uma distribuição de probabilidade fixa $P(x, \cdot)$, ou seja, para todo $x_i, x_j \in \Omega$ temos*

$$\mathbb{P}\{X_{t+1} = x_j | H_{t-1} \cap \{X_t = x_i\}\} = \mathbb{P}\{X_{t+1} = x_j | X_t = x_i\} = P(x_i, x_j),$$

em que $H_{t-1} = \bigcap_{s=0}^{t-1} \{X_s = x_s\}$ indica a sequência cronológica de realizações das variáveis aleatórias $(X_0, X_1, \ldots, X_{t-1})$ de uma Cadeia de Markov até o tempo $t-1$, de modo que

$$\begin{aligned} H_0 &= \{X_0 = x_0\} \\ H_1 &= \{X_0 = x_0\} \cap \{X_1 = x_1\} \\ &\vdots \\ H_{t-1} &= \{X_0 = x_0\} \cap \cdots \cap \{X_{t-1} = x_{t-1}\}. \end{aligned}$$

Em outras palavras, a probabilidade do processo passar para o estado x_j no tempo $t+1$, dado que no tempo t estava no estado x_i e no tempo $t-1$ estava no estado x_{t-1} e assim sucessivamente, é igual à probabilidade do processo passar para o estado x_j no tempo $t+1$, dado que no tempo t estava no estado x_i, visto que a probabilidade de transição depende apenas do estado atual. Tal processo é conhecido como *markoviano*, ou processo com perda de memória.

Definição 1.9. *Uma matriz de transição P_{ij} é uma matriz de ordem $n \times n$, em que o elemento da i-ésima linha e da j-ésima coluna é exatamente a probabilidade do processo passar do estado i para o estado j em uma unidade de tempo t, com $i, j \in \Omega$. Para isso, escrevemos*

$$P_{ij} = \mathbb{P}(X_{t+1} = j | X_t = i).$$

Definição 1.10. *Uma Cadeia de Markov é dita homogênea no tempo quando as distribuições de probabilidade são independentes do tempo t, ou seja, para todo tempo $t \in \mathbb{N}, \mathbb{P}(X_{t+1} = j | X_t = i)$ é constante.*

Observação 1.1. *Devido ao fato do processo markoviano ser um caso particular de processo estocástico, temos que P_{ij} é uma matriz estocástica, ou seja, P_{ij} é uma matriz $n \times n$ em que todas as entradas são não negativas e todas as linhas têm a soma das suas entradas iguais a* 1*, isto é,*

$$\sum_{y \in \Omega} \mathbb{P}(x, y) = 1; \text{ para todo } x \in \Omega.$$

Para outras características de matrizes estocásticas, ver em (ROSS, 1996), página 163.

Exemplo 1.6. *Certa máquina de calcular utiliza apenas os dígitos* 0 *e* 1*. Ela deve transmitir um desses dígitos a cada estágio. Além disso, em cada estágio, há uma probabilidade p de que o dígito transmitido seja alterado, enquanto que há uma probabilidade $q = 1 - p$ de que o digito não seja modificado.*

Podemos representar o processo da seguinte forma:

$$P_{ij} = \begin{pmatrix} q & p \\ p & q \end{pmatrix}.$$

Note que $\Omega = \{0, 1\}$ e $(X_0, X_1, \ldots)$ é a sequência de dígitos utilizados pela máquina. A primeira linha da matriz P_{ij} representa a distribuição condicional do estado X_{t+1} dado que $X_t = 0$, enquanto que a segunda linha é a distribuição condicional do estado X_{t+1} dado que $X_t = 1$.

Definição 1.11. *Em uma Cadeia de Markov P_{ij}, o vetor estado em uma observação é um vetor linha que representa a probabilidade de observação de cada um dos estados para algum tempo t. Por exemplo*

$$\mu_t = (\xi_1, \xi_2, \ldots, \xi_n),$$

em que $\sum_{i=1}^{n} \xi_i = 1$ e $\xi_i \geq 0$ para $i = 1, 2, \ldots, n$.

Com relação ao Exemplo 1.6, temos que a representação do vetor estado para o tempo $t = 1$, considerando que a máquina inicie com o digito 0, é dada por:

$$\mu_1 = (\mathbb{P}\{x_1 = 0 | x_0 = 0\}, \mathbb{P}\{x_1 = 1 | x_0 = 0\}) = \begin{pmatrix} q & p \end{pmatrix}.$$

Definição 1.12. *Um vetor $\mu_0 = (\xi_1, \xi_2, \ldots, \xi_n)$ é chamado vetor inicial se $\sum_{i=1}^{n} \xi_i = 1$ e $\xi_i \geq 0$ para $i = 1, 2, \ldots, n$.*

O vetor inicial se refere à distribuição de probabilidade no tempo 0*. Geralmente, o processo inicia seguindo alguma distribuição de probabilidade, mas há casos em que a cadeia inicia necessariamente em um determinado estado. Nestes casos, μ_0 tem o número* 1 *na coordenada correspondente ao estado inicial e zero nas demais coordenadas.*

Exemplo 1.7. *Em uma determinada cidade, se hoje está ensolarado, amanhã terá 80% de chances de estar ensolarado. Se hoje está nublado, amanhã terá 60% de chances de estar nublado. Queremos encontrar a probabilidade de que daqui a três dias esteja ensolarado, supondo que hoje está ensolarado.*

Para calcular as probabilidades dos estados para o próximo tempo, basta multiplicar o vetor referente ao estado atual pela matriz de transição do processo, obtendo assim o vetor estado para o tempo seguinte, ou seja,

$$\mu_{t+1} = \mu_t \cdot P.$$

Note que, para o tempo $t = 0$, o vetor estado inicial é igual a $\mu_0 = \begin{pmatrix} 1 & 0 \end{pmatrix}$, visto que o processo observado se inicia em um dia ensolarado.

$$\mu_1 = \begin{pmatrix} 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0,8 & 0,2 \\ 0,4 & 0,6 \end{pmatrix} = \begin{pmatrix} 0,8 & 0,2 \end{pmatrix}$$

$$\mu_2 = \begin{pmatrix} 0,8 & 0,2 \end{pmatrix} \cdot \begin{pmatrix} 0,8 & 0,2 \\ 0,4 & 0,6 \end{pmatrix} = \begin{pmatrix} 0,72 & 0,28 \end{pmatrix}$$

$$\mu_3 = \begin{pmatrix} 0,72 & 0,28 \end{pmatrix} \cdot \begin{pmatrix} 0,8 & 0,2 \\ 0,4 & 0,6 \end{pmatrix} = \begin{pmatrix} 0,688 & 0,312 \end{pmatrix}.$$

Portanto, para o terceiro dia, temos uma probabilidade que faça sol de 68,8%.

Em certos casos, não é interessante fazer uma análise passo a passo do processo, mas apenas onde o processo se encontra após algum tempo. Com o resultado seguinte, podemos calcular as distribuições de probabilidades da matriz P e o vetor estado para qualquer tempo $t \in \mathbb{N}$, sem a necessidade de calcular as probabilidades anteriores.

Do Exemplo 1.7, sabemos que para encontrar as distribuições do vetor estado para o tempo seguinte, calculamos $\mu_{t+1} = \mu_t \cdot P$. Logo, segue que

$$\begin{gathered} \mu_1 = \mu_0 \cdot P \\ \mu_2 = \mu_1 \cdot P = (\mu_0 \cdot P) \cdot P = \mu_0 \cdot P^2 \\ \mu_3 = \mu_2 \cdot P = (\mu_0 \cdot P^2) \cdot P = \mu_0 \cdot P^3 \\ \vdots \\ \mu_t = \mu_0 \cdot P^t. \end{gathered}$$

Deste modo temos que, em um processo markoviano, as distribuições de uma matriz P para algum tempo $t \in \mathbb{N}$ são dadas por $(P_{ij})^t$.

Observação 1.2. *Dada uma Cadeia de Markov em Ω com matriz de transição P, a probabilidade de transição no n-ésimo passo satisfaz*

$$P_{ij}^n = \sum_{k \in \Omega} P_{ik}^{n-1} \cdot P_{kj}.$$

Note que pode existir diferentes caminhos em que é possível partir do estado i e chegar ao estado j passando por um estado intermediário k no passo $n-1$. Logo, temos que a probabilidade de partir do estado i e chegar ao estado j é igual a soma de todas as possibilidades de chegar a esse objetivo. Veremos na Proposição 1.1, que esse será um caso particular da Equação de Chapman-Kolmogorov, em que através dela é possível calcular as probabilidades de transição após $n+m$ passos, particionando o processo no tempo.

Proposição 1.1 (Equação de Chapman-Kolmogorov)**.** *Seja P uma matriz de transição de uma Cadeia de Markov homogênea no tempo. Então para quaisquer $n, m \geq 0$, temos*

$$P_{ij}^{m+n} = \sum_{k \in \Omega} P_{ik}^{m} P_{kj}^{n}.$$

Demonstração.

$$\begin{aligned}
P_{ij}^{m+n} &= \mathbb{P}(X_{m+n} = j | X_0 = i) \\
&= \sum_{k \in \Omega} \frac{\mathbb{P}(X_{m+n} = j, X_m = k, X_0 = i)}{\mathbb{P}(X_0 = i)} \\
&= \sum_{k \in \Omega} \frac{\mathbb{P}(X_m = k, X_0 = i)}{\mathbb{P}(X_0 = i)} \cdot \frac{\mathbb{P}(X_{m+n} = j, X_m = k, X_0 = i)}{\mathbb{P}(X_m = k, X_0 = i)} \\
&= \sum_{k \in \Omega} \mathbb{P}(X_m = k | X_0 = i) \cdot \mathbb{P}(X_{m+n} = j | X_m = k, X_0 = i) \\
&= \sum_{k \in \Omega} \mathbb{P}(X_m = k | X_0 = i) \cdot \mathbb{P}(X_{m+n} = j | X_m = k) \\
&= \sum_{k \in \Omega} P_{ik}^{m} P_{kj}^{n}.
\end{aligned}$$

Na passagem da primeira para segunda linha, foi utilizada a fórmula de probabilidade condicional para eventos independentes, que afirma que, $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$. Na terceira linha multiplicamos pela fração $\frac{\mathbb{P}(X_m = k, X_0 = i)}{\mathbb{P}(X_m = k, X_0 = i)}$, e em seguida, utilizamos novamente a fórmula de probabilidade condicional mencionada. Por fim, na passagem da quarta para quinta linha, foi utilizada a Propriedade de Markov.

Com isso, concluímos que a probabilidade da Cadeia de Markov ir do estado i para o j em $m+n$ passos é igual à soma das probabilidades de alcançar um estado intermediário k em m passos, multiplicadas pelas probabilidades de ir do k ao j em n passos. □

1.3 Classificações de Estados e Cadeias

Definição 1.13. *Seja uma Cadeia de Markov P com espaço de estados finito $\{x_1, x_2, \ldots, x_n\} \in \Omega$. Dizemos que x_j é alcançável por x_i se é possível que o processo se mova de x_i para x_j em um número finito de passos. Denotamos isso por $x_i \to x_j$. Matematicamente, isso significa que existe $n \in \mathbb{N}$ tal que*

$$P^n(x_i, x_j) > 0.$$

Se também vale que $x_j \to x_i$, *então dizemos que esses estados se comunicam e escrevemos* $x_i \leftrightarrow x_j$.

Definição 1.14. *Uma Cadeia de Markov com espaço de estados* Ω *é dita irredutível quando todos os estados da cadeia se comunicam, ou seja*

$$x \leftrightarrow y; \qquad \forall x, y \in \Omega.$$

Definição 1.15. *Seja* P *uma Cadeia de Markov com espaço de estados* Ω *e um estado* $x \in \Omega$. *Temos que o conjunto dos tempos de retorno de* x *é formado pelas unidades de tempo em que é possível sair do estado* x *e retornar ao mesmo estado após* t *passos. Representamos isso por*

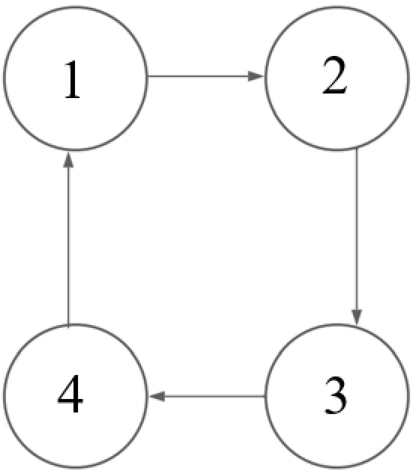
$$\tau(x) \coloneqq \{t \geq 1 : P^t(x, x) > 0\}.$$

Definição 1.16. *O período de um determinado estado* x *é dado pelo* MDC *dos tempos de retorno de* x, *e denotamos por* $d(x) = MDC(\tau(x))$.

Definição 1.17. *Um estado* $x \in \Omega$ *é chamado de periódico se a cadeia de eventos pode retornar ao estado apenas em múltiplos de algum número inteiro maior que* 1, *ou seja,* $d(x) > 1$.

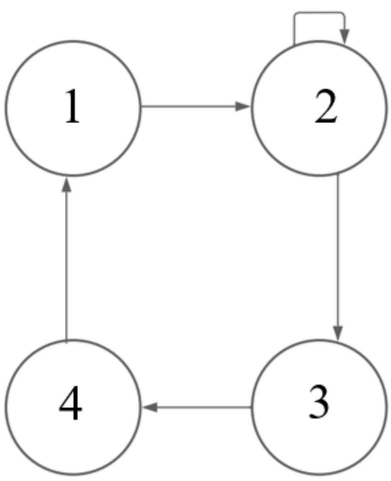
Exemplo 1.8. *Considere as Cadeias de Markov com as seguintes transições de estados:*

Figura 1.1: Cadeia 01



Fonte: Autoria própria (2023).

Figura 1.2: Cadeia 02



Fonte: Autoria própria (2023).

Observe que, para a Figura 1.1, temos que todos os estados são periódicos, pois partindo de qualquer estado, apenas para o conjunto dos tempos de retorno $\tau(x) = 4t$ *com* $t \in \mathbb{N}$, *a cadeia de eventos poderá retornar para o estado inicial. Por outro lado, para a* *cadeia da Figura 1.2, os estados não são periódicos pois, observe por exemplo que* 4 *e* 5 *fazem parte do conjunto dos tempos de retorno dos estados e* $MDC(4, 5) = 1$

Definição 1.18. *Um estado é chamado de aperiódico quando não é periódico, ou seja, quando* $d(x) = 1$.

Definição 1.19. *Uma Cadeia de Markov* P *com espaço de estados* Ω *é chamada de aperiódica se, para todo* $x \in \Omega$, *o* $MDC(\tau(x)) = 1$. *Caso o* $MDC(\tau(x)) > 1$, *então dizemos que* P *é periódica.*

Definição 1.20. *Dizemos que uma Cadeia de Markov é transitiva quando, para cada par de estados* $(x, y) \in \Omega \times \Omega$, *existe uma bijeção* $\varphi = \varphi_{(x,y)} : \Omega \to \Omega$, *tal que*

$$\varphi(x) = y \ \text{ e } \ P(z, w) = P(\varphi(z), \varphi(w)), \forall\, z, w \in \Omega.$$

Exemplo 1.9. *Considere a situação apresentada no Exemplo 1.6 e suponha que as probabilidades* p *e* q *são respectivamente* $\frac{1}{3}$ *e* $\frac{2}{3}$. *Logo, podemos reescrever a matriz* P *como*

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

Considerando $\varphi : \{0, 1\} \to \{0, 1\}$, *tal que* $\varphi(0) = 1$ *e* $\varphi(1) = 0$. *Então, temos que*

$$\begin{aligned} P(0,1) &= P(\varphi(0), \varphi(1)) = P(1,0) = \frac{1}{3}, \\ P(1,0) &= P(\varphi(1), \varphi(0)) = P(0,1) = \frac{1}{3}, \\ P(0,0) &= P(\varphi(0), \varphi(0)) = P(1,1) = \frac{2}{3}, \\ P(1,1) &= P(\varphi(1), \varphi(1)) = P(0,0) = \frac{2}{3}. \end{aligned}$$

Note que existe uma bijeção $\varphi_{(x,y)} : \Omega \to \Omega$. *Desta forma, pela Definição 1.20, concluímos que* P *é uma cadeia transitiva.*

Proposição 1.2. *Se* P *é uma Cadeia de Markov irredutível com espaço de estados* Ω, *então* $MDC(\tau(x)) = MDC(\tau(y)), \forall x, y \in \Omega$.

Demonstração. Sejam $x, y \in \Omega$ estados quaisquer de uma Cadeia de Markov irredutível. Então existem números inteiros não negativos r e l tais que $P^r(x, y) > 0$ e $P^l(y, x) > 0$.

Considerando $m = r + l$, temos que

$$P^m(x, x) = P^{r+l}(x, x) \geq P^r(x, y) \cdot P^l(y, x) > 0.$$

De maneira analóga obtemos,

$$P^m(y, y) = P^{r+l}(y, y) \geq P^l(y, x) \cdot P^r(x, y) > 0.$$

A primeira desigualdade indica que a probabilidade de partir do estado x, e chegar ao estado x após m passos, é maior ou igual a probabilidade de sair do estado x para o estado y em r passos, em seguida retornar para o estado x em l passos, dado que existem outros estados onde é possível fazer esse percurso, além do y. Vale o análogo para a segunda desigualdade.

Com isso, podemos concluir que $m \in \tau(x) \cap \tau(y)$, uma vez que ela faz parte do conjunto dos tempos de retorno de ambos os estados. Seja k um inteiro não negativo com $k \in \tau(y)$, logo $P^k(y, y) > 0$. Dessa forma,

$$P^{k+m}(x, x) = P^{r+k+l}(x, x) \geq P^r(x, y) \cdot P^k(y, y) \cdot P^l(y, x) > 0.$$

Então temos que $k + m \in \tau(x)$, logo $k \in \tau(x) - m$, o que implica que $\tau(y) \subset \tau(x) - m$.

Considerando um inteiro não negativo a, com $a \in \tau(y)$, temos que $a = b - m$ para algum $b \in \tau(x)$. Logo $MDC(\tau(x))|b \Rightarrow MDC(\tau(x))|a + m$ e portanto $MDC(\tau(x))|a$, visto que $MDC(\tau(x))|m$. Então, tem-se que $MDC(\tau(x))$ divide todos os elementos de $\tau(y)$. Logo $MDC(\tau(x)) \leq MDC(\tau(y))$. Por um argumento inteiramente paralelo, podemos obter que $MDC(\tau(y)) \leq MDC(\tau(x))$. Deste modo, concluímos que $MDC(\tau(x)) = MDC(\tau(y))$. □

Proposição 1.3. *Se P é uma matriz de transição de uma Cadeia de Markov aperiódica e irredutível, com espaço de estados em Ω, então existe um inteiro não-negativo r tal que $P^r(x, y) > 0$ para todo $x, y \in \Omega$.*

Demonstração. Para esta demonstração, será utilizado o seguinte resultado de Teoria de Números: "qualquer conjunto de inteiros não-negativos, fechado para adição e que o MDC de todos os seus elementos é igual a 1, deve conter todos os inteiros não negativos, exceto uma quantidade finita deles." A prova desse resultado pode ser encontrada em (LEVIN; PERES; WILMER, 2009), página 20. Como $x \in \Omega$ e P aperiódica, temos que $MDC(\tau(x)) = 1$. Sabemos que $\tau(x)$ é fechado para adição, pois dados s e $t \in \tau(x)$, vale que $P^{s+t}(x, x) \geq P^s(x, x) \cdot P^t(x, x) > 0$. Logo $s + t \in \tau(x)$. Pelo resultado citado anteriormente, existe $l \in \tau(x)$ tal que $q \geq l \Rightarrow q \in \tau(x)$. Como a cadeia é irredutível, então para um $y \in \Omega$, existe um inteiro não negativo $r(x, y)$, tal que $P^r(x, y) > 0$. Portanto, para $q \geq l + r$, temos que

$$P^q(x, y) \geq P^{q-r}(x, x) \cdot P^r(x, y) > 0.$$

Dado que $q-r \geq l$, então $q-l \in \tau(x)$. Deste modo, para todo $q \geq l+r, P^q(x, y) > 0$. Sendo $l' := l + \max_{y \in \Omega} r(x, y)$, temos que $P^q(x, y) > 0, \forall q \geq l$. Com isso, para todo $x, y \in \Omega$, fazemos $q \geq \max_{x \in \Omega} l'$, mostrando que $P^q(x, y) > 0, \forall x, y \in \Omega$.

Deste modo, concluímos que em uma Cadeia de Markov irredutível, a partir de algum tempo, qualquer estado $y \in \Omega$ será acessível por todos os estados de Ω, garantindo assim que a matriz P não possuirá entradas nulas. □

1.4 Distribuições Estacionárias

Em Cadeias de Markov, as distribuições de probabilidade da matriz P podem se aproximar de uma condição de equilíbrio com valores fixos. Veremos no Capítulo 2, no Teorema da Convergência, que toda matriz P aperiódica e irredutível convergirá para uma distribuição constante. Mas primeiramente, será estabelecida uma noção de distância entre distribuições. Para esta seção, contudo, iremos apenas tratar de uma distribuição π que é invariante sob multiplicação pela matriz P^n com $n \in \mathbb{N}$.

Definição 1.21. *Seja P uma Cadeia de Markov com espaço de estados Ω. O vetor estado π é chamado de distribuição estacionária se satisfaz a igualdade*

$$\pi P = \pi.$$

Definição 1.22. *Seja x um estado qualquer de uma Cadeia de Markov com espaço de estados Ω. Definimos tempo de alcance como o tempo mínimo para que a cadeia alcance o estado x. Ou seja,*

$$\tau_x := \min\{t \geq 0 : X_t = x\}.$$

Para tempo de alcance estritamente positivo, escrevemos

$$\tau_x^+ := \min\{t \geq 1 : X_t = x\}.$$

Observação 1.3. *Utilizaremos a notação $\mathbb{E}_x$ e $\mathbb{P}_x$, com $x \in \Omega$, para indicar respectivamente a esperança e a probabilidade dado que o processo foi iniciado no estado x, ou seja, para $X_0 = x$.*

Lema 1.1. *Sejam $x, y \in \Omega$ estados quaisquer de uma Cadeia de Markov irredutível. Então $\mathbb{E}_x(\tau_y^+) < \infty$.*

Demonstração. Sendo P irredutível, sabemos que, para quaisquer $x, y \in \Omega$, existe um $s > 0$ e um $\epsilon > 0$, tal que $P^s(x, y) > \epsilon$. Dado $r \geq s$, então para qualquer variável aleatória X_t, a probabilidade da cadeia alcançar o estado y entre intervalo de tempo t e $t + r$, é no mínimo ϵ. Logo, vale que

$$\mathbb{P}\{X_s = y, t < s \leq t + r\} > \epsilon.$$

Pelo complementar do evento, temos

$$\mathbb{P}\{X_s \neq y, \forall\, t < s \leq t + r\} \leq 1 - \epsilon.$$

Note que, se $\tau_y^+ > n$, temos $X_t \neq y$, $\forall\, t \in (0, n]$, então, para $k > 0$, vale

$$\begin{aligned}
\mathbb{P}_x\{\tau_y^+ > kr\} &= \mathbb{P}_x\{X_t \neq y, \forall 0 < t \leq kr\} \\
&= \mathbb{P}_x\{X_t \neq y, \forall 0 < t \leq (k-1)r\}\mathbb{P}_x\{X_t \neq y, \forall (k-1)r < t \leq kr\} \\
&\leq \mathbb{P}_x\{X_t \neq y, \forall 0 < t \leq (k-1)r\}(1-\epsilon) \\
&\leq \mathbb{P}_x\{X_t \neq y, \forall 0 < t \leq (k-2)r\}(1-\epsilon)^2 \\
&\vdots \\
&\leq \mathbb{P}_x\{X_t \neq y, \forall 0 < t \leq 0\}(1-\epsilon)^k \\
&= (1-\epsilon)^k.
\end{aligned}$$

A passagem da penúltima para a última linha é obtida através da verdade por vacuidade, onde temos $\mathbb{P}_x\{X_t \neq y, \forall 0 < t \leq 0\} = 1$.

Note que, para qualquer variável aleatória X, de valores inteiros não-negativos, temos

$$\begin{aligned}
\mathbb{E}(Y) &= \sum_{t=0}^{\infty} t\,\mathbb{P}(Y = t) \\
&= 1 \cdot \mathbb{P}(Y = 1) + 2 \cdot \mathbb{P}(Y = 2) + 3 \cdot \mathbb{P}(Y = 3) + \cdots \\
&= (\mathbb{P}(Y = 1) + \mathbb{P}(Y = 2) + \mathbb{P}(Y = 3) + \cdots) + (\mathbb{P}(Y = 2) + \mathbb{P}(Y = 3) + \cdots) + \cdots \\
&= \sum_{t=0}^{\infty} \mathbb{P}(Y > t).
\end{aligned}$$

Além disso, segue que $\mathbb{P}\{\tau_y^+ > t\}$ é uma função decrescente com relação a t, pois

$$\begin{aligned}
\mathbb{P}\{\tau_y^+ > t + 1\} &\leq \mathbb{P}\{\tau_y^+ > t + 1\} + \mathbb{P}\{\tau_y^+ = t + 1\} \\
&= \mathbb{P}\{\tau_y^+ \geq t + 1\} \\
&= \mathbb{P}\{\tau_y^+ > t\}.
\end{aligned}$$

Desta forma, com os resultados obtidos, fazemos

$$\mathbb{E}_x(\tau_y^+) = \sum_{t \geq 0} \mathbb{P}_x\{\tau_y^+ > t\} \leq \sum_{k \geq 0} r\mathbb{P}_x\{\tau_y^+ > kr\} \leq \sum_{k \geq 0} r(1-\epsilon)^k < \infty,$$

e como, por definição, $\epsilon \in (0, 1]$, então $(1 - \epsilon) \in [0, 1)$. Logo, temos uma série geométrica convergente, o que implica que $\mathbb{E}_x(\tau_y^+) < \infty$. $\square$

Teorema 1.1. *Seja uma Cadeia de Markov irredutível com matriz de transição P. Então existe distribuição estacionária π em Ω tal que $\pi P = \pi$, com $\pi(x) > 0, \forall x \in \Omega$.*

Demonstração. Fixando um estado $z \in \Omega$ e um estado arbitrário $y \in \Omega$, definimos $\tilde{\pi}(y) \coloneqq \mathbf{E}_z(\text{número de visitas a um estado } y \text{ antes de retornar para } z)$.

Deste modo, temos que

$$\begin{aligned}\tilde{\pi}(y) &:= \mathbf{E}_z \\ &= \sum_{t=0}^{\infty} \mathbb{P}_z\{X_t = y, \tau_z^+ > t\},\end{aligned} \tag{1.1}$$

dado que o número médio de visitas ao estado y antes de retornar para o estado z é igual à soma de todas as probabilidades de alcançar y no tempo anterior ao tempo de retorno do estado z.

A irredutibilidade da cadeia garante que $\tilde{\pi}(y) > 0$. Além disso, para cada estado y, $\tilde{\pi}(y) \leq \mathbf{E}_z \cdot \tau_z^+$ e pelo Lema 1.1, sabemos que $\tilde{\pi}(y) < \infty$.

Veremos a seguir que $\tilde{\pi}(y)$ é estacionário, iniciando com a seguinte notação:

$$(\tilde{\pi}P)(y) = \sum_{x\in\Omega} \tilde{\pi}(x)P(x,y) = \sum_{x\in\Omega}\sum_{t=0}^{\infty} \mathbb{P}_z\{X_t = x, \tau_z^+ > t\}P(x,y). \tag{1.2}$$

Como o evento $\{\tau_z^+ > t\} = \{\tau_z^+ \geq t+1\}$ é apenas determinado pelas variáveis aleatórias $X_0, X_1, \ldots, X_t$, sendo portanto, independente do evento $X_{t+1} = y$, então ao condicionar $X_t = x$, obtemos

$$\begin{aligned}&\mathbb{P}_z\{X_t = x, X_{t+1} = y, \tau_z^+ \geq t+1\} \\ &\quad= \mathbb{P}_z\{X_t = x, \tau_z^+ \geq t+1\}\mathbb{P}_z\{X_{t+1} = y | X_t = x, \tau_z^+ \geq t+1\} \\ &\quad= \mathbb{P}_z\{X_t = x, \tau_z^+ \geq t+1\}P(x,y).\end{aligned} \tag{1.3}$$

Observe que, o terceiro termo da equação (1.3) é equivalente ao último termo equação (1.2). Desta forma, ao utilizar o primeiro termo da equação (1.3) e inverter a ordem do somatório, obtemos

$$\begin{aligned}(\tilde{\pi}P)(y) &= \sum_{x\in\Omega} \tilde{\pi}(x)P(x,y) \\ &= \sum_{t=0}^{\infty}\sum_{x\in\Omega} \mathbb{P}_z\{X_t = x, X_{t+1} = y, \tau_z^+ \geq t+1\} \\ &= \sum_{t=0}^{\infty}\sum_{x\in\Omega} \mathbb{P}_z\{X_{t+1} = y | X_t = x\}\mathbb{P}_z\{X_t = x, \tau_z^+ \geq t+1\} \\ &= \sum_{t=0}^{\infty} \mathbb{P}_z\{X_{t+1} = y, \tau_z^+ \geq t+1\} \\ &= \sum_{t=1}^{\infty} \mathbb{P}_z\{X_t = y, \tau_z^+ \geq t\}.\end{aligned} \tag{1.4}$$

A terceira igualdade é justificada pelo fato de que a soma interior para as probabilidades $\{X_t = x\}$ é igual a 1.

Note que equação (1.4) é equivalente à equação (1.1), logo

$$\begin{aligned}
(\tilde{\pi}P)(y) &= \sum_{t=1}^{\infty} \mathbb{P}_z\{X_t = y, \tau_z^+ \geq t\} \\
&= \tilde{\pi}(y) - \mathbb{P}_z\{X_0 = y, \tau_z^+ > 0\} + \sum_{t=1}^{\infty} \mathbb{P}_z\{X_t = y, \tau_z^+ = t\} \\
&= \tilde{\pi}(y) - \mathbb{P}_z\{X_0 = y\} + \mathbb{P}_z\{X_{\tau_z^+} = y\} \\
&= \tilde{\pi}(y).
\end{aligned}$$

A passagem da terceira para quarta linha é obtida considerando dois casos: $z = y$ e $z \neq y$. Para $z = y$, temos que a probabilidade do processo iniciar no estado z dado que $X_0 = z$, é igual a 1. Além disso, temos que a probabilidade do processo iniciar no estado z e estar no estado z em $X_{\tau_z^+}$ também é igual a 1. Desta forma, os dois últimos termos da terceira igualdade se cancelam, resultando apenas em $\tilde{\pi}(y)$.

Para o caso $z \neq y$, por uma justificativa inteiramente análoga, temos que ambas as probabilidades são iguais a zero, resultando novamente em $\tilde{\pi}(y)$. Portanto, conclui-se que $(\tilde{\pi}P)(y) = \tilde{\pi}(y)$, ou seja, $\pi P = \pi$. □

Proposição 1.4. *Seja P uma matriz de transição de uma Cadeia de Markov irredutível, com espaço de estados Ω. Então existe uma única distribuição de probabilidade que satisfaz $\pi \cdot P = \pi$.*

Demonstração. Suponha que existe mais de uma distribuição estacionária. Digamos que $\pi_1(x)$ e $\pi_2(x)$ com $x \in \Omega$, sejam duas delas. Considere que existe um C constante real de modo que $\frac{\pi_1(x)}{\pi_2(x)} = C$, então para todo x em Ω, temos que $\pi_1(x) = \pi_2(x)$, pois ambas são distribuições de probabilidade proporcionais em que a soma das entradas resulta em 1. Considere então que $\frac{\pi_1(x)}{\pi_2(x)}$ não é constante, então existe $x \in \Omega$ tal que $\frac{\pi_1(x)}{\pi_2(x)}$ tenha valor reduzido. Além disso, considere $w \in \Omega$ com $\frac{\pi_1(x)}{\pi_2(x)} < \frac{\pi_1(w)}{\pi_2(w)}$. Pela irredutibilidade da cadeia sabemos que, $P(w, x) > 0$. Deste modo, temos

$$\begin{aligned}
\pi_1(x) &= \sum_{y \in \Omega} \pi_1(y) \cdot P(y, x) \\
&= \sum_{y \in \Omega} \frac{\pi_1(y)}{\pi_2(y)} \cdot \pi_2(y) \cdot P(y, x) \\
&> \sum_{y \in \Omega} \frac{\pi_1(x)}{\pi_2(x)} \cdot \pi_2(y) \cdot P(y, x) \\
&= \frac{\pi_1(x)}{\pi_2(x)} \cdot \sum_{y \in \Omega} \pi_2(y) \cdot P(y, x) = \pi_1(x),
\end{aligned}$$

formando uma contradição.

A primeira igualdade é obtida utilizando a hipótese das distribuições serem estacionárias. Para a segunda igualdade multiplicamos por $\frac{\pi_2(y)}{\pi_2(y)}$, a irredutibilidade da cadeia garante que $\pi_2(y)$ tenha todas as entradas positivas. A desigualdade é obtida atráves da existência do elemento w e para última igualdade, novamente utilizamos a hipótese mencionada. $\square$

Proposição 1.5. *Seja P uma Cadeia de Markov transitiva com espaço de estados Ω. Então a distribuição de probabilidade uniforme em Ω é estacionária para P.*

Demonstração. Dados dois estados fixos $x, y \in \Omega$, seja $\varphi : \Omega \to \Omega$ uma bijeção que preserva probabilidade de trasição para cada $\varphi(x) = y$ e seja U a distribuição uniforme em Ω. Então

$$\begin{aligned}\sum_{z\in\Omega} U(z)P(z,x) &= \sum_{z\in\Omega} U(\varphi(z))P(\varphi(z),\varphi(x))\\ &= \sum_{z\in\Omega} U(\varphi(z))P(\varphi(z),y)\\ &= \sum_{w\in\Omega} U(w)P(w,y),\end{aligned}$$

em que $\varphi(z) = w$.

Observe que, quando o processo inicia em uma distribuição uniforme, partindo de qualquer estado z e avança em um passo, a probabilidade de alcançar o próximo estado é igual para cada estado de Ω. Como $\sum\limits_{x\in\Omega}\sum\limits_{z\in\Omega} U(z)P(z,x) = 1$, então

$$\sum_{z\in\Omega} U(z)P(z,x) = \frac{1}{|\Omega|} = U(x).$$

Conclui-se assim, que a distribuição uniforme U é estacionária para P. $\square$

Exemplo 1.10. *Um país é dividido em três regiões demográficas. Foi observado que, a cada ano, 5% dos moradores da região 1 mudam para a região 2 e 5% mudam para a região 3. Dos moradores da região 2, 15% mudam para a região 1 e 10% mudam para a região 3. Finalmente, dos moradores da região 3, 10% mudam para a região 1 e 5% mudam para a região 2.*

Com as informações do texto acima, iremos encontrar a distribuição de probabilidade estacionária π.

Observe que para o tempo $t = 0$, a matriz P é dada por

$$P = \begin{pmatrix} 0,9 & 0,05 & 0,05 \\ 0,15 & 0,75 & 0,1 \\ 0,1 & 0,05 & 0,85 \end{pmatrix}.$$

Temos que P é uma Cadeia de Markov irredutível, pois todos os estados da cadeia se comunicam. Logo, existe um vetor de distribuição estacionária.

Pelas Definições 1.21 e 1.11, sabemos respectivamente que $\pi P = \pi$ e $\sum_{i=1}^{n} \pi_i = 1$. Então vale que,

$$\begin{pmatrix} \pi_1 & \pi_2 & \pi_3 \end{pmatrix} \cdot \begin{pmatrix} 0,9 & 0,05 & 0,05 \\ 0,15 & 0,75 & 0,1 \\ 0,1 & 0,05 & 0,85 \end{pmatrix} = \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 \end{pmatrix},$$

$$\pi_1 + \pi_2 + \pi_3 = 1.$$

Assim, obtemos o sistema

$$\begin{cases} 0,9\pi_1 + 0,15\pi_2 + 0,1\pi_3 = \pi_1 \\ 0,05\pi_1 + 0,75\pi_2 + 0,05\pi_3 = \pi_2 \\ 0,05\pi_1 + 0,1\pi_2 + 0,85\pi_3 = \pi_3 \\ \pi_1 + \pi_2 + \pi_3 = 1. \end{cases}$$

Ao resolver o sistema com aproximações em três casas decimais, temos que

$$\pi = \begin{pmatrix} 0,542 & 0,166 & 0,292 \end{pmatrix}.$$

1.5 Reversibilidade e Tempo Reverso

Definição 1.23. *Uma Cadeia de Markov P em Ω é chamada de reversível, se para alguma distribuição π em Ω, satisfaz*

$$\pi(x) \cdot P(x,y) = \pi(y) \cdot P(y,x), \qquad \forall x, y \in \Omega. \tag{1.5}$$

A equação (1.5) é chamada de equação detalhada de balanço. Ela nos diz que, para Cadeia de Markov reversível, a probabilidade do processo iniciar em um estado x, e em um passo finalizar no estado y, é igual à probabilidade de ocorrer o processo reverso.

Proposição 1.6. *Seja P uma matriz de transição de uma Cadeia de Markov em Ω e π uma distribuição qualquer que, com P satisfaz a equação detalhada de balanço. Então π é uma distribuição estacionária de P.*

Demonstração. Ao somar os termos da equação detalhada de balanço, obtemos que para todo $y \in \Omega$,

$$\sum_{y\in\Omega} \pi(y) \cdot P(y,x) = \sum_{y\in\Omega} \pi(x) \cdot P(x,y) = \pi(x) \cdot \sum_{y\in\Omega} P(x,y) = \pi(x),$$

dado que a matriz P é estocástica. □

Definição 1.24. *Seja P uma Cadeia de Markov irredutível e seja π uma distribuição estacionária que com P satisfaz a equação detalhada de balanço. Então, o tempo reverso de P é dado pela seguinte matriz de transição:*

$$\hat{P}(x,y) := \frac{\pi(y)\cdot P(y,x)}{\pi(x)}.$$

Afirmação 1.1. *Temos que a matriz $\hat{P}$ é estocástica pois*

$$\sum_{y\in\Omega}\hat{P}(x,y) = \sum_{y\in\Omega}\frac{\pi(y)\cdot P(y,x)}{\pi(x)} = \frac{1}{\pi(x)}\sum_{y\in\Omega}\pi(y)\cdot P(y,x) = \frac{1}{\pi(x)}\cdot\pi(x) = 1.$$

Proposição 1.7. *Seja (X_t) uma Cadeia de Markov irredutível em Ω, com matriz de transição P, e distribuição estacionária π, e seja $(\hat{X}_t)$ a cadeia de Markov em tempo reverso com matriz $\hat{P}$. Então π é estacionária para $\hat{P}$, e para quaisquer $x_0, x_1, \ldots, x_n \in \Omega$, temos*

$$\pi(x_0)P(x_0,x_1)\cdots P(x_{n-1},x_n) = \pi(x_n)\hat{P}(x_n,x_{n-1})\cdots\hat{P}(x_1,x_0).$$

Demonstração. Para mostrar que π é estacionária para $\hat{P}$, fazemos

$$\sum_{y\in\Omega}\pi(y)\hat{P}(y,x) = \sum_{y\in\Omega}\pi(y)\frac{\pi(x)\cdot P(x,y)}{\pi(y)} = \pi(x)\sum_{y\in\Omega}P(x,y) = \pi(x).$$

Para provar que as probabilidades das duas trajetórias são iguais, note que pela Definição 1.24 temos que para cada índice i de Ω, vale

$$P(x_{i-1},x_i) = \frac{\pi(x_i)\hat{P}(x_i,x_{i-1})}{\pi(x_{i-1})},$$

então, segue que

$$\begin{aligned}
&\pi(x_0)P(x_0,x_1)P(x_1,x_2)\cdots P(x_{n-1},x_n)\\
&= \pi(x_0)\cdot\frac{\hat{P}(x_1,x_0)\pi(x_1)}{\pi(x_0)}\cdot\frac{\hat{P}(x_2,x_1)\pi(x_2)}{\pi(x_1)}\cdots\frac{\hat{P}(x_n,x_{n-1})\pi(x_n)}{\pi(x_{n-1})}\\
&= \pi(x_n)\cdot\hat{P}(x_n,x_{n-1})\cdots\hat{P}(x_2,x_1)\cdot\hat{P}(x_1,x_0),
\end{aligned}$$

como queríamos demonstrar. □

Capítulo 2

Teorema de Convergência em Variação Total e Tempos de Mistura

Ao longo deste capítulo, será estabelecido uma noção de distância entre distribuições de probabilidade, cujos conceitos apresentados serão utilizados no Teorema da Convergência em Variação Total, onde veremos que toda matriz P de uma Cadeia de Markov irredutível e aperiódica convergirá para uma distribuição estacionária. Neste capítulo, também iremos tratar sobre Tempo de Mistura, no qual será introduzido um conceito que define o tempo necessário para que a distância entre duas distribuições seja suficientemente pequeno.

2.1 Distância em Variação Total

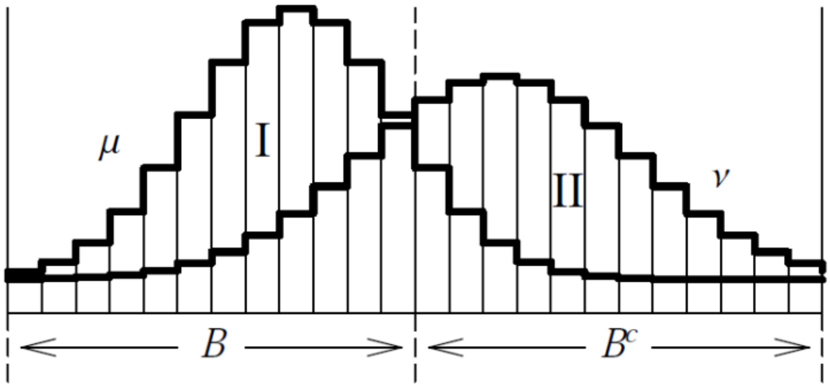
Definição 2.1. *Sejam duas distribuições de probabilidade μ, $\nu \in \Omega$. A distância em variação total entre elas é definida por*

$$\|\mu - \nu\|_{VT} \coloneqq \max_{A \subset \Omega} |\mu(A) - \nu(A)|.$$

Em outros termos, a distância em variação total é a maior distância entre probabilidades que duas distribuições podem assumir para o mesmo evento.

Na Figura 2.1 podemos observar a ilustração de duas distribuições de probabilidades μ e ν, definidas em um conjunto discreto finito dado por $\Omega = B \cup B^c$.

Figura 2.1: Distribuições de probabilidades μ e ν



Fonte: Levin; Peres; Wilmer (2009).

Para o conjunto B, temos que $\mu(x) > \nu(x)$ e para B^c tem-se $\mu(x) \leq \nu(x)$. Além disso, podemos concluir que as áreas I e II são iguais, pois $\sum_{x\in\Omega} \mu(x) = \sum_{x\in\Omega} \nu(x) = 1$, e isso significa que, $\mu(B) - \nu(B) = \nu(B^c) - \mu(B^c)$.

A distância em variação total entre duas probabilidades é o valor máximo da diferença que duas medidas de probabilidade assumem para um único evento. Contudo, essa definição nem sempre é a mais conveniente para determinar a distância entre duas distribuições. A seguinte proposição fornece duas caracterizações alternativas para $\|\mu - \nu\|_{VT}$.

Proposição 2.1. *Sejam duas distribuição de probabilidade μ e $\nu \in \Omega$. Então*

$$\|\mu - \nu\|_{VT} = \frac{1}{2} \cdot \sum_{x\in\Omega} |\mu(x) - \nu(x)|.$$

Demonstração. Seja $B := \{x \mid \mu(x) \geq \nu(x)\}$ e tomando $A \subset \Omega$ como sendo qualquer evento, então

$$\begin{aligned}
\mu(A) - \nu(A) &= \mu(A \setminus B) + \mu(A \cap B) - \nu(A \setminus B) - \nu(A \cap B) \\
&= \mu(A \cap B) - \nu(A \cap B) + \mu(A \setminus B) - \nu(A \setminus B) \\
&\leq \mu(A \cap B) - \nu(A \cap B) \\
&\leq \mu(A \cap B) - \nu(A \cap B) + (\mu(B \setminus A) - \nu(B \setminus A)) \\
&= \mu(A \cap B) + \mu(B \setminus A) - (\nu(A \cap B) + \nu(B \setminus A)) \\
&= \mu(B) - \nu(B).
\end{aligned}$$

Note que podemos escrever os conjuntos A e B respectivamente como, $(A \setminus B) \cup (B \cap A)$ e $(B \setminus A) \cup (B \cap A)$. Além disso, podemos definir $B^c := \{x \mid \mu(x) < \nu(x)\}$. Assim, para a primeira desigualdade, temos que para todo $x \in A \cap B^c$, vale $\mu(x) - \nu(x) < 0$, logo a diferença entre as probabilidades não pode diminuir quando esses elementos são retirados. Para a segunda desigualdade, incluindo mais elementos de B, a diferença entre as probabilidades não diminui, pois em B vale que $\mu(x) - \nu(x) \geq 0$.

Com um argumento análogo, obtemos a desigualdade

$$\nu(A) - \mu(A) \leq \nu(B^c) - \mu(B^c).$$

Como $\max|\mu(A) - \nu(A)| = \dfrac{1}{2} \cdot [\mu(B) - \nu(B) + \nu(B^c) - \mu(B^c)]$, temos que

$$\begin{aligned} \|\mu - \nu\|_{VT} &= \frac{1}{2} \cdot [\mu(B) - \nu(B) + \nu(B^c) - \mu(B^c)] \\ &= \frac{1}{2} \cdot \sum_{x \in \Omega} |\mu(x) - \nu(x)|. \end{aligned}$$

□

Observação 2.1. *Atráves da demonstração da Proposição 2.1, também podemos escrever a distância de variação total como*

$$\|\mu - \nu\|_{VT} = \sum_{\mu(x) \geq \nu(x)} [\mu(x) - \nu(x)],$$

visto que $\|\mu - \nu\|_{VT} = \dfrac{1}{2} \cdot \sum_{x \in \Omega} |\mu(x) - \nu(x)|$ *e* $\sum_{\mu(x) \geq \nu(x)} [\mu(x) - \nu(x)] = \sum_{\nu(x) > \mu(x)} [\nu(x) - \mu(x)]$, *logo, temos*

$$\begin{aligned} \|\mu - \nu\|_{VT} &= \frac{1}{2} \cdot \left[\sum_{\mu(x) \geq \nu(x)} [\mu(x) - \nu(x)] + \sum_{\nu(x) > \mu(x)} [\nu(x) - \mu(x)] \right] \\ &= \sum_{\mu(x) \geq \nu(x)} [\mu(x) - \nu(x)]. \end{aligned}$$

Afirmação 2.1. *Sejam μ, ν, η distribuições de probabilidade em Ω. Então a distância em variação total entre elas satisfaz a desigualdade triangular*

$$\|\mu - \nu\|_{VT} \leq \|\mu - \eta\|_{VT} + \|\eta - \nu\|_{VT}.$$

Para verificar a Afirmação 2.1, tomamos $x \in \Omega$. Como $\mu(x), \nu(x), \eta(x)$ assumem valores reais, vale que

$$\begin{aligned} |\mu(x) - \nu(x)| &= |\mu(x) - \eta(x) + \eta(x) - \nu(x)| \\ &\leq |\mu(x) - \eta(x)| + |\eta(x) - \nu(x)|. \end{aligned}$$

Logo, segue que

$$\sum_{x \in \Omega} |\mu(x) - \nu(x)| \leq \sum_{x \in \Omega} \big[|\mu(x) - \eta(x)| + |\eta(x) - \nu(x)| \big]. \tag{2.1}$$

Pela Proposição 2.1, sabemos que $\|\mu - \nu\|_{VT} = \dfrac{1}{2} \cdot \sum_{x \in \Omega} |\mu(x) - \nu(x)|$, isso implica $2 \cdot \|\mu - \nu\|_{VT} = \sum_{x \in \Omega} |\mu(x) - \nu(x)|$. Ao substituir esse resultado na desigualdade 2.1, temos que $2 \cdot \|\mu - \nu\|_{VT} \leq 2 \cdot \big[\|\mu - \eta\|_{VT} + \|\eta - \nu\|_{VT} \big]$. Dividindo ambos os lados por 2 obtemos $\|\mu - \nu\|_{VT} \leq \|\mu - \eta\|_{VT} + \|\eta - \nu\|_{VT}$, como desejado.

2.2 Teorema de Convergência em Variação Total

Com os resultados apresentados na seção 2.1, iremos demonstrar um dos principais teoremas deste trabalho, o Teorema de Convergência em Variação Total. Ele nos garante que toda matriz aperiódica e irredutível irá convergir para uma distribuição estacionária, de modo que todas as suas linhas se aproximarão, em noção de distância, da distribuição estacionária π, apresentada na Definição 1.21.

Teorema 2.1. *Seja P uma Cadeia de Markov irredutível e aperiódica com espaço de estados Ω e distribuição estacionária π. Então existem constantes $0 < a < 1$ e $C > 0$ tais que*

$$d(t) := \max_{x\in\Omega} \|P^t(x,\cdot) - \pi\|_{VT} \leq Ca^t.$$

Demonstração. Como P é por hipótese irredutível e aperiódica, pela Proposição 1.3, temos que existe um $n \geq 0$ tal que P^n possui todas a entradas positivas. Então para algum $\delta > 0$ suficientemente pequeno, vale

$$P^n(x,y) \geq \delta\pi(y); \quad \forall x, y \in \Omega.$$

Sendo $\Pi = \begin{pmatrix} \pi \\ \vdots \\ \pi \end{pmatrix}$ uma matriz $|\Omega| \times |\Omega|$ em que cada linha representa a distribuição estacionária de π, então segue que Π é uma matriz estocástica.

Seja $\theta = 1 - \delta$, definimos uma nova matriz Q que satisfaz

$$P^n = (1-\theta)\Pi + \theta Q.$$

Logo,

$$Q = \frac{P^n - (1-\theta)\Pi}{\theta}.$$

Note que Q é estocástica, pois

$$\sum_{y\in\Omega} Q(x,y) = \frac{\displaystyle\sum_{y\in\Omega} P^n(x,y) - (1-\theta)\sum_{y\in\Omega}\Pi(x,y)}{\theta} = \frac{1-(1-\theta)}{\theta} = 1.$$

Afirmação 2.2. *Se M é uma matriz estocástica com espaço de estados Ω, então $M\Pi = \Pi$.*

De fato, sendo $i, j \in \Omega$, podemos escrever cada elemento de $(M\Pi)_{ij}$ como

$$(M\Pi)_{ij} = \sum_{1\leq a\leq|\Omega|} M_{ia}\Pi_{aj} = \pi(j) \sum_{1\leq a\leq|\Omega|} M_{ia} = \pi(j) = \Pi_{ij}.$$

A segunda e a quarta igualdades são obtidas pelo fato de a matriz Π ter como linhas o vetor π. Logo $\Pi_{aj} = \Pi_{ij} = \pi(j)$, pois dependem apenas da coluna j. A terceira igualdade é justificada pelo fato de M ser estocástica.

Vamos mostrar por indução que para qualquer $k \in \mathbb{N}\backslash\{0\}$, vale

$$P^{rk} = (1-\theta^k)\Pi + \theta^k Q^k. \tag{2.2}$$

Para $k = 1$, obtemos $P^r = (1-\theta)\Pi + \theta Q$. Portanto, é verificada a base de indução.

Assumindo que a Equação 2.2 vale para $k = n$, temos $P^{rn} = (1-\theta^n)\Pi + \theta^n Q^n$. Então, para $(n+1)$ temos

$$\begin{aligned}
P^{r(n+1)} &= P^{rn}P^r \\
&= [(1-\theta^n)\Pi + \theta^n Q^n]P^r \\
&= (1-\theta^n)\Pi P^r + \theta^n Q^n P^r \\
&= (1-\theta^n)\Pi P^r + \theta^n Q^n[(1-\theta)\Pi + \theta Q] \\
&= (1-\theta^n)\Pi P^r + \theta^n(1-\theta)Q^n\Pi + \theta^{n+1}Q^{n+1} \\
&= (1-\theta^n)\Pi + \theta^n(1-\theta)\Pi + \theta^{n+1}Q^{n+1} \\
&= (1-\theta^n)\Pi + (\theta^n - \theta^{n+1})\Pi + \theta^{n+1}Q^{n+1} \\
&= (1-\theta^{n+1})\Pi + \theta^{n+1}Q^{n+1},
\end{aligned}$$

provando por indução que $P^{rk} = (1-\theta^k)\Pi + \theta^k Q^k$ para todo $k \in \mathbb{N}\backslash\{0\}$. A segunda igualdade usa a hipótese de indução, a quarta igualdade é explicada pelo fato de $P^r = (1-\theta)\Pi + \theta Q$. Para a sexta igualdade utilizamos a Afirmação 2.2 e o fato de π ser estacionária.

Ao multiplicar a Equação 2.2 por P^j, obtemos

$$\begin{aligned}
P^{rk+j} &= (1-\theta^k)\Pi P^j + \theta^k Q^k P^j \qquad (2.3)\\
&= (1-\theta^k)\Pi + \theta^k Q^k P^j \\
&= \Pi - \theta^k\Pi + \theta^k Q^k P^j \\
&= \Pi + \theta^k(Q^k P^j - \Pi),
\end{aligned}$$

deste modo, tem-se que $P^{rk+j} - \Pi = \theta^k(Q^k P^j - \Pi)$. Para qualquer linha x de ambos os lados da equação, vale $(P^{rk+j} - \Pi)(x,\cdot) = \theta^k(Q^k P^j - \Pi)(x,\cdot)$. Com isso, temos que

$$\begin{aligned}
\|P^{rk+j}(x,\cdot) - \pi\|_{VT} &= \|P^{rk+j}(x,\cdot) - \Pi(x,\cdot)\|_{VT} \\
&= \frac{1}{2}\sum_{y\in\Omega} |(P^{rk+j} - \Pi)(x,y)| \\
&= \frac{\theta^k}{2}\sum_{y\in\Omega} |(Q^k P^j - \Pi)(x,y)| \\
&= \theta^k\|(Q^k P^j(x,\cdot) - \Pi(x,\cdot)\|_{VT} \\
&\leq \theta^k.
\end{aligned}$$

A segunda igualdade é explicada pela Proposição 2.1. A terceira igualdade é obtida através do resultado da Equação 2.3. Em seguida, utilizamos novamente a Proposição 2.1. A desigualdade é justificada pelo fato de a variação total entre duas distribuições de probabilidade ser limitada no intervalo [0, 1].

Considere a e C constantes tais que $a = \sqrt[r]{\theta}$ e $C = a^{-r}$. Para qualquer t inteiro positivo, pelo algoritmo da divisão, temos que existem $r, j \in \mathbb{N}$ e k fixo, tais que $t = rk+j$ com $0 \leq j < r$. Então, segue que

$$\begin{aligned}
\|P^t(x,\cdot) - \pi\|_{VT} &= \|P^{rk+j}(x,\cdot) - \pi\|_{VT} \\
&\leq \theta^k \\
&= a^{rk} \\
&\leq a^{rk} a^{j-r} \\
&= a^{-r} a^{rk+j} \\
&= Ca^t.
\end{aligned}$$

Como a desigualdade obtida vale para todo $x \in \Omega$, então para qualquer $t > 0$, obtemos

$$\max_{x \in \Omega} \|P^t(x,\cdot) - \pi\|_{VT} \leq Ca^t.$$

Observe que, quando $t \to \infty$, o lado direito da desigualdade converge para 0 dado que C é uma constante positiva e $a \in (0,1)$. Deste modo, o Teorema da Convergência em Variação Total nos garante que se P é uma matriz aperiódica e irredutível com distribuição estacionária π, então as linhas da matriz P^t convergirão para π quanto $t \to \infty$. □

2.3 Tempo de Mistura

Definição 2.2. *Seja P uma Cadeia de Markov em Ω. Definimos o Tempo de Mistura de P como sendo a primeira unidade de tempo em que P se aproxima da distribuição estacionária π dentro de um parâmetro $\epsilon \in [0,1]$, ou seja,*

$$t_{mix}(\epsilon) := \min\{t : d(t) \leq \epsilon\}. \tag{2.4}$$

Observe que Equação 2.4 fornece uma medida de quanto tempo leva para que todas as linhas da matriz de transição P se aproxime, em certa quantidade, da distribuição π. É muito frequente na literatura que o parâmetro estabelecido seja igual a um quarto, isto é

$$t_{mix} := t_{mix}\left(\frac{1}{4}\right).$$

Definição 2.3. *Seja P uma Cadeia de Markov com espaço de estados Ω e distribuição estacionária π. Definimos a função $\bar{d}(t)$ como,*

$$\bar{d}(t) := \max_{x,y \in \Omega} \|P^t(x,\cdot) - P^t(y,\cdot)\|_{VT}.$$

Lema 2.1. *Dadas as funções $d(t)$ e $\bar{d}(t)$, apresentadas no enunciado do Teorema 2.1 e na Definição 2.3, então*

$$d(t) \leq \bar{d}(t) \leq 2d(t).$$

Demonstração. A segunda desigualdade é obtida através da desigualdade triangular da distância em variação total, como demonstrado abaixo:

$$\begin{aligned}
\bar{d}(t) &= \max_{x,y\in\Omega} \|P^t(x,\cdot) - P^t(y,\cdot)\|_{VT} \\
&\leq \max_{x,y\in\Omega} \left[\|P^t(x,\cdot) - \pi\|_{VT} + \|P^t(y,\cdot) - \pi\|_{VT}\right] \\
&\leq \max_{x\in\Omega} \|P^t(x,\cdot) - \pi\|_{VT} + \max_{y\in\Omega} \|P^t(y,\cdot) - \pi\|_{VT} \\
&= 2d(t).
\end{aligned}$$

Para provar que $d(t) \leq \bar{d}(t)$, note primeiramente que, por hipótese, π é estacionário para P, então para qualquer conjunto $A \subseteq \Omega$, vale que

$$\pi(A) = \sum_{y\in\Omega} \pi(y) P^t(y, A). \tag{2.5}$$

Assim, temos

$$\begin{aligned}
d(t) &= \max_{x\in\Omega} \|P^t(x,\cdot) - \pi\|_{VT} \\
&= \max_{A\subseteq\Omega} |P^t(x, A) - \pi(A)| \\
&= \max_{A\subseteq\Omega} \left| P^t(x, A) - \sum_{y\in\Omega} \pi(y) \cdot P^t(y, A) \right| \\
&= \max_{A\subseteq\Omega} \left| \sum_{y\in\Omega} \pi(y) \cdot [\ P^t(x, A) - P^t(y, A)] \right| \\
&= \max_{A\subseteq\Omega} \sum_{y\in\Omega} \pi(y) \cdot |P^t(x, A) - P^t(y, A)| \\
&\leq \sum_{y\in\Omega} \pi(y) \cdot \max_{A\subseteq\Omega} |P^t(x, A) - P^t(y, A)| \\
&= \sum_{y\in\Omega} \pi(y) \cdot \|P^t(x,\cdot) - P^t(y,\cdot)\|_{VT} \\
&\leq \max_{x,y\in\Omega} \|P^t(x,\cdot) - P^t(y,\cdot)\|_{VT} \\
&= \bar{d}(t).
\end{aligned}$$

A primeira igualdade é obtida através do enunciado do Teorema 2.1. A terceira igualdade é explicada pela equação 2.5. A quarta iqualdade é justificada da seguinte forma: $\sum_{y\in\Omega} \pi(y) \cdot P^t(x, A) = P^t(x, A) \sum_{y\in\Omega} \pi(y) = P^t(x, A)$. Para a primeira desigualdade utilizamos o fato de que o máximo da soma de números reais não é maior do que a soma dos máximos de números reais. Em seguida, foi utilizada a Definição 2.1 de distância em variação total. Como a soma dos elementos de $\pi(y)$ não supera o valor de 1, então temos que $\pi(y)$ multiplicado pela norma da diferença de números reais não será maior do que o máximo da norma da diferença entre números reais, obtendo assim a desigualdade desejada. $\square$

Definição 2.4. *Seja P uma Cadeia de Markov em Ω, definimos distância de separação $s_x(t)$, como*

$$s_x(t) := \max_{y\in\Omega} \left[1 - \frac{P^t(x, y)}{\pi(y)}\right].$$

Para o valor máximo que um estado x pode assumir em $s_x(t)$, escrevemos

$$s(t) := \max_{x\in\Omega} s_x(t).$$

Lema 2.2. *A distância de separação $s_x(t)$ satisfaz*

$$\|P^t(x, \cdot) - \pi\|_{VT} \leq s_x(t),$$

ou seja,

$$d(t) \leq s(t).$$

Demonstração. Calculando diretamente, temos que

$$\begin{aligned}\|P^t(x, \cdot) - \pi\|_{VT} = \sum_{y\in\Omega} [\pi(y) - P^t(x, y)] &= \sum_{y\in\Omega} \pi(y) \left[1 - \frac{P^t(x, y)}{\pi(y)}\right] \\ &\leq \max_{y\in\Omega} \left[1 - \frac{P^t(x, y)}{\pi(y)}\right] = s_x(t).\end{aligned}$$

$\square$

Capítulo 3

Teoria Espectral

Ao longo deste capítulo, iremos construir uma relação entre Cadeias de Markov e Teoria Espectral, principal objetivo do trabalho. Iniciaremos com o estudo de autovalores, autovetores e produto interno e então passaremos para tempo de relaxamento em que, com base nas definições apresentadas, serão encontrados um limitante superior e um limitante inferior para o tempo de mistura de cadeias reversíveis, irredutíveis e aperiódicas. Por fim, encontraremos limitantes superiores para distâncias entre probabilidades de cadeias reversíveis e reversíveis, transitivas.

3.1 Cadeia de Markov Reversível e Teoria Espectral

Definição 3.1. *Seja T uma transformação linear tal que $T : \mathbb{R}^n \to \mathbb{R}^n$. Então um escalar λ é chamado de autovalor de T se existe um vetor não-nulo v, tal que*

$$T(v) = \lambda v. \tag{3.1}$$

O vetor não-nulo que satisfaz a equação (3.1) é chamado de autovetor de T. Além disso, para obter tal igualdade, cada autovetor v deve estar associado ao seu correspondente autovalor λ.

Lema 3.1. *Seja P a matriz de transição de uma Cadeia de Markov finita. Segue que*

(i) Se λ é um autovalor de P, então $|\lambda| \leq 1$.

(ii) Se P é irredutível, o espaço vetorial de autovetores correspondentes ao autovalor 1 é um espaço unidimensional gerado pelo vetor coluna $\mathbf{1} := (1, 1, ..., 1)^T$.

(iii) Se P é irredutível e aperiódica, então -1 não é um autovalor de P.

Demonstração. Seja v um autovetor de P com autovalor correspondente λ, e seja $\|v\|_\infty \coloneqq \max_{y\in\Omega}|v(y)|$. Então para todo $x \in \Omega$, vale que

$$\begin{aligned}\|Pv\|_\infty(x) = \|Pv(x)\|_\infty &= \left\|\sum_{y\in\Omega} P(x,y)v(y)\right\|_\infty \\ &\leq \left\|\sum_{y\in\Omega} P(x,y)\right\|_\infty \|v\|_\infty = \|v\|_\infty.\end{aligned}$$

Como x é arbitrário, concluimos que $\|Pv\|_\infty \leq \|v\|_\infty$. A desigualdade é obtida pelo fato de $v(y)$ ser limitado pelo $\max_{y\in\Omega}|v(y)| \coloneqq \|v\|_\infty$, em seguida, como $\sum_{y\in\Omega} P(x,y)v(y) = 1$, obtemos a última igualdade. Segue que $\|Pv(x)\|_\infty = \|\lambda v(x)\|_\infty$ e $\|Pv(x)\|_\infty \leq \|v\|_\infty$, então, $|\lambda| \cdot \|v\|_\infty = \|\lambda v\|_\infty = \|Pv\|_\infty \leq \|v\|_\infty$, o que implica $|\lambda| \leq 1$, concluindo assim o primeiro item.

Para a verificação do item (ii), note que, pela unicidade da distribuição estacionária, há apenas um espaço de soluções para

$$P\pi = \pi = 1\pi.$$

Como P é estocástica, segue que

$$P \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = 1 \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

Logo, temos que o autovetor gerado pelo autovalor 1 é $(1, 1, ..., 1)^T$, que é único pela Proposição 1.4.

Para o item (iii), suponhamos que -1 é um autovalor associado ao autovetor v não-nulo da Cadeia de Markov P. Pelo Teorema 2.1, sabemos que a matriz P^t é convergente, o que implica a convergência de $P^t v$. Contudo para $\lambda = -1$, temos $P^t v - (-1)^t v \xrightarrow{t\to\infty} 0$. Note que $(-1)^t v$ não é convergente, formando assim uma contradição. $\square$

Definição 3.2. *Seja $\langle \cdot, \cdot \rangle$ o produto interno usual em $\mathbb{R}^{|\Omega|}$, dado por $\langle v_i, v_j \rangle = \sum_{x\in\Omega} v_i(x)v_j(x)$, $\forall\, i, j \in \{1, 2, \ldots, |\Omega|\}$. Definimos um novo produto interno $\langle \cdot, \cdot \rangle_\pi$, em que*

$$\langle v_i, v_j \rangle_\pi \coloneqq \sum_{x\in\Omega} v_i(x)v_j(x)\pi(x).$$

Lema 3.2. *Seja P uma matriz de transição reversível com respeito ao vetor estacionário π. Então*

(i) O produto interno do espaço $(\mathbb{R}^{|\Omega|}, \langle \cdot, \cdot \rangle_\pi)$ tem uma base ortonormal de autovetores reais, dados por $\{v_k\}_{k=1}^{|\Omega|}$, associados aos autovalores reais $\{\lambda_k\}$.

(ii) A matriz P pode ser decomposta em

$$\frac{P^t(x,y)}{\pi(x)} = \sum_{k=1}^{|\Omega|} v_k(x)v_k(y)\lambda_k^t.$$

(iii) O autovetor v_1 correspondente ao autovalor 1 pode ser tomado como o vetor constante $\mathbf{1}$. Neste caso, obtemos

$$\frac{P^t(x,y)}{\pi(y)} = 1 + \sum_{k=2}^{|\Omega|} v_k(x)v_k(y)\lambda_k^t. \tag{3.2}$$

Demonstração. Para provar o item (i), inicialmente definimos uma matriz A, tal que $A(x,y) \coloneqq \pi(x)^{1/2}P(x,y)\pi(y)^{-1/2}$, para cada $x, y \in \Omega$. Dado que P é reversível, então segue que A é uma matriz simétrica, pois

$$\begin{aligned} A(x,y) &= \pi(x)^{-1/2}[\pi(x)P(x,y)]\pi(y)^{-1/2} \\ &= \pi(x)^{-1/2}[\pi(y)P(y,x)]\pi(y)^{-1/2} \\ &= \pi(y)^{1/2}P(y,x)\pi(x)^{-1/2} = A(y,x). \end{aligned}$$

O teorema espectral para matrizes simétricas garante que o produto interno do espaço $(\mathbb{R}^\Omega, \langle\cdot,\cdot\rangle)$ tem base ortonormal de autovetores $\{\varphi_k\}_{k=1}^{|\Omega|}$ com autovalores correspondentes λ_k. Para a demonstração deste resultado, ver (LIMA, 2014), página 257. Além disso, note que $\sqrt{\pi}$ é um autovetor de A com autovalor associado 1, dado que

$$\begin{aligned} (A\sqrt{\pi})(y) &= \sum_{x\in\Omega} A(y,x)\pi(x)^{1/2} \\ &= \sum_{x\in\Omega} A(x,y)\pi(x)^{1/2} \\ &= \sum_{x\in\Omega} \pi(x)^{1/2}\pi(x)^{1/2}P(x,y)\pi(y)^{-1/2} \\ &= \pi(y)^{-1/2}\sum_{x\in\Omega} \pi(x)P(x,y) = \pi(y)^{-1/2}\pi(y) = \sqrt{\pi}(y). \end{aligned}$$

Seja D_π uma matriz diagonal tal que $D_\pi(x,x) = \pi(x)$ para cada $x \in \Omega$. Então podemos escrever a matriz A como $A = D_\pi^{1/2}PD_\pi^{-1/2}$. Se para cada índice k, definimos $v_k \coloneqq D_\pi^{-1/2}\varphi_k$, então cada v_k é um autovetor da matriz P com autovalor correspondente λ_k, como demonstrado abaixo:

$$\begin{aligned} Pv_k = PD_\pi^{-1/2}\varphi_k &= D_\pi^{-1/2}(D_\pi^{1/2}PD_\pi^{-1/2})\varphi_k \\ &= D_\pi^{-1/2}A\varphi_k \\ &= D_\pi^{-1/2}\lambda_k\varphi_k \\ &= \lambda_k v_k. \end{aligned}$$

Segue pela definição de v_k que $\varphi_k = D_\pi^{1/2} v_k$, logo para cada i, j, vale

$$\langle \varphi_i, \varphi_j \rangle = \langle D_\pi^{1/2} v_i, D_\pi^{1/2} v_j \rangle,$$

sendo $v_i = (\mu_1, \mu_2, \ldots, \mu_{|\Omega|})$ e $v_j = (\eta_1, \eta_2, \ldots, \eta_{|\Omega|})$. Então pela definição de D_π, obtemos

$$\begin{aligned}
\langle D_\pi^{1/2} v_i, D_\pi^{1/2} v_j \rangle &= D_\pi^{1/2}\mu_1 D_\pi^{1/2}\eta_1 + D_\pi^{1/2}\mu_2 D_\pi^{1/2}\eta_2 + \cdots + D_\pi^{1/2}\mu_{|\Omega|} D_\pi^{1/2}\eta_{|\Omega|} \\
&= D_\pi \mu_1 \eta_1 + D_\pi \mu_2 \eta_2 + \cdots + D_\pi \mu_{|\Omega|} \eta_{|\Omega|} \\
&= \pi(x_1)\mu_1\eta_1 + \pi(x_2)\mu_2\eta_2 + \cdots + \pi(x_{|\Omega|})\mu_{|\Omega|}\eta_{|\Omega|} \\
&= \sum_{x \in \Omega} v_i(x) v_j(x) \pi(x) = \langle v_i, v_j \rangle_\pi.
\end{aligned}$$

Logo, o conjunto $\{v_k\}_{k=1}^{|\Omega|}$ forma uma base ortonormal com relação ao produto interno $\langle \cdot, \cdot \rangle_\pi$, finalizando assim, o item (i).

Para demonstrar o item (ii), definimos uma função δ_y, tal que

$$\delta_y(x) = \begin{cases} 1 & \text{se } \ y = x \\ 0 & \text{se } \ y \neq x \end{cases}.$$

Segue de um teorema conhecido de Álgebra Linear que, se V é um espaço vetorial finito com base ortornomal $\langle v_1, v_2, \cdots, v_n \rangle$, então, qualquer vetor $u \in V$ pode ser escrito da forma

$$u = \frac{\langle u, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 + \frac{\langle u, v_2 \rangle}{\langle v_2, v_2 \rangle} v_2 + \ldots + \frac{\langle u, v_n \rangle}{\langle v_n, v_n \rangle} v_n.$$

Para a demonstração deste teorema, ver (LIPSCHUTZ; LIPSON, 2011), página 256.

Utilizando este resultado de Álgebra Linear, podemos reescrever δ_y como

$$\delta_y = \sum_{k=1}^{|\Omega|} \langle \delta_y, v_k \rangle_\pi v_k = \sum_{k=1}^{|\Omega|} v_k(y) \pi(y) v_k. \tag{3.3}$$

Como $P^t v_k = \lambda_k^t v_k$ para qualquer $k \in \Omega$, então segue que

$$P^t(x, y) = (P^t \delta_y)(x) = \sum_{k=1}^{|\Omega|} v_k(y) \pi(y) P^t v_k(x) = \sum_{k=1}^{|\Omega|} v_k(y) \pi(y) \lambda_k^t v_k(x).$$

Dividindo ambos os lados por $\pi(y)$, obtemos a equação

$$\frac{P^t(x, y)}{\pi(y)} = \sum_{k=1}^{|\Omega|} v_k(x) v_k(y) \lambda_k^t,$$

como queríamos demonstrar.

A demonstração do item (iii) segue da observação feita anteriormente, na qual vimos que $\sqrt{\pi}$ é um autovetor de A com associado autovalor 1. Deste modo, definimos $\lambda_1 = 1$ e $\varphi_1 = \sqrt{\pi}$. Assim, pela definição de v_k e D_π, temos

$$v_1 = \pi(x)^{-1/2}\pi(x)^{1/2} = 1, \; \forall x \in \Omega.$$

Como $\displaystyle\sum_{k=1}^{|\Omega|} v_k(x)v_k(y)\lambda_k^t = v_1(x)v_1(y)\lambda_1^t + \sum_{k=2}^{|\Omega|} v_k(x)v_k(y)\lambda_k^t$, então

$$\sum_{k=1}^{|\Omega|} v_k(x)v_k(y)\lambda_k^t = 1 + \sum_{k=2}^{|\Omega|} v_k(x)v_k(y)\lambda_k^t.$$

Conclui-se assim o item (iii). □

O estudo sobre Cadeias de Markov e Teoria Espectral é muito importante na criação de ferramentas que possibilitam encontrar limitantes para distâncias entre probabilidades em determinados tipos de cadeias. Veremos nas seções seguintes, alguns tipos de limitantes que são encontrados a partir desta relação entre autovalores e autovetores de uma matriz de transição.

3.2 Tempo de Relaxamento

Iniciaremos a seção estabelecendo conceitos e definições básicos e então, a partir dos resultados construídos na Seção 3.1, encontraremos um limitante superior e um limitante inferior para o tempo de mistura de cadeias reversíveis, aperiódicas e irredutíveis.

Sendo P uma matriz de transição de uma Cadeia de Markov reversível, podemos ordenar os autovalores de P como

$$1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_{|\Omega|} \geq -1. \tag{3.4}$$

Definição 3.3. *Denotamos* $\lambda_* \coloneqq \max\{|\lambda| : \lambda$ *é um autovalor de* P, *com* $\lambda \neq 1\}$.

Definição 3.4. *Dada P uma Cadeia de Markov reversível, chamamos de lacuna espectral a diferença entre os dois maiores autovalores de P, ou seja,*

$$\gamma = 1 - \lambda_2.$$

Além disso, definimos lacuna espectral absoluta como

$$\gamma_* = 1 - \lambda_*.$$

Pelo Lema 3.1 item (iii), temos que se P é irredutível e aperiódica, $\gamma_* > 0$.

Definição 3.5. *Seja P uma Cadeia de Markov reversível com lacuna espectral absoluta γ_*. Denotamos o tempo de relaxamento, t_{rel}, como*

$$t_{rel} := \frac{1}{\gamma_*}.$$

Teorema 3.1. *Seja P uma matriz de transição de uma Cadeia de Markov reversível, irredutível e aperiódica, com espaço de estados Ω e seja $\pi_{\min} := \min_{x\in\Omega} \pi(x)$. Então*

$$t_{mix}(\epsilon) \leq \log\left(\frac{1}{\epsilon\pi_{\min}}\right) t_{rel}.$$

Demonstração. Ao utilizar a igualdade (3.2) substituindo λ_k por λ_* e aplicando a desigualdade de Cauchy-Schwarz, temos

$$\left|\frac{P^t(x,y)}{\pi(y)} - 1\right| \leq \sum_{k=2}^{|\Omega|} |v_k(x)v_k(y)|\lambda_*^t \leq \lambda_*^t \left[\sum_{k=2}^{|\Omega|} v_k^2(x) \sum_{k=2}^{|\Omega|} v_k^2(y)\right]^{1/2}. \tag{3.5}$$

Pela equação (3.3) e utilizando a ortonormalidade de v_k, obtemos

$$\begin{aligned} \pi(x) = \langle \delta_x, \delta_x \rangle_\pi &= \left\langle \sum_{j=1}^{|\Omega|} v_j(x)\pi(x)v_j, \sum_{i=1}^{|\Omega|} v_i(x)\pi(x)v_i \right\rangle_\pi \\ &= \sum_{j=1}^{|\Omega|}\sum_{i=1}^{|\Omega|} v_j(x)v_i(x)\pi^2(x)\langle v_j, v_i\rangle = \pi(x)^2 \sum_{k=1}^{|\Omega|} v_k(x)^2. \end{aligned}$$

Logo, $\sum_{k=1}^{|\Omega|} v_k(x)^2 = \pi(x)^{-1}$, o que implica que $\sum_{k=2}^{|\Omega|} v_k(x)^2 \leq \pi(x)^{-1}$, ou seja, temos um limitante superior para a soma dos autovetores $\{v_k(x)^2\}_{k=2}^{|\Omega|}$. Assim, pela equação (3.5), temos que

$$\lambda_*^t \left[\sum_{k=2}^{|\Omega|} v_k^2(x) \sum_{k=2}^{|\Omega|} v_k^2(y)\right]^{1/2} \leq \lambda_*^t \left[\frac{1}{\pi(x)}\frac{1}{\pi(y)}\right]^{1/2} = \lambda_*^t \left[\frac{1}{\sqrt{\pi(x)}}\frac{1}{\sqrt{\pi(y)}}\right].$$

Deste modo, obtemos

$$\left|\frac{P^t(x,y)}{\pi(y)} - 1\right| \leq \frac{\lambda_*^t}{\sqrt{\pi(x)\pi(y)}} \leq \frac{\lambda_*^t}{\pi_{\min}} = \frac{(1-\gamma_*)^t}{\pi_{\min}} \leq \frac{e^{-\gamma_* t}}{\pi_{\min}}.$$

A segunda desigualdade é obtida através da definição de $\pi_{\min}$. Para a terceira desigualdade, utilizamos o fato de que para qualquer $x \in [0,1]$, vale $(1-x) \leq \epsilon^{-x}$ pois, para $x = 0$, obtemos $(1-0) \leq \epsilon^{-0}$ e para qualquer $x \in (0,1]$, temos que $\frac{d}{dx}(1-x) = -1$ é menor que $\frac{d}{dx}(\epsilon^{-x}) = -\epsilon^{-x}$.

Segue, pelo Lema 2.2, que $d(t) \leq s_x(t)$. Portanto, temos que $d(t) \leq \pi_{\min}^{-1} e^{-\gamma_* t}$. Pela definição de tempo de mistura, temos que $t_{mix}(\epsilon) \coloneqq \min\{t : d(t) \leq \epsilon\}$. Assim, para $t = t_{mix}(\epsilon)$, vale

$$d(t) \leq \epsilon \leq \frac{e^{-\gamma_* t_{mix}(\epsilon)}}{\pi_{\min}}.$$

Logo

$$\begin{aligned} e^{-\gamma_* t_{mix}(\epsilon)} &\geq \epsilon \pi_{\min} \\ -\gamma_* t_{mix}(\epsilon) &\geq \log(\epsilon \pi_{\min}). \end{aligned}$$

Deste modo,

$$\gamma_* t_{mix}(\epsilon) \leq -\log(\epsilon \pi_{\min}) = \log(\epsilon \pi_{\min})^{-1} = \log\left(\frac{1}{\epsilon \pi_{\min}}\right).$$

Com isso, concluímos que

$$t_{mix}(\epsilon) \leq \log\left(\frac{1}{\epsilon \pi_{\min}}\right) t_{rel},$$

como queríamos demonstrar. □

Teorema 3.2. *Para uma Cadeia de Markov reversível, irredutível e aperiódica, temos*

$$t_{mix}(\epsilon) \geq (t_{rel} - 1) \log\left(\frac{1}{2\epsilon}\right).$$

Demonstração. Considere v um autovetor de P com autovalor associado $\lambda \neq 1$. Como P é irredutível e aperiódica, pelo Lema 3.1 temos também que $\lambda \neq -1$. Além disso, como os autovalores de P formam uma base ortonormal em relação a $\langle \cdot, \cdot \rangle_\pi$, e pelo Lema 3.2 sabemos que $\mathbf{1}$ é um autovetor, então $\langle \mathbf{1}, v \rangle_\pi = \sum_{y \in \Omega} \pi(y) v(y) = 0$. Com isso, obtemos

$$\begin{aligned} |\lambda^t v(x)| &= \left| \sum_{y \in \Omega} P^t(x,y) v(x) \right| \\ &= \left| \sum_{y \in \Omega} P^t(x,y) v(x) - \sum_{y \in \Omega} \pi(y) v(y) \right| \\ &= \left| \sum_{y \in \Omega} \left[P^t(x,y) v(x) - \pi(y) v(y) \right] \right| \\ &\leq \sum_{y \in \Omega} \left| P^t(x,y) v(x) - \pi(y) v(y) \right| \\ &\leq \|v\|_\infty \sum_{y \in \Omega} |P^t(x,y) - \pi(y)| = \|v\|_\infty 2d(t). \end{aligned}$$

Tomando x tal que $|v(x)| = \|v\|_\infty$, temos

$$|\lambda^t|\|v\|_\infty = |\lambda^t v(x)| \leq \|v\|_\infty 2d(t) \Longrightarrow |\lambda^t| \leq 2d(t).$$

Logo, para o tempo de mistura $t = t_{mix}(\epsilon)$, vale

$$\lambda_*^{t_{mix}(\epsilon)} \leq 2\epsilon.$$

Ou seja, $\dfrac{1}{2\epsilon} \leq \dfrac{1}{\lambda_*^{t_{mix}(\epsilon)}}$, implicando que

$$\log\left(\frac{1}{2\epsilon}\right) \leq -t_{mix}(\epsilon)\log\lambda_*.$$

Deste modo, segue que

$$t_{mix}(\epsilon) \geq -\log\left(\frac{1}{2\epsilon}\right)\frac{1}{\log(\lambda_*)} = -\log\left(\frac{1}{2\epsilon}\right)\frac{1}{\log(1-\gamma_*)}. \tag{3.6}$$

Note que, para qualquer $x \in [0,1]$, vale que $-(1-x)\log(1-x) \leq x$, pois, se $x = 0$, obtemos $-(1-0)\log(1-0) \leq 0$, e para qualquer $x \in (0,1]$, $\frac{d}{dx}(-(1-x)\log(1-x)) = \log(1-x)+1$, é sempre menor que $\frac{d}{dx}(x) = 1$. Como $\gamma_* \in [0,1]$, então

$$\gamma_* \geq -(1-\gamma_*)\log(1-\gamma_*).$$

Ou seja, temos que $\dfrac{1}{\log(1-\gamma_*)} \leq -\left(\dfrac{1-\gamma_*}{\gamma_*}\right)$, implicando em

$$\frac{1}{\log(1-\gamma_*)} \leq -\left(\frac{1}{\gamma_*} - 1\right) = -(t_{rel} - 1). \tag{3.7}$$

Utilizando a inequação (3.7) em (3.6), obtemos

$$t_{mix}(\epsilon) \geq \log\left(\frac{1}{2\epsilon}\right)(t_{rel} - 1),$$

como queríamos demonstrar. □

Note que, os dois limitantes para o tempo de mistura apresentados nos Teoremas 3.1 e 3.2, representam respectivamente um valor superior e inferior para as distâncias de probabilidade $P(x,\cdot)$ e $\pi(\cdot)$ dentro de uma parâmetro ϵ, que por sua vez, é um parâmetro estabelecido para que estas distâncias de probabilidade em cadeias convergentes se aproximem da distribuição estacionária π. Além disso, temos que tais limitantes têm uma relação de proporcionalidade com o tempo de relaxamento visto que, quanto menor o tempo de relaxamento, mais o limitante vai se aproximar do valor superior de convergência. Por outro lado, quanto menor o tempo de relaxamento, menos o limitante se aproximará do valor inferior de convergência. Para valor de tempo de relaxamento grande, obtemos um resultado análogo.

3.3 Limitantes na norma l_2

Para esta seção, encontraremos limitantes superiores para distâncias entre probabilidades $P(x,\cdot)$ e $\pi(\cdot)$ em cadeias do tipo reversíveis e transitivas. Para tanto, iremos utilizar alguns resultados conhecidos do espaço l^p e da Probabilidade, como a Desigualdade de Jensen, que pode ser encontrado em (DURRETT, 2010).

Definição 3.6. *Dado o espaço vetorial $\mathbb{R}^{|\Omega|}$, definimos a norma $l^p(\pi)$ para cada $p \geq 0$, como*

$$\|v\|_p \coloneqq \left[\sum_{x\in\Omega} |v(x)|^p \pi(x)\right]^{1/p}.$$

Definição 3.7. *Sejam P uma Cadeia de Markov com espaço de estados Ω e distribuição estacionária π e $X : \Omega \to \mathbb{R}$ uma variável aleatória. Então, a esperança da variável aleatória com respeito à distribuição de probabilidade π é dada por*

$$\mathbb{E}_\pi(X) = \sum_{x\in\Omega} X(x)\pi(x).$$

Teorema 3.3. *Seja P uma Cadeia de Markov reversível com autovetores $\{v_k\}_{k=1}^{|\Omega|}$, associados aos autovalores $1 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_{|\Omega|} \geq -1$ e base ortonormal com respeito a $\langle\cdot,\cdot\rangle_\pi$. Então*

(i)

$$4\|P^t(x,\cdot) - \pi\|_{VT}^2 \leq \left\|\frac{P^t(x,\cdot)}{\pi(\cdot)} - 1\right\|_2^2 = \sum_{k=2}^{|\Omega|} v_k(x)^2 \lambda_k^{2t}.$$

(ii) Se P for uma cadeia transitiva, então

$$4\|P^t(x,\cdot) - \pi\|_{VT}^2 \leq \left\|\frac{P^t(x,\cdot)}{\pi(\cdot)} - 1\right\|_2^2 = \sum_{k=2}^{|\Omega|} \lambda_k^{2t}.$$

Demonstração. Inicialmente provaremos um resultado importante para esta demonstração, em que veremos que a distância entre as probabilidades $P(x,\cdot)$ e $\pi(\cdot)$ na norma l_1 é menor ou igual do que na norma l_2, ou seja,

$$\|P(x,\cdot) - \pi(\cdot)\|_1 \leq \|P(x,\cdot) - \pi(\cdot)\|_2.$$

Observe que, pela Proposição 2.1 e pela Definição 3.6, vale que

$$\begin{aligned}
2\|P^t(x,\cdot) - \pi(\cdot)\|_{VT} &= \sum_{y\in\Omega} |P^t(x,y) - \pi(y)| \\
&= \sum_{y\in\Omega} \pi(y) \left|\frac{P^t(x,y)}{\pi(y)} - 1\right| \\
&= \left\|\frac{P^t(x,y)}{\pi(y)} - 1\right\|_1.
\end{aligned}$$

Elevando ambos os termos ao quadrado, obtemos a igualdade

$$4\|P^t(x,\cdot)-\pi(\cdot)\|^2 = \left[\sum_{y\in\Omega}\pi(y)\left|\frac{P^t(x,y)}{\pi(y)}-1\right|\right]^2 = \left\|\frac{P^t(x,y)}{\pi(y)}-1\right\|_1^2. \tag{3.8}$$

Considerando $4\|P^t(x,\cdot)-\pi(\cdot)\|^2$ como uma esperança e utilizando a desigualdade de Jensen para $X(x)=x^2$. Então

$$\begin{aligned}
4\|P^t(x,\cdot)-\pi(\cdot)\|_{VT}^2 &= \left[\sum_{y\in\Omega}\pi(y)\left|\frac{P^t(x,y)}{\pi(y)}-1\right|\right]^2\\
&= \left[\mathbb{E}_\pi\left|\frac{P^t(x,y)}{\pi(y)}-1\right|\right]^2\\
&\leq \mathbb{E}_\pi\left[\left|\frac{P^t(x,y)}{\pi(y)}-1\right|^2\right]\\
&= \sum_{y\in\Omega}\pi(y)\left|\frac{P^t(x,y)}{\pi(y)}-1\right|^2\\
&= \left\|\frac{P^t(x,\cdot)}{\pi(y)}-1\right\|_2^2.
\end{aligned}$$

Pelo Lema 3.2 item(iii), segue que

$$\left\|\frac{P^t(x,\cdot)}{\pi(\cdot)}-1\right\|_2^2 = \left\|\sum_{k=2}^{|\Omega|}v_k(x)v_k\lambda_k^t.\right\|_2^2 = \sum_{k=2}^{|\Omega|}v_k(x)^2\lambda_k^{2t}. \tag{3.9}$$

Por fim, como a distância em variação total equivale à metade da distância na norma l_1, segue que

$$\left\|\frac{P^t(x,\cdot)}{\pi(\cdot)}-1\right\|_1^2 = 4\|P^t(x,\cdot)-\pi\|_{VT}^2 \leq \left\|\frac{P^t(x,\cdot)}{\pi(\cdot)}-1\right\|_2^2. \tag{3.10}$$

Assim, temos, pelas equações (3.9) e (3.10), que

$$4\|P^t(x,\cdot)-\pi\|_{VT}^2 \leq \left\|\frac{P^t(x,\cdot)}{\pi(\cdot)}-1\right\|_2^2 = \sum_{k=2}^{|\Omega|}v_k(x)^2\lambda_k^{2t},$$

concluindo assim o item (i).

Para o item (ii), considere uma Cadeia de Markov transitiva. Então pela Proposição 1.5, temos que a distribuição estacionária π é uniforme. Logo, para quaisquer $x,y\in\Omega$, vale

$$\begin{aligned}
\pi(y) &= \sum_{y\in\Omega}\pi(x)P^t(x,y)\\
&= \pi(\cdot)\sum_{y\in\Omega}P^t(x,y).
\end{aligned}$$

Portanto, segue que

$$\frac{1}{\pi(\cdot)} = \sum_{y\in\Omega} P^t(x,y)\frac{1}{\pi(y)}.$$

Como $\frac{1}{\pi(\cdot)}$ é constante para qualquer estado de Ω, dado que π é uma distribuição uniforme, então segue que o lado direito da última igualdade também é constante para quaisquer $x, y \in \Omega$. Logo, pela equação (3.9) e utilizando o resultado anterior, temos que para qualquer $x_0 \in \Omega$, vale

$$\left\|\frac{P^t(x_0,\cdot)}{\pi(\cdot)} - 1\right\|_2^2 = \sum_{k=2}^{|\Omega|} v_k(x)^2\lambda_k^{2t}. \tag{3.11}$$

Ao somar ambos os termos da equação (3.11) para todo $x \in \Omega$, obtemos

$$\sum_{x\in\Omega}\left\|\frac{P^t(x_0,\cdot)}{\pi(\cdot)} - 1\right\|_2^2 = \sum_{k=2}^{|\Omega|}\sum_{x\in\Omega} v_k(x)^2\lambda_k^{2t}. \tag{3.12}$$

Note que o lado esquerdo da equação é uniforme, então bastar calcular a distância de separação apenas uma vez e multiplicar pela cardinalidade de Ω. Além disso, multiplicando e dividindo por $\pi(x) = \frac{1}{|\Omega|}$ o lado direito da equação (3.12), obtemos

$$|\Omega|\left\|\frac{P^t(x_0,\cdot)}{\pi(\cdot)} - 1\right\|_2^2 = |\Omega|\sum_{k=2}^{|\Omega|}\left[\sum_{x\in\Omega} v_k(x)^2\pi(x)\right]\lambda_k^{2t}.$$

Como $\|v\|_2 = 1$, então a soma interior do lado direito da equação também será igual a 1. Logo,

$$\left\|\frac{P^t(x_0,\cdot)}{\pi(\cdot)} - 1\right\|_2^2 = \sum_{k=2}^{|\Omega|}\lambda_k^{2t}. \tag{3.13}$$

Ao combinar as equanções (3.13) com (3.10), obtém-se o resultado desejado. □

Nesta seção, vimos dois tipos de limitantes superiores, um para cadeias reversíveis, e o outro para cadeias reversíveis e transitivas. Note que são limitantes similares, de fato, pois a demonstração do item (ii), decorre de alguns resultados construídos no item (i). Além disso, para cada item, podemos utilizar dois tipos de limitantes que são equivalentes.

Capítulo 4

Considerações Finais

O presente trabalho buscou apresentar um estudo sobre Cadeias de Markov associado à Teoria Espectral. O trabalho foi desenvolvido de forma progressiva, iniciando com conceitos básicos de Probabilidade até conceitos mais elaborados que envolviam técnicas de demonstrações mais avançadas.

O estudo sobre autovalores, autovetores e produto interno foi bastante útil para criação de ferramentas que possibilitaram encontrar limitantes para distância entre probabilidades em determinados tipos de cadeia visto que, através destas ferramentas foi possível determinar limitantes superiores e inferiores para o tempo de mistura de cadeias reversíveis, irredutíveis e aperiódicas. Além disso, para cadeias reversíveis, vimos outro tipo de ferramenta que possibilitou encontrar limitantes superiores para distância entre probabilidades a partir dos autovetores e autovalores associados.

Conclui-se que o resultado deste trabalho foi bastante satisfatório, dado que através deste estudo foi possível explorar o conteúdo Cadeias de Markov em um tópico pouco difundido. Além disso, foi possível apresentar resultados matemáticos claros e acessíveis que permitiram um desenvolvimento e maturidade matemática bastante significativos para este tópico de Probabilidade.

Referências

CHU, Y.; HOUGH, R. *Markov Chains mixing time, random walk on Groups and the 15 puzzle problem*. New York: Stony Brook University Department of Mathematics, 2020.

DABBS, B. *Markov Chains and mixing times*. [S.l.]: Manuscript, 2009.

DURRETT, R. *Probability*: theory and examples. 4. ed. São Paulo: Cambridge University Press, 2010.

FRANCO, T. *Princípios de Combinatória e Probabilidade*. 1. ed. Rio de Janeiro: IMPA, 2020.

FREEDMAN, A. *Convergence Theorem for finite Markov Chains*. *Proc. REU*, 2017.

HAZZAN, S. *Fundamentos da Matemática 5*: Combinatória Probabilidade. 8. ed. São Paulo: Atual, 2013.

HOWARD, A.; RORRRES, C. *Álgebra Linear com aplicações*. 10. ed. Porto Alegre: Bookman, 2012.

LEVIN, D. A.; PERES, Y.; WILMER, E. L. *Markov Chains and mixing times*. Providence: American Mathematical Society, 2009.

LIMA, E. L. *Álgebra Linear*: Coleção Matemática Universitária. 1. ed. Rio de Janeiro: IMPA, 2014.

LIPSCHUTZ, S.; LIPSON, M. L. *Álgebra Linear*. 4. ed. Porto Alegre: Bookman, 2011.

ROSS, S. M. *Stochastic Processes*. 2. ed. New York - Chichester - Brisbane - Toronto - Singapore: John Wiley & Sons, 1996.

SILVA, D. S. *Convergência em variação total de Cadeias de Markov*. 2013. 60f. Trabalho de Conclusão de Curso (Monografia) - Instituto de Matemática e Estatística, Universidade Federal da Bahia, Salvador, 2013.