

Filipe Ribeiro Carneiro

**Cadeias de Markov:
Tempo de Mistura, *Cutoff* e Redes**

Vitória - Espírito Santo, Brasil

19 de fevereiro de 2016

Filipe Ribeiro Carneiro

**Cadeias de Markov:
Tempo de Mistura, *Cutoff* e Redes**

Dissertação apresentada ao Programa de Pós-graduação em Matemática da Universidade Federal do Espírito Santo PPG-MAT/UFES, como parte dos requisitos para obtenção do título de Mestre em Matemática. Orientador: Prof^o Fábio Júlio da Silva Valentim .

Universidade Federal do Espírito Santo – UFES
Departamento de Matemática
Programa de Pós-Graduação em Matemática

Orientador: Fábio Júlio da Silva Valentim

Vitória - Espírito Santo, Brasil
19 de fevereiro de 2016

Dados Internacionais de Catalogação-na-publicação (CIP)
(Biblioteca Central da Universidade Federal do Espírito Santo, ES, Brasil)

C289c Carneiro, Filipe Ribeiro, 1990-
Cadeias de Markov : tempo de mistura, *cutoff* e redes / Filipe
Ribeiro Carneiro. – 2016.
85 f. : il.

Orientador: Fábio Júlio da Silva Valentim.
Dissertação (Mestrado em Matemática) – Universidade
Federal do Espírito Santo, Centro de Ciências Exatas.

1. Redes elétricas. 2. Processo estocástico. 3.
Probabilidades. 4. Markov, Processos de. I. Valentim, Fábio Júlio
da Silva. II. Universidade Federal do Espírito Santo. Centro de
Ciências Exatas. III. Título.

CDU: 51



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
Centro de Ciências Exatas
Programa de Pós-Graduação em Matemática

“Cadeias de Markov: Tempo de Mistura, Cutoff e Redes”

Filipe Ribeiro Carneiro

Dissertação submetida ao Programa de Pós-Graduação em Matemática da Universidade Federal do Espírito Santo como requisito parcial para a obtenção do título de Mestre em Matemática.

Aprovada em 19/02/2016 por:

Assinatura manuscrita em azul de Fábio Júlio da Silva Valentim.

Fábio Júlio da Silva Valentim – UFES

Assinatura manuscrita em azul de João Paulo Costalonga.

João Paulo Costalonga – UFES

Assinatura manuscrita em azul de Freddy Rolando Hernandez Romero.

Freddy Rolando Hernandez Romero – UFF

A toda a minha família de sangue e de coração.

Agradecimentos

Primeiramente agradeço a Deus por ter me conduzido muito além do que imaginei, por ter me dado força e paciência para concluir mais essa fase em minha vida.

Agradeço aos meus pais, Jasson e Maria Sebastiana (mais conhecida como Tiana) pelo amor e carinho incondicional, sabendo me compreender e estimular nos momentos difíceis. Vocês são para mim um exemplo de vida e harmonia. Agradeço também aos meus irmãos Thiago e Lorena por me compreender nas ausências, pelo amor e paciência dados a mim.

Agradeço aos meus amigos que foram para mim uma segunda família. Agradeço ao Alexandre, Stéfani, Galo (vulgo Marcus Vinicius) e Manu (vulgo Emanuela) que estão comigo desde o início da graduação, sempre estando próximo e ouvindo as reclamações e dificuldades. A Roberta e Franciane, pelo carinho único que me deram deste o final da graduação e ao longo do mestrado. Ao Weverthon pelas longas conversas de orientação. Ao Solon pelo apoio e conversas na sala dos discentes.

Agradeço a Juliana por ter sido uma grata surpresa na minha vida, seu amor, paciência, atenção comigo foram primordiais.

Agradeço, de forma muito especial, ao meu orientador, Professor Fábio Júlio, pela enorme paciência nas minhas falhas e erros. Por sua disponibilidade que teve comigo ao longo dessa caminhada. Como diria *Sir* Isaac Newton: "Se eu vi longe, foi por estar sobre ombros de gigantes" e você, Fábio, foi esse ombro sob o qual eu estava apoiado. Acreditando em mim mais do que eu mesmo, sempre me estimulando. Tornando um grande exemplo de professor.

Agradeço a CAPES pelo apoio financeiro.

Resumo

Este trabalho aborda Cadeias de Markov em tempo discreto e espaço de estados finito. Tratamos a convergência desses objetos e para tal definimos a distância de variação total e estudamos algumas de suas propriedades, além de formas de estimar o tempo de mistura, como por exemplo, usando os autovalores da matriz de transição. Apresentamos a relação biunívoca entre Cadeias de Markov Reversíveis e as Redes Elétricas, além de como a teoria de redes pode ajudar no contexto de Cadeias de Markov. Definimos o Fenômeno de *Cutoff*, mostramos alguns resultados e concluímos com o Contra-exemplo de Aldous.

Palavras-chaves: Cadeias de Markov. Tempo de mistura. *Cutoff*. Redes Elétricas.

Abstract

This work deals with Markov chains in discrete time and finite state space. We treat the convergence of these objects, and define the total variation distance and studied some of their properties , as well as ways of estimating the mixing time, for example, using the eigenvalues of the transition matrix. Present is the one by one relationship between reversible Markov chains and graphs, and how the network theory can help in Markov Chains context. We also define and show some results concerning the so called Cutoff phenomenon, concluding by exhibiting a counter-example due to Aldous.

Key-words: Markov chain. Mixing times. Cutoff. Electrical networks

Sumário

	INTRODUÇÃO	15
1	CONCEITOS INICIAIS DE CADEIAS DE MARKOV	19
1.1	Cadeia de Markov Finita	19
1.2	Cadeia Irredutível e Aperiódica	23
1.3	Distribuição Estacionária	26
1.4	Cadeia Reversível	28
2	CONVERGÊNCIA DE CADEIA DE MARKOV	33
2.1	Distância de Variação Total	33
2.2	Distância Padrão	36
2.3	O Teorema da Convergência e o Tempo de Mistura	39
3	LIMITANDO O TEMPO DE MISTURA	45
3.1	A Representação Espectral	45
3.2	O Tempo de Relaxação	48
3.3	Relação de Abertura	52
4	O FENÔMENO <i>CUTTOFF</i>	55
4.1	Definição e Primeiro Resultados	55
4.2	Contra-exemplo do Aldous	60
5	PASSEIO ALEATÓRIO E REDE	65
5.1	Passeio aleatório no Grafo	65
5.2	Redes e Cadeias de Markov Reversíveis	67
5.3	Funções Harmônicas	70
5.4	Voltagem e Fluxo de Corrente	72
5.5	Resistência Efetiva	75
5.6	Robustez	82
	REFERÊNCIAS	87

Introdução

Desde o seu surgimento, no início do século XX, as Cadeias de Markov tem ganhado muito espaço em frente de pesquisa, seja pela beleza de sua teoria ou pela sua capacidade de atingir uma grande quantidade de áreas da ciência, tais como Física, Economia, Meteorologia, Biologia, Engenharia e as mais diversa áreas da Matemática e ciências afim.

Cadeias de Markov é um exemplo de processo estocástico em que, a cada unidade de tempo, um elemento da cadeia é escolhido sob uma distribuição de probabilidade, de forma que o passado é independente do futuro, o que significa que somente o estado atual é relevante para a predição do estado seguinte. Esta propriedade citada é chamada de propriedade de Markov ou propriedade markoviana em homenagem ao matemático russo Andrei Andreyevich Markov (1856-1922).

O objetivo principal desta dissertação é introduzir rigorosamente os conceitos de Cadeias de Markov e seus desdobramentos, por exemplo, convergência de cadeias de Markov, o fenômeno de *cutoff* e redes elétricas.

Iremos considerar cadeias de Markov a tempo discreto e espaço de estado finito. O nosso primeiro foco será a convergência de cadeias de Markov exposto no Capítulo 2. Quando estamos querendo estudar a convergência de um objeto a primeira preocupação ou questionamento é saber em qual métrica este objeto está convergindo. Desta forma, iniciamos o Capítulo 2 definindo e explorando as propriedades da distância de variação total, assim sendo, já estamos aptos a enunciar o resultado primordial desse capítulo:

Teorema da Convergência 2.3.1 Suponha que P seja irredutível e aperiodica, com distribuição estacionária π . Então existe uma constante $\alpha \in (0, 1)$ e $C > 0$ tal que,

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

Para que possamos entender o Teorema da Convergência temos que compreender bem as suas hipóteses, ou seja, irredutibilidade e aperiodicidade. De maneira informal, podemos dizer que irredutibilidade nos garante que a cadeia não irá ficar restrita a apenas alguns estados, ou seja, é possível alcançar qualquer estado a partir de um estado inicial, mesmo que para isso se faça necessário um tempo t_0 . A aperiodicidade é caracterizada pelo fato de poder atingir qualquer estado independente de períodos de tempo, por exemplo, não precisa aguardar um tempo par para atingir um estado estando num estado inicial. Esses hipóteses são conceitos iniciais sobre cadeias de Markov que são apresentados e entendidos no Capítulo 1, juntamente com eles, vale salientar, são estudados os conceitos de distribuição estacionária e reversibilidades, além de outros conceitos.

Por causa do Teorema da Convergência definimos o tempo de mistura $t_{mix}(\varepsilon)$ que é

um parâmetro que mede quão próximo estamos da convergência dessa cadeia com um erro de ε . Em seguinte vemos, no Capítulo 3, formas de estimar este parâmetro. A primeira forma é olhando para os autovalores da matriz da transição e, desta forma, definimos o tempo de relaxação t_{rel} . Com esse novo parâmetro alcançamos uma limitação inferior e superior para o $t_{mix}(\varepsilon)$, como segue abaixo,

$$(t_{rel} - 1) \log \left(\frac{1}{2\varepsilon} \right) \leq t_{mix}(\varepsilon) \leq t_{rel} \log \left(\frac{1}{\varepsilon \min_{x \in \Omega} \pi(x)} \right).$$

Por fim, terminamos esse capítulo expondo a relação de abertura que utiliza a geometria do espaço de estado para estimar o tempo de mistura.

No Capítulo 4 apresentamos o Fenômeno de *Cutoff* que vai tratar o caso de sequências de cadeias de Markov e o comportamento abrupto na curva do gráfico *tempo* \times *distância estacionária*. Ainda sobre esse capítulo, trabalhamos a relação sobre três conceitos dessa vertente: a definição de *cutoff* e suas equivalências; a definição de *pré-cutoff* e a condição produto. Durante algum tempo, se acreditou que a condição produto era equivalente a definição de *cutoff*, porém Aldous mostrou que essa afirmação é falsa com um contra-exemplo. Iremos construir, no final desse capítulo, um contra-exemplo com a mesma finalidade que é uma modificação do contra-exemplo original do Aldous.

Quando estamos com a hipótese de reversibilidade na cadeia de Markov sempre podemos fazer uma associação biunívoca com redes elétricas, que de forma simplificada é um grafo cuja arestas possuem pesos que chamaremos de condutâncias. Visto isso, definimos, com rigor matemático, os objetos que normalmente aparecem em redes, como o fluxo de corrente, a voltagem e a resistência. Vemos também que com essas definições ainda se preservam os resultados de redes, por exemplo, a redução da rede para uma rede mais simples levando em conta que as resistências estão em série ou em paralelo. O resultado abaixo motiva esse estudo.

Proposição 5.5.3 Para quaisquer $a, z \in \Omega$ vale,

$$P_a\{\tau_z < \tau_a^+\} = \frac{1}{C(a)\mathcal{R}(a \leftrightarrow z)} = \frac{\mathcal{C}(a \leftrightarrow z)}{C(a)}.$$

De maneira informal a proposição acima diz que estando no estado a a probabilidade de chegar no estado z antes de passar no estado a é simplesmente calcular a Condutância efetiva, denotada por $\mathcal{C}(a \leftrightarrow z)$, ou a Resistência efetiva, denotada por $\mathcal{R}(a \leftrightarrow z)$ e o termo $C(a)$ que é somar as condutâncias no ponto a . Informalmente, a Condutância efetiva é a condutância da rede equivalente após a redução para uma rede mais simples ficando apenas os pontos a e z e a Resistência efetiva é, simplesmente, o inverso multiplicativo da Condutância efetiva. Terminamos este capítulo definindo o que seria uma família de rede robusta e mostrando que existe uma família de redes uniformemente limitadas que não é robusta.

Algumas provas de resultados serão omitidas para deixar o texto direto, pois para a demonstração de alguns resultados seria necessária a apresentação de mais teoria, tirando assim, o foco deste texto. Para os capítulos 1 e 2 foram usados as referências [1, 3, 5] já para o capítulo 3, foi usado apenas a referência [3]. Já para o capítulo 4 utilizamos as referências [3, 7]. Para o leitor que deseja mais informações acerca do fenômeno de *cutoff* recomendamos a leitura do artigo [7]. Por fim, para o capítulo 5, as referências são [2–4, 6], ao leitor que gostaria de se aprofundar mais sobre esse assunto, recomendamos a leitura de [4].

1 Conceitos Iniciais de Cadeias de Markov

Neste capítulo iremos apresentar a definição de Cadeia de Markov e suas características. Para facilitar o entendimento, estes novos conceitos serão sempre seguidos por exemplos.

Supomos que o leitor já esteja habituado com linguagem da Teoria da Probabilidade. Iremos recordar alguns conceitos de forma rápida apenas com o intuito de fixar a notação e unificar a linguagem. Aos que necessitem de maiores detalhes recomendamos o livro *Probabilidade: Um Curso de Nível Intermediário* de Barry R. James para leitores inexperientes e, para uma segunda leitura, o livro clássico *A Course in Probability Theory* de Kai L. Chung.

1.1 Cadeia de Markov Finita

Seja Ω um conjunto finito. Denotemos $\Omega = \{x_1, x_2, \dots, x_{|\Omega|}\}$, onde $|\Omega|$ é a cardinalidade de Ω . Cada $x \in \Omega$ será chamado de estado de Ω e Ω de espaço de estados.

Definição 1.1.1. *Um vetor $\mu = (\mu_i : \mu_i \in [0, 1])$ é uma distribuição de probabilidade em Ω (ou simplesmente distribuição em Ω) se:*

- i) $\mu_i \geq 0$;
- ii) $\sum \mu_i = 1$ (*Propriedade de massa total*).

Fixemos um Espaço de Probabilidade (Ω, \mathcal{F}, P) onde $\mathcal{F} = \{A : A \subset \Omega\}$, ou seja \mathcal{F} é o conjunto das partes de Ω , e P é uma distribuição de probabilidade em Ω .

Definição 1.1.2 (Matriz Estocástica). *Dizemos que uma matriz quadrada $P = (p_{ij})$ é **Estocástica** se cada linha é uma distribuição em Ω .*

Note que P por vezes denota a distribuição de probabilidade e por outras denota a matriz estocástica, mas o contexto deixará clara a diferença. Nesta dissertação vamos considerar uma correspondência entre matrizes de transição e diagramas conforme ilustrado nos exemplos abaixo. Para simplificar o desenho, iremos omitir a seta de um estado nele mesmo, sendo que o mesmo pode ser sabido usando que a soma de uma distribuição de probabilidade é 1.

$$P_1 = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} \quad P_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1/2 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}.$$

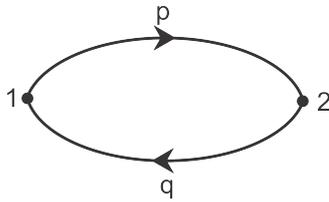


Figura 1 – Cadeia de dois estados correspondente a matriz P_1

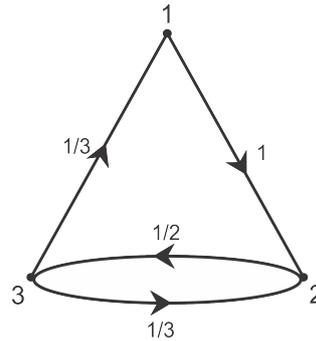


Figura 2 – Cadeia de três estados correspondente a matriz P_2

Denote por \mathbb{N} o conjunto $\mathbb{N} = \{0, 1, 2, 3, \dots\}$ e recorde que a definição de probabilidade condicional é dados $A, B \in \mathcal{F}$ então

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

desde que $P(B) > 0$. Caso $P(B) = 0$ a probabilidade de $P(A | B) = P(A)$.

Definição 1.1.3 (Cadeia de Markov). *Uma cadeia de Markov com espaço de estado finito Ω a tempo discreto \mathbb{N} é uma sequência de variáveis aleatórias $(X_n)_{n \geq 0}$, $X_n : \Omega \rightarrow \Omega$, tal que*

$$P\{X_{t+1} = y | X_t = x_t, \dots, X_0 = x_0\} = P\{X_{t+1} = y | X_t = x_t\} \quad (1.1)$$

para todo $t \in \mathbb{N}$, $x_0, x_1, \dots, x_t \in \Omega$. Em algumas situações consideremos o tempo finito, ou seja, $t \in \{0, 1, 2, \dots, N\}$ ao invés de \mathbb{N} .

Informalmente uma cadeia de Markov com espaço de estado finito Ω e tempo discreto é um processo estocástico que, para cada $y \in \Omega$ e $t \in \mathbb{N}$, a probabilidade condicional do processo estar no estado y no instante $t + 1$ sabendo tudo que ocorreu até o instante t é a mesma probabilidade condicional conhecendo apenas o estado no instante t .

A equação (1.1) é chamada de propriedade de Markov. De forma intuitiva, o conhecimento das variáveis aleatórias X_0, X_1, \dots, X_{t-1} não acrescenta nenhuma informação para prever o valor de X_{t+1} , em outras palavras, condicionando o presente, o passado é independente do futuro.

Vamos considerar que o valor do lado direito de (1.1) independe de t , a dependência se dá apenas em x, y . Neste caso dizemos que a cadeia de Markov é homogênea no tempo. Podemos condensar toda a dinâmica do processo em uma matriz quadrada de ordem $|\Omega| \times |\Omega|$ que iremos definir a seguir.

Definição 1.1.4. *Dado uma cadeia de Markov $(X_n)_{n \geq 0}$, homogênea no tempo, a matriz de transição P é definida por*

$$P\{X_{t+1} = y | X_t = x\} := P(x, y) = p_{xy}$$

Neste trabalho, vamos considerar apenas cadeias homogêneas no tempo. Sem grandes dificuldades vemos que a matriz de transição é uma matriz estocástica, pois qualquer linha de P é uma distribuição $P(x, \cdot)$, logo as suas entradas são não-negativas e $\sum_{y \in \Omega} P(x, y) = 1$ para todo $x \in \Omega$. Abaixo veremos, como exemplo, um caso simples de Cadeia de Markov.

Exemplo 1.1.5 (Cadeia de Dois Estados). *Considere que Ω possua dois estados e iremos nomeá-los como estados 1 e 2 e representá-lo como no diagrama abaixo. A dinâmica consiste assim: suponha que exista uma moeda em cada estado e o lançamento da moeda é que decide se troca de estado ou não, se ao lançar a moeda der cara muda de estado e permanece se der coroa.*

Contudo também vamos supor que as moedas não sejam justas, assim considere que a moeda do estado 1 tem probabilidade p de sair cara, enquanto a moeda do estado 2 tem probabilidade q de sair cara.

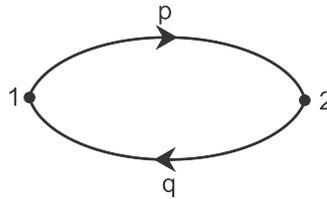


Figura 3 – Diagrama de uma Cadeia de Markov de dois Estados

Seja (X_0, X_1, \dots) a sequência de estados ocupados ao longo do tempo e com a dinâmica explicada acima podemos escrever a matriz de transição

$$P = \begin{pmatrix} P(1,1) & P(1,2) \\ P(2,1) & P(2,2) \end{pmatrix} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

Um questionamento natural é qual a distribuição de X_t para $t \geq 1$. Vamos supor que estamos começando a nossa dinâmica no estado $X_0 = 1$, e queremos saber os seguintes valores, $P\{X_2 = 1|X_0 = 1\}$ e $P\{X_2 = 2|X_0 = 1\}$. Esta questão pode ser resolvida usando propriedades de probabilidade condicional, pois

$$\begin{aligned} P\{X_2 = 1|X_0 = 1\} &= \\ P\{X_2 = 1|X_1 = 1\}P\{X_1 = 1|X_0 = 1\} &+ P\{X_1 = 2|X_0 = 1\}P\{X_2 = 1|X_1 = 2\} \\ &= (1-p)(1-p) + pq \end{aligned}$$

e analogamente, $P\{X_2 = 2|X_0 = 1\} = (1-p)p + p(1-q)$. Poderíamos continuar a escrever expressões como essa, porém quanto maior for o tempo, maior também fica a expressão, porém podemos usar uma abordagem mais eficaz. Podemos escrever estas informações num vetor de distribuição,

$$\mu_t = (P\{X_t = 1|X_0 = 1\}, P\{X_t = 2|X_0 = 1\}). \quad (1.2)$$

Como estamos supondo que começamos no estado $X_0 = 1$, então $\mu_0 = (1, 0)$ e $\mu_1 = (1 - p, p) = \mu_0 P$. Note que pelas contas que fizemos acima resulta que $\mu_2 = \mu_1 P = \mu_0 P^2$. Repetindo o processo, conseguimos descobrir qual o próximo passo da distribuição μ_t com a seguinte expressão:

$$\mu_t = \mu_{t-1} P = \mu_0 P^t \text{ para todo } t \geq 1. \quad (1.3)$$

De um modo geral, temos o seguinte resultado cuja prova pode ser encontrada na referência [5] na seção 1 do capítulo 1.

Teorema 1.1.6. *Seja $(X_n)_{n \geq 0}$ uma cadeia de Markov com distribuição inicial μ e $|\Omega| = N$. Então para quaisquer $n, m \geq 0$, vale:*

$$i) P\{X_n = j\} = (\mu P^n)_j.$$

$$ii) P\{X_n = j \mid X_0 = i\} = P\{X_{m+n} = j \mid X_m = i\} = p_{ij}^{(n)}$$

Um de nossos interesses é entender como se comporta a distribuição μ_t quando o tempo tende ao infinito, isso será mais estudado na Seção 1.3.

Este resultado formaliza a resposta ao seguinte questionamento: qual a probabilidade de após n passos estar no estado y sabendo que começou no estado x ? E com isso percebemos um pouco da importância da matriz de transição como ferramenta para auxiliar na resposta da pergunta.

Teorema 1.1.7. *Uma sequência de variáveis aleatórias $(X_n)_{n \geq 0}$ é uma cadeia de Markov se, e somente se, para todo $x_1, x_2, \dots, x_N \in \Omega$ vale*

$$P\{X_0 = x_0, X_1 = x_1, \dots, X_N = x_N\} = \mu_0 p_{x_0 x_1} p_{x_1 x_2} \cdots p_{x_{N-1} x_N}. \quad (1.4)$$

em que μ_0 é a distribuição inicial.

Demonstração. Suponha que $(X_n)_{n \geq 0}$ é uma cadeia de Markov e fixa $N \in \mathbb{N}$. Usando a definição de probabilidade condicional, podemos escrever

$$\begin{aligned} P\{X_0 = x_0, \dots, X_N = x_N\} &= \\ P\{X_N = x_N \mid X_0 = x_0, \dots, X_{N-1} = x_{N-1}\} &P\{X_0 = x_0, \dots, X_{N-1} = x_{N-1}\}. \end{aligned}$$

Prosseguindo de forma e usando a propriedade de Markov (1.1), conseguimos,

$$\begin{aligned} P\{X_0 = x_0, \dots, X_N = x_N\} &= \\ P\{X_N = x_N \mid X_0 = x_0, \dots, X_{N-1} = x_{N-1}\} &P\{X_0 = x_0, \dots, X_{N-1} = x_{N-1}\} \\ = P\{X_N = x_N \mid X_0 = x_0, \dots, X_{N-1} = x_{N-1}\} &\cdots P\{X_1 = x_1 \mid X_0 = x_0\} P\{X_0 = x_0\} \\ = P\{X_N = x_N \mid X_{N-1} = x_{N-1}\} &\cdots P\{X_1 = x_1 \mid X_0 = x_0\} P\{X_0 = x_0\} \\ = \mu_0 p_{x_0 x_1} p_{x_1 x_2} \cdots p_{x_{N-1} x_N}. \end{aligned}$$

Onde tomamos $\mu_0 = P\{X_0 = x_0\}$ e o resultado segue. Por outro lado, dado que $(X_n)_{n \geq 0}$ satisfaz (1.4) vamos mostrar que $(X_n)_{n \geq 0}$ é uma cadeia de Markov. Se vale (1.4), em particular, $P\{X_0 = x_0\} = \mu_0$ e, para cada $n \in \mathbb{N}$ temos,

$$P\{X_{n+1} = x_{n+1} \mid X_0 = x_0, \dots, X_n = x_n\} = \frac{P\{X_0 = x_0, \dots, X_{n+1} = x_{n+1}\}}{P\{X_0 = x_0, \dots, X_n = x_n\}} = p_{x_n x_{n+1}}.$$

Logo $(X_n)_{n \geq 0}$ é uma cadeia de Markov. □

1.2 Cadeia Irredutível e Aperiódica

Definição 1.2.1 (Irredutível). *Uma cadeia de Markov com matriz de transição P é dita **Irredutível** se para quaisquer dois estados $x, y \in \Omega$ existe um inteiro positivo t , possivelmente dependendo de x e y , tal que $P^t(x, y) > 0$. Caso contrário, chamamos de **Redutível**.*

Isso significa que uma cadeia de Markov é irredutível se é possível obter a partir de qualquer estado um outro estado qualquer apenas iterando a cadeia. No geral iremos sempre estudar cadeias com esta propriedade. Para entender melhor esta propriedade vejamos os exemplos a seguir.

Exemplo 1.2.2. *Considere uma cadeia de Markov com o espaço de estados $\{1, 2, 3, 4, 5\}$ e a seguinte matriz de transição,*

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/4 & 3/4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix}$$

$$\text{onde } P_1 = \begin{pmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix} \text{ e } P_2 = \begin{pmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{pmatrix}.$$

Dependendo do estado inicial de x_0 o estudo da cadeia com a matriz P se resume no estudo apenas de P_1 ou P_2 , ou seja, se x_0 for um estado pertencente a $\{1, 2\}$, os estados posteriores irão pertencer apenas a $\{1, 2\}$ e assim a matriz de transição relevante é P_1 , de forma análoga se x_0 pertencer a $\{2, 3, 4, 5\}$. A partir deste exemplo, não fica difícil de acreditar que caso a matriz de transição da cadeia tenha um bloco de zeros (similar ao exemplo) então a cadeia é redutível.

Exemplo 1.2.3. *Considere uma cadeia com três estados, digamos $\Omega = \{1, 2, 3\}$ com o seguinte diagrama*

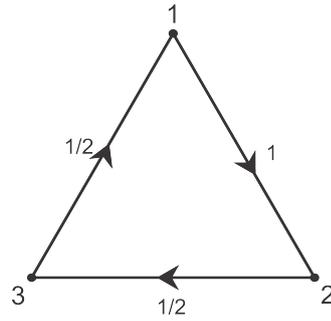


Figura 4 – Diagrama de uma Cadeia de Markov de três Estados

e matriz de transição

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix}.$$

Vamos mostrar que esta cadeia de Markov é irredutível olhando apenas a potência da sua matriz de transição, sem grandes dificuldades temos,

$$P^2 = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/4 & 1/4 & 1/2 \\ 1/4 & 1/2 & 1/4 \end{pmatrix}, P^3 = \begin{pmatrix} 1/4 & 1/4 & 1/2 \\ 1/4 & 3/8 & 3/8 \\ 1/8 & 1/2 & 3/8 \end{pmatrix}.$$

Agora usando a definição 1.2.1 vemos que para o inteiro $t = 3$ já temos que $P^3(x, y) > 0$ para todo $x, y \in \Omega$.

Iremos sempre dizer que a matriz (ou matriz de transição) possui uma dada propriedade, por exemplo a matriz é irredutível, quando queremos dizer que a cadeia de Markov com matriz de transição P é irredutível, achamos que esta forma completa é mais pedante sem nenhum acréscimo, a menos quando houver a possibilidade de confusão, usaremos sempre a forma mais enxuta.

Vamos introduzir mais uma característica, suponha que estamos no espaço de estado $\Omega = \{0, 1, 2\}$ e que $X_0 = 0$, e suponha que podemos apenas retornar para o estado 0 num tempo, digamos, par. Não é muito difícil de acreditar que este tipo de cadeia exista. Neste caso, o estado 0 apresenta um comportamento bem peculiar, o tempo de visita ao estado 0 demonstra uma periodicidade. Abaixo vamos definir esta característica rigorosamente e logo após um exemplo para melhorar o entendimento.

Definição 1.2.4 (Período). *Considere o conjunto $\mathcal{T}(x) := \{t \geq 1 \mid P^t(x, x) > 0\}$ o conjunto de tempos da cadeia em que é possível retornar a posição inicial x . O **Período** do estado x é definido como o máximo divisor comum do conjunto $\mathcal{T}(x)$. Para uma cadeia irredutível, o período da cadeia é definido como sendo o período que é comum a todos os estados. A cadeia será chamada de **Aperiódica** se todos os estados tem período 1. Se uma cadeia não é aperiódica a chamaremos de periódica.*

Exemplo 1.2.5. *Suponha que se tem uma máquina que pode estar em três possíveis situações: 1 - funcionando; 2 - em reparo; 3 - inoperante. Essa máquina é observada apenas quando há mudança de estado. Vamos supor que se a máquina esta funcionando então tem a mesma chance de ficar em reparo ou inoperante, e se a máquina esta em reparo ou inoperante ela volte a funcionar. Desta forma podemos modelar este problema com uma cadeia de Markov com a seguinte matriz de transição,*

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

Não é difícil de perceber que está cadeia é periódica com período 2 e vale

$$P^{2n} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{pmatrix}, \quad P^{2n-1} = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Cadeias e estados que são periódicos são sinônimo de desinteresse na teoria de cadeias de Markov, pois sem essa hipótese de aperiodicidade a distribuição $P^t(x_0, \cdot)$ não converge quando $t \rightarrow \infty$ para qualquer $x_0 \in \Omega$, veremos isso melhor no Teorema da Convergência 2.3.1. Felizmente, uma simples modificação na matriz de transição já soluciona este problema. Dada a matriz de transição P , considere a matriz $P_L := \frac{I+P}{2}$, onde I é a matriz identidade de ordem $|\Omega| \times |\Omega|$. É fácil ver que a matriz P_L é uma matriz estacionária e $P_L(x, x) > \frac{1}{2}$ para todos os estados $x \in \Omega$, logo a matriz P_L é aperiódica. Chamamos P_L como a **Forma Lazy** de P .

Proposição 1.2.6. *Se P é aperiódica e irredutível então existe um inteiro r tal que $P^r(x, y) > 0$ para quaisquer $x, y \in \Omega$.*

A demonstração dessa proposição pode ser encontrada na seção 3 do capítulo 1 da referência [3].

Vale a pena salientar a diferença entre a definição 1.2.1 e a Proposição 1.2.6. Dados $x, y \in \Omega$ a definição 1.2.1 nos garante que em algum momento t_0 a probabilidade $P^{t_0}(x, y) > 0$, mas não temos a garantia que no tempo seguinte (ou em qualquer outro) essa probabilidade continua positiva (veja o Exemplo 1.2.5). Diferentemente, a Proposição 1.2.6 nos garante que a partir do tempo r a probabilidade $P^r(x, y) > 0$ para todo $x, y \in \Omega$, ou seja, temos probabilidade positiva de saltar em r passos de um estado para o outro para quaisquer dois estados de Ω . Olhando para a matriz temos que quando ela é (apenas) irredutível e escolhendo uma entrada, segue da definição que vai existir um tempo que essa entrada vai ser positiva sem nenhuma garantia do tempo seguinte, contudo a Proposição 1.2.6 nos garante que a partir de um determinado tempo todas as entradas da matriz serão positivas.

1.3 Distribuição Estacionária

Iremos definir um vetor de distribuição ao qual chamaremos de Distribuição Estacionária que será invariante pela ação de P com o tempo. Podemos pensar que a definição deste vetor é motivada pelo Exemplo 1.1.5 quando o $t \rightarrow \infty$ na expressão $\mu_t = \mu_{t-1}P$, ou seja, o que acontece com qualquer vetor de distribuição inicial quando o tempo tende a infinito. Será que qualquer distribuição inicial converge para uma distribuição estacionária? Mais adiante, será visto que toda cadeia de Markov, com algumas propriedades, tende a se aproximar da sua distribuição estacionária quando o tempo tende ao infinito.

Definição 1.3.1. *Uma distribuição π em Ω satisfazendo a equação*

$$\pi = \pi P \tag{1.5}$$

*será chamada de **distribuição estacionária** da cadeia de Markov com matriz de transição P . Também podemos escrever (1.5) elemento a elemento como segue*

$$\pi(y) = \sum_{x \in \Omega} \pi(x)P(x, y) \quad \forall y \in \Omega.$$

Podemos pensar em (1.5) que estamos procurando o autovetor da matriz transposta de P correspondente ao autovalor 1. Um dos nossos objetivos é demonstrar que sob certas condições, as cadeias de Markov convergem para as suas distribuições estacionárias. Contudo, vamos antes nos familiarizar com o este conceito e provar sob quais condições a distribuição estacionária existe e se é possível falar em unicidade.

Exemplo 1.3.2. *Como primeiro exemplo, vamos encontrar a distribuição estacionária de uma cadeia de dois estados (ver Exemplo 1.1.5). Vamos escrever $\pi = (\pi(1), \pi(2))$, daí*

$$(\pi(1), \pi(2)) = (\pi(1), \pi(2)) \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}.$$

Lembrando que π é uma distribuição, então a soma de seus elementos é sempre igual a 1, ou seja, $\pi(1) + \pi(2) = 1$. Logo ficamos com o seguinte sistema de equações

$$\begin{cases} \pi(1) = \pi(1)(1-p) + \pi(2)q \\ \pi(2) = \pi(1)p + \pi(2)(1-q) \\ 1 = \pi(1) + \pi(2) \end{cases}$$

Segue daí, o seguinte resultado

$$\pi(1) = \frac{q}{p+q}, \quad \pi(2) = \frac{p}{p+q}.$$

Exemplo 1.3.3. *Se P é uma matriz de transição simétrica, ou seja, $P(x, y) = P(y, x)$, então a distribuição uniforme é estacionária para P .*

De fato, escrevendo $n = |\Omega|$ teremos que $\pi(x) = \frac{1}{n}$ para todo $x \in \Omega$, pois

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

Proposição 1.3.4. *Seja P uma matriz de transição de uma cadeia de Markov irredutível. Então existe uma única distribuição estacionária π em Ω tal que $\pi = \pi P$ e $\pi(x) > 0$ para todo $x \in \Omega$.*

Demonstração. A prova da existência será omitida e pode ser vista na seção 5 do capítulo 1 de [3]. Mas é simples provar a fato que $\pi(x) > 0$ para todo $x \in \Omega$. Com efeito, suponha, por contradição, que existe $x \in \Omega$ tal que $\pi(x) = 0$.

Como P é irredutível existe $y \in \Omega \setminus \{x\}$ com $P(y, x) > 0$ e como vale que

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

então $\pi(y) = 0$. Novamente como P é irredutível existe $z \in \Omega \setminus \{x, y\}$ tal que $P(z, x) > 0$ ou $P(z, y) > 0$. Assim se $P(z, y) > 0$ temos que

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

então $\pi(z) = 0$, análogo para $P(z, x) > 0$.

Desta forma, conseguimos estender este procedimento para todos os elementos de Ω , portanto $\pi(\omega) = 0$ para todo $\omega \in \Omega$, o que é uma absurdo com o fato de π ser uma distribuição.

Agora suponha por contradição que existem duas distribuições, digamos π_1 e π_2 .

Se existe $C \in \mathbb{R}$ tal que $\frac{\pi_1(x)}{\pi_2(x)} = C$ para todo $x \in \Omega$, então temos que $C = 1$, pois reescrevendo e aplicando o somatório em x temos,

$$1 = \sum_x \pi_1(x) = C \sum_x \pi_2(x) = C.$$

Se $\frac{\pi_1(x)}{\pi_2(x)}$ não for constante existe um estado $x_0 \in \Omega$ que minimiza este quociente, ou seja, $\frac{\pi_1(x_0)}{\pi_2(x_0)} \leq \frac{\pi_1(\omega)}{\pi_2(\omega)}$ para todo $\omega \in \Omega$. Por outro lado, existe ao menos um $\omega \in \Omega$ tal que vale $\frac{\pi_1(x_0)}{\pi_2(x_0)} < \frac{\pi_1(\omega)}{\pi_2(\omega)}$, assim usando a hipótese de estacionalidade de π_1 , irredutibilidade de P e multiplicando e dividindo por um termo conveniente,

$$\pi_1(x_0) = \sum_y \pi_1(y)P(y, x_0) = \sum_y \frac{\pi_1(y)}{\pi_2(y)} \pi_2(y)P(y, x_0) > \sum_y \frac{\pi_1(x_0)}{\pi_2(x_0)} \pi_2(y)P(y, x_0)$$

Novamente, usando a estacionalidade de π_2 ,

$$\pi_1(x_0) > \frac{\pi_1(x_0)}{\pi_2(x_0)} \sum_y \pi_2(y)P(y, x_0) = \pi_1(x_0),$$

o que é um absurdo! Portanto, o quociente de entre as duas distribuições é sempre constante o que resulta que $\pi_1 = \pi_2$, como queríamos provar. \square

1.4 Cadeia Reversível

Aqui veremos um tipo especial de cadeia de Markov, as que possuem uma propriedade de simetria ao longo do tempo da seguinte forma, se estudarmos o comportamento da cadeia ao longo do tempo de t_0 até t_n é o mesmo que estudar do tempo contrário, t_n até t_0 . Este tipo de propriedade será ricamente usado no capítulo 5, por enquanto iremos ver alguns conceitos iniciais.

Definição 1.4.1 (Cadeia Reversível). *Seja $(X_n)_{n \geq 0}$ uma cadeia de Markov com matriz de transição P e uma distribuição π em Ω . Chamamos de **Equações de Balanço Detalhado**, abreviado por EBD, se satisfaz*

$$\pi(x)P(x, y) = \pi(y)P(y, x) \text{ para todo } x, y \in \Omega.$$

*Se uma cadeia de Markov satisfaz a EBD para alguma distribuição então vamos chama-la de **reversível**.*

Proposição 1.4.2. *Seja P a matriz de transição de uma cadeia de Markov no espaço de estados Ω . Qualquer distribuição π que satisfaça a EBD é estacionária para P .*

Demonstração. Vamos verificar que sabendo que vale EBD, ou seja, $\pi(x)P(x, y) = \pi(y)P(y, x)$ chegamos que $\pi = \pi P$. De fato, aplicando a somatório na condição de EBD,

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

Agora, como a somatória das probabilidades é 1, resulta que

$$\begin{aligned} \pi(x) \sum_y P(x, y) &= \sum_{y \in \Omega} \pi(y)P(y, x) \\ \pi(x) &= \sum_{y \in \Omega} \pi(y)P(y, x) \end{aligned}$$

ou seja, vale que $\pi = \pi P$, como gostaríamos de provar. \square

Repare que na nossa definição da EBD não fizemos a exigência de π ser estacionária, contudo o nosso primeiro resultado nos mostra que se satisfaz as EBD então π é estacionária, desde que P seja uma matriz de transição. Assim, muitas vezes, é mais simples verificar a EBD para garantir que uma distribuição em particular é estacionária. Além disso, usando a definição das equações de balanço detalhado teremos,

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

o que podemos reescrever da seguinte forma,

$$P_\pi\{X_0 = x_0, X_1 = x_1, \dots, X_n = x_n\} = P_\pi\{X_0 = x_n, X_1 = x_{n-1}, \dots, X_n = x_0\}.$$

Em outras palavras, se uma cadeia (X_t) é reversível e tem distribuição inicial estacionária, então as distribuições dos vetores aleatórios (X_0, X_1, \dots, X_n) e $(X_n, X_{n-1}, \dots, X_0)$ coincidem, o que explica a terminologia reversível.

Exemplo 1.4.3. Se P é uma matriz de transição da cadeia de Markov reversível com respeito a π , então P^2 também é uma matriz de transição de uma cadeia de Markov reversível com respeito a π .

Com efeito, primeiro vamos escrever a P^2 como um somatório e usando o fato de P ser reversível teremos

$$\pi(x)P^2(x, y) = \pi(x) \sum_z P(x, z)P(z, y) = \sum_z \pi(x)P(x, z)P(z, y).$$

Agora, usando a hipótese de ser reversível e comutando os termos do produto de forma conveniente

$$\pi(x)P^2(x, y) = \sum_z \pi(z)P(z, x)P(z, y) = \sum_z \pi(z)P(z, y)P(z, x).$$

Por fim, usando novamente a reversibilidade,

$$\sum_z \pi(z)P(z, y)P(z, x) = \sum_z \pi(y)P(y, z)P(z, x) = \pi(y)P^2(y, x).$$

Vamos ver agora um exemplo onde a cadeia de Markov possui uma distribuição estacionária mas não vale a equação de balanço detalhado.

Exemplo 1.4.4. Considere um passeio aleatório tendencioso no 4-ciclo, com a seguinte dinâmica: uma partícula se move no sentido horário com probabilidade p e com probabilidade $1-p$ no sentido anti-horário.

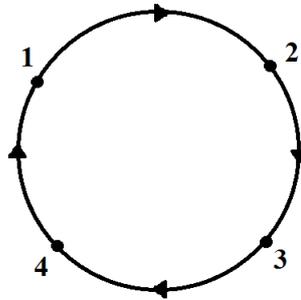


Figura 5 – 4-ciclo

Sem grandes dificuldades vemos que a matriz de transição é

$$P = \begin{pmatrix} 0 & p & 0 & 1-p \\ 1-p & 0 & p & 0 \\ 0 & 1-p & 0 & p \\ p & 0 & 1-p & 0 \end{pmatrix}.$$

Observe que esta cadeia é irredutível, logo temos a garantia da existência e unicidade da distribuição estacionária e resolvendo a equação matricial $\pi = \pi P$ resulta que $\pi = (1/4 \ 1/4 \ 1/4 \ 1/4)$. Porém,

$$\pi(2)P(2, 1) = \frac{1}{4}(1-p) \neq \frac{1}{4}p = \pi(1)P(1, 2)$$

sempre que $p \neq \frac{1}{2}$. Logo não vale a equação de balanço detalhado (quando $p \neq \frac{1}{2}$).

Como já vimos, nem todas as cadeias de Markov são reversíveis. No entanto podemos criar uma cadeia de Markov reversível a partir de qualquer cadeia de Markov exigindo apenas irreduzibilidade.

Definição 1.4.5. A *reversão do tempo* de uma cadeia de Markov irreduzível com matriz de transição P e distribuição estacionária π é uma cadeia com a matriz de transição dada por

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

A exigência da cadeia inicial ser irreduzível que vai nos garantir que a matriz $\hat{P}(x, y)$ seja estocástica. Para tanto temos que mostrar que \hat{P} satisfaz a definição 1.1.2. Sem nenhuma dificuldade, percebemos que para quaisquer $x, y \in \Omega$ vale

$$\hat{P}(x, y) \geq 0$$

e como vale que $\pi = \pi P$ teremos,

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

O resultado seguinte explica o porque da nomenclatura de reversão do tempo.

Proposição 1.4.6. Seja (X_t) uma cadeia de Markov irreduzível com matriz de transição P e distribuição estacionária π . Escreva (\hat{X}_t) para cadeia de reversão do tempo com matriz de transição \hat{P} . Então π é estacionária para \hat{P} e para quaisquer $x_0, x_1, \dots, x_t \in \Omega$ temos,

$$P_\pi\{X_0 = x_0, X_1 = x_1, \dots, X_t = x_t\} = P_\pi\{\hat{X}_0 = x_t, \hat{X}_1 = x_{t-1}, \dots, \hat{X}_t = x_0\}$$

ou escrevendo de forma equivalente,

$$\pi(x_0)P(x_0, x_1) \cdots P(x_{t-1}, x_t) = \pi(x_t)\hat{P}(x_t, x_{t-1}) \cdots \hat{P}(x_1, x_0)$$

Demonstração. Primeiro vamos mostrar que vale $\pi\hat{P} = \pi$, logo π será estacionária para \hat{P} . Com efeito,

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

Para a segunda parte, escrevemos

$$\pi(x_0)P(x_0, x_1) \cdots P(x_{t-1}, x_t)$$

e multiplicamos e dividimos por $\prod_{i=1}^t \pi(x_i)$, e reagrupando teremos

$$\frac{\pi(x_0)P(x_0, x_1)}{\pi(x_1)} \cdot \frac{\pi(x_1)P(x_1, x_2)}{\pi(x_2)} \cdots \frac{\pi(x_{t-1})P(x_{t-1}, x_t)}{\pi(x_t)} \cdot \pi(x_t)$$

agora usando a definição de \hat{P} segue o resultado. \square

Este resultado nos informa que qualquer trajetória em uma cadeia de Markov irredutível P que começa com a distribuição estacionária π tem a mesma probabilidade de ocorrer que a trajetória reversa em uma cadeia de Markov \hat{P} , iniciada com a mesma distribuição estacionária π .

Por fim, o nosso próximo resultado nos mostra uma equivalência de cadeia reversível com tempo de reversão de Markov.

Proposição 1.4.7. *Uma cadeia de Markov com matriz de transição P é reversível se, e somente se, $\hat{P} = P$.*

Demonstração. Primeiro suponha que a cadeia seja reversível. Usando que vale $\pi(x)P(x, y) = \pi(y)P(y, x)$ teremos para quaisquer $x, y \in \Omega$,

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

Por outro lado, se vale que $P = \hat{P}$ basta usar a definição de \hat{P} e verificar que vale a equação de balanço detalhado, logo a cadeia é reversível. \square

Corolário 1.4.8. *Toda cadeia de Markov irredutível com dois estados é reversível.*

Demonstração. Escrevendo $\Omega = \{x, y\}$ basta notar que vale $\hat{P}(x, x) = P(x, x)$ (isso vale para qualquer cadeia) e como a soma da linha é 1 então $\hat{P}(x, y) = P(x, y)$, portanto $\hat{P} = P$. \square

2 Convergência de Cadeia de Markov

Neste capítulo tem como resultado principal o Teorema da Convergência 2.3.1, que nos mostra quais condições são necessárias para a convergência de uma cadeia de Markov para a distribuição estacionária. Mas para estudar convergência é necessário primeiro definir uma métrica, desta forma esse é o primeiro assunto a ser explorado nesse capítulo, além de estudarmos propriedades dessa métrica.

2.1 Distância de Variação Total

Vamos começar estudando uma forma de "medir" a distância entre duas distribuições que é explicitamente probabilística: a distância entre as distribuições μ e ν é a diferença máxima entre as probabilidades atribuídas no mesmo evento, como veremos abaixo.

Definição 2.1.1. A *distância de variação total* entre duas distribuições de probabilidade μ e $\nu \in \Omega$ é definido por

$$\|\mu - \nu\|_{TV} = \max_{A \subset \Omega} |\mu(A) - \nu(A)|. \quad (2.1)$$

Exemplo 2.1.2. Vamos voltar no Exemplo 1.1.5. Sua matriz de transição é

$$\begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

e sua distribuição estacionária é $\pi = \left(\frac{q}{p+q}, \frac{p}{p+q}\right)$. Vamos supor que começamos no estado 1, ou seja, $\mu_0 = (1, 0)$

Como só existem apenas dois estados, existem apenas quatro possibilidade de eventos $A \subset \Omega$. Por isso é fácil ver que

$$\|\mu_t - \pi\|_{TV} = P^t(1, 1) - \pi(1) = \pi(2) - P^t(1, 2).$$

Usar a definição de distância de variação total (2.1) nem sempre é a maneira mais fácil de calcular a distância de duas distribuições, por isso vamos mostrar formas alternativas de calcular a distância de variação total.

Proposição 2.1.3. Sejam μ e ν distribuições em Ω . Então

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

Demonstração. Considere o conjunto $B = \{x : \mu(x) \geq \nu(x)\}$ e seja $A \subset \Omega$ para qualquer evento. Lembrando que podemos escrever $A = (A \cap B) \dot{\cup} (A \cap B^c)$ ¹ conseguimos a seguinte

¹ A notação $X \dot{\cup} Y$ quer dizer união disjunta, ou seja, $X \cup Y$ com $X \cap Y = \emptyset$.

relação

$$\begin{aligned}
\mu(A) - \nu(A) &= [\mu(A \cap B) + \mu(A \cap B^c)] - [\nu(A \cap B) + \nu(A \cap B^c)] & (2.2) \\
&= [\mu(A \cap B) - \nu(A \cap B)] + \underbrace{[\mu(A \cap B^c) - \nu(A \cap B^c)]}_{\leq 0} \\
&\leq \mu(A \cap B) - \nu(A \cap B).
\end{aligned}$$

De forma análoga podemos escrever $B = (A \cap B) \dot{\cup} (A^c \cap B)$ e temos que

$$\begin{aligned}
\mu(B) - \nu(B) &= [\mu(A \cap B) - \nu(A \cap B)] + \underbrace{[\mu(A^c \cap B) - \nu(A^c \cap B)]}_{\geq 0} \\
&\geq \mu(A \cap B) - \nu(A \cap B)
\end{aligned}$$

logo,

$$\mu(A) - \nu(A) \leq \mu(A \cap B) - \nu(A \cap B) \leq \mu(B) - \nu(B). \quad (2.3)$$

Como A é arbitrário, podemos escrever que $\mu(A^c) - \nu(A^c) \leq \mu(B) - \nu(B)$ e disso segue que $\nu(A) - \mu(A) \leq \mu(B) - \nu(B)$, portanto

$$|\mu(A) - \nu(A)| \leq \mu(B) - \nu(B).$$

Por outro lado temos

$$\mu(B) - \nu(B) = (1 - \mu(B^c)) - (1 - \nu(B^c)) = \nu(B^c) - \mu(B^c).$$

Mais ainda, somando de forma conveniente resulta que

$$2|\mu(A) - \nu(A)| \leq (\mu(B) - \nu(B)) + (\nu(B^c) - \mu(B^c)).$$

Agora tomando $A = B$ (ou B^c) e o máximo resulta

$$\|\mu - \nu\|_{TV} = \frac{1}{2} [\mu(B) - \nu(B) + \nu(B^c) - \mu(B^c)] = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|.$$

□

Corolário 2.1.4. *Sejam μ e ν duas distribuições em Ω , logo*

$$\|\mu - \nu\|_{TV} = \sum_{\substack{x \in \Omega \\ \mu(x) \geq \nu(x)}} [\mu(x) - \nu(x)].$$

Demonstração. Como vale que

$$\sum_{\mu(x) \geq \nu(x)} [\mu(x) - \nu(x)] = \sum_{\nu(x) \geq \mu(x)} [\nu(x) - \mu(x)]$$

e usando o teorema anterior teremos

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

□

Observe que o sinal de \geq no subíndice do somatório pode ser substituído por $>$ pois se existe $x_0 \in \Omega$ tal que $\mu(x_0) = \nu(x_0)$, então $\mu(x_0) - \nu(x_0) = 0$ e portanto não interfere no valor do somatório.

Um outro fato importante que podemos observar facilmente usando a Proposição 2.1.3 é que a distância de variação total é de fato uma métrica definida em Ω , pois satisfaz:

- i) *Positividade*, ou seja, $\|\mu - \nu\|_{TV} \geq 0$ para toda distribuição μ, ν . Segue diretamente da definição.
- ii) *Simétrica*, ou seja, $\|\mu - \nu\|_{TV} = \|\nu - \mu\|_{TV}$. Novamente, segue pela definição ou se preferir pela Proposição 2.1.3
- iii) *Desigualdade triangular*. Segue da Proposição 2.1.3 e a desigualdade triangular para números reais.

Proposição 2.1.5. *Sejam μ e ν distribuições em Ω . Considere o conjunto*

$$S = \left\{ \sum_{x \in \Omega} f(x)\mu(x) - \sum_{x \in \Omega} f(x)\nu(x) \mid \max_{x \in \Omega} \|f(x)\| \leq 1 \right\}$$

então

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sup S$$

Demonstração. Suponha que f satisfaz $\max_{x \in \Omega} \|f(x)\| \leq 1$, assim,

$$\frac{1}{2} \left| \sum_{x \in \Omega} f(x)\mu(x) - \sum_{x \in \Omega} f(x)\nu(x) \right| \leq \frac{1}{2} \sum_{x \in \Omega} |f(x)| |\mu(x) - \nu(x)| \leq \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|$$

Passando o supremo sobre as funções temos que $\frac{1}{2} \sup S \leq \|\mu - \nu\|_{TV}$. Por outro lado, vamos definir a função

$$f^*(x) = \begin{cases} 1 & \text{se } \mu(x) \geq \nu(x); \\ -1 & \text{se } \mu(x) < \nu(x). \end{cases}$$

Daí,

$$\begin{aligned}
\frac{1}{2} \left[\sum_{x \in \Omega} f^*(x) \mu(x) - \sum_{x \in \Omega} f^*(x) \nu(x) \right] &= \frac{1}{2} \sum_{x \in \Omega} f^*(x) [\mu(x) - \nu(x)] \\
&= \frac{1}{2} \left[\sum_{\mu(x) \leq \nu(x)} [\mu(x) - \nu(x)] + \sum_{\nu(x) > \mu(x)} [\nu(x) - \mu(x)] \right] \\
&= \frac{1}{2} [\|\mu - \nu\|_{TV} + \|\nu - \mu\|_{TV}] \\
&= \|\mu - \nu\|_{TV}.
\end{aligned}$$

Ou seja, achamos uma cota superior para o supremo e uma função onde atinge esta cota, portanto vale a igualdade. \square

2.2 Distância Padrão

Estamos interessados em limitar a distância entre $P^t(x_0, \cdot)$ e π para qualquer $x_0 \in \Omega$. Este é um dos nossos principais objetivos nesse capítulo.

Definição 2.2.1. *Sejam $x, y \in \Omega$ definimos,*

$$d(t) := \max_{x \in \Omega} \|P^t(x, \cdot) - \pi(\cdot)\|_{TV} \quad (2.4)$$

e

$$\bar{d}(t) := \max_{x, y \in \Omega} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV}. \quad (2.5)$$

Nos próximos resultados iremos ver algumas propriedades desses dois objetos definidos acima e também algumas desigualdades que os relacionam.

Proposição 2.2.2. *Seja P uma matriz de transição com distribuição estacionária π , então*

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

Demonstração. Basta usar o Corolário 2.1.4 e propriedade de máximo como segue

$$d(t) = \sum_{\substack{y \in \Omega \\ P^t(x, y) < \pi(y)}} [\pi(y) - P^t(x, y)] = \sum_{\substack{y \in \Omega \\ P^t(x, y) < \pi(y)}} \pi(y) \left[1 - \frac{P^t(x, y)}{\pi(y)} \right] \leq \max_y \left[1 - \frac{P^t(x, y)}{\pi(y)} \right].$$

Portanto segue o resultado. \square

Proposição 2.2.3. *Considere $d(t)$ e $\bar{d}(t)$ como definidos previamente então,*

$$d(t) \leq \bar{d}(t) \leq 2d(t). \quad (2.6)$$

Demonstração. Usando a desigualdade triangular e as definições de $d(t)$ e $\bar{d}(t)$ temos

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

Para mostrar que $d(t) \leq \bar{d}(t)$, primeiro note que, π sendo distribuição estacionária, temos que $\pi(A) = \sum_{y \in \Omega} \pi(y)P^t(y, A)$ para qualquer conjunto A . Usando este fato, escrevemos

$$d(t) = \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{TV} = \max_{x \in \Omega} \max_{A \subset \Omega} |P^t(x, A) - \pi(A)| = \max_{x \in \Omega} \max_{A \subset \Omega} \left| P^t(x, A) - \sum_{y \in \Omega} \pi(y)P^t(y, A) \right|.$$

Usando de forma conveniente a propriedade de módulo e que $1 = \sum_y \pi(y)$, logo

$$d(t) = \max_{x \in \Omega} \max_{A \subset \Omega} \left| \sum_{y \in \Omega} \pi(y)[P^t(x, A) - P^t(y, A)] \right| \leq \max_{x \in \Omega} \max_{A \subset \Omega} \sum_{y \in \Omega} \pi(y) |P^t(x, A) - P^t(y, A)|.$$

Podemos usar o fato que o máximo de uma soma não é maior que a soma dos máximos e o fato que a média ponderada de um conjunto nunca é maior que o seu elemento máximo, para obtermos

$$\begin{aligned}d(t) &\leq \max_{x \in \Omega} \max_{A \subset \Omega} \sum_{y \in \Omega} \pi(y) |P^t(x, A) - P^t(y, A)| \\ &\leq \max_{x \in \Omega} \sum_{y \in \Omega} \pi(y) \max_{A \subset \Omega} |P^t(x, A) - P^t(y, A)| \\ &\leq \max_{x \in \Omega} \sum_{y \in \Omega} \pi(y) \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \\ &\leq \max_{x,y \in \Omega} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \\ &= \bar{d}(t).\end{aligned}$$

□

Proposição 2.2.4. *Para todo $s, t > 0$ vale $\bar{d}(s+t) \leq \bar{d}(s)\bar{d}(t)$.*

Esta demonstração faz uso da teoria de acoplamento e o mesmo não será abordado neste texto, porém o leitor interessado pode encontrar a demonstração na seção 4 do capítulo 4 da referência [3] ou na seção 4 do capítulo 2 de [1].

Infelizmente, $d(t)$ não possui esta propriedade, abaixo daremos um contra-exemplo, porém $d(t)$ e $\bar{d}(t)$ são funções não-crescente, como veremos na proposição a seguir.

Exemplo 2.2.5. O Exemplo 1.1.5 é um caso em que não vale a desigualdade $d(t+s) \leq d(t)d(s)$. Neste exemplo temos a seguinte matriz de transição

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

com distribuição estacionária $\pi = \left(\frac{q}{p+q}, \frac{p}{p+q}\right)$.

Primeiramente, iremos escrever P da seguinte forma, $P = QDQ^{-1}$ em que D é uma matriz diagonal e os elementos da diagonal principal são os autovalores de P , que por sua vez são: $\lambda_1 = 1$ e $\lambda_2 = 1 - p - q$. Considere que $0 < p < q < 1$. É necessário que, tanto p quanto q não assumam o valor 0 pois senão a matriz ficaria redutível, nem o valor 1 pois senão a matriz ficaria periódica.

Com uma conta simples, resulta que o autoespaço de λ_1 e λ_2 são, respectivamente, $[(1, 1)]$ e $[(-p/q, 1)]$, daí teremos que

$$Q = \begin{pmatrix} 1 & -p/q \\ 1 & 1 \end{pmatrix} \quad e \quad Q^{-1} = \begin{pmatrix} \frac{q}{p+q} & \frac{p}{p+q} \\ \frac{-q}{p+q} & \frac{q}{p+q} \end{pmatrix}.$$

Assim

$$\begin{aligned} P &= QDQ^{-1} \\ &= \begin{pmatrix} 1 & -p/q \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1-p-q \end{pmatrix} \begin{pmatrix} \frac{q}{p+q} & \frac{p}{p+q} \\ \frac{-q}{p+q} & \frac{q}{p+q} \end{pmatrix}. \end{aligned}$$

Logo,

$$\begin{aligned} P^t &= QD^tQ^{-1} \\ &= \begin{pmatrix} 1 & -p/q \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & (1-p-q)^t \end{pmatrix} \begin{pmatrix} \frac{q}{p+q} & \frac{p}{p+q} \\ \frac{-q}{p+q} & \frac{q}{p+q} \end{pmatrix} \\ &= \begin{pmatrix} \frac{q + p(1-p-q)^t}{p+q} & \frac{p - p(1-p-q)^t}{p+q} \\ \frac{q - q(1-p-q)^t}{p+q} & \frac{p + q(1-p-q)^t}{p+q} \end{pmatrix}. \end{aligned}$$

Note que, $-1 < 1 - p - q < 1$ e recorde que $\Omega = \{1, 2\}$, logo

$$\begin{aligned} d(t) &= \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{TV} \\ &= \max_{x \in \Omega} \left[\frac{1}{2} \sum_{y \in \Omega} |P^t(x, y) - \pi(y)| \right] \\ &= \max_{x \in \Omega} \left[\frac{1}{2} (|P^t(x, 1) - \pi(1)| + |P^t(x, 2) - \pi(2)|) \right] \\ &= \max_{x \in \Omega} \left[\frac{1}{2} \left(\left| P^t(x, 1) - \frac{q}{p+q} \right| + \left| P^t(x, 2) - \frac{p}{p+q} \right| \right) \right] \\ &= \frac{1}{2} \left(\frac{q |1-p-q|^t}{p+q} + \frac{q |1-p-q|^t}{p+q} \right) \\ &= \frac{q |1-p-q|^t}{p+q}. \end{aligned}$$

Assim,

$$\begin{aligned} d(s)d(t) &= \left(\frac{q |1 - p - q|^s}{p + q} \right) \left(\frac{q |1 - p - q|^t}{p + q} \right) \\ &= \frac{q}{p + q} \left(\frac{q |1 - p - q|^{t+s}}{p + q} \right) \\ &= \frac{q}{p + q} d(t + s). \end{aligned}$$

Por fim, basta responder a seguinte pergunta: para quais valores de p e q teremos $d(t + s) \leq d(t)d(s)$? Ora, supondo $d(t + s) \leq d(t)d(s)$ implica dizer que $\frac{q}{p+q} \geq 1$, ou seja, $p \leq 0$ que é uma contradição pela nossa consideração de $0 < p < q < 1$.

Proposição 2.2.6. *As funções $d(t)$ e $\bar{d}(t)$ são não-crescentes.*

Demonstração. Primeiro vamos mostrar para a função $\bar{d}(t)$, para tanto vamos usar a Proposição 2.2.4 e o fato que $0 \leq \bar{d}(t) \leq 1$ como segue

$$\bar{d}(t + s) \leq \bar{d}(t)\bar{d}(s) \leq \bar{d}(t).$$

Para a $d(t)$ vamos usar a Proposição 2.1.3, escrevendo $1 = \sum_z P(x, z)$ e usando de forma conveniente o máximo, temos

$$\begin{aligned} d(t + s) = \max_{x \in \Omega} \|P^{t+s}(x, \cdot) - \pi\|_{TV} &= \max_{x \in \Omega} \frac{1}{2} \sum_y |P^{t+s}(x, y) - \pi(y)| \\ &= \max_{x \in \Omega} \frac{1}{2} \sum_y \left| \sum_z P^t(x, z)P^s(z, y) - \sum_z P^t(x, z)\pi(y) \right| \\ &\leq \max_{x \in \Omega} \frac{1}{2} \sum_y \sum_z P^t(x, z) |P^s(z, y) - \pi(y)| \\ &\leq \max_{x \in \Omega} \sum_z P^t(x, z) \max_{\tilde{z}} \left(\frac{1}{2} \sum_y |P^s(\tilde{z}, y) - \pi(y)| \right) \\ &= d(s). \end{aligned}$$

□

2.3 O Teorema da Convergência e o Tempo de Mistura

Agora já conhecemos e estamos bem familiarizados com todas as hipóteses e ferramentas (irredutibilidade, aperiodicidade e a distância de variação total) que são necessárias para a demonstração do Teorema da Convergência, que garante a convergência de todas as linhas de $P^t(x, \cdot)$ para a distribuição estacionária, ou seja, garante sob algumas condições que qualquer distribuição $P^t(x, \cdot)$, com $x \in \Omega$ se aproxima da distribuição estacionária π . A demonstração é feita basicamente com operações com matrizes e vetores e indução matemática.

Ao final dessa seção tratamos também o tempo de mistura que é um parâmetro que mede o quão próximo a distribuição $P^t(x, \cdot)$, com $x \in \Omega$ está próxima da distribuição estacionária π .

Teorema 2.3.1 (Teorema da Convergência). *Suponha que P seja irredutível e aperiódica, com distribuição estacionária π . Então existe uma constante $\alpha \in (0, 1)$ e $C > 0$ tal que,*

$$d(t) = \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{TV} \leq C\alpha^t. \quad (2.7)$$

Demonstração. Considere P sendo irredutível e aperiódica, pela Proposição 1.2.6 existe r tal que P^r tem entradas estritamente positivas. Além disso, como P é irredutível, pela Proposição 1.3.4, $\pi(y) > 0$ para todo $y \in \Omega$, assim defina Π como sendo a matriz com $|\Omega|$ linhas, sendo cada linha o vetor de π , logo Π é uma matriz estocástica de ordem $|\Omega| \times |\Omega|$. Para algum $\delta > 0$ suficientemente pequeno, temos

$$P^r(x, y) \geq \delta\pi(y) > 0$$

para todos $x, y \in \Omega$. Tome $\theta = 1 - \delta > 0$. A equação

$$P^r = (1 - \theta)\Pi + \theta Q \quad (2.8)$$

define uma matriz estocástica Q . Com efeito, pela definição de matriz estocástica basta mostrar que todos os elementos da matriz Q são não-negativos e a soma dos elementos de cada linha é igual a 1. Para tanto, usando (2.8) e $P^r(x, y) \geq \delta\pi(y)$ logo,

$$Q(x, y) = \frac{P^r(x, y) - (1 - \theta)\Pi(x, y)}{\theta} = \frac{P^r(x, y) - \delta\pi(y)}{\theta} \geq 0,$$

e para todo $x \in \Omega$ temos

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

Iremos precisar das duas afirmações seguintes, a primeira é: se M é uma matriz estocástica, então $M\Pi = \Pi$. De fato, sejam $x, y \in \Omega$, daí

$$(M\Pi)(x, y) = \sum_{k \in \Omega} M(x, k)\Pi(k, y) = \pi(y) \sum_{k \in \Omega} M(x, k) = \pi(y) = \Pi(x, y).$$

A segunda afirmação é: se $\pi M = \pi$, então $\Pi \cdot M = \Pi$. De fato, como cada linha de Π é formada pelo vetor π a demonstração é imediata.

Agora, vamos usar indução sobre k para mostrar que

$$P^{rk} = (1 - \theta^k)\Pi + \theta^k Q^k. \quad (2.9)$$

Se $k = 1$ o resultado é imediato por (2.8). Assumindo que (2.9) é verdadeiro para $k = n$, vamos demonstrar que é válido para $k = n + 1$. Com efeito,

$$P^{r(n+1)} = P^{rn} P^r = [(1 - \theta^n)\Pi + \theta^n Q^n] P^r. \quad (2.10)$$

Expandindo P^r no lado direito da expressão anterior e usando (2.8)

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

Usando que $\Pi P^r = \Pi$ e $Q^n \Pi = \Pi$ temos que,

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

Isto prova a validade de (2.9). Multiplicando ambos os lados de (2.9) por P^j e reagrupando os termos temos

$$P^{rk+j} - \Pi = \theta^k (Q^k P^j - \Pi). \quad (2.11)$$

Para completar a prova, vamos usar a Proposição 2.1.3, ou seja, somando os valores em módulo dos elementos da linha x_0 (arbitrário) em ambos os lados de (2.11) e dividindo por dois. Portanto, para qualquer $x_0 \in \Omega$ temos

$$\begin{aligned} \frac{1}{2} \sum_{x \in \Omega} |P^{rk+j}(x_0, x) - \Pi(x_0, x)| &= \frac{\theta^k}{2} \sum_{x \in \Omega} |(Q^k P^j)(x_0, x) - \Pi(x_0, x)| \\ &= \theta^k \|(Q^k P^j)(x_0, \cdot) - \pi\|_{TV} \\ &\leq \theta^k \end{aligned} \quad (2.12)$$

em que a desigualdade acima ocorre do fato que a métrica de variação total entre duas distribuições varia no intervalo $[0,1]$. Em particular, dado $t \in \mathbb{N}$, existem $j, k \in \mathbb{N}$ com $0 \leq j < r$ tal que $t = rk + j$ e obtemos por (2.12),

$$\|P^t(x_0, \cdot) - \pi\|_{TV} \leq \theta^{\frac{t-j}{k}} = (\sqrt[k]{\theta})^t \theta^{-\frac{j}{k}} = \alpha^t \left(\frac{1}{\theta}\right)^{\frac{j}{k}} \leq \alpha^t \left(\frac{1}{\theta}\right) = C\alpha^t,$$

onde tomamos $\alpha = \sqrt[k]{\theta}$ e $C = \frac{1}{\theta}$. □

Por causa do Teorema 2.3.1, a distribuição π também é chamada de **distribuição de equilíbrio**.

É útil introduzir um parâmetro que mede o tempo necessário para uma cadeia de Markov estar próxima da sua distribuição estacionária com um erro escolhido, o que motiva a próxima definição.

Definição 2.3.2. *Seja $\varepsilon \in [0, 1]$, o tempo de mistura de uma cadeia de Markov é definido por*

$$t_{mix}(\varepsilon) := \min\{t \mid d(t) \leq \varepsilon\}$$

e

$$t_{mix} := t_{mix}(1/4).$$

Vamos concluir essa seção com algumas observações sobre o tempo de mistura.

- (i) Não é difícil de perceber que a exigência de $0 \leq \varepsilon \leq 1$ é necessária visto que a distância de variação total varia também de 0 a 1.
- (ii) [Uma relação entre tempos de mistura com taxas distintas] A partir das Proposições 2.2.3 e 2.2.4, segue que quando c e t são números inteiros não negativos quaisquer, implica que,

$$d(ct) \leq \bar{d}(ct) \leq \bar{d}(t)^c. \quad (2.13)$$

Agora, usando (2.13) e a Proposição 2.2.3 nos mostra que, para qualquer l inteiro não negativo,

$$d(lt_{mix}(\varepsilon)) \leq \bar{d}(lt_{mix}(\varepsilon)) \leq \bar{d}(t_{mix}(\varepsilon))^l \leq [2d(t_{mix}(\varepsilon))]^l \leq (2\varepsilon)^l. \quad (2.14)$$

Repare que, em (2.14), o interessante é termos $0 < \varepsilon < 1/2$, pois já sabemos que a distância de variação total é no máximo 1. Agora, dado $0 < \delta < 1$ tal que vale

$$d(lt_{mix}(\varepsilon)) \leq (2\varepsilon)^l \leq \delta.$$

Usando a definição do tempo de mistura resulta que $t_{mix}(\delta) \leq lt_{mix}(\varepsilon)$, por outro lado

$$(2\varepsilon)^l \leq \delta \text{ resulta em } l \log_2 2\varepsilon \leq \log_2 \delta \text{ portanto } l \geq \frac{\log_2 \delta}{\log_2 2\varepsilon} = \log_{2\varepsilon} \delta.$$

Resumindo, temos que

$$t_{mix}(\delta) \leq lt_{mix}(\varepsilon)$$

sempre que $l \geq \log_{2\varepsilon} \delta$, $0 < \varepsilon < 1/2$ e $0 < \varepsilon < 1$. Mas como estamos querendo a melhor cota possível para l vamos tomar a igualdade, logo vale que

$$t_{mix}(\delta) \leq \log_{2\varepsilon} \delta \cdot t_{mix}(\varepsilon).$$

Um caso particular que vale a pena ressaltar é quando relacionamos o $t_{mix}(\varepsilon)$ com o t_{mix} , desta forma temos

$$t_{mix}(\varepsilon) \leq \log_2 \varepsilon^{-1} t_{mix}.$$

Está última desigualdade pode servir para justificar a escolha de $\varepsilon = 1/4$ na definição 2.3.2.

(iii) [$t_{mix}(\varepsilon)$ e o Teorema da Convergência 2.3.1] O Teorema da Convergência 2.3.1 permite-nos obter uma cota para o tempo de mistura $t_{mix}(\varepsilon)$. Recorde a anotação utilizada no Teorema da Convergência 2.3.1 e observe, por (2.7),

$$t_{mix}(\varepsilon) \leq \frac{1}{\log_2 \alpha} \log_2 \left(\frac{\varepsilon}{C} \right)$$

Usando que $\alpha = \sqrt[r]{\theta}$ e $C = \frac{1}{\theta}$ obtemos que

$$t_{mix}(\varepsilon) \leq r \left[\frac{\log_2 \varepsilon}{\log_2 \theta} + 1 \right].$$

Em particular, uma estimativa é dada considerando

$$r = \min\{t \in \mathbb{N} \mid P^t(x, y) > 0 \text{ para todo } x, y \in \Omega\}$$

e $\theta = 1 - \delta$ onde

$$\delta = \min_{x, y \in \Omega} \frac{P^r(x, y)}{\pi(y)}.$$

3 Limitando o tempo de mistura

Conseguimos sempre descrever uma cadeia de Markov finita através de uma matriz, que chamamos de matriz de transição, então é natural fazer um estudo sobre estas matrizes. Desta forma, veremos que a partir dos autovalores da matriz de transição conseguimos uma ótima estimativa do tempo de mistura, que é extraordinário de se pensar, pois apenas calculando os autovalores da matriz de transição conseguimos saber de antemão quantas vezes precisamos iterar a cadeia para que esta esteja próxima da distribuição estacionária.

Além disso, vamos definir um novo produto interno que permite que a base de autovetores seja ortonormal, como veremos no Lema 3.1.2.

Ao final iremos estimar o tempo de mistura usando a geometria do espaço de estado que será muito útil no capítulo 4.

3.1 A Representação Espectral

Começamos com alguns fatos elementares sobre autovalores de matrizes de transição.

Proposição 3.1.1. *Seja P a matriz de transição de uma cadeia de Markov finita*

- (i) *Se λ é um autovalor de P , então $|\lambda| \leq 1$;*
- (ii) *Se P é irredutível, o autoespaço do autovetor correspondente ao autovalor 1 é unidimensional e é gerado pelo vetor coluna $\mathbf{1} := (1, 1, \dots, 1)^t$;*
- (iii) *Se P é irredutível e aperiódica, então -1 não é um autovalor de P .*

Demonstração. Primeiro afirmamos que $\|Pf\|_\infty \leq \|f\|_\infty$, onde $\|f\|_\infty = \max_{x \in \Omega} |f(x)|$. Com efeito, temos

$$\|Pf\|_\infty = \max_{x \in \Omega} \left| \sum_{y \in \Omega} P(x, y) f(y) \right| \leq \max_{x \in \Omega} \sum_{y \in \Omega} P(x, y) |f(y)| \leq \max_{x \in \Omega} |f(y)| = \|f\|_\infty.$$

Agora suponha que f seja um autovetor da matriz P com autovalor λ , então vale que $Pf = \lambda f$ e, usando a afirmação acima conseguimos que

$$|\lambda| \|f\|_\infty = \|\lambda f\|_\infty = \|Pf\|_\infty \leq \|f\|_\infty.$$

Portanto $|\lambda| \leq 1$, logo demonstramos o item (i). Para o item (ii), precisamos do seguinte fato: 1 é um autovalor de P se, e somente se, 1 também é autovalor de P_L (a forma *lazy* de P). De fato, suponha que P admita o autovalor 1. Daí temos, $Pf = f$ para alguma $f \neq 0$ e como por definição $P_L = \frac{1}{2}P + \frac{1}{2}I$ segue que

$$P_L f = \frac{1}{2}Pf + \frac{1}{2}If = \frac{1}{2}f + \frac{1}{2}f = f.$$

Logo 1 é um autovalor de P_L . Reciprocamente, suponha agora que $P_L f = f$, logo

$$\begin{aligned} P_L f &= \frac{1}{2} P f + \frac{1}{2} I f = f \\ \frac{1}{2} P f &= \frac{1}{2} f \\ P f &= f. \end{aligned}$$

Seja 1 o autovalor de P associado ao autovetor f , logo $P f = f$. Vamos mostrar que $f(x) = 1$ para todo $x \in \Omega$, note que vale $P^t f(x) = f(x)$, pois

$$P^t f(x) = P^{t-1}(P f(x)) = P^{t-1} f(x) = \dots = f(x).$$

Como P é irredutível então existe a distribuição estacionária π . Agora, usando a equivalência $P^t f = f \Leftrightarrow P_L^t f = f$ temos

$$\begin{aligned} |f(x) - \mathbf{E}_\pi f| &= |P^t f(x) - \mathbf{E}_\pi f| \\ &= |P_L^t f(x) - \mathbf{E}_\pi f| \\ &= \left| \sum_{y \in \Omega} P_L^t(x, y) f(y) - \sum_{y \in \Omega} f(y) \pi(y) \right| \\ &\leq \max_{z \in \Omega} |f(z)| \sum_{y \in \Omega} |P_L^t(x, y) - \pi(y)| \\ &\leq 2 \max_{z \in \Omega} |f(z)| \|P_L^t(x, \cdot) - \pi\|_{TV}. \end{aligned}$$

Como P_L é irredutível (pois P é irredutível) e aperiódica, então podemos usar o Teorema da Convergência 2.3.1

$$|f(x) - \mathbf{E}_\pi f| \leq 2 \max_{z \in \Omega} |f(z)| \|P_L^t(x, \cdot) - \pi\|_{TV} \rightarrow 0$$

Portanto $f(x) = \mathbf{E}_\pi f$, ou seja, $f(x)$ é constante para todo $x \in \Omega$, com isso termina a demonstração do (ii). Para o item (iii), suponha por contradição que P seja irredutível e aperiódico com autovalor -1 . Note que conseguimos as seguintes relações usando o Teorema da Convergência 2.3.1

$$\begin{aligned} f(x) &= P^{2t} f(x) \xrightarrow{t \rightarrow \infty} \mathbf{E}_\pi f; \\ -f(x) &= P^{2t+1} f(x) \xrightarrow{t \rightarrow \infty} \mathbf{E}_\pi f. \end{aligned}$$

Logo para t suficientemente grande, temos $f(x) = \mathbf{E}_\pi f = -f(x)$, portanto, $f(x) = 0$. Contradição pois f é um autovetor, ou seja, $f \neq 0$. \square

Denote por $\langle \cdot, \cdot \rangle$ o produto interno usual em \mathbb{R}^Ω , dado por $\langle f, g \rangle = \sum_{x \in \Omega} f(x)g(x)$. Mas é de nosso interesse adotar um produto interno que deixa a base de autovetores ortonormais como será visto no próximo resultado, assim é preciso de outro produto interno, denotado por $\langle \cdot, \cdot \rangle_\pi$ e definido por

$$\langle f, g \rangle_\pi = \sum_{x \in \Omega} f(x)g(x)\pi(x).$$

Proposição 3.1.2. *Seja P reversível com respeito à distribuição π .*

(i) *O espaço com produto interno $(\mathbb{R}^\Omega, \langle \cdot, \cdot \rangle_\pi)$ tem uma base ortonormal de autovetores $\{f_j\}_{j=1}^{|\Omega|}$ corresponde aos autovalores reais $\{\lambda_j\}$;*

(ii) *A matriz pode ser decomposta como*

$$\frac{P^t(x, y)}{\pi(y)} = \sum_{j=1}^{|\Omega|} f_j(x) f_j(y) \lambda_j^t;$$

(iii) *O autovetor f_1 correspondente ao autovalor 1 pode ser tomado como sendo o vetor constante $\mathbf{1}$, e nesse caso*

$$\frac{P^t(x, y)}{\pi(y)} = 1 + \sum_{j=2}^{|\Omega|} f_j(x) f_j(y) \lambda_j^t. \quad (3.1)$$

Demonstração. Defina $A(x, y) = \pi^{1/2}(x) \pi^{-1/2}(y) P(x, y)$, onde $\pi^{1/2}$ é o vetor cujas entradas são as raízes das entradas de π e observe que P é reversível se, e somente se, A é simétrico, pois

$$A(x, y) = A(y, x) \Leftrightarrow \left(\frac{\pi(x)}{\pi(y)} \right)^{\frac{1}{2}} P(x, y) = \left(\frac{\pi(y)}{\pi(x)} \right)^{\frac{1}{2}} P(y, x) \Leftrightarrow \pi(x) P(x, y) = \pi(y) P(y, x).$$

Logo, pelo Teorema Espectral segue que A possui uma base de autovetores ortogonais $\{\varphi_j\}_{j=1}^{|\Omega|}$ tal que φ_j é um autovetor associado ao autovalor λ_j . Note que $\pi^{1/2}(x)$ é um autovetor associado ao autovalor 1, pois

$$A\pi^{1/2}(x) = \sum_{y \in \Omega} A(x, y) \pi^{1/2}(y) = \sum_{y \in \Omega} \left(\frac{\pi(x)}{\pi(y)} \right)^{\frac{1}{2}} P(x, y) \pi^{1/2}(y) = \pi^{1/2}(y)$$

assim definimos $\varphi_1 := \pi^{1/2}(x)$ e $\lambda_1 = 1$. Denote D_π como a matriz diagonal dada por $D_\pi(x, x) = \pi(x)$, assim podemos escrever $A = D_\pi^{1/2} P D_\pi^{-1/2}$. Se definimos $f_j := D_\pi^{-1/2} \varphi_j$ então f_j é um autovetor de P com autovalor λ_j , pois

$$P f_j = \left(D_\pi^{-1/2} A D_\pi^{1/2} \right) D_\pi^{-1/2} \varphi_j = D_\pi^{-1/2} \lambda_j \varphi_j = \lambda_j f_j.$$

Embora os autovetores $\{f_j\}$ não sejam ortogonais usando o produto interno usual, os mesmos são ortogonais usando $\langle \cdot, \cdot \rangle_\pi$, pois

$$\langle \varphi_i, \varphi_j \rangle = \langle D_\pi^{1/2} f_i, D_\pi^{1/2} f_j \rangle = \langle f_i, f_j \rangle_\pi.$$

Isso prova o item (i). Considere δ_y como a função

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

Considerando $(\mathbb{R}^\Omega, \langle \cdot, \cdot \rangle_\pi)$ com a base ortonormal de autovetores $\{f_j\}_{j=1}^{|\Omega|}$, a função δ_y pode ser escrita como

$$\delta_y = \sum_{j=1}^{|\Omega|} \langle \delta_y, f_j \rangle_\pi f_j = \sum_{j=1}^{|\Omega|} f_j(y) \pi(y) f_j.$$

Como $P^t f_j = \lambda_j^t f_j$ e $P^t(x, y) = (P^t \delta_y)(x)$,

$$P^t(x, y) = \sum_{j=1}^{|\Omega|} f_j(y) \pi(y) \lambda_j^t f_j(x).$$

Dividindo essa equação por $\pi(y)$, completamos a prova de (ii). O item (iii), segue diretamente do item (ii). \square

3.2 O Tempo de Relaxação

Já sabemos pela Proposição 3.1.1 que os autovalores de uma matriz de transição estão no intervalo $[-1, 1]$, assim podemos nomeá-los da seguinte forma,

$$1 = \lambda_1 > \lambda_2 \geq \dots \geq \lambda_{|\Omega|} \geq -1 \quad (3.2)$$

Definição 3.2.1. Chamamos de **buraco espectral absoluto**, a diferença

$$\gamma_\star := 1 - \lambda_\star,$$

onde $\lambda_\star := \max\{|\lambda| : \lambda \text{ é um autovalor de } P, \lambda \neq 1\}$. E o **buraco espectral** de uma cadeia reversível é definido por $\gamma := 1 - \lambda_2$.

Se tivermos que P é irredutível e aperiódica, segue pela Proposição 3.1.1 que $\gamma_\star > 0$. Como consequência da seguinte proposição, temos que os dois valores acima definidos coincidem no caso de cadeias na forma Lazy.

Proposição 3.2.2. Os autovalores de uma matriz na forma lazy assumem valores apenas entre zero e um. Portanto para uma matriz lazy temos que $\gamma_\star = \gamma$.

Demonstração. Suponha que λ_i seja um autovalor qualquer de P_L , a forma lazy da matriz P , logo podemos escrever $P_L v_i = \lambda_i v_i$, onde v_i é o autovetor associado a λ_i . Por outro lado, temos

$$P_L = \frac{P + I}{2}.$$

Daí,

$$\begin{aligned} \lambda_i v_i &= P_L v_i = \frac{P v_i + v_i}{2} \\ P v_i &= 2\lambda_i v_i - v_i = (2\lambda_i - 1)v_i. \end{aligned}$$

Portanto, $2\lambda_i - 1$ é um autovalor de P , pela Proposição 3.1.1 temos que $-1 \leq 2\lambda_i - 1 \leq 1$, ou seja, os autovalores de uma matriz na forma lazy assumem valores $0 \leq \lambda_i \leq 1$. \square

Definição 3.2.3. O tempo de relaxação t_{rel} da cadeia de Markov reversível com o buraco espectral absoluto γ_* é definido por

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

Um significado operacional do tempo de relaxação vem do próximo resultado. Mas antes temos que relembrar o que é a Variância de uma variável aleatória X . Por definição a variância de X é

$$Var X = E(X - EX)^2 = EX^2 - (EX)^2.$$

Contudo no nosso contexto fica da seguinte forma

$$Var_\pi(f) = E_\pi(f - E_\pi(f))^2 = E_\pi(f)^2 - (E_\pi(f))^2 = \langle f - E_\pi(f), f - E_\pi(f) \rangle_\pi.$$

Proposição 3.2.4.

$$Var_\pi(P^t f) \leq (1 - \gamma_*)^{2t} Var_\pi(f). \quad (3.3)$$

Demonstração. Primeiro vamos mostrar que $E_\pi(P^t f) = E_\pi(f) = \langle f, f_1 \rangle_\pi$ onde $f_1 = \mathbf{1} = (1, 1, \dots, 1)^t$. Com efeito,

$$\begin{aligned} E_\pi(P^t f) &= \sum_{x \in \Omega} (P^t f)(x) \pi(x) = \sum_{x \in \Omega} \sum_{y \in \Omega} P^t(x, y) f(y) \pi(x) \\ &= \sum_{y \in \Omega} f(y) \sum_{x \in \Omega} \pi(x) P^t(x, y) = \sum_{y \in \Omega} f(y) \pi(y) \\ &= E_\pi f = \langle f, f_1 \rangle_\pi. \end{aligned}$$

Usando o item (ii) da Proposição 3.1.2 conseguimos escrever $P^t f = \sum_{j=1}^{|\Omega|} \langle f, f_j \rangle_\pi f_j \lambda_j^t$. Disto temos

$$P^t f - E_\pi P^t f = \sum_{j=1}^{|\Omega|} \langle f, f_j \rangle_\pi f_j \lambda_j^t - \langle f, f_1 \rangle_\pi = \sum_{j=2}^{|\Omega|} \langle f, f_j \rangle_\pi f_j \lambda_j^t.$$

Agora da definição de variância e o fato de $\{f_j\}_{j=1}^{|\Omega|}$ é uma base ortonormal, segue que

$$\begin{aligned} Var_\pi(P^t f) &= \langle P^t f - E_\pi(P^t f), P^t f - E_\pi(P^t f) \rangle_\pi = \left\langle \sum_{j=2}^{|\Omega|} \langle f, f_j \rangle_\pi f_j \lambda_j^t, \sum_{j=2}^{|\Omega|} \langle f, f_j \rangle_\pi f_j \lambda_j^t \right\rangle_\pi \\ &= \sum_{j=2}^{|\Omega|} \langle f, f_j \rangle_\pi^2 \lambda_j^{2t} \leq \lambda_*^{2t} \sum_{j=2}^{|\Omega|} \langle f, f_j \rangle_\pi^2. \end{aligned}$$

Agora basta mostrar que $\sum_{j=2}^{|\Omega|} \langle f, f_j \rangle_\pi^2 = Var_\pi(f)$, com efeito,

$$\sum_{j=2}^{|\Omega|} \langle f, f_j \rangle_\pi^2 = \sum_{j=1}^{|\Omega|} \langle f, f_j \rangle_\pi^2 - \langle f, f_1 \rangle_\pi^2 = E_\pi(f^2) - (E_\pi(f))^2 = Var_\pi(f).$$

Portanto, segue o resultado. \square

Vamos mostrar como podemos limitar superiormente e inferiormente o tempo de mistura sabendo de antemão o tempo de relaxação. Em particular, conseguimos estimar o tempo de mistura encontrando os autovalores da matriz de transição.

Teorema 3.2.5. *Seja P uma matriz de transição reversível e irredutível de uma cadeia de Markov com espaço de estado Ω , e considere $\pi_{\min} = \min_{x \in \Omega} \pi(x)$. Então*

$$t_{\text{mix}}(\varepsilon) \leq \log\left(\frac{1}{\varepsilon\pi_{\min}}\right) t_{\text{rel}}. \quad (3.4)$$

Demonstração. Usando o item (iii) da Proposição 3.1.2 e aplicando a desigualdade de Cauchy-Schwarz temos

$$\left| \frac{P^t(x, y)}{\pi(y)} - 1 \right| \leq \sum_{j=2}^{|\Omega|} |f_j(x)f_j(y)| \lambda_\star^t \leq \lambda_\star^t \left[\sum_{j=2}^{|\Omega|} f_j^2(x) \sum_{j=2}^{|\Omega|} f_j^2(y) \right]^{\frac{1}{2}}. \quad (3.5)$$

Usando (3.5) e a ortogonalidade de $\{f_j\}$ temos

$$\pi(x) = \langle \delta_x, \delta_x \rangle_\pi = \left\langle \sum_{j=1}^{|\Omega|} f_j(x)\pi(x)f_j, \sum_{j=1}^{|\Omega|} f_j(x)\pi(x)f_j \right\rangle_\pi = \pi(x)^2 \sum_{j=1}^{|\Omega|} f_j(x)^2.$$

onde δ_x é a função auxiliar definida na demonstração da Proposição 3.1.2. Consequentemente, $\sum_{j=2}^{|\Omega|} f_j(x)^2 \leq \pi(x)^{-1}$. Isto, juntamente com (3.5) implica que

$$\left| \frac{P^t(x, y)}{\pi(y)} - 1 \right| \leq \frac{\lambda_\star^t}{\sqrt{\pi(x)\pi(y)}} \leq \frac{\lambda_\star^t}{\pi_{\min}} = \frac{(1 - \gamma_\star)^t}{\pi_{\min}} \leq \frac{e^{-\gamma_\star t}}{\pi_{\min}}. \quad (3.6)$$

onde a última desigualdade usamos $1 - x \leq e^{-x}$ para todo $x \geq 0$ que pode ser provada usando ferramentas de cálculo, por exemplo, observando que a função $f(x) = e^{-x} + x - 1$ é crescente quando $x \geq 0$ e $f(0) = 0$.

Aplicando o Proposição 2.2.2 temos que $d(t) \leq \pi_{\min}^{-1} e^{-\gamma_\star t}$. Fazendo algumas manipulações algébricas e usando a definição de $t_{\text{mix}}(\varepsilon)$, temos que o resultado segue como esta abaixo,

$$\begin{aligned} \varepsilon &\leq \pi_{\min}^{-1} e^{-\gamma_\star t_{\text{mix}}(\varepsilon)} \\ \varepsilon\pi_{\min} &\leq e^{-\gamma_\star t_{\text{mix}}(\varepsilon)} \\ \log(\varepsilon\pi_{\min}) &\leq -\gamma_\star t_{\text{mix}}(\varepsilon) \\ \gamma_\star t_{\text{mix}}(\varepsilon) &\leq -\log(\varepsilon\pi_{\min}) = \log\left(\frac{1}{\varepsilon\pi_{\min}}\right) \\ t_{\text{mix}}(\varepsilon) &\leq \frac{1}{\gamma_\star} \log\left(\frac{1}{\varepsilon\pi_{\min}}\right) = t_{\text{rel}} \log\left(\frac{1}{\varepsilon\pi_{\min}}\right). \end{aligned}$$

□

Teorema 3.2.6. *Para uma cadeia de Markov reversível, irredutível e aperiódica*

$$t_{\text{mix}}(\varepsilon) \geq (t_{\text{rel}} - 1) \log\left(\frac{1}{2\varepsilon}\right). \quad (3.7)$$

Demonstração. Suponha que f seja um autovetor de P com autovalor $\lambda \neq 1$, então temos $Pf = \lambda f$. Visto que os autovetores são ortogonais com relação a $\langle \cdot, \cdot \rangle_\pi$ e $\mathbf{1}$ um autovetor, então $\sum_{y \in \Omega} \pi(y)f(y) = \langle \mathbf{1}, f \rangle_\pi = 0$. Disto segue que

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

Com esta desigualdade, podemos obter um limitante inferior para o tempo de mistura. Tomando x apropriado para $|f(x)| = \|f\|_\infty$ proporciona que

$$|\lambda|^t \leq 2d(t). \quad (3.8)$$

Portanto, $|\lambda|^{t_{mix}(\varepsilon)} \leq 2d(t_{mix}(\varepsilon)) \leq 2\varepsilon$, daí segue que aplicando o logaritmo e arrumando a desigualdade anterior conseguimos,

$$\begin{aligned} t_{mix}(\varepsilon) \log(|\lambda|) &\leq \log(2\varepsilon) \\ t_{mix}(\varepsilon) \log\left(\frac{1}{|\lambda|}\right) &\geq \log\left(\frac{1}{2\varepsilon}\right). \end{aligned} \quad (3.9)$$

Usando a desigualdade $\log(1+x) \leq x$ para todo $x \geq 0$, obtemos que $\log\left(\frac{1}{|\lambda|}\right) = \log\left(1 + \frac{1}{|\lambda|} - 1\right) \leq \frac{1}{|\lambda|} - 1$, daí, multiplicando esta última desigualdade por $t_{mix}(\varepsilon)$, resulta

$$t_{mix}(\varepsilon) \left(\frac{1}{|\lambda|} - 1\right) \geq t_{mix}(\varepsilon) \log\left(\frac{1}{|\lambda|}\right). \quad (3.10)$$

Juntando as desigualdades (3.9) e (3.10), temos

$$t_{mix}(\varepsilon) \left(\frac{1}{|\lambda|} - 1\right) \geq t_{mix}(\varepsilon) \log\left(\frac{1}{|\lambda|}\right) \geq \log\left(\frac{1}{2\varepsilon}\right).$$

Passando para λ_* e ajustando os termos da desigualdade como abaixo, segue o resultado.

$$\begin{aligned} t_{mix}(\varepsilon) \left(\frac{1}{\lambda_*} - 1\right) &\geq \log\left(\frac{1}{2\varepsilon}\right) \\ t_{mix}(\varepsilon) \left(\frac{1 - \lambda_*}{\lambda_*}\right) &\geq \log\left(\frac{1}{2\varepsilon}\right) \\ t_{mix}(\varepsilon) &\geq \left(\frac{\lambda_*}{1 - \lambda_*}\right) \log\left(\frac{1}{2\varepsilon}\right) \\ t_{mix}(\varepsilon) &\geq \left(\frac{1}{1 - \lambda_*} - 1\right) \log\left(\frac{1}{2\varepsilon}\right) \\ t_{mix}(\varepsilon) &\geq (t_{rel} - 1) \log\left(\frac{1}{2\varepsilon}\right). \end{aligned}$$

□

3.3 Relação de Abertura

Nesta seção vamos limitar o tempo de mistura explorando a geometria do espaço de estado Ω . Uma abertura em Ω mostra a dificuldade de alcançar alguns lugares, limitando a velocidade de convergência.

Definição 3.3.1. *Considere P como a matriz de transição de uma cadeia de Markov irreduzível e aperiódica com distribuição estacionária π . A **medida aresta** Q é definida como*

$$Q(x, y) := \pi(x)P(x, y), \quad Q(A, B) = \sum_{x \in A, y \in B} Q(x, y).$$

A **relação de abertura**¹ de um conjunto S é definido por

$$\Phi(S) = \frac{Q(S, S^c)}{\pi(S)} \quad (3.11)$$

onde $\pi(S) = \sum_{x \in S} \pi(x)$ enquanto a relação de abertura de toda a cadeia é

$$\Phi_* := \min_{S : \pi(S) \leq \frac{1}{2}} \Phi(S).$$

Temos que $Q(A, B)$ é a probabilidade de se mover do conjunto A para o conjunto B em uma única etapa quando começamos com a distribuição estacionária. Logo, a seguir veremos a relação entre o tempo de mistura com a relação de abertura.

Teorema 3.3.2. *Considere a Relação de Abertura como definido acima, então vale*

$$t_{mix} \geq \frac{1}{4\Phi_*}.$$

Demonstração. Primeiro vamos fixar as seguintes notações

$$\pi_S(A) = \pi(A \cap S) \text{ e } \mu_S(A) = \frac{\pi_S(A)}{\pi(S)}.$$

Usando o Corolário 2.1.4 podemos escrever

$$\pi(S) \|\mu_S P - \mu_S\|_{TV} = \pi(S) \sum_{\substack{y \in \Omega \\ \mu_S P(y) \geq \mu_S(y)}} [\mu_S P(y) - \mu_S(y)].$$

Mas,

$$\pi(S) \mu_S P(y) = \pi(S) \sum_{x \in \Omega} \frac{\pi_S(x)}{\pi(S)} P(x, y) = \pi_S P(y)$$

$$\text{e } \pi(S) \mu_S(y) = \pi(S) \frac{\pi_S(y)}{\pi(S)} = \pi_S(y), \text{ então}$$

$$\pi(S) \|\mu_S P - \mu_S\|_{TV} = \sum_{\substack{y \in \Omega \\ \pi_S P(y) \geq \pi_S(y)}} [\pi_S P(y) - \pi_S(y)]. \quad (3.12)$$

¹ O termo em inglês é bottleneck ratio

Por outro lado, $\pi_S(x) = \pi(x) > 0$ se $x \in S$ e $\pi(x) = 0$ se $x \notin S$. Disto segue que, para todo $y \in S$,

$$\pi_S P(y) = \sum_{x \in \Omega} \pi_S(x) P(x, y) = \sum_{x \in S} \pi_S(x) P(x, y) \leq \sum_{x \in \Omega} \pi(x) P(x, y) = \pi(y). \quad (3.13)$$

Em contrapartida, como $\pi_S(y) = 0$ se $y \in S^c$,

$$\pi_S P(y) \geq 0 = \pi_S(y) \text{ para todo } y \in S^c. \quad (3.14)$$

Combinando as equações (3.13) e (3.14) mostra que podemos trocar o índice da somatória de (3.12) por

$$\pi(S) \|\mu_S P - \mu_S\|_{TV} = \sum_{y \in S^c} [\pi_S P(y) - \pi_S(y)] = \sum_{y \in S^c} \pi_S P(y) = \sum_{y \in S^c} \sum_{x \in S} \pi(x) P(x, y) = Q(S, S^c).$$

Portanto, dividindo a igualdade anterior por $\pi(S)$ segue que $\|\mu_S P - \mu_S\|_{TV} = \Phi(S)$. Agora, usando a desigualdade triangular, temos

$$\begin{aligned} \|\mu_S P^t - \mu_S\|_{TV} &= \sum_{u=0}^{t-1} \|\mu_S P^{u+1} - \mu_S P^u\|_{TV} \\ &= \sum_{u=0}^{t-1} \|P^u\|_{TV} \cdot \|\mu_S P - \mu_S\|_{TV} \\ &\leq t \Phi(S). \end{aligned} \quad (3.15)$$

Assumindo que $\pi(S) \leq 1/2$. É imediato que $\mu_S(S^c) = 0$ e disto

$$\|\mu_S - \pi\|_{TV} \geq \pi(S^c) - \mu_S(S^c) = 1 - \pi(S) \geq \frac{1}{2}.$$

Usando novamente a desigualdade triangular, temos

$$\frac{1}{2} \leq \|\mu_S - \pi\|_{TV} \leq \|\mu_S - \mu_S P^t\|_{TV} + \|\mu_S P^t - \pi\|_{TV}. \quad (3.16)$$

Tomando $t = t_{mix}$ e usando a definição do tempo de mistura na equação (3.16) resulta que

$$\frac{1}{2} \leq t_{mix} \Phi(S) + \frac{1}{4}.$$

Reorganizando e minimizando em S segue o resultado. \square

Teorema 3.3.3. *Seja P a matriz de uma cadeia de Markov irreduzível, aperiódica e reversível. Então, vale que*

$$\frac{\Phi_\star^2}{2} \leq \gamma \leq 2\Phi_\star$$

onde $\gamma = 1 - \lambda_2$ é o buraco espectral.

Para a demonstração desse teorema é necessária uma teoria que não será abordada neste texto por uma questão de foco. Mas pode ser encontrada na referência [3] na seção 3 do capítulo 13.

Vamos concluir essa seção com duas observações.

(i) Revendo a demonstração do Teorema 3.3.2, também conseguimos o resultado

$$t_{mix}(\varepsilon) \geq \frac{1 - 2\varepsilon}{2\Phi_\star}.$$

(ii) Em algumas casos é melhor ver o resultado do Teorema 3.3.3 como segue

$$\frac{1}{2\Phi_\star} \leq t_{rel} \leq \frac{2}{\Phi_\star^2}.$$

4 O Fenômeno *Cutoff*

Até agora estudamos a dinâmica de uma cadeia de Markov e vimos as suas propriedades, porém podemos pensar como seria uma sequência de cadeias de Markov, ou seja, uma sequência cujo elementos são sequência de variáveis aleatórias. Iremos ver como fica a sua convergência para o equilíbrio nesse novo contexto.

A forma mais simples de pensar numa sequência de cadeia de Markov é pensar que o espaço de estado Ω aumenta a cada elemento da sequência de cadeias. É claro que, dependendo do problema, podemos ter as mais diversas situações.

Como o nome do capítulo sugere vamos estudar o fenômeno *cutoff* que nos fornece um comportamento abrupto na curva do gráfico tempo \times distância estacionária.

4.1 Definição e Primeiro Resultados

Definição 4.1.1. *Considere uma sequência de cadeias de Markov indexadas por $n = 1, 2, 3, \dots$ e o tempo de mistura da n ésima cadeia será denotado por $t_{mix}^n(\varepsilon)$. Dizemos que esta sequência de cadeias apresenta *cutoff* se, para todo $0 < \varepsilon < 1$ vale*

$$\lim_{n \rightarrow \infty} \frac{t_{mix}^n(\varepsilon)}{t_{mix}^n(1 - \varepsilon)} = 1. \quad (4.1)$$

Uma observação que decorre da definição é que podemos restringir $0 < \varepsilon < \frac{1}{2}$ por causa do limite da definição ser 1. Porém vamos adotar como está nas referências [3].

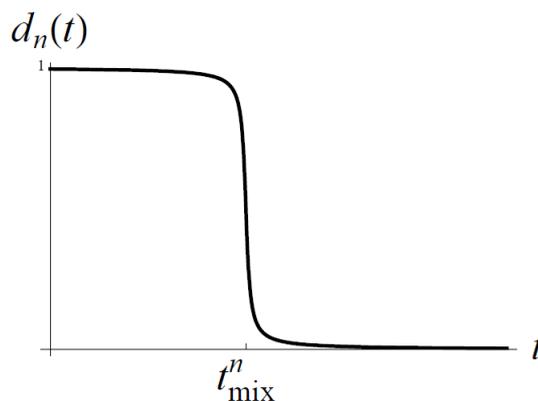


Figura 6 – O gráfico de uma sequência de cadeias de *Cutoff*, quando visto na escala de tempo de t_{mix}^n adequado se aproxima de uma função degrau

O cálculo para encontrar o tempo de mistura, $t_{mix}^n(\varepsilon)$, na maioria das vezes é muito trabalhoso o que dificulta identificar se uma sequência de cadeias de Markov apresenta *cutoff*. O nosso próximo resultado caracteriza se uma sequência de cadeias possui *cutoff*

através da distância para a estacionariedade. Daqui em diante, iremos sempre denotar $t_{mix}^n(\varepsilon)$, t_{rel}^n e d_n , como o n -ésimo tempo de mistura, tempo de relaxação e distância para a estacionariedade, respectivamente, de uma sequência de cadeias de Markov.

Proposição 4.1.2. *Uma sequência de cadeias de Markov apresenta cutoff se, e somente se,*

$$\lim_{n \rightarrow \infty} d_n(ct_{mix}^n) = \begin{cases} 1, & \text{se } 0 < c < 1, \\ 0, & \text{se } c > 1. \end{cases}$$

Demonstração. Suponha que vale (4.1). Seja $\varepsilon > 0$, por definição, dado $\gamma > 0$, existe $n_0 \in \mathbb{N}$ tal que, se $n > n_0$, vale

$$t_{mix}^n(\varepsilon) < (\gamma + 1) t_{mix}^n(1 - \varepsilon).$$

Fazendo uma escolha apropriada para ε , temos que

$$t_{mix}^n(1 - \varepsilon) \leq t_{mix}^n \leq t_{mix}^n(\varepsilon).$$

Juntando estas duas equações, conseguimos

$$t_{mix}^n(\varepsilon) < (\gamma + 1) t_{mix}^n(1 - \varepsilon) < (\gamma + 1) t_{mix}^n \text{ para todo } n > n_0.$$

Como d_n é uma função não-crescente teremos

$$d_n((\gamma + 1) t_{mix}^n) \leq d_n(t_{mix}^n(\varepsilon)) \leq \varepsilon \text{ para todo } n > n_0.$$

Portanto, $\lim_{n \rightarrow \infty} d_n(c t_{mix}^n) = 0$, onde estamos tomando $c = \gamma + 1 > 1$. De forma análoga mostra que $\lim_{n \rightarrow \infty} d_n(c t_{mix}^n) = 1$ se $0 < c < 1$.

Reciprocamente, suponha que vale

$$\lim_{n \rightarrow \infty} d_n(ct_{mix}^n) = \begin{cases} 1, & \text{se } 0 < c < 1, \\ 0, & \text{se } c > 1. \end{cases}$$

Dado $\gamma > 0$ teremos que $\lim_{n \rightarrow \infty} d_n((\gamma + 1) t_{mix}^n) = 0$. Por definição, dado $\varepsilon > 0$ existe $n_1 \in \mathbb{N}$ tal que $d_n((\gamma + 1) t_{mix}^n) < \varepsilon$ sempre que $n > n_1$. Logo por definição do tempo de mistura temos

$$t_{mix}^n(\varepsilon) \leq (1 + \gamma)t_{mix}^n \text{ para } n > n_1.$$

Analogamente, $\lim_{n \rightarrow \infty} d_n((1 - \gamma)t_{mix}^n) = 1$. Logo por definição de limite, existe $n_2 \in \mathbb{N}$, tal que para $n > n_2$, $d_n((1 - \gamma)t_{mix}^n) > 1 - \varepsilon$, portanto,

$$t_{mix}^n(1 - \varepsilon) \geq (1 - \gamma)t_{mix}^n \text{ para } n > n_2.$$

Logo, para $n > \max\{n_1, n_2\}$

$$\frac{t_{mix}^n(\varepsilon)}{t_{mix}^n(1 - \varepsilon)} \leq \frac{1 + \gamma}{1 - \gamma} \xrightarrow{\gamma \rightarrow 0} 1.$$

Por outro lado, para $0 < \varepsilon < \frac{1}{2}$ temos que

$$1 \leq \frac{t_{mix}^n(\varepsilon)}{t_{mix}^n(1-\varepsilon)}.$$

Portante, segue o resultado. □

Existe uma forma equivalente de definir o fenômeno de *cuttloff* que é adotada por outros autores e referências, o nosso próximo resultado mostra esta forma equivalente.

Proposição 4.1.3. *A definição do fenômeno de Cuttloff é equivalente a*

$$t_{mix}^n(\varepsilon) - t_{mix}^n(1-\varepsilon) = o(t_{mix}^n) \text{ para todo } 0 < \varepsilon < \frac{1}{4}.$$

Demonstração. Suponha que vale a definição de *cuttloff* e seja $0 < \varepsilon < \frac{1}{4}$, assim temos que vale $\frac{1}{4} < 1 - \varepsilon$ e disto implica que

$$\frac{t_{mix}^n(1-\varepsilon)}{t_{mix}^n} \leq 1.$$

Por outro lado,

$$\begin{aligned} 0 \leq \frac{t_{mix}^n(\varepsilon) - t_{mix}^n(1-\varepsilon)}{t_{mix}^n} &= \left[\frac{t_{mix}^n(\varepsilon)}{t_{mix}^n(1-\varepsilon)} - 1 \right] \cdot \frac{t_{mix}^n(1-\varepsilon)}{t_{mix}^n} \\ &\leq \frac{t_{mix}^n(\varepsilon)}{t_{mix}^n(1-\varepsilon)} - 1 \\ &\xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Portanto, $t_{mix}^n(\varepsilon) - t_{mix}^n(1-\varepsilon) = o(t_{mix}^n)$ para todo $0 < \varepsilon < \frac{1}{4}$.

Reciprocamente, suponha que vale para $0 < \varepsilon < \frac{1}{4}$,

$$t_{mix}^n(\varepsilon) - t_{mix}^n(1-\varepsilon) = o(t_{mix}^n).$$

Temos que, para todo $0 < \varepsilon < \frac{1}{2}$, vale a expressão

$$\frac{t_{mix}^n(\varepsilon)}{t_{mix}^n(1-\varepsilon)} \geq 1.$$

Além disso, usando a hipótese temos que dado $\xi > 0$ e para n suficientemente grande vale e se $0 < \varepsilon < \frac{1}{4}$

$$t_{mix}^n(1-\varepsilon) \geq t_{mix}^n(\varepsilon) - \xi t_{mix}^n.$$

mas, como $0 < \varepsilon < \frac{1}{4}$ implica que $\frac{t_{mix}^n}{t_{mix}^n(\varepsilon)} \geq 1$ temos

$$\begin{aligned} 1 \leq \frac{t_{mix}^n(\varepsilon)}{t_{mix}^n(1-\varepsilon)} &\leq \frac{t_{mix}^n(\varepsilon)}{t_{mix}^n(\varepsilon) - \xi t_{mix}^n} = \frac{1}{1 - \xi \frac{t_{mix}^n}{t_{mix}^n(\varepsilon)}} \\ &\leq \frac{1}{1 - \xi} \xrightarrow{\xi \rightarrow 0} 1. \end{aligned}$$

Portanto vale para todo $0 < \varepsilon < \frac{1}{4}$,

$$\lim_{n \rightarrow \infty} \frac{t_{mix}^n(\varepsilon)}{t_{mix}^n(1 - \varepsilon)} = 1.$$

Note que vale,

$$1 = \lim_{n \rightarrow \infty} \frac{t_{mix}^n(\varepsilon)}{t_{mix}^n(1 - \varepsilon)} = \lim_{n \rightarrow \infty} \frac{t_{mix}^n(1 - \varepsilon)}{t_{mix}^n(\varepsilon)}$$

disto resulta que a nossa tese também é válida para $\frac{3}{4} < \varepsilon < 1$.

Agora considere $\frac{1}{4} < \varepsilon \leq \frac{1}{2}$, assim basta considerar $0 < \varepsilon' < \frac{1}{4} < \varepsilon$ que resulta

$$1 \leq \frac{t_{mix}^n(\varepsilon)}{t_{mix}^n(1 - \varepsilon)} \leq \frac{t_{mix}^n(\varepsilon')}{t_{mix}^n(1 - \varepsilon')} \xrightarrow{n \rightarrow \infty} 1.$$

Então, usando o mesmo argumento anterior implica que para todo $0 < \varepsilon < 1$ vale,

$$\lim_{n \rightarrow \infty} \frac{t_{mix}^n(\varepsilon)}{t_{mix}^n(1 - \varepsilon)} = 1.$$

□

Definição 4.1.4 (Janela de *cutoff*). *Uma sequência de cadeias de Markov com cutoff tem uma janela de tamanho $\{\omega_n\}$ se*

- 1) $\omega_n = o(t_{mix}^n)$;
- 2) Para todo $0 < \varepsilon < 1$ existe $c_\varepsilon > 0$ tal que

$$t_{mix}^n(\varepsilon) - t_{mix}^n(1 - \varepsilon) \leq c_\varepsilon \omega_n \text{ para todo } n.$$

Uma conta simples implica que as condições 1 e 2 da definição de janela de *cutoff* não implica na definição de *cutoff*, isso é também claramente visto usando a Proposição 4.1.3. Na verdade a janela de *cutoff* é, graficamente, a queda abrupta que vemos no gráfico, veja a Figura 6. Abaixo iremos definir um outro fenômeno que também pode ocorrer.

Definição 4.1.5 (Pré-Cutoff). *Uma sequência de cadeias de Markov possui pré-cutoff se satisfaz*

$$\sup_{0 < \varepsilon < \frac{1}{2}} \limsup_{n \rightarrow \infty} \frac{t_{mix}^n(\varepsilon)}{t_{mix}^n(1 - \varepsilon)} < \infty.$$

Segue da definição, sem dificuldades, que qualquer sequência de cadeias de Markov que apresenta o *cutoff* apresenta também o pré-*cutoff*, mas claramente a recíproca é falsa, para facilitar o nosso entendimento, veja um gráfico de uma sequência de cadeias que exibem o pré-*cutoff* (Figura 7).

Recorde do capítulo 3 o conceito de tempo de relaxação que é

$$t_{rel} = \frac{1}{\gamma} = \frac{1}{1 - \lambda_2}$$

onde γ é o buraco espectral para uma cadeia reversível e λ_2 é o segundo maior autovalor da matriz de transição.

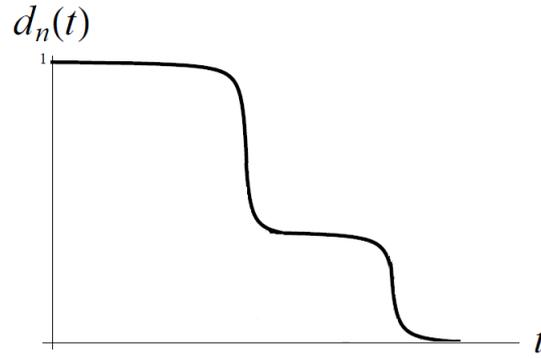


Figura 7 – O gráfico de uma sequência de cadeias com pré-*cuttoff*, quando visto na escala de tempo de t_{mix}^n adequado também se aproxima de uma função degrau, só não tem a obrigatoriedade de ser apenas um degrau.

Definição 4.1.6 (Condição Produto). *Considere uma sequência de cadeias de Markov irredutível, reversível e na forma lazy. Dizemos que esta família de cadeias satisfaz a condição produto se $(1 - \lambda_2^n)t_{mix}^n \rightarrow \infty$, ou de forma equivalente, $t_{rel}^n = o(t_{mix}^n)$.*

A condição produto aparece bastante na literatura atual, isso se justifica não só pela sua importância na teoria, mas também pela conjectura feita por Yuval Peres no ano de 2004. Ele conjecturou que a condição produto era suficiente e necessária para ocorrer o fenômeno de *cuttoff*. Porém, a condição produto não implica, necessariamente, em *cuttoff*. Aldous construiu um contra-exemplo relevante que nos aponta que apenas a condição produto não é suficiente para nos garantir que ocorra o *cuttoff* como será visto mais adiante. Abaixo vamos ver mais alguns resultados.

Proposição 4.1.7. *Considere uma sequência de cadeias de Markov irredutíveis, aperiódicas e reversíveis. Se $\frac{t_{mix}^n}{t_{rel}^n}$ for limitado superiormente, então a sequência não apresenta o pré-*cuttoff*.*

Demonstração. Usando a hipótese podemos escrever que $\frac{t_{rel}^n - 1}{t_{mix}^n} \geq c$ para algum $c \in \mathbb{R}$, e pelo Teorema 3.2.6 conseguimos,

$$\begin{aligned} \frac{t_{mix}^n(\varepsilon)}{t_{mix}^n(1 - \varepsilon)} &\geq \frac{(t_{rel}^n - 1) \log\left(\frac{1}{2\varepsilon}\right)}{t_{mix}^n} \\ &\geq c \log\left(\frac{1}{2\varepsilon}\right). \end{aligned}$$

Como $\log\left(\frac{1}{2\varepsilon}\right)$ explode para o infinito quando o $\varepsilon \rightarrow 0^-$, portanto teremos que

$$\sup_{0 < \varepsilon < \frac{1}{2}} \limsup_{n \rightarrow \infty} \frac{t_{mix}^n(\varepsilon)}{t_{mix}^n(1 - \varepsilon)} = \infty$$

disto, por definição, a sequência de cadeias de Markov não apresenta o pré-*cuttoff*, logo não apresenta o *cuttoff*. \square

Corolário 4.1.8. *Considere uma sequência de cadeias de Markov irredutíveis, aperiódicas e reversíveis. Se esta sequência apresenta cutoff então vale a condição produto.*

Demonstração. Basta notar que a definição de *cutoff* implica em pré-*cutoff* e usar a proposição anterior. \square

A melhor leitura que podemos ter da Proposição 4.1.7 é olhando a contra-positiva que seria a hipótese de pré-*cutoff* implica na condição produto. Vale a pena salientar também que a Corolário 4.1.8 nos diz que a luz da hipótese do fenômeno de cutoff implica também na condição produto.

4.2 Contra-exemplo do Aldous

Neste capítulo apresentamos alguns conceitos relacionados ao fenômeno *cutoff*, a saber, a definição de *cutoff* e suas equivalências; a definição de pré-*cutoff* e a condição produto. Vimos também como estes conceitos se relacionam com os resultados expostos. Abaixo, segue um diagrama que exhibe, de forma breve, do que estamos falando.

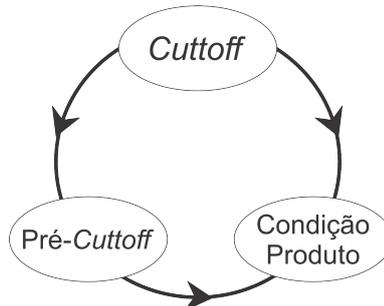


Figura 8 – Relação entre: *cutoff*, pré-*cutoff* e condição produto.

Claramente os conceitos de *cutoff* e pré-*cutoff* não são equivalentes. E, como já foi mencionado, Aldous construiu um contra-exemplo que mostra uma sequência de cadeias de Markov que satisfaz a condição produto mas não exhibe o fenômeno *cutoff* (veja o capítulo 18 da referência [3]). A referência [6] mostra uma variação do contra-exemplo do Aldous. Iremos exhibir uma sequência de cadeias de Markov que tem a condição produto, porém não apresenta o fenômeno de *cutoff* baseado no contra-exemplo da referência [6].

Exemplo 4.2.1. *Considere a sequência de cadeias (Ω_n, P_n, π_n) , onde $\Omega_n = \{-10n, -10n+2, -10n+4, \dots, -2, 0, 1, 2, 3, 4, \dots, 2n, 2n+1\}$. As probabilidades são dadas da seguinte maneira: de ficar no próprio ponto $P_n(x, x) = 1/2$ se x é par, $P_n(x, x) = 3/4$ se x é ímpar e $x < 2n+1$ e $P_n(2n+1, 2n+1) = 9/10$; de saltar para os vizinhos*

$$P_n(2i, \min\{2i+2, 2n+1\}) = 2P_n(2i, 2i-2) = 2P_n(2i-1, 2i+1) = 4P_n(2i-1, \max\{2i-3, 0\}) = \frac{1}{3}$$

Finalmente, os pontos que estão faltando são $P_n(-10n, -10n + 2) = 1/2$, $P_n(0, 2) = P_n(0, 1) = 2P_n(0, -2) = 1/5$ e $P_n(2n + 1, 2n) = P_n(2n + 1, 2n - 1) = 1/20$. Veja a figura 9.

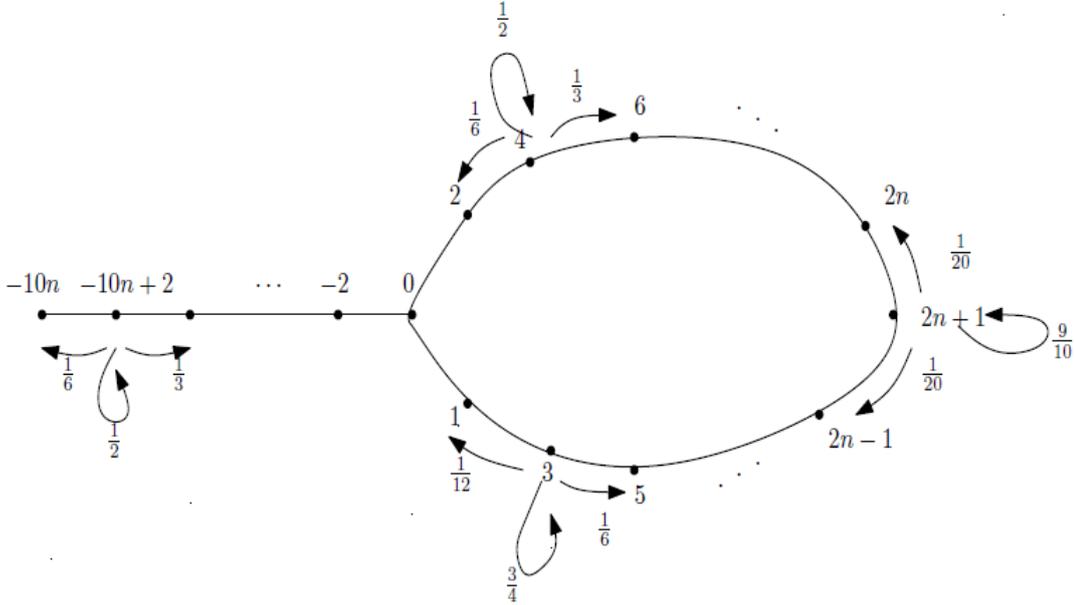


Figura 9 – Contra-exemplo. Fonte referência [7] (2015).

Sem grandes problemas vemos que P_n é irredutível e aperiódica (pois P_n é *lazy*). Como P_n é irredutível a Proposição 1.3.4 garante que existe a distribuição estacionária π_n para cada $n \in \mathbb{N}$. E uma maneira é tentar ver se as EBD são satisfeitas, em cujo caso a medida estacionária também será reversível. A ideia é deixar todas as entradas de π_n em função do valor de $\pi_n(-10n)$ e depois usar o fato que a somatória de todas as entradas de π_n é 1 e descobrir o valor de $\pi_n(-10n)$ e, conseqüentemente, o valor das demais entradas.

Relembrando que uma distribuição estacionária satisfaz as EBD se para todo $x, y \in \Omega$ vale,

$$\pi_n(x)P_n(x, y) = \pi_n(y)P_n(y, x).$$

Pensando dessa forma, a primeira equação será,

$$\begin{aligned} \pi_n(-10n)P_n(-10n, -10n + 2) &= \pi_n(-10n + 2)P_n(-10n + 2, -10n) \\ \pi_n(-10n + 2) &= 2\pi_n(-10n). \end{aligned}$$

E continuando de forma recorrente, temos

$$\pi_n(-10n + 2k) = 2^k \pi_n(-10n) \text{ para todo } k \in \{1, 2, \dots, 5n - 1\}. \quad (4.2)$$

Se continuarmos com a recorrência, conseguimos calcular a distribuição no ponto 0, ficando da seguinte forma

$$\pi_n(0) = \frac{5}{3} 2^{5n} \pi_n(-10n). \quad (4.3)$$

Agora, para os pontos pares e ímpares seguintes a distribuição estacionária fica da seguinte forma, para $k \in \{1, 2, 3, \dots, n\}$

$$\pi_n(2k) = 2^{k-1}2^{5n+1}\pi_n(-10n); \quad (4.4)$$

$$\pi_n(2k-1) = 2^{k-1}2^{5n+2}\pi_n(-10n). \quad (4.5)$$

Por fim, continuando com a recorrência, tanto pelos pares quanto pelos ímpares, atingimos o último ponto com a seguinte expressão

$$\pi_n(2n+1) = \frac{5}{3}2^{6n+2}\pi_n(-10n). \quad (4.6)$$

Afim de encontrar o valor de $\pi_n(-10n)$ basta usar o fato que a soma das entradas da distribuição estacionária é 1, ou seja,

$$\begin{aligned} 1 = & \left[\sum_{k=0}^{5n-1} 2^k \pi_n(-10n) \right] + \left[\frac{5}{3}2^{5n} \pi_n(-10n) \right] + \left[\sum_{k=1}^n 2^{k-1}2^{5n+1} \pi_n(-10n) \right] \\ & + \left[\sum_{k=1}^n 2^{k-1}2^{5n+2} \pi_n(-10n) \right] + \left[\frac{5}{3}2^{6n+2} \pi_n(-10n) \right]. \end{aligned}$$

Daí, resolvendo esses somatórios e reagrupando os termos, temos

$$\begin{aligned} \frac{1}{\pi_n(-10n)} &= [2^{5n} - 1] + \left[\frac{5}{3}2^{5n} \right] + [(2^n - 1)2^{5n+1}] + [(2^n - 1)2^{5n+2}] + \left[\frac{5}{3}2^{6n+2} \right] \\ &= \frac{19}{3}2^{6n+1} - \frac{10}{3}2^{5n} - 1 \end{aligned}$$

Portanto, conseguimos as seguintes estimativas

$$3 \cdot 2^{6n+2} = \frac{18}{3}2^{6n+1} < \frac{1}{\pi_n(-10n)} < \frac{20}{3}2^{6n+1} = \frac{5}{3}2^{6n+3}. \quad (4.7)$$

Desta forma, temos a estimativa de $\pi_n(-10n) > \frac{3}{5} \frac{1}{2^{6n+3}}$ e disto, usando a equação (4.6), temos

$$\pi_n(2n+1) = \frac{5}{3}2^{6n+2}\pi_n(-10n) > \frac{1}{2}. \quad (4.8)$$

Denote $S_n = \Omega \setminus \{2n+1\}$ e vamos usar a relação de abertura para estimar o tempo de relaxação t_{rel} . Com efeito, usando a definição da relação de abertura e o fato que a cadeia é reversível, temos que,

$$\begin{aligned} \Phi^{(n)}(S_n) &= \frac{Q(S_n, S_n^c)}{\pi_n(S_n)} = \frac{\sum_{x \in S_n, y \in S_n^c} Q(x, y)}{\pi_n(S_n)} \\ &= \frac{\sum_{x \in S_n, y \in S_n^c} \pi_n(x)P_n(x, y)}{1 - \pi_n(2n+1)} = \frac{\sum_{x \in S_n, y \in S_n^c} \pi_n(y)P_n(y, x)}{1 - \pi_n(2n+1)} \\ &= \frac{\pi_n(2n+1)}{1 - \pi_n(2n+1)} [P_n(2n+1, 2n) + P_n(2n+1, 2n-1)] = \frac{\pi_n(2n+1)}{1 - \pi_n(2n+1)} \cdot \frac{1}{10}. \end{aligned}$$

Como $\pi_n(S_n^c) = \pi_n(2n+1) > 1/2$ então $\pi_n(S_n) \leq 1/2$ e disto segue que

$$\Phi_\star^{(n)} = \frac{\pi_n(2n+1)}{1 - \pi_n(2n+1)} \cdot \frac{1}{10}. \quad (4.9)$$

Usando a equação (4.8) resulta que $1/2 < \pi_n(2n+1) < 1$, logo

$$\Phi_\star^{(n)} > \frac{1}{10}. \quad (4.10)$$

Do item (ii) da observação do Teorema 3.3.3, conseguimos

$$t_{rel} \leq \frac{2}{(\Phi_\star^{(n)})^2} < 200.$$

Portanto, $t_{rel} = O(1)$, em particular vale a condição produto. Agora vamos mostrar que não vale o *cuttoff*. Com efeito, vamos mostrar que falha $t_{mix}^n(\varepsilon) - t_{mix}^n(1-\varepsilon) = o(t_{mix}^n)$ para todo $0 < \varepsilon < 1/4$. Para isso, uma forma é mostrando que

$$\frac{t_{mix}^n}{t_{mix}^n(\varepsilon) - t_{mix}^n(1-\varepsilon)} \leq C$$

para algum $C > 0$ e para todo $n \in \mathbb{N}$. Combinando os Teoremas 3.2.5 e 3.2.6 resulta que

$$(t_{rel}^n - 1) \log\left(\frac{1}{2\varepsilon}\right) \leq t_{mix}^n(\varepsilon) \leq t_{rel}^n \log\left(\frac{1}{\varepsilon\pi_n(-10n)}\right). \quad (4.11)$$

Disto,

$$t_{mix}^n(\varepsilon) - t_{mix}^n(1-\varepsilon) \geq (t_{rel}^n - 1) \log\left(\frac{1}{2\varepsilon}\right) - t_{rel}^n \log\left(\frac{1}{(1-\varepsilon)\pi_n(-10n)}\right). \quad (4.12)$$

Usando novamente o item (ii) da observação do Teorema 3.3.3 na equação (4.12)

$$t_{mix}^n(\varepsilon) - t_{mix}^n(1-\varepsilon) \geq \left(\frac{1}{2\Phi_\star^{(n)}} - 1\right) \log\left(\frac{1}{2\varepsilon}\right) - \frac{2}{(\Phi_\star^{(n)})^2} \log\left(\frac{1}{(1-\varepsilon)\pi_n(-10n)}\right). \quad (4.13)$$

Por outro lado, também vamos limitar o t_{mix}^n usando os mesmos argumentos, teremos que

$$t_{mix}^n \leq \log\left(\frac{4}{\pi_n(-10n)}\right) \cdot \frac{2}{(\Phi_\star^{(n)})^2}. \quad (4.14)$$

Dividindo a equação (4.14) pela equação (4.13) teremos,

$$\frac{t_{mix}^n}{t_{mix}^n(\varepsilon) - t_{mix}^n(1-\varepsilon)} \leq \frac{4 \log\left(\frac{4}{\pi_n(-10n)}\right)}{\Phi_\star^{(n)}(1 - \Phi_\star^{(n)}) \log\left(\frac{1}{2\varepsilon}\right) - 4 \log\left(\frac{1}{(1-\varepsilon)\pi_n(-10n)}\right)}. \quad (4.15)$$

Usando as equações (4.8), (4.9) e (4.10) conseguimos a seguinte relação,

$$\Phi_\star^{(n)}(1 - \Phi_\star^{(n)}) > \frac{1}{10}(1 - \Phi_\star^{(n)}) = \frac{10 - 11\pi_n(2n+1)}{100[1 - \pi_n(2n+1)]} > \frac{10 - 11\pi_n(2n+1)}{50}. \quad (4.16)$$

Substituindo a cota da equação (4.16) na expressão (4.15) e usando a desigualdade $\log(x) < x$ para todo $x > 0$, logo,

$$\begin{aligned} \frac{t_{mix}^n}{t_{mix}^n(\varepsilon) - t_{mix}^n(1 - \varepsilon)} &\leq \frac{\frac{16}{\pi_n(-10n)}}{\Phi_\star^{(n)}(1 - \Phi_\star^{(n)}) + 4\pi_n(-10n)} \leq \frac{\frac{16}{\pi_n(-10n)}}{\frac{50}{10 - 11\pi_n(2n + 1)} + 4\pi_n(-10n)} \\ &= \frac{\frac{16}{\pi_n(-10n)}}{\frac{50 + 4\pi_n(-10n)[10 - 11\pi_n(2n + 1)]}{10 - 11\pi_n(2n + 1)}} \\ &= \frac{16}{\pi_n(-10n)} \cdot \frac{10 - 11\pi_n(2n + 1)}{50 + 40\pi_n(-10n) - 44\pi_n(-10n)\pi_n(2n + 1)} \end{aligned}$$

Por fim, vamos usar a equação (4.6) e (4.7), obtemos,

$$\begin{aligned} \frac{t_{mix}^n}{t_{mix}^n(\varepsilon) - t_{mix}^n(1 - \varepsilon)} &< 16 \cdot \frac{\left[\frac{10}{\pi_n(-10n)} - \frac{11\pi_n(2n + 1)}{\pi_n(-10n)} \right]}{40\pi_n(-10n) - 44\pi(-10n)\pi_n(2n + 1)} \\ &< 16 \cdot \frac{\left[10\frac{5}{3}2^{6n+3} - 11\frac{5}{3}2^{6n+2} \right]}{40\frac{5}{3}2^{6n+3} - 44\pi_n(-10n)\frac{5}{3}2^{6n+2}\pi_n(-10n)} \\ &= \frac{144}{80 - 44[\pi_n(-10n)]^2} \xrightarrow{n \rightarrow \infty} = \frac{9}{5}, \end{aligned}$$

como queríamos demonstrar. No artigo [7] mostra que esse contra-exemplo possui duas quedas, caracterizando o pré-*cutoff*, veja a figura 4.2

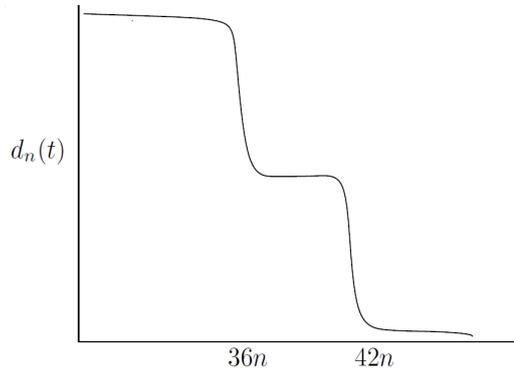


Figura 10 – O decaimento do contra-exemplo do Aldous. Fonte: referência [7] (2015).

5 Passeio Aleatório e Rede

Neste capítulo iremos ver a relação entre cadeias de Markov reversíveis e redes elétricas. Iremos definir redes sem levar em conta os fenômenos empíricos, porém veremos que os resultados alcançados serão os mesmos. A grande vantagem dessa associação é que a teoria de redes elétricas é bem conhecida e contida, assim pegamos um problema de probabilidade que é difícil de resolver, associamos e resolvemos com a teoria de redes, depois voltamos com a resposta para o contexto probabilidade.

Para isso, precisamos compreender bem sobre passeio aleatório em Grafos, uma vez que, para o nosso contexto, uma rede é um grafo cujas arestas possuem pesos.

5.1 Passeio aleatório no Grafo

Vamos começar esta seção com um exemplo.

Exemplo 5.1.1. *Seja $\Omega = \mathbb{Z} \setminus n\mathbb{Z} = \mathbb{Z}_n$ o toro discreto 1-dimensional com n pontos. Denote \mathbb{Z}_n por $\{0, 1, 2, 3, \dots, n-1\}$. Considere a matriz de transição*

$$P(j, k) = \begin{cases} 1/2 & \text{se } k \equiv j + 1 \pmod{n} \\ 1/2 & \text{se } k \equiv j - 1 \pmod{n} \\ 0 & \text{caso contrário} \end{cases}$$

a cadeia de Markov (X_t) é também chamada de passeio aleatório no n -ciclo. Podemos imaginar que a dinâmica da cadeia ocorre sobre um círculo e os estados são pontos sobre o círculo igualmente espaçados.

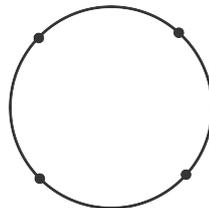


Figura 11 – Passeio aleatório no 4-ciclo.

Num passeio aleatório no 4-ciclo a matriz de transição fica com a seguinte forma:

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

O passeio aleatório no n -ciclo é um exemplo simples, porém importante de cadeia de Markov em grafos.

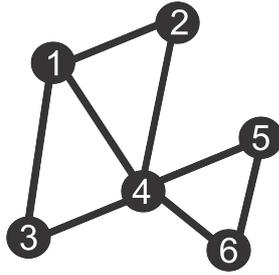


Figura 12 – Grafo com o conjunto de vértices $\{1, 2, 3, 4, 5, 6\}$ e 8 arestas.

Definição 5.1.2 (Grafo). Um **Grafo** $G = (V, E)$ consiste em um conjunto de vértices V (também chamados de pontos ou nós) e de um conjunto de arestas E (também chamado de elos), onde os elementos de E são pares não-ordenados de vértices: $E \subset \{\{x, y\} : x, y \in V\}$. Quando $\{x, y\} \in E$, escrevemos que $x \sim y$ e dizemos que x é vizinho de y (ou y é vizinho de x). O **grau do vértice** x , denotado por $\text{grau}(x)$ é o número de vizinhos de x . Um grafo é **conexo** se é possível chegar em qualquer vértice, a partir de um vértice inicial, movendo apenas pelas arestas, caso contrário chamamos o grafo de **desconexo**.

De forma mais intuitiva, podemos imaginar que V é um conjunto de pontos onde x e y são ligados por uma linha se, e somente se, $\{x, y\} \in E$.

Definição 5.1.3 (Passeio Aleatório simples em G). Dado um grafo $G = (V, E)$ definimos o **passeio aleatório simples em G** como a cadeia de Markov com espaço de estado V e matriz de transição

$$P(x, y) = \begin{cases} \frac{1}{\text{grau}(x)} & \text{se } y \sim x; \\ 0 & \text{caso contrário.} \end{cases}$$

Exemplo 5.1.4. Considere o grafo G como sendo exibido na figura 12. A matriz de transição no passeio aleatório simples é dada pela matriz

$$P = \begin{pmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & 0 & \frac{1}{5} & \frac{1}{5} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}.$$

Um das vantagens de passeios aleatórios simples em grafos é a facilidade para encontrar a distribuição estacionária. Considere um grafo $G = (V, E)$ e $y \in V$. Lembrando que encontramos a distribuição estacionária através de $\pi P = \pi$, temos

$$\sum_{x \in V} \text{grau}(x) P(x, y) = \sum_{x \sim y} \frac{\text{grau}(x)}{\text{grau}(x)} = \text{grau}(y).$$

Como a distribuição estacionária é uma medida de probabilidade basta normalizar pelo fator $\sum_{y \in V} \text{grau}(y) = 2|E|$ (este fato não é difícil de checar, basta se convencer que cada aresta é contada duas vezes). Assim conclui-se

$$\pi(y) = \frac{\text{grau}(y)}{2|E|} \text{ para todo } y \in \Omega.$$

ou seja, cada entrada da distribuição é proporcional ao grau. No Exemplo 5.1.4 a distribuição estacionária é

$$\pi = \left(\frac{3}{16} \frac{2}{16} \frac{2}{16} \frac{5}{16} \frac{2}{16} \frac{2}{16} \right).$$

Vamos também definir mais dois tipos de grafos que iremos precisar no final desse capítulo. É claro que existem os tipos mais variados de grafos, mas iremos restringir aos tipos que iremos precisar para compreender uma construção que será vista na última seção desse capítulo.

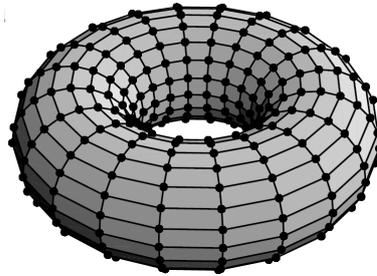


Figura 13 – Um exemplo de toro: 2-toro \mathbb{Z}_{20}^2 . Fonte: referência [3] (2001).

Definição 5.1.5. Um grafo é chamado *d-regular* se cada vértice tem o mesmo número de arestas d . Um toro *d-dimensional* com n^d pontos é um grafo do qual o conjunto de seus vértices é o produto cartesiano

$$\mathbb{Z}_n^d = \underbrace{\mathbb{Z}_n \times \mathbb{Z}_n \times \cdots \times \mathbb{Z}_n}_{d \text{ vezes}}.$$

Dizemos que os pontos $x = (x_1, \dots, x_d)$ e $y = (y_1, \dots, y_d)$ do toro são vizinhos se para algum $j \in \{1, 2, \dots, d\}$ temos $x_i = y_i$ para todo $i \neq j$ e $x_j = y_j \equiv \pm 1 \pmod{n}$. Veja Figura (13).

É interessante fazer algumas observações sobre o grafo *d-regular*, o primeiro é que vale a seguinte igualdade $2|E| = d \cdot |V|$ e, a segunda é que a sua distribuição estacionária é a uniforme com $\pi(y) = \frac{1}{|V|}$ para todo $y \in V$.

5.2 Redes e Cadeias de Markov Reversíveis

As Redes Elétricas trazem uma diferente linguagem para as cadeias de Markov reversíveis. Veremos que existe uma relação biunívoca sobre estes dois conceitos. Esta nova

linguagem é bem útil para o estudo de cadeias de Markov reversíveis porque ganhamos o conhecimento das leis físicas de rede elétricas.

Definição 5.2.1 (Rede). Uma **Rede** é um grafo finito conexo $G = (V, E)$ dotado de números não-negativos $C(e)$, para $e \in E$. $C(e)$ é chamado de condutância que está associado a aresta do grafo. A inversa $R(e) = \frac{1}{C(e)}$ é chamado de resistência da aresta e denotamos uma rede como o par $(G, \{C(e)\})$ e chamaremos os vértices de G de nós.

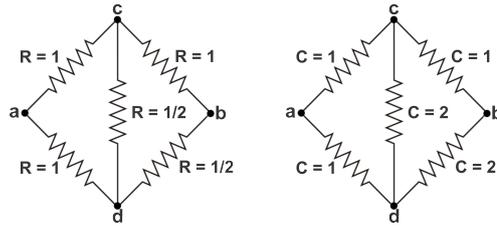


Figura 14 – Temos a mesma rede, mas a direita esta representada com a condutância e a esquerda com a resistência.

Vamos mostrar que podemos associar uma rede a uma cadeia de Markov reversível e, também vale a volta. Considere uma rede $(G, \{C(e)\})$ e defina a matriz de transição como

$$P(x, y) := \begin{cases} \frac{C(x, y)}{C(x)}, & y \sim x \\ 0, & \text{caso contrário} \end{cases}$$

onde $C(x, y) = C(\{x, y\})$, $\{x, y\} \in E$ e $C(x) = \sum_{y: y \sim x} C(x, y)$.

Para termos um melhor entendimento, vamos mostrar como fica a cadeia associada a rede exibida da Figura 14. Neste caso, $C(a) = 2$, $C(b) = 3$, $C(c) = 4$, $C(d) = 5$ e a matriz de transição P associada é

$$P = \begin{pmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{4} & \frac{1}{4} & 0 & \frac{2}{4} \\ \frac{1}{5} & \frac{2}{5} & \frac{2}{5} & 0 \end{pmatrix}.$$

Este processo é chamado de passeio aleatório em G com peso $C(x, y)$. Definindo a distribuição π por

$$\pi(x) := \frac{C(x)}{C_G}$$

onde $C_G = \sum_{x \in V} C(x)$. Assim, na rede do nosso exemplo, temos $\pi = \left(\frac{2}{14} \frac{3}{14} \frac{4}{14} \frac{5}{14}\right)$, podemos também checar que nesse exemplo vale $\pi P = \pi$.

De forma geral, π é uma distribuição estacionária e P é reversível, pois satisfaz a equação de balanço detalhado,

$$\begin{aligned}
\pi(x)P(x, y) &= \frac{C(x)}{C_G} \frac{C(x, y)}{C(x)} = \frac{C(x, y)}{C_G} \\
&= \frac{C(y, x)}{C_G} = \frac{C(y)}{C_G} \frac{C(y, x)}{C(y)} \\
&= \pi(y)P(y, x).
\end{aligned}$$

Assim temos que a distribuição π é estacionária para P . Agora vamos mostrar que podemos associar uma cadeia de Markov reversível com uma rede.

Considere P uma matriz de transição de uma cadeia de Markov irreduzível e reversível no espaço de estado (finito) Ω e π sua distribuição estacionária. Defina os vértices (ou nós) como os estados de Ω e a aresta $\{x, y\}$ quando $P(x, y) > 0$ e pela reversibilidade da cadeia teremos $P(y, x) > 0$. Agora defina a condutância da aresta $\{x, y\}$ como

$$C(x, y) = \pi(x)P(x, y). \quad (5.1)$$

Como a cadeia é reversível então vale que $C(x, y) = C(y, x)$. Por fim, defina os pesos como

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

Desta forma, mostramos que o estudo de cadeias de Markov reversível é equivalente ao estudo de passeio aleatório em Redes.

Exemplo 5.2.2. *Vamos determinar a rede elétrica da cadeia de Markov reversível cuja matriz de transição é dada por*

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{3}{4} & 0 & \frac{1}{4} \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

e a distribuição estacionária é $\pi = (\frac{1}{16} \ \frac{4}{16} \ \frac{6}{16} \ \frac{4}{16} \ \frac{1}{16})$.

A matriz de transição nós mostra toda a dinâmica. Assim sendo, vamos nomear os estados do grafo como sendo $\{a, b, c, d, e\}$. Desta forma não fica difícil de perceber que cada vértice tem dois vizinhos com exceção dos vértices extremos, a saber: a, e . Calculando as condutividades pela fórmula (5.1), temos

$$C(a, b) = \frac{1}{16}, \quad C(b, c) = \frac{3}{16}, \quad C(c, d) = \frac{3}{16}, \quad C(d, e) = \frac{1}{16}.$$

E o esquema elétrico da rede pode ser desenhado como mostra a Figura 15.

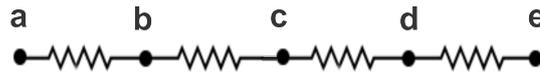


Figura 15 – O esquema elétrico da rede do Exemplo 5.2.2.

5.3 Funções Harmônicas

Iremos assumir ao longo desta seção que a cadeia de Markov é irreduzível e, não necessariamente, a cadeia de Markov será reversível.

Definição 5.3.1. *Uma função $h : \Omega \rightarrow \mathbb{R}$ é dita harmônica com relação a matriz de transição P no ponto $x \in \Omega$ se*

$$h(x) = \sum_{y \in \Omega} P(x, y)h(y). \quad (5.2)$$

Dizemos que a função é harmônica em $D \subset \Omega$ se é harmônica para todo $x \in D$.

Podemos reescrever (5.2) na forma matricial como $Ph = h$, assim as funções harmônicas são aquelas invariantes pela multiplicação de P . Quando P é uma matriz de transição de um passeio aleatório no grafo, dizemos que $h(x)$ é uma média dos valores de h na vizinhança do vértice x .

Definição 5.3.2 (Tempo de Alcance). *Para $x \in \Omega$, definimos **tempo de alcance** como sendo*

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

De forma natural também podemos definir o Tempo de Alcance de um conjunto $B \subset \Omega$ como

$$\tau_B := \min\{t \geq 0 : X_t \in B\}.$$

Informalmente, o tempo de alcance é o primeiro tempo no qual a cadeia visita o estado x . É interessante também o caso de quanto tempo a cadeia demora para visitar o estado $x \in \Omega$ após ter saído do mesmo.

Definição 5.3.3 (Tempo de Primeiro Retorno). *Quando $X_0 = x$ definimos como **tempo de primeiro retorno** como*

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

Conseguimos usar o tempo de alcance para estimar o tempo de mistura desde que colocamos uma hipótese sobre a distribuição estacionária π , como iremos mostrar abaixo. Considere $B \subset \Omega$, $x \in \Omega$ e um tempo $t_0 \in \{1, 2, 3, \dots\}$, assim

$$\begin{aligned} \|P^{t_0}(x, \cdot) - \pi\| &\geq \pi(B) - P^{t_0}(x, B) \\ &\geq \pi(B) - P_x\{\tau_B \leq t_0\} \\ &\geq P_x\{\tau_B > t_0\} + \pi(B) - 1. \end{aligned}$$

Logo, podemos concluir que se $\pi(B) \simeq 1$ e $P_x\{\tau_B > t_0\} \geq \delta$ para algum $\delta > 0$ conseguimos a relação,

$$t_{mix}(\delta) \geq t_0. \quad (5.3)$$

Proposição 5.3.4. *Sejam (X_t) uma cadeia de Markov irredutível com matriz de transição P e $B \subset \Omega$ e $h_B : B \rightarrow \mathbb{R}$. A função*

$$\begin{aligned} h : \Omega &\longrightarrow \mathbb{R} \\ x &\longmapsto E_x h_B(X_{\tau_B}) \end{aligned}$$

é a única extensão de h_B tal que $h(x) = h_B(x)$ se $x \in B$ e h é harmônica para P em todo $x \in \Omega \setminus B$.

Demonstração. Vamos mostrar que $h(x) = E_x h_B(X_{\tau_B})$ é uma extensão de h_B e é harmônica. Sem grandes dificuldades percebemos que $h(x) = h_B(x)$ para todo $x \in B$. Falta mostrar que para $x \in \Omega \setminus B$ a função h é harmônica. Para tanto, tome $x \in \Omega \setminus B$, então

$$h(x) = E_x h_B(X_{\tau_B}) = \sum_y P(x, y) E_x [h_B(X_{\tau_B}) | X_1 = y].$$

Observe que $x \in \Omega \setminus B$ implica que $\tau_B \geq 1$. Pela propriedade forte de Markov podemos escrever

$$\begin{aligned} E_x [h_B(X_{\tau_B}) | X_1 = y] &= E[h_B(X_{\tau_B}) | X_1 = y, X_0 = x] \\ &= E[h_B(X_{\tau_B}) | X_1 = y] \\ &= E_y [h_B(X_{\tau_B})] \\ &= h(y). \end{aligned}$$

Portanto,

$$\begin{aligned} h(x) &= \sum_{y \in \Omega} P(x, y) E_x [h_B(X_{\tau_B}) | X_1 = y] \\ &= \sum_{y \in \Omega} P(x, y) h(y) \end{aligned}$$

logo h é uma função harmônica para todo $x \in \Omega \setminus B$.

Para justificar a unicidade, note que pela Proposição 3.1.1 item (ii), o autoespaço Λ_1 de P associado ao autovetor 1 tem dimensão 1. Como a função harmônica $h \in \Lambda_1$ e $h|_B \equiv h_B$ temos que h é única. □

Proposição 5.3.5 (Princípio do Máximo). *Suponha que P é uma matriz de transição irredutível. Seja $B \subset \Omega$ um conjunto não vazio do espaço de estado e, suponha que $h : \Omega \rightarrow \mathbb{R}$ é uma função não constante e harmônica para todo estado $x \notin B$ e $h(y) = \max_{x \in \Omega} h(x)$. Então $y \in B$.*

Demonstração. Suponha $h(y_0) = \max_{x \in \Omega} h(x)$ e $y_0 \in \Omega \setminus B$. Como h é não constante existe $x_0 \in \Omega$ tal que $h(x_0) < h(y_0)$. Sendo P irredutível existe um caminho y_0, y_1, \dots, y_n com $y_n = x_0$ tal que $P(y_i, y_{i+1}) > 0$ e $i = 0, 1, \dots, n-1$. Seja $k = \min_{1 \leq i \leq n} \{h(y_i) < h(y_0)\}$. Vale:

$$\begin{aligned} h(y_0) = h(y_{k-1}) &= \sum_{z \in \Omega} P(y_{k-1}, z)h(z) \\ &= P(y_{k-1}, y_k)h(y_k) + \sum_{z \in \Omega \setminus \{y_k\}} P(y_{k-1}, z)h(z) \\ &< P(y_{k-1}, y_k)h(y_{k-1}) + (1 - P(y_{k-1}, y_k))h(y_{k-1}) \\ &= h(y_{k-1}) = h(y_0) \end{aligned}$$

Absurdo! Logo $h(y) = \max_{x \in \Omega} h(x)$ implica que $y \in B$. □

5.4 Voltagem e Fluxo de Corrente

A partir de agora iremos introduzir rigorosamente alguns conceitos comuns na linguagem de redes elétricas, sem necessariamente, se basear no comportamento físico das partículas ou elementos que compõem o circuito.

Definição 5.4.1. *Considere uma rede $(G; \{C(e)\})$. Iremos distinguir dois nós $a, z \in \Omega$ e os chamaremos de fonte e solvedouro, respectivamente, da rede. Uma função v que é harmônica em $V \setminus \{a, z\}$ será chamada de voltagem.*

A Proposição 5.3.4 nos garante que conhecendo apenas os valores de $v(a)$ e $v(z)$ podemos determinar completamente a única função voltagem v . Podemos pensar que estes dois pontos da rede são os terminais de uma bateria ou algo similar. Como uma forma de simplificar, vamos escolher a voltagem unitária, ou seja, $v(a) = 1$ e $v(z) = 0$, pois dada qualquer outra voltagem que não seja constante basta fazer uma transformação afim que obtemos a voltagem unitária. Com efeito, dada uma voltagem v_0 tal que $v_0(a) = k_1$ e $v_0(z) = k_2$, queremos encontrar constantes $c_1, c_2 \in \mathbb{R}$ tal que vale

$$v = c_1 v_0 + c_2$$

onde v é a voltagem unitária. Aplicando a equação acima nos nós a e z conseguimos o seguinte sistema

$$\begin{cases} c_1 k_1 + c_2 = 1, \\ c_1 k_2 + c_2 = 0, \end{cases}$$

cujas soluções são $c_1 = \frac{1}{k_1 - k_2}$ e $c_2 = \frac{k_2}{k_2 - k_1}$.

Definição 5.4.2. Uma aresta orientada $\vec{e} = \vec{xy}$ é uma ordenação no par do nós (x, y) . Um fluxo θ é uma função sobre as arestas orientadas com a propriedade de antisimetria, ou seja, $\theta(\vec{xy}) = -\theta(\vec{yx})$. Para o fluxo θ , definimos o divergente de θ em x por

$$\text{div}(\theta(x)) = \sum_{y:y \sim x} \theta(\vec{xy}).$$

Um fato bem simples é que, se somarmos todos os $\text{div} \theta(x)$ com $x \in V$ o resultado é zero, de fato,

$$\sum_{x \in V} \text{div}(\theta(x)) = \sum_{x \in V} \sum_{y:y \sim x} \text{div} \theta(\vec{xy}) = \sum_{\{x,y\} \in E} [\theta(\vec{xy}) + \theta(\vec{yx})] = 0. \quad (5.4)$$

Dizemos que θ é um fluxo de a até z se:

i) **Lei de Kierchhoff para Nós :**

$$\text{div} \theta(x) = 0 \text{ para todo } x \notin \{a, z\}; \quad (5.5)$$

ii) $\text{div} \theta(a) \geq 0$.

Definimos a **intensidade do fluxo** θ de a até z como $\|\theta\| := \text{div} \theta(a)$, observe que (5.4) implica que $\text{div} \theta(a) = -\text{div} \theta(z)$. Até o momento apenas o fluxo foi definido em arestas orientadas. A condutância e a resistência são definidas sem orientação, mas não há problemas em defini-las, para uma conveniente notação futura, com o aspecto de arestas orientadas, assim ficamos com $C(\vec{xy}) = C(\vec{yx}) = C(x, y)$ e $R(\vec{xy}) = R(\vec{yx}) = R(x, y)$.

A nossa próxima definição é motivada por uma importante lei da Física, a Lei de Ohm. Este resultado é uma resposta linear a diferença de voltagem descrita como:

Lei de Ohm: Se $x \sim y$ o fluxo da corrente de x para y satisfaz

$$v(x) - v(y) = i(\vec{xy})R(\vec{xy})$$

onde o fluxo da corrente i é definido a seguir.

Definição 5.4.3. Considere um voltagem v em uma rede, o **fluxo da corrente**, ou simplesmente **corrente**, denotado por i , é definido com orientação nas arestas como

$$i(x, y) = i(\vec{xy}) = \frac{v(x) - v(y)}{R(\vec{xy})} = C(\vec{xy})[v(x) - v(y)].$$

Matematicamente, a Lei de Ohm define a corrente em termos da voltagem, porém na física ou na engenharia a corrente é a taxa de variação das cargas elétricas pelo tempo, ou seja, em um meio condutor é a quantidade de elétrons que estão fluindo por uma unidade de tempo. Esta abordagem utiliza ricamente ferramentas do Cálculo Diferencial.

Se convencionarmos que sempre $v(a) \geq v(z)$ conseguimos mostrar que toda corrente i é também um fluxo de a até z . Sem grandes dificuldades vemos pela definição que i é uma

função antissimétrica. Usando o Princípio do Máximo 5.3.5 temos que $v(a) = \max v(x)$, logo as diferenças $v(a) - v(x) \geq 0$ e com isso resulta que

$$\operatorname{div} i(a) = \sum_{x:x \sim a} i(a, x) = \sum_{x:x \sim a} C(a, x)[v(a) - v(x)] \geq 0.$$

Por fim, basta mostrar que i satisfaz a Lei de Kirchhoff para Nós para todo $x \notin \{a, z\}$. Para tanto, lembrando que $P(x, y) = \frac{C(x, y)}{C(x)}$ e que a voltagem é uma função harmônica no complementar de $\{a, z\}$, então,

$$\begin{aligned} \operatorname{div} i(x) &= \sum_{y:y \sim x} i(\vec{xy}) \\ &= \sum_{y:y \sim x} C(\vec{xy})[v(x) - v(y)] \\ &= C(x)v(x) - \sum_{y:y \sim x} C(x)P(x, y)v(y) \\ &= C(x)v(x) - C(x) \sum_{y:y \sim x} P(x, y)v(y) \\ &= C(x)v(x) - C(x)v(x) = 0. \end{aligned}$$

Agora suponha que o nossa rede seja um caminho fechado, ou seja, o último vértice coincide com o primeiro e não tenha auto-interseção. Se somarmos todas as diferenças de voltagens teremos uma soma telescópica com o resultado igual a zero, este resultado é uma importante lei conhecida como,

Lei de Kirchhoff para Círculo: Se $x_1 \sim x_2 \sim \dots \sim x_n \sim x_{n+1} = x_1$ com $x_1, \dots, x_n \in V$ é um círculo, então, escrevendo $\vec{e}_i = (x_i, x_{i+1})$,

$$\sum_{i=1}^n i(\vec{e}_i)R(\vec{e}_i) = 0. \quad (5.6)$$

Se o fluxo θ de a até z satisfaz a Lei do Círculo 5.6, então temos na verdade uma corrente como veremos no próximo resultado. Desta forma, conseguimos que toda corrente é um fluxo de a até z e, vale a volta se adicionarmos uma condição.

Proposição 5.4.4. *Se θ é uma fluxo de a até z satisfazendo a lei do círculo,*

$$\sum_{i=j}^m R(\vec{e}_j)\theta(\vec{e}_j) = 0$$

para qualquer $\vec{e}_1, \vec{e}_2, \dots, \vec{e}_m$ e se $\|\theta\| = \|i\|$, então $\theta = i$.

Demonstração. Defina a função auxiliar $f := \theta - i$. Vamos mostrar que f satisfaz a Lei do Círculo 5.6 e a Lei de Kirchhoff para Nós 5.5, pois,

$$\sum_{j=1}^m R(\vec{e}_j)f(\vec{e}_j) = \sum_{j=1}^m R(\vec{e}_j)\theta(\vec{e}_j) - \sum_{j=1}^m R(\vec{e}_j)i(\vec{e}_j) = 0 - 0 = 0$$

e para $x \notin \{a, z\}$, vale

$$\operatorname{div} f(x) = \sum_{y:y \sim x} \theta(\vec{xy}) - \sum_{y:y \sim x} i(\vec{xy}) = 0 - 0 = 0.$$

Agora suponha que exista uma aresta $\vec{e}_1 = (x_0, x_1)$ tal que $f(\vec{e}_1) > 0$. Como vale a Lei de Kirchoff para Nó em x_1 , ou seja, $\text{div } f(x_1) = 0$, existe um x_2 tal que $f(\vec{e}_2) > 0$ com $\vec{e}_2 = (x_1, x_2)$. Repetindo o processo e, como o grafo é finito, em algum momento teremos que um vértice irá se repetir, assim teremos uma trajetória fechada, ou seja, um círculo. Reordenando os índices (se necessário) temos um círculo com arestas $\vec{e}_{j_1}, \vec{e}_{j_2}, \dots, \vec{e}_{j_m}$ com $x_{j_1} = x_{j_{m+1}}$. O que é uma contradição com a Lei do Círculo, pois

$$f(\vec{e}_{j_1}) + f(\vec{e}_{j_2}) + \dots + f(\vec{e}_{j_m}) > 0.$$

□

5.5 Resistência Efetiva

Vamos começar esta seção com a definição abaixo.

Definição 5.5.1 (Resistência Efetiva). *A Resistência Efetiva entre os vértices a e z é definida por*

$$\mathcal{R}(a \leftrightarrow z) := \frac{v(a) - v(z)}{\|i\|}.$$

Paralelamente, também definimos a Condutância Efetiva como

$$\mathcal{C}(a \leftrightarrow z) := \frac{1}{\mathcal{R}(a \leftrightarrow z)}.$$

Agora estamos com todos os conceitos e definições necessárias para obter mais uma relação entre probabilidade e redes elétricas. Essa relação é dada pelos próximos resultados.

Lema 5.5.2. *Para quaisquer $x, a, z \in \Omega$ vale*

$$P_x\{\tau_z < \tau_a\} = \frac{v(a) - v(x)}{v(a) - v(z)}.$$

Demonstração. Pela Proposição 5.3.4, a função

$$\begin{aligned} f : \Omega &\longrightarrow [0, 1] \\ x &\longmapsto E_x \mathbf{1}_{\{X_{\tau_{\{a,z\}}} = z\}} = P_x\{\tau_z < \tau_a\} \end{aligned}$$

é a única função harmônica em $\Omega \setminus \{a, z\}$ com $f(a) = 0$ e $f(z) = 1$. Por outro lado, a função

$$\begin{aligned} g : \Omega &\longrightarrow [0, 1] \\ x &\longmapsto \frac{v(a) - v(x)}{v(a) - v(z)} \end{aligned}$$

também é harmônica em $\Omega \setminus \{a, z\}$ e com $g(a) = 0$ e $g(z) = 1$. De fato como a voltagem é uma função harmônica no complementar de $\{a, z\}$ teremos

$$\begin{aligned} g(x) &= \frac{v(a) - v(x)}{v(a) - v(z)} = \frac{\sum_y P(x, y)v(a) - \sum_y P(x, y)v(y)}{v(a) - v(z)} \\ &= \sum_y P(x, y) \frac{[v(a) - v(y)]}{v(a) - v(z)} = \sum_y P(x, y)g(y). \end{aligned}$$

Logo a função g é harmônica no complementar de $\{a, z\}$ e pela unicidade garantida pela Proposição 5.3.4 segue que

$$P_x\{\tau_z < \tau_a\} = \frac{v(a) - v(x)}{v(a) - v(z)}.$$

□

Proposição 5.5.3. Para quaisquer $a, z \in \Omega$ vale

$$P_a\{\tau_z < \tau_a^+\} = \frac{1}{C(a)\mathcal{R}(a \leftrightarrow z)} = \frac{\mathcal{C}(a \leftrightarrow z)}{C(a)}.$$

Demonstração. Usando o Lema 5.5.2 temos,

$$\begin{aligned} P_a\{\tau_z < \tau_a^+\} &= \sum_{x \in V} P(a, x)P_x\{\tau_z < \tau_a\} \\ &= \sum_{x: x \sim a} \frac{C(a, x)}{C(a)} \cdot \frac{v(a) - v(x)}{v(a) - v(z)} \\ &= \frac{\sum_{x: x \sim a} i(\vec{ax})}{C(a)[v(a) - v(z)]} = \frac{\|i\|}{C(a)[v(a) - v(z)]} \\ &= \frac{1}{C(a)\mathcal{R}(a \leftrightarrow z)} = \frac{\mathcal{C}(a \leftrightarrow z)}{C(a)}. \end{aligned}$$

□

Como podemos calcular a resistência efetiva (ou a condutância efetiva)? Queremos que a nossa rede seja equivalente a uma rede mais simples no sentido que podemos transformá-la em uma rede apenas com os nós a e z , é natural pensar que isso seja possível fazendo algumas transformações na rede. Iremos apresentar dois tipos, quando as resistências estão em série e quando estão em paralelo.

I. SÉRIE. Duas resistências R_1 e R_2 em série é equivalente a uma única resistência $R_{eq} = R_1 + R_2$. (Duas condutâncias C_1 e C_2 em série é equivalente a uma única condutância $\frac{1}{C_{eq}} = \frac{1}{C_1} + \frac{1}{C_2}$)

A corrente é a mesma em todos os nós, ou seja, $i(a, b) = i(b, z)$. Com efeito, como a voltagem é harmônica e estamos supondo a voltagem unitária podemos escrever

$$v(b) = P(b, a)v(a) + P(b, z)v(z) = \frac{C(b, a)}{C(b)} = \frac{R_2}{R_1 + R_2}$$

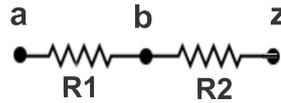


Figura 16 – Associação em série

logo, usando a definição podemos calcular as correntes

$$i(a, b) = \frac{v(a) - v(b)}{R(a, b)} = \frac{1 - \frac{R_2}{R_1 + R_2}}{R_1} = \frac{1}{R_1 + R_2}$$

e

$$i(b, z) = \frac{v(b) - v(z)}{R(b, z)} = \frac{\frac{R_2}{R_1 + R_2}}{R_2} = \frac{1}{R_1 + R_2}$$

como queríamos mostrar.

Assim temos que $v(a) - v(b) = iR_1$ e $v(b) - v(z) = iR_2$, daí somando as duas igualdades resulta que $v(a) - v(z) = i(R_1 + R_2)$. Por outro lado, gostaríamos que $v(a) - v(z) = iR_{eq}$, logo vale que

$$R_{eq} = R_1 + R_2.$$

Este processo pode ser repetido para qualquer quantidade finita, assim se R_1, R_2, \dots, R_N são resistências em série então podemos escrever que

$$R_{eq} = \sum_{n=1}^N R_n.$$

I. PARALELO. Duas resistências R_1 e R_2 em paralelo é equivalente a uma única resistência $\frac{1}{R_{eq}} = \frac{1}{R_1} + \frac{1}{R_2}$. (Duas condutâncias C_1 e C_2 em paralelo é equivalente a uma única condutância $C_{eq} = C_1 + C_2$)

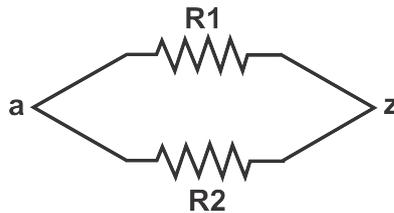


Figura 17 – Associação em paralelo

Neste caso, basta olhar como grafo, ou seja, temos duas arestas ligando os mesmo dois pontos, portanto a condutância equivalente é simplesmente a soma. logo vale que

$$\frac{1}{R_{eq}} = \frac{1}{R_1} + \frac{1}{R_2}.$$

Este processo pode ser repetido para qualquer quantidade finita, assim se R_1, R_2, \dots, R_N são resistências em paralelo então podemos escrever que

$$\frac{1}{R_{eq}} = \sum_{n=1}^N \frac{1}{R_n}.$$

Observe que a rigor a nossa definição de grafo não permite aresta em paralelo.

Exemplo 5.5.4. *Suponha que todas as arestas da rede seguinte tenha resistência iguais. Qual é o valor de $P_a\{\tau_z < \tau_a^+\}$? Seguindo as transformações indicadas na figura, obtemos que $\mathcal{R}(a \leftrightarrow z) = 3/2$, donde*

$$P_a\{\tau_z < \tau_a^+\} = \frac{1}{C(a)\mathcal{R}(a \leftrightarrow z)} = \frac{1}{3}.$$

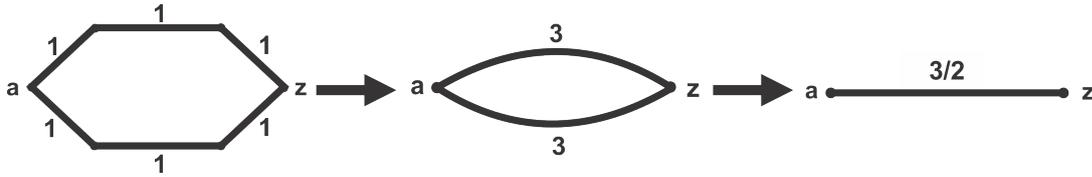


Figura 18 – As transformações para reduzir a rede

Exemplo 5.5.5. *Fixe $\alpha > 0$ com $\alpha \neq 1$ e considere os vértices $\{0, 1, 2, \dots, n\}$ com as condutâncias em série com valor de $C(k-1, k) = \alpha^k$ para $k = 1, 2, \dots, n$.*

Temos que $P(0, 1) = P(n, n-1) = 1$ e

$$P(k, k+1) = \frac{C(k, k+1)}{C(k)} = \frac{\alpha^{k+1}}{\alpha^k + \alpha^{k+1}} = \frac{\alpha}{1 + \alpha} := p;$$

$$P(k, k-1) = \frac{C(k, k-1)}{C(k)} = \frac{\alpha^k}{\alpha^k + \alpha^{k+1}} = \frac{1}{1 + \alpha} := 1 - p.$$

Vamos reduzir para uma rede equivalente mais simples, como mostra a Figura 19. Vamos calcular as resistências R_1 e R_2 como uma soma de uma PG finita.

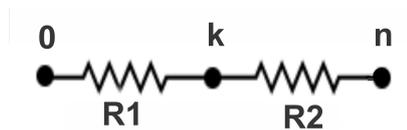


Figura 19 – As transformações para reduzir a rede

$R_1 = \sum_{j=1}^k \frac{1}{\alpha^j} = \frac{1 - \alpha^{-k}}{\alpha - 1}$; e $R_2 = \sum_{j=1}^n \frac{1}{\alpha^j} - R_1 = \frac{1 - \alpha^{-n}}{\alpha - 1} - \frac{1 - \alpha^{-k}}{\alpha - 1} = \frac{\alpha^{-k} - \alpha^{-n}}{\alpha - 1}$.
Como vale que

$$\begin{aligned} P_k\{\tau_n < \tau_0\} &= \frac{v(0) - v(k)}{v(0) - v(n)} = \frac{R_1}{R_1 + R_2} \\ &= \frac{\alpha^{-k} - 1}{\alpha^{-n} - 1}. \end{aligned}$$

Em particular, para um passeio aleatório tendencioso para se mover para a direita (ou qualquer outra direção) com probabilidade p

$$P_k\{\tau_n < \tau_0\} = \frac{\left[\frac{1-p}{p}\right]^k - 1}{\left[\frac{1-p}{p}\right]^n - 1}.$$

Os próximos resultados e definições tem o intuito de dar mais formas de calcular a Resistência efetiva ou ao menos uma boa estimativa.

Definição 5.5.6 (Energia). *A energia do fluxo θ é dada por*

$$\mathcal{E}(\theta) := \sum_e [\theta(e)]^2 R(e).$$

Teorema 5.5.7 (Princípio de Thomson). *Para qualquer grafo conexo finito, vale*

$$\mathcal{R}(a \leftrightarrow z) = \inf\{\mathcal{E}(\theta) : \theta \text{ é um fluxo unitária de } a \text{ para } z\}.$$

O único fluxo no qual o ínfimo é atingido é o fluxo de corrente unitária.

Demonstração. Considere C o conjunto das fluxos unitárias. Podemos interpretar C como um subconjunto de \mathbb{R}^n , com $n = |E|$. Com efeito, como o grafo é finito podemos enumerar as arestas como e_1, e_2, \dots, e_n e podemos escrever cada elemento θ de C como $(\theta(\vec{e}_1), \theta(\vec{e}_2), \dots, \theta(\vec{e}_n))$.

Sem perda de generalidade, podemos considerar os fluxos que $|\theta(e)| \leq M$ para todo $e \in E$ e algum $M \in \mathbb{R}$, pois nosso interesse é tomar o ínfimo. Com isso, conseguimos, usando a norma do máximo que $C \subset B[0, M]$, a bola fechada de \mathbb{R}^n centrada na origem e raio M .

Agora tome (θ_n) uma sequência de fluxos unitários de a até z em $B[0, M]$ tal que

$$\lim_{n \rightarrow \infty} \mathcal{E}(\theta_n) = \inf_{\theta} \mathcal{E}(\theta).$$

Sendo a bola compacta, existe uma subsequência (θ_{n_k}) que converge para algum $\theta^* \in B[0, M]$. Mas, por continuidade da função soma temos que $\theta^* \in C$ e θ^* é o fluxo de a até z que realiza o ínfimo da energia.

A unicidade de θ é garantida pela Lei do Círculo 5.4.4. Dado um ciclo cuja arestas são $\vec{e}_{i_1}, \vec{e}_{i_2}, \dots, \vec{e}_{i_k}$ defina a função auxiliar γ antissimétrica descrita como:

$$\gamma(\vec{e}) = \begin{cases} 1, & \text{se } \vec{e} = \vec{e}_{i_j} \text{ } 1 \leq j \leq k. \\ 0, & \text{caso contrário.} \end{cases}$$

Temos que γ satisfaz a Lei dos Nós, de fato,

$$\operatorname{div} \gamma(x_i) = \sum_{y: y \sim x_i} \gamma(x_i, y) = \gamma(x_i, x_{i+1}) + \gamma(x_i, x_{i-1}) = \gamma(x_i, x_{i+1}) - \gamma(x_{i-1}, x_i) = 0$$

e sem dificuldades, vemos que $\operatorname{div} \gamma(a) = 0$, portanto γ é um fluxo de a até z , mas $\sum_{j=1}^k \gamma(\vec{e}_{i_j}) = n \neq 0$. Para qualquer $\delta \in \mathbb{R}$ e como o fluxo θ é o ínfimo teremos que

$$\begin{aligned} 0 &\leq \mathcal{E}(\theta + \delta\gamma) - \mathcal{E}(\theta) \\ &= \sum_{j=1}^k \left[\theta(\vec{e}_{i_j}) + \delta\gamma(\vec{e}_{i_j}) \right]^2 R(\vec{e}_{i_j}) - \sum_{j=1}^k \theta^2(\vec{e}_{i_j}) R(\vec{e}_{i_j}) \\ &= 2\delta \sum_{j=1}^n \theta(\vec{e}_{i_j}) R(\vec{e}_{i_j}) + \delta^2 \sum_{j=1}^k R(\vec{e}_{i_j}). \end{aligned}$$

Escrevendo $l := \sum_{j=1}^n R(\vec{e}_j)$ ficamos com a expressão

$$0 \leq 2\delta \sum_{j=1}^n \theta(\vec{e}_j) R(\vec{e}_j) + l\delta^2. \quad (5.7)$$

Agora tomando $\delta > 0$ e dividindo ambos os lados em (5.7)

$$0 \leq 2 \sum_{j=1}^k \theta(\vec{e}_{i_j}) R(\vec{e}_{i_j}) + l\delta$$

e fazendo $\delta \downarrow 0$ implica

$$0 \leq \sum_{j=1}^k \theta(\vec{e}_{i_j}) R(\vec{e}_{i_j}).$$

Analogamente, tomando $\delta < 0$ e dividindo ambos os lados em (5.7) e fazendo $\delta \uparrow 0$ teremos

$$0 \geq \sum_{j=1}^k \theta(\vec{e}_{i_j}) R(\vec{e}_{i_j}).$$

Portanto, $\sum_{j=1}^k \theta(\vec{e}_{i_j}) R(\vec{e}_{i_j}) = 0$ verificando que θ satisfaz a Lei do Circulo 5.4.4. Agora basta mostrar que o fluxo de corrente unitária i satisfaz $\mathcal{E}(i) = \mathcal{R}(a \leftrightarrow z)$, com efeito

$$\begin{aligned} \sum_e R(e)[i(e)]^2 &= \frac{1}{2} \sum_x \sum_y R(x, y) \left[\frac{v(x) - v(y)}{R(x, y)} \right]^2 \\ &= \frac{1}{2} \sum_x \sum_y C(x, y) [v(x) - v(y)]^2 \\ &= \frac{1}{2} \sum_x \sum_y i(\vec{x}\vec{y}) [v(x) - v(y)]. \end{aligned}$$

Observe que usando a Lei de Nós temos que $\sum_y i(\vec{x}\vec{y})v(y) = 0$ se $x \notin \{a, z\}$ daí

$$\begin{aligned} \frac{1}{2} \sum_x \sum_y i(\vec{x}\vec{y})v(x) &= \frac{1}{2} \left[\sum_y i(\vec{a}\vec{y})v(a) + \sum_y i(\vec{z}\vec{y})v(z) \right] \\ &= \frac{1}{2} [\|i\|v(a) - \|i\|v(z)]. \end{aligned}$$

E como $-\frac{1}{2} \sum_x \sum_y i(\vec{x}\vec{y})v(y) = \frac{1}{2} \sum_y \sum_x i(\vec{y}\vec{x})v(y)$ e usando o argumento anterior teremos

$$\sum_e R(e)[i(e)]^2 = \|i\|[v(a) - v(z)] = \mathcal{R}(a \leftrightarrow z)$$

onde usamos que $\|i\| = 1$.

□

Se $R = \{R(e) : e \in E' \subset E\}$ é um subconjunto do conjunto das resistências nas arestas do grafo G , escrevemos $\mathcal{R}(a \leftrightarrow z; R)$ para denotar a resistência efetiva nesse subconjunto.

Teorema 5.5.8 (Lei de Monotonicidade de Rayleigh). *Se $R = \{R(e)\}$ e $R' = \{R'(e)\}$ são conjuntos de resistências nas arestas do mesmo grafo G e se $R(e) \leq R'(e)$ para todo $e \in E$, então*

$$\mathcal{R}(a \leftrightarrow z, R) \leq \mathcal{R}(a \leftrightarrow z, R').$$

Demonstração. Este resultado segue diretamente do Princípio de Thomson 5.5.7, pois como $R(e) \leq R'(e)$, então

$$\begin{aligned} \inf_{\theta} \sum_e [\theta(e)]^2 R(e) &\leq \inf_{\theta} \sum_e [\theta(e)]^2 R'(e) \\ \mathcal{R}(a \leftrightarrow z, R) &\leq \mathcal{R}(a \leftrightarrow z, R'). \end{aligned}$$

□

Uma técnica dada por Nash-Williams nós dá uma simples, porém útil, limitação inferior da resistência efetiva. Chamaremos de $\Pi \subset V$ como *corte por aresta* separando a de z se qualquer caminho de a para z contém alguma aresta (x, y) em Π .

Proposição 5.5.9. *Se $\{\Pi_k\}$ são edge cutsets disjuntos com nós separados a e z , então*

$$\mathcal{R}(a \leftrightarrow z) \geq \sum_k \left(\sum_{e \in \Pi_k} C(e) \right)^{-1}.$$

A desigualdade acima é chamada de desigualdade de Nash-Williams.

Demonstração. Seja θ um fluxo unitário de a até z . Para cada k , usando a desigualdade de Cauchy-Schwarz, teremos

$$\begin{aligned} \sum_{e \in \Pi_k} C(e) \sum_{e \in \Pi_k} R(e) \theta^2(e) &\geq \left(\sum_{e \in \Pi_k} \sqrt{C(e)} \sqrt{R(e)} |\theta(e)| \right)^2 \\ &= \left(\sum_{e \in \Pi_k} |\theta(e)| \right)^2. \end{aligned}$$

Mais ainda,

$$\begin{aligned} \sum R(e) \theta^2(e) &\geq \sum_k \sum_{e \in \Pi_k} R(e) \theta^2(e) \\ &\geq \sum_k \left(\sum_{e \in \Pi_k} C(e) \right)^{-1}. \end{aligned}$$

o resultado segue aplicando o Princípio de Thomson 5.5.7.

□

5.6 Robustez

Esta seção é baseada no artigo [6]. Um fato natural de se pensar é, dada uma rede $(G, \{C(e)\})$ o que acontece, se houverem algumas perturbações na taxas da condutância, como isso vai influenciar na rede como um todo? Como é a sensibilidade do tempo de mistura quando multiplicamos as condutâncias por um fator limitado, ou seja, se modificarmos as condutâncias como vai se comportar a convergência. Desta forma convém definir.

Definição 5.6.1 (Robustez). *Dizemos que uma família de redes $(G_n, \{C_n(e)\})$ é robusta se para toda constante $C > 0$ existe uma constante $K > 0$ tal que se multiplicarmos as condutâncias por um fator de até $C > 0$, os correspondentes tempo de mistura são preservados por um fator até $K > 0$.*

Iremos denotar por $t_{mix}(G, v) = \min\{t : \|P^t(v, \cdot) - \pi(\cdot)\|_{VT} \leq \frac{1}{4}\}$ e chamaremos de tempo de mistura do grafo G por

$$t_{mix}(G) = \max_{v \in V} t_{mix}(G, v).$$

O próximo teorema é o resultado chave dessa seção, nele garante que existe uma família de redes que não são robustas, ou seja, existem redes nas quais quando perturbamos as condutâncias perdemos o controle do tempo de mistura.

Teorema 5.6.2. *Existe uma família de redes uniformemente limitado que não são robustas. Mais ainda, existe uma sequência de grafos $\{G_n\}$ com grau máximo 10 e $|V(G_n)| = n$ bem como uma regra para mudar a condutância até um fator de 2, tal que o tempo de mistura em G_n mudará por um fator de pelo menos $\frac{c \log n}{\log \log n}$ onde c é uma constante.*

Construção

Vamos construir uma família de grafos com fator que atinja $\frac{\log n}{\log \log n}$ como mencionado no Teorema 5.6.2.

Considere uma árvore binária T que começa em O e tem altura K (estamos supondo que K é suficientemente grande), como a nossa árvore é binária, ou seja, de cada vértice sai duas arestas. Vamos chamá-las de direita e esquerda, assim definimos como R o conjunto de todos os pontos a direita e, analogamente, como L todos vértices a esquerda. Definimos o conjunto H_i como os pontos da i -ésima linha da árvore T , logo $|H_i| = 2^{i-1}$ (considerando como cardinalidade o número de pontos de vértices do conjunto), mais ainda, tome $l := 100 \log K$, e escrevemos

$$\mathcal{H} := \bigcup_{j=K/4l}^{K/2l} H_{jl}.$$

Provavelmente não teremos que l seja inteiro, mas pense como um inteiro mais próximo. O conjunto \mathcal{H} tem como extremo as linhas $H_{\frac{K}{4}}$ e $H_{\frac{K}{2}}$ e todas as linhas intermediárias que tem como índice um múltiplo (inteiro) de l , portanto \mathcal{H} não é um conjunto muito grande visto que l não é um número pequeno.

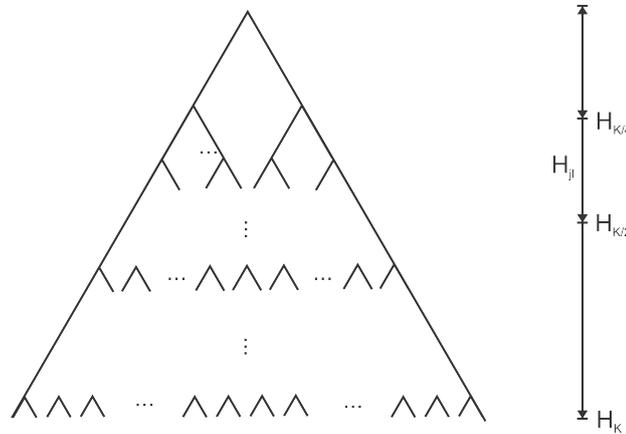


Figura 20 – Passeio aleatório no 4-ciclo

Para cada $v \in \mathcal{H}$, defina T_v como a subárvore binária de T que começa em v e tem altura l , sem dificuldade vemos que $T_v \cap T_w = \emptyset$ desde que $v \neq w$. Também defina $\Gamma_{v,w}$ como o único caminho que liga os pontos v e w e determinamos

$$A_v := \{w \in T_v : l/4 \leq |\Gamma_{v,w}| \leq l/2, \quad ||\Gamma_{v,w} \cap L| \cap |\Gamma_{v,w} \cap R|| \leq \sqrt{l}\}$$

como sendo a coleção de vértices equilibrados na subárvore T_v .

A primeira condição do conjunto garante que os vértices equilibrados se encontrem no segundo quartil de T_v e a segunda condição nos assegura que a quantidade de vezes que se anda para a direita e esquerda sejam próxima, não ultrapassando \sqrt{l} .

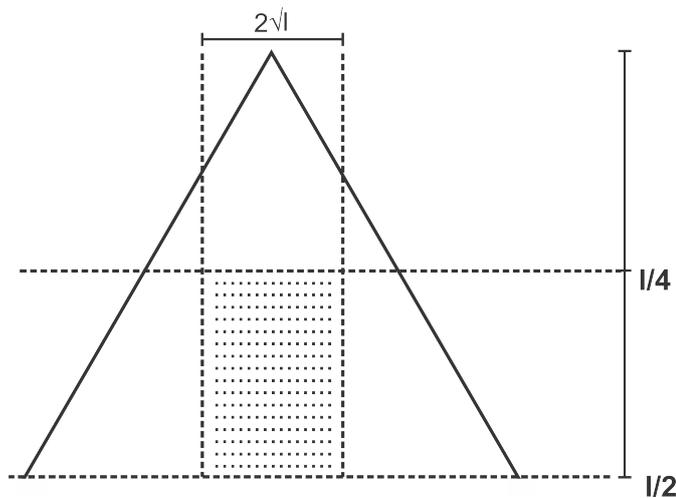


Figura 21 – Passeio aleatório no 4-ciclo

Denote por $\mathcal{A} := \cup_{v \in \mathcal{H}} A_v$. Para cada $u \in \mathcal{A}$, vamos anexar um 3D toro B_u de volume K , ou seja, o conjunto \mathbb{Z}_n^3 com $n = \sqrt[3]{K}$. Por fim, escrevemos $G^* = (V^*, E^*)$ como um

grafo 3-regular de tamanho $K^2 2^K$ e vamos anexá-lo na árvore T no nível H_K . Vamos escrever $G = (V, E)$ como o grafo final obtido pela construção.

Agora vamos mostrar que $|V| = (1 + O(1/K))K^2 2^K$. De fato, os vértices de T é uma simples soma de uma P.G. de razão 2, dando a quantidade de $2^K - 1$ vértices e $|V^*| = K^2 2^K$. Já a quantidade de vértices foi anexado um toro B_u pode ser pensando como: em cada subárvore T_v temos $\frac{l}{4} 2\sqrt{l}$ vértices equilibrados e o total de árvores é, aproximadamente, $2^{\frac{3K}{4}-l}$ que é a soma de uma P.G. finita de razão 2^l e como o volume do toro é K , temos o seguinte total de vértices

$$\begin{aligned} |V| &= 2^K - 1 + K^2 2^K + \frac{l^{3/4}}{2} \cdot 2^{(3K/4)-l} \cdot K \\ &= K^2 2^K \left(1 + \frac{1}{K^2} - \frac{1}{K^2 2^K} + \frac{l^{3/2} \cdot 2^{-(K/4)-l-1}}{K} \right) \end{aligned}$$

como K e l são fixos não é difícil perceber que podemos escrever $|V| = (1 + O(1/K))K^2 2^K$.

Agora vamos modificar as condutâncias de G com a regra abaixo

$$C(u, v) = \begin{cases} 2, & (u, v) \in T \text{ com } u \text{ parente de } v \text{ e } v \in L \\ 1, & \text{caso contrário} \end{cases}$$

e denotaremos $\tilde{G} = (\tilde{V}, \tilde{E})$ a rede com essa nova condutância. Com os próximos resultados vamos verificar que o tempo de mistura de G e \tilde{G} diferem por uma fator desejado.

Lema 5.6.3. *O tempo de mistura em G satisfaz $t_{mix}(G) \geq cK^2$ para uma constante $c > 0$.*

Demonstração. Seja τ o tempo de alcance do conjunto V^* para um passeio aleatório (S_t) em G . Primeiro vamos limitar $E_o \tau$ inferiormente. Denote por G' o grafo G sem ter sido anexado o toro 3D.

Primeiro considere o passeio aleatório (S'_t) em G' . Denote por N_i como o número de vezes que (S'_t) visita $H_i \cap \mathcal{A}$ antes de alcançar o conjunto V^* e denote $\mathcal{N} = \sum_i N_i$. Sem dúvida (S'_t) visita ao menos um vértice de H_i antes de atingir H_k .

Agora vamos analisar $|H_i \cap \mathcal{A}|$. Para alguns $i = 1, 2, \dots, k$ é possível que $|H_i \cap \mathcal{A}| = \emptyset$. Mas vamos avaliar os casos que

$$0 < 1 \leq |H_i \cap \mathcal{A}| \leq |H_i| = 2^{i-1}$$

logo,

$$\frac{1}{2^{i-1}} \leq \frac{|H_i \cap \mathcal{A}|}{|H_i|} \leq 1$$

e como $\frac{1}{2^{K-1}} \leq \frac{1}{2^{i-1}}$ então tomando $c = \frac{1}{2^{K-1}} > 0$ teremos que $\frac{|H_i \cap \mathcal{A}|}{|H_i|} \geq c \geq 0$. Assim,

$$E_o \mathcal{N} = \sum_{i=1}^K E_o N_i \geq \sum_{i=1}^K \frac{|H_i \cap \mathcal{A}|}{|H_i|} \geq cK.$$

Não é difícil de perceber que qualquer (S_t) pode ser decomposto por (S'_t) e por um passeio aleatório no toro 3D anexo em G .

Recorde do fato que a esperança do tempo de retorno para a origem de um passeio aleatório é o volume total normalizado pelo grau da origem. Assim todo tempo de (S_t) para visitar o estado $v \in \mathcal{A}$, o tempo médio que leva o passeio aleatório para um passeio em B_v é no mínimo $K/2$. Portanto conseguimos,

$$E_o\tau \geq E_o\mathcal{N} \frac{K}{2} \geq \frac{cK^2}{2}$$

Tome v^* tal que $E_{v^*}\tau := \max_{v \in V} E_v\tau$ e usando a propriedade de Markov conseguimos que

$$\begin{aligned} E_{v^*}\tau &= E_{v^*} \left(\tau \mid \tau < \frac{cK^2}{20} \right) + E_{v^*} \left(\tau \mid \tau \geq \frac{cK^2}{20} \right) \\ &\leq \frac{cK^2}{20} + E_{v^*} \left(\tau \mid \tau \geq \frac{cK^2}{20} \right) \\ &\leq \frac{cK^2}{20} + P_{v^*} \left\{ \tau \geq \frac{cK^2}{20} \right\} E_{v^*}\tau. \end{aligned}$$

Disto temos que

$$P_{v^*} \left\{ \tau \geq \frac{cK^2}{20} \right\} \geq \frac{3}{4}.$$

Como a distribuição estacionária em V^* satisfaz $\pi(V^*) \geq 1 - 1/K \geq 9/10$, obtemos que $t_{mix}(G, v^*) \geq cK^2/20$, como queríamos. \square

Lema 5.6.4. *O tempo de mistura em \tilde{G} satisfaz $t_{mix}(\tilde{G}) = O(K \log K)$.*

Demonstração. Para provar esse resultado é necessário o Princípio dos Grandes Desvios o que sai do escopo desse trabalho, por isso vamos deixar apenas a referência [6]. \square

Demonstração do Teorema da Robustez 5.6.2. Vamos chamar de n o número de elementos do nosso grafo, ou seja,

$$n = |V| = (1 + O(1/K))K^2 2^K < 2^K \cdot 2^K$$

portanto temos que $\frac{1}{2} \log n < K$. Juntando os dois lemas anteriores, conseguimos

$$\frac{t_{mix}(\tilde{G})}{t_{mix}(G)} \leq \frac{K \log K}{cK^2} = \frac{\log K}{cK} < \frac{\log \log n}{\tilde{c} \log n},$$

donde segue o resultado. \square

Referências

- [1] Aldous, David; Fill, J. *Reversible Markov Chains and Random Walks on Graphs*. Em elaboração, disponível em <http://www.stat.berkeley.edu/~aldous/RWG/book.html>
- [2] Doyle, Peter G.; Snell, J. Laurie. *Random walks and electrical networks*. 2000. arXiv:math/0001057v1
- [3] Levin, David A.; Peres, Yuval; Wilmer, Elizabeth L. *Markov Chains and Mixing Times*, 1ª edição, Providence: American Mathematical Society, 2001
- [4] Lyons, Russell; Peres, Yuval. *Probability on Trees and Networks*. Em elaboração, disponível em <https://www.uni-due.de/~hm0110/book.pdf>
- [5] Norris, J. R.; *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. 1ª edição, 1997
- [6] Peres, Yuval; Ding, Jian. *Sensitivity of mixing times*. Electron Commun. Probab. 18 (2013), 1-6.
- [7] Peres, Yuval; Basu, Riddhipratim; Hermon, Jonathan. *Characterization of cutoff for reversible Markov chains*. 2015. arXiv:1409.3250v4.