

Universidade Federal do Rio Grande do Sul
Instituto de Matemática e Estatística
Programa de Pós-Graduação em Matemática

**Tempos de Mistura de Cadeias de Markov:
Relações com outros indicadores
e estimativas para o passeio aleatório no toro**

Dissertação de Mestrado

Humberto de Lima

Porto Alegre, Abril de 2024.

Dissertação submetida por Humberto de Lima como requisito parcial para a obtenção do grau de Mestre em Matemática pelo Programa de Pós-Graduação em Matemática do Instituto de Matemática e Estatística da Universidade Federal do Rio Grande do Sul.

Professor Orientador:

Ricardo Misturini (UFRGS)

Professora Coorientadora:

Susana Frómeta Fernández (UFRGS)

Banca Examinadora:

Adriana Neumann de Oliveira (UFRGS)

Marco Aymone (UFMG)

Rodrigo Marinho (UFSM)

19 de abril de 2024

Resumo

Esta dissertação explora conceitos fundamentais das cadeias de Markov, abordando sua construção e conceitos básicos, passando por resultados clássicos como o teorema de convergência, e dedicando-se especialmente ao estudo de indicadores relacionados à cadeia, em especial o tempo de mistura, o tempo de relaxamento e o tempo de acerto. O objetivo deste trabalho é compilar e estruturar diferentes resultados na área, alguns bastante recentes, explorando as relações entre esses indicadores, e utilizando o exemplo do passeio aleatório no toro para a visualização dessas grandezas e melhor compreensão de suas propriedades. Além disso, busca fornecer um material ao mesmo tempo abrangente e de fácil entendimento para os leitores, numa área relativamente recente e com grandes avanços na última década.

Palavras-chave: Cadeias de Markov; Indicadores de cadeias; Tempo de mistura; Tempo de relaxamento; Tempo de acerto.

Abstract

This dissertation explores fundamental concepts of Markov chains, addressing their construction and basic concepts, covering classical results such as the convergence theorem, and especially focusing on the study of indicators related to the chain, particularly mixing time, relaxation time, and hitting time. The objective of this work is to compile and structure different results in the field, some quite recent, exploring the relationships between these indicators, and using the example of the random walk on the torus to visualize these quantities and better understand their properties. Additionally, it aims to provide material that is both comprehensive and easily understandable for readers in a relatively recent area with significant advancements in the past decade.

Keywords: Markov chains; Chain indicators; Mixing time; Relaxation time; Hitting time.

Agradecimentos

Agradeço primeiramente à minha mãe, que sempre me apoiou incondicionalmente e nunca mediu esforços para que eu chegasse onde cheguei. Ela sabe o quanto a valorizo por isso. Não posso deixar de agradecer também à minha família, especialmente aos meus tios Paulo, Vanda e There, e aos meus primos Marcelo, Maurício, Patrícia e Gabriele, que sempre foram base e inspiração para mim.

Agradeço aos amigos que fiz na graduação e no mestrado, especialmente Mikaela, Andressa e Alessandra, que estiveram ao meu lado nos momentos de maior questionamento e foram essenciais para que eu persistisse e alcançasse este objetivo.

Ao meu orientador Ricardo e à minha coorientadora Susana, agradeço pela paciência e pelo apoio ao longo deste período. Também agradeço ao meu primeiro orientador, Álvaro, que, mesmo após minha mudança de área, continuou sendo uma valiosa fonte de conselhos e boas conversas.

Não posso deixar de agradecer aos meus amigos de infância, com quem tive a sorte de conviver diariamente até a faculdade e que, mesmo com a distância, continuam sendo uma parte importante da minha vida. Vocês desempenharam um papel crucial no meu crescimento pessoal.

Finalmente, à minha namorada Aline, que, de tudo o que a UFRGS me proporcionou, é certamente o mais importante. Agradeço não só aqui, mas diariamente, por termos nos conhecido e por combinarmos tão bem. Também agradeço por me entender e me aturar por todo esse tempo.

Sumário

Introdução	7
1 Cadeias de Markov: Definições Básicas	9
1.1 Definição de um cadeia de Markov	9
1.2 Existência da cadeia de Markov	10
1.3 Propriedades fraca e forte de Markov	12
1.4 Irredutibilidade e estrutura de classes	13
1.5 Tempo de acerto	14
1.6 Recorrência e transiência	16
1.7 Distribuição estacionária	17
1.7.1 No caso S finito	20
1.8 Reversibilidade	20
1.9 Passeios aleatórios em grafos	21
1.10 Representação via sistema dinâmico aleatório	22
1.11 Cadeias lazy	23
2 Teorema de Convergência e Tempos de Mistura	25
2.1 Distância de variação total	25
2.2 Acoplamentos	27
2.3 Teorema de convergência	32
2.4 Tempo de mistura	37
2.5 Tempo de mistura considerando outras distâncias	40
3 Passeios Aleatórios em Redes	44
3.1 Funções harmônicas	44
3.2 Redes elétricas e o passeio aleatório associado	47
3.3 Voltagem e fluxo de corrente	50
3.4 Resistência efetiva	54
3.5 Simplificando redes	57
3.5.1 Redes em multigrafos	58
3.5.2 Simplificação de redes	59
3.6 Energia	67
3.7 Conjunto separador entre a e z	70
4 Outros Indicadores das Cadeias de Markov	73
4.1 Tempo do alvo aleatório	73
4.2 Tempo de acerto	75
4.3 Tempo de comutação	76
4.4 Espectro	80
4.5 Tempo de relaxamento	84

5	Relações Entre os Indicadores	86
5.1	Relação entre t_{mix} e t_{hit}	86
5.2	Relações entre t_{mix} e t_{rel}	89
5.3	Relação entre t_{\odot} e o espectro	93
5.4	Relações ente t_{hit} , t_{rel} e probabilidades de retorno	94
5.4.1	Resultados para o passeio aleatório lazy em um grafo	94
5.4.2	Resultados em um contexto mais geral: reversibilidade	105
6	Exemplos Clássicos e o Passeio Aleatório no Toro	110
6.1	Construção dos exemplos	110
6.1.1	Ruína do jogador	110
6.1.2	Urna de Pólya	111
6.1.3	Passeio aleatório no n -ciclo	114
6.1.4	Passeio no toro \mathbb{Z}_n^d	114
6.2	Tempo de mistura no toro	115
6.2.1	t_{mix} no n -ciclo \mathbb{Z}_n	115
6.2.2	t_{mix} no toro \mathbb{Z}_n^d	117
6.3	Tempo de relaxamento no toro	119
6.3.1	t_{rel} no n -ciclo \mathbb{Z}_n	119
6.3.2	t_{rel} no toro \mathbb{Z}_n^d	121
6.4	Tempo de acerto no toro	121
6.4.1	Cotas inferiores para o caso multidimensional	125
6.4.2	Cotas superiores para o caso multidimensional com $d \geq 3$	127
6.4.3	Cotas superiores para o caso multidimensional com $d = 2$	130
A	Pequeno Resumo de Alguns Resultados	133

Introdução

O conceito de Cadeias de Markov foi introduzido no início do século XX pelo matemático russo Andrey Markov [2]. Informalmente, uma cadeia de Markov é um processo estocástico com perda de memória, de modo que estando em um estado x no tempo t , o estado ocupado no tempo $t + 1$ é escolhido de acordo com uma distribuição de probabilidade que depende unicamente de x , sendo independente da sequência de estados que o precederam. Conforme ficará claro na Definição 1.2 toda a dinâmica da cadeia pode ser descrita por uma matriz de transição P .

Uma questão clássica neste contexto é estudar a evolução no tempo de uma determinada cadeia e compreender a velocidade de convergência para uma distribuição estacionária quando $t \rightarrow \infty$. Nas últimas décadas cresceu o interesse por cadeias com espaços de estados de cardinalidade finita mas muito grande, surgindo assim uma nova abordagem assintótica: uma distância específica ε é fixada e o número de passos requeridos para a cadeia estar ε próxima da distribuição estacionária é chamado *tempo de mistura* (t_{mix}). O objetivo é entender como o tempo de mistura cresce como função do tamanho do espaço de estados. Tal ferramenta tem-se mostrado de grande valia especialmente em ciência da computação e em física estatística visto que em diversos algoritmos de simulação, amostragem ou contagem aproximada, o tempo de mistura relaciona-se diretamente com o tempo de execução do algoritmo.

É natural nos referirmos ao tempo de mistura como um *indicador* da cadeia. Neste trabalho apresentaremos mais alguns indicadores como o tempo de acerto da cadeia t_{hit} e o tempo de relaxamento t_{rel} , além de explorarmos as relações entre eles. Intuitivamente, t_{hit} é o número médio de passos necessários para atingir um estado específico da cadeia. Já o t_{rel} é definido em termos do espectro da matriz de transição e está intimamente relacionado com a velocidade exponencial em t da convergência para a medida estacionária. Vamos trabalhar quase exclusivamente com espaços de estados finitos, uma vez que o principal interesse é analisar o comportamento desses indicadores em relação ao tamanho do espaço de estados. Apesar disso, no primeiro capítulo definiremos e estudaremos as propriedades de cadeias de Markov em espaço de estados S enumeráveis. Fizemos tal escolha por três motivos: primeiro porque o caso S enumerável cobre o caso S finito, segundo porque o trabalho adicional para tal construção é bem pequeno, e terceiro porque poderemos aproveitar alguns resultados clássicos enunciados em [7] e [12] sem precisar alterar suas demonstrações. Alguns poucos resultados do primeiro capítulo valem exclusivamente para espaços de estados finitos, o que deixaremos claro em seus enunciados.

Dividimos este texto em seis capítulos: no primeiro fixamos a notação utilizada na sequência do trabalho e relembramos os principais conceitos de Cadeias de Markov. Nessa parte inicial, muitos resultados clássicos são apresentados sem demonstração, mas indicamos referências. Também apresentamos alguns objetos importantes como os passeios aleatórios em grafos, a representação via sistema dinâmico aleatório e a versão *lazy* das cadeias. O segundo capítulo destina-se à apresentação do Teorema de Convergência, um resultado clássico do estudo de cadeias de Markov, além da introdução do primeiro indicador estudado neste trabalho: o tempo de mistura, t_{mix} .

O terceiro capítulo é onde, motivados pelos conceitos físicos de circuitos elétricos, introduzimos o conceito de resistência efetiva entre dois nós de um passeio aleatório em um grafo. Tal capítulo quebra um pouco a sequência desta dissertação, mas apresenta uma ferramenta

não apenas elegante, mas extremamente útil para a obtenção dos próximos resultados. Além disso, essa rica analogia proporciona aos passeios aleatórios intuições físicas vindas dos circuitos elétricos, e vice-versa. Ainda neste capítulo, dedicamos um esforço considerável para demonstrar três leis físicas aplicadas ao nosso contexto. Essas leis afirmam que, em um circuito elétrico, as condutâncias em paralelo são somadas, as resistências em série também são somadas, e que, sob certas condições, podemos “colar” dois nós sem alterar as condutâncias da rede. Embora esses resultados sejam intuitivos e possam ser demonstrados de forma simples através de álgebra, não encontramos na literatura outra obra que detalhe essas contas tão minuciosamente como fizemos aqui.

No quarto capítulo nos dedicamos a apresentação de outros indicadores das cadeias, e no quinto estudamos algumas relações entre eles e o t_{mix} . A maioria dessas relações estão presentes no livro texto [9], mas também incluímos alguns resultados mais recentes do artigo [13].

Para ilustrar esses indicadores em uma situação concreta, no último capítulo obteremos cotas para t_{mix} , t_{rel} e t_{hit} em um exemplo fundamental: o passeio aleatório no toro \mathbb{Z}_n^d . Para obtenção dessas cotas precisaremos estudar alguns exemplos clássicos preliminares: a ruína do jogador e a urna de Polya. Optamos por deixar a construção destes e de outros exemplos para o último capítulo. Tentamos referenciá-lo sempre que possível, mas ainda assim recomendamos ao leitor recorrer a este capítulo sempre que precisar de uma visualização dos conceitos.

Tentamos não deixar de lado nenhum resultado ou definição importante para a construção dos resultados propostos, visando deixar este texto o mais completo possível. Para facilitar a consulta, uma esquematização resumida dos principais resultados, referente aos indicadores t_{mix} , t_{rel} e t_{hit} , apresentados neste trabalho pode ser encontrada no Apêndice.

Capítulo 1

Cadeias de Markov: Definições Básicas

Neste capítulo relembremos o conceito de cadeia de Markov e exploraremos as propriedades fraca e forte de Markov. Também definiremos os principais conceitos envolvidos, como irreducibilidade, aperiodicidade e reversibilidade, além de definir medida invariante. Além disso, recordaremos alguns resultados clássicos desta teoria, em destaque o fato de que uma cadeia de Markov irreduzível em um espaço de estados finito possui uma única distribuição invariante.

No final ainda introduziremos alguns conceitos que serão úteis e recorrentes neste trabalho, como cadeias *lazy* e representação via sistema dinâmico aleatório, além de definir o passeio aleatório em um grafo qualquer.

1.1 Definição de um cadeia de Markov

Um **processo estocástico** no espaço de estados S é uma coleção de variáveis aleatórias $(X_t)_{t=0}^{\infty}$ tomando valores em S . O indexador t pode ser pensado como o tempo, com X_t representando a posição do processo no instante t . Trabalharemos exclusivamente em processos a tempo discreto, de modo que a notação $(X_t)_{t=0}^{\infty}$ (ou ainda $(X_t)_{t \geq 0}$) indicará que t está variando no conjunto $\{0, 1, 2, \dots\}$. Quando não causar confusão suprimiremos os índices, denotando este processo simplesmente por (X_t) .

Neste trabalho estaremos interessados especialmente em **cadeias de Markov** (homogêneas no tempo), as quais são processos estocásticos com perda de memória, ou seja, onde a posição do processo no tempo $t + 1$ depende apenas de um elemento aleatório e da posição em que ele se encontra no tempo t , independentemente dos estados ocupados nos instantes anteriores a t , e do próprio tempo t . Clarificaremos tal conceito formalizando-o. No que segue, S é um espaço de estados finito ou enumerável.

Definição 1.1. Dizemos que $p : S \times S \rightarrow [0, 1]$ é uma coleção de **probabilidades de transição** se para todo $x \in S$, $p(x, \cdot)$ for uma distribuição de probabilidade em S , isto é, se $\sum_{y \in S} p(x, y) = 1$.

A matriz $P = (p(x, y))_{x, y \in S}$ onde a x -ésima linha é dada pelo vetor $p(x, \cdot)$ tem em cada linha uma distribuição de probabilidade, e portanto a soma de seus elementos é unitária. A matriz que goza desta propriedade é dita **estocástica**.

Definição 1.2. Uma família $(X_t)_{t=0}^{\infty}$ de variáveis aleatórias definidas em um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ tomando valores em S é uma **cadeia de Markov** com distribuição inicial μ e matriz de transição P , ou simplesmente cadeia de Markov- (μ, P) , se as seguintes propriedades forem válidas:

- (i) para todo $x_0 \in S$, $\mathbb{P}(X_0 = x_0) = \mu(x_0)$;

(ii) $\mathbb{P}[X_{t+1} = x_{t+1} \mid X_0 = x_0, X_1 = x_1, \dots, X_t = x_t] = \mathbb{P}[X_{t+1} = x_{t+1} \mid X_t = x_t] = p(x_t, x_{t+1})$
 para quaisquer $x_0, x_1, \dots, x_{t+1} \in S$ e $t \geq 0$.

Em palavras, a propriedade (i) garante que o processo começa com distribuição μ , no sentido de que, olhando no tempo $t = 0$, a probabilidade de estar em um sítio x é dada por $\mu(x)$. Já a propriedade (ii) nos diz que **a dinâmica do processo é determinada pela matriz P** : se em um certo instante t o processo está no estado x , então no instante $t + 1$ o processo estará em um estado escolhido aleatoriamente de acordo com a distribuição $p(x, \cdot)$, independentemente dos estados visitados nos instantes anteriores a t . Essa independência é o que chamamos de *perda de memória* da cadeia.

As condições (i)+(ii) podem ser substituídas pela condição equivalente (iii) abaixo que estabelece a probabilidade de que o processo siga o caminho x_0, x_1, \dots, x_t .

(iii) Para todo $t > 0$ e quaisquer $x_0, x_1, \dots, x_t \in S$,

$$\mathbb{P}[X_0 = x_0, X_1 = x_1, \dots, X_t = x_t] = \mu(x_0)p(x_0, x_1) \dots p(x_{t-1}, x_t). \quad (1.1)$$

É possível também definir cadeia de Markov em termos de probabilidades condicionais com respeito às σ -álgebras geradas pelo processo, conforme é feito em [7]. Para isso recordemos que dado um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$, uma sigma álgebra $\mathcal{G} \subset \mathcal{F}$ e uma variável aleatória X com $\mathbb{E}|X| < \infty$, a **esperança condicional de X dada a sigma álgebra \mathcal{G}** , denotada por $\mathbb{E}[X|\mathcal{G}]$, é a única variável aleatória Y que satisfaz:

- Y é \mathcal{G} -mensurável
- Para todo $A \in \mathcal{G}$, $\int_A X d\mathbb{P} = \int_A Y d\mathbb{P}$

Intuitivamente, pensamos em \mathcal{G} como a informação que temos disponível: para cada $A \in \mathcal{G}$, sabemos se A aconteceu ou não. A esperança condicional $\mathbb{E}[X|\mathcal{G}]$ representa então “o melhor palpite” para o valor de X dada a informação provida pela σ -álgebra \mathcal{G} .

Dado um evento $A \in \mathcal{F}$ define-se a probabilidade condicional $\mathbb{P}(A|\mathcal{G})$ pela igualdade $\mathbb{P}(A|\mathcal{G}) = \mathbb{E}(\mathbf{1}_A|\mathcal{G})$.

Denotemos por \mathcal{F}_t a σ -álgebra gerada $\mathcal{F}_t = \sigma(X_0, X_1, \dots, X_t)$. A σ -álgebra \mathcal{F}_t representa toda a informação que temos sobre o processo até o instante t .

Proposição 1.3. *Uma sequência de variáveis aleatórias $(X_t)_{t \geq 0}$ definidas em um espaço $(\Omega, \mathcal{F}, \mathbb{P})$, tomando valores em S , é uma cadeia de Markov com matriz de transição P se, e somente se, para quaisquer $t \geq 0$ e $y \in S$, valer*

$$\mathbb{P}[X_{t+1} = y \mid \mathcal{F}_t] = P(X_t, y). \quad (1.2)$$

Para uma demonstração o leitor pode consultar [7, Teorema 6.1.1].

1.2 Existência da cadeia de Markov

Na Definição 1.2 caracterizamos uma sequência de variáveis aleatórias $(X_t)_{t=0}^{\infty}$ como uma cadeia de Markov- (μ, P) a partir da distribuição inicial μ e da família das probabilidades de transição $\{p(x, y); x, y \in S\}$. Para que tal definição esteja coerente, precisamos garantir que dadas μ e P , existem um espaço Ω com uma probabilidade \mathbb{P} , e variáveis aleatórias $X_t : \Omega \rightarrow S$ satisfazendo as condições desta definição.

Mostraremos agora como o espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ pode ser construído. Fixado o espaço de estados S , denotamos por $\Omega = S^{\mathbb{N}} = \{\omega = (\omega_0, \omega_1, \dots); \omega_t \in S \forall t \geq 0\}$ o espaço das trajetórias, e nele definimos, para cada $t \in \mathbb{N}$, a projeção

$$\begin{aligned} X_t : \Omega &\rightarrow S \\ \omega &\mapsto \omega_t \end{aligned}$$

Como a variável aleatória X_t representa a posição em que a cadeia se encontra no tempo t , os elementos $\omega \in \Omega$ passam a representar todas as possíveis trajetórias que a cadeia pode seguir neste espaço de estados. A partir dessa interpretação, o conjunto

$$\begin{aligned} [X_t = x] &= \{\omega \in \Omega; X_t(\omega) = x\} = \{\omega \in \Omega; \omega_t = x\} \\ &= \{\omega = (\omega_0, \dots, \omega_{t-1}, x, \omega_{t+1}, \dots) \in \Omega; \omega_r \in S \forall r \neq t\} \end{aligned}$$

consiste de todas as trajetórias tais que, no tempo t , a cadeia se encontra na posição $x \in S$.

Para cada subconjunto finito $\{t_0, t_1, \dots, t_k\} \subset \mathbb{N}$ ($k \geq 0$) definimos a projeção

$$\begin{aligned} (X_{t_0}, X_{t_1}, \dots, X_{t_k}) : \Omega &\rightarrow S^{k+1} \\ \omega &\mapsto (\omega_{t_0}, \omega_{t_1}, \dots, \omega_{t_k}) \end{aligned}$$

e chamamos de **cilindros** os subconjuntos de $S^{\mathbb{N}} = \Omega$ que têm a forma

$$(X_0, X_1, \dots, X_t)^{-1}(A) = \{\omega \in \Omega; (X_0, X_1, \dots, X_t)(\omega) \in A\}$$

para algum $t \geq 0$ e algum conjunto $A \subseteq S^{t+1}$.

Denotamos por $\mathcal{B}(S^{\mathbb{N}})$ a σ -álgebra gerada pelos cilindros. E para cada $t \geq 0$ denotamos por $\mathcal{P}(S^{t+1})$ a σ -álgebra das partes de S^{t+1} .

Considere a seguinte versão do Teorema de Extensão de Kolmogorov, cuja demonstração pode ser consultada em [3, Teorema 1.17] ou, em uma forma mais geral, em [7, Apêndice A.3].

Teorema 1.4 (Teorema de Extensão de Kolmogorov). *Seja $\mu_t : \mathcal{P}(S^{t+1}) \rightarrow [0, 1]$, $t = 0, 1, \dots$, uma sequência consistente de probabilidades, isto é, para cada t*

$$\mu_{t+1}(A_1 \times \dots \times A_t \times S) = \mu_t(A_1 \times \dots \times A_t),$$

então existe uma única probabilidade $\mathbb{P} : \mathcal{B}(S^{\mathbb{N}}) \rightarrow [0, 1]$ tal que

$$\mathbb{P}((X_0, X_1, \dots, X_t) \in E) = \mu_t(E), \tag{1.3}$$

para todo $t \geq 0$ e todo $E \subseteq S^{t+1}$.

Dada uma família de probabilidades de transição $\{p(x, y); x, y \in S\}$ e uma distribuição inicial μ , definimos para cada $t \geq 0$ uma probabilidade μ_t em S^{t+1} atribuindo um peso a cada elemento $x = (x_0, x_1, \dots, x_t) \in S^{t+1}$ por

$$\mu_t(x) = \mu(x_0)p(x_0, x_1) \dots p(x_{t-1}, x_t).$$

Verifica-se que essa família de probabilidades é consistente e portanto, o Teorema de Extensão de Kolmogorov garante que existe uma única probabilidade \mathbb{P} em $(S^{\mathbb{N}}, \mathcal{B}(S^{\mathbb{N}}))$ de modo que as projeções X_t satisfazem (1.1). Assim, as variáveis aleatórias projeção $X_t : \Omega \rightarrow S$ definidas por $X_t(\omega) = \omega_t$ formam uma cadeia de Markov- (μ, P) . Note que \mathbb{P} é uma medida no espaço das trajetórias, que depende da matriz P e da distribuição inicial μ . Sempre que quisermos enfatizar a distribuição inicial escreveremos \mathbb{P}_μ ao invés de \mathbb{P} . A esperança com respeito a \mathbb{P}_μ é denotada por \mathbb{E}_μ . Quando $\mu = \delta_x$, escrevemos simplesmente \mathbb{P}_x e \mathbb{E}_x .

Cálculo de probabilidades com produto matricial

Fixada P , a matriz de transição da cadeia de Markov, como o espaço de estados S é finito ou enumerável, uma medida $\mu : S \rightarrow [0, +\infty]$ pode ser vista como um vetor linha ($\mu(x)$; $x \in S$), permitindo-nos trabalhar com o produto μP , que é outro vetor linha, o qual representa a distribuição da cadeia no instante $t = 1$, de fato

$$\mathbb{P}_\mu(X_1 = z) = \sum_{y \in S} \mathbb{P}_\mu(X_0 = y, X_1 = z) = \sum_{y \in S} \mu(y)p(y, z) = (\mu P)(z).$$

Com indução em t se mostra também que $\mathbb{P}_x[X_t = y] = P^t(x, y)$ e, mais geralmente, $\mathbb{P}_\mu[X_t = y] = (\mu P^t)(y)$.

1.3 Propriedades fraca e forte de Markov

A caracterização de uma cadeia de Markov através da identidade (1.2) permite-nos formalizar alguns resultados chave para o desenvolvimento da teoria. Antes disso, precisamos definir os operadores *shift*.

Lembremos que as variáveis aleatórias X_t que representam a cadeia de Markov estão definidas no espaço das sequências infinitas $\Omega = S^{\mathbb{N}}$ cujos elementos são as trajetórias

$$w = (\omega_0, \omega_1, \omega_2, \dots)$$

com $\omega_t \in S$, e são caracterizadas como as projeções $X_t : \Omega \rightarrow S$ com $X_t(\omega) = \omega_t$ (e o que faz com que $(X_t)_{t \geq 0}$ seja Markov- (μ, P) é a probabilidade \mathbb{P}_μ definida neste espaço). Para cada $t \in \mathbb{N}$, definimos o operador *shift* $\theta_t : \Omega \rightarrow \Omega$ por

$$\theta_t \omega = (\omega_t, \omega_{t+1}, \omega_{t+2}, \dots),$$

isto é, para todo $s \in \mathbb{N}$, $(\theta_t \omega)_s = \omega_{t+s}$, a coordenada $t + s$ de ω . Então, considerando como antes $\mathcal{F}_t = \sigma(X_0, \dots, X_t)$, podemos enunciar o seguinte resultado:

Teorema 1.5 (Propriedade Fraca de Markov). *Seja $Y : \Omega \rightarrow \mathbb{R}$ limitada e mensurável considerando a σ -álgebra dos cilindros. Denotando por $\mathbb{E}_{X_t}[Y]$ a variável aleatória $h(X_t)$, onde $h(x) = \mathbb{E}_x[Y]$, tem-se*

$$\mathbb{E}_\mu[Y \circ \theta_t \mid \mathcal{F}_t] = \mathbb{E}_{X_t}[Y].$$

A demonstração deste resultado utiliza-se do Teorema $\pi - \lambda$ e do Teorema da classe monótona, e pode ser encontrada em [7, Teorema 6.3.1].

Neste contexto, o operador θ_t tem o significado de ignorar as primeiras t coordenadas da trajetória. A propriedade de Markov pode ser interpretada intuitivamente do seguinte modo: sabendo o que acontece até o tempo t , é como se o processo recomeçasse no tempo t , esquecendo tudo o que houve anteriormente e seguindo a partir de onde está com dinâmica dada pela matriz P . Em particular, se soubermos que $X_t = x \in S$, então a cadeia $(X_{s+t})_{s \geq 0}$ será Markov- (δ_x, P) , e independente das variáveis aleatórias X_0, \dots, X_t . É dessa forma mais intuitiva que a propriedade de Markov é apresentada em [12, Teorema 1.1.2].

É por conta do comentário acima que tanto a propriedade fraca de Markov quanto a forte (que veremos adiante) são muitas vezes referidas como “perda de memória da cadeia”.

Podemos ainda generalizar a propriedade fraca de Markov enunciada acima, substituindo o tempo t determinístico por uma variável aleatória τ que satisfaz certas condições. Para isso, lembramos que uma sequência de σ -álgebras $\{\mathcal{G}_t\}_{t=0}^\infty$ é dita uma **filtração** se para todo t valer $\mathcal{G}_t \subset \mathcal{G}_{t+1}$, e definimos:

Definição 1.6 (Tempo de Parada). *Uma variável aleatória $\tau : \Omega \rightarrow \mathbb{N} \cup \{\infty\}$ é dita um tempo de parada com respeito a filtração $\{\mathcal{G}_t\}_{t \geq 0}$ se, para todo $t \geq 0$ tivermos $[\tau = t] \in \mathcal{G}_t$.*

Em particular, a sequência de σ -álgebras $\mathcal{F}_t = \sigma(X_0, \dots, X_t)$, onde cada \mathcal{F}_t representa a informação dada pela cadeia até o tempo t , é uma filtração. E dizer que $[\tau = t] \in \mathcal{F}_t$ significa que com a informação da cadeia até o tempo t , sabemos dizer se $\tau = t$.

Definição 1.7. *Se τ é um tempo de parada para uma cadeia de Markov, definimos a σ -álgebra*

$$\mathcal{F}_\tau := \{A \subset \Omega; A \cap \{\tau = t\} \in \mathcal{F}_t \text{ para todo } t\},$$

que intuitivamente, representa a informação que temos até o instante aleatório τ .

Assim estamos prontos para enunciar o seguinte teorema, cuja demonstração é consequência direta da propriedade fraca de Markov, e pode ser encontrada em [7, Teorema 6.3.4]

Teorema 1.8 (Propriedade Forte de Markov). *Sejam $(X_t)_{t \geq 0}$ cadeia de Markov em S , $\Omega = S^{\mathbb{N}}$ o espaço das trajetórias, τ tempo de parada e $Y_t : \Omega \rightarrow \mathbb{R}$ mensurável de maneira que $|Y_t| \leq M$, para todo $t \geq 0$. Então, no evento $\{\tau < \infty\}$,*

$$\mathbb{E}_\mu[Y_\tau \circ \theta_\tau \mid \mathcal{F}_\tau] = \mathbb{E}_{X_\tau}[Y_\tau],$$

onde $\mathbb{E}_{X_\tau}[Y_\tau]$ denota o valor da função $\varphi(x, t) = \mathbb{E}_x[Y_t]$ calculada em $(x, t) = (X_\tau, \tau)$.

Do ponto de vista intuitivo, a propriedade forte de Markov nos diz que se $(X_t)_{t \geq 0}$ é Markov- (μ, P) e τ um tempo de parada da cadeia, então, condicionado a $\tau < \infty$ e a $X_\tau = x$, $(X_{\tau+t})_{t \geq 0}$ é Markov- (δ_x, P) e independente das variáveis aleatórias X_0, \dots, X_τ .

Observação 1.9. *Até agora definimos e enunciamos os principais resultados tanto do ponto de vista mais rigoroso, seguindo [7], como do ponto de vista mais intuitivo/informal, seguindo [12] e [9]. Tivemos tal cuidado para que no futuro possamos utilizar a abordagem que achamos mais útil para os propósitos do texto, procurando não perder a intuição sobre os objetos em questão, e muito menos a segurança advinda do rigor matemático.*

1.4 Irredutibilidade e estrutura de classes

Seja $(X_t)_{t \geq 0}$ cadeia de Markov no espaço de estados S finito ou enumerável. Para $x, y \in S$, dizemos que y é acessível a partir de x , e escrevemos $x \rightarrow y$, se $\mathbb{P}_x[X_t = y \text{ para algum } t \geq 0] > 0$. Também dizemos que x e y se comunicam, e escrevemos $x \leftrightarrow y$, se $x \rightarrow y$ e $y \rightarrow x$. Isso define uma relação de equivalência no conjunto S , dividindo-o em classes de equivalência.

A **cadeia** é dita **irredutível** se existir uma única classe de equivalência, isto é, se $x \rightarrow y$ para quaisquer $x, y \in S$. Neste caso é possível partindo de um estado acessar qualquer outro.

O próximo resultado nos trás equivalências importantes para a definição de um estado ser acessível por outro. Sua demonstração é bem simples e pode ser encontrada em [12, Teorema 1.2.1].

Proposição 1.10. *São equivalentes:*

- i. $x \rightarrow y$;
- ii. $P^t(x, y) > 0$ para algum $t \geq 0$;
- iii. existem $t \geq 0$ e $x = x_0, x_1, \dots, x_t = y$ tais que $P(x_0, x_1)P(x_1, x_2) \cdots P(x_{t-1}, x_t) > 0$.

Para $x \in S$, consideramos $\mathcal{T}(x) := \{t \geq 1; P^t(x, x) > 0\}$ o conjunto dos tempos em que é possível retornar a x . Chamamos de **período** do estado x o número $d(x) := \text{mdc } \mathcal{T}(x)$. Se $d(x) = 1$, então x é dito um ponto aperiódico.

Observação 1.11. *Este conceito pode parecer artificial introduzido neste momento, mas será fundamental quando discutirmos o teorema de convergência no Capítulo 2. O leitor interessado pode verificar o Exemplo 2.13, onde verificamos que no passeio aleatório no n -ciclo (com n par) o estado 0 possui período 2. Por enquanto vamos verificar alguns resultados relacionados, que também serão úteis futuramente.*

Lema 1.12. *Sejam $(X_t)_{t \geq 0}$ cadeia de Markov sobre S e $x, y \in S$. Se $x \leftrightarrow y$, então $d(x) = d(y)$.*

Demonstração. De fato, como $x \leftrightarrow y$, existem $r, s > 0$ tais que $P^r(x, y) > 0$ e $P^s(y, x) > 0$, em particular, temos que $P^{r+s}(y, y) \geq P^s(y, x)P^r(x, y) > 0$, e assim $d(y) \mid r + s$. Além disso, se t é tal que $P^t(x, x) > 0$, vale $P^{r+s+t}(y, y) \geq P^s(y, x)P^t(x, x)P^r(x, y) > 0$, de modo que $d(y) \mid r + s + t$.

Em particular, obtemos que $d(y) \mid (r + s + t) - (r + s) = t$, ou seja, mostramos que para todo $t > 0$ tal que $P^t(x, x) > 0$, vale $d(y) \mid t$, ou seja, $d(y)$ é um divisor comum de todos os elementos do conjunto $\{t > 0; P^t(x, x) > 0\}$, o que garante que $d(y) \leq d(x)$, visto que este último é o maior divisor comum deste mesmo conjunto.

Analogamente mostra-se que $d(x) \leq d(y)$, o que conclui a afirmação. \square

Como consequência, se uma cadeia é irredutível, todos os pontos possuem o mesmo período, o qual é denominado período da cadeia. Se todos os estados de uma cadeia irredutível possuírem período 1, dizemos que tal cadeia é **aperiódica**.

Lema 1.13. *Se x é um ponto aperiódico, então existe um $t_0 \in \mathbb{N}$ tal que $P^t(x, x) > 0$ para todo $t \geq t_0$.*

A demonstração deste lema é bem simples, bastando combinar o resultado anterior com o seguinte fato de teoria dos números: Se $A \subset \mathbb{N}$ é um conjunto fechado para somas, com $\text{mdc}(A) = 1$, então existe um natural m_0 tal que $A \supset \{m_0, m_0 + 1, \dots\}$.

Ainda, se o espaço de estados for finito, e a cadeia for irredutível e aperiódica, combinando o lema anterior com a Proposição 1.10 temos o seguinte resultado, que pode ser encontrado em [9, Proposição 1.7]:

Proposição 1.14. *Se P é uma cadeia irredutível e aperiódica em um espaço de estados finito, então existe um inteiro r_0 tal que $P^r(x, y) > 0$ para todos $x, y \in S$ e todo $r \geq r_0$.*

1.5 Tempo de acerto

Uma variável aleatória de fundamental importância no desenvolvimento da teoria é o tempo de acerto de um conjunto, também chamado de *hitting time*, definido como segue:

Definição 1.15. *Para $A \subset S$, e $(X_t)_{t \geq 0}$ cadeia de Markov em S , definimos o **tempo de acerto** do conjunto A por*

$$\tau_A := \min\{t \geq 0; X_t \in A\},$$

ou seja, o primeiro momento em que o processo chega no conjunto A . Com abuso de notação, quando $A = \{x\} \subset S$, escrevemos τ_x .

Também definimos a variável aleatória $\tau_x^+ := \min\{t > 0; X_t = x\}$, que quando $X_0 = x$ é interpretada como o **tempo do primeiro retorno** ao sítio x .

Lema 1.16. Para quaisquer dois estados x e y de uma cadeia de Markov irredutível em um espaço de estados finito, $\mathbb{E}_x[\tau_y^+] < \infty$.

Ou seja, saindo de um estado x , o tempo esperado para chegar em um estado y é finito. Em particular, como $\tau_y \leq \tau_y^+$, vale também $\mathbb{E}_x[\tau_y] < \infty$.

Demonstração. Do fato de P ser irredutível temos que, para quaisquer $x, y \in S$, existe um $t = t(x, y) > 0$ tal que $P^t(x, y) > 0$. E sendo S finito, tomando máximos e mínimos, temos que existe um inteiro $r > 0$ e um real $\varepsilon > 0$ de modo que, para todos $z, w \in S$, existe um $j \leq r$ tal que $P^j(z, w) > \varepsilon$.

Agora note que, para $w \in S$ qualquer

$$\mathbb{P}_w(\tau_y^+ \leq r) = \mathbb{P}_w(\cup_{j=1}^r [X_j = y]) \geq \mathbb{P}_w[X_j = y] = P^j(w, y),$$

para todo $0 < j \leq r$. Em particular, como existe um $j \leq r$ tal que $P^j(w, y) > \varepsilon$, segue que

$$\mathbb{P}_w(\tau_y^+ \leq r) > \varepsilon. \quad (1.4)$$

Da propriedade Fraca de Markov segue que, para $k > 0$ qualquer,

$$\begin{aligned} \mathbb{P}_x(\tau_y^+ > kr) &= \mathbb{E}_x \left[\mathbb{E}_x \left[\mathbf{1}_{\{\tau_y^+ > kr\}} \mid \mathcal{F}_{(k-1)r} \right] \right] \\ &= \mathbb{E}_x \left[\mathbb{E}_x \left[\mathbf{1}_{\{\tau_y^+ > (k-1)r\}} \cdot \left(\mathbf{1}_{\{\tau_y^+ > r\}} \circ \theta_{(k-1)r} \right) \mid \mathcal{F}_{(k-1)r} \right] \right] \\ &= \mathbb{E}_x \left[\mathbf{1}_{\{\tau_y^+ > (k-1)r\}} \mathbb{E}_x \left[\mathbf{1}_{\{\tau_y^+ > r\}} \circ \theta_{(k-1)r} \mid \mathcal{F}_{(k-1)r} \right] \right] \\ &= \mathbb{E}_x \left[\mathbf{1}_{\{\tau_y^+ > (k-1)r\}} \mathbb{E}_{X_{(k-1)r}} \left[\mathbf{1}_{\{\tau_y^+ > r\}} \right] \right] \\ &= \mathbb{E}_x \left[\mathbf{1}_{\{\tau_y^+ > (k-1)r\}} \mathbb{P}_{X_{(k-1)r}} \left[\tau_y^+ > r \right] \right]. \end{aligned} \quad (1.5)$$

De (1.4) temos que $\mathbb{P}_w(\tau_y^+ > r) \leq 1 - \varepsilon$ para todo $w \in S$, de modo que a variável aleatória $\mathbb{P}_{X_{(k-1)r}} \left[\tau_y^+ > r \right]$ é limitada por $1 - \varepsilon$, portanto, o lado direito de (1.5) é limitado por

$$\mathbb{E}_x \left[\mathbf{1}_{\{\tau_y^+ > (k-1)r\}} (1 - \varepsilon) \right] \leq (1 - \varepsilon) \mathbb{P}_x \left[\tau_y^+ > (k-1)r \right].$$

Repetidas aplicações desta desigualdade nos dão que

$$\mathbb{P}_x(\tau_y^+ > kr) \leq (1 - \varepsilon)^k$$

Por outro lado, como τ_y^+ é uma variável aleatória não negativa,

$$\mathbb{E}_x[\tau_y^+] = \sum_{t \geq 0} \mathbb{P}_x(\tau_y^+ > t) = \sum_{k \geq 0} \left(\sum_{j=0}^{r-1} \mathbb{P}_x(\tau_y^+ > kr + j) \right)$$

Como $\mathbb{P}_x(\tau_y^+ > t)$ é uma função decrescente de t , os r termos do somatório de dentro podem ser limitados por $\mathbb{P}_x(\tau_y^+ > kr)$, donde segue que

$$\mathbb{E}_x[\tau_y^+] \leq \sum_{k \geq 0} (r \mathbb{P}_x(\tau_y^+ > kr)) \leq r \sum_{k \geq 0} (1 - \varepsilon)^k < \infty.$$

□

Cabe observar que na demonstração acima justificamos que

$$\mathbb{P}_x(\tau_y^+ > kr) \leq (1 - \varepsilon)\mathbb{P}_x[\tau_y^+ > (k-1)r].$$

através da propriedade fraca de Markov, “quebrando” a cadeia no tempo $(k-1)r$ com o operador $\theta_{(k-1)r}$. Tal igualdade é intuitiva se pensarmos que o lado esquerdo indica a probabilidade de saindo de x , não visitar y nos primeiros kr passos; mas isto acontecer é a mesma coisa que saindo de x não chegar em y nos primeiros $(k-1)r$ passos e então, como a cadeia recomeça independente dos passos anteriores e com a mesma dinâmica, sair de onde estiver, no caso $X_{(k-1)r}$, e não chegar em y nos primeiros r passos. Este último termo é limitado por $1 - \varepsilon$, enquanto que o outro é a probabilidade que temos no lado direito.

1.6 Recorrência e transiência

Definição 1.17. *Seja P uma cadeia de Markov no espaço de estados S finito ou enumerável. Um estado $x \in S$ é dito **recorrente** se $\mathbb{P}_x[\tau_x^+ < \infty] = 1$, e dito **transiente** caso contrário, isto é, se $\mathbb{P}_x[\tau_x^+ < \infty] < 1$.*

Como consequência do Lema 1.16, se a cadeia de Markov for irredutível em um espaço de estados finito, então para qualquer $x \in S$, $\mathbb{E}_x[\tau_x^+] < \infty$, consequentemente $\mathbb{P}_x[\tau_x^+ < \infty] = 1$ e x é recorrente.

Com a afirmação acima, pode parecer que tal conceito é desnecessário para o objetivo principal deste trabalho, que são cadeia de Markov em espaços de estados finitos. Isto não é verdade, uma vez que vários resultados destes capítulos iniciais serão discutidos para espaços de estados enumeráveis, e mesmo se os adaptássemos para espaços de estados finitos, o conceito de recorrência permearia suas demonstrações.

Para facilitar notação nesta seção, definimos $\rho_{xy} := \mathbb{P}_x[\tau_y^+ < \infty]$, isto é, a probabilidade, saindo de x , de chegar em y em tempo finito. Consequentemente, um estado é recorrente se, e somente se, $\rho_{xx} = 1$. Os próximos resultados são apenas uma reorganização do conteúdo de [7, Seção 6.4].

Proposição 1.18. *Definindo $N_y = \sum_{t \geq 1} \mathbf{1}_{\{X_t=y\}}$ o número de visitas ao estado y em tempos positivos, temos que $\mathbb{E}_x[N_y] = \frac{\rho_{xy}}{1 - \rho_{yy}}$. Em particular, temos que $y \in S$ é recorrente se, e somente se, $\mathbb{E}_y[N_y] = \infty$.*

Com base neste resultado, podemos demonstrar uma das propriedades mais interessantes da recorrência: o fato dela ser contagiosa, conforme afirma no seguinte resultado:

Proposição 1.19. *Se $x \in S$ é recorrente, e $\rho_{xy} = \mathbb{P}_x[\tau_y^+ < \infty] > 0$, então y também é recorrente. Além disso, neste caso $\rho_{yx} = \mathbb{P}_y[\tau_x^+ < \infty] = 1$.*

Em particular, note que se a cadeia for irredutível, ou todos os estados são recorrentes, ou todos são transientes. De fato, observe que para $x, y \in S$ quaisquer, a irredutibilidade garante a existência de um $t > 0$ tal que $P^t(x, y) > 0$, logo $\rho_{xy} = \mathbb{P}_x(\tau_y^+ < \infty) \geq \mathbb{P}_x(X_t = y) = P^t(x, y) > 0$. Assim, se existir um estado $x \in S$ recorrente, como qualquer $y \in S$ satisfaz $\rho_{xy} > 0$, o resultado anterior garante que y também é recorrente.

Além disso, se S é finito deve possuir ao menos um estado recorrente, pois se não fosse teríamos $\rho_{yy} < 1$ para todo $y \in S$, de modo que

$$\mathbb{E}_x[N_y] = \frac{\rho_{xy}}{1 - \rho_{yy}} < \infty$$

para todo $x \in S$, o que implicaria por sua vez, como S é finito, que

$$\infty > \sum_{y \in S} \mathbb{E}_x[N_y] = \sum_{y \in S} \sum_{t=1}^{\infty} P^t(x, y) = \sum_{t=1}^{\infty} \sum_{y \in S} P^t(x, y) = \sum_{t=1}^{\infty} 1 = \infty,$$

um absurdo.

Os dois comentários acima fornecem uma demonstração alternativa para o fato de que em uma cadeia de Markov irreduzível em um espaço de estados finito todos os estados são recorrentes.

1.7 Distribuição estacionária

Uma medida μ em S é dita **estacionária** (ou **invariante**) para a cadeia de Markov P se satisfaz

$$\mu P = \mu,$$

Tal equação pode ser reescrita, elemento a elemento, por:

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

Chamaremos de distribuição em S qualquer medida de probabilidade em S , isto é, uma medida tal que $\sum_{x \in S} \mu(x) = 1$. Em particular, note que se μ for uma distribuição invariante então para todo $t \geq 0$,

$$\mathbb{P}_\mu(X_t = x) = \mu P^t(x) = \mu(x),$$

ou seja, ao começar com a distribuição estacionária μ , em qualquer instante de tempo a probabilidade de encontrar-se no sítio x vai ser a mesma, a saber, $\mu(x)$. Na Seção 1.9 veremos nosso primeiro exemplo de obtenção de uma distribuição invariante, naquele caso, para o passeio aleatório em um grafo.

No caso em que S é finito, usando um resultado de topologia, existe um prova bem simples (mas não construtiva) para o seguinte:

Teorema 1.20. *Se S é finito, toda cadeia de Markov em S admite uma distribuição invariante.*

Demonstração. Suponha que $|S| = n$. O conjunto

$$\Lambda = \left\{ \mu : S \rightarrow [0, 1]; \sum_{x \in S} \mu(x) = 1 \right\}$$

das medidas de probabilidade em S é um compacto convexo e a transformação $T : \Lambda \rightarrow \Lambda$, dada por $T(\mu) = \mu P$, é contínua. Assim, o resultado segue do Teorema do Ponto Fixo de Brouwer¹. \square

Considerado-se uma cadeia com dois estados absorventes percebe-se que a distribuição invariante pode não ser única. Mas veremos no Teorema 1.10 que temos unicidade se a cadeia for irreduzível.

Discutiremos agora, seguindo a abordagem de [12], uma construção probabilística de medidas estacionárias para cadeias com espaço de estados enumeráveis. Depois particularizaremos os resultados para o nosso caso de interesse, onde S é finito.

¹O Teorema do Ponto Fixo de Brouwer afirma que toda função contínua de um compacto convexo em si mesmo possui um ponto fixo.

Teorema 1.21. *Seja P uma cadeia de Markov irredutível e recorrente em um espaço de estados S finito ou enumerável. Fixado $y \in S$, definimos a medida $\mu^y : S \rightarrow \mathbb{R}$ por*

$$\mu^y(x) = \mathbb{E}_y \left[\sum_{t=0}^{\tau_y^+ - 1} \mathbf{1}_{\{X_t=x\}} \right], \quad (1.6)$$

ou seja, saindo de y o número esperado de visitas a x antes de retornar a y . Então:

- i. $\mu^y(y) = 1$;
- ii. $\mu^y P = \mu^y$;
- iii. $0 \leq \mu^y(x) < \infty$ para todo $x \in S$.

Demonstração. A demonstração do item (i) é trivial a partir da definição. Para o item (iii) observamos que cada termo do somatório é uma função indicadora, que assume valores 0 ou 1, assim todos os termos são positivos, sendo o somatório uma variável aleatória positiva, portanto com esperança maior ou igual a 0. Para a outra desigualdade, fixamos $x \in S$, e observamos que como P é irredutível, existe um $m > 0$ tal que $P^m(x, y) > 0$, logo utilizando os itens (i) e (ii), temos que

$$1 = \mu^y(y) = \mu^y P^m(y) = \sum_{z \in S} \mu^y(z) P^m(z, y) \geq \mu^y(x) P^m(x, y),$$

donde seque que

$$\mu^y(x) \leq \frac{1}{P^m(x, y)} < \infty.$$

O item (ii) é um caso particular do Lema 4.4, com $\mu = \nu = \delta_y$ e $\tau = \tau_y^+$. Neste caso valem trivialmente as condições do lema: $\mathbb{P}_\mu(X_\tau = \cdot) = \mathbb{P}_y(X_{\tau_y^+} = \cdot) = \delta_y = \nu$, e $\mathbb{P}_\mu(\tau < \infty) = \mathbb{P}_y(\tau_y^+ < \infty) = 1$ uma vez que y é recorrente. \square

Observação 1.22. *Optamos por não demonstrar aqui o item (ii) do teorema, visto que sua demonstração é uma simples adaptação da exibida para o Lema 4.4 na Seção 4.3. O leitor que quiser ver tal demonstração agora não terá dificuldade, visto que as ferramentas lá utilizadas já foram todas construídas e o raciocínio não difere dos utilizados até o presente momento.*

O teorema anterior garante que toda cadeia irredutível e recorrente possui medida invariante. O próximo resultado, cuja demonstração pode ser encontrada em [12, Teorema 1.7.6], garante que ela é única a menos de multiplicação por escalar.

Teorema 1.23. *Seja P irredutível e ν uma medida invariante para P com $\nu(y) = 1$. Então $\nu \geq \mu^y$. Se além disso P for recorrente, então $\nu = \mu^y$.*

Resumindo os resultados anteriores, toda cadeia de Markov irredutível e recorrente em um espaço de estados S finito ou enumerável possui uma medida invariante, que é única a menos de multiplicação por escalar. Quando o espaço não for finito, pode não ser possível normalizar tal medida, de modo que não existe uma distribuição invariante. Um exemplo clássico é o passeio aleatório simples em \mathbb{Z} onde anda-se para a direita com probabilidade $1/2$ ou para a esquerda também com probabilidade $1/2$. É fácil ver que a medida uniforme em \mathbb{Z} , dada por $\mu(z) = 1$, para todo $z \in \mathbb{Z}$, é invariante. Mas como \mathbb{Z} é infinito essa medida não pode ser normalizada para se tornar uma probabilidade, de modo que não existe uma distribuição invariante. Visando tratar estes casos, introduzimos o conceito de recorrência positiva:

Definição 1.24. Um estado recorrente x em um espaço de estados S finito ou enumerável é dito **recorrente positivo** se $\mathbb{E}_x[\tau_x^+] < \infty$, e **recorrente nulo** se $\mathbb{E}_x[\tau_x^+] = \infty$.

Com isso podemos exprimir uma condição para existência de distribuição invariante.

Teorema 1.25. Para uma cadeia de Markov P irredutível, são equivalentes:

- i. Todos os estados são recorrentes positivos;
- ii. Algum estado é recorrente positivo;
- iii. Existe uma distribuição invariante π .

Além disso, se vale qualquer uma das alternativas (e portanto todas),

$$\pi(x) = \frac{1}{\mathbb{E}_x[\tau_x^+]} \quad (1.7)$$

e, em particular, a distribuição invariante é única.

Demonstração. (i) \Rightarrow (ii) Trivial.

(ii) \Rightarrow (iii) Fixamos $y \in S$ recorrente positivo. Em particular y é recorrente, o que de acordo com a demonstração do Teorema 1.21 basta para garantir que μ^y definida em (1.6) seja uma medida invariante para P . Além disso, graças ao teorema de Fubini, temos que

$$\sum_{x \in S} \mu^y(x) = \sum_{x \in S} \mathbb{E}_y \left[\sum_{t=0}^{\tau_y^+ - 1} \mathbf{1}_{\{X_t=x\}} \right] = \mathbb{E}_y \left[\sum_{t=0}^{\tau_y^+ - 1} \sum_{x \in S} \mathbf{1}_{\{X_t=x\}} \right] = \mathbb{E}_y[\tau_y^+] < \infty. \quad (1.8)$$

Deste modo podemos normalizar μ^y , obtendo a distribuição de probabilidade

$$\pi(x) = \frac{\mu^y(x)}{\sum_{z \in S} \mu^y(z)} = \frac{\mu^y(x)}{\mathbb{E}_y[\tau_y^+]},$$

que por ser igual a μ^y multiplicada por uma constante, continua sendo invariante para P . Portanto, vale (iii).

(iii) \Rightarrow (i) Suponha que existe π distribuição invariante, e fixe $z \in S$ qualquer. Como π é uma probabilidade, $\sum_{y \in S} \pi(y) = 1$, logo existe um $y \in S$ com $\pi(y) \neq 0$. Além disso, como P é irredutível, existe um $r > 0$ tal que $P^r(y, z) > 0$, logo

$$\pi(z) = (\pi P^r)(z) = \sum_{w \in S} \pi(w) P^r(w, z) \geq \pi(y) P^r(y, z) > 0,$$

e fica bem definida

$$\gamma(x) := \frac{\pi(x)}{\pi(z)}.$$

Note que γ é invariante por ser a multiplicação de π pelo escalar $1/\pi(z)$. Além disso, $\gamma(z) = 1$ e o Teorema 1.23 garante-nos que $\gamma \geq \mu^z$. Utilizando as mesmas contas de equação (1.8), obtemos

$$\mathbb{E}_z[\tau_z^+] = \sum_{y \in S} \mu^z(y) \leq \sum_{y \in S} \gamma(y) = \sum_{y \in S} \frac{\pi(y)}{\pi(z)} = \frac{1}{\pi(z)} < \infty, \quad (1.9)$$

o que mostra que z é recorrente positivo.

Para finalizar, note que se vale uma, e portanto todas as alternativas, então P é recorrente e o Teorema 1.23 garante também que vale a igualdade $\gamma = \mu^z$. Deste modo a desigualdade (1.9) é na verdade uma igualdade, valendo

$$\mathbb{E}_z[\tau_z^+] = \frac{1}{\pi(z)}.$$

□

1.7.1 No caso S finito

Se o espaço de estados S for finito, basta que a cadeia seja irredutível para que tenhamos as conclusões obtidas acima. De fato, conforme observamos no início da seção anterior, sob a hipótese de irredutibilidade o Lema 1.16 garante que para qualquer $y \in S$, $\mathbb{E}_y[\tau_y^+] < \infty$, ou seja, que y é recorrente positivo. Então do Teorema 1.25 temos que existe uma única distribuição invariante π dada pela equação (1.7). Para resumir isto, enunciaremos o seguinte teorema:

Teorema 1.26. *Seja P uma cadeia de Markov irredutível em um espaço de estados S finito. Então a distribuição π em S dada por*

$$\pi(x) = \frac{1}{\mathbb{E}_x[\tau_x^+]} \quad (1.10)$$

é a única distribuição invariante.

1.8 Reversibilidade

Dizemos que uma cadeia de Markov P é **reversível** se existir uma medida $\mu : S \rightarrow \mathbb{R}$ tal que, para todos $x, y \in S$, vale a equação do balanço detalhado

$$\mu(x)P(x, y) = \mu(y)P(y, x). \quad (1.11)$$

Neste caso, para todo $x \in S$,

$$\mu P(y) = \sum_{x \in S} \mu(x)P(x, y) = \sum_{x \in S} \mu(y)P(y, x) = \mu(y) \sum_{x \in S} P(y, x) = \mu(y), \quad (1.12)$$

mostrando que se P é reversível com respeito à medida μ , então μ é uma medida invariante para P .

Note também que é consequência da equação do balanço detalhado que para quaisquer $x_0, x_1, \dots, x_t \in S$,

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

Pela definição de uma cadeia de Markov, tal igualdade pode ser reescrita como

$$\mathbb{P}_\pi((X_0, \dots, X_t) = (x_0, \dots, x_t)) = \mathbb{P}_\pi((X_0, \dots, X_t) = (x_t, \dots, x_0)), \quad (1.13)$$

indicando que se a distribuição inicial for a estacionária, então a distribuição de (X_0, X_1, \dots, X_t) é a mesma de $(X_t, X_{t-1}, \dots, X_0)$. Em tradução livre, [1] caracterizam este comportamento como “olhando para o filme da cadeia e depois olhando ele de trás para frente, é impossível dizer qual é qual”. Tal observação justifica chamarmos tais cadeias de reversíveis.

Observação 1.27. *Em [11, Exercício 2.1] os autores exploram as principais propriedades de uma cadeia reversível.*

Se P e π satisfazem a equação do balanço detalhado (1.11), utilizando que $P^2(x, y) = \sum_{z \in S} P(x, z)P(z, y)$ é simples mostrar que para quaisquer estados $x, y \in S$, vale também

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

A partir disso, uma simples indução garante que, para todo $t \geq 1$ e todos estados $x, y \in S$, vale

$$\pi(x)P^t(x, y) = \pi(y)P^t(y, x). \quad (1.14)$$

1.9 Passeios aleatórios em grafos

Grafos constituem uma ferramenta importantíssima no estudo de cadeias de Markov uma vez que grande parte das cadeias podem ser vistas como um passeio aleatório (simples ou ponderado) em um grafo, simplesmente considerando o espaço de estados como o conjunto dos vértices e dando pesos às arestas de acordo com a matriz de transição P , conforme veremos no Capítulo 3. Nesta seção vamos apenas definir os conceitos básicos deste objeto, os quais utilizaremos no decorrer deste trabalho. Para manter a concordância com as principais literaturas da área, chamaremos de “grafo” o objeto que na álgebra é conhecido como “grafo simples” (sem *loops* ou arestas paralelas), e mais adiante nos referiremos ao que na álgebra é conhecido como “grafo” pelo nome do “multigrafo”.

Definição 1.28. *Um grafo $G = (V, E)$ é um par onde V é um conjunto de vértices e $E \subset \{\{x, y\}; x, y \in V, x \neq y\}$ é um conjunto de pares não ordenados de vértices distintos, que chamaremos de arestas.*

No nosso texto, estamos interessados em grafos em que V é um conjunto finito. Se existe uma aresta ligando os vértices x e y , isto é, $\{x, y\} \in E$, dizemos que x e y são **vizinhos**, e escrevemos $x \sim y$. Como a aresta não possui orientação, é claro que neste caso y também é vizinho de x . Dizemos que o **grau** d_x de um vértice x é o número de vizinhos que ele possui.

Dado um grafo $G = (V, E)$ como definido acima, existe uma cadeia de Markov com espaço de estados V naturalmente associada, cuja matriz de transição é dada por

$$P(x, y) := \begin{cases} \frac{1}{d_x}, & \text{se } y \sim x \\ 0, & \text{caso contrário} \end{cases}. \quad (1.15)$$

A dinâmica desta cadeia caracteriza-se por, estando em um estado $x \in V$, escolher um dos caminhos (arestas) disponíveis e então seguir por ele até o próximo vértice. Tal cadeia é conhecida como **passeio aleatório simples em G** .

A matriz P do passeio aleatório simples no grafo G , definida em (1.15), é reversível com respeito à medida $\mu(x) = d_x$. De fato, se $y \sim x$,

$$\mu(x)P(x, y) = d_x \cdot \frac{1}{d_x} = 1 = d_y \cdot \frac{1}{d_y} = \mu(y)P(y, x),$$

ao passo que se $y \not\sim x$, então $\mu(x)P(x, y) = 0 = \mu(y)P(y, x)$. Em qualquer caso, estão satisfeitas as equações do balanço detalhado (1.11). Em particular, conforme vimos na Seção 1.8, isso garante que μ é invariante para P . Além disso, como $\sum_{y \in V} \mu(y) = \sum_{y \in V} d_y = 2|E| < \infty$, podemos normalizar a medida μ para obter

$$\pi(x) = \frac{\mu(x)}{\sum_{y \in V} \mu(y)} = \frac{d_x}{2|E|} \quad (1.16)$$

uma distribuição invariante para P .

O grafo $G = (V, E)$ é dito **conexo** quando dados dois vértices quaisquer, é possível ir de um até outro através de uma sequência de arestas. É imediato da definição que G é conexo se, e somente se o passeio aleatório simples em G definido pela matriz P em (1.15) é irredutível.

A definição de grafo dada acima é a mais simples possível, mas pode ser estendida de modo a permitir *loops* (arestas que ligam um vértice a ele mesmo) e *arestas paralelas* (quando mais de uma aresta conecta um único par de vértices x, y). Permitindo tais elementos, o passeio aleatório simples no grafo continua com a mesma interpretação de, estando em um vértice

x , escolher aleatoriamente e uniformemente um caminho para seguir, mas não pode mais ser definido com $P(x, y) = 1/d_x$, uma vez que a quantidade de caminhos possíveis saindo do vértice x pode ser maior que d_x (já que loops e arestas paralelas são opções a seguir, mas não alteram o grau do vértice). Uma alternativa seria definir um conceito como *grau efetivo* de um vértice, que conta quantas arestas saem de x , e colocar $P(x, y) = 1/(\text{grau efetivo})$. Entretanto, não vamos dar atenção a tais sutilezas aqui, uma vez que elas serão resolvidas de modo mais simples no Capítulo 3 quando tivermos definido o conceito de redes elétricas.

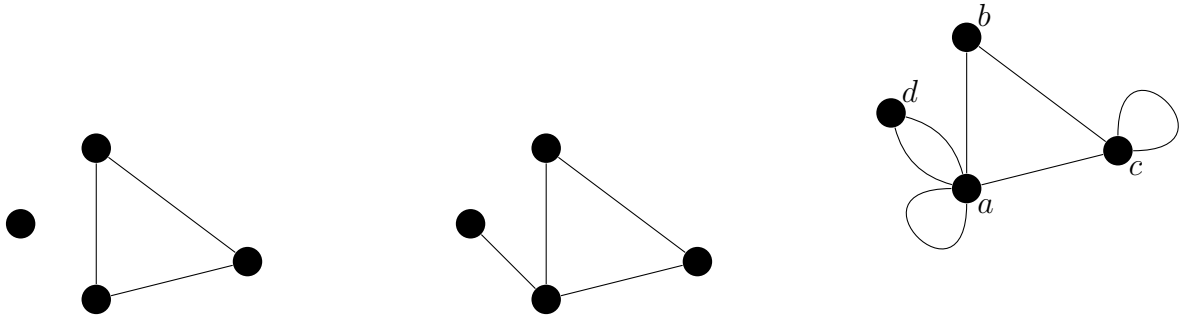


Figura 1.1: Da esquerda para a direita: grafo não-conexo, grafo conexo, e grafo com loops e arestas paralelas.

Na Figura 1.1 $d_a = 3$, mas quando se está no estado a existem 5 possíveis caminhos a seguir. Neste caso particular, o passeio aleatório simples neste grafo tem matriz de transição

$$P = \begin{bmatrix} 1/5 & 1/5 & 1/5 & 2/5 \\ 1/2 & 0 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

Observação 1.29. *Se ao estar em um vértice x escolhermos o caminho a seguir dando pesos diferentes a cada caminho (ao invés de escolher uniformemente entre os caminhos possíveis), teremos um passeio aleatório ponderado no grafo. Este conceito será melhor explorado quando definirmos o conceito de redes no Capítulo 3, mas poderia ser definido sem o auxílio destas ferramentas.*

1.10 Representação via sistema dinâmico aleatório

Considere o passeio aleatório simples no n -ciclo visto na Seção 6.1.3, cuja matriz de transição é

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

Seja Z uma variável aleatória que assume os valores 1 e -1 com igual probabilidade, isto é, $\mathbb{P}(Z = -1) = \mathbb{P}(Z = 1) = 1/2$. Se a cadeia encontra-se no estado $x \in \mathbb{Z}_n$, então o próximo estado será $x + Z \pmod{n}$, ou seja, para todo $y \in \mathbb{Z}_n$,

$$\mathbb{P}(x + Z \pmod{n} = y) = P(x, y).$$

Mais que isso, se considerarmos Z_1, Z_2, \dots uma sequência de variáveis aleatórias independentes, todas elas com a mesma distribuição de Z , e definirmos recursivamente a cadeia

$$Y_t := Y_{t-1} + Z_t \pmod{n},$$

então a dinâmica desta cadeia será a mesma do passeio aleatório simples no n -ciclo.

Observação 1.30. *A recorrência não está bem definida pois não definimos Y_0 . Mas isso não altera a dinâmica, apenas a distribuição inicial. Mais precisamente, se definirmos que Y_0 possui distribuição μ , então $(Y_t)_{t \geq 0}$ vai ser o passeio aleatório no n -ciclo com distribuição inicial μ .*

O que fizemos acima com o passeio aleatório no n -ciclo pode ser feito com qualquer cadeia em um espaço de estados finito, e é conhecida como **representação via sistema dinâmico aleatório**. Formalizemos:

Definição 1.31. *Uma representação via sistema dinâmico aleatório de uma cadeia de Markov com matriz de transição P no espaço de estados S é uma função $f : S \times \Lambda \rightarrow S$, juntamente com uma variável aleatória Z assumindo valores em Λ , satisfazendo*

$$\mathbb{P}(f(x, Z) = y) = P(x, y).$$

É fácil verificar que se Z_1, Z_2, \dots é uma sequência de variáveis aleatórias independentes, cada uma com distribuição Z , e se definirmos a variável aleatória X_0 como tendo distribuição μ e sendo independente de $(Z_t)_{t \geq 1}$, então a sequência (X_0, X_1, \dots) definida para todo $t \geq 1$ por

$$X_t = f(X_{t-1}, Z_t) \tag{1.17}$$

é uma cadeia de Markov com matriz de transição P e distribuição inicial μ . De fato, pela independência,

$$\mathbb{P}(X_{t+1} = y | X_t = x) = \mathbb{P}(f(X_t, Z) = y | X_t = x) = \mathbb{P}(f(x, Z) = y) = P(x, y).$$

Em particular, isso mostra que toda representação via sistema dinâmico aleatório de uma matriz de transição P gera, através da definição (1.17), uma cadeia de Markov com matriz de transição P . Além disso, toda matriz de transição possui ao menos uma representação via sistema dinâmico aleatório. Para demonstrar tal afirmação consideramos P uma matriz de transição em um espaço de estados finito $S = \{x_1, \dots, x_n\}$, tomamos $\Lambda = [0, 1]$ e Z a variável uniforme neste intervalo. Colocamos $F_{j,k} := \sum_{i=1}^k P(x_j, x_i)$, e definimos

$$f(x_j, z) := x_k \quad \text{se } F_{j,k-1} < z \leq F_{j,k}.$$

Então temos que

$$\mathbb{P}(f(x_j, Z) = x_k) = \mathbb{P}(F_{j,k-1} < Z \leq F_{j,k}) = P(x_j, x_k),$$

mostrando que tal f , juntamente com a variável aleatória Z uniforme em Λ , é uma representação via sistema dinâmico aleatório de P .

Para finalizar esta seção, observamos que a representação via sistema dinâmico aleatório é uma maneira muito útil para descrever uma cadeia, uma vez que fica claro como ela evolui de um instante de tempo para o seguinte. Além disso, essa evolução vem com base em uma fonte de aleatoriedade externa Z que independe da cadeia. Este último fato será extremamente útil ao construirmos acoplamentos no próximo capítulo, onde estaremos interessados em duas cadeias com mesma matriz de transição P que começam com distribuições iniciais diferentes, e para isso, basta que as atualizemos ambas as cadeias pela mesma fonte de aleatoriedade Z .

1.11 Cadeias lazy

Aqui apresentaremos duas definições que serão amplamente utilizadas neste texto. Primeiramente dizemos que uma cadeia P é **lazy** se $P(x, x) \geq 1/2$ para todo $x \in S$. Intuitivamente isso

significa que a cada salto, temos ao menos 50% de probabilidade de não mudarmos de estado, e por isso o nome *preguiçosa*.

Além disso, dada uma matriz de transição P , se I for a matriz identidade $|S| \times |S|$, então a matriz $Q = \frac{P+I}{2}$ também será estocástica. A cadeia de Markov com matriz de transição Q é denominada **versão lazy de P** . Sua interpretação é bem simples: a cada passo joga uma moeda, se o resultado for cara, move-se de acordo com a dinâmica da P , já se o resultado for coroa, fica parado no mesmo lugar.

É imediato da álgebra matricial que toda cadeia lazy Q é também a versão lazy da cadeia $P = 2Q - I$. Para um primeiro exemplo de cadeia lazy, visite se Seção 6.1.3.

Versões lazy de uma cadeia mantém a informação sobre a dinâmica da cadeia original, além de serem sempre aperiódicas, já que pela definição $Q(x, x) \geq 1/2 > 0$ para todo $x \in S$, logo $Q^t(x, x) > 0$ para todo $t \geq 1$ e portanto $d(x) = \text{mdc } \mathcal{T}(x) = 1$.

Capítulo 2

Teorema de Convergência e Tempos de Mistura

O capítulo anterior foi dedicado à apresentação de alguns resultados gerais sobre cadeias de Markov. Neste capítulo definiremos o primeiro indicador das cadeias de Markov, o tempo de mistura, o qual mede o tempo necessário para que a distribuição da cadeia esteja suficientemente próxima da distribuição estacionária.

Para que essa ideia de tempo de mistura faça sentido é necessário que a distribuição da cadeia no tempo t convirja para alguma distribuição. Isso será garantido pelo teorema de convergência (Teorema 2.14). Além disso, para falarmos no conceito de proximidade, precisamos escolher uma distância no espaço das distribuições. As opções são muitas, conforme veremos no final deste capítulo, mas uma das mais úteis é a distância de variação total.

2.1 Distância de variação total

Definição 2.1. A distância de variação total entre duas distribuições de probabilidade μ e ν em S é definida por

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

De acordo com a definição ela é a maior diferença entre as probabilidades que as duas distribuições associam a um evento simples. Nas palavras de [1], “é o maior erro que pode cometer ao usar a ‘distribuição errada’ para calcular a probabilidade de um evento”.

Utilizar a definição não é o jeito mais simples de calcular a distância de variação total entre duas distribuições. Para isso temos o seguinte resultado.

Proposição 2.2. Sejam μ e ν distribuições de probabilidade em S , então

$$\|\mu - \nu\|_{TV} = \sum_{\substack{x \in S, \\ \mu(x) \geq \nu(x)}} [\mu(x) - \nu(x)] = \frac{1}{2} \sum_{x \in S} |\mu(x) - \nu(x)|$$

Demonstração. Sejam $B := \{x \in S; \mu(x) \geq \nu(x)\}$ e $A \subset S$ um evento qualquer. Temos que

$$\mu(A) - \nu(A) \leq \mu(A \cap B) - \nu(A \cap B). \quad (2.2)$$

Para ver isso observamos que da definição de B , $\mu(x) - \nu(x) < 0$ para todo $x \in A \cap B^C$, logo

$$\mu(A \cap B^C) - \nu(A \cap B^C) = \sum_{x \in A \cap B^C} \mu(x) - \sum_{x \in A \cap B^C} \nu(x) = \sum_{x \in A \cap B^C} [\mu(x) - \nu(x)] < 0.$$

Portanto, utilizando que o evento A se decompõe como $A = (A \cap B) \dot{\cup} (A \cap B^C)$,

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

Também temos que

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

uma vez que pela definição de B , $\mu(x) - \nu(x) \geq 0$ para todo $x \in B$, de modo que adicionar termos deste tipo não faz um somatório decrescer, e assim

$$\mu(A \cap B) - \nu(A \cap B) = \sum_{x \in A \cap B} [\mu(x) - \nu(x)] \leq \sum_{x \in B} [\mu(x) - \nu(x)] = \mu(B) - \nu(B).$$

Das equações (2.2) e (2.3) temos que

$$\mu(A) - \nu(A) \leq \mu(A \cap B) - \nu(A \cap B) \leq \mu(B) - \nu(B). \quad (2.4)$$

Por um raciocínio análogo podemos mostrar que

$$\nu(A) - \mu(A) \leq \nu(B^C) - \mu(B^C). \quad (2.5)$$

Das duas equações anteriores temos que $M := \mu(B) - \nu(B)$ e $N := \nu(B^C) - \mu(B^C)$ são duas cotas superiores para $|\mu(A) - \nu(A)|$, qualquer que seja $A \subset S$, logo são, em virtude da definição (2.1), cotas superiores para $\|\mu - \nu\|_{TV}$. Mas note que estas duas cotas são iguais, uma vez que sendo μ e ν distribuições de probabilidade,

$$M - N = [\mu(B) - \nu(B)] - [\nu(B^C) - \mu(B^C)] = [\mu(B) + \mu(B^C)] - [\nu(B) + \nu(B^C)] = \mu(S) - \nu(S) = 0.$$

Além disso, de acordo com as equações (2.4) e (2.5), tal cota é atingida por $|\mu(A) - \nu(A)|$ quando $A = B$ (ou $A = B^C$). Isso tudo nos garante que

$$\|\mu - \nu\|_{TV} = \max_{A \subset S} |\mu(A) - \nu(A)| = M = N. \quad (2.6)$$

Em particular, utilizando a definição do conjunto B , segue de (2.6) que

$$\|\mu - \nu\|_{TV} = M = \mu(B) - \nu(B) = \sum_{x \in B} [\mu(x) - \nu(x)] = \sum_{\substack{x \in S, \\ \mu(x) \geq \nu(x)}} [\mu(x) - \nu(x)],$$

mostrando a primeira das igualdades desejadas.

Além disso, como $M = N$, vale $M = (M + N)/2$, isto é,

$$M = \frac{1}{2} [\mu(B) - \nu(B) + \nu(B^C) - \mu(B^C)] = \frac{1}{2} \left[\sum_{x \in B} [\mu(x) - \nu(x)] + \sum_{x \in B^C} [\nu(x) - \mu(x)] \right].$$

Como pela definição de B , $\mu(x) - \nu(x) \geq 0$ para todo $x \in B$ e $\nu(x) - \mu(x) \geq 0$ para todo $x \in B^C$, o lado direito acima é igual a

$$\frac{1}{2} \left[\sum_{x \in B} |\mu(x) - \nu(x)| + \sum_{x \in B^C} |\mu(x) - \nu(x)| \right] = \frac{1}{2} \sum_{x \in S} |\mu(x) - \nu(x)|,$$

de modo que obtemos, a partir da equação (2.6), que

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

□

A distância de variação total, assim como se espera de uma distância, satisfaz a desigualdade triangular. Isso é consequência imediata da Proposição 2.2 juntamente com a desigualdade triangular para números reais.

Corolário 2.3. *Sejam μ , ν e η distribuições sobre S . Então*

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

Outro resultado interessante para calcular a distância de variação total é dado pela seguinte proposição. Como não o utilizaremos neste trabalho, optamos por omitir sua prova, que pode ser encontrada em [9, Proposição 4.5].

Proposição 2.4. *Sejam μ e ν distribuições sobre S . Então elas satisfazem*

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sup \left\{ \sum_{x \in S} f(x)\mu(x) - \sum_{x \in S} f(x)\nu(x); \max_{x \in S} |f(x)| \leq 1 \right\}$$

2.2 Acoplamentos

Acoplamentos constituem uma ferramenta poderosíssima, conforme observaremos neste capítulo. Começamos esta seção definindo o conceito de acoplamento de duas distribuições, para então estender tal ideia para acoplar cadeias de Markov.

Definição 2.5. *Dadas duas distribuições de probabilidade μ e ν , um **acoplamento** entre elas é um par de variáveis aleatórias (X, Y) definidas no mesmo espaço de probabilidade, de modo que X tem distribuição marginal μ enquanto que Y tem distribuição marginal ν , isto é, $\mathbb{P}(X = x) = \mu(x)$ e $\mathbb{P}(Y = y) = \nu(y)$.*

Note que não há restrições quanto ao comportamento conjunto das variáveis aleatórias do acoplamento, dando-nos bastante liberdade para defini-las. Isso fica claro quando olhamos para o seguinte exemplo.

Exemplo 2.6. *Considere que μ e ν são as distribuições de duas moedas honestas no conjunto $\{0, 1\}$, isto é, $\mu(0) = \mu(1) = \frac{1}{2}$ e $\nu(0) = \nu(1) = \frac{1}{2}$. Temos (ao menos) dois acoplamentos para μ e ν :*

1. *Define (X, Y) como um par de variáveis aleatórias independentes, cada uma representando o lançamento de uma moeda honesta, assim:*

$$\begin{aligned} \mathbb{P}(X = x) &= \frac{1}{2} = \mu(x), & \forall x \in \{0, 1\} \\ \mathbb{P}(Y = y) &= \frac{1}{2} = \nu(y), & \forall y \in \{0, 1\}. \end{aligned}$$

Pela independência, o comportamento conjunto do par (X, Y) é dado por

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y) = \frac{1}{4},$$

para todo $x, y \in \{0, 1\}$.

2. Defina X como sendo o lançamento de uma moeda honesta, ou seja, $\mathbb{P}(X = x) = \frac{1}{2}$ para todo $x \in \{0, 1\}$, e defina $Y = X$. Neste caso,

$$\begin{aligned}\mathbb{P}(X = x) &= \frac{1}{2} = \mu(x), & \forall x \in \{0, 1\} \\ \mathbb{P}(Y = y) &= \mathbb{P}(X = y) = \frac{1}{2} = \nu(y), & \forall y \in \{0, 1\}.\end{aligned}$$

Diferentemente de antes, o comportamento conjunto deste par é dado por

$$\begin{aligned}\mathbb{P}(X = 0, Y = 0) &= \frac{1}{2} \\ \mathbb{P}(X = 1, Y = 1) &= \frac{1}{2} \\ \mathbb{P}(X = 0, Y = 1) &= \mathbb{P}(X = 1, Y = 0) = 0.\end{aligned}$$

Esses são dois possíveis acoplamentos para as distribuições μ e ν , mas podem existir muitos outros. A única coisa que podemos afirmar trivialmente é que o acoplamento independente como em (i) sempre existe. E assim, tendo garantida a existência de ao menos um acoplamento, fica bem enunciado o seguinte teorema.

Proposição 2.7. *Sejam μ e ν duas distribuições em S , então*

$$\|\mu - \nu\|_{TV} = \inf \{ \mathbb{P}(X \neq Y); (X, Y) \text{ é um acoplamento de } \mu \text{ e } \nu \}. \quad (2.7)$$

Demonstração. Começamos observando que para qualquer acoplamento (X, Y) de μ e ν , e para qualquer evento $A \subset S$, vale

$$\mathbb{P}(X \in A) = \mathbb{P}(X \in A, Y \in A) + \mathbb{P}(X \in A, Y \notin A) \leq \mathbb{P}(Y \in A) + \mathbb{P}(X \in A, Y \notin A),$$

logo

$$\mu(A) - \nu(A) = \mathbb{P}(X \in A) - \mathbb{P}(Y \in A) \leq \mathbb{P}(X \in A, Y \notin A) \leq \mathbb{P}(X \neq Y).$$

Então, de acordo com a definição de $\|\mu - \nu\|_{TV}$ em (2.1), temos que

$$\|\mu - \nu\|_{TV} = \max_{A \subset S} |\mu(A) - \nu(A)| \leq \inf \{ \mathbb{P}(X \neq Y); (X, Y) \text{ é um acoplamento de } \mu \text{ e } \nu \}.$$

Assim sendo, para concluirmos a demonstração, basta construirmos um acoplamento (X, Y) de μ e ν tal que $\mathbb{P}(X \neq Y)$ seja exatamente $\|\mu - \nu\|_{TV}$.

Para isso considerarmos a Figura 2.1, em especial as regiões I, II e III em $S \times \mathbb{R}$. Observe que de acordo com a Proposição 2.2,

$$\begin{aligned}\text{Área I} &= \sum_{x \in B} [\mu(x) - \nu(x)] = \mu(B) - \nu(B) = \|\mu - \nu\|_{TV}, \\ \text{Área II} &= \sum_{x \in B} [\nu(x) - \mu(x)] = \nu(B^C) - \mu(B^C) = \|\mu - \nu\|_{TV}, \\ \text{Área III} &= 1 - \text{Área I} = 1 - \|\mu - \nu\|_{TV}.\end{aligned}$$

Escolhemos aleatoriamente, de maneira uniforme, um ponto Q_1 na união das regiões I e III, e definimos X como a coordenada x de Q_1 . Agora

- (i) Se Q_1 pertence à região III, colocamos $Y = X$;
- (ii) Se Q_1 pertence à região I, então escolhemos aleatoriamente, de maneira uniforme e independente de Q_1 , um ponto Q_2 da região II, e tomamos Y como sendo a coordenada x de Q_2 .

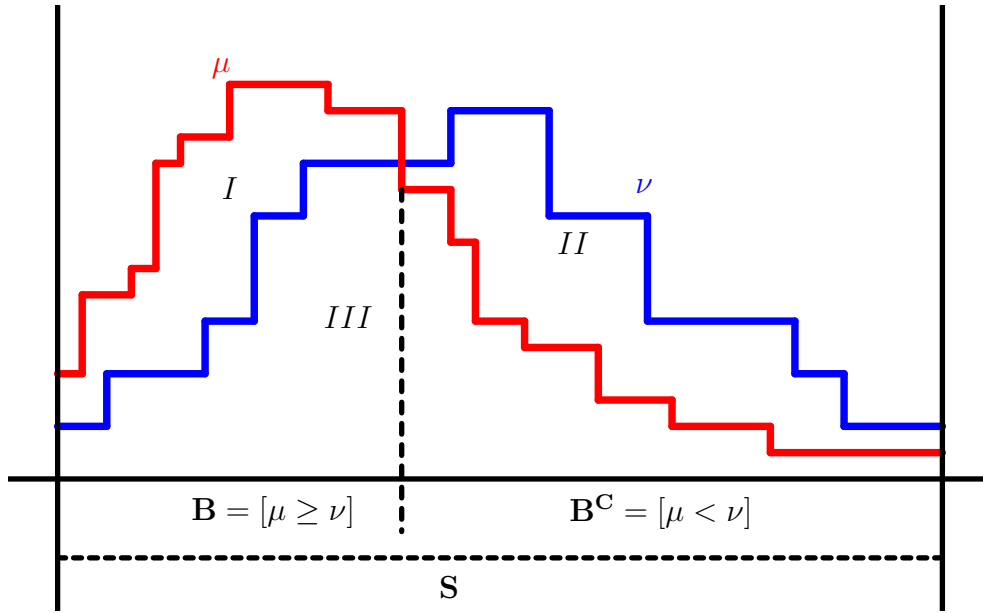


Figura 2.1: Distribuições μ e ν . Reordenamos o espaço de estados S finito a fim de separar os conjuntos $B = \{x \in S; \mu(x) \geq \nu(x)\}$ e B^C .

Pela figura fica fácil observar que as regiões I e II são “disjuntas no eixo x ”, no sentido de que um ponto que pertence à região I tem coordenada x diferente de qualquer ponto que pertença à região II. Isso faz com que $X \neq Y$ no caso (ii). Assim $\mathbb{P}(X = Y) = \text{Área III} = 1 - \|\mu - \nu\|_{TV}$, e $\mathbb{P}(X \neq Y) = \text{Área I} = \|\mu - \nu\|_{TV}$.

Por fim note que como a construção de X e Y é baseada nas distribuições de μ e ν temos que $\mathbb{P}(X = x) = \mu(x)$.

Para mostrar que $\mathbb{P}(Y = y) = \nu(y)$ comecemos observando que

$$\mathbb{P}(X = Y = y) = \mu(y) \wedge \nu(y)$$

pois essa probabilidade corresponde à probabilidade do ponto Q_1 ser escolhido na região III e ter abscissa igual a y . Agora consideremos dois casos. Se $\mu(y) \geq \nu(y)$ temos que

$$\begin{aligned} \mathbb{P}(Y = y) &= \mathbb{P}(Y = y, X = Y) + \mathbb{P}(Y = y, X \neq Y) \\ &= \mu(y) \wedge \nu(y) + 0 = \nu(y). \end{aligned}$$

Já se $\mu(y) < \nu(y)$, considerando $p = \text{Área III}$, temos que

$$\begin{aligned} \mathbb{P}(Y = y) &= \mathbb{P}(Y = y, X = Y) + \mathbb{P}(Y = y, X \neq Y) \\ &= \mu(y) \wedge \nu(y) + \mathbb{P}(X \neq Y) \mathbb{P}(Y = y | Y \neq X) \\ &= \mu(y) + (1 - p) \frac{\nu(y) - \mu(y)}{1 - p} = \nu(y) \end{aligned}$$

Portanto, mostramos que existe um acoplamento (X, Y) satisfazendo $\mathbb{P}(X \neq Y) = \|\mu - \nu\|_{TV}$, o que conclui a demonstração. \square

Na construção acima é fácil nos convenceremos que $\mathbb{P}(X = x) = \mu(x)$, mas nem tanto de aceitarmos que $\mathbb{P}(Y = y) = \nu(y)$. Baseando-nos em [9, Proposição 4.7], podemos fornecer uma construção mais rigorosa deste acoplamento que satisfaz $\mathbb{P}(X \neq Y) = \|\mu - \nu\|_{TV}$.

Demonstração alternativa. Começamos definindo o número real $p := 1 - \|\mu - \nu\|_{TV}$. Tal valor é exatamente a área da região III, logo pode ser expresso também por $\sum_{x \in S} \mu(x) \wedge \nu(x)$. Então jogamos uma moeda com probabilidade p de sair cara.

(i) Se sair cara, escolhamos um valor Z aleatoriamente com distribuição

$$\gamma_{III}(x) = \frac{\mu(x) \wedge \nu(x)}{p},$$

e então colocamos $X = Y = Z$.

(ii) Se sair coroa, sorteamos um valor para X de acordo com a distribuição

$$\gamma_I(x) = \begin{cases} \frac{\mu(x) - \nu(x)}{\|\mu - \nu\|_{TV}}, & \text{se } \mu(x) > \nu(x) \\ 0, & \text{caso contrário} \end{cases},$$

e independentemente, sorteamos um valor para Y com distribuição

$$\gamma_{II}(x) = \begin{cases} \frac{\nu(x) - \mu(x)}{\|\mu - \nu\|_{TV}}, & \text{se } \nu(x) > \mu(x) \\ 0, & \text{caso contrário} \end{cases}.$$

A proposição 2.2 garante que γ_I e γ_{II} são distribuições.

Note que, saindo coroa, $X \neq Y$ uma vez que γ_I e γ_{II} são positivas em subconjuntos diferentes de S . Assim $X = Y$ se e somente se sai cara na moeda, ou seja, $\mathbb{P}(X = Y) = p = 1 - \|\mu - \nu\|_{TV}$ e $\mathbb{P}(X \neq Y) = 1 - p = \|\mu - \nu\|_{TV}$.

Além disso, observe que o evento $[X = x]$ pode ocorrer de duas maneiras: saindo cara (com probabilidade p) e então obtendo x através da distribuição γ_{III} , ou saindo coroa (com probabilidade $1 - p$) e neste caso obtendo x através da distribuição γ_I . Utilizando o mesmo raciocínio para o evento $[Y = x]$, o qual depende da moeda e das distribuições γ_{III} e γ_{II} , obtemos que

$$\begin{aligned} \mathbb{P}(X = x) &= p \cdot \gamma_{III}(x) + (1 - p) \cdot \gamma_I(x) = \mu(x) \\ \mathbb{P}(Y = x) &= p \cdot \gamma_{III}(x) + (1 - p) \cdot \gamma_{II}(x) = \nu(x), \end{aligned}$$

onde a última igualdade em cada uma das equações é consequência direta das definições de γ_I , γ_{II} e γ_{III} .

Isso mostra que (X, Y) é um acoplamento de μ e ν que satisfaz $\mathbb{P}(X \neq Y) = \|\mu - \nu\|_{TV}$, como queríamos. \square

Acima definimos o acoplamento de duas distribuições μ e ν como um par de variáveis aleatórias (X, Y) satisfazendo $\mathbb{P}(X = x) = \mu(x)$ e $\mathbb{P}(Y = y) = \nu(y)$. Tal ferramenta apresentou-se extremamente útil para obtenção de cotas para distância de variação total entre estas medidas seguindo a Proposição 2.7. Nosso objetivo agora é estender este conceito, acoplando não mais duas distribuições, mas sim duas trajetórias de uma mesma cadeia de Markov.

Definição 2.8. *Dada uma matriz estocástica P (a fonte de aleatoriedade de uma cadeia de Markov), um **acoplamento de cadeias de Markov** com matriz de transição P é um processo $(X_t, Y_t)_{t \geq 0}$ tal que $(X_t)_{t \geq 0}$ e $(Y_t)_{t \geq 0}$ são duas cadeias de Markov com matriz de transição P .*

Note que nada é pedido em relação à distribuição inicial das cadeias $(X_t)_{t \geq 0}$ e $(Y_t)_{t \geq 0}$, assim como também não exigimos nada em relação ao comportamento conjunto de (X_t, Y_t) . Além disso, nada garante a priori que o processo $(X_t, Y_t)_{t \geq 0}$ seja uma cadeia de Markov no espaço de estados $S \times S$.

Assim como no acoplamento de distribuições, existem inúmeras maneiras de realizar o acoplamento de cadeias de Markov, com a escolha ideal dependendo dos objetivos envolvidos. Tal flexibilidade o torna uma ferramenta extremamente útil ao permitir enxergar um dado problema por outro ângulo, conforme fica claro no exemplo a seguir.

Exemplo 2.9. *Considere o passeio aleatório simples no segmento $\{0, 1, \dots, n\}$, isto é, a cadeia de Markov que se move para cima e para baixo com igual probabilidade, e que quando se encontra nos extremos e tenta se mover para fora do intervalo, fica parada. Sua matriz de transição é dada por:*

$$\begin{aligned} \text{Se } 0 < i < n, \text{ então} & \quad P(i, i-1) = P(i, i+1) = \frac{1}{2} \\ \text{Se } i = n, \text{ então} & \quad P(i, i-1) = P(i, i) = \frac{1}{2} \\ \text{Se } i = 0, \text{ então} & \quad P(i, i) = P(i, i-1) = \frac{1}{2} \end{aligned} \quad (2.8)$$

Vamos mostrar que se $x \leq y$ então $P^t(x, n) \leq P^t(y, n)$.

A abordagem será fixar $x \leq y$ e construir um acoplamento, ou seja, duas cadeias de Markov no mesmo espaço de probabilidade e com a mesma matriz P definida em (2.8), mas com uma delas começando em x e nos dando informação sobre $P^t(x, n)$, enquanto que a outra começa em y e nos dá informação sobre $P^t(y, n)$. Além disso vamos construí-las de modo com que possamos obter uma relação entre estas quantidades.

Para isso, fixados $x \leq y$, consideramos $\Delta_1, \Delta_2, \dots$ uma sequência de variáveis aleatória i.i.d (independentes e identicamente distribuídas) com

$$\mathbb{P}(\Delta_i = 1) = \frac{1}{2} = \mathbb{P}(\Delta_i = -1)$$

Definimos a primeira cadeia (X_t) começando em x e a segunda, (Y_t) , começando em y , com ambas seguindo a mesma regra de atualização:

- Se $\Delta_t = +1$, então as cadeias movem-se para cima, se possível, ou então, se estiverem no extremo superior, ficam paradas;
- Se $\Delta_t = -1$, então as cadeias movem-se para baixo, se possível, ou então, se estiverem no extremo inferior, ficam paradas.

Assim fica claro que a matriz de transição de ambas as cadeias é a matriz P definida em (2.8), logo a probabilidade da cadeia $(X_t)_{t \geq 0}$ estar no tempo t no estado n é $\mathbb{P}(X_t = n) = P^t(x, n)$, enquanto que a probabilidade da cadeia $(Y_t)_{t \geq 0}$ estar no tempo t no estado n é $\mathbb{P}(Y_t = n) = P^t(y, n)$.

Outra propriedade que nos interessa deste acoplamento é que a cadeia $(X_t)_{t \geq 0}$ nunca assumirá valores maiores que a cadeia $(Y_t)_{t \geq 0}$. Isto ocorre pelo fato de $(X_t)_{t \geq 0}$ começar abaixo de $(Y_t)_{t \geq 0}$ e ambas serem atualizadas pela mesma regra, de modo que vão manter a mesma distância entre si até que uma fique parada em alguma das extremidades, momento em que a outra cadeia poderá se aproximar da primeira. Portanto $(X_t)_{t \geq 0}$ ficará sempre abaixo de $(Y_t)_{t \geq 0}$, até que elas eventualmente se encontrem e comecem a andar juntas, já que seguem a mesma regra de atualização.

Como consequência do fato anterior, sempre que $(X_t)_{t \geq 0}$ estiver no estado n , $(Y_t)_{t \geq 0}$ também vai estar, ou seja, $[X_t = n] \subset [Y_t = n]$. Portanto,

$$P^t(x, n) = \mathbb{P}[X_t = n] \leq \mathbb{P}[Y_t = n] = P^t(y, n)$$

mostrando a desigualdade afirmada anteriormente.

O leitor atento deve ter notado que pelo modo com que definimos as cadeias acima, estamos “por baixo dos panos” nos utilizando da representação via sistema dinâmico aleatório da cadeia.

Observação 2.10. Se $(X_t)_{t \geq 0}$ e $(Y_t)_{t \geq 0}$ são cadeias de Markov acopladas com $X_0 = x$ e $Y_0 = y$, vamos escrever $\mathbb{P}_{x,y}$ para a probabilidade no espaço das trajetórias em que ambas estão definidas.

O próximo teorema exemplifica o poder dos acoplamentos ao utilizá-los para obter um cota para a distância de variação total entre as distribuições $P^t(x, \cdot)$ e $P^t(y, \cdot)$.

Proposição 2.11. Sejam P uma matriz de transição no espaço de estados S e $(X_t, Y_t)_{t \geq 0}$ um acoplamento de cadeias de Markov com $X_0 = x$ e $Y_0 = y$, construído de modo que a partir do momento que as cadeias $(X_t)_{t \geq 0}$ e $(Y_t)_{t \geq 0}$ se encontram elas seguem juntas para sempre, isto é, tal que se $X_s = Y_s$ então para todo $t \geq s$, $X_t = Y_t$. Definimos o tempo de coalescência das duas cadeias por

$$\tau_{couple} = \min\{t; X_s = Y_s \ \forall s \geq t\} \quad (2.9)$$

Então, para todo t fixo, vale

$$\|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \leq \mathbb{P}_{x,y}[\tau_{couple} > t]$$

Demonstração. Note que pela definição das cadeias $(X_t)_{t \geq 0}$ e $(Y_t)_{t \geq 0}$, $P^t(x, \cdot) = \mathbb{P}_{x,y}(X_t = \cdot)$ e $P^t(y, \cdot) = \mathbb{P}_{x,y}(Y_t = \cdot)$, de modo que para cada t fixo, o par de variáveis aleatórias (X_t, Y_t) é um acoplamento das distribuições $P^t(x, \cdot)$ e $P^t(y, \cdot)$, no sentido definido da Definição 2.5. Consequentemente a Proposição 2.7 garante que

$$\|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \leq \mathbb{P}_{x,y}(X_t \neq Y_t).$$

Como escolhemos os acoplamentos de modo tal que a partir do momento que eles se encontram vão andar juntos para sempre, temos que $\mathbb{P}_{x,y}(\tau_{couple} > t) = \mathbb{P}_{x,y}(X_t \neq Y_t)$, valendo portanto

$$\|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \leq \mathbb{P}_{x,y}(\tau_{couple} > t).$$

□

Observação 2.12. Já havíamos passado por isto no Exemplo 2.9, mas na demonstração acima fica ainda mais clara uma relação fundamental entre os acoplamentos de distribuições e os acoplamentos de cadeias de Markov: Se $(X_t, Y_t)_{t \geq 0}$ for um acoplamento de cadeias de Markov com matriz de transição P , tal que $X_0 = x$ e $Y_0 = y$, então para cada t fixo, o par de variáveis aleatórias (X_t, Y_t) é um acoplamento das distribuições $P^t(x, \cdot)$ e $P^t(y, \cdot)$.

2.3 Teorema de convergência

Cadeias de Markov são caracterizadas pela perda de memória, ou seja, a cada passo esquecer completamente o passado e levar em conta apenas onde está e para onde pode saltar (veja Seção 1.3). Deste modo é natural que com o passar do tempo perca-se a noção de onde a cadeia começou, ficando apenas com a informação da dinâmica. Faz sentido então que se a distribuição da cadeia no tempo t convergir para alguma distribuição específica, esta dependa exclusivamente da matriz P . Mostraremos nesta seção que se estiverem satisfeitas as condições necessárias para que haja convergência, ela se dará para a distribuição invariante da matriz de transição P definida anteriormente. Tal resultado é conhecido como **teorema de convergência**, e possui uma versão mais geral para espaços de estados enumeráveis, e uma versão mais fina para espaços de estados finitos, onde obtemos uma cota para a velocidade dessa convergência. Vamos dedicar esta seção a apresentar estas duas versões, juntamente com algumas demonstrações delas. Porém começamos apresentando um exemplo em que não pode haver convergência.

Exemplo 2.13. *Considere o passeio aleatório simples no n -ciclo (apresentado na Seção 6.1.3) com n par, isto é, $P(x, y) = P(y, x) = \frac{1}{2}$ se $y \equiv x \pm 1 \pmod{n}$. Defina $X_0 = 0$. Neste caso temos que $X_1 = 1$ ou $X_1 = n - 1$, em qualquer caso, X_1 vai ser ímpar. Seguindo esta análise X_2 vai ser um estado par, e assim por diante. Em particular, $\mathbb{P}_0(X_t = 0) = 0$ para t ímpar, e $\mathbb{P}_0(X_t = 0) > 0$ para t par, de modo que o estado 0 tem período 2. Mais que isso, é intuitivo (e possível mostrar) que $\mathbb{P}_0(X_t = 0)$ não se aproxima de 0 quanto t aumenta, de modo que $\mathbb{P}_0(X_t = 0)$ não pode convergir quando $t \rightarrow \infty$.*

Este é um exemplo em que não vale a conclusão do teorema de convergência que apresentaremos a seguir. Sua falha está na ausência da hipótese de aperiodicidade, que quando satisfeita fornece o seguinte resultado:

Teorema 2.14 (Teorema de Convergência). *Seja P uma cadeia irreduzível e aperiódica, com distribuição invariante π . Então para quaisquer estados $x, y \in S$,*

$$\lim_{t \rightarrow \infty} P^t(x, y) = \pi(y).$$

Demonstração. Fixamos um estado $x \in S$ e uma cadeia $(X_t)_{t \geq 0}$ Markov- (δ_x, P) . Tomamos $(Y_t)_{t \geq 0}$ outra cadeia de Markov em S , independente da primeira, com mesma matriz de transição P e com distribuição inicial π .

O processo $(X_t, Y_t)_{t \geq 0}$ é um acoplamento de cadeias de Markov com matriz de transição P . Além disso, ele é também uma cadeia de Markov em $S \times S$ com matriz de transição

$$\tilde{P}((x_1, y_1), (x_2, y_2)) = P(x_1, x_2)P(y_1, y_2), \quad (2.10)$$

que começa com $X_0 = x$ e $Y_0 \sim \pi$. Seja $\tilde{\mathbb{P}} = \tilde{\mathbb{P}}_{\delta_x \times \pi}$ a probabilidade no espaço das trajetórias de $S \times S$ induzida por esta cadeia.

Começamos observando que \tilde{P} é irreduzível. Para isso, fixamos dois estados $(x_1, y_1), (x_2, y_2) \in S \times S$. Pela irreduzibilidade da P temos que existem $k, l > 0$ tais que $P^k(x_1, x_2) > 0$ e $P^l(y_1, y_2) > 0$. Além disso, como x_1 e y_1 são aperiódicos, o Lema 1.13 garante que existem $t_1, t_2 \in \mathbb{N}$ tais que para todo $t \geq t_1$, $P^t(x_1, x_1) > 0$, e para todo $t \geq t_2$, $P^t(y_1, y_1) > 0$. Assim,

$$\begin{aligned} \tilde{P}^{k+l+t_1+t_2}((x_1, y_1), (x_2, y_2)) &= P^{k+l+t_1+t_2}(x_1, x_2)P^{k+l+t_1+t_2}(y_1, y_2) \\ &\geq P^{l+t_1+t_2}(x_1, x_1)P^k(x_1, x_2)P^{k+t_1+t_2}(y_1, y_1)P^l(y_1, y_2) > 0 \end{aligned}$$

o que mostra que existe um $t = l + k + t_1 + t_2 > 0$ tal que $\tilde{P}^t((x_1, y_1), (x_2, y_2)) > 0$, e como isso vale para qualquer par de estados em $S \times S$, \tilde{P} é irreduzível.

Também temos que \tilde{P} possui distribuição invariante $\tilde{\pi}(x, y) = \pi(x)\pi(y)$. Para isto basta observarmos que para todo estado $(x, y) \in S \times S$, como π é invariante para P ,

$$\begin{aligned} (\tilde{\pi}\tilde{P})(x, y) &= \sum_{(u,v) \in S \times S} \tilde{\pi}(u, v)\tilde{P}((u, v), (x, y)) \\ &= \sum_{u \in S} \sum_{v \in S} \pi(u)\pi(v)P(u, x)P(v, y) \\ &= \left[\sum_{u \in S} \pi(u)P(u, x) \right] \left[\sum_{v \in S} \pi(v)P(v, y) \right] \\ &= [(\pi P)(x)] [(\pi P)(y)] \\ &= \pi(x)\pi(y) \\ &= \tilde{\pi}(x, y) \end{aligned}$$

Como \tilde{P} é irredutível e possui distribuição invariante π , o Teorema 1.25 garante que todos os estados são recorrentes positivos, em particular recorrentes, logo para todo $z \in S$ vale que $\tilde{\mathbb{P}}_{(z,z)}(\tau_{(z,z)}^+ < \infty) = 1$.

Ao mesmo tempo, como \tilde{P} é irredutível, para todo par $(x, y) \in S \times S$, existe um $h > 0$ tal que $\tilde{P}^h((z, z), (x, y)) > 0$, e assim $\rho_{(z,z)}(x, y) = \tilde{\mathbb{P}}_{(z,z)}[\tau_{(x,y)}^+ < \infty] \geq \tilde{\mathbb{P}}_{(z,z)}[(X_h, Y_h) = (x, y)] > 0$. Daqui segue, pela proposição 1.19 que

$$\tilde{\mathbb{P}}_{(x,y)}[\tau_{(z,z)}^+ < \infty] = \rho_{(x,y)}(z, z) = 1.$$

Como isto vale para todo par $(x, y) \in S \times S$, vale que

$$\tilde{\mathbb{P}}_{\delta_x \times \pi}[\tau_{(z,z)}^+ < \infty] = 1.$$

Definindo $T = \inf\{t \geq 0; X_t = Y_t\}$ o primeiro tempo em que as cadeias $(X_t)_{t \geq 0}$ e $(Y_t)_{t \geq 0}$ se encontram, $[T < \infty] = \cup_{z \in S} [\tau_{(z,z)}^+ < \infty]$. E como união de eventos com probabilidade 1 também tem probabilidade 1, vale $\tilde{\mathbb{P}}[T < \infty] = 1$.

Como $(X_t)_{t \geq 0}$ e $(Y_t)_{t \geq 0}$ encontram-se no tempo T , pela perda de memória ambas as cadeias possuem a mesma distribuição para $t \geq T$, isto é $\tilde{\mathbb{P}}(X_t = y, T \leq t) = \tilde{\mathbb{P}}(Y_t = y, T \leq t)$. Para ver isso observamos que graças à propriedade de Markov e independência de $(X_t)_{t \geq 0}$ e $(Y_t)_{t \geq 0}$, $\tilde{\mathbb{P}}(X_s = y | X_r = x, T = r) = \tilde{\mathbb{P}}(X_s = y | X_r = x)$, além disso, como $(X_t)_{t \geq 0}$ e $(Y_t)_{t \geq 0}$ têm a mesma matriz de transição P , $\tilde{\mathbb{P}}(X_s = y | X_r = x) = \tilde{\mathbb{P}}(Y_s = y | Y_r = x)$, portanto somando sobre todos os tempos e todos os estados em que pode ocorrer o primeiro encontro das duas cadeias, obtemos

$$\begin{aligned} \tilde{\mathbb{P}}(X_t = y, T \leq t) &= \sum_{r=1}^t \sum_x \tilde{\mathbb{P}}(T = r, X_r = x, X_t = y) \\ &= \sum_{r=1}^t \sum_x \tilde{\mathbb{P}}(T = r, X_r = x) \tilde{\mathbb{P}}(X_t = y | X_r = x) \\ &= \sum_{r=1}^t \sum_x \tilde{\mathbb{P}}(T = r, Y_r = x) \tilde{\mathbb{P}}(Y_t = y | Y_r = x) \\ &= \tilde{\mathbb{P}}(Y_t = y, T \leq t). \end{aligned}$$

Assim temos que

$$\begin{aligned} \tilde{\mathbb{P}}(X_t = y) &= \tilde{\mathbb{P}}(X_t = y, T \leq t) + \tilde{\mathbb{P}}(X_t = y, T > t) \\ &= \tilde{\mathbb{P}}(Y_t = y, T \leq t) + \tilde{\mathbb{P}}(X_t = y, T > t) \\ &\leq \tilde{\mathbb{P}}(Y_t = y) + \tilde{\mathbb{P}}(X_t = y, T > t), \end{aligned}$$

e analogamente, $\tilde{\mathbb{P}}(Y_t = y) \leq \tilde{\mathbb{P}}(X_t = y) + \tilde{\mathbb{P}}(Y_t = y, T > t)$, logo,

$$|\tilde{\mathbb{P}}(X_t = y) - \tilde{\mathbb{P}}(Y_t = y)| \leq \tilde{\mathbb{P}}(X_t = y, T > t) + \tilde{\mathbb{P}}(Y_t = y, T > t).$$

Portanto, somando para todo $y \in S$, obtemos

$$\sum_y |\tilde{\mathbb{P}}(X_t = y) - \tilde{\mathbb{P}}(Y_t = y)| \leq 2\tilde{\mathbb{P}}(T > t) \rightarrow 0,$$

quando $t \rightarrow \infty$, uma vez que $\tilde{\mathbb{P}}[T < \infty] = 1$.

Como $Y_0 \sim \pi$ a distribuição estacionária, $\tilde{\mathbb{P}}(Y_t = y) = \pi(y)$, e como $X_0 = x$, então $\tilde{\mathbb{P}}(X_t = y) = P^t(x, y)$ de modo que a convergência acima se reescreve como

$$\sum_y |P^t(x, y) - \pi(y)| \leq 2\tilde{\mathbb{P}}(T > t) \rightarrow 0. \quad (2.11)$$

Como esta é uma soma de termos positivos que converge a 0, cada termo individualmente também deve convergir para 0, ou seja

$$|P^t(x, y) - \pi(y)| \rightarrow 0$$

quando $t \rightarrow \infty$, concluindo a demonstração do teorema. \square

Vamos aproveitar parte da construção feita acima para apresentar uma outra prova do teorema de convergência, onde é dado maior enfoque ao acoplamento utilizado.

Demonstração Alternativa. Fixados um estado $x \in S$ e uma cadeia $(X_t)_{t \geq 0}$ Markov- (δ_x, P) . Tomamos $(Y_t)_{t \geq 0}$ a cadeia de Markov em S que começa com distribuição π e evolui com base na matriz P , independente de $(X_t)_{t \geq 0}$ até o momento que encontra esta cadeia, então a partir dali começa a andar junto com $(X_t)_{t \geq 0}$. É claro que $(Y_t)_{t \geq 0}$ é uma cadeia de Markov- (π, P) , logo $(X_t, Y_t)_{t \geq 0}$ é um acoplamento de cadeias de Markov com matriz de transição P . Escrevemos $\bar{\mathbb{P}} = \bar{\mathbb{P}}_{\delta_x \times \pi}$ para a probabilidade no espaço das trajetórias em que ambas as cadeias estão definidas.

Definimos assim como antes $T := \inf\{t \geq 0; X_t = Y_t\}$ o primeiro tempo em que as cadeias $(X_t)_{t \geq 0}$ e $(Y_t)_{t \geq 0}$ se encontram. Até o tempo T , o acoplamento definido acima (que induz a probabilidade $\bar{\mathbb{P}}$) e o acoplamento definido na demonstração anterior (que induz a probabilidade $\tilde{\mathbb{P}}$) são iguais, de modo que vale

$$\bar{\mathbb{P}}(T < \infty) = \tilde{\mathbb{P}}(T < \infty) = 1,$$

conforme vimos na demonstração anterior. Além disso, pela forma com que construímos o acoplamento $\bar{\mathbb{P}}$, vale a igualdade de eventos $[X_t \neq Y_t] = [T > t]$.

Para cada t fixo, $\bar{\mathbb{P}}(X_t = \cdot) = P^t(x, \cdot)$ e $\bar{\mathbb{P}}(Y_t = \cdot) = \pi(y)$, de modo que para cada t , $(X_t, Y_t)_{t \geq 0}$ é um acoplamento das distribuições $P^t(x, \cdot)$ e π . Então é consequência das Proposições 2.2 e 2.7 que

$$\frac{1}{2} \sum_{y \in S} |P^t(x, y) - \pi(y)| = \|P^t(x, \cdot) - \pi\|_{TV} \leq \bar{\mathbb{P}}(X_t \neq Y_t) = \bar{\mathbb{P}}(T > t) \rightarrow 0 \quad (2.12)$$

quando $t \rightarrow \infty$. E assim como na demonstração anterior, como temos uma soma de termos positivos convergindo para 0, cada termo converge para 0 individualmente, obtendo que para todo $y \in S$,

$$\lim_{t \rightarrow \infty} P^t(x, y) = \pi(y).$$

\square

Cabe observar que nas demonstrações acima foi provado também que

$$\lim_{t \rightarrow \infty} \|P^t(x, \cdot) - \pi\|_{TV} = 0.$$

Na segunda demonstração isso fica claro na equação (2.12), mas na primeira isso também é imediato quando combinamos a equação (2.11) com a Proposição 2.2.

No caso em que o espaço de estados S é finito, um pequeno passo a mais no teorema anterior permite obtermos uma cota para a velocidade dessa convergência.

Teorema 2.15 (Teorema de Convergência para Espaço de Estados Finito). *Seja P uma cadeia irredutível e aperiódica em um espaço de estados S finito, com distribuição invariante π . Então existem constantes $C \leq \infty$ e $\alpha \in (0, 1)$ tais que*

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

Demonstração. Seja, como na primeira demonstração do teorema de convergência, $(X_t, Y_t)_{t \geq 0}$ um acoplamento de duas cadeias de Markov independentes, e $\tilde{\mathbb{P}}$ a probabilidade no espaço que as trajetórias de ambas estejam definidas. Os subíndices que atribuiremos a $\tilde{\mathbb{P}}$ indicarão as distribuições iniciais de X_0 e Y_0 .

Pela Proposição 1.14 existe um $r > 0$ tal que $P^r(x, y) > 0$ para todo $x, y \in S$. Pondo $h = \min_{x, y \in S} P^r(x, y) > 0$, temos que para todo $x, y \in S$, pela independência de $(X_t)_{t \geq 0}$ e $(Y_t)_{t \geq 0}$,

$$\tilde{\mathbb{P}}_{x,y}(X_r = z, Y_r = z) = \tilde{\mathbb{P}}_{x,y}(X_r = z) \tilde{\mathbb{P}}_{x,y}(Y_r = z) = P^r(x, z) P^r(y, z) > h^2$$

para todo $z \in S$. Assim

$$\tilde{\mathbb{P}}_{x,y}(X_r = Y_r) = \sum_{z \in S} \tilde{\mathbb{P}}_{x,y}(X_r = z, Y_r = z) > |S|h^2 := 1 - \varepsilon$$

Em particular note que $\varepsilon \in (0, 1)$, haja vista que $|S|h < 1$ (se não fosse, o modo com que h foi definido implicaria P não estocástica). Além disso, como $h \leq 1$ temos que $h^2 < h$, e portanto vale $|S|h^2 < 1$.

Prosseguindo, para $T = \inf\{t \geq 0; X_t = Y_t\}$ como na demonstração do teorema de convergência, note que

$$\tilde{\mathbb{P}}_{x,y}[T \leq r] \geq \tilde{\mathbb{P}}_{x,y}[X_r = Y_r] \geq 1 - \varepsilon,$$

assim vale

$$\tilde{\mathbb{P}}_{x,y}[T > r] \leq \varepsilon,$$

e pela perda de memória (a cada r passos recomeça a cadeia) segue que para todo k ,

$$\tilde{\mathbb{P}}_{x,y}[T > kr] \leq \varepsilon^k.$$

Para $t \geq 0$, escrevemos $t = kr + j$ para algum $j < r$ e algum $k \geq 0$. Como $\tilde{\mathbb{P}}_{x,y}[T > t]$ é decrescente em t , vale

$$\tilde{\mathbb{P}}_{x,y}[T > t] = \tilde{\mathbb{P}}_{x,y}[T > kr + j] \leq \tilde{\mathbb{P}}_{x,y}[T > kr] \leq \varepsilon^k.$$

Mas isso vale para todo $x, y \in S$, logo vale a desigualdade com $\tilde{\mathbb{P}}_{\delta_x \times \pi}$. Assim, temos que

$$\tilde{\mathbb{P}}_{\delta_x \times \pi}[T > t] \leq \varepsilon^k = \varepsilon^{\frac{t-j}{r}} = (\varepsilon^{1/r})^t \varepsilon^{r/j} \leq \frac{C}{2} \alpha^t,$$

onde $C = 2 \max_{0 \leq j < r} \varepsilon^{r/j}$ e $\alpha = \varepsilon^{1/r}$.

Por fim, da equação (2.11) e da Proposição 2.2, segue que

$$\|P^t(x, \cdot) - \pi\|_{TV} = \frac{1}{2} \sum_{y \in S} |\mathbb{P}_x(X_t = y) - \pi(y)| \leq 2\tilde{\mathbb{P}}(T > t) \leq C\alpha^t.$$

□

2.4 Tempo de mistura

O teorema de convergência garante para cadeias irredutíveis e aperiódicas que, independente do estado em que começar, a distribuição da cadeia no tempo t converge para a distribuição estacionária, a qual depende apenas da dinâmica da cadeia dada pela matriz P . A versão do teorema de convergência para espaços de estados finitos inclusive nos dá uma cota exponencial para a velocidade desta convergência. Nosso objetivo agora passa a ser refinar este resultado, especialmente para espaços de estados finitos onde podemos analisar como o tamanho do espaço influencia neste tempo.

Como esta convergência garantida é dada em função da distância de variação total, nada melhor que a utilizarmos como métrica. Para isso definimos

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

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

onde π é a distribuição estacionária da cadeia.

Proposição 2.16. *A relação entre as medidas d e \bar{d} definidas acima é dada por*

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

Demonstração. Pela Proposição 2.3, a distância de variação total satisfaz a desigualdade triangular, valendo portanto, para todo $x, y \in S$,

$$\|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \leq \|P^t(x, \cdot) - \pi\|_{TV} + \|P^t(y, \cdot) - \pi\|_{TV} \leq 2d(t).$$

Consequentemente $\bar{d}(t) = \max_{x, y \in S} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \leq 2d(t)$.

Para a outra desigualdade Utilizamos a definição de distância de variação total dada em (2.1). Note que sendo π a distribuição estacionária, $\pi(x) = (\pi P^t)(x) = \sum_{y \in S} \pi(y) P^t(y, x)$ para todo $x \in S$ e todo $t > 0$. Então para qualquer conjunto $A \subset S$, somando sobre todos os $x \in A$, obtemos

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

Ao mesmo tempo π é uma distribuição, logo $\sum_y \pi(y) = 1$, e portanto,

$$P^t(x, A) = P^t(x, A) \sum_y \pi(y) = \sum_y \pi(y) P^t(x, A).$$

Destes dois fatos, juntamente com a definição de distância de variação total em (2.1) e a

definição de $\bar{d}(t)$, obtemos

$$\begin{aligned}
 |P^t(x, A) - \pi(A)| &= \left| \sum_{y \in S} \pi(y) P^t(x, A) - \sum_{y \in S} \pi(y) P^t(y, A) \right| \\
 &= \left| \sum_{y \in S} \pi(y) [P^t(x, A) - P^t(y, A)] \right| \\
 &\leq \sum_{y \in S} \pi(y) |P^t(x, A) - P^t(y, A)| \\
 &\leq \sum_{y \in S} \pi(y) \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \\
 &\leq \sum_{y \in S} \pi(y) \bar{d}(t) \\
 &= \bar{d}(t) \sum_{y \in S} \pi(y) \\
 &= \bar{d}(t).
 \end{aligned} \tag{2.13}$$

Utilizando mais uma vez a definição da distância de variação total, concluimos em virtude de (2.13) que

$$d(t) = \max_{x \in S} \|P^t(x, \cdot) - \pi\|_{TV} = \max_{x \in S} \max_{A \subset S} |P^t(x, A) - \pi(A)| \leq \bar{d}(t).$$

□

Lema 2.17. *A função $\bar{d}(t)$ é sub-multiplicativa, isto é,*

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

Demonstração. Dados s, t inteiros positivos. Fixamos $x, y \in S$ e consideramos (X_s, Y_s) o acoplamento das distribuições $P^s(x, \cdot)$ e $P^s(y, \cdot)$ que satisfaz

$$\|P^s(x, \cdot) - P^s(y, \cdot)\|_{TV} = \mathbb{P}(X_s \neq Y_s)$$

cuja existência é garantida pela Proposição 2.7. Em particular temos que

$$\bar{d}(s) = \max_{u, v \in S} \|P^s(u, \cdot) - P^s(v, \cdot)\|_{TV} \geq \mathbb{P}(X_s \neq Y_s). \tag{2.14}$$

Pela definição de acoplamento $\mathbb{P}(X_s = \cdot) = P^s(x, \cdot)$, então condicionando sobre os possíveis estados em que o processo encontra-se no s -ésimo passo e somando sobre todas as possibilidades, obtemos que para $w \in S$ qualquer,

$$P^{s+t}(x, w) = \sum_{z \in S} P^s(x, z) P^t(z, w) = \sum_{z \in S} \mathbb{P}(X_s = z) P^t(z, w) = \mathbb{E}[P^t(X_s, w)].$$

Analogamente mostra-se que $P^{s+t}(y, w) = \mathbb{E}[P^t(Y_s, w)]$ para todo $w \in S$. Deste modo, para um conjunto $A \subset S$, somando sobre $w \in A$, obtemos

$$\begin{aligned}
 P^{s+t}(x, A) - P^{s+t}(y, A) &= \mathbb{E} [P^t(X_s, A) - P^t(Y_s, A)] \\
 &= \mathbb{E} [(P^t(X_s, A) - P^t(Y_s, A)) \mathbf{1}_{\{X_s \neq Y_s\}}] \\
 &\leq \mathbb{E} [\bar{d}(t) \mathbf{1}_{\{X_s \neq Y_s\}}]
 \end{aligned} \tag{2.15}$$

uma vez que pela definição de \bar{d} como o máximo, $P^t(X_s, A) - P^t(Y_s, A) \leq \bar{d}(t)$.

Como a equação (2.15) vale para todo $A \subset S$, combinando com (2.14) temos que

$$\|P^{s+t}(x, \cdot) - P^{s+t}(y, \cdot)\|_{TV} = \max_{A \subset S} |P^{s+t}(x, A) - P^{s+t}(y, A)| \leq \bar{d}(t)\mathbb{P}(X_s \neq Y_s) \leq \bar{d}(t)\bar{d}(s).$$

E como isso vale para $x, y \in S$ quaisquer, tomando o máximo obtemos a desigualdade desejada. \square

Corolário 2.18. *A função $\bar{d}(t)$ goza das seguintes propriedades:*

- (a) $\bar{d}(t)$ é não crescente em t ;
- (b) $\bar{d}(ct) \leq \bar{d}(t)^c$ para todos c e t inteiros positivos.

Demonstração. Para o item (a), basta observar que se $t_2 > t_1$, então existe um inteiro positivo s tal que $t_2 = t_1 + s$, logo segue do resultado anterior que

$$\bar{d}(t_2) = \bar{d}(t_1 + s) \leq \bar{d}(t_1)\bar{d}(s) \leq \bar{d}(t_1),$$

já que $0 \leq \bar{d}(s) \leq 1$ pela definição de \bar{d} através da distância de variação total.

Para o item (b) façamos por indução. Fixamos t inteiro positivo. Para $c = 1$ vale trivialmente. Agora suponha que o resultado vale para c , isto é, que $\bar{d}(ct) \leq \bar{d}(t)^c$, e vamos mostrar que vale para $c + 1$. De fato, é consequência do lema 2.17 e da hipótese de indução que

$$\bar{d}((c+1)t) = \bar{d}(ct + t) \leq \bar{d}(ct) \cdot \bar{d}(t) \leq \bar{d}(t)^c \cdot \bar{d}(t) = \bar{d}(t)^{c+1},$$

mostrando que vale para $c + 1$, o que conclui a indução. \square

Agora já temos um bom entendimento das métricas d e \bar{d} , podendo definir o tempo de mistura do seguinte modo:

Definição 2.19. *Seja P uma cadeia de Markov com distribuição estacionária π e*

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

*como posto anteriormente. Definimos o **tempo de mistura** da cadeia por*

$$t_{\text{mix}}(\varepsilon) := \min\{t; d(t) \leq \varepsilon\},$$

e

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

Note que fixado $\varepsilon > 0$, o tempo de mistura $t_{\text{mix}}(\varepsilon)$ é o menor tempo t tal que $d(t) = \max_{x \in S} \|P^t(x, \cdot) - \pi\|_{TV} \leq \varepsilon$, isto é, o primeiro instante em que, independente de onde tenha começado, a distância de variação total entre a distribuição da cadeia neste tempo e a distribuição estacionária seja menor que ε . Em outras palavras, ele indica o tempo necessário para que a distribuição da cadeia fique suficientemente próxima da distribuição estacionária.

Por ser uma característica global da cadeia, chamaremos tal objeto de *indicador da cadeia*. Este é o primeiro indicador construído, mas não será o único. No Capítulo 4 definiremos mais indicadores, e no Capítulo 5 estudaremos as relações entre eles.

Ainda observamos que o teorema de convergência para espaços de estados finitos garante a existência do tempo de mistura. Mais que isso, ele já gera uma cota para este indicador, uma vez que garante que $d(t) \leq C\alpha^t$, logo para $t \geq \log_\alpha(\frac{\varepsilon}{C})$ temos que $d(t) \leq \varepsilon$, o que é suficiente para mostrar que $t_{\text{mix}}(\varepsilon) \leq \log_\alpha(\frac{\varepsilon}{C})$. Neste trabalho ainda vamos melhorar esta cota,

principalmente em alguns casos particulares, onde a relacionaremos com o tamanho do espaço de estados da cadeia.

Finalizamos esta seção com duas desigualdades envolvendo tempo de mistura e uma cota para t_{mix} , a qual será extremamente útil em resultados futuros.

Como $t_{\text{mix}}(\varepsilon) = \min\{t; d(t) \leq \varepsilon\}$, devemos ter $d(t_{\text{mix}}(\varepsilon)) \leq \varepsilon$. Combinando isso com a Proposição 2.16 e o Corolário 2.18 temos que, para todo l inteiro positivo,

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

Em particular, para $\varepsilon = \frac{1}{4}$ vale

$$d(lt_{\text{mix}}) \leq 2^{-l}$$

Donde obtemos, com $l = \lceil \log_2 \varepsilon^{-1} \rceil$, que

$$d(\lceil \log_2 \varepsilon^{-1} \rceil t_{\text{mix}}) \leq 2^{-\lceil \log_2 \varepsilon^{-1} \rceil} \leq 2^{-\log_2 \varepsilon^{-1}} = 2^{\log_2 \varepsilon} = \varepsilon$$

Ou seja, $t_0 = \lceil \log_2 \varepsilon^{-1} \rceil t_{\text{mix}}$ satisfaz $d(t_0) < \varepsilon$, logo da definição de $t_{\text{mix}}(\varepsilon)$ segue que $t_{\text{mix}}(\varepsilon) \leq t_0 = \lceil \log_2 \varepsilon^{-1} \rceil t_{\text{mix}}$.

O próximo resultado cumpre a promessa que fizemos quando enunciamos a Proposição 2.11.

Proposição 2.20. *Suponha que para cada par de estados $x, y \in S$ existe um acoplamento $(X_t, Y_t)_{t \geq 0}$ com $X_0 = x$ e $Y_0 = y$. Para cada um desses acoplamentos, seja τ_{couple} o **tempo de coalescência** definido na equação (2.9). Então*

$$d(t) \leq \max_{x, y \in S} \mathbb{P}_{x, y}(\tau_{\text{couple}} > t),$$

e conseqüentemente

$$t_{\text{mix}} \leq \lceil 4 \max_{x, y \in S} \mathbb{E}_{x, y}[\tau_{\text{couple}}] \rceil.$$

Demonstração. A primeira desigualdade é imediata da Proposição 2.16 e da Proposição 2.11, uma vez que

$$d(t) \leq \bar{d}(t) = \max_{x, y \in S} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \leq \max_{x, y \in S} \mathbb{P}_{x, y}(\tau_{\text{couple}} > t).$$

Utilizando a desigualdade de Chebyshev nesta desigualdade, obtemos

$$d(t) \leq \max_{x, y \in S} \mathbb{P}_{x, y}(\tau_{\text{couple}} > t) \leq \frac{1}{t} \max_{x, y \in S} \mathbb{E}_{x, y}(\tau_{\text{couple}}).$$

Assim para $t \geq 4 \max_{x, y} \mathbb{E}_{x, y}(\tau_{\text{couple}})$ temos que $d(t) \leq \frac{1}{4}$, donde concluímos que $t_{\text{mix}} \leq \lceil 4 \max_{x, y \in S} \mathbb{E}_{x, y}[\tau_{\text{couple}}] \rceil$. \square

2.5 Tempo de mistura considerando outras distâncias

Na seção anterior utilizamos a distância de variação total entre as distribuições $P^t(x, \cdot)$ e a distribuição estacionária π para definir o tempo de mistura, visto que o teorema de convergência garantia que tal distância convergia para 0. Este teorema é mais geral, uma vez que garante também que para todo $x \in S$, $P^t(x, \cdot)$ converge pontualmente para π quando $t \rightarrow \infty$, de modo que a função

$$\frac{P^t(x, \cdot)}{\pi(\cdot)} - 1$$

de S em \mathbb{R} deve convergir pontualmente para a função nula quando $t \rightarrow \infty$. Deste modo, escolher qualquer norma no espaço das funções $S \rightarrow \mathbb{R}$ e analisar como a norma desta função decresce quando $t \rightarrow \infty$ também tratá informações sobre a velocidade da convergência.

Nesta seção vamos estar interessados em normas que levem em consideração a distribuição estacionária π , em especial as normas $\ell^p(\pi)$ para $1 \leq p \leq \infty$, definidas para uma função $f : S \rightarrow \mathbb{R}$ por

$$\|f\|_p := \begin{cases} \left[\sum_{y \in S} |f(y)|^p \pi(y) \right]^{1/p}, & 1 \leq p < \infty \\ \max_{y \in S} |f(y)|, & p = \infty \end{cases}.$$

Lembre que dadas duas funções $f, g : S \rightarrow \mathbb{R}$, o produto interno usual entre elas é definido por

$$\langle f, g \rangle = \sum_{x \in S} f(x)g(x).$$

Também podemos considerar o produto interno ponderado pela distribuição π , que é dado por

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

Da definição da norma $\ell^2(\pi)$ fica claro que ela provém desse produto interno.

Para uma matriz irredutível P em S com distribuição estacionária π , definimos

$$q_t(x, y) := \frac{P^t(x, y)}{\pi(y)},$$

e observamos que se P é reversível com respeito à π então $q_t(x, y) = q_t(y, x)$. Também observamos que com o produto interno ponderado pela π definido acima, vale

$$\langle q_t(x, \cdot), 1 \rangle_\pi = \sum_y q_t(x, y)\pi(y) = \sum_y P^t(x, y) = 1. \quad (2.16)$$

Definimos a distância $d^{(p)}$ por

$$d^{(p)}(t) := \max_{x \in S} \|q_t(x, \cdot) - 1\|_p = \max_{x \in S} \left\| \frac{P^t(x, \cdot)}{\pi(\cdot)} - 1 \right\|_p. \quad (2.17)$$

Pelo que comentamos no início desta seção, ela constitui uma boa medida para visualizar a velocidade da convergência garantida pelo Teorema de Convergência, uma vez que para cada $x \in S$ a função $\frac{P^t(x, \cdot)}{\pi(\cdot)} - 1$ converge pontualmente para a função nula.

Pensando nisso definimos o **ℓ^p -tempo de mistura** como sendo

$$t_{\text{mix}}^{(p)}(\varepsilon) := \inf \{ t \geq 0; d^{(p)}(t) \leq \varepsilon \},$$

e

$$t_{\text{mix}}^{(p)} := t_{\text{mix}}^{(p)}\left(\frac{1}{2}\right).$$

O parâmetro $t_{\text{mix}}^{(\infty)}$ é geralmente chamado de **tempo de mistura uniforme**.

A distância de variação total definida anteriormente é um caso particular das distâncias $\ell^p(\pi)$ definidas acima, uma vez que vale $d^{(1)}(t) = 2d(t)$. Para ver isto, basta observar que da

Proposição 2.2, temos que

$$\begin{aligned}
 \|P^t(x, \cdot) - \pi\|_{TV} &= \frac{1}{2} \sum_{y \in S} |P^t(x, y) - \pi(y)| \\
 &= \frac{1}{2} \sum_{y \in S} \left| \frac{P^t(x, y)}{\pi(y)} - 1 \right| \pi(y) \\
 &= \frac{1}{2} \sum_{y \in S} |q_t(x, y) - 1| \pi(y) \\
 &= \frac{1}{2} \|q_t(x, \cdot) - 1\|_1.
 \end{aligned} \tag{2.18}$$

Tomando o máximo para $x \in S$, segue a igualdade desejada. Por esta igualdade é que escolhemos $t_{\text{mix}}^{(p)} := t_{\text{mix}}^{(p)}(1/2)$ ao invés de utilizar $1/4$, copiando o que fizemos para definir $t_{\text{mix}} = t_{\text{mix}}(1/4)$. A escolha atual garante-nos que

$$\begin{aligned}
 t_{\text{mix}}^{(1)} &= t_{\text{mix}}^{(1)} \left(\frac{1}{2} \right) = \min \left\{ t; d^{(1)}(t) < \frac{1}{2} \right\} \\
 &= \min \left\{ t; 2d(t) < \frac{1}{2} \right\} \\
 &= \min \left\{ t; d(t) < \frac{1}{4} \right\} = t_{\text{mix}} \left(\frac{1}{4} \right) = t_{\text{mix}}.
 \end{aligned}$$

Para finalizar esta seção, observamos que para cada $1 \leq p \leq \infty$, a distância $d^{(p)}$ também é sub-multiplicativa, isto é,

$$d^{(p)}(t + s) \leq d^{(p)}(t) d^{(p)}(s),$$

para quaisquer s, t inteiros positivos. A demonstração deste fato pode ser encontrada em [9, Lema 4.18], como consequência da desigualdade de Hölder. Além disso, uma vez que as normas ℓ^p são não decrescentes, isto é, para uma função $f : S \rightarrow \mathbb{R}$ fixa, a função $p \mapsto \|f\|_p$ é não decrescente para $p \geq 1$, é imediato da definição de $d^{(p)}(t)$ em (2.17) que as medidas $d^{(p)}$ satisfazem

$$2d(t) = d^{(1)}(t) \leq d^{(2)}(t) \leq d^{(\infty)}(t). \tag{2.19}$$

As normas ℓ^2 e ℓ^∞ ainda possuem uma relação mais precisa quando a cadeia é reversível:

Proposição 2.21. *Para uma cadeia de Markov reversível,*

$$d^{(\infty)}(2t) = [d^{(2)}(t)]^2 = \max_{x \in S} q_{2t}(x, x) - 1. \tag{2.20}$$

Demonstração. Começamos observando que $P^{2t}(x, y) = \sum_z P^t(x, z)P^t(z, y)$, logo

$$q_{2t}(x, y) = \frac{P^{2t}(x, y)}{\pi(y)} = \sum_z \frac{1}{\pi(y)} P^t(x, z)P^t(z, y) = \sum_z \frac{P^t(x, z)}{\pi(z)} \frac{P^t(z, y)}{\pi(y)} \pi(z).$$

Ao mesmo tempo, como P é reversível, já vimos que $q_t(y, z) = q_t(z, y)$, e assim vale

$$\begin{aligned}
 \langle q_t(x, \cdot), q_t(y, \cdot) \rangle_\pi &= \sum_z q_t(x, z)q_t(y, z)\pi(z) \\
 &= \sum_z q_t(x, z)q_t(z, y)\pi(z) \\
 &= \sum_z \frac{P^t(x, z)}{\pi(z)} \frac{P^t(z, y)}{\pi(y)} \pi(z).
 \end{aligned}$$

Combinando as duas equações anteriores obtemos que

$$q_{2t}(x, y) = \langle q_t(x, \cdot), q_t(y, \cdot) \rangle_\pi. \quad (2.21)$$

Por outro lado, utilizando a linearidade do produto interno juntamente com a equação (2.16), obtemos

$$\begin{aligned} \langle q_t(x, \cdot) - 1, q_t(y, \cdot) - 1 \rangle_\pi &= \langle q_t(x, \cdot), q_t(y, \cdot) \rangle_\pi - \langle 1, q_t(y, \cdot) \rangle_\pi + \langle q_t(x, \cdot), 1 \rangle_\pi + \langle -1, -1 \rangle_\pi \\ &= \langle q_t(x, \cdot), q_t(y, \cdot) \rangle_\pi - 1 - 1 + 1, \end{aligned}$$

o que e combinando com a igualdade anterior nos dá

$$\langle q_t(x, \cdot) - 1, q_t(y, \cdot) - 1 \rangle_\pi = q_{2t}(x, y) - 1. \quad (2.22)$$

Em particular, como a norma $l^2(\pi)$ provém desse produto interno, fazendo $x = y$ na igualdade acima, obtemos que

$$\|q_t(x, \cdot) - 1\|_2^2 = \langle q_t(x, \cdot) - 1, q_t(x, \cdot) - 1 \rangle_\pi = q_{2t}(x, x) - 1,$$

donde segue que

$$d^{(2)}(t) = \max_x \|q_t(x, \cdot) - 1\|_2 = \max_x [(q_{2t}(x, x) - 1)^{1/2}] = \max_x [(q_{2t}(x, x) - 1)]^{1/2},$$

mostrando que vale a última igualdade em (2.20).

Para a outra igualdade, utilizando Cauchy-Schwarz juntamente com a equação (2.22) obtemos

$$\begin{aligned} |q_{2t}(x, y) - 1|^2 &= |\langle q_t(x, \cdot) - 1, q_t(y, \cdot) - 1 \rangle_\pi|^2 \\ &\leq \langle q_t(x, \cdot) - 1, q_t(x, \cdot) - 1 \rangle_\pi \langle q_t(y, \cdot) - 1, q_t(y, \cdot) - 1 \rangle_\pi \\ &= (q_{2t}(x, x) - 1)(q_{2t}(y, y) - 1), \end{aligned}$$

donde segue que

$$|q_{2t}(x, y) - 1| \leq \sqrt{q_{2t}(x, x) - 1} \sqrt{q_{2t}(y, y) - 1}.$$

Consequentemente

$$\begin{aligned} d^{(\infty)}(2t) &= \max_x \|q_{2t}(x, \cdot) - 1\|_\infty \\ &= \max_{x,y} |q_{2t}(x, y) - 1| \\ &\leq \max_{x,y} \sqrt{q_{2t}(x, x) - 1} \sqrt{q_{2t}(y, y) - 1} \\ &= \max_x q_{2t}(x, x) - 1. \end{aligned} \quad (2.23)$$

A última equação diz que

$$\max_{x,y} |q_{2t}(x, y) - 1| \leq \max_x q_{2t}(x, x) - 1,$$

mas se isso vale, então o máximo do lado esquerdo é obtido com $y = x$, e portanto vale a igualdade. Ou seja, a equação (2.23) garante que

$$d^{(\infty)}(2t) = \max_x q_{2t}(x, x) - 1,$$

o que conclui a demonstração. \square

Capítulo 3

Passeios Aleatórios em Redes

Do ponto de vista da física, uma rede elétrica¹ é um conjunto de componentes elétricos, como resistores, capacitores e indutores, que podem ser dispostos de diferentes maneiras. Quando aplicamos uma voltagem (ou diferença de potencial) entre dois pontos uma corrente é gerada neste circuito, e cada componente influencia o comportamento desta corrente de acordo com suas características como resistência, capacitância e indutância. Essas ideias são a base da construção das **redes**: um objeto matemático que copia as características dos circuitos elétrico da física, trazendo sua intuição para o estudo de cadeias de Markov.

Neste capítulo definiremos rigorosamente o que é uma rede do ponto de vista matemático. Após isso também definiremos o passeio aleatório associado à rede e mostraremos que toda cadeia de Markov reversível em um espaço de estados finito pode ser vista como tal (Teorema 3.11).

A construção deste capítulo é baseada na abordagem vista em [9], mas também utilizamos muitas ideias vistas em [6], [12, Seção 4.3] e [11]. Citamos em especial este último, que recomendamos por conter vários resultados interessantes, mas que achamos melhor omitir por serem paralelos ao objetivo do nosso texto.

3.1 Funções harmônicas

Começamos apresentando o conceito de função harmônica, que constitui uma importante ferramenta para o estudo de cadeias de Markov.

Definição 3.1. *Seja P a matriz de transição de uma cadeia de Markov irredutível em um espaço de estados finito S . Dizemos que uma função $h : S \rightarrow \mathbb{R}$ é **harmônica** no vértice $x \in S$ (em relação à P) se*

$$h(x) = Ph(x) = \sum_{y \in S} P(x, y)h(y). \quad (3.1)$$

Se h for harmônica para todo estado $x \in \mathcal{D}$, dizemos simplesmente que h é harmônica em \mathcal{D} .

Quando P é a matriz de transição do passeio aleatório simples em um grafo $G = (V, E)$ definido na Seção 1.9, isto é, $P(x, y) = \frac{1}{d_x} \mathbf{1}_{\{y \sim x\}}$, de acordo com a equação (3.1), uma função h é harmônica em $x \in V$ se

$$h(x) = \sum_{y \in V} P(x, y)h(y) = \sum_{y \sim x} \frac{1}{d_x} h(y) = \frac{1}{d_x} \sum_{y \sim x} h(y),$$

ou seja, se $h(x)$ for a média simples dos valores de h nos vértices vizinhos de x .

¹Para uma primeira visão dos conceitos de rede elétrica, recomendamos [5]

Generalizando, quando P é uma cadeia de Markov qualquer no espaço de estados S , a mesma equação nos diz que h é harmônica em $x \in S$ se

$$h(x) = \sum_{y \in S} P(x, y)h(y) = \sum_{y \in S} h(y)\mathbb{P}(X_1 = y|X_0 = x) = \mathbb{E}[h(X_1)|X_0 = x] = \mathbb{E}_x[h(X_1)],$$

que pode ser interpretado como $h(x)$ sendo a média dos valores de h nos vizinhos em que ela pode acessar em 1 passo, ponderada pela probabilidade de ir para cada um deles².

A seguir veremos alguns resultados clássicos para funções harmônicas, os quais terão grande utilidade futura.

Lema 3.2. *Seja P uma cadeia de Markov irredutível nos espaço de estados S finito. Se h é harmônica em S , então ela é constante.*

Demonstração. Como S é finito deve existir um estado $x_0 \in S$ tal que $h(x_0) = M$ é maximal. Afirmamos que para todo $z \in S$ tal que $P(x_0, z) > 0$, devemos ter $h(z) = M$. De fato, se não fosse teríamos um $z \in S$ com $P(x_0, z) = \varepsilon > 0$ e $h(z) < M$, conseqüentemente, como h é harmônica, teríamos

$$\begin{aligned} h(x_0) &= \sum_{y \in S} P(x_0, y)h(y) = P(x_0, z)h(z) + \sum_{y \neq z} P(x_0, y)h(y) \\ &\leq \varepsilon h(z) + M \sum_{y \neq z} P(x_0, y) \\ &= \varepsilon h(z) + M(1 - \varepsilon) < M, \end{aligned}$$

um absurdo.

Agora para $y \in S$ qualquer, como P é irredutível, existe uma sequência $x_0, x_1, \dots, x_n = y$ tal que $P(x_i, x_{i+1}) > 0$. Aplicando o argumento acima para cada par (x_i, x_{i+1}) , mostra-se que

$$h(y) = h(x_{n-1}) = \dots = h(x_1) = h(x_0) = M,$$

Concluindo que $h(y) = M$ para todo $y \in S$, sendo a função h constante. \square

Com as mesmas ideias da demonstração acima, podemos demonstrar o Princípio do Máximo para funções harmônicas. Este resultado, juntamente com o Princípio da Unicidade que demonstraremos a seguir são amplamente discutidos em [6].

Lema 3.3 (Princípio do Máximo). *Seja P uma cadeia de Markov irredutível em S finito. Sejam ainda $A \subset S$ e $h : S \rightarrow \mathbb{R}$ harmônica em $S \setminus A$. Então $\max_{x \in S} h(x) = \max_{x \in A} h(x)$.*

Demonstração. Pela inclusão de conjuntos é claro que $\max_{x \in S} h(x) \geq \max_{x \in A} h(x)$. O que queremos mostrar é que não vale a desigualdade estrita, que por sua vez é equivalente a termos $\max_{x \in S \setminus A} h(x) > \max_{x \in A} h(x)$. Suponha por absurdo que isso vale.

Seja $M = \max_{x \in S \setminus A} h(x) > \max_{x \in A} h(x)$. Como S é finito, $S \setminus A$ também é, logo tal máximo é atingido, existindo $y \in S \setminus A$ com $h(y) = M$. Pelo mesmo argumento da demonstração do Lema 3.2, $h(z) = M$ para todo $z \in S$ tal que $P(y, z) > 0$ \blacklozenge . Note que precisamos da harmonicidade de h em y para obter tal resultado.

²No contexto contínuo, $h : \mathbb{R}^n \rightarrow \mathbb{R}$ é dita harmônica quando $\Delta h = 0$. Funções harmônicas nesse contexto satisfazem a propriedade do valor médio, isto é, para qualquer $r > 0$, o valor de $h(x)$ coincide com o valor médio de f na bola de centro x e raio r , que também é o valor médio de h na esfera de centro x e raio r . Isso é análogo à definição de função harmônica que está sendo dada aqui no contexto discreto: o valor de $h(x)$ é igual ao valor médio (ponderado) de h nos pontos vizinhos de x .

Tomemos $w \in A$ e um caminho $y = x_0, x_1, \dots, x_l = w$ com $x_i \notin A$ ($0 \leq i < l - 1$) tal que $P(x_i, x_{i+1}) > 0$ para todo i . Tal caminho existe, pois fixado $z \in A$ qualquer, como a cadeia é irredutível, existe um caminho $y = x_0, x_1, \dots, x_l, \dots, x_m = z$ com $P(x_i, x_{i+1}) > 0$ para todo i , bastando tomar w o primeiro x_i deste caminho tal que $x_i \in A$.

Como $x_0, x_1, \dots, x_{l-1} \in S \setminus A$, h é harmônica nestes vértices, e repetidas aplicações de \diamond garantem que

$$h(w) = h(x_{l-1}) = \dots = h(x_1) = h(x_0) = h(y) = M,$$

de modo que, $\max_{x \in A} h(x) \geq h(w) = M = \max_{x \in S \setminus A} h(x)$, chegando a uma contradição. Portanto devemos ter

$$\max_{x \in A} h(x) = \max_{x \in S} h(x).$$

□

Lema 3.4 (Princípio da Unicidade). *Seja P uma cadeia irredutível em S finito, $A \subset S$, e suponha que $f, g : S \rightarrow \mathbb{R}$ são harmônicas em $S \setminus A$, e que $f = g$ em A . Então $f = g$ em todo S .*

Demonstração. Considere a função $f - g$. Como f e g são harmônicas em $S \setminus A$, a função $f - g$ também é, uma vez que para todo $x \in S \setminus A$,

$$\begin{aligned} \sum_{y \in S} P(x, y)(f - g)(y) &= \sum_{y \in S} P(x, y)f(y) - \sum_{y \in S} P(x, y)g(y) \\ &= f(x) - g(x) \\ &= (f - g)(x). \end{aligned}$$

Além disso, pela construção, $(f - g)(x) = 0$ para todo $x \in A$. Então é consequência do Princípio do Máximo que

$$\max_{x \in S} (f - g)(x) = \max_{x \in A} (f - g)(x) = 0,$$

o que garante que $f - g \leq 0$ em S .

Analogamente mostra-se que $g - f \leq 0$ em S , concluindo que $f = g$ em S . □

Proposição 3.5. *Seja $(X_t)_{t \geq 0}$ uma cadeia de Markov irredutível em S finito, com matriz de transição P . Seja ainda $B \subset S$ e $h_B : B \rightarrow \mathbb{R}$ uma função definida em B . Então a função $h : S \rightarrow \mathbb{R}$ definida por*

$$h(x) = \mathbb{E}_x[h_B(X_{\tau_B})]$$

é a única extensão de h_B para S que satisfaz

- $h(x) = h_B(x)$ para todo $x \in B$;
- h é harmônica para P em $S \setminus B$.

Demonstração. Seja τ_B o primeiro tempo que a cadeia chega no conjunto B . $\tau_B < \infty$ uma vez que o espaço de estados finito e $\mathbb{E}_x[\tau_y] < \infty$ para todo $x, y \in S$. Então $X_{\tau_B} \in B$ e h_B está bem definida para X_{τ_B} .

Para mostrar que $h = h_B$ em B , fixamos $x \in B$ qualquer e observamos que condicional ao evento $X_0 = x$, com probabilidade 1 $\tau_B = 0$, logo $X_{\tau_B} = X_0 = x$ e

$$h_B(X_{\tau_B}) = h_B(x),$$

donde segue que

$$h(x) = \mathbb{E}_x h_B(X_{\tau_B}) = h_B(x).$$

Para mostrar que h é harmônica em $S \setminus B$ fixamos $x \in S \setminus B$, e observamos que

$$h(x) = \mathbb{E}_x h_B(X_{\tau_B}) = \mathbb{E}_x h(X_{\tau_B}),$$

uma vez que $X_{\tau_B} \in B$ e, conforme visto acima, $h = h_B$ em B . Assim, pela lei da esperança total,

$$h(x) = \sum_{y \in S} P(x, y) \mathbb{E}_x [h(X_{\tau_B}) \mid X_1 = y].$$

Como $x \in S \setminus B$ e $\tau_B \geq 1$, a propriedade de Markov garante que

$$\mathbb{E}_x [h(X_{\tau_B}) \mid X_1 = y] = \mathbb{E}_y [h(X_{\tau_B})] = h(y).$$

Para se convencer disso observe que como a trajetória não atinge B no tempo 0, e como o argumento da esperança depende apenas do estado em que a trajetória primeiro atinge o conjunto B (e não do tempo), um shift de tempo 1 na cadeia não altera tal resultado.

Assim, das duas últimas equações ficamos com

$$h(x) = \sum_{y \in S} P(x, y) h(y),$$

o que conclui que h é harmônica para todo $x \in S \setminus B$.

Resta agora mostrar a unicidade, mas isso é imediato uma vez que se h e \tilde{h} são harmônicas em $S \setminus B$ com $h = h_B$ e $\tilde{h} = h_B$ em B , elas satisfazem $h = \tilde{h}$ em B , e então o princípio da unicidade (Lema 3.4) garante que $h = \tilde{h}$ em todo S . \square

Observação 3.6. *A harmonicidade de h poderia ser demonstrada usando a versão formal da propriedade fraca de Markov. Para isso observamos que*

$$\begin{aligned} h(x) &= \mathbb{E}_x h(X_{\tau_B}) = E_x [\mathbb{E}_x [h(X_{\tau_B}) \mid \mathcal{F}_1]] \\ &= E_x [\mathbb{E}_x [h(X_{\tau_B}) \circ \theta_1 \mid \mathcal{F}_1]] \\ &= \mathbb{E}_x [\mathbb{E}_{X_1} [h(X_{\tau_B})]], \end{aligned}$$

uma vez que como $X_0 \notin B$, se der este shift só vai mudar o tempo em que a trajetória atinge B , e não o estado em que ela atinge. Então utilizando a lei da esperança total, obtemos

$$\begin{aligned} h(x) &= \sum_y P(x, y) \mathbb{E}_x [\mathbb{E}_{X_1} [h(X_{\tau_B}) \mid X_1 = y]] \\ &= \sum_y P(x, y) \mathbb{E}_x [\mathbb{E}_y [h(X_{\tau_B})]] \\ &= \sum_y P(x, y) \mathbb{E}_y [h(X_{\tau_B})] \\ &= \sum_y P(x, y) h(y), \end{aligned}$$

onde a penúltima igualdade é consequência do argumento da esperança \mathbb{E}_x ser uma constante.

3.2 Redes elétricas e o passeio aleatório associado

Nesta seção, baseando-nos na intuição física dos circuitos elétricos, formalizaremos o conceito de redes do ponto de vista matemático. Antes da definição, relembremos o conceito de grafo visto na Seção 1.9 como um par $G = (V, E)$ onde V é o conjunto dos vértices e $E \subset \{\{x, y\}; x, y \in V, x \neq y\}$ o conjunto das arestas. Vamos estar interessados basicamente no mesmo objeto, com a pequena diferença de que agora permitiremos loops, ou seja, que exista uma aresta ligando um vértice nele mesmo, mais precisamente, permitiremos que $\{x, x\} \in E$.

Definição 3.7. Uma *rede* é um grafo finito, conexo e não direcional $G = (V, E)$ juntamente com um conjunto qualquer de números reais não negativos $\{c(e)\}_{e \in E}$ associados a cada aresta $e = \{x, y\} \in E$, os quais chamamos de **condutância**.

Fazendo um paralelo com a física, podemos considerar que cada aresta do grafo se comporta de maneira análoga a um resistor elétrico. Quando aplicamos uma diferença de potencial entre dois vértices, a corrente que flui através de cada aresta depende de sua condutância e resistência equivalentes. Vamos formalizar melhor este conceito na próxima seção ao definirmos corrente.

Quando não houver confusão, vamos denotar por

$$c(x, y) := c(\{x, y\}) \quad (3.2)$$

a condutância da aresta que liga os vértices x e y . Então, como o grafo é não-direcional, esta é a mesma aresta que liga os pontos y e x (isto é, $\{x, y\} = \{y, x\}$), e vale

$$c(y, x) = c(x, y).$$

Isso evidencia que a condutância é definida sobre a aresta, não dependendo do sentido na qual a mesma é percorrida, mas apenas dos vértices que ela une.

Definição 3.8. A **resistência** de uma aresta $e \in E$ é definida como o inverso da condutância, isto é

$$r(e) = \frac{1}{c(e)}.$$

A partir de agora, denotaremos uma rede por $(G, \{c(e)\}_{e \in E})$. Chamaremos os vértices de G de nós, e como já fixamos na Seção 1.9, denotaremos por $x \sim y$ para indicar que x e y são vizinhos, ou seja, que $\{x, y\} \in E$.

Até o final desta seção vamos mostrar que cada rede $(G, \{c(e)\}_{e \in E})$ possui uma cadeia de Markov reversível naturalmente associada. Mostraremos também que vale a volta, isto é, a toda cadeia de Markov reversível existe uma rede tal que o passeio aleatório a ela associado é exatamente a cadeia original.

Começamos fixando uma rede $(G, \{c(e)\})$, e definindo

$$c(x) := \sum_{y \sim x} c(x, y) \quad (3.3)$$

a condutância total no vértice x . Então definimos

$$P(x, y) := \frac{c(x, y)}{c(x)}. \quad (3.4)$$

É claro que $P(x, y) \geq 0$ e que $\sum_{y \in V} P(x, y) = 1$, de modo que P é uma matriz estocástica e portanto define uma cadeia de Markov sobre V , a qual é conhecida como **passeio aleatório em G ponderado pelas condutâncias nas arestas** $\{c(e)\}_{e \in E}$, ou alternativamente como **a cadeia de Markov associada à rede** $(G, \{c(e)\})$.

Esta cadeia pode ser pensada como uma generalização do passeio aleatório em um grafo definido na seção 1.9. Agora, quando encontra-se em um vértice x , ao invés de ter igual probabilidade de seguir cada aresta ($P(x, y) = 1/d_x$ como era no passeio aleatório simples), a probabilidade é proporcional à condutância de cada aresta ($P(x, y) = \frac{c(x, y)}{c(x)}$), sendo que uma aresta com maior condutância (e menor resistência) é mais provável de ser seguida. No nosso paralelo com redes elétricas, isso significa que a quantidade de corrente que flui por cada resistência é proporcional à sua condutância.

Em particular, o passeio aleatório simples $G = (V, E)$ é um caso especial em que todos os pesos são 1, ou seja, $c(e) = 1$ para todo $e \in E$. Neste caso

$$c(x) = \sum_{y \sim x} c(x, y) = \sum_{y \sim x} 1 = d_x$$

e

$$P(x, y) = \frac{c(x, y)}{c(x)} = \frac{1}{d_x}$$

se $\{x, y\} \in E$, exatamente como tínhamos na Seção 1.9.

A construção acima pode ser resumida na seguinte definição:

Definição 3.9. *A cada grafo $G = (V, E)$ podemos escolher um conjunto qualquer de valores positivos $\{c(e)\}_{e \in E}$ associados às arestas, os quais chamamos de condutâncias, definindo assim uma rede $(G, \{c(e)\})$. Existe uma **cadeia de Markov naturalmente associada à rede** $(G, \{c(e)\})$ cujas probabilidades de transição são dadas por*

$$P(x, y) = \frac{c(x, y)}{c(x)}, \quad (3.5)$$

onde $c(x) = \sum_{y \sim x} c(x, y)$. Tal cadeia também é conhecida como **passeio aleatório em G ponderado pelas condutâncias** $\{c(e)\}_{e \in E}$.

Teorema 3.10. *O passeio aleatório definido em (3.5) é reversível com respeito à medida de probabilidade π definida por*

$$\pi(x) := \frac{c(x)}{c_G}, \quad (3.6)$$

onde $c_G = \sum_{x \in V} c(x)$.

Demonstração. Note que $c(x)$ define uma medida em V , e que a definição de π acima é feita exatamente para normalizar a medida c , de modo que π é uma distribuição de probabilidade em V .

Para mostrar que a cadeia é reversível, basta verificar que satisfaz a equação do balanço detalhado (1.11). De fato,

$$\pi(x)P(x, y) = \frac{c(x)}{c_G} \frac{c(x, y)}{c(x)} = \frac{1}{c_G} c(x, y) = \frac{1}{c_G} c(y, x) = \frac{c(y, x)}{c(y)} \frac{c(y)}{c_G} = P(y, x)\pi(y),$$

uma vez que $c(x, y) = c(y, x)$. Portanto este processo é reversível com respeito à distribuição π definida em (3.6). Além disso, conforme observamos na seção 1.8, isto garante que tal π é a distribuição estacionária da cadeia. \square

Antes de prosseguirmos, note que

$$c_G := \sum_{x \in V} c(x) = \sum_{x \in V} \sum_{y \sim x} c(x, y). \quad (3.7)$$

Em particular, se o grafo não possuir loops ($c(x, x) = 0$ para todo $x \in V$), cada aresta $e = \{x, y\}$ é contada duas vezes no somatório em (3.7), e como $c(e) := c(x, y) = c(y, x)$, temos que

$$c_G = \sum_{x \in V} \sum_{y \sim x} c(x, y) = \sum_{x \in V} \sum_{y \in V \setminus \{x\}; y \sim x} c(x, y) = 2 \sum_{e \in E} c(e). \quad (3.8)$$

É claro que havendo *loops*, as arestas $\{x, x\}$ são contadas apenas uma vez em (3.7), não valendo esta última igualdade.

Observamos também que para fins práticos, é equivalente retirar uma aresta do grafo ou atribuir condutância 0 para tal aresta. Deste modo, se for interessante, podemos assumir que o grafo $G = (V, E)$ é completo, isto é, $\{x, y\} \in E$ para todo par $x, y \in V$, mas que algumas destas arestas possuam condutância 0.

Finalizamos esta seção demonstrando o resultado que citamos na introdução deste capítulo e que justifica todo o trabalho que tivemos para definir rigorosamente o conceito de redes elétricas.

Teorema 3.11. *Toda cadeia de Markov reversível em um espaço de estados finito pode ser vista como um passeio aleatório ponderado em uma rede.*

Demonstração. A prova baseia-se em construir uma rede cuja cadeia de Markov associada seja exatamente a cadeia reversível original.

Para isso, fixe uma cadeia de Markov P (em um espaço de estados finito S) reversível com respeito à distribuição π , isto é, uma matriz de transição P satisfazendo a equação do balanço detalhado

$$\pi(x)P(x, y) = \pi(y)P(y, x)$$

para todos $x, y \in S$.

Começamos definindo o grafo $G = (V, E)$. Para isso fixe o conjunto dos vértices como sendo $V = S$, e diremos que $e = \{x, y\}$ é uma aresta se $P(x, y) > 0$. Pela reversibilidade, $P(y, x) = \frac{\pi(x)}{\pi(y)}P(x, y)$, de modo que $P(x, y) > 0$ se, e somente se $P(y, x) > 0$. Assim não há perigo de haver arestas não direcionais e fica bem definido o grafo $G = (V, E)$.

Definimos a condutância na aresta $\{x, y\} \in E$ por

$$c(x, y) = \pi(x)P(x, y).$$

Pela reversibilidade

$$c(y, x) = \pi(y)P(y, x) = \pi(x)P(x, y) = c(x, y),$$

de modo que a condutância depende apenas da aresta $e = \{x, y\} = \{y, x\}$ e não do sentido no qual ela é percorrida, estando portanto bem definida.

Os passos acima definem uma rede $(G, \{c(e)\}_{e \in E})$. Nela a condutância em cada vértice x é

$$c(x) = \sum_{y \sim x} c(x, y) = \sum_{y \sim x} \pi(x)P(x, y) = \pi(x) \sum_{y \sim x} P(x, y) = \pi(x) \sum_{y \in S} P(x, y) = \pi(x).$$

Então, pela Definição 3.9, a cadeia de Markov naturalmente associada à rede é dada pela matriz de transição

$$\tilde{P}(x, y) = \frac{c(x, y)}{c(x)} = \frac{\pi(x)P(x, y)}{\pi(x)} = P(x, y),$$

mostrando que a cadeia de Markov associada a rede $(G, \{c(e)\}_{e \in E})$ que acabamos de definir é exatamente a cadeia original P . Em outras palavras, dada uma cadeia de Markov reversível P , construímos uma rede $(G, \{c(e)\}_{e \in E})$ cujo passeio aleatório associado é exatamente a cadeia original, demonstrando assim o teorema. □

3.3 Voltagem e fluxo de corrente

Para esta seção fixamos uma rede $(G, \{c(e)\}_{e \in E})$. Fixados dois pontos $a, z \in V$, uma **voltagem** é uma função W definida nos vértices da rede, harmônica em $V \setminus \{a, z\}$ com relação ao passeio

aleatório associado à rede. É consequência de Proposição 3.5 que, a partir do momento que escolhermos os valores $W(a)$ e $W(z)$, existe uma única extensão harmônica em $V \setminus \{a, z\}$, de modo que os valores $W(a), W(z)$ determinam completamente a voltagem.

A rede foi definida em grafos não orientados, com as arestas podendo ser percorridas em ambos os sentidos. Para alguns conceitos precisamos diferenciar quando percorremos a aresta $e = \{x, y\}$ no sentido de x para y de quando a percorremos no sentido oposto. Para isso vamos definir a **aresta orientada** \vec{xy} como sendo a aresta $\{x, y\}$ no sentido de x para y . Podemos também denotar esta aresta pelo par ordenado (x, y) .

Condutância e resistência são dois conceitos definidos inicialmente nas arestas não orientadas, dependendo apenas da aresta e não do sentido na qual a mesma é percorrida, tendo o mesmo valor para os dois sentidos da aresta e podendo ser estendidos para arestas orientadas como segue:

$$\begin{aligned} c(\vec{xy}) &= c(\vec{yx}) = c(x, y), \\ r(\vec{xy}) &= r(\vec{yx}) = r(x, y). \end{aligned}$$

Definimos um **fluxo** θ como sendo uma função antissimétrica nas arestas orientadas da rede, isto é, para todo $\{x, y\} \in E$,

$$\theta(\vec{xy}) = -\theta(\vec{yx}).$$

E o **divergente** de θ no ponto $x \in V$ é definido por

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

Note que o divergente de θ em x considera todo fluxo que “entra” em x e todo fluxo de “sai” de x , ou seja, considera o fluxo total em x . Além disso, para qualquer fluxo θ , vale

$$\sum_{x \in V} \operatorname{div} \theta(x) = \sum_{x \in V} \sum_{y \sim x} \theta(\vec{xy}) = \sum_{\{x, y\} \in E} [\theta(\vec{xy}) + \theta(\vec{yx})] = 0, \quad (3.9)$$

uma vez que na soma à esquerda da segunda igualdade cada aresta aparece nos dois sentidos, então reordenando esta soma obtemos o somatório da direita, onde cada termo é nulo visto que θ é antissimétrica por definição. Essa equação pode ser interpretada fisicamente como a necessidade de não haver fluxo sobrando no sistema.

Fixados dois pontos $a, z \in V$, dizemos que θ é um **fluxo entre a e z** se satisfaz

$$\operatorname{div} \theta(x) = 0, \quad \forall x \in V \setminus \{a, z\}. \quad (3.10)$$

Adiante veremos que isso significa, do ponto de vista físico, que nenhuma carga é criada ou destruída em um nó que não seja a ou z . A simples combinação das equações (3.9) e (3.10) garante que um fluxo entre a e z satisfaz

$$\operatorname{div} \theta(z) = -\operatorname{div} \theta(a). \quad (3.11)$$

Para um fluxo θ entre a e z , definimos sua **força** como sendo $\|\theta\| := \operatorname{div} \theta(a)$. E dizemos que um fluxo θ entre a e z é **unitário** quando $\|\theta\| = 1$.

Definição 3.12. Fixada uma rede $(G, \{c(e)\}_{e \in E})$, escolhermos dois estados $a, z \in V$ e uma voltagem W . Definimos a **corrente** associada à voltagem W por

$$I(\vec{xy}) := c(\vec{xy})[W(x) - W(y)]. \quad (3.12)$$

Note que pela própria definição a corrente satisfaz $I(x, y) = -I(y, x)$, de modo a ser um fluxo. Além disso, como W é harmônica em $V \setminus \{a, z\}$, para todo $x \in V \setminus \{a, z\}$,

$$\begin{aligned} \operatorname{div} I(x) &= \sum_{y \sim x} I(\vec{xy}) \\ &= \sum_{y \sim x} c(x, y)[W(x) - W(y)] \\ &= c(x) \sum_{y \sim x} P(x, y)W(x) - c(x) \sum_{y \sim x} P(x, y)W(y) \\ &= W(x)c(x) - c(x)W(x) = 0, \end{aligned}$$

mostrando que I é um fluxo entre a e z .

Além disso, é imediato da definição de corrente que se as arestas orientadas $\vec{e}_1, \dots, \vec{e}_m$ formam um ciclo orientado (isto é, $x_0, x_1, \dots, x_m \in V$, $\vec{e}_i = (x_{i-1}, x_i)$ e $x_m = x_0$), vale

$$\sum_{i=1}^m r(\vec{e}_i)I(\vec{e}_i) = \sum_{i=1}^m r(x_{i-1}, x_i)c(x_{i-1}, x_i)[W(x_{i-1}) - W(x_i)] = \sum_{i=1}^m [W(x_{i-1}) - W(x_i)] = 0$$

A definição de corrente e as propriedades acima são geralmente chamadas de “leis”, enunciadas abaixo:

- **Lei de Ohm:** Se $x \sim y$, a corrente $I(x, y)$ de x para y satisfaz

$$W(x) - W(y) = I(x, y)r(x, y).$$

- **Lei dos Nós de Kirchhoff:** A corrente é um fluxo entre a e z .
- **Lei das Malhas de Kirchhoff:** Se $x_0 \sim x_1 \sim \dots \sim x_{m-1} \sim x_m = x_0$ formam um ciclo, então

$$\sum_{i=1}^m r(x_{i-1}, x_i)I(x_{i-1}, x_i) = 0.$$

Note que sendo W uma voltagem entre a e z , deduzimos acima que uma função que satisfaz a Lei de Ohm também satisfaz as duas leis de Kirchhoff. Mas também vale a volta: Se uma função antissimétrica nas arestas satisfaz as duas leis de Kirchhoff, então ela satisfaz a Lei de Ohm, sendo assim um fluxo de corrente.

Fazendo um paralelo dos conceitos definidos acima com seus correspondentes na física como é feito em [11] (em questão de nomenclatura, a única diferença é que a função que aqui definimos como voltagem é referida como diferença de potencial), temos a seguinte interpretação: se conectarmos uma bateria a uma rede nos vértices a e z de modo que a diferença de potencial seja de $W(a)$ volts no vértice a e $W(z)$ volts no vértice z , então a diferença de potencial W nos vértices entrará em um estado de equilíbrio e uma corrente I fluirá através das arestas. Essas funções devem satisfazer as leis de Ohm e Kirchhoff. A Lei de Ohm nos diz que a corrente é diretamente proporcional à diferença de potencial entre os vértices e inversamente proporcional à resistência das arestas. Podemos interpretar a corrente como a taxa de fluxo de carga por unidade de tempo, de modo que a Lei dos Nós de Kirchhoff nos diz que a carga não pode ser criada ou destruída em nenhum nó que não sejam os extremos a ou z , ou seja, onde possuímos uma fonte externa ligada.

Continuando com o formalismo matemático para os conceitos da física, temos o seguinte resultado:

Proposição 3.13. *Seja θ um fluxo entre a e z satisfazendo a Lei das Malhas de Kirchhoff*

$$\sum_{i=1}^m r(\vec{e}_i)\theta(\vec{e}_i) = 0$$

para todo ciclo $\vec{e}_1, \dots, \vec{e}_m$, e tal que $\|\theta\| = \|I\|$. Então $\theta = I$.

Em particular, isso mostrar que existe um único fluxo de corrente unitário.

Demonstração. Considere $f := \theta - I$, e note que f é um fluxo entre a e z , isto é, $\text{div } f(x) = 0$ para todo $x \in S \setminus \{a, z\}$. De fato, para $x \in S \setminus \{a, z\}$, como tanto θ quanto I são fluxos entre a e z , $\text{div } \theta(x) = 0$ e $\text{div } I(x) = 0$, logo

$$\text{div } f(x) = \sum_{y \sim x} f(\vec{x}\vec{y}) = \sum_{y \sim x} [\theta(\vec{x}\vec{y}) - I(\vec{x}\vec{y})] = \sum_{y \sim x} \theta(\vec{x}\vec{y}) - \sum_{y \sim x} I(\vec{x}\vec{y}) = \text{div } \theta(x) - \text{div } I(x) = 0.$$

Além disso, $\text{div } f(x) = 0$ inclusive para $x = a, z$. De fato, para $x = a$, temos que $\text{div } \theta(a) = \|\theta\| = \|I\| = \text{div } I(a)$, logo vale

$$\text{div } f(a) = \sum_{y \sim a} f(\vec{a}\vec{y}) = \sum_{y \sim a} [\theta(\vec{a}\vec{y}) - I(\vec{a}\vec{y})] = \sum_{y \sim a} \theta(\vec{a}\vec{y}) - \sum_{y \sim a} I(\vec{a}\vec{y}) = \text{div } \theta(a) - \text{div } I(a) = 0.$$

E para $x = z$, como f é um fluxo entre a e z , a equação (3.11) garante que

$$\text{div } f(z) = -\text{div } f(a) = 0.$$

Também temos que tal f satisfaz a Lei das Malhas de Kirchhoff. De fato, seja $\vec{e}_1, \dots, \vec{e}_m$ um ciclo orientado (isto é, $\vec{e}_i = (x_{i-1}, x_i)$ e $x_m = x_0$). Como tanto θ quanto I satisfazem a Lei das Malhas, temos respectivamente que

$$\begin{aligned} \sum_{i=1}^m r(\vec{e}_i)\theta(\vec{e}_i) &= 0 \\ \text{e } \sum_{i=1}^m r(\vec{e}_i)I(\vec{e}_i) &= 0, \end{aligned}$$

donde segue que

$$\sum_{i=1}^m r(\vec{e}_i)f(\vec{e}_i) = \sum_{i=1}^m r(\vec{e}_i)[\theta(\vec{e}_i) - I(\vec{e}_i)] = \sum_{i=1}^m r(\vec{e}_i)\theta(\vec{e}_i) - \sum_{i=1}^m r(\vec{e}_i)I(\vec{e}_i) = 0.$$

Agora podemos mostrar que $f = 0$. Por absurdo, suponha que existe uma aresta orientada $\vec{e}_1 = (x_0, x_1)$ tal que $f(\vec{e}_1) > 0$. Como mostramos que

$$0 = \text{div } f(x_1) = \sum_{y \sim x_1} f(\vec{x}_1\vec{y}),$$

e este somatório contém o termo $f(\vec{x}_1\vec{x}_0) = -f(\vec{x}_0\vec{x}_1) = -f(\vec{e}_1) < 0$, precisa também conter algum termo positivo para ser nulo, logo deve existir um $x_2 \sim x_1$ tal que $f(\vec{x}_1\vec{x}_2) > 0$. Chamaremos $\vec{e}_2 := (x_1, x_2)$. Repetindo o mesmo argumento para \vec{e}_2 , e continuando este processo, vamos obter uma sequência de arestas orientadas onde f é estritamente positiva. Como a cadeia é finita, eventualmente esta sequência vai repetir algum nó, fechando assim um ciclo. Mas tal ciclo viola a Lei das Malhas, uma vez que nele

$$\sum r(\vec{e}_i)f(\vec{e}_i) > 0,$$

já que $r(\vec{e}_i) > 0$ por definição e ele foi construído de modo que $f(\vec{e}_i) > 0$. Chegamos assim a um absurdo, que surgiu do fato de supor que $f(\vec{e}_1) > 0$ para alguma aresta orientada \vec{e}_1 . Portanto devemos ter $f = 0$ e $\theta = I$. \square

Cabe notar que na demonstração acima, o que faz com que tenhamos $f = 0$ é essencialmente o fato de $\operatorname{div} f(a) = 0$. É normal que $\operatorname{div} f(x) = 0$ para todo $x \notin \{a, z\}$, mas quando temos também que $\operatorname{div} f(a) = 0$, isso significa que não há nenhum fluxo entrando na rede, o que faz com que o fluxo seja nulo em todos os pontos. Isso concorda com o que esperamos intuitivamente que aconteça em uma rede elétrica na física.

3.4 Resistência efetiva

Na seção anterior, fixada uma rede $(G, \{c(e)\})$ e dois pontos $a, z \in V$, definimos uma voltagem entre a e z como sendo qualquer função harmônica em $V \setminus \{a, z\}$ com relação ao passeio aleatório associado à rede. Note que fixados os valores para $W(a)$ e $W(z)$, digamos α e β respectivamente, a Proposição 3.5 juntamente com o Princípio da Unicidade, garantem que a única extensão harmônica em $V \setminus \{a, z\}$ é dada por

$$W(x) = \mathbb{E}_x[f(X_{\tau_{\{a,z\}}})],$$

onde $f : \{a, z\} \rightarrow \mathbb{R}$ é tal que $f(a) = \alpha$, $f(z) = \beta$. Como $f(X_{\tau_{\{a,z\}}}) \in \{\alpha, \beta\}$, temos pela definição de esperança que

$$\begin{aligned} \mathbb{E}_x[f(X_{\tau_{\{a,z\}}})] &= \alpha \mathbb{P}_x[f(X_{\tau_{\{a,z\}}}) = \alpha] + \beta \mathbb{P}_x[f(X_{\tau_{\{a,z\}}}) = \beta] \\ &= \alpha \mathbb{P}_x[X_{\tau_{\{a,z\}}} = a] + \beta \mathbb{P}_x[X_{\tau_{\{a,z\}}} = z] \\ &= \alpha \mathbb{P}_x[\tau_a < \tau_z] + \beta \mathbb{P}_x[\tau_z < \tau_a]. \end{aligned}$$

Em outras palavras, fixados os valores de $W(a)$ e $W(z)$, a única extensão harmônica em $V \setminus \{a, z\}$ é dada por

$$W(x) = W(a)\mathbb{P}_x(\tau_a < \tau_z) + W(z)\mathbb{P}_x(\tau_z < \tau_a), \quad (3.13)$$

ou seja, esta é a única voltagem entre a e z .

Fixamos uma rede $(G, \{c(e)\})$ e dois pontos $a, z \in V$. Cada voltagem W induz na rede um fluxo de corrente I por

$$I(\vec{xy}) = c(\vec{xy})[W(x) - W(y)].$$

Afirmamos que a razão

$$\frac{W(a) - W(z)}{\|I\|}$$

não depende da voltagem W escolhida. Ou seja, embora o numerador e o denominador dependam ambos dos valores escolhidos para $W(a)$, $W(z)$, a razão entre eles independe da escolha destes valores.

De fato, escolhida uma voltagem W qualquer

$$\begin{aligned} \|I\| &= \operatorname{div} I(a) \\ &= \sum_{y \sim a} I(\vec{ay}) \\ &= \sum_{y \sim a} c(a, y)[W(a) - W(y)] \\ &= \sum_{y \sim a} c(a, y)[W(a) - (W(a)\mathbb{P}_y(\tau_a < \tau_z) + W(z)\mathbb{P}_y(\tau_z < \tau_a))] \\ &= \sum_{y \sim a} c(a, y) [W(a)[1 - \mathbb{P}_y(\tau_a < \tau_z)] - W(z)\mathbb{P}_y(\tau_z < \tau_a)]. \end{aligned}$$

Agora note que $1 - \mathbb{P}_y(\tau_a < \tau_z) = \mathbb{P}_y(\tau_z \leq \tau_a) = \mathbb{P}_y(\tau_z < \tau_a)$ uma vez que a e z são distintos, logo

$$\begin{aligned} \|I\| &= \sum_{y \sim a} c(a, y) [W(a) \mathbb{P}_y(\tau_z < \tau_a)] - W(z) \mathbb{P}_y(\tau_z < \tau_a) \\ &= [W(a) - W(z)] \sum_{y \sim a} c(a, y) \mathbb{P}_y(\tau_z < \tau_a) \\ &= [W(a) - W(z)] \sum_{y \sim a} c(a) P(a, y) \mathbb{P}_y(\tau_z < \tau_a) \\ &= [W(a) - W(z)] c(a) \mathbb{P}_a(\tau_z < \tau_a^+). \end{aligned}$$

donde segue que

$$\frac{W(a) - W(z)}{\|I\|} = \frac{1}{\sum_{y \sim a} c(a, y) \mathbb{P}_y(\tau_z < \tau_a)} = \frac{1}{c(a) \mathbb{P}_a(\tau_z < \tau_a^+)}, \quad (3.14)$$

mostrando que a razão $\frac{W(a) - W(z)}{\|I\|}$ não depende dos valores escolhidos para $W(a), W(z)$, logo não depende da voltagem W escolhida, mas sim da rede e dos dois estados escolhidos. Cabe portanto a seguinte definição:

Definição 3.14. Dada uma rede $(G, \{c(e)\})$ e dois nós $a, z \in V$, definimos a **resistência efetiva** entre a e z por

$$\mathcal{R}(a \leftrightarrow z) := \frac{W(a) - W(z)}{\|I\|}, \quad (3.15)$$

onde W é qualquer voltagem entre a e z , e I é o fluxo de corrente associada à voltagem W .

Naturalmente, definimos a **condutância efetiva** como sendo o inverso da resistência

$$\mathcal{C}(a \leftrightarrow z) = \frac{1}{\mathcal{R}(a \leftrightarrow z)} = c(a) \mathbb{P}_a(\tau_z < \tau_a^+) = \sum_{y \sim a} c(a, y) \mathbb{P}_y(\tau_z < \tau_a). \quad (3.16)$$

O conceito de resistência efetiva pode ser estendido para um par de conjuntos disjuntos A e B . Para isso dizemos que W é uma voltagem entre A e B se W for harmônica em $V \setminus (A \cup B)$. Com as mesmas ferramentas de antes mostra-se que se $W(a) = \alpha$ para todo $a \in A$ e $W(b) = \beta$ para todo $b \in B$, então $W(x) = \alpha \mathbb{P}_x(\tau_A < \tau_B) + \beta \mathbb{P}_x(\tau_B < \tau_A)$. Fixada tal voltagem W , ela induz uma corrente por $I(\vec{x}\vec{y}) = c(x, y)[W(x) - W(y)]$ (quando nos referimos a conjuntos, uma corrente é um fluxo entre A e B no sentido de satisfazer $\text{div } I(x) = 0$ para todo $x \in V \setminus (A \cup B)$, cuja força é definida por $\|I\| = \sum_{a \in A} I(a)$). Esta corrente satisfaz

$$\begin{aligned} \|I\| &= \sum_{a \in A} \text{div } I(a) = \sum_{a \in A} \sum_{y \sim a} I(a\vec{y}) \\ &= \sum_{a \in A} \sum_{y \sim a} c(a, y) [W(a) - W(y)] \\ &= \sum_{a \in A} \sum_{y \sim a} c(a, y) [\alpha - (\alpha \mathbb{P}_y(\tau_A < \tau_B) + \beta \mathbb{P}_y(\tau_B < \tau_A))] \\ &= \sum_{a \in A} \sum_{y \sim a} c(a, y) [\alpha(1 - \mathbb{P}_y(\tau_A < \tau_B)) - \beta \mathbb{P}_y(\tau_B < \tau_A)]. \end{aligned}$$

Como os conjuntos A e B são disjuntos, $1 - \mathbb{P}_y(\tau_A < \tau_B) = \mathbb{P}_y(\tau_B < \tau_A)$. Utilizando tal fato no lado direito da igualdade acima ficamos com

$$\begin{aligned} \|I\| &= \sum_{a \in A} \sum_{y \sim a} c(a, y)(\alpha - \beta) \mathbb{P}_y(\tau_B < \tau_A) \\ &= (\alpha - \beta) \sum_{a \in A} \sum_{y \sim y} c(a, y) \mathbb{P}_y(\tau_B < \tau_A) \\ &= (\alpha - \beta) \sum_{a \in A} \sum_{y \sim a} c(a) P(a, y) \mathbb{P}_y(\tau_B < \tau_A) \\ &= (\alpha - \beta) \sum_{a \in A} c(a) \mathbb{P}_a(\tau_B < \tau_A^+), \end{aligned}$$

donde concluímos que a razão $\frac{\alpha - \beta}{\|I\|}$, onde α é o valor de W no conjunto A e β é o valor assumido por W no conjunto B , não depende da escolha da voltagem W .

Com base nisso definimos a **resistência efetiva** entre os conjuntos A e B por

$$\mathcal{R}(A \leftrightarrow B) := \frac{\alpha - \beta}{\|I\|} = \frac{1}{\sum_{a \in A} c(a) \mathbb{P}_a(\tau_B < \tau_A^+)},$$

e a condutância efetiva entre os conjuntos A e B por

$$\mathcal{C}(A \leftrightarrow B) := \frac{1}{\mathcal{R}(A \leftrightarrow B)} = \sum_{a \in A} c(a) \mathbb{P}_a(\tau_B < \tau_A^+).$$

É claro a partir das definições que $\mathcal{R}(a \leftrightarrow z)$ é o mesmo objeto que $\mathcal{R}(A \leftrightarrow B)$ quando $A = \{a\}$ e $B = \{z\}$. Visando simplificar a interpretação física, não exploraremos neste primeiro momento a resistência efetiva entre dois conjuntos, porém todos os resultados que obtivermos na sequência para um par de nós podem ser estendidos para um par de conjuntos.

É consequência do princípio do máximo que para todo $x \in V \setminus \{a, z\}$ os valores de $W(x)$ encontram-se no intervalo delimitado por $W(a)$ e $W(z)$. Deste modo, se $W(a) > W(z)$, temos que para todo $y \sim a$, $W(y) \leq W(a)$, logo $I(a\vec{y}) = c(a\vec{y})[W(a) - W(y)] \geq 0$, significando que do vértice a só “sai” corrente. Ao mesmo tempo, para todo $y \sim z$, $W(y) > W(z)$ e $I(y\vec{z}) \geq 0$, significando que no vértice z só “chega” corrente. Neste caso podemos escrever que a corrente I vai de a para z . Além disso, $\|I\| = \text{div } I(a) > 0$ e a resistência efetiva dada pela expressão (3.15) é positiva.

Escolhemos o nome resistência efetiva pelo fato de que, se substituíssemos toda a rede $(G, \{c(e)\})$ por uma nova rede contendo apenas os nós a e z e uma aresta os ligando com resistência $\mathcal{R}(a \leftrightarrow z)$, e então aplicássemos em ambas as redes uma mesma voltagem W (entre a e z), a quantidade de corrente fluindo de a para z , $\|I\|$, seria igual em ambos os casos.

A prova dessa afirmação é bastante simples. Seja I a corrente definida na rede original, que conforme vimos em (3.15), satisfaz $\|I\| = \frac{W(a) - W(z)}{\mathcal{R}(a \leftrightarrow z)}$. E seja \tilde{I} a corrente definida nesta nova rede, que contém apenas os nós a e z , a partir da voltagem W . Pela definição de corrente em (3.12),

$$\tilde{I}(a\vec{z}) = \frac{W(a) - W(z)}{\mathcal{R}(a \leftrightarrow z)},$$

logo

$$\|\tilde{I}\| = \text{div } \tilde{I}(a) = \sum_{y \sim a} \tilde{I}(a\vec{y}) = \tilde{I}(a\vec{z}) = \frac{W(a) - W(z)}{\mathcal{R}(a \leftrightarrow z)},$$

mostrando que $\|I\| = \|\tilde{I}\|$.

Definição 3.15. A função de Green de um passeio aleatório no tempo de parada τ é definida por

$$G_\tau(a, x) := \mathbb{E}_a \left(\sum_{t=0}^{\infty} \mathbf{1}_{\{X_t=x, \tau>t\}} \right), \quad (3.17)$$

representando o número esperado de visitas ao sítio x antes do tempo de parada τ quando o passeio começa no vértice a .

Lema 3.16. Seja $G_{\tau_B}(a, x)$ a função de Green definida acima em relação ao tempo de parada τ_B que indica o instante em que o passeio chega ao conjunto B pela primeira vez. Então

$$G_{\tau_B}(a, a) = c(a)\mathcal{R}(a \leftrightarrow B).$$

Demonstração. Recordamos que uma variável aleatória com distribuição geométrica de parâmetro p descreve a quantidade de repetições de um experimento de Bernoulli(p) necessárias para obtenção do primeiro sucesso. Tal variável aleatória tem esperança $\frac{1}{p}$.

Supondo que $X_0 = a$, a variável aleatória $\sum_{t=0}^{\infty} \mathbf{1}_{\{X_t=a, \tau_B>t\}}$ que conta o número de visitas ao sítio a antes de chegar em B possui distribuição geométrica de parâmetro $p = \mathbb{P}_a(\tau_B < \tau_a^+)$. Para se convencer disso considere o experimento de Bernoulli que representa chegar ao conjunto B antes de voltar ao sítio a , e tem probabilidade de sucesso $p = \mathbb{P}_a(\tau_B < \tau_a^+)$. Se $T \sim \text{Geom}(p)$ então:

- $T = 1$ significa obter sucesso na primeira tentativa, ou seja, chegar em B antes de retornar ao vértice a pela primeira vez. Neste caso passamos por a apenas em $t = 0$ e $\sum_{t=0}^{\infty} \mathbf{1}_{\{X_t=a, \tau_B>t\}} = 1$;
- $T = n$ significa que obtivemos sucesso apenas na n -ésima tentativa, ou seja, precisamos voltar $n - 1$ vezes ao vértice a antes de chegar em B pela primeira vez. Como começamos no vértice a , isso significa que $\sum_{t=0}^{\infty} \mathbf{1}_{\{X_t=a, \tau_B>t\}} = n$.

Deste modo temos que

$$G_{\tau_B}(a, a) := \mathbb{E}_a \left(\sum_{t=0}^{\infty} \mathbf{1}_{\{X_t=a, \tau_B>t\}} \right) = \frac{1}{\mathbb{P}_a(\tau_B < \tau_a^+)}.$$

Combinando com a identidade (3.16) o resultado segue. □

3.5 Simplificando redes

A identidade (3.16) mostra que as propriedades físicas de uma rede como a resistência efetiva podem fornecer informações a respeito da cadeia de Markov associada. Nesta seção veremos como é possível simplificar uma rede sem alterar suas propriedades do mesmo modo que podemos fazer na física quando somamos resistências em série e condutâncias em paralelo. Para tal é útil ampliarmos um pouco a definição de rede.

3.5.1 Redes em multigrafos

Redes foram definidas sobre grafos (possivelmente com loops) definidos na seção 1.9. Podemos estender sua definição para multigrafos, os quais permitem que dois vértices x e y possuam mais de uma aresta conectando-os. Neste caso não faz mais sentido denotarmos uma aresta que liga x e y por $e_1 = \{x, y\}$, já que podem existir inúmeras arestas ligando estes dois pontos. Passamos então a denotar por $E(xy)$ o conjunto de todas as arestas que ligam os pontos x e y , e escrevemos $e_1 \in E(xy)$. Cabe notar que $E(xy) = E(yx)$ como igualdade de conjuntos, visto que as arestas não possuem sentido. Ainda, se quisermos falar em arestas orientadas, as quais diferenciaremos por uma seta, denotaremos por $E^+(xy)$ o conjunto das arestas orientadas que saem de x e chegam em y . Assim, se $e \in E(xy)$, ficam naturalmente associadas as arestas orientadas $\vec{e} \in E^+(xy)$ que indica percorrer a aresta e no sentido de x para y , e $-\vec{e} \in E^+(yx)$ que indica percorrer a mesma aresta no sentido contrário. Por fim denotaremos o conjunto de todas as arestas conectadas em um ponto x por $E_x = \cup_{y \in V} E(xy)$, e o conjunto de todas as arestas orientadas que partem de x por $E_x^+ = \cup_{y \in V} E^+(xy)$.

Para definir uma rede em um grafo bastava atribuir condutâncias às arestas. Como isto dependia das arestas, e não dos pares de vértices, assim mantemos nos multigrafos. Neste caso, a equação (3.2) não faz mais sentido matemático, mas podemos manter sua interpretação definindo

$$c(x, y) = \sum_{e \in E(xy)} c(e), \quad (3.18)$$

a soma das condutâncias em todas as arestas que ligam x e y . Neste caso é claro que $c(x, y) = c(y, x)$. A condutância total no vértice x também não pode mais ser definida por (3.3), mas mantém a mesma interpretação se definirmos

$$c(x) = \sum_{e \in E_x} c(e) = \sum_{y \sim x} \sum_{e \in E(xy)} c(e). \quad (3.19)$$

Com estas novas definições, vale a mesma relação entre a condutância total no vértice x e a condutância das arestas que ligam x e y que vimos na equação (3.3):

$$c(x) = \sum_{y \in V} c(x, y) \quad (3.20)$$

A construção acima generaliza o conceito de redes para multigrafos, sendo as redes definidas em grafos um caso particular onde existe apenas uma aresta ligando dois vértices.

A cada rede $(G, \{c(e)\})$ definida em um multigrafo G , existe uma cadeia de Markov naturalmente associada, construída do seguinte modo: Para cada aresta e que liga x e y , define

$$P^e(x, y) := \frac{c(e)}{c(x)}$$

a probabilidade de transitar de x para y através da aresta e . Naturalmente define-se a probabilidade de transitar de x para y como a soma das probabilidades de, a partir de x , percorrer cada uma das arestas que liga x a y , isto é,

$$P(x, y) = \sum_{e \in E(xy)} P^e(x, y). \quad (3.21)$$

Neste caso, vale

$$P(x, y) = \sum_{e \in E(xy)} \frac{c(e)}{c(x)} = \frac{1}{c(x)} \sum_{e \in E(xy)} c(e) = \frac{c(x, y)}{c(x)}. \quad (3.22)$$

A cadeia definida acima é reversível com respeito à probabilidade

$$\pi(x) = \frac{c(x)}{c_G},$$

onde $c_G = \sum_{x \in V} c(x)$.

Cabe observar que o multigrafo é apenas uma ferramenta para permitirmos arestas paralelas, mas no final das contas estes caminhos a mais não são estudados, já que a única coisa que nos interessa é a probabilidade de ir de um vértice para outro. Claramente, como todo grafo é um multigrafo, a toda cadeia reversível existe uma rede em um multigrafo associada.

Todos os conceitos e resultados obtidos para redes em grafos podem ser estendidos para multigrafos, tomando as devidas precauções. Não tomaremos tempo com tais verificações, que são facilmente obtidas a partir das definições feitas nesta seção. Entretanto destacaremos a extensão do conceito de corrente, que era definida nas arestas para redes em grafos, e assim manteremos para redes em multigrafos.

Definição 3.17. Fixada uma rede $(G, \{c(e)\}_{e \in E})$, escolhamos dois estados $a, z \in V$ e uma voltagem W . Definimos a **corrente** associada a W pondo, para cada aresta orientada $\vec{e} \in E^+(x, y)$

$$I(\vec{e}) := c(\vec{e})[W(x) - W(y)]. \quad (3.23)$$

Ainda podemos manter a definição de corrente entre dois vértices pondo

$$I(\vec{x}\vec{y}) = \sum_{\vec{e} \in E^+(xy)} I(\vec{e}),$$

e o divergente de I num vértice x continua sendo expresso por

$$\operatorname{div} I(x) = \sum_{\vec{e} \in E_x^+} I(\vec{e}) = \sum_{y \in V} \sum_{\vec{e} \in E^+(xy)} I(\vec{e}) = \sum_{y \in V} I(\vec{x}\vec{y}).$$

Também lembremos, por curiosidade, que uma função h é dita **harmônica** no vértice $x \in S$ (em relação a P) se

$$h(x) = Ph(x) = \sum_{y \in V} P(x, y)h(y).$$

Se P for a matriz do passeio aleatório associado ao multigrafo $(G, \{c(e)\})$, com a notação desta seção, podemos reescrever esta expressão em função das arestas, obtendo que h é harmônica em x se

$$h(x) = \sum_{y \in V} P(x, y)h(y) = \sum_{y \in V} \sum_{e \in E(xy)} P^e(x, y)h(y).$$

3.5.2 Simplificação de redes

No decorrer desta seção estaremos assumindo que G é um grafo (não um multigrafo), com exceção na Proposição 3.18 das duas arestas e_1 e e_2 conectando os vértices v_1 e v_2 . Tais resultados são facilmente estendidos (apenas com uma notação mais complicada) para o caso geral, onde G é um multigrafo, permitindo arestas paralelas em qualquer vértice.

Proposição 3.18 (Condutâncias em paralelo são somadas). *Suponha que os vértices $v_1, v_2 \in V$ de uma rede $(G, \{c(e)\})$ sejam conectados por duas arestas e_1 e e_2 , com condutâncias $c(e_1)$ e $c(e_2)$ respectivamente. Considere $(\tilde{G}, \{\tilde{c}(e)\})$ uma cópia da rede original, onde as arestas e_1, e_2 são substituídas por uma nova aresta e com condutância $\tilde{c}(e) = c(e_1) + c(e_2)$. Se $\tilde{\mathcal{R}}(a \leftrightarrow z)$*

representar a resistência entre a e z na nova rede e $\mathcal{R}(a \leftrightarrow z)$ a resistência entre a e z na rede original, vale $\tilde{\mathcal{R}}(a \leftrightarrow z) = \mathcal{R}(a \leftrightarrow z)$.

Mais precisamente, se W for uma voltagem entre a e z em $(G, \{c(e)\})$, então W continua sendo voltagem entre a e z em $(\tilde{G}, \{\tilde{c}(e)\})$. Além disso, na nova rede, a corrente associada a W continua igual em todas as arestas com exceção de e , onde a corrente passa a ser expressa por $\tilde{I}(\vec{e}) = I(\vec{e}_1) + I(\vec{e}_2)$.



Figura 3.1: Duas arestas paralelas podem ser substituídas por uma única aresta, cuja condutância é a soma das condutâncias nas arestas originais.

Demonstração. Seja $G = (V, E)$ o grafo da rede original e $\tilde{G} = (\tilde{V}, \tilde{E})$ o grafo da rede criada pela simplificação acima. Nenhum vértice foi criado ou destruído, de modo que $\tilde{V} = V$. Já em relação às arestas, substituímos as arestas e_1 e e_2 pela aresta e , de modo que $\tilde{E} = (E \setminus \{e_1, e_2\}) \cup \{e\}$.

Denotando por \tilde{c} as condutâncias nas arestas da nova rede, por construção, $\tilde{c}(f) = c(f)$ para todo $f \in \tilde{E}$ com $f \neq e$ e $\tilde{c}(e) = c(e_1) + c(e_2)$. Afirmamos que para todo par de estados $x, y \in \tilde{V}$ a condutância entre eles continua a mesma, isto é, $\tilde{c}(x, y) = c(x, y)$. Para ver isto separamos em casos: Se $x \notin \{v_1, v_2\}$, então para qualquer $y \in \tilde{V}$ nenhuma aresta de x para y foi alterada em relação à rede original, de modo que a condutância entre estes dois vértices mantém-se igual; para $x = v_1$ ($x = v_2$ é análogo) e $y \neq v_2$ também não houve alteração de arestas, e a condutância mantém-se igual; e por fim, para $x = v_1$ e $y = v_2$, temos que em relação à rede original as arestas e_1 e e_2 foram substituídas pela aresta e , de modo que $\tilde{c}(v_1, v_2) = \tilde{c}(e) = c(e_1) + c(e_2) = c(v_1, v_2)$, mostrando que a condutância também mantém-se igual.

É consequência imediata do fato acima e da equação (3.20) que a condutância total em cada vértice $v \in \tilde{V}$ dada por $\tilde{c}(v) = \sum_{y \in \tilde{V}} \tilde{c}(x, y)$ não se altera, isto é, $\tilde{c}(v) = c(v)$ para todo $v \in \tilde{V}$. Assim, se P e \tilde{P} são as matrizes de transição dos passeios aleatórios associados às redes $(G, \{c(e)\})$ e $(\tilde{G}, \{\tilde{c}(e)\})$ respectivamente, temos que para todo $x, y \in \tilde{V} = V$,

$$\tilde{P}(x, y) = \frac{\tilde{c}(x, y)}{\tilde{c}(x)} = \frac{c(x, y)}{c(x)} = P(x, y).$$

Definimos $\tilde{W} : \tilde{V} \rightarrow \mathbb{R}$ por $\tilde{W}(x) = W(x)$ para todo $x \in \tilde{V} = V$. Como $\tilde{P} = P$ o fato de W ser harmônica em relação à P em $V \setminus \{a, z\}$ faz com que \tilde{W} seja harmônica em relação à \tilde{P} em $\tilde{V} \setminus \{a, z\}$. Assim \tilde{W} é uma voltagem entre a e z na rede $(\tilde{G}, \{\tilde{c}(e)\})$ onde define uma corrente dada por

$$\tilde{I}(\vec{f}) := \tilde{c}(\vec{f})[\tilde{W}(x) - \tilde{W}(y)]$$

para todos os vértices $x, y \in \tilde{V}$ e toda aresta orientada $\vec{f} \in E^+(xy)$.

Na rede original a corrente associada à voltagem W é

$$I(\vec{f}) := c(\vec{f})[W(x) - W(y)],$$

logo, para toda aresta orientada \vec{f} com $f \in \tilde{E}$ e $f \neq e$, como $\tilde{c}(f) = c(f)$ e $\tilde{W} = W$, é claro que $\tilde{I}(\vec{f}) = I(\vec{f})$. Já para $f = e$ temos

$$\tilde{I}(\vec{e}) = \tilde{c}(e)[\tilde{W}(v_1) - \tilde{W}(v_2)] = [c(e_1) + c(e_2)][W(v_1) - W(v_2)] = I(e_1) + I(e_2).$$

Assim sendo, temos que se $a \neq v_1, v_2$,

$$\|\tilde{I}\| = \operatorname{div} \tilde{I}(a) = \sum_{\vec{f} \in \tilde{E}_a^+} \tilde{I}(\vec{f}) = \sum_{\vec{f} \in E_a^+} I(\vec{f}) = \operatorname{div} I(a) = \|I\|,$$

uma vez que \tilde{E}_a não contém a aresta e . E se $a = v_1$ ($a = v_2$ é análogo),

$$\begin{aligned} \|\tilde{I}\| &= \operatorname{div} \tilde{I}(a) = \sum_{\vec{f} \in \tilde{E}_a^+} \tilde{I}(\vec{f}) \\ &= \sum_{\vec{f} \in \tilde{E}_a^+, f \neq e} \tilde{I}(\vec{f}) + \tilde{I}(\vec{e}) \\ &= \sum_{\vec{f} \in E_a^+, f \neq e_1, e_2} I(\vec{f}) + I(\vec{e}_1) + I(\vec{e}_2) \\ &= \sum_{\vec{f} \in E_a^+} I(\vec{f}) = \operatorname{div} I(a) = \|I\|. \end{aligned}$$

Em qualquer caso, $\|\tilde{I}\| = \|I\|$, de modo que a resistência efetiva entre a e z nesta nova rede satisfaz

$$\tilde{\mathcal{R}}(a \leftrightarrow z) = \frac{\tilde{W}(a) - \tilde{W}(z)}{\|\tilde{I}\|} = \frac{W(a) - W(z)}{\|I\|} = \mathcal{R}(a \leftrightarrow z),$$

ou seja, é igual a resistência efetiva entre a e z na rede original, mostrando que a simplificação da rede definida no enunciado não altera a resistência efetiva. \square

Proposição 3.19 (Resistências em Série são somadas). *Suponha que os vértices $v_1, v_2, v_3 \in V$ de uma rede $(G, \{c(e)\})$ sejam conectados em série, isto é, $e_1 = \{v_1, v_2\}$, $e_2 = \{v_2, v_3\} \in E$ e v_2 não possui nenhuma outra aresta além das duas acima. Considere $(\tilde{G}, \{\tilde{c}(e)\})$ uma cópia da rede original, onde o vértice v_2 é excluído, e as arestas e_1, e_2 são substituídas por uma única aresta $e = \{v_1, v_3\}$ com resistência $r(e) = r(e_1) + r(e_2)$. Se $\tilde{\mathcal{R}}(a \leftrightarrow z)$ representar a resistência efetiva entre a e z na nova rede e $\mathcal{R}(a \leftrightarrow z)$ a resistência efetiva entre a e z na rede original, vale $\tilde{\mathcal{R}}(a \leftrightarrow z) = \mathcal{R}(a \leftrightarrow z)$.*

Mais precisamente, se W for uma voltagem entre a e z em $(G, \{c(e)\})$, então definindo \tilde{W} igual a W nos vértices da nova rede, ele é voltagem entre a e z em $(\tilde{G}, \{\tilde{c}(e)\})$. Além disso, se I é a corrente associada à W na rede original e \tilde{I} é a corrente associada a \tilde{W} na nova rede, então $\tilde{I} = I$ em todas as arestas herdadas da rede original, e a corrente na aresta criada é dada por $\tilde{I}(\vec{e}) = I(\vec{e}_1) = I(\vec{e}_2)$.



Figura 3.2: Duas arestas em série podem ser substituídas por uma única aresta, cuja resistência é a soma das resistências das arestas originais.

Demonstração. Façamos o caso $v_1, v_3 \notin \{a, z\}$. Seja $G = (V, E)$ o grafo da rede original e $\tilde{G} = (\tilde{V}, \tilde{E})$ o grafo da rede criada pela simplificação acima. Quanto aos vértices, a única alteração é que excluímos o vértice v_2 , ficando com $\tilde{V} = V \setminus \{v_2\}$, enquanto que para as arestas, temos que $\tilde{E} = (E \cup \{e\}) \setminus \{e_1, e_2\}$.

Denotando por \tilde{c} as condutâncias nas arestas da nova rede, por construção, $\tilde{c}(f) = c(f)$ para todo $f \in \tilde{E}$ com $f \neq e$. Já para a aresta e , como $\tilde{r}(e) = r(e_1) + r(e_2) = \frac{1}{c(e_1)} + \frac{1}{c(e_2)}$, temos que

$$\tilde{c}(e) = \frac{1}{\tilde{r}(e)} = \frac{1}{\frac{1}{c(e_1)} + \frac{1}{c(e_2)}} = \frac{1}{\frac{c(e_1) + c(e_2)}{c(e_1)c(e_2)}} = \frac{c(e_1)c(e_2)}{c(e_1) + c(e_2)}. \quad (3.24)$$

Para a condutância total nos vértices, como as únicas modificações encontram-se entre os vértices v_1 e v_3 , temos que $\tilde{c}(x) = c(x)$ para todo $x \in \tilde{V}$ com $x \neq v_1, v_3$. Para o vértice v_1 temos que, em relação à condutância original, foi retirada a aresta e_1 que o ligava ao vértice v_2 com condutância $c(e_1)$, e adicionada a aresta e ligando-o ao vértice v_3 , com condutância dada por (3.24), assim

$$\begin{aligned} \tilde{c}(v_1) &= c(v_1) - c(e_1) + \tilde{c}(e) \\ &= c(v_1) - c(e_1) + \frac{c(e_1)c(e_2)}{c(e_1) + c(e_2)} \\ &= c(v_1) - \frac{c(e_1)^2}{c(e_1) + c(e_2)} \\ &= c(v_1) - \frac{c(e_1)^2}{c(v_2)} \\ &= \frac{c(v_1)c(v_2) - c(e_1)^2}{c(v_2)}, \end{aligned} \quad (3.25)$$

uma vez que como v_2 é vizinho, na rede original, apenas de v_1 e v_3 pelas arestas e_1 e e_2 respectivamente, temos que $c(v_2) = c(e_1) + c(e_2)$. Analogamente, para o vértice v_3 temos que

$$\tilde{c}(v_3) = \frac{c(v_3)c(v_2) - c(e_2)^2}{c(v_2)}.$$

Para a matriz de transição \tilde{P} do passeio aleatório associado à rede $(\tilde{G}, \{\tilde{c}(e)\})$ temos que $\tilde{P}(x, y) = P(x, y)$ para todo $x, y \in \tilde{V}$ com $x \neq v_1, v_2$, uma vez que neste caso, nem a condutância total no vértice x nem a condutância na respectiva aresta foi alterada. Para o caso $x = v_1$ e $y \in \tilde{V} \setminus \{v_3\}$ (raciocínio análogo para $x = v_3, y \in \tilde{V} \setminus \{v_1\}$), temos que a condutância da aresta segue inalterada, enquanto que a condutância total em $x = v_1$ é dada por (3.25), assim ficamos com

$$\tilde{P}(v_1, y) = \frac{\tilde{c}(v_1, y)}{\tilde{c}(v_1)} = c(v_1, y) \cdot \frac{c(v_2)}{c(v_1)c(v_2) - c(e_1)^2} = \frac{c(v_1)c(v_2)P(v_1, y)}{c(v_1)c(v_2) - c(e_1)^2}, \quad (3.26)$$

uma vez que $P(v_1, y) = \frac{c(v_1, y)}{c(v_1)}$. E finalmente, para o caso $x = v_1, y = v_3$, temos que $\tilde{c}(v_1, v_3) = \tilde{c}(e)$, logo das equações (3.24) e (3.25) e do fato já observado de $c(v_2) = c(e_1) + c(e_2)$, obtemos que

$$\tilde{P}(v_1, v_3) = \frac{\tilde{c}(e)}{\tilde{c}(v_1)} = \frac{c(e_1)c(e_2)}{c(v_1)c(v_2) - c(e_1)^2}. \quad (3.27)$$

Isso determina completamente a matriz \tilde{P} do passeio aleatório associado à rede $(\tilde{G}, \{\tilde{c}(e)\})$.

Pela construção de $\tilde{V} = V \setminus \{v_2\}$, fica bem definida $\tilde{W} : \tilde{V} \rightarrow \mathbb{R}$ por $\tilde{W}(v) = W(v)$ para todo $v \in \tilde{V}$. Verificamos que ela é uma voltagem na rede $(\tilde{G}, \{\tilde{c}(e)\})$ mostrando que é harmônica em $\tilde{V} \setminus \{a, z\}$. Para isso fixamos $x \in \tilde{V} \setminus \{a, z\}$. Se $x \neq v_1, v_3$, então $\tilde{P}(x, y) = P(x, y)$ para todo $y \sim x$, e como $\tilde{W}(y) = W(y)$, a harmonicidade de \tilde{W} em x é imediata da harmonicidade da W em x . Já se $x = v_1$ ($x = v_3$ é análogo), como os vizinhos de v_1 em $\tilde{V} \setminus \{v_3\}$ são os mesmo

vizinhos que ele possui em $V \setminus \{v_2\}$ (veja a Figura 3.2), a harmonicidade de W na rede original garante-nos que

$$\begin{aligned}
 \sum_{y \in \tilde{V}} \tilde{P}(v_1, y) \tilde{W}(y) &= \tilde{P}(v_1, v_3) W(v_3) + \sum_{y \in \tilde{V} \setminus \{v_3\}} \tilde{P}(v_1, y) W(y) \\
 &\stackrel{(3.26)}{=} \tilde{P}(v_1, v_3) W(v_3) + \frac{c(v_1)c(v_2)}{c(v_1)c(v_2) - c(e_1)^2} \sum_{y \in \tilde{V} \setminus \{v_3\}} P(v_1, y) W(y) \\
 &= \tilde{P}(v_1, v_3) W(v_3) + \frac{c(v_1)c(v_2)}{c(v_1)c(v_2) - c(e_1)^2} \sum_{y \in V \setminus \{v_2\}} P(v_1, y) W(y) \\
 &= \tilde{P}(v_1, v_3) W(v_3) + \frac{c(v_1)c(v_2)}{c(v_1)c(v_2) - c(e_1)^2} [W(v_1) - P(v_1, v_2) W(v_2)]. \quad (3.28)
 \end{aligned}$$

Como W é harmônica também em v_2 na rede original, cujos únicos vizinhos são os vértices v_1 e v_3 , temos que

$$W(v_2) = P(v_2, v_1) W(v_1) + P(v_2, v_3) W(v_3) = \frac{c(e_1)}{c(v_2)} W(v_1) + \frac{c(e_2)}{c(v_2)} W(v_3). \quad (3.29)$$

Então, utilizando na equação (3.28) a expressão para $W(v_2)$ obtida em (3.29) e a forma para $\tilde{P}(v_1, v_3)$ da equação (3.27), ficamos com

$$\begin{aligned}
 \sum_{y \in \tilde{V}} \tilde{P}(v_1, y) \tilde{W}(y) &= \frac{c(e_1)c(e_2)}{c(v_1)c(v_2) - c(e_1)^2} W(v_3) + \frac{c(v_1)c(v_2)}{c(v_1)c(v_2) - c(e_1)^2} W(v_1) \\
 &\quad - \frac{c(v_1)c(v_2)P(v_1, v_2)}{c(v_1)c(v_2) - c(e_1)^2} W(v_2) \\
 &= \frac{c(e_1)c(e_2)}{c(v_1)c(v_2) - c(e_1)^2} W(v_3) + \frac{c(v_1)c(v_2)}{c(v_1)c(v_2) - c(e_1)^2} W(v_1) \\
 &\quad - \frac{c(v_1)c(v_2)P(v_1, v_2)}{c(v_1)c(v_2) - c(e_1)^2} \left[\frac{c(e_1)}{c(v_2)} W(v_1) + \frac{c(e_2)}{c(v_2)} W(v_3) \right] \\
 &= \frac{1}{c(v_1)c(v_2) - c(e_1)^2} [c(e_1)c(e_2)W(v_3) + c(v_1)c(v_2)W(v_1) \\
 &\quad - c(v_1)P(v_1, v_2)c(e_1)W(v_1) - c(v_1)P(v_1, v_2)c(e_2)W(v_3)].
 \end{aligned}$$

Reagrupando os termos, e utilizando que $P(x, y)c(x) = c(x, y)$, obtemos que

$$\begin{aligned}
 \sum_{y \in \tilde{V}} \tilde{P}(v_1, y) \tilde{W}(y) &= \frac{1}{c(v_1)c(v_2) - c(e_1)^2} [W(v_1)[c(v_1)c(v_2) - c(v_1)P(v_1, v_2)c(e_1)] \\
 &\quad - W(v_3)[c(e_1)c(e_2) - c(v_1)P(v_1, v_2)c(e_2)] \\
 &= \frac{1}{c(v_1)c(v_2) - c(e_1)^2} [W(v_1)[c(v_1)c(v_2) - c(v_1, v_2)c(e_1)] \\
 &\quad - W(v_3)[c(e_1)c(e_2) - c(v_1, v_2)c(e_2)] \\
 &= \frac{1}{c(v_1)c(v_2) - c(e_1)^2} [W(v_1)[c(v_1)c(v_2) - c(e_1)^2] \\
 &\quad - W(v_3)[c(e_1)c(e_2) - c(e_1)c(e_2)] \\
 &= W(v_1),
 \end{aligned}$$

mostrando que \tilde{W} é harmônica em v_1 .

Esta voltagem \tilde{W} gera uma corrente em $(\tilde{G}, \{\tilde{c}(e)\})$ dada por

$$\tilde{I}(x\vec{y}) = \tilde{c}(x\vec{y})[\tilde{W}(x) - \tilde{W}(y)].$$

Comparando com o fluxo de corrente I associado à voltagem W na rede original, temos que $\tilde{I}(f) = I(f)$ para toda aresta $f \in \tilde{E} \setminus \{e\}$, uma vez que a condutância destas arestas não foi alterada e $\tilde{W} = W$. Além disso, para a aresta e recém criada, como $\tilde{W} = W$ temos que

$$\tilde{I}(\vec{e}) = \tilde{I}(v_1\vec{v}_3) = \tilde{c}(e)[\tilde{W}(v_1) - \tilde{W}(v_3)] = \frac{c(e_1)c(e_2)}{c(e_1) + c(e_2)}[W(v_1) - W(v_3)].$$

Ao mesmo tempo, para a corrente I na rede original, em função da harmonicidade de W em v_2 (equação (3.29)), temos que para $\vec{e}_1 = v_1\vec{v}_2$,

$$\begin{aligned} I(\vec{e}_1) &= c(e_1)[W(v_1) - W(v_2)] \\ &= c(e_1) \left[W(v_1) - \left[\frac{c(e_1)}{c(v_2)}W(v_1) + \frac{c(e_2)}{c(v_2)}W(v_3) \right] \right] \\ &= c(e_1) \left[\frac{W(v_1)c(v_2) - W(v_1)c(e_1) - W(v_3)c(e_2)}{c(v_2)} \right] \\ &= c(e_1) \left[\frac{W(v_1)[c(e_1) + c(e_2)] - W(v_1)c(e_1) - W(v_3)c(e_2)}{c(e_1) + c(e_2)} \right] \\ &= c(e_1) \left[\frac{[W(v_1) - W(v_3)]c(e_2)}{c(e_1) + c(e_2)} \right] \\ &= \frac{c(e_1)c(e_2)}{c(e_1) + c(e_2)} [W(v_1) - W(v_3)] \end{aligned}$$

Analogamente, verifica-se que

$$I(e_2) = I(v_2\vec{v}_3) = \frac{c(e_1)c(e_2)}{c(e_1) + c(e_2)}[W(v_1) - W(v_3)],$$

mostrando que $\tilde{I}(\vec{e}) = I(\vec{e}_1) = I(\vec{e}_2)$.

Como estamos supondo $a \neq v_1, v_3$, temos pelo que vimos antes que $\tilde{I}(a\vec{y}) = I(a\vec{y})$ para todo $y \in V$, de modo que

$$\|\tilde{I}\| = \operatorname{div} \tilde{I}(a) = \sum_{y \in \tilde{V}} \tilde{I}(a\vec{y}) = \sum_{y \in V} I(a\vec{y}) = \operatorname{div} I(a) = \|I\|. \quad (3.30)$$

Portanto, como $\tilde{W} = W$,

$$\tilde{\mathcal{R}}(a \leftrightarrow z) = \frac{\tilde{W}(a) - \tilde{W}(z)}{\|\tilde{I}\|} = \frac{W(a) - W(z)}{\|I\|} = \mathcal{R}(a \leftrightarrow z),$$

mostrando que tal simplificação não altera a resistência efetiva entre a e z . \square

Observação 3.20. Para o caso $v_1 = a$ (os demais casos não cobertos na demonstração seguem a mesma argumentação), as contas continuam iguais, assim como a construção da matriz \tilde{P} , da voltagem \tilde{W} e da corrente \tilde{I} , com a diferença de que não é necessário verificar a harmonicidade de \tilde{W} em a . Além disso a relação $\tilde{I}(\vec{e}) = I(\vec{e}_1) = I(\vec{e}_2)$, que também pode ser escrita como $\tilde{I}(a\vec{v}_3) = I(a\vec{v}_2) = I(v_2\vec{v}_3)$, continua valendo. Assim, como a única alteração em relação ao vértice $a = v_1$ de G para \tilde{G} é a substituição da aresta $\vec{e}_1 = a\vec{v}_2$ pela aresta $\vec{e} = a\vec{v}_3$, e como $\tilde{I}(a\vec{v}_3) = I(a\vec{v}_2)$ e $\tilde{I}(a\vec{y}) = I(a\vec{y})$ para todo $y \sim a$, continua valendo a igualdade dos somatórios em (3.30). Deste modo $\|\tilde{I}\| = \|I\|$ e valem todas as conclusões do teorema neste caso.

Proposição 3.21 (Colagem). *Em uma rede $(G, \{c(e)\})$, a colagem de dois vértices v e w consiste em identificá-los por um novo vértice, denotado por $\delta = v\tilde{w}$, e manter todas as arestas existentes com as condutâncias originais. Tal processo define um nova rede $\{\tilde{G}, \{\tilde{c}(e)\}\}$, possivelmente com loops e arestas paralelas. Se W for uma voltagem entre a e z na rede original satisfazendo $W(v) = W(w)$, então a função \tilde{W} definida por $\tilde{W}(\delta) = W(v) = W(w)$ e $\tilde{W}(y) = W(y)$ para todo $y \neq \delta$ é uma voltagem entre a e z na nova rede. Como todas as arestas originais continuam presentes, a corrente nelas induzida por \tilde{W} é a mesma corrente que era induzida por W . Além disso, a resistência $\tilde{\mathcal{R}}(a \leftrightarrow z)$ entre a e z na nova rede é igual a resistência $\mathcal{R}(a \leftrightarrow z)$ entre a e z na rede original.*



Figura 3.3: O processo de colagem de dois vértices mantém todas as arestas do grafo original.

Observação 3.22. *O processo de colagem mantém todas as arestas, sendo a condição $W(v) = W(w)$ necessária para que $\tilde{W} = W$ continue harmônica nos vizinhos de v, w .*

Para exemplificar considere como exemplo a rede da Figura 3.3. Suponha que $W(v) \neq W(w)$. Fazendo a colagem destes vértices, definimos $\tilde{W}(x) = W(x)$ para todo $x \neq \delta$. Afirmando que não há nenhum modo de definir $\tilde{W}(\delta)$ de modo que \tilde{W} seja harmônica em $\tilde{V} \setminus \{a, z\}$. Para isso, observe que como W é harmônica em d e em y , devemos ter

$$\begin{aligned} W(y) &= P(y, w)W(w) \\ W(d) &= P(d, v)W(v) + P(d, b)W(b) \end{aligned}$$

Na nova rede, como as arestas se mantêm, $\tilde{c}(d) = c(d)$, e devemos ter $\tilde{P}(d, \delta) = P(d, v)$ e $\tilde{P}(d, b) = P(d, b)$, e o mesmo raciocínio se aplica ao vértice y , de modo que \tilde{W} vai ser harmônica em a e em y se satisfizer

$$\begin{aligned} \tilde{W}(y) &= \tilde{P}(y, \delta)\tilde{W}(\delta) = P(y, w)\tilde{W}(\delta) \\ \tilde{W}(d) &= \tilde{P}(d, \delta)\tilde{W}(\delta) + \tilde{P}(d, b)\tilde{W}(b) = P(d, v)\tilde{W}(\delta) + P(d, b)W(b). \end{aligned}$$

Como $\tilde{W}(y) = W(y)$ e $\tilde{W}(a) = W(a)$, a combinação das equações anteriores diz que devemos ter $\tilde{W}(\delta) = W(w)$, e $\tilde{W}(\delta) = W(v)$, mas isto é impossível uma vez que $W(w) \neq W(v)$. Assim tal \tilde{W} não pode ser harmônica em $\tilde{V} \setminus \{a, z\}$.

*Porém é importante destacar que na rede gerada pela colagem vai existir um **outra** função \tilde{W} harmônica, que obrigatoriamente será diferente da W original.*

Demonstração. Seja $G = (V, E)$ o grafo da rede original e $\tilde{G} = (\tilde{V}, \tilde{E})$ o grafo da rede criada pela colagem. Vamos supor por simplicidade que o G não possui loops (o caso geral apenas adiciona alguns termos às equações, mas a abordagem é a mesma). Retiramos os vértices v e w , e adicionamos δ , de modo que $\tilde{V} = (V \setminus \{v, w\}) \cup \{\delta\}$. Já em relação às arestas, todas foram mantidas, porém as que tinham v ou w como extremos passaram a ter o vértice δ no lugar, mais precisamente, se $v\vec{y} \in E$, então na nova rede ela foi substituída pela aresta $\delta\vec{y} \in \tilde{E}$, com a mesma condutância.

Em relação à condutância total em um vértice $x \in \tilde{V}$, $x \neq \delta$, as arestas que partem dele não mudaram (podem ter trocado o ponto final, que passou de v ou w para δ , mas não alteraram sua condutância), logo $\tilde{c}(x) = c(x)$. E para o vértice δ que foi obtido pela colagem de v com w , como ele herdou todas as arestas que partem de v e todas as que partem de w , então $\tilde{c}(\delta) = c(v) + c(w) - c(v, w)$, onde este último termo surge para compensar o fato de tanto em $c(v)$ quando em $c(w)$ estarmos contando a aresta $\{v, w\}$ (se ela existir), porém na nova rede ela aparece apenas uma vez em forma de loop.

Seja W uma voltagem entre a e z na rede original satisfazendo $W(v) = W(w)$. Definimos $\tilde{W} : \tilde{V} \rightarrow \mathbb{R}$ por $\tilde{W}(x) = W(x)$ para todo $x \in \tilde{V}$, $x \neq \delta$, e $\tilde{W}(\delta) = W(v) = W(w)$. Afirmamos que \tilde{W} é harmônica em $\tilde{V} \setminus \{a, z\}$.

De fato, para os vértices $x \in \tilde{V} \setminus \{a, z\}$ com $x \neq \delta$, todos os seus vizinhos $y \neq \delta$ nesta nova rede têm $\tilde{W}(y) = W(y)$ e tiveram a aresta mantida com a mesma condutância, então como $\tilde{c}(x) = c(x)$, vale que $\tilde{P}(x, y) = P(x, y)$. Já para o vizinho $y = \delta$, $\tilde{W}(\delta) = W(v) = W(w)$, mas $\tilde{P}(x, \delta) = \frac{c(x, v) + c(x, w)}{c(x)} = P(x, v) + P(x, w)$ uma vez que agora existem as arestas correspondentes a $\{x, v\}$ e a $\{x, w\}$ ligando estes dois pontos (tecnicamente, é um multigrafo agora, mas não precisamos nos preocupar muito com isso). Deste modo, temos que

$$\begin{aligned} \sum_{y \in \tilde{V}} \tilde{P}(x, y) \tilde{W}(y) &= \sum_{y \in \tilde{V}, y \neq \delta} P(x, y) W(y) + \tilde{P}(x, \delta) \tilde{W}(\delta) \\ &= \sum_{y \in V, y \neq v, w} P(x, y) W(y) + [P(x, v) + P(x, w)] W(v) \\ &= \sum_{y \in V, y \neq v, w} P(x, y) W(y) + P(x, v) W(v) + P(x, w) W(w) \\ &= \sum_{y \in V} P(x, y) W(y), \end{aligned}$$

e como W é harmônica em x na rede original, este último somatório é igual a $W(x) = \tilde{W}(x)$, mostrando que \tilde{W} é harmônica em x na nova rede.

Para o vértice $x = \delta$, como estamos supondo que o grafo original não possui loops, $\tilde{c}(\delta, \delta) = c(v, w)$. Além disso, como todas as arestas foram mantidas, δ possui as arestas de v e as de w , de modo que $\tilde{c}(\delta, y) = c(v, y) + c(w, y)$ para $y \neq \delta$. Deste modo, $\tilde{P}(\delta, \delta) = \frac{c(v, w)}{\tilde{c}(\delta)}$ e $\tilde{P}(\delta, y) = \frac{c(v, y) + c(w, y)}{\tilde{c}(\delta)}$ para todo $y \neq \delta$, e consequentemente

$$\begin{aligned} \sum_{y \in \tilde{V}} \tilde{P}(\delta, y) \tilde{W}(y) &= \sum_{y \in \tilde{V}, y \neq \delta} \left[\frac{c(v, y) + c(w, y)}{\tilde{c}(\delta)} \right] W(y) + \frac{c(v, w)}{\tilde{c}(\delta)} \tilde{W}(\delta) \\ &= \frac{1}{\tilde{c}(\delta)} \left[\sum_{y \in \tilde{V} \setminus \{\delta\}} c(v, y) W(y) + \sum_{y \in \tilde{V} \setminus \{\delta\}} c(w, y) W(y) + c(v, w) \tilde{W}(\delta) \right] \\ &= \frac{1}{\tilde{c}(\delta)} \left[c(v) \sum_{y \in V \setminus \{v, w\}} P(v, y) W(y) + c(w) \sum_{y \in V \setminus \{v, w\}} P(w, y) W(y) + c(v, w) \tilde{W}(\delta) \right] \\ &= \frac{1}{\tilde{c}(\delta)} [c(v)[W(v) - P(v, v)W(v) - P(v, w)W(w)] \\ &\quad + c(w)[W(w) - P(w, v)W(v) - P(w, w)W(w)] + c(v, w) \tilde{W}(\delta), \end{aligned}$$

onde a última igualdade é consequência da harmonicidade de W em v e em w . Então, utilizando que $\tilde{W}(\delta) = W(v) = W(w)$ e que $P(v, v) = P(w, w) = 0$ já que G não possui loops, obtemos

que

$$\begin{aligned}
 & \sum_{y \in \tilde{V}} \tilde{P}(\delta, y) \tilde{W}(y) \\
 &= \frac{1}{\tilde{c}(\delta)} [c(v) \tilde{W}(\delta) - c(v) P(v, w) \tilde{W}(\delta) + c(w) \tilde{W}(\delta) - c(w) P(w, v) \tilde{W}(\delta) + c(v, w) \tilde{W}(\delta)] \\
 &= \frac{\tilde{W}(\delta)}{\tilde{c}(\delta)} [c(v) - c(v, w) + c(w) - c(w, v) + c(v, w)] \\
 &= \tilde{W}(\delta),
 \end{aligned}$$

uma vez que $\tilde{c}(\delta) = c(v) + c(w) - c(v, w)$. Isso conclui que \tilde{W} é harmônica em $\tilde{V} \setminus \{a, z\}$, sendo então uma voltagem entre a e z .

Esta voltagem \tilde{W} define uma corrente \tilde{I} na rede $(\tilde{G}, \{\tilde{c}(e)\})$ por

$$\tilde{I}(\vec{f}) := \tilde{c}(\vec{f}) [\tilde{W}(x) - \tilde{W}(y)]$$

para todos os vértices $x, y \in \tilde{V}$ e toda aresta orientada $\vec{f} \in E^+(xy)$.

Como todas as arestas foram mantidas com a mesma condutância (no máximo trocamos os pontos v e w por δ), e a voltagem continua a mesma em todos os vértices (inclusive $\tilde{W}(\delta) = W(v) = W(w)$), em cada aresta a corrente \tilde{I} possui o mesmo valor que a corrente I tinha na respectiva aresta da rede original. Deste modo, $\|\tilde{I}\| = \text{div } \tilde{I}(a) = \text{div } I(a) = \|I\|$, e então $\tilde{\mathcal{R}}(a \leftrightarrow z) = \mathcal{R}(a \leftrightarrow z)$. □

3.6 Energia

Utilizaremos esta seção para apresentar o conceito de energia de um fluxo e utilizá-lo como ferramenta para o estudo da resistência efetiva. Tal conceito é bem amplo e possui várias outras aplicações que podem ser consultadas em [11] e em [8].

Relembremos que um fluxo θ foi definido como uma função antissimétrica nas arestas orientadas. Deste modo $\theta(x\vec{y})^2 = \theta(y\vec{x})^2$ e fica bem definido $\theta(e)^2$ para toda aresta e não orientada. Partindo disso definimos a **energia** de um fluxo θ por

$$\mathcal{E}(\theta) := \sum_{e \in E} [\theta(e)]^2 r(e). \quad (3.31)$$

Quando I é o fluxo de corrente unitário entre a e z associado à voltagem W ,

$$\mathcal{E}(I) = \mathcal{R}(a \leftrightarrow z). \quad (3.32)$$

Para verificar isso observamos que da definição de fluxo de corrente e de energia,

$$\begin{aligned}
 \mathcal{E}(I) &= \sum_{e \in E} [I(e)]^2 r(e) = \frac{1}{2} \sum_x \sum_y r(x, y) \left[\frac{W(x) - W(y)}{r(x, y)} \right]^2 \\
 &= \frac{1}{2} \sum_x \sum_y c(x, y) [W(x) - W(y)]^2 \\
 &= \frac{1}{2} \sum_x \sum_y [W(x) - W(y)] I(\vec{xy}) \\
 &= \frac{1}{2} \left[\sum_x \sum_y W(x) I(\vec{xy}) - \sum_x \sum_y W(y) I(\vec{xy}) \right] \\
 &= \frac{1}{2} \left[\sum_x \sum_y W(x) I(\vec{xy}) + \sum_y \sum_x W(y) I(\vec{yx}) \right] \\
 &= \sum_x W(x) \sum_y I(\vec{xy}).
 \end{aligned}$$

Ao mesmo tempo a lei de Kirchhoff garante que $0 = \operatorname{div} I(x) = \sum_y I(\vec{xy})$ para todo $x \neq a, z$.

E como I é um fluxo entre a e z , também temos que $\sum_y I(\vec{ay}) = \operatorname{div} I(a) = \|I\|$ e que

$\sum_y I(\vec{zy}) = \operatorname{div} I(z) = -\|I\|$, de modo que a igualdade anterior pode ser reescrita como

$$\begin{aligned}
 \mathcal{E}(I) &= \sum_{x \neq a, z} W(x) \sum_y I(\vec{xy}) + W(a) \sum_y I(\vec{ay}) + W(z) \sum_y I(\vec{zy}) \\
 &= \|I\| [W(a) - W(z)].
 \end{aligned}$$

Finalmente, como estamos supondo $\|I\| = 1$, o resultado segue da definição de resistência efetiva em 3.14.

O Princípio de Thomson nos diz que a energia de qualquer fluxo unitário entre a e z é uma cota superior para a resistência efetiva entre estes dois vértices.

Teorema 3.23 (Princípio de Thomson). *Para um grafo conexo e finito,*

$$\mathcal{R}(a \leftrightarrow z) = \inf \{ \mathcal{E}(\theta); \theta \text{ é um fluxo unitário entre } a \text{ e } z \}. \quad (3.33)$$

Além disso, conforme vimos anteriormente, tal ínfimo é atingido quando $\theta = I$ é o fluxo de corrente unitário.

Demonstração. Já mostramos em (3.32) que $\mathcal{E}(I) = \mathcal{R}(a \leftrightarrow z)$. Agora se θ é um fluxo unitário qualquer entre a e z , ele pode ser escrito como $\theta = I + \delta$, onde δ é um fluxo com $\operatorname{div} \delta(a) =$

$0 = \operatorname{div} \delta(z)$ e que satisfaz $\operatorname{div} \delta(x) = 0$ para todo $x \notin \{a, z\}$. Logo

$$\begin{aligned}
 \mathcal{E}(\theta) &= \sum_{e \in E} r(e)[I(e) + \delta(e)]^2 \\
 &= \mathcal{E}(I) + \mathcal{E}(\delta) + \frac{1}{2} \sum_{e \in E} 2r(e)i(e)\delta(e) \\
 &= \mathcal{E}(I) + \mathcal{E}(\delta) + \sum_{(x,y) \in E} [W(x) - W(y)]\delta(x,y) \\
 &= \mathcal{E}(I) + \mathcal{E}(\delta) + 2 \sum_{(x,y) \in E} W(x)\delta(x,y) \\
 &= \mathcal{E}(I) + \mathcal{E}(\delta) + 2 \sum_x W(x) \operatorname{div} \delta(x) \\
 &= \mathcal{E}(I) + \mathcal{E}(\delta) \\
 &\geq \mathcal{E}(I).
 \end{aligned}$$

A argumentação acima mostra inclusive que se $\theta \neq I$ (neste caso $\delta \neq 0$ e $\mathcal{E}(\delta) > 0$) vale a desigualdade estrita, mostrando que I é o único fluxo unitário que atinge o ínfimo em (3.33). \square

Se $r = \{r(e)\}_{e \in E}$ é o conjunto dos valores das resistências associadas às arestas de um grafo G , escrevemos $\mathcal{R}(a \leftrightarrow z, r)$ para denotar a resistência efetiva entre a e z neste grafo. Se alterarmos as resistências das arestas de forma controlada, então podemos controlar também a resistência efetiva, conforme garante o seguinte teorema.

Teorema 3.24 (Lei da Monotonicidade de Rayleigh). *Sejam $r = \{r(e)\}$ e $r' = \{r'(e)\}$ dois conjuntos de resistências para as arestas de um mesmo grafo $G = (V, E)$, satisfazendo $r(e) \leq r'(e)$ para todo $e \in E$. Então,*

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

Demonstração. Note que para todo θ fluxo unitário entre a e z fixo, vale

$$\sum_e r(e)\theta(e)^2 \leq \sum_e r'(e)\theta(e)^2,$$

portanto $\inf_{\theta} \sum_e r(e)\theta(e)^2 \leq \inf_{\theta} \sum_e r'(e)\theta(e)^2$, donde segue pelo Princípio de Thomson que $\mathcal{R}(a \leftrightarrow z, r) \leq \mathcal{R}(a \leftrightarrow z, r')$. \square

Corolário 3.25. *Adicionar uma aresta não aumenta a resistência efetiva $\mathcal{R}(a \leftrightarrow z)$. Além disso, se a aresta adicionada não é incidente ao vértice a , isso não diminui a probabilidade de escape $\mathbb{P}_a(\tau_z < \tau_a^+) = \frac{1}{c(a)\mathcal{R}(a \leftrightarrow z)}$.*

Demonstração. Adicionar uma aresta e_0 à rede é equivalente a considerar que tal aresta já existia com condutância $c(e_0) = 0$ e alterar sua condutância para um valor $c'(e_0) > 0$. Em relação à resistência, como esta é a única aresta alterada, o que está sendo feito é substituir o conjunto $\{r(e)\}$ pelo conjunto $\{r'(e)\}$, onde

$$\begin{aligned}
 r'(e_0) &< r(e_0), \\
 r'(e) &= r(e) \quad \forall e \neq e_0.
 \end{aligned}$$

Em particular $r'(e) \leq r(e)$ para toda aresta $e \in E$ e o Teorema 3.24 garante que $\mathcal{R}(a \leftrightarrow z, r) \geq \mathcal{R}(a \leftrightarrow z, r')$, mostrando que a resistência efetiva não aumentou ao se adicionar a aresta e_0 .

Para a segunda afirmação, note que pela identidade (3.16),

$$\mathbb{P}_a(\tau_z < \tau_a^+) = \frac{1}{c(a)\mathcal{R}(a \leftrightarrow z)}.$$

Se a aresta adicionada não é incidente ao vértice a , então $c(a)$ segue inalterada enquanto que $\mathcal{R}(a \leftrightarrow z)$ não aumenta, de modo que a probabilidade de escape não pode diminuir. \square

Corolário 3.26. *A operação de colagem de vértices não aumenta a resistência efetiva*

Demonstração. Fixada uma rede $(G, \{c(e)\})$ e dois vértices $v, w \in V$, realizamos a colagem destes obtendo o vértice δ e uma nova rede $(\tilde{G}, \{\tilde{c}(e)\})$. Mostraremos que para todo θ fluxo unitário entre a e z em $(G, \{c(e)\})$ existe um fluxo $\tilde{\theta}$ de a para z unitário em $(\tilde{G}, \{\tilde{c}(e)\})$ tal que $\mathcal{E}(\theta) = \mathcal{E}(\tilde{\theta})$.

Seja $(G, \{c(e)\})$ uma rede, e seja θ um fluxo unitário de a para z , isto é $\|\theta\| = \text{div } \theta(a) = 1$ e $\text{div } \theta(x) = 0$ para todo $x \neq a, z$.

O processo de colagem não cria nem destrói arestas, então seja $\tilde{\theta}$ o fluxo naturalmente associado (cada aresta mantém o mesmo fluxo que a original, só muda a maneira como chama algumas arestas). Afirmamos que $\tilde{\theta}$ é um fluxo unitário de a para z .

Para isso, note que se $x \neq \delta$, nenhuma de suas arestas incidentes foi alterada, e o divergente segue inalterado, ou seja, $\text{div } \tilde{\theta}(x) = \text{div } \theta(x)$. Já se $x = \delta$, como as arestas não foram destruídas,

$$\begin{aligned} \text{div } \tilde{\theta}(\delta) &= \sum_{y \sim \delta} \tilde{\theta}(\delta y) \\ &= \sum_{y \sim v, y \neq w} \theta(vy) + \sum_{y \sim w, y \neq v} \theta(wy) \\ &= \sum_{y \sim v} \theta(vy) + \sum_{y \sim w} \theta(wy) - \theta(vw) - \theta(wv) \\ &= \sum_{y \sim v} \theta(vy) + \sum_{y \sim w} \theta(wy) \\ &= \text{div } \theta(v) + \text{div } \theta(w). \end{aligned}$$

Se a colagem envolveu o vértice a (análogo quando ela envolve o vértice z), este foi colado com um vértice w com $\text{div } \theta(w) = 0$, e portanto $\|\tilde{\theta}\| = \text{div } \tilde{\theta}(a) = \text{div } \theta(a) + \text{div } \theta(w) = \text{div } \theta(a) = \|\theta\| = 1$, de modo que o fluxo é unitário. Além disso, para todo $x \neq a, z$, como não foi alterada nenhuma das duas arestas, vale $\text{div } \tilde{\theta}(x) = \text{div } \theta(x) = 0$, mostrando que é um fluxo de a para z .

Já se a colagem não envolveu os vértices a ou z , então $\|\tilde{\theta}\| = \text{div } \tilde{\theta}(a) = \text{div } \theta(a) = \|\theta\| = 1$, mostrando que ele é unitário. E além disso, vale $\text{div } \tilde{\theta}(\delta) = \text{div } \theta(v) + \text{div } \theta(w) = 0$ e $\text{div } \tilde{\theta}(x) = \text{div } \theta(x) = 0$ para todos os demais vértices que não sejam a, z e vértice colado, de modo que $\tilde{\theta}$ é um fluxo de a para z .

Como nenhuma arestas foi criada ou destruída (lembre que a aresta que ligava v com w foi transformada em um loop, com mesma condutância da aresta original), e mostramos acima que $\tilde{\theta} = \theta$, então $\mathcal{E}(\tilde{\theta}) = \mathcal{E}(\theta)$. Isso mostra que todo fluxo unitário de a para z na rede original também é fluxo unitário de a para z na rede depois da colagem, com a mesma energia, mostrando que o ínfimo do Princípio de Thomson não pode aumentar com a operação de colagem. \square

3.7 Conjunto separador entre a e z

Nesta seção apresentamos um método extremamente útil na obtenção de cotas inferiores para a resistência efetiva. Começamos dizendo que $\Pi \subset E$ é um **conjunto separador de a e z** se todo caminho de a para z contém obrigatoriamente uma aresta de Π .

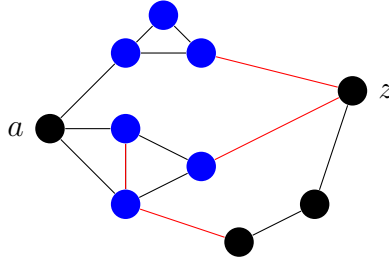


Figura 3.4: Em vermelho Π , um conjunto separador de a e z . Não é possível transitar entre a e z sem passar por Π . Os vértices em azul são aqueles em que é possível chegar a partir de a sem passar por Π .

A quantidade de fluxo que passa por um conjunto separador entre dois vértices a e z limita a força do fluxo, conforme garante o seguinte resultado:

Lema 3.27. *Se θ é um fluxo de a para z , e Π é um conjunto separando a de z , então*

$$\|\theta\| \leq \sum_{e \in \Pi} |\theta(e)|$$

Observação 3.28. *O conjunto de todas as arestas que partem de a , e o conjunto de todas as arestas que chegam a z são dois conjuntos separadores naturais. Note que no primeiro caso, não necessariamente vale a igualdade $\sum_{y \sim a} \theta(\vec{a}y) = \|\theta\| = \sum_{e \in \Pi} |\theta(e)|$, pois embora as arestas sejam as mesmas, pode acontecer de $\theta(\vec{a}y) < 0$, tornando a soma da esquerda menor que a da direita, mesmo ambas sendo sobre as mesmas arestas.*

Demonstração. Considere E o conjunto dos vértices x tais que possamos ir da a até x sem passar por Π (veja a Figura 3.4). Pela construção, os pontos de E são separados dos pontos fora de E por uma aresta $e \in \Pi$, o que garante que

$$\sum_{x \in E} \operatorname{div} \theta(x) = \sum_{x \in E} \sum_y \theta(\vec{x}y) = \sum_{x \in E} \sum_{y \notin E} \theta(\vec{x}y) + \sum_{x \in E} \sum_{y \in E} \theta(\vec{x}y) = \sum_{x \in E} \sum_{y \notin E} \theta(\vec{x}y) \leq \sum_{e \in \Pi} |\theta(e)|.$$

Na equação acima, no segundo somatório antes da terceira igualdade, cada aresta aparece duas vezes, uma em cada sentido, então pela anti-simetria de θ o somatório é nulo. Além disso, a última desigualdade surge não apenas pela introdução do módulo, mas também pelo fato de que podem existir arestas em Π que ligam dois vértices de E ou de E^C . Em particular, isso nos diz que essa cota pode ser melhorada com a escolha de um Π “melhor”.

Por outro lado note que $z \notin E$, e como θ é um fluxo de a para z a Lei de Kirchhoff garante que $\operatorname{div} \theta(x) = 0$ para todo $x \in E$ com $x \neq a$, portanto,

$$\sum_{x \in E} \operatorname{div} \theta(x) = \operatorname{div} \theta(a) + \sum_{x \in E, x \neq a} \operatorname{div} \theta(x) = \operatorname{div} \theta(a) = \|\theta\|.$$

Juntando as duas equações anteriores, segue o resultado desejado. \square

Proposição 3.29 (Desigualdade de Nash-Williams). *Se $\{\Pi_k\}_k$ é uma coleção de conjuntos separadores de a e z disjuntos, então*

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

Demonstração. Fixado θ um fluxo unitário de a para z . Para cada k , a desigualdade de Cauchy-Schwarz e o lema anterior garantem que

$$\sum_{e \in \Pi_k} c(e) \sum_{e \in \Pi_k} r(e) \theta(e)^2 \geq \left(\sum_{e \in \Pi_k} \sqrt{c(e)} \sqrt{r(e)} |\theta(e)| \right)^2 = \left(\sum_{e \in \Pi_k} |\theta(e)| \right)^2 \geq \|\theta\|^2 = 1,$$

donde segue que

$$\sum_{e \in \Pi_k} r(e) \theta(e)^2 \geq \left(\sum_{e \in \Pi_k} c(e) \right)^{-1}.$$

Como $\{\Pi_k\}$ são disjuntos e $\cup_k \Pi_k \subset E$, vale

$$\mathcal{E}(\theta) = \sum_{e \in E} r(e) \theta(e)^2 \geq \sum_k \sum_{e \in \Pi_k} r(e) \theta(e)^2 \geq \sum_k \left(\sum_{e \in \Pi_k} c(e) \right)^{-1}.$$

Como a desigualdade acima vale para todo fluxo unitário de a para z , o princípio de Thomson garante que

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

□

Capítulo 4

Outros Indicadores das Cadeias de Markov

Neste capítulo serão apresentados novos indicadores das cadeias de Markov: o tempo do alvo aleatório (t_{\odot}), o tempo de acerto da cadeia (t_{hit}), o tempo de comutação (t_{com}) e o tempo de relaxamento (t_{rel}). Tais indicadores, juntamente com o tempo de mistura (t_{mix} já definido) fornecem valiosas informações a respeito do processo, e estão relacionados entre si pelos resultados que apresentaremos no próximo capítulo.

Iniciamos este capítulo introduzindo uma notação que nos será bastante útil. Lembremos que na Definição 1.15 introduzimos o tempo de acerto do estado $a \in S$, $\tau_a = \min\{t \geq 0 : X_t = a\}$. Se π é uma distribuição em S , pela lei da esperança total vale

$$\mathbb{E}_{\pi}[\tau_a] = \sum_{x \in S} \mathbb{E}_{\pi}[\tau_a | X_0 = x] \mathbb{P}_{\pi}(X_0 = x) = \sum_{x \in S} \mathbb{E}_x[\tau_a] \pi(x).$$

O lado direito da igualdade acima representa uma média dos valores de $\mathbb{E}_x[\tau_a]$ com x sendo ponderado pela distribuição π , ou em outras palavras, representa o valor esperado para $\mathbb{E}_x[\tau_a]$ quando o argumento x é escolhido aleatoriamente a partir da distribuição π . Então, com um abuso de notação escreveremos

$$\mathbb{E}_a[\tau_{\pi}] := \sum_{y \in S} \mathbb{E}_a[\tau_y] \pi(y),$$

que interpretamos como o valor esperado para $\mathbb{E}_a[\tau_y]$ quando y for escolhido a partir da distribuição π . Ainda nesta linha, fica também definido

$$\mathbb{E}_{\pi}[\tau_{\pi}] := \sum_{x \in S} \sum_{y \in S} \mathbb{E}_x[\tau_y] \pi(x) \pi(y).$$

4.1 Tempo do alvo aleatório

Definição 4.1. *Para uma cadeia de Markov $(X_t)_{t \geq 0}$ com distribuição estacionária π , definimos*

$$t_{\odot}^a := \sum_{x \in S} \mathbb{E}_a[\tau_x] \pi(x) = \mathbb{E}_a[\tau_{\pi}].$$

Interpretamos t_{\odot}^a como o tempo esperado para a cadeia, começando em a , atingir um alvo escolhido aleatoriamente a partir de π . Se a cadeia for irredutível, o Lema 1.16 garante que $\mathbb{E}_a[\tau_x] < \infty$ para todo x , e conseqüentemente que t_{\odot}^a é finito.

Lema 4.2 (Lema do alvo aleatório). *Seja $(X_t)_{t \geq 0}$ uma cadeia de Markov irredutível em um espaço de estados finito S , com distribuição estacionária π . Então t_{\odot}^a não depende de $a \in S$.*

Demonstração. Definimos a função $h : S \rightarrow \mathbb{R}$ por $h(a) := t_{\circlearrowleft}^a = \sum_{x \in S} \mathbb{E}_a[\tau_x] \pi(x)$. Graças ao Lema 3.2 é suficiente mostrar que h é harmônica em S . Visando isso, definimos as funções auxiliares $h_x(a) := \mathbb{E}_a[\tau_x]$, de modo que $h(a) = \sum_{x \in S} h_x(a) \pi(x)$.

Começamos observando que se $x \neq a$, quando condicionamos em $X_1 = y$ já temos dado um passo e a perda de memória nos permite reiniciar a contagem do tempo necessário para chegar em x , de modo que vale

$$\mathbb{E}_a[\tau_x | X_1 = y] = 1 + \mathbb{E}_y[\tau_x].$$

Assim sendo, condicionando no primeiro passo, a lei da esperança total nos garante que, para $x \neq a$,

$$\begin{aligned} h_x(a) &= \mathbb{E}_a[\tau_x] = \sum_{y \in S} \mathbb{E}_a[\tau_x | X_1 = y] \mathbb{P}_a(X_1 = y) \\ &= \sum_{y \in S} [1 + \mathbb{E}_y[\tau_x]] P(a, y) \\ &= \sum_{y \in S} P(a, y) + \sum_{y \in S} h_x(y) P(a, y) \\ &= 1 + (Ph_x)(a). \end{aligned}$$

Para $x = a$, a mesma argumentação de antes mostra que

$$\mathbb{E}_a[\tau_a^+] = \sum_{y \in S} \mathbb{E}_a[\tau_a^+ | X_1 = y] P(a, y) = \sum_{y \in S} (1 + \mathbb{E}_y[\tau_a]) P(a, y) = 1 + (Ph_a)(a).$$

Como $\mathbb{E}_a[\tau_a^+] = \pi(a)^{-1}$ de acordo com o Teorema 1.25, vale

$$(Ph_a)(a) = \mathbb{E}_a[\tau_a^+] - 1 = \frac{1}{\pi(a)} - 1.$$

Em resumo, as contas acima nos mostram que

$$(Ph_x)(a) = \begin{cases} h_x(a) - 1 & \text{se } x \neq a \\ \frac{1}{\pi(a)} - 1 & \text{se } x = a \end{cases}, \quad (4.1)$$

donde segue que

$$\begin{aligned} (Ph)(a) &= \sum_{x \in S} (Ph_x)(a) \pi(x) \\ &= \sum_{x \neq a} [h_x(a) - 1] \pi(x) + \left(\frac{1}{\pi(a)} - 1 \right) \pi(a) \\ &= \sum_{x \neq a} h_x(a) \pi(x) - \sum_{x \neq a} \pi(x) + (1 - \pi(a)) \\ &= \sum_{x \neq a} h_x(a) \pi(x) - (1 - \pi(a)) + (1 - \pi(a)) \\ &= \sum_{x \neq a} h_x(a) \pi(x) \\ &= h(a), \end{aligned}$$

onde, na última igualdade, usamos que $h_a(a) = 0$.

As contas acima valem para $a \in S$ qualquer, mostrando que h é harmônica em S e portanto constante de acordo com o Lema 3.2, o que conclui que $t_{\circlearrowleft}^a = h(a)$ não depende de a . \square

Graças ao resultado acima, fica bem definido o **tempo do alvo aleatório** da cadeia $(X_t)_{t \geq 0}$ por

$$t_{\odot} := t_{\odot}^a,$$

para qualquer $a \in S$. Como $\sum_{a \in S} \pi(a) = 1$, podemos escrever

$$t_{\odot} = \sum_{a \in S} t_{\odot} \pi(a) = \sum_{a \in S} t_{\odot}^a \pi(a) = \sum_{a \in S} \left(\sum_{y \in S} \mathbb{E}_a[\tau_y] \pi(y) \right) \pi(a) = \mathbb{E}_{\pi}[\tau_{\pi}], \quad (4.2)$$

permitindo interpretar t_{\odot} como o tempo esperado para viajar entre dois estados x e y sorteados aleatoriamente e independentemente a partir da distribuição π .

Na Seção 5.3 vamos obter um fórmula para t_{\odot} em termos dos autovalores de P .

4.2 Tempo de acerto

Novamente para uma cadeia de Markov $(X_t)_{t \geq 0}$ com distribuição estacionária π , definimos

$$t_{\text{hit}} := \max_{x, y \in S} \mathbb{E}_x[\tau_y]$$

como sendo o pior cenário, isto é, o maior tempo esperado para viajar ente dois estados da cadeia. Este indicador é chamado de **tempo de acerto** da cadeia, e é denotado por t_{hit} visto que esta é a notação difundida, proveniente do inglês *hitting time*.

É claro a partir da definição que $t_{\odot} \leq t_{\text{hit}}$. Também é possível construir uma cota no sentido contrário.

Lema 4.3. *Seja $(X_t)_{t \geq 0}$ uma cadeia de Markov irredutível no espaço de estados S com distribuição estacionária π . Então*

$$t_{\text{hit}} \leq 2 \max_{w \in S} \mathbb{E}_{\pi}[\tau_w].$$

Demonstração. Fixados os estados $a, y \in S$, ao forçar a cadeia a passar em um estado $x \in S$ qualquer, aumentamos o tempo deste percurso, de modo que vale

$$\mathbb{E}_a[\tau_y] \leq \mathbb{E}_a[\tau_x] + \mathbb{E}_x[\tau_y] \quad (4.3)$$

para todo $x \in S$. Assim

$$\mathbb{E}_a[\tau_y] = \sum_{x \in S} \mathbb{E}_a[\tau_y] \pi(x) \leq \sum_{x \in S} \mathbb{E}_a[\tau_x] \pi(x) + \sum_{x \in S} \mathbb{E}_x[\tau_y] \pi(x) = \mathbb{E}_a[\tau_{\pi}] + \mathbb{E}_{\pi}[\tau_y]. \quad (4.4)$$

Mas note que o Lema 4.2 garante que

$$\mathbb{E}_a[\tau_{\pi}] = t_{\odot}^a = t_{\odot} = \mathbb{E}_{\pi}[\tau_{\pi}] = \sum_{x \in S} \sum_{y \in S} \pi(x) \pi(y) \mathbb{E}_x[\tau_y].$$

Desenvolvendo o lado direito, obtemos que

$$\begin{aligned} \mathbb{E}_a[\tau_{\pi}] &= \sum_{y \in S} \pi(y) \sum_{x \in S} \pi(x) \mathbb{E}_x[\tau_y] \\ &= \sum_{y \in S} \pi(y) \mathbb{E}_{\pi}[\tau_y] \\ &\leq \max_{w \in S} \mathbb{E}_{\pi}[\tau_w] \sum_{y \in S} \pi(y) \\ &= \max_{w \in S} \mathbb{E}_{\pi}[\tau_w]. \end{aligned} \quad (4.5)$$

Além disso, é claro que

$$\mathbb{E}_\pi[\tau_y] \leq \max_{w \in S} \mathbb{E}_\pi[\tau_w]. \quad (4.6)$$

As equações (4.5) e (4.6) mostram que os dois termos do lado direito de (4.4) são limitados por $\max_w \mathbb{E}_\pi[\tau_w]$, garantindo assim que $\mathbb{E}_a[\tau_y] \leq 2 \max_{w \in S} \mathbb{E}_\pi[\tau_w]$. E como isso vale para $a, y \in S$ quaisquer, concluímos que

$$t_{\text{hit}} = \max_{a, y \in S} \mathbb{E}_a[\tau_y] \leq 2 \max_{w \in S} \mathbb{E}_\pi[\tau_w].$$

□

4.3 Tempo de comutação

O **tempo de comutação** entre dois estados a e b de uma cadeia de Markov indica o tempo esperado para ir de a até b e então voltar para a . Formalmente, fixados estes dois estados a e b , definimos a variável aleatória

$$\tau_{a,b} := \min\{t \geq \tau_b; X_t = a\}$$

no espaço das trajetórias, e definimos o tempo de comutação por

$$t_{a \leftrightarrow b} := \mathbb{E}_a[\tau_{a,b}].$$

Também definimos o **máximo tempo de comutação** como sendo

$$t_{\text{com}} := \max_{a,b \in S} t_{a \leftrightarrow b}.$$

Observe que condicionando em $X_0 = a$, a definição de $\tau_{a,b}$ nos permite reescrevê-lo como $\tau_b + \tau_a \circ \theta_{\tau_b}$. Ao mesmo tempo, como τ_b é um tempo de parada, a propriedade forte de Markov garante que

$$\mathbb{E}_a[\tau_a \circ \theta_{\tau_b}] = \mathbb{E}_a[\mathbb{E}_a[\tau_a \circ \theta_{\tau_b} | \mathcal{F}_{\tau_b}]] = \mathbb{E}_a[\mathbb{E}_{X_{\tau_b}}[\tau_a]] = \mathbb{E}_a[\mathbb{E}_b[\tau_a]] = \mathbb{E}_b[\tau_a].$$

Deste modo temos que

$$t_{a \leftrightarrow b} = \mathbb{E}_a[\tau_{a,b}] = \mathbb{E}_a[\tau_b + \tau_a \circ \theta_{\tau_b}] = \mathbb{E}_a[\tau_b] + \mathbb{E}_a[\tau_a \circ \theta_{\tau_b}] = \mathbb{E}_a[\tau_b] + \mathbb{E}_b[\tau_a]. \quad (4.7)$$

Tal igualdade é natural visto que ambos os lados da igualdade estão medindo o tempo necessário para saindo de a chegar em b e então retornar ao estado a .

O próximo resultado é uma generalização do Teorema 1.21.

Lema 4.4. *Seja $(X_t)_{t \geq 0}$ uma cadeia de Markov com matriz de transição P em um espaço de estados S finito ou enumerável. Sejam μ e ν duas distribuições de probabilidade em S , e τ um tempo de parada satisfazendo $\mathbb{P}_\mu(\tau < \infty) = 1$ e $\mathbb{P}_\mu(X_\tau = \cdot) = \nu$. Considere*

$$\rho(x) := \mathbb{E}_\mu \left(\sum_{t=0}^{\tau-1} \mathbf{1}_{\{X_t=x\}} \right). \quad (4.8)$$

Então $\rho P = \rho - \mu + \nu$. Em particular, se $\mu = \nu$ vale $\rho P = \rho$. Ainda, se $\mu = \nu$ e $\mathbb{E}_\mu(\tau) < \infty$, então $\frac{\rho}{\mathbb{E}_\mu(\tau)}$ é a distribuição estacionária para P .

Demonstração. Mostrar a igualdade $\rho P = \rho - \mu + \nu$ resume-se a algumas contas. Começamos observando que

$$\begin{aligned} (\rho P)(x) &= \sum_y P(y, x) \rho(y) \\ &= \sum_y P(y, x) \mathbb{E}_\mu \left(\sum_{t=0}^{\tau-1} \mathbf{1}_{\{X_t=y\}} \right) \\ &= \sum_y P(y, x) \mathbb{E}_\mu \left(\sum_{t=0}^{\infty} \mathbf{1}_{\{X_t=y\}} \mathbf{1}_{\{\tau>t\}} \right). \end{aligned}$$

A linearidade da esperança e a mudança de variáveis $s = t + 1$ nos dão

$$\begin{aligned} (\rho P)(x) &= \sum_y P(y, x) \sum_{s=1}^{\infty} \mathbb{E}_\mu \left(\mathbf{1}_{\{X_{s-1}=y\}} \mathbf{1}_{\{\tau>s-1\}} \right) \\ &= \sum_y \sum_{s=1}^{\infty} \mathbb{E}_\mu \left(\mathbf{1}_{\{X_{s-1}=y\}} \mathbf{1}_{\{\tau>s-1\}} P(X_{s-1}, x) \right), \end{aligned}$$

onde a última igualdade se dá pela presença da indicadora. Além disso, a propriedade fraca de Markov nos diz que

$$\mathbb{E}_\mu[\mathbf{1}_{\{X_1=x\}} \circ \theta_{s-1} | \mathcal{F}_{s-1}] = \mathbb{E}_{X_{s-1}}(\mathbf{1}_{\{X_1=x\}}) = P(X_{s-1}, x).$$

Utilizando tal resultado na equação anterior, juntamente com o fato de $\mathbf{1}_{\{X_{s-1}=y\}}$ e $\mathbf{1}_{\{\tau>s-1\}}$ serem \mathcal{F}_{s-1} -mensuráveis, obtemos que

$$\begin{aligned} (\rho P)(x) &= \sum_y \sum_{s=1}^{\infty} \mathbb{E}_\mu \left(\mathbb{E}_\mu \left((\mathbf{1}_{\{X_{s-1}=y\}} \mathbf{1}_{\{\tau>s-1\}} \cdot (\mathbf{1}_{\{X_1=x\}} \circ \theta_{s-1}) | \mathcal{F}_{s-1} \right) \right) \\ &= \sum_y \sum_{s=1}^{\infty} \mathbb{E}_\mu \left(\mathbb{E}_\mu \left((\mathbf{1}_{\{X_{s-1}=y\}} \mathbf{1}_{\{\tau>s-1\}} \mathbf{1}_{\{X_s=x\}} | \mathcal{F}_{s-1} \right) \right) \\ &= \sum_y \sum_{s=1}^{\infty} \mathbb{E}_\mu \left((\mathbf{1}_{\{X_{s-1}=y\}} \mathbf{1}_{\{\tau>s-1\}} \mathbf{1}_{\{X_s=x\}}) \right). \end{aligned}$$

Trocando a ordem dos somatórios, e utilizando o fato de $\sum_{y \in S} \mathbf{1}_{\{X_{s-1}=y\}} = 1$, ficamos com

$$\begin{aligned} (\rho P)(x) &= \sum_{s=1}^{\infty} \mathbb{E}_\mu \left(\mathbf{1}_{\{X_s=x\}} \mathbf{1}_{\{\tau>s-1\}} \sum_y \mathbf{1}_{\{X_{s-1}=y\}} \right) \\ &= \sum_{s=1}^{\infty} \mathbb{E}_\mu \left(\mathbf{1}_{\{X_s=x\}} \mathbf{1}_{\{\tau>s-1\}} \right) \\ &= \mathbb{E}_\mu \left(\sum_{s=1}^{\infty} \mathbf{1}_{\{X_s=x\}} \mathbf{1}_{\{\tau \geq s\}} \right) \\ &= \mathbb{E}_\mu \left(\sum_{s=1}^{\tau} \mathbf{1}_{\{X_s=x\}} \right). \end{aligned}$$

Por fim, reescrevendo a última igualdade, e utilizando a hipótese de que $\mathbb{P}_\mu(X_\tau = \cdot) = \nu$, obtemos que

$$\begin{aligned} (\rho P)(x) &= \mathbb{E}_\mu \left(\sum_{s=0}^{\tau-1} \mathbf{1}_{\{X_s=x\}} - \mathbf{1}_{\{X_0=x\}} + \mathbf{1}_{\{X_\tau=x\}} \right) \\ &= \mathbb{E}_\mu \left(\sum_{s=0}^{\tau-1} \mathbf{1}_{\{X_s=x\}} \right) - \mathbb{P}_\mu(X_0 = x) + \mathbb{P}_\mu(X_\tau = x) \\ &= \rho(x) - \mu(x) + \nu(x), \end{aligned}$$

mostrando assim a primeira afirmação do enunciado. Daqui é imediato que se $\mu = \nu$, então $\rho P = \rho$, ou seja, que ρ é medida invariante para P .

Para concluir, note que

$$\begin{aligned} \sum_{x \in S} \rho(x) &= \sum_{x \in S} \mathbb{E}_\mu \left(\sum_{t=0}^{\tau-1} \mathbf{1}_{\{X_t=x\}} \right) \\ &= \mathbb{E}_\mu \left(\sum_{t=0}^{\tau-1} \sum_{x \in S} \mathbf{1}_{\{X_t=x\}} \right) \\ &= \mathbb{E}_\mu \left(\sum_{t=0}^{\tau-1} 1 \right) \\ &= \mathbb{E}_\mu(\tau) \end{aligned}$$

Portanto, como ρ é medida invariante, sua normalização $\frac{\rho}{\sum_{x \in S} \rho(x)} = \frac{\rho}{\mathbb{E}_\mu(\tau)}$ (se $\mathbb{E}_\mu \tau < \infty$) é distribuição invariante para P . \square

Note que quando $\mu = \delta_a$, ρ definida em (4.8) é exatamente a função de Green $G_\tau(a, \cdot)$ definida em (3.17), uma vez que

$$\rho(x) = \mathbb{E}_a \left(\sum_{t=0}^{\tau-1} \mathbf{1}_{\{X_t=x\}} \right) = \mathbb{E}_a \left(\sum_{t=0}^{\infty} \mathbf{1}_{\{X_t=x, \tau > t\}} \right) = G_\tau(a, x). \quad (4.9)$$

Corolário 4.5. *Sejam $(X_t)_{t \geq 0}$ uma cadeia de Markov irredutível com distribuição invariante π e seja τ um tempo de parada satisfazendo $\mathbb{P}_a(X_\tau = a) = 1$, então para todo $x \in S$*

$$\frac{G_\tau(a, x)}{\mathbb{E}_a(\tau)} = \pi(x).$$

Demonstração. Tomando $\mu = \nu = \delta_a$, a hipótese $\mathbb{P}_a(X_\tau = a) = 1$ garante que $\mathbb{P}_\mu(X_\tau = \cdot) = \nu$ e que $\mathbb{E}_a(\tau) < \infty$. Assim sendo, em decorrência do Lema anterior e da identidade (4.9) o resultado segue. \square

Seja $(G, \{c(e)\})$ uma rede e $\mathcal{R}(a \leftrightarrow b)$ a resistência efetiva entre os vértices a e b , definida na Seção 3.4. Visto que c_G é uma constante que depende do grafo como um todo, o seguinte resultado indica que $t_{a \leftrightarrow b}$ é diretamente proporcional à resistência efetiva entre os nós a e b .

Proposição 4.6. *Seja $(G, \{c(e)\})$ uma rede e $(X_t)_{t \geq 0}$ o passeio aleatório associado. Então para todos os nós $a, b \in V$,*

$$t_{a \leftrightarrow b} = c_G \mathcal{R}(a \leftrightarrow b).$$

Demonstração. Sabemos da Proposição 3.10 que $\pi(x) = \frac{c(x)}{c_G}$ é distribuição estacionária para esta cadeia. Ao mesmo tempo $\tau_{a,b}$ satisfaz $\mathbb{P}_a(X_{\tau_{a,b}} = a) = 1$, de modo que o Corolário acima garante que

$$\pi(x) = \frac{G_{\tau_{a,b}}(a, x)}{\mathbb{E}_a(\tau_{a,b})}.$$

Juntando estas duas equações e aplicando em $x = a$, ficamos com

$$\frac{G_{\tau_{a,b}}(a, a)}{\mathbb{E}_a(\tau_{a,b})} = \frac{c(a)}{c_G}. \quad (4.10)$$

Por outro lado observamos que

$$G_{\tau_{a,b}}(a, a) = G_{\tau_b}(a, a),$$

uma vez que entre os tempos τ_b e $\tau_{a,b}$ a cadeia não pode visitar a nenhuma vez, pois se isto ocorresse, este seria o próprio tempo $\tau_{a,b}$. Juntando tal fato com a definição do tempo de comutação por $t_{a \leftrightarrow b} = \mathbb{E}_a(\tau_{a,b})$, a equação (4.10) pode ser reescrita como

$$\frac{G_{\tau_b}(a, a)}{t_{a \leftrightarrow b}} = \frac{c(a)}{c_G}.$$

O resultado segue ao utilizarmos a identidade $G_{\tau_b}(a, a) = c(a)\mathcal{R}(a \leftrightarrow b)$ obtida no Lema 3.16. \square

Corolário 4.7. *A resistência \mathcal{R} satisfaz a desigualdade triangular, isto é, se a, b e c são vértices, então*

$$\mathcal{R}(a \leftrightarrow c) \leq \mathcal{R}(a \leftrightarrow b) + \mathcal{R}(b \leftrightarrow c).$$

Demonstração. Basta combinarmos as desigualdades

$$\begin{aligned} \mathbb{E}_a(\tau_c) &\leq \mathbb{E}_a(\tau_b) + \mathbb{E}_b(\tau_c), \\ \mathbb{E}_c(\tau_a) &\leq \mathbb{E}_c(\tau_b) + \mathbb{E}_b(\tau_a) \end{aligned}$$

vistas na demonstração do Lema 4.3 com a Proposição 4.6 e a igualdade (4.7) que obtemos

$$\begin{aligned} \mathcal{R}(a \leftrightarrow c) &= \frac{t_{a \leftrightarrow c}}{c_G} \\ &= \frac{\mathbb{E}_a(\tau_c) + \mathbb{E}_c(\tau_a)}{c_G} \\ &\leq \frac{[\mathbb{E}_a(\tau_b) + \mathbb{E}_b(\tau_c)] + [\mathbb{E}_c(\tau_b) + \mathbb{E}_b(\tau_a)]}{c_G} \\ &= \frac{\mathbb{E}_a(\tau_b) + \mathbb{E}_b(\tau_a)}{c_G} + \frac{\mathbb{E}_b(\tau_c) + \mathbb{E}_c(\tau_b)}{c_G} \\ &= \frac{t_{a \leftrightarrow b}}{c_G} + \frac{t_{b \leftrightarrow c}}{c_G} \\ &= \mathcal{R}(a \leftrightarrow b) + \mathcal{R}(b \leftrightarrow c). \end{aligned}$$

\square

Para finalizar esta seção observe que

$$\begin{aligned} t_{com} = \max_{a,b} t_{a \leftrightarrow b} &= \max_{a,b} [\mathbb{E}_a(\tau_b) + \mathbb{E}_b(\tau_a)] \\ &\leq \max_{a,b} \mathbb{E}_a(\tau_b) + \max_{a,b} \mathbb{E}_b(\tau_a) \\ &= 2t_{hit}. \end{aligned} \quad (4.11)$$

Se, por alguma simetria, a cadeia satisfizer a propriedade $\mathbb{E}_a(\tau_b) = \mathbb{E}_b(\tau_a)$ para todos os estados $a, b \in V$, então $\max_{a,b} [\mathbb{E}_a(\tau_b) + \mathbb{E}_b(\tau_a)] = 2 \max_{a,b} \mathbb{E}_a(\tau_b)$, e a desigualdade em (4.11) é na verdade uma igualdade. Neste caso, podemos utilizar t_{com} para obter cotas superiores para t_{hit} (e não apenas cotas inferiores, como é no caso geral). Além disso, neste caso temos que $\mathbb{E}_a(\tau_b) = \frac{1}{2} t_{a \leftrightarrow b}$. Isso será útil, na Seção 6.4 quando formos obter cotas para o t_{hit} no exemplo do passeio aleatório no toro.

4.4 Espectro

Outro importante indicador das cadeias de Markov é o tempo de relaxamento. Ele é definido diretamente a partir do espectro da matriz de transição, fato que não deve surpreender-nos visto que tal matriz determina toda a dinâmica do processo. Antes de defini-lo, vamos explorar um pouco o espectro das matrizes de transição, especialmente nos casos em que a cadeia associada é reversível (neste caso, com certo abuso de notação, escreveremos que a matriz P é reversível).

Definição 4.8. *Para uma matriz de transição P no espaço de estados S , uma função $f : S \rightarrow \mathbb{C}$ é uma autofunção associada autovalor $\lambda \in \mathbb{C}$ se satisfaz $Pf = \lambda f$.*

O produto da matriz P pela função $f : S \rightarrow \mathbb{C}$ pode ser visto como o produto de uma matriz por um vetor a partir do momento que identificamos a função f com um vetor coluna em $\mathbb{C}^{|S|}$, cuja x -ésima entrada é dada por $f(x)$.

Proposição 4.9. *Seja P a matriz de transição de uma cadeia de Markov num espaço de estados S finito.*

- (i) 1 é autovalor de P , com autofunção $f = \mathbf{1} := (1, 1, \dots, 1)^T$;
- (ii) Se λ é um autovalor de P , então $|\lambda| \leq 1$;
- (iii) Se P é irredutível, o autoespaço associado ao autovalor 1 é o espaço unidimensional gerado pelo vetor coluna $\mathbf{1} := (1, 1, \dots, 1)^T$;
- (iv) Se P é irredutível e aperiódica, então -1 não é autovalor de P .

Demonstração. (i) Basta observar que pelo fato de P ser estocástica

$$P\mathbf{1}(x) = \sum_y P(x, y)\mathbf{1}(y) = \sum_y P(x, y) = 1 = \mathbf{1}(x)$$

para todo $x \in S$, mostrando que $P\mathbf{1} = \mathbf{1}$.

(ii) Considere a norma $\|\cdot\|_\infty$ dada por $\|f\|_\infty = \max_{x \in S} |f(x)|$. Como P estocástica, para toda função f e todo estado $x \in S$,

$$|Pf(x)| = \left| \sum_{y \in S} P(x, y)f(y) \right| \leq \sum_{y \in S} P(x, y)|f(y)| \leq \|f\|_\infty \sum_{y \in S} P(x, y) = \|f\|_\infty, \quad (4.12)$$

de modo que tomando o máximo sobre $x \in S$, obtemos $\|Pf\|_\infty \leq \|f\|_\infty$.

Se λ é autovalor de P com autofunção associada $f \neq 0$, isto é, satisfazendo $Pf = \lambda f$. Pela equação (4.12) temos que

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

mostrando que $|\lambda| \leq 1$.

(iii) Seja $[1]$ o autoespaço associado ao autovalor 1. Do item (i) temos que $\text{span}\{\mathbf{1}\} \subset [1]$. Resta-nos mostrar a inclusão contrária quando P é irredutível. Mas isto é imediato visto que se $f \in [1]$ é uma autofunção associada ao autovalor 1, satisfaz $Pf = f$ e portanto é harmônica em relação à P , de modo que o Lema 3.2 garante que f é constante, e portanto $f \in \text{span}\{\mathbf{1}\}$.

(iv) Sendo P irredutível e aperiódica, a Proposição 1.14 garante que existe um r_0 tal que, para todo $r \geq r_0$, $P^r(x, y) > 0$, $\forall x, y \in S$. Isso vale em particular para uma potência $2r > r_0$, mostrando que

$$P^{2r}(x, y) > 0 \quad \forall x, y \in S,$$

o que garante que a matriz P^2 é irredutível.

Suponha por absurdo que -1 é autovalor de P . Neste caso existe uma autofunção $f \neq 0$ tal que $Pf = -f$. Em particular,

$$P^2f = P(Pf) = P(-f) = -Pf = f,$$

de modo que f é uma autofunção da matriz (também estocástica) P^2 , associada ao autovalor 1. Como P^2 é irredutível, o item anterior nos diz que f deve ser constante, ou seja, $f = c\mathbf{1}$, para algum $c \in \mathbb{C}$. Ao mesmo tempo, como f satisfaz $Pf = -f$, deve valer $cP\mathbf{1} = -c\mathbf{1}$, com esta igualdade valendo coordenada a coordenada. Assim fixado um $x \in S$ qualquer, devemos ter

$$c = c \sum_{y \in S} P(x, y) \cdot 1 = (cP\mathbf{1})_x = (-c\mathbf{1})_x = -c,$$

o que mostra que $c = 0$, absurdo uma vez que $f \neq 0$. \square

No restante desta seção estudaremos essencialmente matrizes reversíveis, ou seja, que satisfazem a equação (1.11). O próximo resultado garante que tais matrizes possuem uma base de autofunções reais associadas a autovalores reais. Tendo isso em vista, restringiremos nossa atenção a funções $f : S \rightarrow \mathbb{R}$ e relembremos a definição feita na Seção 2.5.

Definição 4.10. Para uma distribuição π em S , define-se o produto interno ponderado por π como sendo

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

E denota-se por $\ell^2(\pi)$ o espaço $(\mathbb{R}^S, \langle \cdot, \cdot \rangle_\pi)$.

Proposição 4.11. Seja P reversível com respeito à distribuição π .

(i) O espaço com produto interno $(\mathbb{R}^S, \langle \cdot, \cdot \rangle_\pi)$ possui uma base ortonormal formada pelas autofunções $\{f_j\}_{j=1}^{|S|}$ associadas a autovalores reais $\{\lambda_j\}_{j=1}^{|S|}$;

(ii) A matriz P pode ser decomposta como

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

(iii) A autofunção f_1 correspondente ao autovalor 1 pode ser tomada como o vetor constante $\mathbf{1}$, de modo que

$$\frac{P^t(x, y)}{\pi(y)} = 1 + \sum_{j=2}^{|S|} f_j(x)f_j(y)\lambda_j^t. \quad (4.13)$$

Este resultado ainda será utilizado em vários momentos deste texto. Cabe aqui observar que a equação (4.13) estabelece que $P^t(x, y)$ é próximo de $\pi(y)$ e o erro relativo dessa aproximação decai exponencialmente com t . A velocidade dessa convergência é determinada pelo segundo maior autovalor λ_2 (em módulo). Isso justifica que o buraco espectral absoluto e consequentemente o t_{rel} , que serão definidos mais adiante, são características da cadeia que estão relacionadas à velocidade da convergência para o equilíbrio.

Demonstração. (i) O fato de P ser reversível com respeito à medida π garante que ele é um operador auto-adjunto no espaço com produto interno $(\mathbb{R}^S, \langle \cdot, \cdot \rangle_\pi)$, isto é, vale que $\langle Pf, g \rangle_\pi = \langle f, Pg \rangle_\pi$. Neste caso, todos os autovalores de P são reais, conforme podemos encontrar em [4, Lema 10.1]. Além disso, o Teorema Espectral para operadores auto-adjuntos em espaços de dimensão finita, que pode ser encontrado tanto em [4, Teorema 10.2] quanto em [10, Teorema 13.6] garante que as autofunções de P formam uma base ortogonal de $(\mathbb{R}^S, \langle \cdot, \cdot \rangle_\pi)$.

(ii) Para cada $y \in S$ considere a função $\delta_y : S \rightarrow \mathbb{R}$ definida por

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

Como $\{f_j\}_{j=1}^{|S|}$ é uma base ortonormal de $(\mathbb{R}^S, \langle \cdot, \cdot \rangle_\pi)$, δ_y pode ser escrita como

$$\delta_y = \sum_{j=1}^{|S|} \langle \delta_y, f_j \rangle_\pi f_j,$$

donde segue, pela definição do produto interno e da função δ_y , que

$$\delta_y = \sum_{j=1}^{|S|} \langle \delta_y, f_j \rangle_\pi f_j = \sum_{j=1}^{|S|} \left(\sum_{x \in S} \delta_y(x) f_j(x) \pi(x) \right) f_j = \sum_{j=1}^{|S|} f_j(y) \pi(y) f_j. \quad (4.14)$$

Sendo λ_j o autovalor associado à autofunção f_j , vale $P^t f_j = \lambda_j^t f_j$, e consequentemente

$$P^t \delta_y = P^t \left(\sum_{j=1}^{|S|} f_j(y) \pi(y) f_j \right) = \sum_{j=1}^{|S|} f_j(y) \pi(y) P^t f_j = \sum_{j=1}^{|S|} f_j(y) \pi(y) \lambda_j^t f_j.$$

Por outro lado, $(P^t \delta_y)(x) = \sum_{z \in S} P^t(x, z) \delta_y(z) = P^t(x, y)$. Combinando com a equação anterior obtemos

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

ou seja,

$$\frac{P^t(x, y)}{\pi(y)} = \sum_{j=1}^{|S|} f_j(x) f_j(y) \lambda_j^t. \quad (4.15)$$

(iii) Na Proposição 4.9 vimos que a autofunção associada ao autovalor 1 é constante. A única função constante com a norma induzida pelo produto interno $\langle \cdot, \cdot \rangle_\pi$ unitária é $\mathbf{1} = (1, 1, \dots, 1)^T$. Assim $f_1 = \mathbf{1}$ e obtemos pela equação (4.15) que

$$\begin{aligned} \frac{P^t(x, y)}{\pi(y)} &= f_1(x) f_1(y) \lambda_1^t + \sum_{j=2}^{|S|} f_j(x) f_j(y) \lambda_j^t \\ &= 1 + \sum_{j=2}^{|S|} f_j(x) f_j(y) \lambda_j^t. \end{aligned}$$

□

Cabe notar que sendo $\{f_j\}_{j=1}^{|S|}$ uma base ortonormal, qualquer função $f : S \rightarrow \mathbb{R}$ pode ser escrita como

$$f = \sum_{j=1}^{|S|} \langle f, f_j \rangle_{\pi} f_j, \quad (4.16)$$

e sendo $\{\lambda_j\}_{j=1}^{|S|}$ os respectivos autovalores, vale $P^t f_j = \lambda_j^t f_j$ para todo $t > 0$, donde segue que

$$P^t f = \sum_{j=1}^{|S|} \langle f, f_j \rangle_{\pi} \lambda_j^t f_j. \quad (4.17)$$

A partir disso é fácil demonstrarmos o seguinte resultado, que apresenta uma caracterização variacional do segundo maior autovalor da matriz de transição.

Corolário 4.12. *Seja P reversível com respeito à distribuição π . Então*

$$\lambda_2 = \sup_{\langle f, 1 \rangle_{\pi} = 0} \frac{\langle f, Pf \rangle_{\pi}}{\langle f, f \rangle_{\pi}}. \quad (4.18)$$

Demonstração. Sejam $1 = \lambda_1 > \lambda_2 > \dots > \lambda_{|S|} \geq -1$ os autovalores de P , e $f_1 = 1, f_2, \dots, f_{|S|}$ as respectivas autofunções que, como vimos acima, formam uma base ortonormal do espaço $(\mathbb{R}^S, \langle \cdot, \cdot \rangle_{\pi})$.

Se $f \in \mathbb{R}^S$ é tal que $\langle f, 1 \rangle_{\pi} = 0$, replicando as equações (4.16) e (4.17), ficamos com

$$f = \sum_{j=2}^{|S|} \langle f, f_j \rangle_{\pi} f_j$$

e

$$Pf = \sum_{j=2}^{|S|} \langle f, f_j \rangle_{\pi} \lambda_j f_j.$$

Utilizando a distributividade do produto interno e o fato de $\{f_1, \dots, f_{|S|}\}$ ser ortonormal, obtemos que

$$\langle f, f \rangle_{\pi} = \left\langle \sum_{j=2}^{|S|} \langle f, f_j \rangle_{\pi} f_j, \sum_{j=2}^{|S|} \langle f, f_j \rangle_{\pi} f_j \right\rangle_{\pi} = \sum_{j=2}^{|S|} \langle f, f_j \rangle_{\pi}^2. \quad (4.19)$$

e que

$$\langle f, Pf \rangle_{\pi} = \left\langle \sum_{j=2}^{|S|} \langle f, f_j \rangle_{\pi} f_j, \sum_{j=2}^{|S|} \langle f, f_j \rangle_{\pi} \lambda_j f_j \right\rangle_{\pi} = \sum_{j=2}^{|S|} \langle f, f_j \rangle_{\pi}^2 \lambda_j \leq \lambda_2 \sum_{j=2}^{|S|} \langle f, f_j \rangle_{\pi}^2, \quad (4.20)$$

uma vez que $\lambda_2 > \lambda_3, \dots, \lambda_{|S|}$. Combinando as equações (4.19) e (4.20) obtemos que λ_2 é cota superior para a razão em (4.18).

Para mostrar que λ_2 é de fato o supremo, basta observar que ele é atingido com $f = f_2$, visto que esta função satisfaz $\langle f, f_1 \rangle_{\pi} = 0$ e $\langle f, Pf \rangle_{\pi} = \langle f, \lambda_2 f \rangle_{\pi} = \lambda_2 \langle f, f \rangle_{\pi}$. □

Definição 4.13. *Se μ é uma distribuição de probabilidade em S e $f : S \rightarrow \mathbb{R}$ é uma função qualquer, definimos a esperança de f com respeito à probabilidade μ por*

$$E_{\mu}(f) = \sum_{x \in S} f(x) \mu(x)$$

Note que E_μ refere-se a uma esperança sobre funções definidas em S , enquanto que \mathbb{E}_μ definida no início deste trabalho refere-se a uma esperança no espaço $S^{\mathbb{N}}$ das trajetórias da cadeia de Markov que inicia com distribuição μ . As duas estão relacionadas por $E_\mu(f) = \mathbb{E}_\mu(f(X_0))$.

Também observe que se P é reversível com respeito a π e φ é uma autofunção associada a um autovalor $\lambda \neq 1$, então

$$\mathbb{E}_\pi(\varphi) = \sum_{x \in S} \varphi(x)\pi(x) = \sum_{x \in S} \varphi(x)\mathbf{1}(x)\pi(x) = \langle \varphi, \mathbf{1} \rangle_\pi = 0,$$

já que pelo que vimos no resultado anterior, as autofunções são ortogonais entre si. Porém esta igualdade vale mesmo sem a hipótese da reversibilidade:

Proposição 4.14. *Sejam P uma matriz de transição com distribuição estacionária π , e φ uma autofunção de P associada a um autovalor $\lambda \neq 1$. Então $E_\pi[\varphi] = 0$.*

Demonstração. Seja φ autofunção associada ao autovalor $\lambda \neq 1$, isto é, $P\varphi = \lambda\varphi$. Multiplicando a esquerda ambos os lados da igualdade por π e utilizando a estacionariedade de π , obtemos que $\pi\varphi = \lambda\pi\varphi$.

Como

$$E_\pi(\varphi) = \sum \varphi(x)\pi(x) = \pi\varphi,$$

a igualdade anterior nos diz que neste caso

$$E_\pi(\varphi) = \lambda E_\pi(\varphi),$$

donde concluímos que $E_\pi(\varphi) = 0$, uma vez que $\lambda \neq 1$. □

4.5 Tempo de relaxamento

Tendo bem definido o espectro da matriz de transição e suas propriedades, podemos definir mais um indicador da cadeia de Markov: o tempo de relaxamento.

Definição 4.15. *Seja P uma matriz de transição e*

$$\lambda_* := \max\{|\lambda|; \lambda \text{ é um autovalor de } P, \lambda \neq 1\}.$$

*Definimos o **buraco espectral absoluto** de P como sendo a diferença $\gamma_* := 1 - \lambda_*$.*

A Proposição 4.9 garante que se P for reversível, 1 é autovalor com multiplicidade 1, e podemos ordenar os autovalores de P em ordem decrescente

$$1 = \lambda_1 > \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_{|S|} \geq -1.$$

A mesma proposição garante que se P for irredutível e aperiódica, então -1 não é autovalor, e portanto $\lambda_* < 1$ e $\gamma_* > 0$.

Definição 4.16. *O **buraco espectral** de uma cadeia reversível é definido por $\gamma := 1 - \lambda_2$, ou seja, é a diferença entre os dois maiores autovalores da matriz.*

Proposição 4.17. *Seja Q uma cadeia de Markov lazy. Todos os autovalores de Q são não negativos e o buraco espectral é igual ao buraco espectral absoluto, isto é, $\gamma_* = \gamma$.*

Demonstração. Conforme vimos na Seção 1.11, Q pode ser vista como a versão lazy da cadeia $P = 2Q - I$. Neste caso, λ é autovalor de P se, e somente se $\frac{\lambda+1}{2}$ é autovalor de Q . Como sabemos que o espectro de P está contido em $[-1, 1]$, o espectro de Q deve estar contido em $[0, 1]$.

Como todos os autovalores são não-negativos, podemos ordená-los como

$$1 = \lambda_1 > \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_{|S|} \geq 0$$

onde fica claro perceber que $\lambda_* = \max\{|\lambda|; \lambda \text{ é um autovalor de } P, \lambda \neq 1\} = \lambda_2$, e portanto $\gamma_* = \gamma$. \square

A equação (4.13) afirma que quanto menor for λ_* ou, equivalentemente, quanto maior for o buraco espectral γ_* , mais rápida será a convergência de $P^t(x, y)$ para $\pi(y)$. Isso motiva a seguinte definição.

Definição 4.18. *O tempo de relaxamento de uma cadeia de Markov reversível é definido por*

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

onde γ_* é o buraco espectral absoluto definido anteriormente.

Veremos na Seção 5.2 que é possível obter cotas inferiores e superiores para t_{mix} em termos de t_{rel} .

Capítulo 5

Relações Entre os Indicadores

Neste capítulo apresentaremos alguns resultados que relacionam os indicadores definidos no capítulo anterior, além do tempo de mistura definido no Capítulo 2 e da probabilidade de retorno $P^t(x, x)$.

5.1 Relação entre t_{mix} e t_{hit}

Começamos desenvolvendo alguns resultados auxiliares.

Proposição 5.1. *Seja P a matriz de transição de uma cadeia de Markov reversível no espaço de estados S com distribuição estacionária π . Então para todo $t \geq 0$ e todo $x \in S$, vale $P^{2t+2}(x, x) \leq P^{2t}(x, x)$.*

Se além das hipóteses acima a matriz P_L for lazy, isto é, $P_L(x, x) \geq 1/2$ para todo x , então para todo $t \geq 0$ e todo $x \in S$ vale $P_L^{t+1}(x, x) \leq P_L^t(x, x)$.

Demonstração. Da Proposição 4.11 sabemos que

$$P^{2t+2}(x, x) = \pi(x) + \pi(x) \sum_{j=2}^{|S|} f_j(x)^2 \lambda_j^{2t+2} = \pi(x) + \pi(x) \sum_{j=2}^{|S|} f_j(x)^2 \lambda_j^{2t} \lambda_j^2.$$

Como $\lambda_j \in [-1, 1]$, temos que $0 \leq \lambda_j^2 \leq 1$, e assim o lado direito da equação acima é limitado por

$$\pi(x) + \pi(x) \sum_{j=2}^{|S|} f_j(x)^2 \lambda_j^{2t},$$

que por sua vez é igual a $P^{2t}(x, x)$ pela mesma Proposição 4.11.

Se a cadeia P_L for lazy, sabemos da Proposição 4.17 que todos os seus autovalores são não negativos, logo $\lambda_j \in [0, 1]$ para todo $j \geq 2$, e novamente fazendo uso da Proposição 4.11 obtemos que

$$P_L^{t+1}(x, x) = \pi(x) + \pi(x) \sum_{j=2}^{|S|} f_j(x)^2 \lambda_j^t \lambda_j \leq \pi(x) + \pi(x) \sum_{j=2}^{|S|} f_j(x)^2 \lambda_j^t = P_L^t(x, x).$$

□

Proposição 5.2. *Seja $(X_t)_{t \geq 0}$ uma cadeia de Markov no espaço de estados S finito, irredutível e aperiódica, com matriz de transição P e distribuição estacionária π . Então, para todo $x \in S$,*

$$\pi(x) \mathbb{E}_\pi(\tau_x) = \sum_{t=0}^{\infty} [P^t(x, x) - \pi(x)]. \quad (5.1)$$

Demonstração. Fixado $x \in S$, definimos para cada $m \geq 0$,

$$\tau_x^{(m)} := \min\{t \geq m; X_t = x\} = m + \tau_x \circ \theta_m,$$

e

$$\mu_m := P^m(x, \cdot) = \mathbb{P}_x(X_m = \cdot).$$

Como as hipóteses do enunciado são as mesmas do Teorema de Convergência, μ_m converge para π quando $m \rightarrow \infty$. Além disso, temos que

$$\mathbb{E}_x[\tau_x^{(m)}] = \mathbb{E}_x[\mathbb{E}_x[\tau_x^{(m)} | \mathcal{F}_m]] = \mathbb{E}_x[\mathbb{E}_x[m + \tau_x \circ \theta_m | \mathcal{F}_m]] = \mathbb{E}_x[m + \mathbb{E}_{X_m}[\tau_x]].$$

E sendo $\mathbb{E}_{X_m}[\tau_x] = f(X_m)$ uma função de variável aleatória com $f : S \rightarrow \mathbb{R}$ dada por $f(y) = \mathbb{E}_y[\tau_x]$, sua esperança pode ser calculada por

$$\begin{aligned} \mathbb{E}_x[\mathbb{E}_{X_m}[\tau_x]] &= \mathbb{E}_x[f(X_m)] \\ &= \sum_{y \in S} \mathbb{P}_x(X_m = y) f(y) \\ &= \sum_{y \in S} \mu_m(y) f(y) \\ &= \sum_{y \in S} \mu_m(y) \mathbb{E}_y[\tau_x] \\ &= \mathbb{E}_{\mu_m}[\tau_x], \end{aligned}$$

de modo que ficamos com

$$\mathbb{E}_x[\tau_x^{(m)}] = m + \mathbb{E}_{\mu_m}[\tau_x]. \quad (5.2)$$

Como a cadeia é irredutível em um espaço de estados finito, o Lema 1.16 juntamente com a igualdade acima garantem-nos que $\mathbb{E}_x[\tau_x^{(m)}] < \infty$, e conseqüentemente, $\mathbb{P}_x[\tau_x^{(m)} < \infty] = 1$. Além disso, sendo $\tau_x^{(m)}$ o primeiro tempo a partir de m tal que a cadeia visita o estado x , é claro que $\mathbb{P}_x(X_{\tau_x^{(m)}} = \cdot) = \delta_x$. Assim podemos aplicar o Lema 4.4 com $\tau = \tau_x^{(m)}$ e $\mu = \nu = \delta_x$, o qual nos garante que

$$\rho(x) = \mathbb{E}_x \left(\sum_{t=0}^{\tau_x^{(m)}-1} 1_{\{X_t=x\}} \right)$$

satisfaz $\rho P = \rho$. Mais do que isso, como vimos que $\mathbb{E}_x(\tau_x^{(m)}) < \infty$, ele garante também que

$$\pi = \frac{\rho}{\mathbb{E}_x(\tau_x^{(m)})}. \quad (5.3)$$

Por outro lado, note que depois do instante $m-1$, a próxima visita a x será exatamente no tempo $\tau_x^{(m)}$. Combinando tal fato com a linearidade da esperança obtemos que

$$\rho(x) = \mathbb{E}_x \left(\sum_{t=0}^{\tau_x^{(m)}-1} 1_{\{X_t=x\}} \right) = \mathbb{E}_x \left(\sum_{t=0}^{m-1} 1_{\{X_t=x\}} \right) = \sum_{t=0}^{m-1} \mathbb{E}_x[1_{\{X_t=x\}}] = \sum_{t=0}^{m-1} P^t(x, x). \quad (5.4)$$

Combinando as equações (5.3) e (5.4), e utilizando a identidade (5.2), obtemos que

$$\sum_{t=0}^{m-1} P^t(x, x) = \rho(x) = \pi(x) \mathbb{E}_x(\tau_x^{(m)}) = \pi(x) [m + \mathbb{E}_{\mu_m}(\tau_x)],$$

donde segue que

$$\sum_{t=0}^{m-1} [P^t(x, x) - \pi(x)] = \pi(x) \mathbb{E}_{\mu_m}(\tau_x).$$

Fazendo $m \rightarrow \infty$ obtemos

$$\sum_{t=0}^{\infty} [P^t(x, x) - \pi(x)] = \pi(x) \mathbb{E}_{\pi}(\tau_x).$$

□

Lema 5.3. *Para uma cadeia de Markov P reversível com respeito à distribuição π , vale para todo $t \geq 0$*

$$\left| \frac{P^t(x, y)}{\pi(y)} - 1 \right| \leq \sqrt{\frac{P^t(x, x)}{\pi(x)} - 1} \sqrt{\frac{P^t(y, y)}{\pi(y)} - 1}.$$

Demonstração. Basta observar que o Lema 4.11 e a desigualdade de Cauchy-Schwarz nos permitem escrever

$$\begin{aligned} \left(\frac{P^t(x, y)}{\pi(y)} - 1 \right)^2 &= \left(\sum_{j=2}^{|S|} f_j(x) f_j(y) \lambda_j^t \right)^2 \\ &= \left(\sum_{j=2}^{|S|} f_j(x) \sqrt{\lambda_j^t} f_j(y) \sqrt{\lambda_j^t} \right)^2 \\ &\leq \left(\sum_{j=2}^{|S|} f_j(x)^2 \lambda_j^t \right) \left(\sum_{j=2}^{|S|} f_j(y)^2 \lambda_j^t \right) \\ &= \left(\frac{P^t(x, x)}{\pi(x)} - 1 \right) \left(\frac{P^t(y, y)}{\pi(y)} - 1 \right). \end{aligned}$$

□

Feito isso, podemos demonstrar o seguinte resultado, que estabelece uma cota para o tempo de mistura em termos do tempo de acerto da cadeia.

Teorema 5.4. *Seja P a matriz de transição de uma cadeia de Markov reversível e com distribuição estacionária π no espaço de estados S finito. Se a cadeia é lazy, isto é, $P(x, x) \geq 1/2$ para todo x , então o tempo de mistura uniforme definido na Seção 2.5 satisfaz*

$$t_{\text{mix}}^{(\infty)}(1/4) \leq 4 \max_{x \in S} \mathbb{E}_{\pi}[\tau_x] + 1.$$

Em particular,

$$t_{\text{mix}}(1/4) \leq t_{\text{mix}}^{(2)}(1/2) = \left\lceil \frac{1}{2} t_{\text{mix}}^{(\infty)}(1/4) \right\rceil \leq 2 \max_{x \in S} \mathbb{E}_{\pi}[\tau_x] + 1. \quad (5.5)$$

Cabe observar que como $\mathbb{E}_{\pi}[\tau_x] \leq t_{\text{hit}}$, a cota acima garante que sob as condições do Teorema 5.4,

$$t_{\text{mix}} \leq 2t_{\text{hit}} + 1,$$

mostrando que t_{hit} é uma cota para o tempo de mistura, conforme comentado acima.

Demonstração. O fato da cadeia ser lazy garante aperiodicidade, permitindo-nos fazer uso da Proposição 5.2, obtendo que para todo $t \geq 0$

$$\pi(x)\mathbb{E}_\pi\tau_x = \sum_{k=0}^{\infty} [P^k(x, x) - \pi(x)] \geq \sum_{k=0}^t [P^k(x, x) - \pi(x)].$$

O fato da cadeia ser lazy também garante, via Proposição 5.1 que

$$P^k(x, x) \geq P^t(x, x)$$

para todo $k \leq t$, donde obtemos que

$$\pi(x)\mathbb{E}_\pi\tau_x \geq t[P^t(x, x) - \pi(x)].$$

Dividindo ambos os lados por $t\pi(x) > 0$, ficamos com

$$\frac{\mathbb{E}_\pi\tau_x}{t} \geq \frac{P^t(x, x)}{\pi(x)} - 1 = \left| \frac{P^t(x, x)}{\pi(x)} - 1 \right|,$$

já que como $P^t(x, x) \rightarrow \pi(x)$ quanto $t \rightarrow \infty$ e $P^t(x, x)$ decresce em t , devemos ter $P^t(x, x) \geq \pi(x)$ para todo t . Deste modo, segue do Lema 5.3 que

$$\max_x \frac{\mathbb{E}_\pi\tau_x}{t} \geq \max_x \left| \frac{P^t(x, x)}{\pi(x)} - 1 \right| = \max_{x, y} \sqrt{\frac{P^t(x, x)}{\pi(x)} - 1} \sqrt{\frac{P^t(y, y)}{\pi(y)} - 1} \geq \max_{x, y} \left| \frac{P^t(x, y)}{\pi(y)} - 1 \right|.$$

O lado direito é exatamente o tempo de mistura uniforme $d^{(\infty)}$ definido na Seção 2.5, de modo que ficamos com

$$\max_x \frac{\mathbb{E}_\pi\tau_x}{t} \geq d^{(\infty)}(t).$$

Portanto, para $t \geq 4 \max_x \mathbb{E}_\pi\tau_x$ vale $d^{(\infty)}(t) < 1/4$, mostrando a cota do enunciado. Para demonstrar (5.5) basta utilizar as relações entre os tempos de mistura ℓ^p já discutidas neste trabalho. \square

5.2 Relações entre t_{mix} e t_{rel}

Conforme já comentamos quando enunciamos a Proposição 4.11, o espectro da matriz de transição, mais precisamente o tempo de relaxamento, gera cotas para a velocidade de convergência para o equilíbrio, representada nos próximos enunciados pelo tempo de mistura. Começamos com uma cota superior.

Teorema 5.5. *Seja P a matriz de transição de uma cadeia de Markov irredutível e reversível com distribuição estacionária π . Pondo $\pi_{\min} = \min_{x \in S} \pi(x)$, vale*

$$t_{\text{mix}}(\varepsilon) \leq \left\lceil t_{\text{rel}} \left(\frac{1}{2} \log \left(\frac{1}{\pi_{\min}} \right) + \log \left(\frac{1}{2\varepsilon} \right) \right) \right\rceil \leq t_{\text{rel}} \log \left(\frac{1}{\varepsilon\pi_{\min}} \right), \quad (5.6)$$

$$t_{\text{mix}}^\infty(\varepsilon) \leq \left\lceil t_{\text{rel}} \log \left(\frac{1}{\varepsilon\pi_{\min}} \right) \right\rceil. \quad (5.7)$$

Demonstração. O item (iii) da Proposição 4.11, juntamente com a desigualdade triangular garantem que

$$\left| \frac{P^t(x, y)}{\pi(y)} - 1 \right| \leq \left| \sum_{j=2}^{|S|} f_j(x) f_j(y) \lambda_j^t \right| \leq \sum_{j=2}^{|S|} |f_j(x) f_j(y)| |\lambda_j|^t.$$

Pela definição, $\lambda_* = \max\{|\lambda|; \lambda \neq 1 \text{ é autovalor}\} \geq |\lambda_j|$ para todo $j \geq 2$. Utilizando este fato na equação anterior, juntamente com a desigualdade de Cauchy-Schwarz, obtemos que

$$\left| \frac{P^t(x, y)}{\pi(y)} - 1 \right| \leq \lambda_*^t \sum_{j=2}^{|S|} |f_j(x)f_j(y)| \leq \lambda_*^t \left[\sum_{j=2}^{|S|} f_j(x)^2 \sum_{j=2}^{|S|} f_j(y)^2 \right]^{1/2}. \quad (5.8)$$

Por outro lado, da equação (4.14) e da ortonormalidade do conjunto $\{f_j\}_{j=1}^{|S|}$,

$$\begin{aligned} \langle \delta_x, \delta_x \rangle_\pi &= \left\langle \sum_{j=1}^{|S|} f_j(x)\pi(x)f_j, \sum_{j=1}^{|S|} f_j(x)\pi(x)f_j \right\rangle_\pi \\ &= \sum_{j=1}^{|S|} f_j(x)^2 \pi(x)^2 \langle f_j, f_j \rangle_\pi \\ &= \pi(x)^2 \sum_{j=1}^{|S|} f_j(x)^2. \end{aligned}$$

Mas também vale, a partir da definição de δ_x , que

$$\langle \delta_x, \delta_x \rangle_\pi = \sum_{y \in S} \delta_x(y)\delta_x(y)\pi(y) = \pi(x).$$

Deste modo, as duas igualdades anteriores garantem que

$$\pi(x)^{-1} = \sum_{j=1}^{|S|} f_j(x)^2 \geq \sum_{j=2}^{|S|} f_j(x)^2. \quad (5.9)$$

Combinando este último resultado com a equação (5.8) ficamos com

$$\left| \frac{P^t(x, y)}{\pi(y)} - 1 \right| \leq \lambda_*^t [\pi(x)^{-1}\pi(y)^{-1}]^{1/2} = \frac{\lambda_*^t}{\sqrt{\pi(x)\pi(y)}} \leq \frac{(1 - \gamma_*)^t}{\pi_{\min}} \leq \frac{e^{-\gamma_* t}}{\pi_{\min}}, \quad (5.10)$$

onde na última desigualdade, simplesmente utilizamos o fato de $e^x \geq 1 + x$ para todo $x \in \mathbb{R}$. Daqui segue que

$$d^{(\infty)}(t) = \max_x \|q_t(x, \cdot) - 1\|_\infty = \max_{x, y} \left| \frac{P^t(x, y)}{\pi(y)} - 1 \right| \leq \frac{e^{-\gamma_* t}}{\pi_{\min}}. \quad (5.11)$$

Como

$$\frac{e^{-\gamma_* t}}{\pi_{\min}} \leq \varepsilon \iff \log\left(\frac{1}{\varepsilon\pi_{\min}}\right) \leq \gamma_* t \iff t \geq \frac{1}{\gamma_*} \log\left(\frac{1}{\varepsilon\pi_{\min}}\right) = t_{\text{rel}} \log\left(\frac{1}{\varepsilon\pi_{\min}}\right),$$

temos que para $t \geq t_{\text{rel}} \log\left(\frac{1}{\varepsilon\pi_{\min}}\right)$, vale $d^{(\infty)}(t) \leq \varepsilon$, o que garante que

$$t_{\text{mix}}^{(\infty)}(\varepsilon) \leq \left\lceil t_{\text{rel}} \log\left(\frac{1}{\varepsilon\pi_{\min}}\right) \right\rceil,$$

mostrando (5.7).

Para mostrar (5.6), temos das equações (2.19) e (2.20) que

$$d(t) \leq \frac{1}{2}d^{(2)}(t) = \frac{1}{2}\sqrt{d^{(\infty)}(2t)}.$$

Assim, para que $d(t) \leq \varepsilon$ é suficiente que $\frac{1}{2}\sqrt{d^{(\infty)}(2t)} \leq \varepsilon$, o que por sua vez é equivalente a mostrar que $d^{(\infty)}(2t) \leq (2\varepsilon)^2$.

Começamos observando que

$$\begin{aligned} \frac{e^{-\gamma_* 2t}}{\pi_{\min}} \leq (2\varepsilon)^2 &\iff \frac{1}{\pi_{\min}(2\varepsilon)^2} \leq e^{-\gamma_* 2t} \\ &\iff \log\left(\frac{1}{\pi_{\min}}\right) + 2\log\left(\frac{1}{2\varepsilon}\right) \leq \gamma_* 2t \\ &\iff \frac{1}{\gamma_*} \left(\frac{1}{2} \log\left(\frac{1}{\pi_{\min}}\right) + \log\left(\frac{1}{2\varepsilon}\right) \right) \leq t. \end{aligned}$$

Assim, se $t \geq t_{\text{rel}} \left(\frac{1}{2} \log\left(\frac{1}{\pi_{\min}}\right) + \log\left(\frac{1}{2\varepsilon}\right) \right)$, a equação anterior combinada com a equação (5.11) garante que

$$d^{(\infty)}(2t) \leq \frac{e^{\gamma_* 2t}}{\pi_{\min}} \leq (2\varepsilon)^2,$$

e como comentamos anteriormente, isto é suficiente para que $d(t) \leq \varepsilon$. Tal raciocínio mostra que

$$t_{\text{mix}}(\varepsilon) \leq \left\lceil t_{\text{rel}} \left(\frac{1}{2} \log\left(\frac{1}{\pi_{\min}}\right) + \log\left(\frac{1}{2\varepsilon}\right) \right) \right\rceil,$$

provando a primeira desigualdade em (5.6). A segunda desigualdade é imediata das propriedades do logaritmo. \square

O resultado acima fornece uma cota superior para o tempo de mistura em termos do tempo de relaxamento. Também podemos obter uma cota no sentido contrário.

Teorema 5.6. *Seja $\lambda \neq 1$ autovalor da matriz de transição P irredutível e aperiódica. Então*

$$t_{\text{mix}}(\varepsilon) \geq \left(\frac{1}{1 - |\lambda|} - 1 \right) \log\left(\frac{1}{2\varepsilon}\right). \quad (5.12)$$

Em particular, para cadeias reversíveis vale

$$t_{\text{mix}}(\varepsilon) \geq (t_{\text{rel}} - 1) \log\left(\frac{1}{2\varepsilon}\right). \quad (5.13)$$

Demonstração. Para $\lambda = 0$ o resultado vale trivialmente. Seja $\lambda \notin \{0, 1\}$ autovalor de P com autofunção associada f . A Proposição 4.14 garante que $\sum_y \pi(y)f(y) = E_\pi[f] = 0$, de modo que para todo $x \in S$,

$$\begin{aligned} |\lambda^t f(x)| &= |P^t f(x)| = \left| \sum_{y \in S} P^t(x, y) f(y) \right| \\ &= \left| \sum_{y \in S} P^t(x, y) f(y) - \sum_{y \in S} \pi(y) f(y) \right| \\ &\leq \sum_{y \in S} |f(y)| |P^t(x, y) - \pi(y)| \\ &\leq \|f\|_\infty \sum_{y \in S} |P^t(x, y) - \pi(y)|. \end{aligned} \quad (5.14)$$

Ao mesmo tempo, da Proposição 2.2 e da definição de $d(t) = \max_x \|P^t(x, \cdot) - \pi\|_{TV}$ sabemos que para todo $x \in S$,

$$\sum_{y \in S} |P^t(x, y) - \pi(y)| = 2\|P^t(x, \cdot) - \pi\|_{TV} \leq 2d(t).$$

Resultado que combinado com a equação (5.14) garante que

$$|\lambda^t f(x)| \leq \|f\|_\infty 2d(t).$$

Como isto vale para todo $x \in S$, vale em particular para o x tal que $|f(x)| = \|f\|_\infty$, garantindo assim que

$$|\lambda|^t \leq 2d(t). \quad (5.15)$$

A desigualdade acima vale para todo $t > 0$, e em particular, para $t = t_{\text{mix}}(\varepsilon) = \inf\{t \geq 0; d(t) \leq \varepsilon\}$,

$$|\lambda|^{t_{\text{mix}}(\varepsilon)} \leq 2d(t_{\text{mix}}(\varepsilon)) \leq 2\varepsilon,$$

donde segue que

$$\frac{1}{2\varepsilon} \leq |\lambda|^{-t_{\text{mix}}(\varepsilon)}.$$

Como $\log(x) \leq x - 1$ para todo $x > 0$,

$$\log\left(\frac{1}{2\varepsilon}\right) \leq t_{\text{mix}}(\varepsilon) \log(|\lambda|^{-1}) \leq t_{\text{mix}}(\varepsilon)(|\lambda|^{-1} - 1),$$

e portanto

$$\begin{aligned} t_{\text{mix}}(\varepsilon) &\geq \left(\frac{1}{\frac{1}{|\lambda|} - 1}\right) \log\left(\frac{1}{2\varepsilon}\right) \\ &= \left(\frac{1}{1 - |\lambda|} - 1\right) \log\left(\frac{1}{2\varepsilon}\right), \end{aligned}$$

mostrando a desigualdade (5.12).

Se a cadeia for reversível, o tempo de relaxamento é definido por $t_{\text{rel}} = \frac{1}{\gamma_*} = \frac{1}{1 - \lambda_*}$, onde $\lambda_* = \max\{|\lambda|; \lambda \neq 1 \text{ é autovalor de } P\}$. Como a desigualdade (5.12) vale para todo $\lambda \neq 1$, vale em particular para o $\tilde{\lambda}$ tal que $\lambda_* = |\tilde{\lambda}|$, de modo que

$$t_{\text{mix}}(\varepsilon) \geq \left(\frac{1}{1 - |\tilde{\lambda}|} - 1\right) \log\left(\frac{1}{2\varepsilon}\right) = \left(\frac{1}{1 - \lambda_*} - 1\right) \log\left(\frac{1}{2\varepsilon}\right) = (t_{\text{rel}} - 1) \log\left(\frac{1}{2\varepsilon}\right).$$

□

As cotas superiores e inferiores para o tempo de mistura obtidas nos resultados anteriores nos permitem o seguinte resultado:

Corolário 5.7. *Seja P uma cadeia de Markov reversível, irredutível e aperiódica. Então*

$$\lim_{t \rightarrow \infty} d(t)^{1/t} = \lambda_*.$$

Demonstração. Da equação (5.15) sabemos que para todo $\lambda \neq 1$ autovalor de P vale $2d(t) \geq |\lambda|^t$, donde segue que

$$d(t)^{1/t} \geq \left(\frac{1}{2}\right)^{1/t} |\lambda| \rightarrow |\lambda|$$

quando $t \rightarrow \infty$. Tal convergência vale para todo autovalor $\lambda \neq 1$, em particular vale também para λ_* , mostrando que

$$\lim_{t \rightarrow \infty} d(t)^{1/t} \geq \lambda_*.$$

Para a desigualdade inversa, recorreremos à equação (5.10) da demonstração do Teorema 5.5, que garante que sendo π a distribuição estacionária,

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

para todo $t > 0$. Conseqüentemente,

$$d^{(\infty)}(2t) = \max_x \|q_{2t}(x, \cdot) - 1\|_\infty = \max_{x, y} \left| \frac{P^{2t}(x, y)}{\pi(y)} - 1 \right| \leq \frac{\lambda_*^{2t}}{\pi_{\min}}.$$

Na mesma demonstração, também usamos que

$$d(t) \leq \frac{1}{2} d^{(2)}(t) = \frac{1}{2} \sqrt{d^{(\infty)}(2t)},$$

logo

$$d(t)^{1/t} \leq \left(\frac{1}{2} \sqrt{\frac{\lambda_*^{2t}}{\pi_{\min}}} \right)^{1/t} = \left(\frac{1}{2\sqrt{\pi_{\min}}} \lambda_*^t \right)^{1/t} = C^{1/t} \lambda_*,$$

que converge para λ_* quando $t \rightarrow \infty$, mostrando a desigualdade no sentido contrário. \square

5.3 Relação entre t_\odot e o espectro

O tempo do alvo aleatório definido na Seção 4.1 por

$$t_\odot^a := \sum_{x \in S} \mathbb{E}_a[\tau_x] \pi(x) = \mathbb{E}_a[\tau_\pi]$$

está diretamente relacionado ao espectro da matriz de transição, conforme vemos a seguir.

Proposição 5.8. *Seja P uma cadeia de Markov aperiódica, reversível e irredutível com distribuição estacionária π , e seja $\{\lambda_i\}_{i=1}^n$ o espectro de P . Então*

$$t_\odot = \sum_{j=2}^{|S|} \frac{1}{1 - \lambda_j}.$$

Demonstração. O Lema 4.2 garante que $t_\odot = t_\odot^a$ para todo $a \in S$, o que nos permitiu escrever

$$t_\odot = \sum_{a \in S} \pi(a) t_\odot = \sum_a \pi(a) t_\odot^a = \sum_{a, x} \pi(a) \pi(x) \mathbb{E}_a[\tau_x] = \sum_x \pi(x) \mathbb{E}_\pi[\tau_x].$$

Por sua vez, a Proposição 5.2 garante que para todo $x \in S$,

$$\pi(x) \mathbb{E}_\pi[\tau_x] = \sum_{t=0}^{\infty} [P^t(x, x) - \pi(x)],$$

de modo que

$$t_{\odot} = \sum_x \sum_{t=0}^{\infty} [P^t(x, x) - \pi(x)].$$

Utilizando o Lema 4.11 ficamos com

$$t_{\odot} = \sum_x \sum_{t=0}^{\infty} \pi(x) \sum_{j=2}^{|S|} f_j(x)^2 \lambda_j^t.$$

Por fim, alterando a ordem dos somatórios e lembrando que $\sum_x f_j(x)^2 \pi(x) = \langle f_j, f_j \rangle_{\pi} = 1$ uma vez que $\{f_j\}_j$ é ortonormal, concluímos que

$$t_{\odot} = \sum_{j=2}^{|S|} \sum_{t=0}^{\infty} \lambda_j^t \sum_x f_j(x)^2 \pi(x) = \sum_{j=2}^{|S|} \sum_{t=0}^{\infty} \lambda_j^t = \sum_{j=2}^{|S|} \frac{1}{1 - \lambda_j}.$$

□

5.4 Relações ente t_{hit} , t_{rel} e probabilidades de retorno

Nesta seção vamos apresentar alguns resultados novos obtidos em [13], que relacionam os indicadores t_{hit} , t_{rel} e as probabilidades de retorno $P^t(x, x)$. Os principais resultados serão observados para o passeio aleatório *lazy* em um grafo G , mas também serão exibidas algumas relações para cadeias reversíveis gerais.

5.4.1 Resultados para o passeio aleatório *lazy* em um grafo

Começamos definindo o passeio aleatório *lazy* em um grafo G com sendo simplesmente a versão *lazy* do passeio aleatório simples no grafo G definido na Seção 1.9. Vamos formalizar.

Considere $G = (V, E)$ um grafo conexo e finito (sem loops ou arestas paralelas, conforme definido na Seção 1.9) com $n = |V| \geq 2$ vértices. Denotamos por $\text{dist}(x, y)$ a distância entre dois vértices x e y do grafo, a qual é definida como o menor número de arestas necessárias para ir de um vértice até outro. Também denotamos o grau de um vértice $x \in V$ por d_x , além de utilizarmos d_{\min} , d_{\max} e d_{avg} para os graus mínimo, máximo e médio do grafo G respectivamente. O passeio aleatório *lazy* (LRW) no grafo G é a cadeia de Markov em V com matriz de transição P dada por

$$P(x, y) := \begin{cases} \frac{1}{2}, & \text{se } y = x; \\ \frac{1}{2d_x}, & \text{se } \{x, y\} \in E; \\ 0, & \text{caso contrário.} \end{cases} \quad (5.16)$$

O fato de G ser conexo garante que P é irredutível, e conforme vimos na equação (1.16), a distribuição

$$\pi(x) := \frac{d_x}{d_{\text{avg}} \cdot n} \quad (5.17)$$

é estacionária para P . Mais do que isso, P e π satisfazem a equação do balanço detalhado, de modo que P é reversível com respeito à π . Assim sendo, de acordo com as Proposições 4.11 e 4.17, P possui espectro real e não-negativo

$$1 = \lambda_1 > \lambda_2 \geq \dots \geq \lambda_n \geq 0,$$

com autofunção $f_1 = \mathbf{1}$ associada ao autovalor $\lambda_1 = 1$. Como todos os autovalores são não negativos, o buraco espectral absoluto γ_* é igual ao buraco espectral $\gamma = 1 - \lambda_2$, e o tempo de relaxamento da cadeia é

$$t_{\text{rel}} = \frac{1}{\gamma_*} = \frac{1}{1 - \lambda_2}.$$

Além disso, considerando o produto interno

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

suas autofunções $\{f_j\}_{j=1}^n$ (associadas aos autovalores $\{\lambda_j\}_{j=1}^n$) formam uma base ortonormal para o espaço $(\mathbb{R}^V, \langle \cdot, \cdot \rangle_\pi)$.

Sabemos do Teorema 3.11 que toda cadeia reversível em um espaço de estados finito pode ser vista como um passeio aleatório ponderado em uma rede. No caso do LRW no grafo $G = (V, E)$ definido pela matriz P em (5.16), a rede associada é a que possui condutâncias $c(a, b) = 1 = c(b, a)$ para cada aresta $\{a, b\} \in E$, $c(a, a) = d_a$ para cada $a \in V$, e $c(a', b') = 0$ para os demais pares $a', b' \in V$. Para ver isto, basta observar que neste caso, cada vértice $x \in V$ possui condutância

$$c(x) = \sum_{y \sim x} c(x, y) = \sum_{y \sim x, y \neq x} c(x, y) + c(x, x) = \sum_{y \sim x, y \neq x} 1 + c(x, x) = 2d_x,$$

de modo que o passeio aleatório associado a esta rede tem matriz de transição

$$\tilde{P}(x, y) = \frac{c(x, y)}{c(x)} = \begin{cases} \frac{1}{2}, & \text{se } y = x; \\ \frac{1}{2d_x}, & \text{se } \{x, y\} \in E, y \neq x; \\ 0, & \text{caso contrário} \end{cases}$$

que é igual a matriz P que define o LRW. Aproveitamos as contas acima para observar que

$$c_G = \sum_{x \in V} c(x) = \sum_{x \in V} 2d_x = 2 \cdot n \cdot d_{\text{avg}},$$

Antes de apresentar os principais resultados desta seção serão necessários alguns resultados preliminares. Começamos definimos que um conjunto de vértices x_0, \dots, x_l é um **caminho** no grafo G se $x_i \in V$ para todo i , e cada par de pontos consecutivos é ligado por uma aresta, isto é, $\{x_i, x_{i+1}\} \in E$ para todo i . Além disso, dizemos que tal caminho é **geodésico** se for o caminho mais curto entre os vértices x_0 e x_l , ou seja, se não existir outro caminho com menos de l vértices.

Proposição 5.9. *Se x_0, \dots, x_l é um caminho geodésico em G , então $l \leq \frac{3n}{D} - 1$, onde $D = \min\{d_x; x \in \{x_0, \dots, x_l\}\}$. Em particular, como $D \geq d_{\min}$, vale $l \leq \frac{3n}{d_{\min}} - 1$.*

Demonstração. Seja x_0, x_1, \dots, x_l um caminho geodésico e $\mathcal{N}(x) := \{y \in V; \{x, y\} \in E\} \cup \{x\}$ o conjunto que contém x juntamente com todos os seus vizinhos. Uma vez que o grafo não possui loops, $|\mathcal{N}(x)| = d_x + 1 \geq D + 1$.

Como o caminho é geodésico, $\mathcal{N}(x_i) \cap \mathcal{N}(x_j) = \emptyset$ para todo $j \geq i + 2$. De fato, conforme fica claro na Figura 5.1, se existisse um vértice $z \in \mathcal{N}(x_i) \cap \mathcal{N}(x_j)$, o caminho x_0, x_1, \dots, x_l não seria mais geodésico, uma vez que $x_0, \dots, x_{i-1}, x_i, z, x_j, x_{j+1}, \dots, x_l$ vai satisfazer $\{x_k, x_{k+1}\} \in E$ para todo k , sendo assim um caminho que liga x_0 a x_l com comprimento menor que o caminho original.

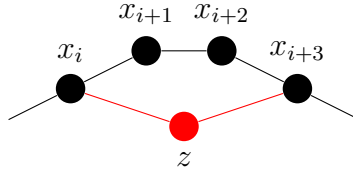


Figura 5.1: Se existisse um nó z nestas condições, o caminho x_0, x_1, \dots, x_l deixaria de ser geodésico.

Assim observe que se $z \in V$ é um vértice qualquer, ele pode aparecer em no máximo 3 conjuntos $\mathcal{N}(x_j)$ distintos. Para exemplificar observe que se $z \in \mathcal{N}(x_0)$, então $z \notin \mathcal{N}(x_j)$ para todo $j \geq 3$, de modo que ele pode aparecer no máximo nos conjuntos $\mathcal{N}(x_0), \mathcal{N}(x_1)$ e $\mathcal{N}(x_2)$. Portanto vale

$$\sum_{i=0}^l |\mathcal{N}(x_i)| \leq 3n.$$

Por outro lado, pelo que vimos anteriormente, $|\mathcal{N}(x)| \geq D + 1$, donde segue que,

$$\sum_{i=0}^l |\mathcal{N}(x_i)| \geq (l + 1)(D + 1),$$

o que unido com a equação anterior nos dá o resultado desejado. \square

O tamanho do grafo G , juntamente com seu grau médio e mínimo, fornece uma cota para o tempo de relaxamento do LRW, conforme mostramos a seguir

Proposição 5.10. *Para o passeio aleatório lazy no grafo G ,*

$$t_{\text{rel}} \leq \max_{x,y \in V} (\mathbb{E}_x[\tau_y] + \mathbb{E}_y[\tau_x]) \leq 6 \left(\frac{d_{\text{avg}}}{d_{\text{min}}} \right) n^2 - 4.$$

Demonstração. Começamos pela primeira desigualdade. Como estamos trabalhando com uma cadeia lazy, $t_{\text{rel}} = \frac{1}{1 - \lambda_2}$. Ao mesmo tempo, o Lema 5.8 garante que

$$t_{\odot} = \sum_{i=2}^n \frac{1}{1 - \lambda_i}.$$

Assim, utilizando a igualdade (4.2) para o tempo de alvo aleatório, temos

$$t_{\text{rel}} = \frac{1}{1 - \lambda_2} \leq \sum_{i=2}^n \frac{1}{1 - \lambda_i} = t_{\odot} = \mathbb{E}_{\pi} \tau_{\pi} = \sum_{a,b \in V} \pi(a)\pi(b) \mathbb{E}_a \tau_b.$$

Para todo $a, b \in V$, $\mathbb{E}_a \tau_b \leq \max_{x,y \in V} (\mathbb{E}_x \tau_y + \mathbb{E}_y \tau_x)$, logo

$$\begin{aligned} t_{\text{rel}} &\leq \sum_{a,b \in V} \pi(a)\pi(b) \max_{x,y \in V} (\mathbb{E}_x \tau_y + \mathbb{E}_y \tau_x) \\ &= \max_{x,y \in V} (\mathbb{E}_x \tau_y + \mathbb{E}_y \tau_x) \sum_{a,b \in V} \pi(a)\pi(b) \\ &= \max_{x,y \in V} (\mathbb{E}_x \tau_y + \mathbb{E}_y \tau_x). \end{aligned}$$

Como definimos na Seção 4.3 o máximo tempo de comutação por $t_{comm} = \max_{a,b \in V} t_{a \leftrightarrow b}$, com $t_{a \leftrightarrow b} = \mathbb{E}_a \tau_b + \mathbb{E}_b \tau_a$, a relação acima pode ser reescrita como

$$t_{rel} \leq \max_{x,y \in V} t_{x \leftrightarrow y} = t_{comm}. \quad (5.18)$$

Para a segunda desigualdade vamos limitar $t_{x \leftrightarrow y}$. Começamos observando que se $a, b \in V$ são dois vértices vizinhos no grafo $G = (V, E)$, então $\mathbb{P}_a(\tau_b < \tau_a^+) \geq P(a, b) = \frac{1}{2d_a}$, uma vez que ir de a para b no primeiro passo faz com que valha $\tau_b < \tau_a^+$, mas não é necessariamente o único jeito (poderia dar voltas no grafo e chegar em b sem voltar ao vértice a). Assim, segue da equação 3.16 que

$$\mathcal{R}(a \leftrightarrow b) = \frac{1}{c(a)\mathbb{P}_a(\tau_b < \tau_a^+)} \leq \frac{1}{(2d_a)\frac{1}{2d_a}} = 1. \quad (5.19)$$

Agora se $x, y \in V$ são dois vértices quaisquer no grafo $G = (V, E)$, tomamos $x_0 = x, x_1, \dots, x_l = y$ um caminho geodésico qualquer. Neste caso $\{x_i, x_{i+1}\} \in E$ para todo i , e pela Proposição 5.9, $l \leq \frac{3n}{d_{min}} - 1$. A desigualdade triangular para resistências efetivas (Corolário 4.7) garante-nos que

$$\mathcal{R}(x \leftrightarrow y) \leq \mathcal{R}(x_0 \leftrightarrow x_1) + \dots + \mathcal{R}(x_{l-1} \leftrightarrow x_l),$$

e como cada termo é limitado superiormente por 1,

$$\mathcal{R}(x \leftrightarrow y) \leq l \leq \frac{3n}{d_{min}} - 1.$$

Da Proposição 4.6 temos que $t_{x \leftrightarrow y} = c_G \mathcal{R}(x \leftrightarrow y)$, logo

$$t_{x \leftrightarrow y} \leq (2 \cdot n \cdot d_{avg}) \cdot \left(\frac{3n}{d_{min}} - 1 \right) = 6n^2 \frac{d_{avg}}{d_{min}} - 2nd_{avg} \leq 6n^2 \frac{d_{avg}}{d_{min}} - 4,$$

uma vez que $n \cdot d_{avg} = \sum_x d_x = 2|E| \geq 2$. A cota superior acima obtida não depende dos vértices x e y , sendo automaticamente uma cota superior para o lado direito da equação (5.18), concluindo a demonstração. \square

Seja

$$G_\tau(y, x) = \mathbb{E}_y \left(\sum_{s=0}^{\tau-1} \mathbf{1}\{X_s = x\} \right)$$

a função de Green definida na equação (3.17), que conta o número esperado de visitas a x até o tempo τ quando começamos no estado y . Introduzimos uma variante desta função, definida por

$$g_\tau(y, x) := \mathbb{E}_y \left(\sum_{s=0}^{\tau} \mathbf{1}\{X_s = x\} \right)$$

que além das passagens anteriores, conta também a passagem por x no tempo τ . Ela pode ser limitada por constantes do grafo como d_{min} , d_{avg} e o tamanho n .

Proposição 5.11. *Considere o passeio aleatório lazy no grafo $G = (V, E)$ e $A \subset V$ não vazio. Fixamos $x \in V \setminus A$ e consideramos o número de retornos ao vértice x até to tempo $\tau_A - 1$*

$$g_{\tau_A-1}(x, x) := \mathbb{E}_x \left[\sum_{s=0}^{\tau_A-1} \mathbf{1}\{X_s = x\} \right].$$

Então

$$\frac{g_{\tau_A-1}(x, x)}{\pi(x)} \leq 9 \left(\frac{d_{avg}n}{d_{min}} \right)^2 (1 - \pi(A)).$$

Demonstração. Fixados um conjunto $A \subset V$ não vazio e $x \in V \setminus A$, o Lema 3.16 garante que

$$G_{\tau_A}(x, x) = c(x)\mathcal{R}(x \leftrightarrow A).$$

Pela definição da rede associada ao LRW, $c(x) = 2d_x$ e $\pi(x) = \frac{c(x)}{c_G} = \frac{c(x)}{2nd_{avg}}$, donde ficamos com

$$G_{\tau_A}(x, x) = \pi(x)(2nd_{avg})\mathcal{R}(x \leftrightarrow A).$$

Como $x \notin A$, é claro pela definição de ambas variantes da função de Green que $G_{\tau_A}(x, x) = g_{\tau_A-1}(x, x)$, valendo portanto

$$\frac{g_{\tau_A-1}(x, x)}{\pi(x)} = 2nd_{avg}\mathcal{R}(x \leftrightarrow A). \quad (5.20)$$

Guardemos esta equação e trabalhemos para limitar seu lado direito. Faremos isso obtendo cotas para $\mathcal{R}(x \leftrightarrow A)$ considerando $B = V \setminus A$ e mostrando que

$$\mathcal{R}(x \leftrightarrow A) \leq \frac{9|B|}{2d_{min}}. \quad (5.21)$$

Começamos com o caso $|B| \leq \frac{2d_{min}}{3}$. Como por hipótese G não possui loops ou arestas paralelas, $x \in B$ possui d_x vizinhos, dos quais no máximo $\frac{2d_{min}}{3} - 1$ estão em B , já que B possui $\frac{2d_{min}}{3}$ vértices, entre eles o próprio x . Assim sendo, x possui ao menos

$$d_x - \left(\frac{2d_{min}}{3} - 1 \right) \geq d_x - \frac{2d_x}{3} = \frac{d_x}{3}$$

vizinhos em $A = V \setminus B$, donde segue que

$$\begin{aligned} \mathcal{R}(x \leftrightarrow A) &= \frac{1}{\mathcal{C}(x \leftrightarrow A)} = \frac{1}{\sum_y c(x, y)\mathbb{P}_y(\tau_A < \tau_x)} \\ &\leq \frac{1}{\sum_{a \in A} c(x, a)} = \frac{1}{\text{número vizinhos de } x \text{ em } A} \leq \frac{3}{d_x}. \end{aligned}$$

Como $\frac{3|B|}{2} \geq 1$,

$$\mathcal{R}(x \leftrightarrow A) \leq \frac{3}{d_x} \leq \frac{3}{d_{min}} \leq \frac{9|B|}{2d_{min}}.$$

Agora para o caso $|B| > \frac{2d_{min}}{3}$, consideramos $y \in A$ o vértice em A mais próximo de x (na distância do grafo). Como cada aresta possui resistência unitária, a desigualdade triangular para resistência efetivas (Corolário 4.7) nos garante que

$$\mathcal{R}(x \leftrightarrow A) \leq \mathcal{R}(x \leftrightarrow y) \leq \text{dist}(x, y),$$

bastando-nos limitar $k = \text{dist}(x, y)$. Para isso, consideramos $x = x_0, x_1, \dots, x_k = y$ um caminho geodésico entre x e y e observamos que x_0, \dots, x_{k-2} devem ter todos os seus vizinhos em B , pois se não fosse, existiria outro vértice em A mais próximo de x que o vértice y .

Assim sendo, o grafo \tilde{G} que contém os vértices x_0, \dots, x_{k-2} juntamente com todos os seus vizinhos e as respectivas arestas está inteiramente contido no conjunto B . Mais que isso, todos os vértices do caminho x_0, \dots, x_{k-2} possuem neste novo grafo \tilde{G} o mesmo grau que possuem no grafo original G . Aplicando a Proposição 5.9 a este caminho no grafo \tilde{G} , obtemos

$$k - 2 \leq \frac{3|B|}{D} - 1$$

onde $D = \min\{d_x; x \in \{x_0, \dots, x_{k-2}\}\} \geq d_{min}$, restando

$$k \leq \frac{3|B|}{d_{min}} + 1,$$

Ainda observamos que como $|B| > \frac{2d_{min}}{3}$, vale

$$|B| + \frac{d_{min}}{3} \leq |B| + \frac{1}{2}|B| = \frac{3|B|}{2}.$$

Assim, voltando para a cota obtida para k , temos que

$$k \leq \frac{3|B|}{d_{min}} + 1 = \left(|B| + \frac{d_{min}}{3}\right) \frac{3}{d_{min}} \leq \frac{9|B|}{2d_{min}},$$

concluindo que $\mathcal{R}(x \leftrightarrow A) \leq \frac{9|B|}{2d_{min}}$.

Verificamos acima que qualquer que seja $|B|$, vale $\mathcal{R}(x \leftrightarrow A) \leq \frac{9|B|}{2d_{min}}$, mostrando assim a equação (5.21).

Seguindo, observamos que como $\pi(B) = \sum_{x \in B} \pi(x) = \sum_{x \in B} \frac{d_x}{d_{avg}n}$, então

$$|B| = \sum_{x \in B} 1 \leq \sum_{x \in B} \frac{d_x}{d_{min}} = \frac{nd_{avg}}{d_{min}} \sum_{x \in B} \frac{d_x}{d_{avg}n} = \frac{nd_{avg}}{d_{min}} \pi(B). \quad (5.22)$$

Para concluir, unindo a equação (5.22) acima e a cota (5.21) mostrada anteriormente, e utilizando isso na equação (5.20), obtemos que

$$\frac{g_{\tau_A-1}(x, x)}{\pi(x)} = (2d_{avg}n)\mathcal{R}(x \leftrightarrow A) \leq (2d_{avg}n) \frac{9|B|}{2d_{min}} \leq \frac{9d_{avg}n}{d_{min}} \frac{nd_{avg}}{d_{min}} \pi(B),$$

e como $B = V \setminus A$, vale $\pi(B) = 1 - \pi(A)$, donde segue o resultado desejado

$$\frac{g_{\tau_A-1}(x, x)}{\pi(x)} \leq 9 \left(\frac{d_{avg}n}{d_{min}}\right)^2 (1 - \pi(A)).$$

□

Antes de prosseguirmos para os próximos resultados vamos definir um conjunto que nos será extremamente útil. Começamos observando que quando trabalhamos com um tempo t determinístico, a função de Green se exprime como

$$g_t(y, x) = \mathbb{E}_y \left(\sum_{s=0}^t \mathbf{1}\{X_s = x\} \right), \quad (5.23)$$

e conta o número esperado de visitas a x até o tempo t , saindo de y . Além disso, como tal soma é finita, é consequência da linearidade da esperança que

$$g_t(y, x) = \mathbb{E}_y \left[\sum_{s=0}^t \mathbf{1}\{x_s = x\} \right] = \sum_{s=0}^t \mathbb{E}_y[\mathbf{1}\{X_s = x\}] = \sum_{s=0}^t P^s(y, x). \quad (5.24)$$

Fixamos $x \in V$ e $\alpha > 1$, e definimos o conjunto

$$A_\alpha := \{y \in V; g_t(y, x) \leq \alpha\pi(x)(t+1)\}. \quad (5.25)$$

Com base na equação (5.23), os elementos $y \in A$ são aqueles que quando partimos dele, o número esperado de visitas a x até um tempo t qualquer é limitado por uma constante que depende de t , a saber $\alpha\pi(x)(t+1)$. Isso significa A_α é o conjunto tal que após o LRW atingi-lo, vai fazer poucos retornos ao vértice x .

Como consequência da igualdade (5.24),

$$\pi(y)g_t(y, x) = \sum_{s=0}^t \pi(y)P^s(y, x).$$

para todo $x, y \in V$. Em especial, trocando x por y , vale

$$\pi(x)g_t(x, y) = \sum_{s=0}^t \pi(x)P^s(x, y).$$

Como P é reversível com respeito à distribuição π , sabemos que para todo $s \geq 0$ vale $\pi(x)P^s(x, y) = \pi(y)P^s(y, x)$, logo, das identidades anteriores, temos que para todos $x, y \in V$ e todo $t \geq 0$,

$$\pi(y)g_t(y, x) = \pi(x)g_t(x, y).$$

Pela definição de A_α , para todo $y \notin A_\alpha$, $\frac{g_t(y, x)}{\alpha(t+1)\pi(x)} \geq 1$, logo,

$$\begin{aligned} 1 - \pi(A_\alpha) &= \pi(A_\alpha^C) = \sum_{y \in A_\alpha^C} \pi(y) \\ &\leq \sum_{y \in A_\alpha^C} \pi(y) \frac{g_t(y, x)}{\alpha(t+1)\pi(x)} \\ &\leq \sum_{y \in V} \frac{\pi(y)g_t(y, x)}{\alpha(t+1)\pi(x)} \\ &= \sum_{y \in V} \frac{\pi(x)g_t(x, y)}{\alpha(t+1)\pi(x)} \\ &= \frac{1}{\alpha(t+1)} \sum_{y \in V} g_t(x, y). \end{aligned}$$

Ao mesmo tempo, como V é finito, segue da definição da função de Green e da linearidade da esperança, que

$$\sum_{y \in V} g_t(x, y) = \sum_{y \in V} \mathbb{E}_x \left[\sum_{s=0}^t \mathbf{1}\{X_s = y\} \right] = \mathbb{E}_x \left[\sum_{s=0}^t \sum_{y \in V} \mathbf{1}\{X_s = y\} \right] = \mathbb{E}_x \left[\sum_{s=0}^t 1 \right] = t + 1,$$

portanto

$$1 - \pi(A_\alpha) \leq \frac{1}{\alpha}, \quad (5.26)$$

mostrando que A_α é grande para α grande.

Os próximos dois resultados permitem limitarmos superiormente $g_t(x, x)$, e serão utilizados para demonstrar, posteriormente, os dois principais resultados desta seção.

Lema 5.12. *Para o passeio aleatório lazy no grafo G , para $x \in V$ e $0 \leq t \leq \lceil t_{\text{rel}} \rceil$,*

$$\frac{g_t(x, x)}{\pi(x)} \leq 6 \frac{d_{\text{avg}} n}{d_{\text{min}}} \sqrt{t + 1}.$$

Demonstração. Fixado $\alpha > 1$, consideramos $A_\alpha = \{y \in V; g_t(y, x) \leq \alpha\pi(x)(t+1)\}$ definido em (5.25). Vimos que $1 - \pi(A_\alpha) \leq 1/\alpha < 1$, o que garante que $A_\alpha \neq \emptyset$.

Afirmamos que

$$\frac{g_t(x, x)}{\pi(x)} \leq \frac{9}{\alpha} \left(\frac{d_{avg}n}{d_{min}} \right)^2 + \alpha(t+1). \quad (5.27)$$

De fato, para $x \in A_\alpha$, da própria definição de A_α temos que

$$\frac{g_t(x, x)}{\pi(x)} \leq \alpha(t+1),$$

e como o outro termo em (5.27) é positivo, a desigualdade vale trivialmente. Já para $x \notin A_\alpha$, começamos por notar que a propriedade forte de Markov aplicada ao tempo de parada τ_{A_α} garante que

$$\begin{aligned} \mathbb{E}_x \left[\sum_{s=\tau_{A_\alpha}}^{\tau_{A_\alpha}+t} \mathbf{1}\{X_s = x\} \right] &= \mathbb{E}_x \left[\mathbb{E}_x \left[\sum_{s=\tau_{A_\alpha}}^{\tau_{A_\alpha}+t} \mathbf{1}\{X_s = x\} \mid \mathcal{F}_{\tau_{A_\alpha}} \right] \right] \\ &= \mathbb{E}_x \left[\mathbb{E}_x \left[\left(\sum_{s=0}^t \mathbf{1}\{X_s = x\} \right) \circ \theta_{\tau_{A_\alpha}} \mid \mathcal{F}_{\tau_{A_\alpha}} \right] \right] \\ &= \mathbb{E}_x \left[\mathbb{E}_{X_{\tau_{A_\alpha}}} \left[\sum_{s=0}^t \mathbf{1}\{X_s = x\} \right] \right] \\ &= \mathbb{E}_x [g_t(X_{\tau_{A_\alpha}}, x)]. \end{aligned}$$

A identidade acima juntamente com a linearidade da esperança nos dá que

$$\begin{aligned} g_t(x, x) &= \mathbb{E}_x \left[\sum_{s=0}^t \mathbf{1}\{X_s = x\} \right] \\ &\leq \mathbb{E}_x \left[\sum_{s=0}^{\tau_{A_\alpha}+t} \mathbf{1}\{X_s = x\} \right] \\ &= \mathbb{E}_x \left[\sum_{s=0}^{\tau_{A_\alpha}-1} \mathbf{1}\{X_s = x\} \right] + \mathbb{E}_x \left[\sum_{s=\tau_{A_\alpha}}^{\tau_{A_\alpha}+t} \mathbf{1}\{X_s = x\} \right] \\ &= g_{\tau_{A_\alpha}-1}(x, x) + \mathbb{E}_x [g_t(X_{\tau_{A_\alpha}}, x)]. \end{aligned}$$

Onde dividindo ambos os lados por $\pi(x)$, ficamos com

$$\frac{g_t(x, x)}{\pi(x)} \leq \frac{g_{\tau_{A_\alpha}-1}(x, x)}{\pi(x)} + \mathbb{E}_x \left[\frac{g_t(X_{\tau_{A_\alpha}}, x)}{\pi(x)} \right]. \quad (5.28)$$

Por outro lado, como $x \notin A_\alpha$ e $A_\alpha \neq \emptyset$, a Proposição 5.11 nos diz que

$$\frac{g_{\tau_{A_\alpha}-1}(x, x)}{\pi(x)} \leq 9 \left(\frac{d_{avg}n}{d_{min}} \right)^2 (1 - \pi(A_\alpha)) \leq 9 \left(\frac{d_{avg}n}{d_{min}} \right)^2 \frac{1}{\alpha}. \quad (5.29)$$

Além disso, como $X_{\tau_{A_\alpha}} \in A_\alpha$, a própria definição de A_α garante que $g_t(X_{\tau_{A_\alpha}}, x) \leq \alpha\pi(x)(t+1)$, logo

$$\mathbb{E}_x \left[\frac{g_t(X_{\tau_{A_\alpha}}, x)}{\pi(x)} \right] \leq \mathbb{E}_x \left[\frac{\alpha\pi(x)(t+1)}{\pi(x)} \right] = \alpha(t+1). \quad (5.30)$$

Deste modo, utilizando as desigualdades (5.29) e (5.30) em (5.28), obtemos

$$\frac{g_t(x, x)}{\pi(x)} \leq 9 \left(\frac{d_{avg}n}{d_{min}} \right)^2 \frac{1}{\alpha} + \alpha(t+1),$$

mostrando que vale (5.27).

Por fim, considerando

$$\alpha = \frac{3d_{avg}n}{d_{min}\sqrt{t+1}}$$

na equação (5.27), obtemos que

$$\begin{aligned} \frac{g_t(x, x)}{\pi(x)} &\leq \frac{9}{\frac{3d_{avg}n}{d_{min}\sqrt{t+1}}} \left(\frac{d_{avg}n}{d_{min}} \right)^2 + \frac{3d_{avg}n}{d_{min}\sqrt{t+1}}(t+1) \\ &= 3 \frac{d_{avg}n}{d_{min}} \sqrt{t+1} + 3 \frac{d_{avg}n}{d_{min}} \sqrt{t+1} \\ &= 6 \frac{d_{avg}n}{d_{min}} \sqrt{t+1}. \end{aligned}$$

□

A escolha de α acima é motivada por este ser um ponto crítico da função $f(\alpha) = \frac{9}{\alpha} \left(\frac{d_{avg}n}{d_{min}} \right)^2 + \alpha(t+1)$, sendo assim um bom chute para ser a menor cota possível obtida através da equação (5.27).

Observe ainda que para valer a construção feita acima é necessário que $\alpha > 1$. Isto é consequência da hipótese $0 \leq t \leq \lceil t_{rel} \rceil$, uma vez que neste caso a Proposição 5.10 garante que

$$t+1 \leq t_{rel} + 2 \leq \left(6 \left(\frac{d_{avg}}{d_{min}} \right) n^2 - 4 \right) + 2 \leq 6 \left(\frac{d_{avg}}{d_{min}} \right) n^2,$$

logo

$$\sqrt{t+1} \leq \sqrt{6} \sqrt{\frac{d_{avg}}{d_{min}}} n,$$

e portanto

$$\alpha = \frac{3d_{avg}n}{d_{min}} \frac{1}{\sqrt{t+1}} \geq \frac{3d_{avg}n}{d_{min}} \frac{1}{\sqrt{6} \sqrt{\frac{d_{avg}}{d_{min}}} n} = \frac{3}{\sqrt{6}} \sqrt{\frac{d_{avg}}{d_{min}}} > 1.$$

Lema 5.13. *Seja P uma cadeia de Markov irredutível em um espaço de estados finito V com distribuição estacionária π . Suponha ainda que P é reversível e que possui espectro não negativo e autofunções $\{f_j\}_{j=1}^n$. Fixados $x \in V$ e $t \geq 0$, temos que,*

$$0 \leq P^t(x, x) - \pi(x) \leq \frac{g_t(x, x) - (t+1)\pi(x)}{t+1}$$

e

$$g_t(x, x) - (t+1)\pi(x) \leq \left(\frac{e}{e-1} \right) g_{(\lceil t_{rel} \rceil - 1) \wedge t}(x, x). \quad (5.31)$$

Demonstração. Comece por observar que a função $f = \frac{1}{\sqrt{\pi(x)}} \delta_x \in \mathbb{R}^V$ satisfaz

$$\langle f_j, f \rangle_\pi = \sum_y f_j(y) f(y) \pi(y) = \sum_y f_j(y) \frac{1}{\sqrt{\pi(x)}} \delta_x(y) \pi(y) = f_j(x) \frac{1}{\sqrt{\pi(x)}} \pi(x) = f_j(x) \sqrt{\pi(x)}.$$

Assim, como o Lema 4.11 garante que

$$P^t(x, x) - \pi(x) = \sum_{j=2}^n f_j(x)^2 \pi(x) \lambda_j^t,$$

temos que

$$P^t(x, x) - \pi(x) = \sum_{j=2}^n \langle f_j, f \rangle_\pi^2 \lambda_j^t. \quad (5.32)$$

Como estamos supondo que o espectro de P é não negativo, e como $\langle f_j, f \rangle_\pi^2$ também é não negativo, segue da equação (5.32) que $P^t(x, x) - \pi(x) \geq 0$. Além disso, como $1 = \lambda_1 > \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_n \geq 0$, cada termo do somatório em (5.32) decresce quando t aumenta, logo $P^t(x, x) - \pi(x)$ decresce com t . Em particular, $P^t(x, x)$ decresce com t , o que significa que para cada t fixo, $P^t(x, x) \leq P^s(x, x)$ para todo $s \leq t$. Se lembrarmos a expressão que tínhamos para a função de Green na equação (5.24), obtemos que

$$g_t(x, x) = \sum_{s=0}^t P^s(x, x) \geq \sum_{s=0}^t P^t(x, x) = (t+1)P^t(x, x),$$

portanto,

$$P^t(x, x) - \pi(x) \leq \frac{g_t(x, x)}{t+1} - \pi(x) = \frac{g_t(x, x) - (t+1)\pi(x)}{t+1},$$

mostrando que vale a primeira afirmação.

Para a outra desigualdade, começamos combinando a expressão para $g_t(x, x)$ obtida em (5.24) com a identidade (5.32) para obter

$$\begin{aligned} g_t(x, x) &= \sum_{s=0}^t P^s(x, x) = \sum_{s=0}^t \left[\pi(x) + \sum_{j=2}^n \langle f_j, f \rangle_\pi^2 \lambda_j^s \right] \\ &= (t+1)\pi(x) + \sum_{j=2}^n \langle f_j, f \rangle_\pi^2 \sum_{s=0}^t \lambda_j^s \\ &= (t+1)\pi(x) + \sum_{j=2}^n \langle f_j, f \rangle_\pi^2 \left(\frac{1 - \lambda_j^{t+1}}{1 - \lambda_j} \right), \end{aligned}$$

isto é,

$$g_t(x, x) - (t+1)\pi(x) = \sum_{j=2}^n \langle f_j, f \rangle_\pi^2 \left(\frac{1 - \lambda_j^{t+1}}{1 - \lambda_j} \right). \quad (5.33)$$

Note agora que para $t = \lceil t_{\text{rel}} \rceil - 1$, como $1 = \lambda_1 > \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_n \geq 0$, vale

$$0 \leq \lambda_i^{t+1} = \lambda_i^{\lceil t_{\text{rel}} \rceil} \leq \lambda_i^{t_{\text{rel}}} \leq \lambda_2^{t_{\text{rel}}} = \lambda_2^{\frac{1}{1-\lambda_2}} \leq e^{-1}$$

para todo $i \geq 2$. Aplicando na relação (5.33) ficamos com

$$\begin{aligned} g_{\lceil t_{\text{rel}} \rceil - 1}(x, x) - \lceil t_{\text{rel}} \rceil \pi(x) &= \sum_{i=2}^n \left(\frac{1 - \lambda_i^{\lceil t_{\text{rel}} \rceil}}{1 - \lambda_i} \right) \langle f_i, f \rangle_\pi^2 \\ &\geq \sum_{i=2}^n \left(\frac{1 - e^{-1}}{1 - \lambda_i} \right) \langle f_i, f \rangle_\pi^2 \\ &= \frac{e-1}{e} \sum_{i=2}^n \frac{1}{1 - \lambda_i} \langle f_i, f \rangle_\pi^2. \end{aligned} \quad (5.34)$$

Como $\lambda_i \in [0, 1)$, vale $\frac{1-\lambda_i^{t+1}}{1-\lambda_i} \leq \frac{1}{1-\lambda_i}$ para todo i . Assim também é consequência de (5.33) que para t qualquer,

$$g_t(x, x) - (t+1)\pi(x) \leq \sum_{i=2}^n \langle f_i, f \rangle_\pi^2 \frac{1}{1-\lambda_i},$$

o que combinando com a equação (5.34) nos dá que

$$g_t(x, x) - (t+1)\pi(x) \leq \frac{e}{e-1} (g_{\lceil t_{\text{rel}} \rceil - 1}(x, x) - \lceil t_{\text{rel}} \rceil \pi(x)) \leq \frac{e}{e-1} g_{\lceil t_{\text{rel}} \rceil - 1}(x, x).$$

A cota acima vale para todo $t \geq 0$, mas para $t \leq \lceil t_{\text{rel}} \rceil - 1$ vale

$$g_t(x, x) - (t+1)\pi(x) \leq g_t(x, x) \leq \frac{e}{e-1} g_t(x, x)$$

uma cota mais fina. Em qualquer caso, mostramos a cota (5.31). \square

Com isto podemos demonstrar os dois resultados principais desta seção, que limitam o tempo de acerto e a probabilidade de retorno do LRW em termos do tempo de relaxamento.

Teorema 5.14. *Para o passeio aleatório lazy no grafo G , para qualquer $x \in V$ e $t \geq 0$:*

$$P^t(x, x) - \pi(x) \leq \frac{10d_x}{d_{\min}} \left(\frac{1}{\sqrt{t+1}} \wedge \frac{\sqrt{t_{\text{rel}}+1}}{t+1} \right)$$

Demonstração. Combinando as duas desigualdades do Lema 5.13 obtemos que

$$P^t(x, x) - \pi(x) \leq \frac{e}{e-1} \frac{g_{(\lceil t_{\text{rel}} \rceil - 1) \wedge t}(x, x)}{t+1}. \quad (5.35)$$

Por outro lado, aplicando o Lema 5.12 para $t = (\lceil t_{\text{rel}} \rceil - 1) \wedge t \leq \lceil t_{\text{rel}} \rceil$ e lembrando que $\pi(x) = \frac{d_x}{d_{\text{avg}}n}$, temos

$$g_{(\lceil t_{\text{rel}} \rceil - 1) \wedge t}(x, x) \leq 6 \frac{d_{\text{avg}}n\pi(x)}{d_{\min}} \sqrt{((\lceil t_{\text{rel}} \rceil - 1) \wedge t) + 1} \leq 6 \frac{d_x}{d_{\min}} \sqrt{(t_{\text{rel}} + 1) \wedge (t + 1)}. \quad (5.36)$$

Então, combinando (5.35) e (5.36), e cotando superiormente $6e/(e-1)$ por 10, obtemos

$$P^t(x, x) - \pi(x) \leq \frac{6e}{e-1} \frac{d_x}{d_{\min}} \frac{\sqrt{(t_{\text{rel}} + 1) \wedge (t + 1)}}{t+1} \leq 10 \frac{d_x}{d_{\min}} \left(\frac{1}{\sqrt{t+1}} \wedge \frac{\sqrt{t_{\text{rel}}+1}}{t+1} \right),$$

o que conclui o teorema. \square

Teorema 5.15. *Para o passeio aleatório lazy no grafo G ,*

$$t_{\text{hit}} \leq \frac{20d_{\text{avg}}}{d_{\min}} n \sqrt{t_{\text{rel}} + 1}$$

Demonstração. Para $x \in V$ a Proposição 5.2 juntamente com a igualdade (5.24) para a função de Green garantem-nos que

$$\begin{aligned} \mathbb{E}_\pi \tau_x &= \frac{1}{\pi(x)} \sum_{s=0}^{\infty} [P^s(x, x) - \pi(x)] \\ &= \frac{1}{\pi(x)} \lim_{t \rightarrow \infty} \left(\sum_{s=0}^t [P^s(x, x) - \pi(x)] \right) \\ &= \frac{1}{\pi(x)} \lim_{t \rightarrow \infty} \left(\sum_{s=0}^t P^s(x, x) - (t+1)\pi(x) \right) \\ &= \frac{1}{\pi(x)} \lim_{t \rightarrow \infty} (g_t(x, x) - (t+1)\pi(x)). \end{aligned} \quad (5.37)$$

O Lema 5.13 nos diz que para todo $t \geq 0$,

$$g_t(x, x) - (t + 1)\pi(x) \leq \frac{e}{e - 1} g_{(\lceil t_{\text{rel}} \rceil - 1) \wedge t}(x, x).$$

Ao mesmo tempo, quando $t \rightarrow \infty$, $(\lceil t_{\text{rel}} \rceil - 1) \wedge t = \lceil t_{\text{rel}} \rceil - 1$, sendo consequência da equação (5.37) que

$$\mathbb{E}_\pi \tau_x \leq \frac{1}{\pi(x)} \lim_{t \rightarrow \infty} \left(\frac{e}{e - 1} g_{(\lceil t_{\text{rel}} \rceil - 1) \wedge t}(x, x) \right) = \frac{e}{e - 1} \frac{g_{\lceil t_{\text{rel}} \rceil - 1}(x, x)}{\pi(x)}.$$

Utilizando o Lema 5.12 com $t = \lceil t_{\text{rel}} \rceil - 1 \leq \lceil t_{\text{rel}} \rceil$ para limitar o lado direito, obtendo

$$\mathbb{E}_\pi \tau_x \leq \frac{e}{e - 1} \left(\frac{6d_{\text{avg}}n}{d_{\text{min}}} \sqrt{\lceil t_{\text{rel}} \rceil} \right) \leq \frac{10d_{\text{avg}}n}{d_{\text{min}}} \sqrt{t_{\text{rel}} + 1}, \quad (5.38)$$

onde limitamos novamente $6e/(e - 1)$ por 10 para simplificar a constante.

Para finalizar, lembramos do Lema 4.3 que $t_{\text{hit}} \leq 2 \max_x \mathbb{E}_\pi \tau_x$, então, como (5.38) vale para todo $x \in V$,

$$t_{\text{hit}} \leq 2 \max_x \mathbb{E}_\pi \tau_x \leq 20 \frac{d_{\text{avg}}n}{d_{\text{min}}} \sqrt{t_{\text{rel}} + 1}.$$

□

5.4.2 Resultados em um contexto mais geral: reversibilidade

Os resultados da seção anterior são válidos para o passeio aleatório lazy em um grafo $G = (V, E)$. Para cadeias de Markov reversíveis com espectro não negativo existem relações semelhantes entre estes indicadores, porém menos finas, valendo

$$t_{\text{hit}} \leq 2 \max_x \left(\frac{1 - \pi(x)}{\pi(x)} \right) t_{\text{rel}}, \quad (5.39)$$

$$\text{e } P^t(x, x) - \pi(x) \leq (1 - t_{\text{rel}}^{-1})^t (1 - \pi(x)). \quad (5.40)$$

Para demonstrarmos tais cotas, construiremos uma matriz P_* satisfazendo as mesmas hipóteses da matriz original P , e tal que

$$\begin{aligned} P_*^t(x, x) - \pi(x) &= (1 - t_{\text{rel}}^{-1})^t (1 - \pi(x)) \\ \text{e } P^t(x, x) - \pi(x) &\leq P_*^t(x, x) - \pi(x). \end{aligned}$$

Deste modo garantiremos que o lado direito de (5.40) é cota superior, e também que é a melhor, uma vez que é atingida por P_* .

Começamos fixando uma matriz P reversível com respeito à distribuição π , e com espectro não negativo $1 = \lambda_1 > \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_n \geq 0$. Neste caso seu tempo de relaxamento é dado por $t_{\text{rel}} = \frac{1}{\gamma}$ onde $\gamma = 1 - \lambda_2$.

Se Π é a matriz definida por $\Pi(x, y) = \pi(y)$ para todo $x, y \in V$, definimos a matriz P_* por

$$P_* = (1 - \gamma)I + \gamma\Pi, \quad (5.41)$$

cuja dinâmica da cadeia de Markov associada é, a cada passo, com probabilidade $1 - \gamma$ ficar parada, e com probabilidade γ mover-se para um estado escolhido aleatoriamente com base na distribuição π . O que motiva a construção de P_* desta forma é exatamente o próximo resultado:

Lema 5.16. *A matriz P_* definida acima possui apenas dois autovalores, a saber, $\lambda_1^* = 1$ com multiplicidade 1 e $\lambda_2^* = \lambda_2 = 1 - \gamma$ com multiplicidade $n - 1$. Assim seu espectro é*

$$1 = \lambda_1^* > \lambda_2^* = \lambda_3^* = \cdots = \lambda_n^* = 1 - \gamma = \lambda_2 \geq 0.$$

Em particular, P_ possui espectro não negativo e o mesmo tempo de relaxamento que a matriz P visto que*

$$t_{\text{rel}}^* = \frac{1}{1 - \lambda_2^*} = \frac{1}{1 - \lambda_2} = t_{\text{rel}}.$$

Demonstração. Sabemos do Lema 4.9 que $\lambda_1^* = 1$ é autovalor de P_* com multiplicidade 1. Agora observe que com $\lambda_2^* = 1 - \gamma = \lambda_2$ temos,

$$P_* - \lambda_2^* I = [(1 - \gamma)I + \gamma\Pi] - (1 - \gamma)I = \gamma\Pi.$$

Como Π possui todas as linhas iguais, $\det(P_* - \lambda_2^* I) = \det(\gamma\Pi) = 0$, mostrando que $\lambda_2^* = 1 - \gamma$ é autovalor de P_* . Mais que isso, tal fato sobre a matriz Π nos diz que ela possui apenas uma linha linearmente independente para n variáveis, possuindo $n - 1$ graus de liberdade, ou seja, $\dim \ker(P_* - \lambda_2^* I) = n - 1$, garantindo-nos que λ_2^* possui multiplicidade geométrica (e consequentemente algébrica) $n - 1$. \square

A matriz P_* também goza de algumas propriedades interessantes, conforme temos nos seguintes lemas:

Lema 5.17. *P_* possui a mesma base de autofunções que a matriz P , além disso, também é reversível com respeito a mesma distribuição π .*

Demonstração. Para a reversibilidade basta verificar a equação do balanço detalhado. Assim, observando que para todo $x, y \in V$ vale $\pi(x)I(x, y) = \pi(y)I(y, x)$ uma vez que $I(x, y) = 1$ apenas para $y = x$, temos que

$$\begin{aligned} \pi(x)P_*(x, y) &= \pi(x)[(1 - \gamma)I(x, y) + \gamma\Pi(x, y)] \\ &= (1 - \gamma)\pi(x)I(x, y) + \gamma\pi(x)\Pi(x, y) \\ &= (1 - \gamma)\pi(y)I(y, x) + \gamma\pi(x)\pi(y) \\ &= \pi(y)[(1 - \gamma)I(y, x) + \gamma\Pi(y, x)] \\ &= \pi(y)P_*(y, x). \end{aligned}$$

Para a outra afirmação do enunciado, começamos por observar que em virtude do Lema 4.9, $f_1^* = (1, 1, \dots, 1)^T = f_1$, restando mostrar que $f_i^* = f_i$ para todo $i \geq 2$.

O Teorema de convergência garante que quando $t \rightarrow \infty$, $P^t f \rightarrow \Pi f$ para toda função $f \in \mathbb{R}^V$. Em particular, se f_i é autofunção de P satisfazendo $P^t f_i = \lambda_i^t f_i$, quando fazemos $t \rightarrow \infty$, temos que $\lambda_i^t \rightarrow 0$ (já que $0 \leq \lambda_i < 1$), e assim

$$\Pi f_i = \lim_{t \rightarrow \infty} P^t f_i = \lim_{t \rightarrow \infty} \lambda_i^t f_i = 0,$$

de modo que para todo $i \geq 2$ ($\lambda_i^* = \lambda_2^* = 1 - \gamma$),

$$P_* f_i = [(1 - \gamma)I + \gamma\Pi]f_i = \lambda_i^* I f_i + \gamma \Pi f_i = \lambda_i^* f_i,$$

mostrando que para $i \geq 2$, f_i é autofunção de P_* associada ao autovalor $\lambda_i^* = \lambda_2$.

Para finalizar, o Lema 4.11 garante que como P_* é reversível com respeito à π , então $\{f_i^*\}_{i=1}^n = \{f_i\}_{i=1}^n$ é base ortonormal de \mathbb{R}^V com respeito ao produto interno π . \square

Lema 5.18. Para toda função $f \in \mathbb{R}^V$ e todo tempo $t \geq 0$, vale

$$\langle f, P^t f \rangle_\pi \leq \langle f, P_*^t f \rangle_\pi.$$

Demonstração. Fixamos $t \geq 0$ e $f \in \mathbb{R}^V$. Conforme vimos na (4.17), como $\{f_j^*\} = \{f_j\}$ é base ortonormal de \mathbb{R}^V , podemos escrever

$$\begin{aligned} P^t f &= \sum_{j=1}^n \langle f, f_j \rangle_\pi f_j \lambda_j^t, \\ P_*^t f &= \sum_{j=1}^n \langle f, f_j^* \rangle_\pi f_j^* (\lambda_j^*)^t, \end{aligned}$$

e também

$$\begin{aligned} f &= \sum_{j=1}^n \langle f, f_j \rangle_\pi f_j, \\ f &= \sum_{j=1}^n \langle f, f_j^* \rangle_\pi f_j^*. \end{aligned}$$

Como as bases são ortonormais em relação ao produto interno $\langle \cdot, \cdot \rangle_\pi$, vale $\langle f_i, f_j \rangle_\pi = \delta_{ij}$ e $\langle f_i^*, f_j^* \rangle_\pi = \delta_{ij}$, assim, da linearidade do produto interno, obtemos que

$$\begin{aligned} \langle f, P^t f \rangle_\pi &= \left\langle \sum_{j=1}^n \langle f, f_j \rangle_\pi f_j, \sum_{i=1}^n \langle f, f_i \rangle_\pi f_i \lambda_i^t \right\rangle_\pi \\ &= \sum_i \sum_j \langle f, f_j \rangle_\pi \langle f, f_i \rangle_\pi \lambda_i^t \langle f_i, f_j \rangle_\pi \\ &= \sum_j \langle f, f_j \rangle_\pi^2 \lambda_j^t. \end{aligned} \tag{5.42}$$

e que

$$\begin{aligned} \langle f, P_*^t f \rangle_\pi &= \left\langle \sum_{j=1}^n \langle f, f_j^* \rangle_\pi f_j^*, \sum_{i=1}^n \langle f, f_i^* \rangle_\pi f_i^* (\lambda_i^*)^t \right\rangle_\pi \\ &= \sum_i \sum_j \langle f, f_j^* \rangle_\pi \langle f, f_i^* \rangle_\pi (\lambda_i^*)^t \langle f_i^*, f_j^* \rangle_\pi \\ &= \sum_j \langle f, f_j^* \rangle_\pi^2 (\lambda_j^*)^t. \end{aligned} \tag{5.43}$$

Como para todo j , $f_j^* = f_j$, então $\langle f, f_j^* \rangle_\pi = \langle f, f_j \rangle_\pi$. Ao mesmo tempo, $\lambda_1^* = \lambda_1$ e para $j \geq 2$, $\lambda_j^* = \lambda_2 \geq \lambda_j$, de modo que vale, para todo j , que $\lambda_j^* \geq \lambda_j$. Assim, segue das equações (5.42) e (5.43) que

$$\langle f, P^t f \rangle_\pi = \sum_j \langle f, f_j \rangle_\pi^2 \lambda_j^t \leq \sum_j \langle f, f_j^* \rangle_\pi^2 (\lambda_j^*)^t = \langle f, P_*^t f \rangle_\pi.$$

□

Com estes resultados, estamos prontos para demonstrar as cotas enunciadas no início desta seção.

Proposição 5.19. *Seja P uma cadeia de Markov irredutível, reversível com respeito à distribuição π , e com espectro não negativo $1 = \lambda_1 > \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_n \geq 0$. Então*

$$t_{\text{hit}} \leq 2 \max_x \left(\frac{1 - \pi(x)}{\pi(x)} \right) t_{\text{rel}}$$

$$\text{e } P^t(x, x) - \pi(x) \leq (1 - t_{\text{rel}}^{-1})^t (1 - \pi(x)).$$

Demonstração. Fixada a matriz P satisfazendo as hipóteses do enunciado, definimos P_* como em (5.41). Dos lemas anteriores sabemos que $\lambda_1^* = 1 = \lambda_1$ e que $\lambda_j^* = \lambda_2$ para todo $j \geq 2$, de modo que vale $\lambda_j^* \geq \lambda_j$ para todo $j = 1, 2, \dots, n$. Ademais, dos lemas anteriores também temos que $f_j^* = f_j$ para todo j . Então, utilizando o Lema 4.11, obtemos que

$$P_*^t(x, x) - \pi(x) = \pi(x) \sum_{j=2}^n f_j^*(x)^2 (\lambda_j^*)^t \geq \pi(x) \sum_{j=2}^n f_j(x)^2 \lambda_j^t = P^t(x, x) - \pi(x). \quad (5.44)$$

O mesmo Lema 4.11 aplicado à matriz P_* nos diz que

$$P_*^t(x, x) = \pi(x) + \pi(x) \sum_{j=2}^n f_j^*(x)^2 (\lambda_j^*)^t,$$

e como $\lambda_j^* = \lambda_2 = 1 - \gamma$ para todo $j \geq 2$, temos que

$$P_*^t(x, x) - \pi(x) = \pi(x) (1 - \gamma)^t \sum_{j=2}^n f_j^*(x)^2. \quad (5.45)$$

Além disso, conforme vimos na demonstração do Teorema 5.5, vale a identidade (5.9), isto é

$$\frac{1}{\pi(x)} = \sum_{j=1}^n f_j^*(x)^2,$$

de modo que vale

$$\sum_{j=2}^n f_j^*(x)^2 = -f_1^*(x)^2 + \sum_{j=1}^n f_j^*(x)^2 = -1 + \frac{1}{\pi(x)} = \frac{1 - \pi(x)}{\pi(x)}. \quad (5.46)$$

Combinando as equações (5.44), (5.45) e (5.46), obtemos que

$$P^t(x, x) - \pi(x) \leq P_*^t(x, x) - \pi(x) = (1 - \gamma)^t (1 - \pi(x)), \quad (5.47)$$

o que mostra a desigualdade (5.40), além de mostrar que tal cota é a melhor possível, uma vez que ela é atingida pela matriz P_* (uma matriz reversível com espectro não negativo).

O fato do espectro de P ser não negativo garante via Lema 4.11 que P é aperiódica. Assim podemos utilizar a Proposição 5.2, que combinada com a desigualdade (5.47) nos dá que

$$\pi(x) \mathbb{E}_\pi \tau_x = \sum_{s=0}^{\infty} (P^s(x, x) - \pi(x)) \leq \sum_{s=0}^{\infty} (1 - \gamma)^s (1 - \pi(x)) = (1 - \pi(x)) \gamma^{-1},$$

ou seja,

$$\mathbb{E}_\pi \tau_x \leq \frac{1 - \pi(x)}{\pi(x)} t_{\text{rel}}.$$

Como P é irredutível, o Lema 4.3 garante que

$$t_{\text{hit}} \leq 2 \max_x \mathbb{E}_\pi \tau_x \leq 2 \max_x \left(\frac{1 - \pi(x)}{\pi(x)} \right) t_{\text{rel}}.$$

□

A partir deste teorema também podemos obter uma cota para o t_{hit} do passeio aleatório *lazy* em um grafo G . Tal cota, embora mais geral, é menos fina que a obtida no Teorema 5.15, visto que a segunda depende de $\sqrt{t_{\text{rel}}}$ enquanto que a primeira depende de t_{rel} .

Corolário 5.20. *Para o passeio aleatório lazy no grafo G ,*

$$t_{\text{hit}} \leq K \frac{d_{\text{avg}} n}{d_{\text{min}}} t_{\text{rel}}, \quad (5.48)$$

para uma certa constante $K > 0$.

Demonstração. Note que $\frac{1 - \pi(x)}{\pi(x)}$ é máximo quando $\pi(x)$ é mínimo. Como no caso do LRW $\pi(x) = \frac{d_x}{d_{\text{avg}} n}$, o denominador é constante, sendo $\pi(x)$ mínimo para $d_x = d_{\text{min}}$. Assim a cota do teorema pode ser expressa por

$$t_{\text{hit}} \leq 2 \max_x \left(\frac{1 - \pi(x)}{\pi(x)} \right) t_{\text{rel}} \leq 2 \left(\frac{1 - \frac{d_{\text{min}}}{d_{\text{avg}} n}}{\frac{d_{\text{min}}}{d_{\text{avg}} n}} \right) t_{\text{rel}} = 2 \left(\frac{d_{\text{avg}} n - d_{\text{min}}}{d_{\text{min}}} \right) t_{\text{rel}} \leq K \frac{d_{\text{avg}} n}{d_{\text{min}}} t_{\text{rel}}.$$

□

Tal resultado encerra nossa exposição das relações entre os indicadores da cadeia. Muitos outros resultados ainda podem ser encontrados, especialmente em [9], mas tornariam este texto mais longo que o necessário. No próximo capítulo nos empenharemos em entender mais sobre estes indicadores ao calculá-los no exemplo do toro.

Capítulo 6

Exemplos Clássicos e o Passeio Aleatório no Toro

Neste capítulo apresentaremos os principais exemplos citados e discutidos ao longo deste trabalho, sendo eles os clássicos exemplo da ruína do jogador e da urna de Pòlya, além dos passeios aleatórios no n -ciclo e no toro. Estes dois últimos serão aproveitados para calcular de forma prática os principais indicadores apresentados no corpo desta dissertação.

6.1 Construção dos exemplos

Começamos definindo as cadeias de Markov relativas aos exemplos citados acima.

6.1.1 Ruína do jogador

No espaço de estados $S = \{0, 1, \dots, n\}$, define-se a matriz de transição P por $P(k, k+1) = P(k, k-1) = \frac{1}{2}$, $P(0, 0) = P(n, n) = 1$, e $P(k, l) = 0$ para todos os demais casos. A cadeia de Markov $(X_t)_{t \geq 0}$ com matriz de transição P é conhecida como a ruína do jogador, pois suas trajetórias representam os caminhos possíveis para um apostador que a cada jogada tem igual chance de ganhar ou perder uma unidade de dinheiro, chegando à fortuna (n dinheiros) ou à ruína (0 dinheiros).

Proposição 6.1. *Seja $(X_t)_{t \geq 0}$ o exemplo da ruína do jogador definido acima. Então para qualquer $k \in S$,*

$$\begin{aligned}\mathbb{P}_k(\tau_n < \tau_0) &= \frac{k}{n}, \\ \mathbb{E}_k[\tau_{\{0, n\}}] &= k(n - k).\end{aligned}$$

Demonstração. Definindo $f(k) := \mathbb{P}_k(\tau_n < \tau_0)$, é claro que $f(0) = 0$ e $f(n) = 1$. Além disso, pela lei da probabilidade total juntamente com a dinâmica da cadeia e sua perda de memória, ao condicionar no primeiro passo, obtemos que para $k \neq 0, n$,

$$\begin{aligned}f(k) &= \mathbb{P}_k(\tau_n < \tau_0) \\ &= \mathbb{P}_k(\tau_n < \tau_0 | X_1 = k-1) \mathbb{P}_k(X_1 = k-1) + \mathbb{P}_k(\tau_n < \tau_0 | X_1 = k+1) \mathbb{P}_k(X_1 = k+1) \\ &= \mathbb{P}_{k-1}(\tau_n < \tau_0) \cdot \frac{1}{2} + \mathbb{P}_{k+1}(\tau_n < \tau_0) \cdot \frac{1}{2} \\ &= \frac{1}{2} f(k-1) + \frac{1}{2} f(k+1),\end{aligned}$$

mostrando que f é harmônica (em relação a matriz P) em $S \setminus \{0, n\}$. Ao mesmo tempo, a função $F(k) := \frac{k}{n}$ também é harmônica neste mesmo espaço e satisfaz $F(0) = 0 = f(0)$ e $F(n) = 1 = f(n)$, de modo que o Princípio da Unicidade (Lema 3.4) garante que $f = F$ em S , mostrando a primeira igualdade do enunciado.

Para a segunda igualdade, definimos a função $g(k) := \mathbb{E}_k[\tau_{\{0,n\}}]$ e observamos que $g(0) = g(n) = 0$ e que para $k \neq 0, n$,

$$g(k) = \mathbb{E}_k[\tau_{\{0,n\}}] = \mathbb{E}_k[\tau_{\{0,n\}} \circ \theta_1 + 1] = 1 + \mathbb{E}_k[\mathbb{E}_k[\tau_{\{0,n\}} \circ \theta_1 \mid \mathcal{F}_1]].$$

Utilizando a propriedade de Markov e a definição da cadeia, ficamos com

$$\begin{aligned} g(k) &= 1 + \mathbb{E}_k[\mathbb{E}_{X_1}[\tau_{\{0,n\}}]] = 1 + \mathbb{E}_k[g(X_1)] = 1 + \sum_{w \in S} g(w) \mathbb{P}_k(X_1 = w) \\ &= 1 + g(k-1) \mathbb{P}_k(X_1 = k-1) + g(k+1) \mathbb{P}_k(X_1 = k+1) \\ &= 1 + \frac{1}{2}g(k-1) + \frac{1}{2}g(k+1). \end{aligned} \tag{6.1}$$

Mas a função $G(k) := k(n-k)$ satisfaz a mesma relação (6.1), além de satisfazer $G(0) = G(n) = 0$. Deste modo, definindo $\eta(k) := G(k) - g(k)$ temos que ela é harmônica (em relação a P) em $S \setminus \{0, n\}$ e satisfaz $\eta(0) = \eta(n) = 0$. Assim como antes, isso basta para concluir, utilizando o Princípio da Unicidade, que $\eta = 0$ em S , mostrando a segunda igualdade do enunciado. \square

6.1.2 Urna de Pólya

Consideremos uma urna contendo inicialmente duas bolas: uma branca e outra preta. A cada passo retiramos aleatoriamente uma bola, e a devolvemos juntamente com outra bola da mesma cor.

Suponha que após termos adicionado k bolas (tendo portanto um total de $k+2$ bolas dentro da urna), a urna contenha j bolas pretas. Então a probabilidade de adicionarmos no próximo passo mais uma bola preta é exatamente a probabilidade de retirarmos da urna uma bola preta, que é dada por $\frac{j}{k+2}$; enquanto que a probabilidade de adicionarmos uma bola branca neste passo é igual à probabilidade de retirarmos uma bola branca da urna, ou seja, $\frac{k+2-j}{k+2}$.

Com a dinâmica definida acima, a sequência de pares ordenados $(B_t, W_t)_{t \geq 0}$ indicando o número de bolas pretas e brancas na urna no tempo t é uma cadeia de Markov no espaço de estados $\{1, 2, \dots\}^2$ cujas transições são dadas por

$$\begin{aligned} P((j, k+2-j), (j+1, k+2-j)) &= \frac{j}{k+2} \\ P((j, k+2-j), (j, k+3-j)) &= \frac{k+2-j}{k+2} \end{aligned}$$

para $j \geq 1$ e $k \geq 0$.

Como a urna começa com 2 bolas, e a cada passo adicionamos mais uma, no tempo k a urna deve contar exatamente $k+2$ bolas, sendo ao menos uma de cada cor. Mas sabemos ainda mais:

Lema 6.2. *Seja B_k o número de bolas pretas na urna de Pólya após a adição de k bolas. A distribuição de B_k é uniforme em $\{1, 2, \dots, k+1\}$.*

Demonstração. A prova será feita por indução. Para $k = 1$ é óbvio uma vez que começando com uma bola de cada cor, no primeiro passo temos $1/2$ de probabilidade de adicionar uma bola

branca e $1/2$ de probabilidade de adicionar uma bola preta, de modo que após a adição desta bola, ficamos com $1/2$ de probabilidade de continuar com uma bola preta (adicionando uma bola da cor branca) e $1/2$ de probabilidade de ficarmos com duas bolas pretas (adicionando uma bola da cor preta), o que mostra que B_1 tem distribuição uniforme em $\{1, 2\}$.

Para o passo de indução, assumimos que B_{k-1} possua distribuição uniforme em $\{1, \dots, k\}$. Para cada $j = 1, 2, \dots, k+1$, condicionando no passo $k-1$, a lei da probabilidade total garante-nos que

$$\begin{aligned} \mathbb{P}(B_k = j) &= \sum_{i=1}^k \mathbb{P}(B_k = j | B_{k-1} = i) \mathbb{P}(B_{k-1} = i) \\ &= \mathbb{P}(B_k = j | B_{k-1} = j-1) \mathbb{P}(B_{k-1} = j-1) + \mathbb{P}(B_k = j | B_{k-1} = j) \mathbb{P}(B_{k-1} = j) \\ &= \mathbb{P}(B_k = j | B_{k-1} = j-1) \frac{1}{k} + \mathbb{P}(B_k = j | B_{k-1} = j) \frac{1}{k}. \end{aligned}$$

Note que na segunda igualdade sobram apenas dois termos já que todos os demais são nulos uma vez que é impossível chegarmos a j bolas pretas no tempo k se não tivermos ou j ou $j-1$ bolas pretas no tempo $k-1$.

No lado direito da igualdade acima, a primeira probabilidade condicional reflete o caso em que estamos com $j-1$ bolas pretas após a adição de $k-1$ bolas e adicionamos mais uma bola preta, o que ocorre com probabilidade $\frac{j-1}{k+1}$; enquanto que a segunda probabilidade condicional reflete o caso em que estamos com j bolas pretas após a adição de $k-1$ bolas, e então não aumentamos o número de bolas pretas, adicionando uma bola da cor branca, o que ocorre com probabilidade $\frac{k+1-j}{k+1}$. Deste modo, o lado direito é igual a

$$\frac{j-1}{k+1} \cdot \frac{1}{k} + \frac{k+1-j}{k+1} \cdot \frac{1}{k} = \frac{1}{k+1},$$

mostrando que B_k tem distribuição uniforme em $\{1, \dots, k+1\}$. \square

Como consequência do resultado anterior, juntamente como o fato da urna conter no tempo k precisamente $k+2$ bolas, temos que para um tempo t fixo, o vetor (B_t, W_t) é uniformemente distribuído sobre o conjunto

$$V_t = \{(x, y); x, y \in \mathbb{Z}, x, y \geq 1, x + y = t + 2\}$$

que representa os pares de números inteiros situados sobre a reta $x + y = t + 2$.

O processo da urna de Pólya pode ser generalizado para d cores. A dinâmica é a mesma de antes: Começamos com uma urna que contém d bolas, uma de cada cor. Então a cada passo uma bola é retirada aleatoriamente da urna, e então devolvida juntamente com uma outra bola da mesma cor da sorteada.

Sendo N_t^i o número de bolas da cor i no tempo t , então mostra-se que o processo $(N_t)_{t \geq 0} = (N_t^1, \dots, N_t^d)_{t \geq 0}$ é uma cadeia de Markov no espaço de estados $\{1, 2, \dots\}^d$.

Lema 6.3. *Seja $(N_t)_{t=0}^\infty$ o processo que representa a Urna de Pólya d -dimensional. Então o vetor N_t tem distribuição uniforme em*

$$V_t = \left\{ (x_1, \dots, x_d); x_i \in \mathbb{Z}, x_i \geq 1 \forall i = 1, \dots, d, e \sum_{i=1}^d x_i = t + d \right\}.$$

Em particular, como $|V_t| = \binom{t+d-1}{d-1}$, vale que

$$\mathbb{P}(N_t = v) = \frac{1}{\binom{t+d-1}{d-1}}$$

para todo $v \in V_t$.

A demonstração deste resultado nada mais é que uma generalização dos argumentos apresentados no Lema 6.2. Para simplificar a notação, representaremos por $\mathbf{1}$ o vetor cujas entradas são todas iguais a 1, e por δ_i o vetor que tem todas entradas nulas com exceção da i -ésima, a qual vale 1. Também vamos dizer que temos a configuração $x = (x_1, \dots, x_d)$ de bolas se tivermos x_1 bolas da cor 1, até x_d bolas da cor d .

Demonstração. Para $t = 1$ o resultado é trivial, pois uma vez que começamos com $N_0 = \mathbf{1}$, isto é, d bolas na urna, uma de cada cor, cada uma tem igual probabilidade de ser retirada. Consequentemente, cada coordenada de N_0 tem igual probabilidade de ter sido acrescida em uma unidade, de modo que N_1 tem distribuição uniforme no conjunto

$$\{\mathbf{1} + \delta_1, \dots, \mathbf{1} + \delta_d\}$$

que facilmente verificamos ser igual ao conjunto V_1 .

Para o passo de indução, supomos que o resultado vale para $t - 1$, isto é, que N_{t-1} tem distribuição uniforme no conjunto V_{t-1} . Para $x = (x_1, \dots, x_d) \in V_t$ qualquer, note que $x - \delta_j \in V_{t-1}$ para todo $j \in \{1, \dots, d\}$. Mais que isso, só é possível o processo $(N_t)_{t \geq 0}$ chegar a x em um passo a partir de algum desses $x - \delta_j$. Logo, pela lei da probabilidade total obtemos que

$$\mathbb{P}(N_t = x) = \sum_{j=1}^d \mathbb{P}(N_t = x | N_{t-1} = x - \delta_j) \mathbb{P}(N_{t-1} = x - \delta_j)$$

As probabilidades condicionais acima referem-se ao evento: Tendo a configuração $x - \delta_j$ de bolas, retirar uma bola da cor j , para então adicionar mais uma bola desta cor e passar para a configuração x de bolas. Como $x - \delta_j \in V_{t-1}$, nesta configuração existem $d + t - 1$ bolas na urna, sendo $x_j - 1$ bolas da cor j , de modo que a probabilidade de retirarmos uma bola desta cor é $\frac{x_j - 1}{d + t - 1}$. Ao mesmo tempo, como N_{t-1} tem distribuição uniforme em V_{t-1} e $x - \delta_j \in V_{t-1}$, vale $\mathbb{P}(N_{t-1} = x - \delta_j) = \frac{1}{|V_{t-1}|}$. Deste modo, a equação acima pode ser reescrita como

$$\begin{aligned} \mathbb{P}(N_t = x) &= \frac{1}{|V_{t-1}|} \sum_{j=1}^d \frac{x_j - 1}{d + t - 1} \\ &= \frac{1}{|V_{t-1}|} \cdot \frac{1}{d + t - 1} \left(\sum_{j=1}^d x_j - d \right) \\ &= \frac{1}{|V_{t-1}|} \cdot \frac{t}{d + t - 1}, \end{aligned} \tag{6.2}$$

mostrando que N_t tem distribuição uniforme em V_t .

Para finalizar mostraremos que $|V_t| = \binom{t+d-1}{d-1}$. De fato, pela construção do conjunto V_t com suas restrições, temos d coordenadas/urnas, cada uma com uma unidade/bola, restando-nos distribuir t unidades nestas d coordenadas (isso fará com que cada coordenada possua ao menos uma unidade, e que a soma de todas as unidades seja igual a $t + d$). Este problema é o mesmo que permutarmos t bolas e $d - 1$ barras em uma fila, uma vez que cada par de barras vai determinar quantas bolas ficam em cada urna. Deste modo, queremos permutar $d + t - 1$ objetos, sendo que as t bolas são iguais e as $d - 1$ barras também são iguais, o que podemos fazer de um total de

$$\frac{(d + t - 1)!}{t!(d - 1)!} = \binom{t + d - 1}{d - 1}$$

maneiras, sendo esta a quantidade de elementos em V_t . □

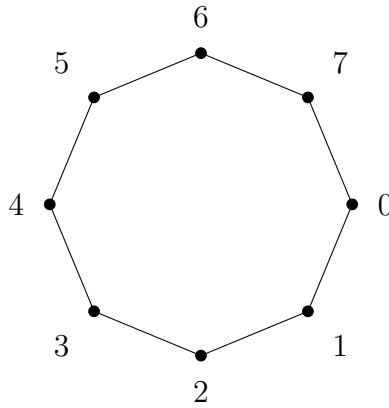


Figura 6.1: O n -ciclo com $n = 8$.

6.1.3 Passeio aleatório no n -ciclo

Para definir o passeio aleatório no n -ciclo, fixamos $n \in \mathbb{N}$ e consideramos $S = \mathbb{Z}_n = \{0, 1, \dots, n-1\}$ o conjunto dos restos da divisão por n . Fixamos também $p, q \geq 0$ com $p + q = 1$, e definimos a seguinte matriz de transição:

$$P(i, j) = \begin{cases} p, & \text{se } j = i + 1 \pmod{n} \\ q, & \text{se } j = i - 1 \pmod{n} \\ 0, & \text{caso contrário} \end{cases}$$

A cadeia de Markov em $S = \mathbb{Z}_n$ com matriz de transição P é conhecida como **passeio aleatório $(p - q)$ -viesado no n -ciclo \mathbb{Z}_n** , que anda no sentido horário com probabilidade p e no sentido anti-horário com probabilidade $q = 1 - p$. No caso particular em que $p = q = 1/2$ temos o **passeio aleatório simples em \mathbb{Z}_n** .

Para enxergar esta cadeia, podemos dispor os estados em um círculo conforme a Figura 6.1, e então pensar que a cada passo uma moeda $(p - q)$ -viesada é lançada. Se o resultado é cara a partícula move-se no sentido horário, enquanto que se o resultado for coroa a partícula move-se no sentido anti-horário.

A distribuição invariante desta cadeia é uniforme $\pi(x) = \frac{1}{n}$ para todo $x \in \mathbb{Z}_n$. Para comprovar isto basta verificar que tal medida satisfaz $\pi P = \pi$. O mesmo raciocínio estende-se para o passeio aleatório lazy $(p - q)$ -viesado no n -ciclo, que é simplesmente a versão *lazy* deste passeio.

6.1.4 Passeio no toro \mathbb{Z}_n^d

O passeio aleatório no Toro é uma generalização do passeio aleatório no n -ciclo definido anteriormente. Antes de mais nada definimos o objeto toro.

Definição 6.4. *O Toro d -dimensional é o grafo cujo conjunto dos vértices é formado pelos pontos do produto cartesiano*

$$\mathbb{Z}_n^d = \mathbb{Z}_n \times \mathbb{Z}_n \cdots \times \mathbb{Z}_n,$$

e onde os vértices $x = (x^1, \dots, x^d)$ e $y = (y^1, \dots, y^d)$ são vizinhos se para algum $j \in \{1, 2, \dots, d\}$ temos $x^i = y^i$ para todo $i \neq j$ e $x^j = y^j \pm 1 \pmod{n}$, ou seja, se eles possuem todas exceto uma coordenada iguais, e neste eles são vizinhos no correspondente n -ciclo.

Definimos o passeio aleatório simples no toro como sendo o passeio aleatório simples no grafo que ele representa, o qual estando em um vértice, pula para cada um de seu vizinhos com

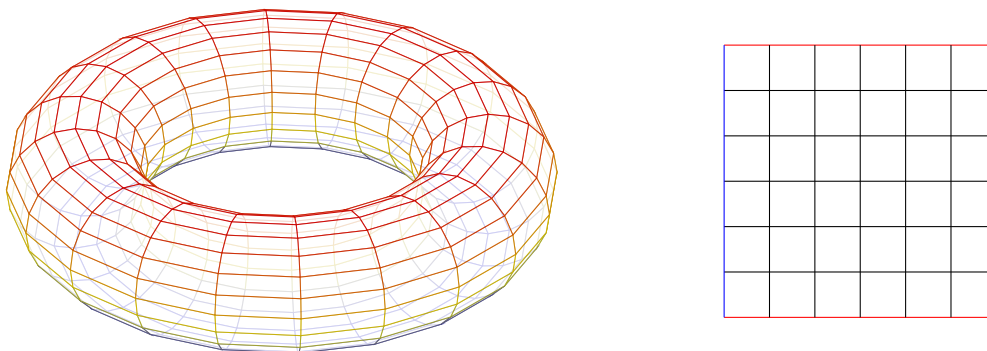


Figura 6.2: Duas representações distintas para o toro. Na esquerda a visualização clássica do toro \mathbb{Z}_{20}^2 . Já na direita temos a visualização do toro \mathbb{Z}_6^2 como o produto cartesiano de $\mathbb{Z}_6 \times \mathbb{Z}_6$ e a identificação dos vértices correspondentes com a mesma cor.

igual probabilidade. Tal processo também pode ser visto do seguinte modo: estando em um vértice $x = (x^1, \dots, x^d)$, escolhe aleatoriamente e uniformemente uma das d coordenadas para mover, então pula com igual probabilidade para cada um de seu vizinhos.

Tendo construídos os exemplos mais importantes, dedicaremos as próximas seções para calcular na prática os principais indicadores das cadeias de Markov definidas no toro. A primeira grandeza definida neste trabalho foi o tempo de mistura (ver Seção 2.4), então nada mais justo que começarmos por ele.

6.2 Tempo de mistura no toro

Começamos por observar que se n for par, o passeio aleatório no n -ciclo só retorna à posição inicial em uma quantidade par de movimentos, conforme vimos no Exemplo 2.13. E o mesmo vale para o passeio aleatório no toro. Em vista disso, os dois processos são periódicos, não valendo o Teorema de Convergência e consequentemente não estando bem definido o tempo de mistura.

Para contornarmos este problema basta em ambos os casos passarmos para a versão *lazy* do passeio, onde a cada passo temos $1/2$ de probabilidade de permanecer parado e $1/2$ de probabilidade de seguir a dinâmica original. Em ambos os casos tal passeio é aperiódico, valendo o teorema de convergência e estando bem definido o tempo de mistura. Começemos com o caso do toro unidimensional, ou seja, o n -ciclo.

6.2.1 t_{mix} no n -ciclo \mathbb{Z}_n

Consideremos o passeio aleatório lazy $(p - q)$ -viesado no n -ciclo, que fica parado com probabilidade $1/2$, move-se no sentido horário com probabilidade $p/2$, e no sentido anti-horário com probabilidade $q/2$, cuja matriz de transição em $S = \mathbb{Z}_n$ é dada por

$$P(i, j) = \begin{cases} p/2, & \text{se } j = i + 1 \pmod{n} \\ q/2, & \text{se } j = i - 1 \pmod{n} \\ 1/2, & \text{se } i = j \pmod{n} \\ 0, & \text{caso contrário} \end{cases}. \quad (6.3)$$

Proposição 6.5 (Tempo de Mistura para o Passeio Aleatório Preguiçoso no n -ciclo). *Para o passeio aleatório lazy $(p - q)$ -viesado no n -ciclo $S = \mathbb{Z}_n$ definido acima (inclusive para*

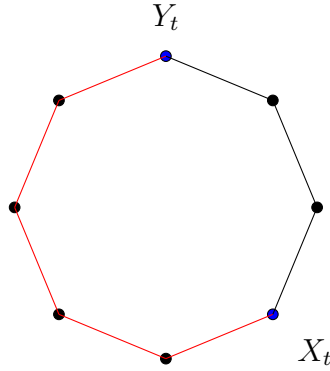


Figura 6.3: O acoplamento $(X_t, Y_t)_{t \geq 0}$ no n -ciclo. O processo $(D_t)_{t \geq 0}$ mede a distância no sentido horário entre X_t e Y_t .

$p = q = 1/2$),

$$\frac{1}{32}n^2 \leq t_{\text{mix}} \leq n^2.$$

Demonstração. A cota superior baseia-se inteiramente na Proposição 2.20 que garante que $t_{\text{mix}} \leq \lceil 4 \max_{x,y} \mathbb{E}_{x,y}[\tau_{\text{couple}}] \rceil$. Nesta expressão observamos que o máximo é tomado sobre todos os estados $x, y \in S$, e não sobre os acoplamentos que têm $X_0 = x$ e $Y_0 = y$. Essa sutileza permite-nos, a cada par de estados $x, y \in S$, escolher um acoplamento conveniente $(X_t, Y_t)_{t \geq 0}$ satisfazendo $X_0 = x$ e $Y_0 = y$, para então calcularmos $\mathbb{E}_{x,y}[\tau_{\text{couple}}]$. Lembremos que a única condição a ser satisfeita pelo acoplamento é que individualmente, $(X_t)_{t \geq 0}$ e $(Y_t)_{t \geq 0}$ tenham matriz de transição P dada por (6.3).

Fixados $x, y \in \mathbb{Z}_n$, construiremos um acoplamento $(X_t, Y_t)_{t \geq 0}$ de modo que as partículas representadas por $(X_t)_{t \geq 0}$ e $(Y_t)_{t \geq 0}$ nunca andem simultaneamente, evitando assim que uma possa “pular” sobre a outra quando elas estiverem em estados vizinhos. Para isso, definimos o acoplamento de seguinte modo: até as duas partículas se encontrarem, a cada unidade de tempo, jogamos uma moeda justa, independente das moedas anteriores, para determinar qual das duas partículas vai saltar e qual vai ficar parada; então a partícula selecionada move-se no sentido horário com probabilidade p e no sentido anti-horário com probabilidade q ; a partir do momento em que as partículas se encontram, elas andam juntas, ficando paradas com probabilidade $1/2$, e andando no sentido horário com probabilidade $p/2$ e no sentido anti-horário com probabilidade $q/2$.

A dinâmica definida acima faz com que individualmente, cada partícula performe o passeio aleatório lazy $(p - q)$ -viesado em \mathbb{Z}_n , de modo que se definirmos $X_0 = x$ e $Y_0 = y$, teremos $(X_t, Y_t)_{t \geq 0}$ o acoplamento desejado.

Para determinar $\mathbb{E}_{x,y}[\tau_{\text{couple}}]$ vamos olhar para o processo $(D_t)_{t \geq 0}$ descrito na Figura 6.3, que a cada tempo t mede a distância no sentido horário entre X_t e Y_t . Abrindo as contas para encontrar a matriz de transição de $(D_t)_{t \geq 0}$ verificamos que sua dinâmica é a mesma da Ruína do Jogador em $\{0, 1, \dots, n\}$ vista na Seção 6.1.1, onde $\tau := \min\{t \geq 0; D_t \in \{0, n\}\}$ satisfaz $\mathbb{E}_k[\tau] = k(n - k)$ para qualquer $k \in \{1, \dots, n\}$.

Pelo modo como foram definidos, $\mathbb{E}_{x,y}[\tau_{\text{couple}}] = \mathbb{E}_k[\tau]$ com k sendo a distância inicial no sentido horário entre x e y . Como isso vale para cada par de estados $x, y \in \mathbb{Z}_n$, temos que

$$\max_{x,y \in \mathbb{Z}_n} \mathbb{E}_{x,y}[\tau_{\text{couple}}] = \max_{k \in \{0, \dots, n\}} \mathbb{E}_k[\tau] = \max_{k \in \{0, \dots, n\}} k(n - k) = \frac{n^2}{4}.$$

Agora basta utilizar a Proposição 2.20 para concluir que $t_{\text{mix}} \leq \lceil 4 \max_{x,y} \mathbb{E}_{x,y}[\tau_{\text{couple}}] \rceil = n^2$.

Para a cota inferior começamos supondo que $X_0 = x_0$. Sendo $(S_t)_{t \geq 0}$ o passeio aleatório

lazy $(p - q)$ -viesado em \mathbb{Z} , ele pode ser representado via sistema dinâmico aleatório¹ como

$$S_t = Z_1 + \dots + Z_t,$$

onde Z_1, \dots, Z_t são variáveis aleatórias independentes satisfazendo $\mathbb{P}(Z_s = -1) = \frac{q}{2}$, $\mathbb{P}(Z_s = 0) = \frac{1}{2}$ e $\mathbb{P}(Z_s = 1) = \frac{p}{2}$. Neste caso $\mathbb{E}Z_s = \frac{p-q}{2}$, $\mathbb{E}[Z_s^2] = \frac{1}{2}$, e

$$\text{Var}[Z_s] = \mathbb{E}[Z_s^2] - [\mathbb{E}Z_s]^2 = \frac{1}{2} - \left(\frac{p-q}{2}\right)^2 = \frac{1}{2} - \frac{1-4pq}{4} = \frac{1}{4} + pq,$$

de modo que $\mathbb{E}(S_t) = t(p - q)/2$ e $\text{Var}(S_t) = t\left(\frac{1}{4} + pq\right) \leq \frac{t}{2}$.

Escrevemos $X_t = S_t \bmod n$, e consideramos ρ a distância no n -ciclo, que é dada pelo menor número de arestas entre os dois vértices. Fixando $\mu_t = \mathbb{E}(S_t) = t(p - q)/2$, definimos o conjunto

$$A_t := \{k; \rho(k, \lfloor x_0 + \mu_t \rfloor \bmod n) \geq n/4\}.$$

Sendo π a distribuição invariante de $(X_t)_{t \geq 0}$ vista na Seção 6.1.3, é claro que $\pi(A) \geq \frac{1}{2}$.

Por outro lado, a desigualdade de Chebyshev garante que

$$\mathbb{P}(X_t \in A_t) \leq \mathbb{P}(|S_t - \mu_t| \geq n/4) \leq \frac{8t}{n^2} < \frac{1}{4}$$

para $t < \frac{n^2}{32}$. Deste modo, segue da definição de $d(t)$ na Seção 2.4 que para $t < \frac{n^2}{32}$,

$$d(t) \geq \pi(A_t) - \mathbb{P}(X_t \in A_t) > \frac{1}{2} - \frac{1}{4},$$

o que garante que $t_{\text{mix}} \geq \frac{n^2}{32}$. □

6.2.2 t_{mix} no toro \mathbb{Z}_n^d

O n -ciclo \mathbb{Z}_n é o caso particular do toro \mathbb{Z}_n^d quando $d = 1$. Generalizaremos as ferramentas utilizadas no cálculo do seu tempo de mistura para calcular este mesmo indicador no caso do passeio aleatório lazy no toro.

Para esta construção será necessário o seguinte resultado, conhecido como identidade de Wald, que também é válido em contextos mais gerais.

Lema 6.6 (Identidade de Wald). *Seja $(T_j)_{j=1}^\infty$ uma sequência de variáveis aleatórias positivas, todas com mesma média $\mathbb{E}(T_j) = m$ e seja N uma variável aleatória com esperança finita, independente da sequência $(T_j)_{j=1}^\infty$, tomando valores inteiros positivos. Então*

$$\mathbb{E}\left(\sum_{j=1}^N T_j\right) = m\mathbb{E}(N)$$

Demonstração.

$$\begin{aligned} \mathbb{E}\left(\sum_{j=1}^N T_j\right) &= \mathbb{E}\left(\sum_{j=1}^\infty T_j \mathbf{1}_{\{N \geq j\}}\right) = \sum_{j=1}^\infty \mathbb{E}(T_j \mathbf{1}_{\{N \geq j\}}) \\ &= \sum_{j=1}^\infty \mathbb{E}(T_j) \mathbb{E}(\mathbf{1}_{\{N \geq j\}}) = m \sum_{j=1}^\infty \mathbb{P}(N \geq j) = m\mathbb{E}(N) \end{aligned}$$

□

¹Veja a definição deste objeto na Seção 1.10.

Proposição 6.7 (Tempo de Mistura para o Passeio Aleatório Preguiçoso no Toro). *Para o passeio aleatório preguiçoso no toro $\mathbb{T} = \mathbb{Z}_n^d$,*

$$\frac{dn^2}{32} \leq t_{\text{mix}} \leq d^2 n^2.$$

Demonstração. Para a cota superior, utilizaremos novamente a Proposição 2.20, a qual garante que $t_{\text{mix}} \leq \lceil 4 \max_{x,y} \mathbb{E}_{x,y}[\tau_{\text{couple}}] \rceil$. Começamos fixando um par de estados $x, y \in \mathbb{T}$ e construímos um acoplamento $(X_t, Y_t)_{t \geq 0}$ tal que $X_0 = x$ e $Y_0 = y$.

Definimos a dinâmica do acoplamento do seguinte modo: a cada passo, escolhe uma das d coordenadas ao acaso; escolhida a coordenada i , se a posição das cadeias $(X_t)_{t \geq 0}$ e $(Y_t)_{t \geq 0}$ concordam nesta coordenada, ou seja, se $X_t^i = Y_t^i$, então movemos as duas cadeias juntas por $+1, -1$ ou 0 nesta coordenada, com probabilidade $1/4, 1/4$ e $1/2$ respectivamente; já se a posição das duas cadeias for diferente nesta coordenada isto é, se $X_t^i \neq Y_t^i$, então jogamos uma moeda justa para definir qual das cadeias se move, e a escolhida se move ± 1 na coordenada i , com sinal escolhido novamente pelo lançamento de uma moeda justa independente da anterior.

A dinâmica acima faz com que, individualmente, $(X_t)_{t \geq 0}$ e $(Y_t)_{t \geq 0}$ possuam a mesma dinâmica do passeio aleatório preguiçoso no toro \mathbb{T} , sendo portanto $(X_t, Y_t)_{t \geq 0}$ um acoplamento com $X_0 = x$ e $Y_0 = y$.

Note que fixada uma coordenada i e escrevendo $X_t = (X_t^1, \dots, X_t^d)$ e $Y_t = (Y_t^1, \dots, Y_t^d)$ para cada t , podemos definir

$$\tau_i = \min\{t \geq 0; X_t^i = Y_t^i\}$$

o tempo necessário para que as duas cadeias do acoplamento concordem na coordenada i .

Se olharmos para a dinâmica do acoplamento $(X_t^i, Y_t^i)_{t \geq 0}$ apenas nos tempos em que a coordenada i é selecionada, este processo que chamaremos de $(\tilde{X}_s^i, \tilde{Y}_s^i)_{s \geq 0}$ tem o mesmo comportamento do acoplamento que utilizamos para o exemplo do n -ciclo na Proposição 6.5, de modo que se $\tilde{\tau}_i$ é o tempo em que \tilde{X}_s^i e \tilde{Y}_s^i se encontram, ele é limitado por $n^2/4$.

Voltando ao processo $(X_t^i, Y_t^i)_{t \geq 0}$, como a coordenada i é selecionada a cada passo com probabilidade $1/d$, este processo espera um tempo T com distribuição geométrica de média d até se mover, de modo que

$$\tau_i = \sum_{j=1}^{\tilde{\tau}_i} T_j$$

onde T_j são variáveis aleatórias independentes com distribuição T e independentes de $\tilde{\tau}_i$.

A identidade de Wald (Lema 6.6) garante então que

$$\mathbb{E}_{x,y}[\tau_i] = \mathbb{E}[T] \mathbb{E}_{x,y}[\tilde{\tau}_i] = \frac{dn^2}{4}.$$

E como

$$\tau_{\text{couple}} = \max_{1 \leq i \leq d} \tau_i \leq \sum_{i=1}^d \tau_i,$$

temos que

$$\mathbb{E}_{x,y}[\tau_{\text{couple}}] \leq \sum_{i=1}^d \mathbb{E}_{x,y}[\tau_i] \leq \sum_{i=1}^d \frac{dn^2}{4} = \frac{d^2 n^2}{4}.$$

Como essa construção vale para qualquer par de estados $x, y \in \mathbb{T}$, o resultado segue da Proposição 2.20.

A demonstração da cota inferior é uma simples adaptação da prova da mesma cota para o passeio aleatório no n -ciclo feita anteriormente.

Suponha que $X_0 = x = (x^1, \dots, x^d)$. Sendo, para cada t , $X_t = (X_t^1, \dots, X_t^d)$, a dinâmica da primeira coordenada X_t^1 é equivalente à dinâmica de $x^1 + S_t \pmod n$ onde $(S_t)_{t \geq 0}$ é o passeio aleatório em \mathbb{Z} , que tem $S_0 = 0$ e que, em cada passo, incrementa uma unidade com probabilidade $\frac{1}{4d}$, diminui uma unidade com probabilidade $\frac{1}{4d}$ ou fica parado com probabilidade $1 - \frac{1}{2d}$. Note que podemos escrever

$$S_t = Z_1 + \dots + Z_t$$

onde $(Z_s)_{s \geq 1}$ são variáveis aleatórias independentes com $\mathbb{P}(Z_s = \pm 1) = \frac{1}{4d}$ e $\mathbb{P}(Z_s = 0) = 1 - \frac{1}{2d}$. Assim fica fácil deduzir que $\text{Var}(Z_s) = \mathbb{E}(Z_s^2) = \frac{1}{2d}$ e portanto $\text{Var}(S_t) = \frac{t}{2d}$.

Seja ρ a distância no ciclo \mathbb{Z}_n , $A = \{k \in \mathbb{Z}_n : \rho(k, x^1) \geq \frac{n}{4}\}$ e $\bar{A} = A \times \mathbb{Z}_n^{d-1}$. Note que $|\bar{A}| \geq \frac{n^d}{2}$, e portanto $\pi(\bar{A}) \geq \frac{1}{2}$. Além disso, pela desigualdade de Chebyshev, obtemos

$$\mathbb{P}_x(X_t \in \bar{A}) = \mathbb{P}_x(X_t^1 \in A) \leq \mathbb{P}\left(|S_t| \geq \frac{n}{4}\right) \leq \frac{16\text{Var}(S_t)}{n^2} = \frac{8t}{dn^2} < \frac{1}{4},$$

se $t < dn^2/32$. Assim, para $t < dn^2/32$ temos

$$d(t) \geq \pi(\bar{A}) - \mathbb{P}_x(X_t \in \bar{A}) > \frac{1}{2} - \frac{1}{4} = \frac{1}{4},$$

o que mostra que $t_{\text{mix}} \geq dn^2/32$. □

6.3 Tempo de relaxamento no toro

O segundo indicador que calcularemos no toro é o tempo de relaxamento, definido na Seção 4.5. Assim como antes começaremos calculando-o no n -ciclo, para depois estender as contas para o caso geral.

6.3.1 t_{rel} no n -ciclo \mathbb{Z}_n

Fixado $n \in \mathbb{N}$ com $n \geq 2$, seja $(X_t)_{t \geq 0}$ o passeio aleatório simples no n -ciclo $S = \mathbb{Z}_n$ definido na Seção 6.1.3, cuja matriz de transição é dada por $P(x, y) = \frac{1}{2}$ se $y \equiv x \pm 1 \pmod n$, e $P(x, y) = 0$ caso contrário. Para determinarmos o tempo de relaxamento desta cadeia basta encontramos o espectro da sua matriz de transição.

Começamos observando que se $f : S \rightarrow \mathbb{R}$ é uma função qualquer, vale para todo $x \in S$

$$Pf(x) = \sum_{z \in S} P(x, z)f(z) = P(x, x-1)f(x-1) + P(x, x+1)f(x+1) = \frac{f(x-1) + f(x+1)}{2}, \tag{6.4}$$

onde as somas em S são feitas mod n . Para cada $j = 0, 1, \dots, n-1$, a função $\varphi_j : S \rightarrow \mathbb{C}$ dada por

$$\varphi_j(k) = \exp\left\{\frac{2\pi i j k}{n}\right\}$$

onde i é a unidade imaginária, satisfaz

$$P\varphi_j(k) = \frac{\varphi_j(k-1) + \varphi_j(k+1)}{2} = \left(\frac{e^{2\pi i j/n} + e^{-2\pi i j/n}}{2}\right) e^{2\pi i j k/n} = \cos\left(\frac{2\pi j}{n}\right) \varphi_j(k),$$

sendo ela autofunção da matriz P , associada ao autovalor $\lambda_j = \cos\left(\frac{2\pi j}{n}\right)$

Estes autovalores não são todos distintos, sendo $\lambda_0 = 1$ e $\lambda_j = \lambda_{n-j}$ para $1 \leq j \leq n-1$. Além disso, se n é par (caso em que a cadeia é periódica) $\lambda_{n/2} = -1$, com autofunção real $\varphi_{n/2}(k) = (-1)^k$.

Como P é matriz real, tanto a parte real quanto a parte imaginária de φ_j são autofunções. Assim, para cada autovalor λ_j , $1 \leq j < n/2$, obtemos o par de autofunções reais linearmente independentes

$$f_j(k) = \cos\left(\frac{2\pi k j}{n}\right) \quad \text{e} \quad g_j(k) = \sin\left(\frac{2\pi k j}{n}\right).$$

Deste modo obtemos n autofunções reais linearmente independentes e concluimos que o espectro de P é o conjunto

$$\left\{ \lambda_j = \cos\left(\frac{2\pi j}{n}\right); j = 0, 1, \dots, n-1 \right\}.$$

que pode ser representado graficamente como o conjunto das projeções das raízes n -ésimas da unidade no eixo real, conforme a Figura 6.4.

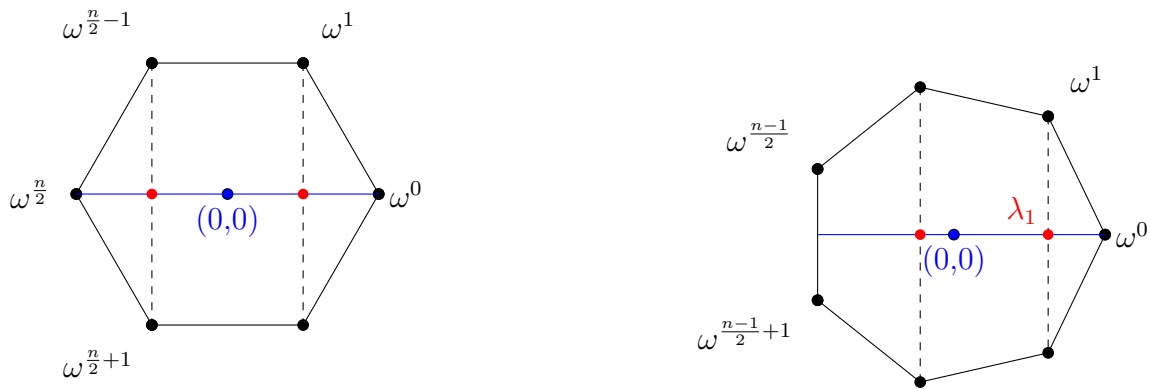


Figura 6.4: Raízes n -ésimas da unidade com n par ($n = 6$) e ímpar ($n = 7$). As projeções destas raízes no eixo horizontal são o espectro da matriz P do passeio aleatório no respectivo n -ciclo.

A partir de agora, em algumas partes do texto, usaremos a notação $f(n) = O(g(n))$ para dizer que existe uma constante C tal que $|f(n)| \leq C|g(n)|$ para todo n .

O segundo maior autovalor é $\lambda_1 = \cos(2\pi/n) = 1 - \frac{2\pi^2}{n^2} + O(n^{-4})$ e portanto o buraco espectral é $\gamma = 1 - \lambda_1 = \frac{2\pi^2}{n^2} + O(n^{-4})$. Quanto ao buraco espectral absoluto, quando n é par temos $\lambda_* = -1$, logo $\gamma_* = 0$. Já quando n é ímpar² $\lambda_* = |\lambda_{(n-1)/2}|$, sendo $\gamma_* = 1 - |\lambda_{(n-1)/2}| = 1 - \cos(\pi/n)$ que é menor que γ mas também é da ordem n^{-2} .

Em ambos os casos (n par ou ímpar) podemos considerar a cadeia lazy $\tilde{P} = \frac{I+P}{2}$ que tem espectro positivo com segundo maior autovalor $\tilde{\lambda}_1 = \frac{1+\lambda_1}{2}$, de modo que o buraco espectral é $\tilde{\gamma} = \gamma/2 = \frac{\pi^2}{n^2} + O(n^{-4})$, que coincide com o absoluto $\tilde{\gamma}_*$. Portanto, o tempo de relaxamento $t_{\text{rel}} = \tilde{\gamma}^{-1}$ da cadeia lazy é da ordem de n^2 , isto é, existem constantes c e C tais que

$$cn^2 \leq t_{\text{rel}} \leq Cn^2.$$

Mais precisamente temos que

$$t_{\text{rel}} = \frac{n^2}{\pi^2}(1 + O(n^{-2})).$$

É interessante notar neste exemplo que mesmo no caso n ímpar temos $\gamma_* < \tilde{\gamma}_*$, ou seja, a versão lazy da cadeia, a qual anda mais lentamente, tem buraco espectral absoluto maior e portanto um tempo de relaxamento menor.

²Para ver que quando n é ímpar $\lambda_* = \lambda_{(n-1)/2}$, olhamos para a representação dos autovalores no círculo unitário da Figura 6.4. O polígono por eles formado contém n vértices, entre eles o vértice $(1,0)$. Como a distância no círculo entre cada par de vértices é igual, a distância entre $w^0 = 1$ e w_1 é a mesma distância entre $w^{(n-1)/2}$ e $w^{(n-1)/2+1}$ que são os dois vértices da “base” deste polígono. E daqui fica claro que a projeção destes dois vértices precisa estar mais próxima do ponto $(-1,0)$ do que a projeção do vértice w^1 precisa estar do vértice $(1,0)$, logo $|\text{Re}(w^{(n-1)/2})| > |\text{Re}(w^1)|$, e portanto $|\lambda_{(n-1)/2}| > |\lambda_1|$.

6.3.2 t_{rel} no toro \mathbb{Z}_n^d

Uma simples generalização do raciocínio da subseção anterior é suficiente para obtermos o tempo de relaxamento do passeio aleatório no toro. Consideremos inicialmente o passeio não-lazy em \mathbb{Z}_n^d . Para $f : \mathbb{Z}_n^d \rightarrow \mathbb{C}$, a ação de P neste caso é

$$Pf(y) = \frac{1}{2d} \sum_{\|z\|=1} f(y+z).$$

Com as mesmas contas que fizemos no caso unidimensional, se mostra que, para cada $x = (x_1, x_2, \dots, x_d) \in \mathbb{Z}_n^d$, a função complexa $\varphi_x : \mathbb{Z}_n^d \rightarrow \mathbb{C}$ dada por

$$\varphi_x(y) = \exp \left\{ \frac{2\pi i x \cdot y}{n} \right\}$$

satisfaz

$$P\varphi_x(y) = \lambda_x \varphi_x(y),$$

onde

$$\lambda_x = \frac{1}{d} \sum_{k=1}^d \cos \left(\frac{2\pi x_k}{n} \right).$$

Assim obtemos que o conjunto $\{\varphi_x; x \in \mathbb{Z}_n^d\}$ é uma base ortogonal de autofunções de P .

O segundo maior autovalor é obtido quando x é um vetor da base canônica $\{e_1, \dots, e_d\}$ de \mathbb{R}^d . Neste caso, temos que $\lambda_{e_j} = \frac{1}{d}(d-1 + \cos(2\pi/n))$. Assim, o buraco espectral é dado por $\gamma = 1 - \lambda_{e_1} = \frac{1 - \cos(2\pi/n)}{d}$.

Considerando a versão lazy, temos que o buraco espectral é $\tilde{\gamma} = \frac{1 - \cos(2\pi/n)}{2d}$. Note que esse valor difere por um fator d daquele obtido no toro unidimensional. Assim, obtemos a seguinte estimativa para o buraco espectral do passeio lazy em \mathbb{Z}_n^d :

$$t_{\text{rel}} = \frac{dn^2}{\pi^2} (1 + O(n^{-2})).$$

6.4 Tempo de acerto no toro

Para finalizar esta exposição dos indicadores, vamos nesta seção encontrar o tempo de acerto do passeio aleatório simples no toro. A rede associada a esta cadeia é definida sobre o grafo do toro, com todas as condutâncias iguais a 1, e sem loops. Como o toro d dimensional possui dn^d arestas, $c_G = 2 \sum_{e \in E} c(e) = 2dn^d$. Além disso, a simetria do passeio aleatório simples no toro faz com que, para quaisquer sítios $a, b \in \mathbb{Z}_n^d$ tenhamos $\mathbb{E}_a \tau_b = \mathbb{E}_b \tau_a$, de modo que $t_{a \leftrightarrow b} = \mathbb{E}_a \tau_b + \mathbb{E}_b \tau_a = 2\mathbb{E}_a \tau_b$. Mas vimos na Proposição 4.6 que

$$t_{a \leftrightarrow b} = c_G \mathcal{R}(a \leftrightarrow b) = 2dn^d \mathcal{R}(a \leftrightarrow b),$$

de modo que para o passeio aleatório simples no toro vale

$$\mathbb{E}_a \tau_b = dn^d \mathcal{R}(a \leftrightarrow b). \quad (6.5)$$

Esta equação, juntamente com a definição do tempo de acerto vista na Seção 4.2

$$t_{\text{hit}} = \max_{x, y \in S} \mathbb{E}_x[\tau_y].$$

serão as ferramentas essenciais desta seção.

Ao contrário do que fizemos nas seções anteriores para a obtenção dos demais indicadores, não vamos dividir esta seção apenas entre caso unidimensional e o caso geral. Ao invés disso demandaremos a primeira subseção para obtenção das cotas inferiores, a segunda para a obtenção das cotas superiores exclusivamente para o caso $d \geq 3$, e fecharemos utilizando a terceira para obter as cotas superiores do caso $d = 2$. Ao final destas teremos demonstrado o seguinte resultado:

Proposição 6.8. *Considere o passeio aleatório simples no toro \mathbb{Z}_n^d . Existem constantes $0 < c_d \leq C_d < \infty$ tais que se x e y estão a uma distância $k \geq 1$, então,*

$$c_d n^d \leq \mathbb{E}_x \tau_y \leq C_d n^d \tag{6.6}$$

uniformemente em k se $d \geq 3$, e

$$c_2 n^2 \log(k) \leq \mathbb{E}_x \tau_y \leq C_2 n^2 \log(k + 1) \tag{6.7}$$

para $d = 2$.

O caso do toro unidimensional foi deixado de fora da proposição acima por ser extremamente simples. De fato, basta observar que identificando os estados 0 e n no n -ciclo, o tempo $\mathbb{E}_x[\tau_y]$ para o passeio aleatório simples no n -ciclo é o mesmo que o tempo esperado para o sucesso ou fracasso da ruína do jogador em $\{0, 1, \dots, n\}$ quando começa-se em $k = \text{dist}(x, y)$, que de acordo com a Proposição 6.1 é dado por $\mathbb{E}_k[\tau_0] = k(n - k)$. Assim

$$t_{\text{hit}} = \max_{x,y} \mathbb{E}_x[\tau_y] = \max_{0 \leq k \leq n} \mathbb{E}_k[\tau_0] = \max_{0 \leq k \leq n} k(n - k).$$

O lado direito da última igualdade é igual a $\frac{n^2}{4}$ quando n é par, e igual a $\frac{n^2 - 1}{4}$ quando n é ímpar, o que nos permite em ambos os casos escrever

$$t_{\text{hit}} = \left\lfloor \frac{n^2}{4} \right\rfloor.$$

Vamos agora focar na demonstração da Proposição 6.8. Para isso o seguinte lema desempenhará um papel fundamental.

Lema 6.9. *Seja B_n o grafo que representa o grid $n \times n$ em dimensão 2, cujos vértices são pares de inteiros (z, w) tais que $1 \leq z, w \leq n$, e cujas arestas são pares de pontos com distância euclidiana unitária, isto é, são da forma $\{x, x + \delta_i\}$ para $i = 1, 2$ com $x, x + \delta_i \in V(B_n)$ o conjunto de vértices do B_n .*

Seja $a = (1, 1)$ o canto inferior esquerdo e $z = (n, n)$ o canto superior direito do quadrado B_n . Se cada aresta de B_n possuir condutância unitária, então a resistência efetiva entre a e z satisfaz

$$\frac{\log(n)}{2} \leq \mathcal{R}(a \leftrightarrow z) \leq 2 \log(n)$$

Demonstração. **Para a cota inferior,** consideramos Π_k , conforme na Figura 6.5 o conjunto de arestas

$$\Pi_k = \{\{v, w\} \in E(B_k); \|v\|_\infty = k, \|w\|_\infty = k + 1\},$$

onde $\|v\|_\infty = \max\{v_1, v_2\}$ para $v = (v_1, v_2)$. Observamos que Π_k contém exatamente $2k$ arestas, cada uma com condutância unitária, logo $\sum_{e \in \Pi_k} c(e) = |\Pi_k| = 2k$.

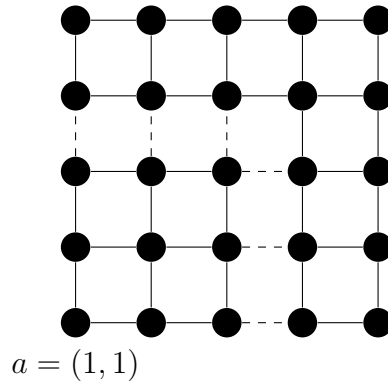


Figura 6.5: O grafo B_5 com o conjunto separador Π_3 representado pelas arestas tracejadas

A construção de Π_k faz com que todo caminho entre a e z contenha obrigatoriamente uma aresta em Π_k , sendo ele um conjunto separador de a e z . Além disso, todos os conjuntos Π_k ($1 \leq k \leq n-1$) são disjuntos, seguindo da Proposição 3.29 que

$$\mathcal{R}(a \leftrightarrow z) \geq \sum_{k=1}^{n-1} \left(\sum_{e \in \Pi_k} c(e) \right)^{-1} = \sum_{k=1}^{n-1} \frac{1}{2k} \geq \frac{\log(n)}{2}.$$

Para a **cota superior**, utilizaremos o Princípio de Thomson (Teorema 3.23), que garante que

$$\mathcal{R}(a \leftrightarrow z) = \inf \{ \mathcal{E}(\theta); \theta \text{ é um fluxo unitário entre } a \text{ e } z \},$$

ou seja, que a energia de qualquer fluxo unitário de a para z é uma cota superior para $\mathcal{R}(a \leftrightarrow z)$. A urna de Pólya nada tem a ver com o processo original do passeio aleatório simples em B_n , mas é peça chave para construirmos um fluxo entre a e z , cuja energia será a cota superior enunciada.

Considere o processo $(N_t)_{t \geq 0} = (B_t, W_t)_{t \geq 0}$, onde as coordenadas indicam a quantidade de bolas pretas e brancas, respectivamente, na Urna de Pólya, no tempo t . Relembramos que este processo começa com $(B_0, W_0) = (1, 1) = a$ e sua dinâmica já foi discutida na Subseção 6.1.2.

Consideramos para cada k , a reta

$$V_k := \{(x_1, x_2) \in \mathbb{Z}^2; x_1 + x_2 = k + 1\}.$$

Então rodamos o processo N_t (que por construção fica restrito ao quadrado B_n) até que ele atinja a diagonal principal V_n .

Seja E_n o conjunto das arestas orientadas no sentido positivo de algum dos eixos, que estejam abaixo da reta V_n , ou seja, na metade inferior esquerda do quadrado B_n . Mais precisamente, o conjunto das arestas orientadas de B_n da forma

$$((x_1, x_2), (x_1, x_2) + \delta_i)$$

para $i = 1, 2$, com $x_1 + x_2 \leq n$. Para cada aresta $e \in E_n$, definimos

$$\begin{aligned} f(e) &:= \mathbb{P}(\text{processo da urna de Pólya passar por } e) \\ &= \mathbb{P}((N_t, N_{t+1}) = e \text{ para algum } t \geq 0). \end{aligned}$$

As arestas da metade superior direita do quadrado B_n não pertencem ao conjunto E_n , logo não tiveram fluxo definido, o que faz sentido uma vez que o processo N_t não chega a passar por

estas já que é interrompido quando atinge a diagonal V_n . Para estas, damos ao fluxo valores simétricos às arestas da metade inferior, pondo

$$f(e) := -f(z - e).$$

Além disso, se e é uma aresta tal que $f(e)$ ainda não tenha sido definido, colocamos

$$f(e) = -f(-e).$$

Isso define f antissimétrica para todas as arestas orientadas do grafo B_n , fazendo com que f seja um fluxo. Na verdade, f é também um fluxo unitário de a para z uma vez que satisfaz

- (i) $\operatorname{div} f(a) = 1 > 0$;
- (ii) $\operatorname{div} f(x) = 0$ para todo $x \notin \{a, z\}$.

Para o item (i) basta observar que as únicas duas arestas que saem de a são $(a, a + \delta_1)$ e $(a, a + \delta_2)$, então

$$\operatorname{div} f(a) = f((a, a + \delta_1)) + f((a, a + \delta_2)).$$

Como tanto $(a, a + \delta_1)$ quanto $(a, a + \delta_2)$ pertencem a E_n , o valor de f nestas arestas é igual a probabilidade do processo da urna de Pólya passar por tais arestas, o qual é positivo. Mais que isso, tal valor é igual a 1 uma vez que o processo começa em a , sendo obrigado a passar por uma destas arestas.

Já para o item (ii), tomemos um vértice $x \notin \{a, z\}$, o qual graças à simetria de f podemos supor sem perda de generalidade que se encontra na metade inferior esquerda do quadrado B_n , isto é, abaixo da diagonal principal V_n . Neste caso temos que

$$\begin{aligned} \operatorname{div} f(x) &= f((x, x + \delta_1)) + f((x, x + \delta_2)) + f((x, x - \delta_1)) + f((x, x - \delta_2)) \\ &= f((x, x + \delta_1)) + f((x, x + \delta_2)) - [f((x - \delta_1, x)) + f((x - \delta_2, x))] \\ &= \mathbb{P}(\text{passar por } \{(x, x + \delta_1), (x, x + \delta_2)\}) - \mathbb{P}(\text{passar por } \{(x - \delta_1, x), (x - \delta_2, x)\}) = 0, \end{aligned}$$

uma vez que passando por $\{(x - \delta_1, x), (x - \delta_2, x)\}$ vai chegar no ponto x , e como o processo da Urna de Pólya aumenta uma unidade em uma das coordenadas a cada passo, deverá também passar por $\{(x, x + \delta_1), (x, x + \delta_2)\}$ para sair de x .

Este último fato citado sobre a urna de Pólya garante que no passo k ($1 \leq k \leq n - 1$), isto é, após a adição de k bolas à urna, o processo encontrar-se-á sobre a reta V_{k+1} . E o Lema 6.2 garante que neste tempo o processo tem igual probabilidade de encontrar-se em cada um dos vértices deste conjunto, sendo ela $\frac{1}{k+1}$ visto que $|V_{k+1}| = k + 1$.

Sabendo que cada aresta tem resistência unitária, a energia do fluxo f pode ser expressa por

$$\mathcal{E}(f) = \sum_{e \in B_n} f(e)^2 r(e) = 2 \sum_{e \in E_n} f(e)^2,$$

uma vez que E_n divide ao meio o quadrado B_n . O lado direito da expressão acima é igual a

$$2 \sum_{k=2}^n \sum_{v \in V_k} \sum_{(u,v) \in E_n} f(u, v)^2$$

visto que aqui somamos sobre todas as diagonais abaixo de V_n , então somamos sobre todos os vértices nesta diagonal, e por fim, sobre todas as arestas que terminam neste vértice, o que é

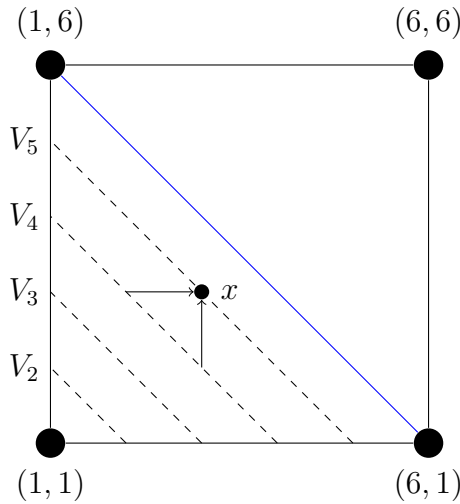


Figura 6.6: O quadrado B_n em preto, a diagonal V_n em azul, e as demais diagonais abaixo de V_n tracejadas. No tempo t o processo da urna de Pólya obrigatoriamente encontra-se sobre a diagonal V_t . Na figura v representa um vértice qualquer de uma destas diagonais, e traçamos todas as arestas que nele terminam.

a mesma coisa que somar sobre todas as arestas de E_n . Isso fica mais claro olhando para a Figura 6.6.

Mas observe que, pela definição do fluxo f e pelo fato já visto do processo N_t encontrar-se na diagonal V_k (somente e obrigatoriamente) no tempo $t = k - 1$, temos que para $v \in V_k$,

$$\begin{aligned} \sum_{(u,v) \in E_n} f(u,v)^2 &\leq \left[\sum_{(u,v) \in E_n} f(u,v) \right]^2 \\ &= \mathbb{P}(N_t = v \text{ para algum } t \geq 0)^2 \\ &= \mathbb{P}(N_{k-1} = v)^2 = \left(\frac{1}{k}\right)^2. \end{aligned}$$

Como $|V_k| = k$, ficamos com

$$\mathcal{E}(f) = 2 \sum_{k=2}^n \sum_{v \in V_k} \sum_{(u,v) \in E_n} f(u,v)^2 \leq 2 \sum_{k=2}^n k \left(\frac{1}{k}\right)^2 = 2 \sum_{k=2}^n \frac{1}{k} \leq 2 \log n.$$

Assim construímos um fluxo f unitário entre a e z tal que $\mathcal{E}(f) \leq 2 \log n$, mostrando pelo Princípio de Thomson que

$$\mathcal{R}(a \leftrightarrow z) \leq \mathcal{E}(f) \leq 2 \log n.$$

□

6.4.1 Cotas inferiores para o caso multidimensional

Fixamos dois nós $x, y \in \mathbb{Z}_n^d$ com distância $k \geq 1$. Para cada $j \geq 0$ consideramos Π_j o conjunto de arestas que conectam os vértices do cubo centrado em x de lado $2j$ com seu complemento, como no exemplo da Figura 6.7. Este conjunto encapsula o vértice x , formando um conjunto separador de x e qualquer nó que esteja fora do cubo. Além disso, para $1 \leq j < k/d$, o cubo C_j não chega até o vértice y , sendo Π_j um conjunto separador entre x e y .

Para estimar a quantidade de arestas no conjunto Π_j observe que tal valor depende essencialmente da quantidade de nós do hipercubo C_j e da dimensão d . Isso fica claro na Figura 6.8,

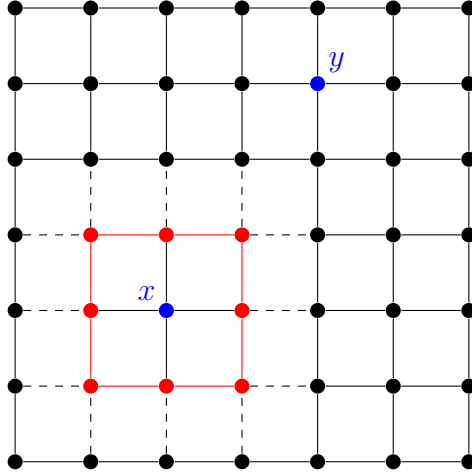


Figura 6.7: Representação do toro \mathbb{Z}_7^2 . Em vermelho o cubo centrado em x de lado 2. As arestas tracejadas formam o conjunto Π_1 .

onde vemos que para $d = 2$, cada nó sobre as arestas do hipercubo possui no máximo duas arestas que o conectam com o complementar do cubo; já para $d = 3$, cada nó sobre as faces do hipercubo possui no máximo 3 arestas que o conectam com o complementar do hipercubo. De modo geral, a dimensão é uma cota superior para este valor.



Figura 6.8: Com $d = 2$ e $d = 3$, para cada possível vértice no cubo centrado em x , a quantidade de arestas que o conectam com o complemento deste cubo.

Além disso, existem $(2j + 1)^{d-1}$ nós em cada uma das hiperfaces do cubo d -dimensional, ao mesmo tempo em que o número de faces é uma função de d que chamaremos $f(d)$. Deste modo,

$$|\Pi_j| \leq d \cdot f(d) \cdot (2j + 1)^{d-1} \leq d \cdot f(d) \cdot 2^{d-1} \cdot j^{d-1} = \tilde{c}_d \cdot j^{d-1}.$$

A Proposição 3.29 nos garante então que

$$\begin{aligned} \mathcal{R}(x \leftrightarrow y) &\geq \sum_{1 \leq j < k/d} \left(\sum_{e \in \Pi_j} c(e) \right)^{-1} \\ &= \sum_{1 \leq j < k/d} (|\Pi_j|)^{-1} \\ &\geq \frac{1}{\tilde{c}_d} \sum_{1 \leq j < k/d} \frac{1}{j^{d-1}} \end{aligned} \quad (6.8)$$

No caso $d = 2$, as equações (6.5) e (6.8) nos garantem que

$$\mathbb{E}_x \tau_y = dn^d \mathcal{R}(x \leftrightarrow y) \geq \frac{dn^d}{\tilde{c}_d} \sum_{1 \leq j < k/d} \frac{1}{j^{d-1}} = \frac{2n^2}{\tilde{c}_2} \sum_{1 \leq j < k/d} \frac{1}{j} \geq c_2 n^2 \log(k)$$

para uma constante c_2 devidamente escolhida, obtendo assim a cota inferior proposta em (6.7).

No caso $d \geq 3$, podemos limitar inferiormente a série do lado direito de (6.8) pelo primeiro termo, obtendo que $\mathcal{R}(x \leftrightarrow y) \geq \tilde{c}_d^{-1}$. Então segue da equação (6.5) que com $c_d = d\tilde{c}_d^{-1}$

$$\mathbb{E}_x \tau_y = dn^d \mathcal{R}(x \leftrightarrow y) \geq dn^d \tilde{c}_d^{-1} = c_d n^d,$$

mostrando a cota inferior proposta em (6.6).

6.4.2 Cotas superiores para o caso multidimensional com $d \geq 3$

Considere $d \geq 3$. Iniciaremos analisando um caso bem específico, que nada mais é do que uma generalização do Lema 6.9. Então com base nele estenderemos os resultados para o caso geral. Para isso precisaremos do seguinte lema de contagem.

Lema 6.10.

$$\sum_{j=1}^{\infty} \binom{k+j}{k}^{-1} = \frac{1}{k-1}$$

Demonstração. Com algum algebrismo, temos que

$$\sum_{j=1}^{\infty} \binom{k+j}{k}^{-1} = \sum_{j=1}^{\infty} \frac{k!j!}{(k+j)!} = \sum_{j=1}^{\infty} \frac{k!}{(j+1)\dots(j+k)} = k \cdot (k-2)! \sum_{j=1}^{\infty} \frac{k-1}{(j+1)\dots(j+k)}.$$

Porém observe que o termo geral da série acima pode ser reescrito como

$$\frac{k-1}{(j+1)\dots(j+k)} = \frac{1}{(j+1)\dots(j+k-1)} - \frac{1}{(j+2)\dots(j+k)},$$

de modo que temos na verdade uma série telescópica, cuja soma é

$$\begin{aligned} & \lim_{N \rightarrow \infty} \sum_{j=1}^N \left[\frac{1}{(j+1)\dots(j+k-1)} - \frac{1}{(j+2)\dots(j+k)} \right] \\ &= \lim_{N \rightarrow \infty} \left[\frac{1}{2 \cdot 3 \cdot \dots \cdot k} - \frac{1}{(N+2)\dots(N+k)} \right] = \frac{1}{k!}. \end{aligned}$$

Portanto, ficamos com

$$\sum_{j=1}^{\infty} \binom{k+j}{k}^{-1} = k \cdot (k-2)! \cdot \frac{1}{k!} = \frac{1}{k-1}.$$

□

Voltando à obtenção da cota, começamos com o caso em que $x, y \in \mathbb{Z}_n^d$ são pontos antipodais de um hipercubo de lado k , com k par, e dimensão d . Graças à simetria do toro, podemos supor sem perda de generalidade que

$$\begin{aligned} x &= (1, 1, \dots, 1) = \mathbf{1} \\ y &= (k+1, k+1, \dots, k+1) = (k+1)\mathbf{1}. \end{aligned}$$

Vamos supor também que $k < n/d$, de modo que o hipercubo não possa envolver o toro. Isso faz com que o caminho mais curto entre x e y seja “pelo lado de dentro” do hipercubo, garantindo que a distância entre $x = \mathbf{1}$ e $y = (k+1)\mathbf{1}$ seja dada por $d \cdot k < n$.

Para cada j considere o hiperplano

$$V_j := \left\{ (x_1, \dots, x_d) \in \mathbb{Z}^d; \sum x_i = j + d \right\}.$$

Rodamos o processo da Urna de Pólya $(N_t)_{t \geq 0}$ (que começa em $\mathbf{1}$ por definição) até que ele atinja o hiperplano

$$V_{kd/2} = \left\{ (x_1, \dots, x_d) \in \mathbb{Z}^d; \sum x_i = \frac{(k+2)d}{2} \right\}.$$

Este processo inicia em $\mathbf{1}$, e é tal que $\sum_{i=1}^d N_i$ aumenta exatamente uma unidade a cada passo, assim no tempo j o processo $(N_t)_{t \geq 0}$ encontrar-se-á obrigatoriamente em algum vértice do hiperplano V_j .

Seja \vec{E}_k o conjunto das arestas orientadas no sentido positivo de algum dos eixos que estão dentro do hiperplano $\{1, \dots, k+1\}^d$ e abaixo do hiperplano $V_{kd/2}$, isto é, das arestas da forma

$$((x_1, \dots, x_i, \dots, x_d) \rightarrow (x_1, \dots, x_i + 1, \dots, x_d)) = (x, x + \delta_i)$$

para algum $i \in \{1, \dots, d\}$, com $1 \leq x_j \leq k+1$ para todo j , com $x_i \leq k$, e com $1 + \sum x_j \leq \frac{(k+2)d}{2}$. Para uma aresta orientada $\vec{e} \in \vec{E}_k$, definimos o fluxo f como sendo a probabilidade do processo da urna de Pólya passar por esta aresta em algum momento, mais precisamente,

$$f(e) := \mathbb{P} \left[((N_t^1, \dots, N_t^d), (N_{t+1}^1, \dots, N_{t+1}^d)) = e, \text{ para algum } t \geq 0 \right].$$

E para uma aresta orientada (a, b) tal que $(b, a) \in \vec{E}_k$, definimos

$$f((a, b)) = -f((b, a)).$$

Isso define f para todas as arestas orientadas na metade do hiperplano $\{1, \dots, k+1\}^d$ abaixo do hiperplano $V_{kd/2}$. Completamos a definição de f dando valores simétricos para as arestas da outra metade deste hiperplano, isto é, se \vec{e} está na outra metade do hiperplano, definimos

$$f(e) := -f((k+1)\mathbf{1} - e).$$

Tal fluxo f acima definido é um fluxo unitário de $\mathbf{1}$ para $(k+1)\mathbf{1}$. A demonstração deste fato se assemelha com o resultado análogo apresentado no Lema 6.9.

Seja $v \in V_j$ para $1 \leq j \leq kd/2$, isto é, um hiperplano abaixo da diagonal. O processo N_t visita v se, e somente se, viaja por uma das arestas orientadas que terminam em v , então, pela definição do fluxo f , vale

$$\sum_{u: (u,v) \in \vec{E}_k} f(u, v)^2 \leq \left[\sum_{u: (u,v) \in \vec{E}_k} f(u, v) \right]^2 = \mathbb{P} [N_t = v \text{ para algum } t \geq 0]^2.$$

Pela definição do processo da urna de Pólya, $N_t \in V_t$ apenas no tempo $t = j$. Além disso, o Lema 6.3 garante-nos que o processo $(N_t)_{t \geq 0}$ tem igual probabilidade de passar por cada um dos $\binom{j+d-1}{d-1}$ vértices do hiperplano V_j , de modo que o lado direito é igual a

$$\mathbb{P} [N_j = v]^2 = \left(\frac{j+d-1}{d-1} \right)^{-2},$$

mostrando que para cada $v \in V_j$ ($1 \leq j \leq kd/2$) vale

$$\sum_{u: (u,v) \in \vec{E}_k} f(u, v)^2 \leq \left(\frac{j+d-1}{d-1} \right)^{-2}. \quad (6.9)$$

Vamos utilizar a relação acima para calcular a energia total do fluxo f . Começamos observando que a energia é calculada sobre todas as arestas do hipercubo com

$$\mathcal{E}(f) = \sum_{e \in \{1, \dots, k+1\}^d} f(e)^2 r(e),$$

que é equivalente a somarmos sobre todas as arestas orientadas no sentido positivo de alguma coordenada, que por sua vez, graças à simetria, é igual a duas vezes a soma sobre todas as arestas em \vec{E}_k , isto é,

$$\mathcal{E}(f) = \sum_{e \in \{1, \dots, k+1\}^d} f(e)^2 r(e) = 2 \sum_{e \in \vec{E}_k} f(e)^2,$$

onde na última igualdade ainda utilizamos que todas as arestas possuem resistência unitária no passeio aleatório simples no toro.

Do mesmo modo que fizemos na Proposição 6.9, somar sobre todas as arestas em \vec{E}_k é a mesma coisa que pegar todos os hiperplanos abaixo de $V_{kd/2}$, pegar todos os seus vértices, e então somar sobre todas as arestas de \vec{E}_k que terminam neste vértice, isto é,

$$\mathcal{E}(f) = 2 \sum_{e \in \vec{E}_k} f(e)^2 = 2 \sum_{j=1}^{kd/2} \sum_{v \in V_j} \sum_{(u,v) \in \vec{E}_k} f(u,v)^2 \leq 2 \sum_{j=1}^{kd/2} \sum_{v \in V_j} \binom{j+d-1}{d-1}^{-2},$$

onde a última desigualdade é simplesmente a utilização de (6.9). Como o hiperplano V_j possui $\binom{j+d-1}{d-1}$ vértices, a desigualdade acima se reescreve como

$$\mathcal{E}(f) \leq 2 \sum_{j=1}^{kd/2} \binom{j+d-1}{d-1}^{-1}.$$

O somatório no direito da desigualdade acima é limitado por

$$\sum_{j=1}^{\infty} \binom{j+d-1}{d-1}^{-1}$$

que tomando $k = d - 1$ e aplicando o Lema 6.10, mostra-se igual a

$$\sum_{j=1}^{\infty} \binom{j+k}{k}^{-1} = \frac{1}{k-1} = \frac{1}{d-2},$$

donde concluímos que

$$\mathcal{E}(f) \leq \frac{2}{d-2}.$$

O princípio de Thomson (Teorema 3.23) nos garante então que

$$\mathcal{R}(\mathbf{1} \leftrightarrow (k+1)\mathbf{1}) \leq \mathcal{E}(f) \leq \frac{2}{d-2}. \quad (6.10)$$

Com isso valendo para qualquer $k < n/d$ par.

Tendo a equação (6.10), podemos partir para o caso geral e obter as cotas para a resistência efetiva entre dois vértices x e y quaisquer no toro \mathbb{Z}_n^d .

Começamos com o caso em que x e y diferem por um inteiro da forma $4k$ em uma única coordenada, com $2k < n/d$. Podemos supor sem perda de generalidade que $x = \mathbf{1}$ e $y =$

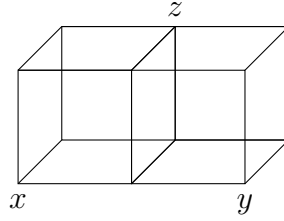


Figura 6.9: x e y diferem em uma única coordenada por $4k$. Consideramos o nó z de modo a podermos utilizar a cota em (6.10).

$\mathbf{1} + (4k)\delta_1 = (4k + 1, 1, \dots, 1)$. Tomamos $z = (2k + 1)\mathbf{1}$ como na Figura 6.9 e temos graças à simetria que $\mathcal{R}(x \leftrightarrow z) = \mathcal{R}(z \leftrightarrow y)$, com ambos limitados pela cota vista em (6.10). A desigualdade triangular para resistências efetivas (Corolário 4.7) garante então que

$$\mathcal{R}(x \leftrightarrow y) \leq \mathcal{R}(x \leftrightarrow z) + \mathcal{R}(z \leftrightarrow y) \leq \frac{4}{d-2}. \quad (6.11)$$

Para o caso em que x e y diferem em uma única coordenada por um outro inteiro da forma $4k - 1, 4k + 1$ ou $4k + 2$, mas ainda com $2k < n/d$, a desigualdade triangular juntamente com a cota anterior garantem que

$$\mathcal{R}(x \leftrightarrow y) \leq \frac{4}{d-2} + 2. \quad (6.12)$$

Agora podemos considerar o caso em que x e y diferem em uma única coordenada uma distância h qualquer. Sem perda de generalidade, $x = \mathbf{1}$ e $y = \mathbf{1} + h\delta_1$. Pela simetria do toro, sabemos que $h < n/2$. Fixemos $\lambda = 2 \cdot \lfloor n/d \rfloor$, e tomemos a sequência de pontos $w_0 = x$, $w_1 = x + \lambda\delta_1$, até $w_m = x + m\lambda\delta_1$, onde m é o maior inteiro tal que $m\lambda < h$. Por simplicidade ainda escrevemos $w_{m+1} = y$. Isso faz com que os pares (w_i, w_{i+1}) difiram em uma única coordenada uma distância λ tal que $\lambda/2 < n/d$, caso em que sabemos pela argumentação anterior que a resistência efetiva entre estes pares de pontos é limitada por $\frac{4}{d-2} + 2$.

Como $h < n/2$, $\lambda = 2\lfloor n/d \rfloor \geq \frac{n}{d}$ e devemos ter $(m+1)\lambda \geq h$, basta que $m+1 = \lceil d/2 \rceil \leq d$. Então pela desigualdade triangular

$$\mathcal{R}(x \leftrightarrow y) \leq \mathcal{R}(w_0 \leftrightarrow w_1) + \dots + \mathcal{R}(w_m \leftrightarrow w_{m+1}) \leq d \left(\frac{4}{d-2} + 2 \right).$$

Finalmente chegamos no caso geral, se x e y são pontos arbitrários, existem vértices $\{z_j\}_{j=0}^d$ com $z_0 = x$ e $z_d = y$, de modo que cada par (z_{j-1}, z_j) difere apenas na j -ésima coordenada. Assim, pela desigualdade triangular,

$$\mathcal{R}(x \leftrightarrow y) \leq \mathcal{R}(x \leftrightarrow z_1) + \mathcal{R}(z_1 \leftrightarrow z_2) + \dots + \mathcal{R}(z_{d-1} \leftrightarrow y) \leq d^2 \left(\frac{4}{d-2} + 2 \right).$$

Por fim, utilizando a equação (6.5), temos que para $a, b \in \mathbb{Z}_n^d$ quaisquer,

$$\mathbb{E}_a \tau_b = dn^d \mathcal{R}(a \leftrightarrow b) \leq d^3 n^d \left(\frac{4}{d-2} + 2 \right) = C_d n^d.$$

6.4.3 Cotas superiores para o caso multidimensional com $d = 2$

Fixamos os vértices x, y no toro bi-dimensional \mathbb{Z}_n^2 . Se eles são dois vértices diagonalmente opostos de um quadrado de lado l , estão a uma distância ℓ_1 de $k = 2l$. Podemos supor sem

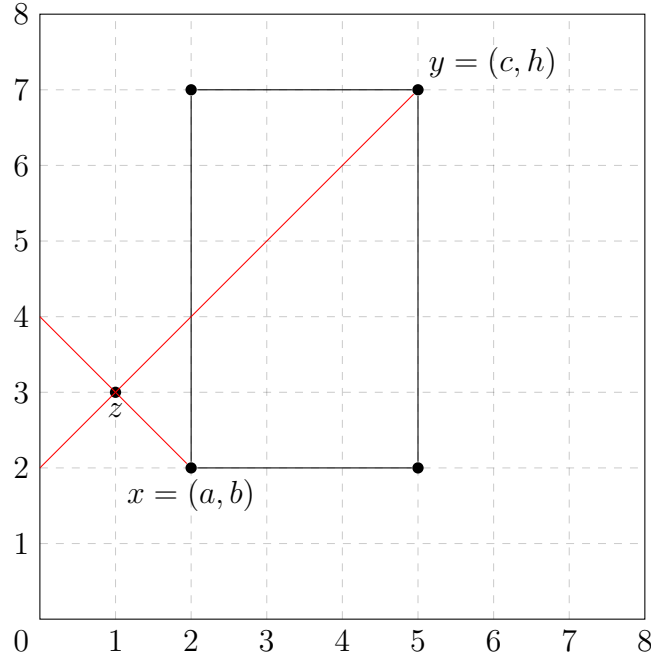


Figura 6.10: Se x e y são vértices de um retângulo não quadrado, consideramos o vértice auxiliar z , e para ele aplicamos os resultados já conhecidos.

perda de generalidade que $x = (1, 1)$ e $y = (l+1, l+1)$, e a cota superior obtida na Proposição 6.9 nos garante que

$$\mathcal{R}(x \leftrightarrow y) \leq 2 \log(l+1) = 2 \log\left(\frac{k+2}{2}\right).$$

Obtemos então, da equação (6.5), que

$$\mathbb{E}_x \tau_y = dn^d \mathcal{R}(x \leftrightarrow y) \leq 2n^2 \cdot 2 \log\left(\frac{k+2}{2}\right) \leq 4n^2 \log(k+1).$$

Para o caso geral, sejam x e y vértices de um retângulo não quadrado como na Figura 6.10. Escrevemos $x = (a, b)$ e $y = (c, h)$. Podemos graças a simetria do toro assumir que $a \leq c$, $b \leq h$ e $c - a \leq h - b$, isto é, que x está “abaixo” e “a esquerda” de y , e que a altura do retângulo é maior que sua base. Neste caso, a distância k entre x e y é

$$k = |c - a| + |h - b| \tag{6.13}$$

Começamos olhando o caso em que $c - a$ e $h - b$ possuem a mesma paridade. Neste caso, a reta com inclinação -1 passando por x , e a reta com inclinação $+1$ passando por y , encontram-se no ponto

$$z = \left(\frac{(a+c) + (b-h)}{2}, \frac{(a-c) + (b+h)}{2} \right).$$

Vamos trabalhar separadamente com os pares (x, z) e (z, y) . Note que x e z são vértices opostos de um quadrado de lado

$$l = x^1 - z^1 = a - \frac{(a+c) + (b-h)}{2} = \frac{2a - a - c - b + h}{2} = \frac{(a-c) + (h-b)}{2},$$

que satisfaz $l \leq k/2$ haja vista a relação (6.13). Assim, do mesmo modo que fizemos anteriormente, a proposição 6.9 nos garante que

$$\mathcal{R}(x \leftrightarrow z) \leq 2 \log(l+1) \leq 2 \log(k+1).$$

Analogamente, observamos que os pontos z e y são vértices opostos de um quadrado de lado

$$l = y^1 - z^1 = h - \frac{(a - c) + (b + h)}{2} = \frac{2h - a + c - b - h}{2} = \frac{(c - a) + (h - b)}{2},$$

satisfazendo também $l \leq k/2$ em razão de (6.13), e pelo mesmo argumento de antes, a Proposição 6.9 garante-nos que

$$\mathcal{R}(y \leftrightarrow z) \leq 2 \log(l + 1) \leq 2 \log(k + 1).$$

Por fim a desigualdade triangular para resistências (Corolário 4.7) nos dá que

$$\mathcal{R}(x \leftrightarrow y) \leq \mathcal{R}(x \leftrightarrow z) + \mathcal{R}(z \leftrightarrow y) \leq 4 \log(k + 1),$$

donde obtemos via equação (6.5) que

$$\mathbb{E}_a \tau_b = dn^d \mathcal{R}(x \leftrightarrow y) \leq 2n^2 \cdot 4 \log(k + 1).$$

Agora para o caso em que $(c - a)$ e $(h - b)$ possuem paridades distintas, consideramos um ponto auxiliar x' que esteja a uma distância unitária de x e mais próximo de y . Para x' vale a construção acima, e a desigualdade triangular nos garante que

$$\mathcal{R}(x \leftrightarrow y) \leq \mathcal{R}(x \leftrightarrow x') + \mathcal{R}(x' \leftrightarrow y) \leq 1 + 4 \log(k + 1) \leq 6 \log(k + 1),$$

donde obtemos pela equação (6.5) que

$$\mathbb{E}_a \tau_b = dn^d \mathcal{R}(x \leftrightarrow y) \leq 2n^2 \cdot 6 \log(k + 1).$$

Analisamos assim todos os casos possíveis, e um ajuste adequado das constantes nos dá como resultado a equação (6.7).

Apêndice A

Pequeno Resumo de Alguns Resultados

Para facilitar a consulta, neste apêndice listamos, de forma bastante resumida, alguns dos principais resultados apresentados no texto, referentes aos indicadores t_{mix} , t_{hit} e t_{rel} . Apresentamos desigualdades que relacionam esses indicadores entre si em contextos gerais e cotas específicas obtidas nos exemplos do passeio aleatório no ciclo \mathbb{Z}_n e no toro \mathbb{Z}_n^d .

Relação entre t_{mix} e t_{hit}

Ver Seção 5.1. Supondo a cadeia *lazy* e reversível,

$$\begin{aligned} t_{\text{mix}}^{(\infty)} &\leq 4 \max_{x \in S} \mathbb{E}_\pi(\tau_x) + 1 \\ \implies t_{\text{mix}} &\leq 2 \max_{x \in S} \mathbb{E}_\pi(\tau_x) + 1 \\ \implies t_{\text{mix}} &\leq 2t_{\text{hit}} + 1 \end{aligned}$$

Relação entre t_{mix} e t_{rel}

Ver Seção 5.2

$$\begin{aligned} t_{\text{mix}}(\varepsilon) &\leq \log\left(\frac{1}{\varepsilon \pi_{\min}}\right) t_{\text{rel}} \\ t_{\text{mix}}(\varepsilon) &\geq (t_{\text{rel}} - 1) \log\left(\frac{1}{2\varepsilon}\right) \end{aligned}$$

Relação entre t_{hit} e t_{rel}

Ver Seção 5.4.

Para o passeio aleatório *lazy* em um grafo, ver Teorema 5.15,

$$t_{\text{hit}} \leq \frac{20d_{\text{avg}}}{d_{\text{min}}} n \sqrt{t_{\text{rel}} + 1}$$

Em um contexto mais geral, supondo reversibilidade e espectro não negativo, ver Proposição 5.19

$$\begin{aligned} t_{\text{hit}} &\leq 2 \max_x \left(\frac{1 - \pi(x)}{\pi(x)} \right) t_{\text{rel}} \\ P^t(x, x) - \pi(x) &\leq (1 - t_{\text{rel}}^{-1})^t (1 - \pi(x)). \end{aligned}$$

t_{mix} para o passeio aleatório lazy no ciclo \mathbb{Z}_n

Ver Seção 6.2.1

$$\frac{n^2}{32} \leq t_{\text{mix}} \leq n^2$$

 t_{mix} para o passeio aleatório lazy no toro \mathbb{Z}_n^d

Ver Seção 6.2.2

$$\frac{dn^2}{32} \leq t_{\text{mix}} \leq d^2n^2$$

 t_{hit} para o passeio aleatório lazy no ciclo \mathbb{Z}_n

Ver Seção 6.4

$$t_{\text{hit}} = \left\lfloor \frac{n^2}{4} \right\rfloor$$

 t_{hit} para o passeio aleatório no toro \mathbb{Z}_n^d

Ver Seção 6.4

$$\begin{aligned} c_2n^2 \log n \leq t_{\text{hit}} \leq C_2n^2 \log n & \quad \text{se } d = 2 \\ c_dn^d \leq t_{\text{hit}} \leq C_dn^d & \quad \text{se } d \geq 3 \end{aligned}$$

No que segue, lembre que a notação $f(n) = O(g(n))$ significa que existe uma constante C tal que $|f(n)| \leq C|g(n)|$ para todo n .

Os resultados a seguir, referentes ao espectro e ao tempo de relaxamento do passeio aleatório no n -ciclo e no toro, são obtidos na Seção 6.3.

Buraco espectral do passeio aleatório (não lazy) no ciclo \mathbb{Z}_n

O buraco espectral é $\gamma = 1 - \lambda_2 = 1 - \cos(2\pi/n) = \frac{2\pi^2}{n^2} + O(n^{-4})$.

Quando n é ímpar o buraco espectral absoluto é $\gamma_\star = 1 - \cos(\pi/n) = \frac{\pi^2}{n^2} + O(n^{-4})$.

Quando n é par o buraco espectral absoluto é nulo, $\gamma_\star = 0$.

Buraco espectral do passeio aleatório lazy no ciclo \mathbb{Z}_n

$$\gamma = \gamma_\star = \frac{1 - \cos(2\pi/n)}{2} = \frac{\pi^2}{n^2} + O(n^{-4})$$

Tempo de relaxamento para o passeio aleatório lazy no ciclo \mathbb{Z}_n

$$cn^2 \leq t_{\text{rel}} \leq Cn^2$$

Mais precisamente:

$$t_{\text{rel}} = \frac{1}{\gamma_\star} = \frac{n^2}{\pi^2} (1 + O(n^{-2}))$$

Buraco espectral do passeio aleatório lazy no toro \mathbb{Z}_n^d

$$\gamma = \gamma_\star = \frac{1 - \cos(2\pi/n)}{2d} = \frac{\pi^2}{dn^2} + O(n^{-4})$$

Tempo de relaxamento para o passeio aleatório lazy no toro \mathbb{Z}_n^d

$$c(d)n^2 \leq t_{\text{rel}} \leq C(d)n^2$$

Mais precisamente:

$$t_{\text{rel}} = \frac{1}{\gamma_\star} = \frac{dn^2}{\pi^2}(1 + O(n^{-2})).$$

Bibliografia

- [1] David Aldous e James Allen Fill. *Reversible Markov Chains and Random Walks on Graphs*. Unfinished monograph, recompiled 2014, available at <http://www.stat.berkeley.edu/~aldous/RWG/book.html>. 2002.
- [2] Gely P. Basharin, Amy N. Langville e Valeriy A. Naumov. “The life and work of A.A. Markov”. Em: *Linear Algebra and its Applications* 386 (2004). Special Issue on the Conference on the Numerical Solution of Markov Chains 2003, pp. 3–26. ISSN: 0024-3795. DOI: <https://doi.org/10.1016/j.laa.2003.12.041>. URL: <https://www.sciencedirect.com/science/article/pii/S0024379504000357>.
- [3] Johel Beltrán. *Cadenas de Markov y Teoria de Potencial*. Publicações matemáticas. IMPA, 2011. ISBN: 9788524403163. URL: https://impa.br/wp-content/uploads/2017/04/28CBM_01.pdf.
- [4] Hamilton Prado Bueno. *Álgebra Linear: Um segundo curso*. SBM, 2006.
- [5] Jearl Walker David Halliday Robert Resnick. *Fundamentos de Física Vol. 3: Eletromagnetismo*. 9^a. Vol. 3. LTC, 2012. ISBN: 978-85-216-1905-5.
- [6] Peter G. Doyle e J. Laurie Snell. *Random Walks and Electric Networks*. 2000. arXiv: [math/0001057](https://arxiv.org/abs/math/0001057) [math.PR].
- [7] Rick Durrett. *Probability: Theory and Examples*. 5^a ed. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2019. DOI: [10.1017/9781108591034](https://doi.org/10.1017/9781108591034).
- [8] Alexandre Gaudillière. *Condenser physics applied to Markov chains - A brief introduction to potential theory*. 2009. URL: <https://arxiv.org/abs/0901.3053>.
- [9] David A. Levin e Yuval Peres. *Markov chains and mixing times (2nd edition)*. American Mathematical Society, 2017.
- [10] Elon Lages Lima. *Álgebra Linear*. 10^a ed. IMPA, 2020.
- [11] Russell Lyons e Yuval Peres. *Probability on Trees and Networks*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2017. DOI: [10.1017/9781316672815](https://doi.org/10.1017/9781316672815).
- [12] James R. Norris. *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1997. DOI: [10.1017/CB09780511810633](https://doi.org/10.1017/CB09780511810633).
- [13] Roberto I. Oliveira e Yuval Peres. “Random walks on graphs: new bounds on hitting, meeting, coalescing and returning”. Em: *2019 Proceedings of the Sixteenth Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*. SIAM. 2019, pp. 119–126.