

**UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS - ICE_x
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA**

Gabriela Araujo Ramalho

**TEMPOS ESTACIONÁRIOS FORTES, ACOPLAMENTOS, ANÁLISE DE
FOURIER E PROPRIEDADES DE MISTURA DE CADEIAS DE
MARKOV**

Belo Horizonte
2022

Gabriela Araujo Ramalho

**TEMPOS ESTACIONÁRIOS FORTES, ACOPLAMENTOS, ANÁLISE DE
FOURIER E PROPRIEDADES DE MISTURA DE CADEIAS DE
MARKOV**

Versão final

Dissertação submetida à banca examinadora, designada pelo Programa de Pós-Graduação em Matemática da UFMG, como requisito parcial para a obtenção do título de mestre em Matemática.

Orientador: Marcelo Richard Hilário

Co-orientador: Charles Almeida

Belo Horizonte
2022

Ramalho, Gabriela Araujo

R165t Tempos estacionários fortes, acoplamentos, análise de fourier e propriedades de mistura de cadeias de Markov / Gabriela Araujo Ramalho. – 2022.
79 f. il.

Orientador: Marcelo Richard Hilário.

Coorientador: Charles Aparecido de Almeida.

Dissertação (mestrado) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Matemática.

Referências: f. 78-79.

1. Matemática – Teses. 2. Markov, Processos de – Teses. 3. Passeio aleatório (Matemática) – Teses. 4. Fourier, Análise de I. Hilário, Marcelo Richard. II. Charles Aparecido de Almeida. III. Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Matemática. IV. Título.

CDU 51 (043)



FOLHA DE APROVAÇÃO


Tempos Estacionários Fortes, Acoplamentos, Análise de Fourier e Propriedades de Mistura de Cadeias de Markov

GABRIELA ARAÚJO RAMALHO

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:




Prof. Marcelo Richard Hilário
UFMG



Prof. Charles Aparecido de Almeida
UFMG



Prof. Bernardo Nunes Borges de Lima
UFMG



Prof. Lucas Henrique Calixto
UFMG



Prof. Marco Vinicius Bahi Aymone
UFMG

Belo Horizonte, 16 de março de 2022.

Agradecimentos

Venho agradecer a todos aqueles que de alguma forma contribuíram para o desenvolvimento desta pesquisa e para que tudo ocorresse da melhor forma possível.

Agradeço ao meu orientador, Professor Marcelo Richard Hilário, por todos os ensinamentos, paciência e disposição para me acompanhar além do que imaginava conseguir. Obrigada, em especial, por acreditar em mim, e por todas descontrações e conversas que foram essenciais, principalmente nos dias mais difíceis. Serei eternamente grata.

Agradeço ao Professor Charles Almeida, pela disponibilidade e pelas essenciais contribuições para a conclusão desse trabalho.

Agradeço ao Professor Bernardo Nunes Borges de Lima, pelas primeiras (e melhores) aulas que me fizeram encantar pela probabilidade e pela orientação e ensinamentos no período da Graduação.

Agradeço à minha família, pelo apoio e esforço incondicional.

Agradeço aos meus queridos amigos e familiares, pelo carinho, pelas boas conversas e por tornarem esses dias mais prazerosos. Vocês sabem o quanto são especiais para mim.

Por fim, agradeço a Deus por ter me dado força e paciência para concluir mais essa fase da minha vida e por ter colocado cada uma dessas pessoas nela, sem elas eu não teria chegado até aqui.

Resumo

Os objetos centrais de estudo nesta dissertação são as cadeias de Markov em espaços de estado finito. O objetivo principal é apresentar técnicas e ferramentas que permitam a obtenção de cotas superiores para taxa de convergência de uma tal cadeia para a sua distribuição estacionária (ou distribuição de equilíbrio). Discutiremos aqui três métodos: os tempos estacionários fortes, os acoplamentos e a análise de Fourier. Durante o percurso, manteremos em mente o exemplo do passeio aleatório no hipercubo para ilustrar o funcionamento das técnicas. Em alguns pontos apresentaremos resultados específicos para este exemplo. As técnicas, no entanto, são aplicáveis em um contexto muito mais amplo para tratar a convergência de cadeias com propriedades de simetria, como os passeios aleatórios em grupos. Além do passeio aleatório no hipercubo, outros exemplos de tais cadeias de Markov aparecem, por exemplo, na modelagem do embaralhamento de cartas que pode ser visto como um passeio aleatório no grupo simétrico.

Palavras-Chaves : cadeias de Markov, tempo estacionário forte, acoplamento, análise de Fourier, passeio aleatório.

Abstract

The central object of study in this dissertation are the Markov chains in finite state spaces. The objective is present techniques and tools that allow the obtention of upper bounds for the convergence rate of such a chain for its stationary distribution (or equilibrium distribution). We will discuss three here: strong stationary times, couplings and Fourier analysis. During the course, we will keep in mind the example of the random walk on the hypercube to illustrate how the techniques work. In some points we will present details for this example. The techniques, however, are implemented in a context of converging chains with symmetry properties, such as random walks in groups. In addition to the random walk in the hypercube, other examples of Markov modeling chains appear, for example, on the shuffling of cards which can be seen as a random walk in symmetric group.

Key Words : Markov chains, strong stationary time, coupling, Fourier analysis, random walk.

Índice

Introdução	9
1 Cadeias de Markov e Convergência	11
1.1 Primeiros Resultados	11
1.2 Convergência para o Equilíbrio	16
2 Cotas superiores para convergência	24
2.1 Tempo Estacionário Forte	24
2.2 Passeio Aleatório no Hipercubo: Tempo Estacionário Forte	27
3 Acoplamento	34
3.1 Acoplamento e Distância de Variação Total	34
3.2 Passeio Aleatório no Hipercubo: Acoplamento	38
4 Propriedades Simétricas	44
4.1 Definições e Consequências Elementares	44
4.2 Consequências das Condições Simétricas	45
4.3 Passeios Aleatórios em Grupos	47
5 Representações de Grupos	59
5.1 Definições Básicas e Resultados da Teoria da Representação	59
5.2 Análise Harmônica no Hipercubo	74
6 Conclusões	77
Referências Bibliográficas	78

Introdução

As cadeias de Markov foram introduzidas pela primeira vez em 1906 pelo matemático russo Andrey Markov [5, Seção 4.2, pág. 15] e são objetos importantes de estudo em probabilidade e nas suas diversas aplicações. Aparecem também em conexão com a física, biologia e outras áreas do conhecimento.

Mais recentemente, matemáticos como David Aldous e Persi Diaconis estudaram intensamente as propriedades da convergência para o equilíbrio de cadeias de Markov como o embaralhamento de cartas [2] e, em geral, passeio aleatórias em grupos. Tanto os métodos espectrais quanto as técnicas probabilísticas, como o acoplamento, desempenharam papéis importantes.

Informalmente, uma cadeia de Markov em um espaço de estados finito Ω é um processo estocástico em que, a cada unidade de tempo t um estado $X_t \in \Omega$ da cadeia é escolhido de acordo com uma distribuição de probabilidade de forma que, condicionado no estado presente, todo o passado é independente do futuro. Em outras palavras, somente o estado atual é relevante para a previsão do estado seguinte. Partindo de uma distribuição inicial para o primeiro estado da cadeia, podemos nos perguntar como evolui a distribuição da posição da cadeia com o passar do tempo.

Uma das linhas de pesquisa importantes na teoria das cadeias de Markov visa quantificar tal convergência, estimando quanto tempo temos que esperar para chegar a uma boa aproximação da medida equilíbrio. Chamamos esse tempo de tempo de mistura da cadeia.

Nesse contexto, David Aldous e Persi Diaconis [10], introduziram pela primeira vez a noção de *cutoff* para descrever a transição abrupta da convergência para estacionariedade. Isso significa que, quando o tamanho do espaço de estados (por exemplo, no processo de embaralhamento de cartas é o número de cartas no baralho) é muito grande, a distância para a distribuição estacionária vai de quase 1 (máximo) a quase 0 em um intervalo em torno do tempo de mistura que é relativamente muito pequeno em relação ao próprio tempo de mistura. Esse comportamento foi observado para uma ampla classe de exemplos naturais. No entanto, não é simples provar com rigor a existência desse rápido decaimento.

Nesta dissertação o enfoque são nas cadeias de Markov com boas propriedades de simetria (que permitem tratar todos os estados da cadeia da mesma forma) como, por exemplo, os passeios aleatórios em grupos. Pretende-se estudar técnicas para obter estimativas para taxa de convergência para a distribuição estacionária em tais cadeias de Markov. Para isso, seguiremos como base principal o artigo [3] D. Aldous e Persi Diaconis.

No Capítulo 1, vamos introduzir o conceito de cadeias de Markov e algumas propriedades básicas. Em seguida, definiremos formalmente o que queremos dizer com estar "próximo ao equilíbrio". Para tanto, introduziremos duas noções úteis de distância entre

distribuições: a distância de separação $s_x(n)$ e a de variação total $d(n)$.

Existem três técnicas disponíveis para obter limites superiores explícitos para $s_x(n)$ e $d(n)$, quando n é finito e grande:

- (1) Tempo estacionário forte
- (2) Acoplamento
- (3) Análise de Fourier

Essas técnicas serão apresentadas nos Capítulos 2, 3 e 5 respectivamente. Elas são fundamentais para a área e suas aplicações práticas têm sido amplamente discutidas na literatura. Aqui, usaremos o exemplo do passeio aleatório no hipercubo n -dimensional, para ilustrar o funcionamento de cada uma dessas três técnicas. Veja a Seção 2.2 para uma descrição precisa desse exemplo.

O capítulo 2 apresenta a técnica de *tempo estacionário forte*. De maneira informal, esses são tempos de parada (portanto aleatórios) que têm a seguinte propriedade: ao observarmos a cadeia parada em um tempo estacionário forte, a sua posição será descrita pela distribuição estacionária e é independente do próprio tempo transcorrido. A observação chave é que a distância entre a distribuição da cadeia em um dado tempo e a distribuição estacionária pode ser cotada pela probabilidade do tempo estacionário forte ainda não ter chegado. Para o nosso exemplo (principal desse trabalho), o passeio aleatório no hipercubo N -dimensional vamos exibir a construção de um tempo estacionário forte.

O Capítulo 3 é totalmente dedicado à técnica do *acomplamento*. Depois de fornecer alguns resultados básicos sobre o assunto mostraremos um análogo do que foi visto no Capítulo 2, agora para a relação entre o acoplamento e a distância de variação total. Discutiremos o mesmo exemplo do passeio aleatório no hipercubo n -dimensional fazendo uma analogia com o exemplo da Urna de Enhefest, que pode ser vista em [10].

O capítulo 4 tratará de propriedades simétricas das cadeias de Markov e suas consequências. Veremos como $d(n)$ e $s_x(n)$ são afetados por essas condições de simetria e como estão relacionados ao segundo maior autovalor da matriz de transição da cadeia.

Por fim, no capítulo 5, descreveremos cotas obtidas através da *Análise de Fourier*. E, novamente trateremos o exemplo do passeio aleatório no hipercubo N -dimensional para entender essa técnica.

Capítulo 1

Cadeias de Markov e Convergência

Neste capítulo apresentaremos uma rápida introdução às Cadeias de Markov a fim de proporcionar uma certa base de entendimento para o leitor. Aqueles familiarizados com as definições e propriedades elementares de Cadeias e Markov podem preferir saltar diretamente para o Capítulo 2. Não temos a intenção de fornecer aqui uma introdução completa e totalmente rigorosa. Para maiores detalhes encorajamos o leitor a consultar as referências [10] e [11] onde o tema é apresentado de maneira introdutória e completa.

1.1 Primeiros Resultados

Consideraremos nesse trabalho cadeias de Markov em espaços de estado finito. Para isso começamos fixando um conjunto finito Ω chamado espaço de estados. Denotemos $\Omega = \{x_1, x_2, \dots, x_{|\Omega|}\}$, onde $|\Omega|$ é a cardinalidade de Ω . Cada ponto $x_i \in \Omega$ será chamado de estado e representa uma possível posição que a cadeia de Markov pode ocupar em um determinado instante de tempo.

Será importante estudar qual é a probabilidade de que a cadeia ocupe uma dada posição em Ω em um determinado tempo. Isso motiva a definição de distribuições de probabilidade sobre Ω .

Definição 1.1. *Uma distribuição de probabilidade μ em um conjunto finito Ω é uma função $\mu : \Omega \rightarrow [0, 1]$ que pode ser representada como um vetor linha e que satisfaz:*

$$(i) \quad \mu(k) \geq 0 \text{ para todo } k \in \Omega,$$

$$(ii) \quad \sum_{k \in \Omega} \mu(k) = 1.$$

Seja (Ω, μ) um espaço de probabilidade finito. Os subconjuntos de Ω são chamados de eventos. A probabilidade de um evento $B \subseteq \Omega$ é dada por

$$\mu(B) = \sum_{x \in B} \mu(x).$$

Dados dois eventos, $A, B \subseteq \Omega$, com $\mu(B) > 0$, a probabilidade condicional de A dado B é definida como

$$\mu(A|B) = \frac{\mu(A \cap B)}{\mu(B)}. \quad (1.1)$$

Dois eventos $A, B \subseteq \Omega$ são independentes se

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

Definição 1.2. *Seja (Ω, \mathcal{F}, P) um espaço de probabilidade finito, onde \mathcal{F} é uma σ -álgebra em Ω . Uma função $X : \Omega \mapsto \mathbb{R}$ é uma variável aleatória se $[X \leq x] = \{\omega \in \Omega; X(\omega) \leq x\}$ é um conjunto mensurável, isto é, $[X \leq x] \in \mathcal{F}$ para todo $x \in \mathbb{R}$.*

Observação 1.3. *Neste texto, geralmente a σ -álgebra considerada, será aquela que contém todos os subconjuntos de Ω . Sendo assim, qualquer função $X : \Omega \rightarrow \mathbb{R}$ se torna mensurável.*

Neste caso, uma variável aleatória pode ser entendida como uma descrição numérica do resultado de um experimento probabilístico.

Um processo estocástico é uma sequência de variáveis aleatórias X_0, X_1, X_2, \dots . A seguir, estaremos interessados em um tipo particular de processos estocásticos.

Definição 1.4. *Uma sequência de variáveis aleatórias (X_0, X_1, \dots) é uma Cadeia de Markov finita com espaço de estados finito Ω se, para todo $t \geq 1$ e $x_1, \dots, x_{t+1} \in \Omega$ vale:*

$$P(X_{t+1} = x_{t+1} | X_t = x_t, \dots, X_0 = x_0) = P(X_{t+1} = x_{t+1} | X_t = x_t). \quad (1.2)$$

As probabilidades condicionais $P(X_{t+1} = x_{t+1} | X_t = x_t)$ são denominadas Probabilidades de Transição e representam, portanto, a probabilidade X_{t+1} assumir o valor x_{t+1} no instante $t + 1$ condicionado em X_t ser igual a x_t no instante t .

A Equação (1.2) nos diz que, a chance de encontrar a cadeia de Markov no estado j no instante $t + 1$ depende somente do estado em que essa cadeia se encontra no instante t . Sendo assim, dado esse último estado, tudo aquilo que ocorreu até o momento t (estados X_0, X_1, \dots, X_{t-1}) é irrelevante para determinar a posição no tempo $t + 1$. Em outras palavras, condicionado no estado presente, o passado é independente do futuro. Esta propriedade é conhecida como a *Propriedade de Markov*.

Existe uma relação estreita entre cadeias de Markov e matrizes do tipo

$$P = \begin{pmatrix} P(1,1) & P(1,2) & \dots & P(1,n) \\ P(2,1) & P(2,2) & \dots & P(2,n) \\ \vdots & \vdots & \ddots & \vdots \\ P(n,1) & \dots & \dots & P(n,n) \end{pmatrix} \quad (1.3)$$

que satisfazem algumas condições.

Definição 1.5. *Dizemos que uma matriz quadrada $P = (x_i, x_j)$ como em (1.3) com $i, j \in \{1, \dots, |\Omega|\}$ é estocástica se cada linha é uma distribuição de probabilidade em Ω . Isto é, para cada $i \in \{1, \dots, |\Omega|\}$ temos que $P(x_i, x_j) \geq 0$ para todo $j \in \{1, \dots, |\Omega|\}$ e que $\sum_{j \in \{1, \dots, |\Omega|\}} P(x_i, x_j) = 1$.*

A matriz de transição de uma cadeia de Markov P é definida por

$$P(x_i, x_j) := P(X_{t+1} = x_j | X_t = x_i), \quad (1.4)$$

sendo $P(x_i, x_j)$ interpretado como a probabilidade de ir do estado x_i para o estado x_j . É importante ressaltar que, a matriz de transição é uma matriz estocástica. Mais ainda,

fixada uma distribuição inicial em Ω a matriz de transição determina completamente a (distribuição da) cadeia de Markov. Em outras palavras, uma cadeia de Markov fica inteiramente determinada uma vez que são especificadas a distribuição de X_0 e a matriz de transição.

Antes de prosseguirmos, passemos a um exemplo simples de cadeia de Markov finita.

Exemplo 1.6. Considere uma cadeia com três estados, digamos $\Omega = \{1, 2, 3\}$ definida através do seguinte diagrama:

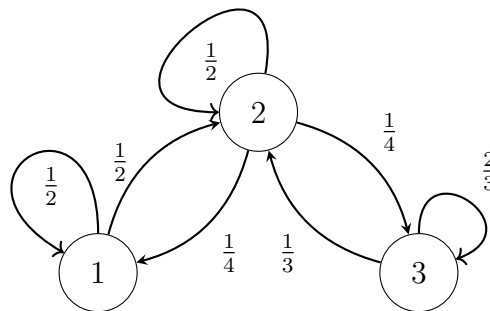


Figura 1.1: Cadeia de três estados correspondentes a matriz P

Assim, a sua matriz de transição P é dada por

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 1/3 & 2/3 \end{pmatrix} \quad (1.5)$$

onde, por exemplo, $P(2, 3) = \frac{1}{4}$, ou seja, a probabilidade de ir do estado 2 para o estado 3 é de $\frac{1}{4}$, como indicado na Figura 1.1 através de uma seta partindo do estado 2 em direção estado 3.

Além da propriedade de Markov, apresentamos a *Propriedade Forte de Markov* que difere da primeira por levar em conta a distribuição do estado ocupado pela cadeia de Markov após um tempo aleatório. Esses são os tempos de parada que definimos a seguir. Seja $\{X_n\}_{n \geq 0}$ um processo estocástico com espaço de estados finito Ω . Uma variável aleatória $\tau : \Omega \rightarrow \mathbb{N} \cup \{\infty\}$ é um tempo de parada se, para todo $m \geq 0$, o evento $\{\tau = m\}$ é mensurável com respeito a X_0, X_1, \dots, X_m .

Intuitivamente, pensamos em um tempo de parada como sendo aquele em que vamos interromper a evolução da cadeia. Mas para que esse tempo seja realmente um tempo de parada, é necessário que a decisão de interromper a evolução seja determinada somente com o conhecimento acumulado até o momento da tomada de decisão. Em outras palavras, para determinar se $\tau = m$ basta analisar as variáveis aleatórias X_0, \dots, X_m .

Podemos agora enunciar a Propriedade Forte de Markov (PFM) cuja demonstração pode ser encontrada em [11, Teorema 1.4.2, pág. 20].

Proposição 1.7 (Propriedade Forte de Markov). *Seja $\{X_n\}_{n \geq 0}$ uma cadeia de Markov com matriz de transição P e τ um tempo de parada para essa matriz, então temos que*

$$\begin{aligned} P(X_{\tau+1} = k_1, \dots, X_{\tau+K} = k_K | X_\tau = x_\tau, \dots, X_0 = x_0) \\ = P(X_{\tau+1} = k_1, \dots, X_{\tau+K} = k_K | X_\tau = x_\tau) \end{aligned}$$

Uma maneira de interpretar a equação acima é notando que, dado o evento $\{\tau = m, X_m = x_m\}$ o futuro $(X_{m+1}, X_{m+2}, \dots)$ é independente do passado (X_0, \dots, X_m) , ou de maneira ainda mais informal, que a cadeia “começa de novo” após um tempo de parada.

Se uma cadeia de Markov tem distribuição inicial μ , isto é, se $P(X_0 = x) = \mu(x)$ para todo $x \in \Omega$, então é possível verificar, por indução, que a distribuição da posição após n passos é dada por

$$P(X_n = k) = \sum_{j \in \Omega} \mu(x) P^n(x, y) = (\mu P^n)(y),$$

onde P^n denota a n -ésima potência da matriz P .

Neste trabalho estamos interessados no comportamento assintótico da distribuição

$$\mu^{(n)} = \mu P^n,$$

quando $n \rightarrow \infty$. Isto é, estamos interessados em determinar qual é o limite, caso esse exista, para distribuição da posição ocupado pela cadeia após n passos.

Definição 1.8. *Uma distribuição π em Ω satisfazendo*

$$\pi = \pi P, \tag{1.6}$$

é chamada de distribuição estacionária da cadeia de Markov com matriz de transição P . Ou seja,

$$\pi(y) = \sum \pi(x) P(x, y) \text{ para todo } y \in \Omega. \tag{1.7}$$

Note que, se uma cadeia de Markov possui uma distribuição estacionária π , tomando essa como a distribuição inicial, implica que a distribuição da cadeia permanece constante, isto é, $\pi P^n = \pi$ para todo n .

Vamos mostrar que sob algumas hipóteses, a cadeia têm uma distribuição estacionária que é o limite $\mu^{(n)}$ quando $n \rightarrow \infty$.

Definição 1.9. *Uma cadeia de Markov com matriz de transição P é dita irredutível se para quaisquer dois estados $x, y \in \Omega$ existe um inteiro positivo t possivelmente dependendo de x e y , tal que $P^t(x, y) > 0$. Caso contrário, chamamos a cadeia de redutível.*

Assim, uma cadeia de Markov é irredutível se é sempre possível ir de um estado para outro, apenas iterando a cadeia. Vamos considerar o exemplo a seguir para entender melhor essa propriedade.

Exemplo 1.10. *A cadeia de Markov no exemplo 1.6 é irredutível. De fato, utilizando (1.5) temos imediatamente:*

$$P^2 = \begin{pmatrix} 3/8 & 1/2 & 1/8 \\ 1/4 & 11/24 & 7/24 \\ 1/12 & 7/18 & 19/36 \end{pmatrix}.$$

Pela Definição, 1.9 vemos que para o inteiro $t = 2$ já temos que $P^2(x, y) > 0$ para todo $x, y \in \Omega$.

Definição 1.11. Seja o conjunto $\tau(x) := \{t \geq 1; P^t(x, x) > 0\}$ o conjunto de tempos em que é possível encontrar a cadeia em sua posição inicial x . O período do estado x é definido como o máximo divisor comum (mdc) do conjunto $\tau(x)$.

O período, para uma cadeia irredutível, é definido como o período comum a todos os estados. A cadeia é chamada de aperiódica se todos os estados tem período 1. Caso contrário, chamamos de periódica.

Exemplo 1.12. Considere uma cadeia com quatro estados, digamos $\Omega = \mathbb{Z}_4 = \{0, 1, 2, 3\}$ o conjunto dos restos módulo 4, definida através do seguinte diagrama:

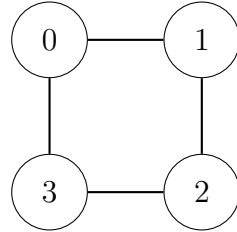


Figura 1.2: Passeio aleatório em \mathbb{Z}_4 é periódica, uma vez que cada passo vai de um estado par para um estado ímpar, ou vice-versa.

Consideramos um passeio aleatório nesse diagrama. A cada etapa, uma moeda é lançada. Se a moeda cair cara, a caminhada dá um passo no sentido horário. Se a moeda cair coroa, a caminhada se moverá um passo no sentido anti-horário. Seja X_t a cadeia de Markov associada e seja $X_0 = v_1$ ou seja, no tempo $t = 0$ o passeio está no vértice v_1 . Assim, para retornar a esse vértice, ele precisa andar um número par de vértices. Isto significa que $P^t(v_1, v_1) > 0$ apenas para $t = 2, 4, 6, \dots$. Consequentemente

$$\text{mdc}\{t > 1 : P^t(x, x) > 0\} = \text{mdc}\{2, 4, 6\} = 2$$

portanto a cadeia é periódica. Já um exemplo de cadeia aperiódica é o caso \mathbb{Z}_7 .

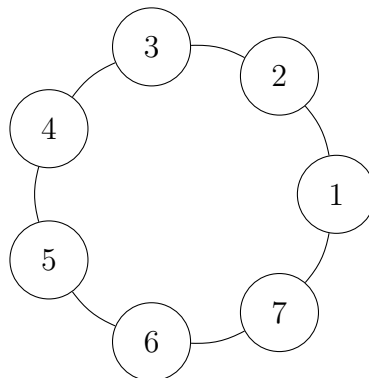


Figura 1.3: O passeio aleatório em \mathbb{Z}_7 é uma cadeia aperiódica, uma vez que o $\text{mdc}\{t \geq 1; P^t(x, x) > 0\} = 1$

De maneira informal, podemos dizer que a irredutibilidade nos garante que a cadeia não irá ficar restrita a apenas alguns estados, ou seja, é possível alcançar qualquer estado a

partir de um estado inicial. Já aperiodicidade garante ser possível atingir qualquer estado depois de passado qualquer valor tempo suficientemente grande. Assim, por exemplo, não é necessário aguardar um tempo par para poder observar a cadeia em um determinado estado.

Proposição 1.13. *Se P é aperiódica e irredutível, então existe um inteiro r tal que $P^r(x_i, x_j) > 0$ para todo $x_i, x_j \in \Omega$.*

A demonstração pode ser vista em [10, Proposição 1.7].

O próximo resultado nos garante que sob certas condições a existência e a unicidade de distribuições estacionárias para uma cadeia de Markov. A sua demonstração pode ser encontrada em [10, Seções 1.5.3 e 1.5.4].

Proposição 1.14. *Seja P uma matriz de transição de uma cadeia de Markov irredutível. Então existe uma única distribuição estacionária π em Ω tal que $\pi = \pi P$ e $\pi(x) > 0$ para todo $x \in \Omega$*

Veremos que as condições de irredutibilidade e aperiodicidade são suficientes para a convergência de uma cadeia de Markov para uma distribuição estacionária. No entanto, para estudar tal convergência é necessário primeiro definir uma noção de distância entre distribuições. Para isso vamos estudar na próxima seção uma maneira de "medir" a distância entre duas distribuições.

1.2 Convergência para o Equilíbrio

Vamos começar estudando uma forma de “medir” a distância entre duas distribuições que é explicitamente probabilística: a distância entre as distribuições μ e ν é a diferença máxima entre as probabilidades atribuídas a um mesmo evento.

Definição 1.15. *A distância de variação total entre duas distribuições de probabilidade μ e $\nu \in \Omega$ é*

$$\| \mu - \nu \|_{TV} = \max_{A \subseteq \Omega} | \mu(A) - \nu(A) | . \quad (1.8)$$

Às vezes é conveniente usar uma forma equivalente para a distância acima.

Proposição 1.16. *Sejam μ e ν distribuições em Ω . Então*

$$\| \mu - \nu \|_{TV} = \frac{1}{2} \sum_{x \in \Omega} | \mu(x) - \nu(x) | . \quad (1.9)$$

Demonstração. Definimos o conjunto $B = \{x; \mu(x) \geq \nu(x)\}$ e seja $A \subseteq \Omega$ um evento. Então pela identidade $A = (A \cap B) \sqcup (A \cap B^c)$ temos que

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

onde \sqcup significa união disjunta e na última desigualdade usamos $\mu(A \cap B^c) - \nu(A \cap B^c) \leq 0$, pela definição de B .

De maneira similar, temos $B = (A \cap B) \sqcup (A^c \cap B)$. Assim, temos que

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

onde, na última desigualdade, usamos $\mu(A^c \cap B) - \nu(A^c \cap B) \geq 0$.

Logo,

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

Similarmente, podemos ver que $\nu(A) - \mu(A) \leq \nu(B^c) - \mu(B^c) = \mu(B) - \nu(B)$.

Assim temos que

$$-(\mu(B) - \nu(B)) \leq \mu(A) - \nu(A) \leq \mu(B) - \nu(B),$$

logo

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

Então

$$\begin{aligned}2 |\mu(A) - \nu(A)| &\leq [\mu(B) - \nu(B) + \nu(B^c) - \mu(B^c)] \\ &= \sum_{x \in B} |\mu(x) - \nu(x)| + \sum_{x \in B^c} |\mu(x) - \nu(x)| \\ &= \sum_{x \in \Omega} |\mu(x) - \nu(x)|.\end{aligned}$$

Como A é um evento arbitrário, a Equação (1.9) nos dá

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

□

A prova da Proposição 1.16 também mostra que

$$\|\mu - \nu\|_{TV} = \sum_{\substack{x \in \Omega \\ \mu(x) > \nu(x)}} [\mu(x) - \nu(x)]. \quad (1.10)$$

Pode-se verificar facilmente, a partir da Equação (1.9), que a distância de variação total é uma métrica no conjunto de todas as distribuições de probabilidade em Ω .

Podemos definir agora algumas quantidades que quantificam o quão distante a cadeia de Markov se encontra do equilíbrio em um dado instante de tempo n . Primeiramente, seja $\pi_n(y) = P^n(x, y) = P(X_n = y)$ com $x, y \in \Omega$ podemos definir

$$d(n) = \|\pi_n^{(x)} - \pi\|_{TV}. \quad (1.11)$$

Maximizando sobre a posição inicial obtemos

$$d^*(n) = \max_{x \in \Omega} \|P^n(x, \cdot) - \pi(\cdot)\|_{TV}. \quad (1.12)$$

Uma outra quantidade de interesse é

$$\bar{d}(n) = \max_{x, y \in \Omega} \| P^n(x, \cdot) - P^n(y, \cdot) \|_{TV} \quad (1.13)$$

que envolve a distância entre a distribuição de duas cadeias de Markov com mesma matriz de transição começando de estados diferentes.

Podemos mostrar a seguinte relação entre as quantidades definidas acima:

Lema 1.17. *Considere $d^*(t)$ e $\bar{d}(t)$ como definidas em (1.12) e (1.13), respectivamente, então*

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

Demonstração. Para provar a primeira desigualdade usaremos a identidade

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

para qualquer conjunto $A \subseteq \Omega$. Esta igualdade segue diretamente da definição de estacionariedade se A for um conjunto com apenas um elemento $\{x\}$. Para obtê-la para A arbitrário, basta somar sobre os elementos em A .

Assim temos

$$\begin{aligned} \| P^t(x, \cdot) - \pi \|_{TV} &= \max_{A \subseteq \Omega} | P^t(x, A) - \pi(A) | \\ &= \max_{A \subseteq \Omega} \left| \sum_{y \in \Omega} \pi(y) [P^t(x, A) - P^t(y, A)] \right| \\ &\leq \max_{A \subseteq \Omega} \sum_{y \in \Omega} \pi(y) | P^t(x, A) - P^t(y, A) | \\ &\leq \sum_{y \in \Omega} \pi(y) \max_{A \subseteq \Omega} | P^t(x, A) - P^t(y, A) | \\ &= \sum_{y \in \Omega} \pi(y) \| P^t(x, \cdot) - P^t(y, \cdot) \|_{TV}. \end{aligned} \quad (1.14)$$

Uma vez que a média ponderada de um conjunto de números nunca é maior do que seu elemento máximo, o lado direito de (1.14) é limitado por $\max_{y \in \Omega} \| P^t(x, \cdot) - P^t(y, \cdot) \|_{TV}$. Tomando o máximo sobre todo x em Ω , temos

$$d^*(t) = \max_{x \in \Omega} \| P^t(x, \cdot) - \pi \|_{TV} \leq \max_{x, y \in \Omega} \| P^t(x, \cdot) - P^t(y, \cdot) \|_{TV} = \bar{d}(t). \quad (1.15)$$

Para mostrar a segunda desigualdade, usaremos a desigualdade triangular para $\| \cdot \|_{TV}$, inserindo π como distribuição intermediária.

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

□

O Lema 1.17 é importante, pois nos permite obter uma cota para $d^*(t)$ em termos de $\bar{d}(t)$ mesmo sem saber qual é a distribuição estacionária.

Outra noção de distância que vamos introduzir é a distância de separação entre duas distribuições de probabilidade. Suponha que μ e ν sejam distribuições de probabilidade em Ω e que $\nu(y) \neq 0$ para todo $y \in \Omega$.

$$s(\mu, \nu) := \max_{y \in \Omega} \left(1 - \frac{\mu(y)}{\nu(y)} \right). \quad (1.16)$$

Equivalentemente,

$$s(\mu, \nu) \text{ é o menor } s \geq 0 \text{ t.q. } \mu = (1 - s)\nu + sq \text{ para algum distribuição } q. \quad (1.17)$$

De fato, escrevendo $\alpha(\mu, \nu) = s(\mu, \nu)^{-1}$,

$$q(x) := \alpha(\mu, \nu)\mu(x) + (1 - \alpha(\mu, \nu))\nu(x)$$

define uma distribuição de probabilidade sobre Ω , uma vez que é uma combinação convexa de distribuições de probabilidade. Seja agora q uma distribuição qualquer e $s \geq 0$ tal que $\mu = (1 - s)\nu + sq$. Então para todo $x \in \Omega$ temos que

$$1 - \frac{\mu(y)}{\nu(y)} = s - s \frac{\mu(y)}{\nu(y)} \leq s$$

para todo $x \in \Omega$.

A noção de separação entre duas distribuições nos permite definir para uma cadeia de Markov com matriz de transição P a separação no tempo t . Para um valor de $x \in \Omega$ fixo, definimos:

$$s_x(t) = s(P^t(x, \cdot), \pi(\cdot)) = \max_{y \in \Omega} \left(1 - \frac{P^t(x, y)}{\pi(y)} \right). \quad (1.18)$$

Maximizando sobre os valores de $x \in \Omega$:

$$s^*(t) = \max_{x \in \Omega} s_x(t) = \max_{x \in \Omega} s(P^t(x, \cdot), \pi(\cdot)) = \max_{x, y \in \Omega} \left(1 - \frac{P^t(x, y)}{\pi(y)} \right). \quad (1.19)$$

Um resultado interessante é que a distância de separação é uma cota superior para a distância de variação total

$$0 \leq \| \mu - \nu \|_{TV} \leq s(\mu, \nu) \leq 1. \quad (1.20)$$

De fato,

$$\begin{aligned} \| \mu - \nu \|_{TV} &= \sum_{y: \nu(y) \geq \mu(y)} (\nu(y) - \mu(y)) = \sum_{y: \nu(y) \geq \mu(y)} \nu(y) \left(1 - \frac{\mu(y)}{\nu(y)} \right) \\ &\leq \sum_{y: \nu(y) \geq \mu(y)} \max_y \left(1 - \frac{\mu(y)}{\nu(y)} \right) = s(\mu, \nu). \end{aligned}$$

Em particular,

$$d^*(t) \leq s_x(t). \quad (1.21)$$

Observação 1.18. Os valores de $d(t)$ e $s_x(t)$ podem diferir bastante. Por exemplo, seja Q a distribuição uniforme em $\Omega - \{l_1\}$ onde $\Omega = \{l_1, l_2, l_3, l_4, l_5, l_6\}$ representa os lados de um dado. Seja $\pi(l_i) = U(l_i) = \frac{1}{|\Omega|}$ para todo $i \in \{1, 2, 3, 4, 5, 6\}$, isto é, π é a distribuição uniforme em Ω . Então $d(1) = \sum_{Q(l_i) \geq \pi(l_i)} (\frac{1}{5} - \frac{1}{6}) = \frac{1}{6}$, mas $s_x(1) = 1$, pois o máximo é atingido quando $Q(l_1) = 0$.

Agora já podemos enunciar o Teorema da Convergência que garante, sob as hipóteses de irredutibilidade e aperiodicidade, a convergência de todas as linhas $P^t(x, \cdot)$, com $x \in \Omega$, para a distribuição estacionária. Veja [10, Teorema 4.9] para uma demonstração.

Teorema 1.19. *Suponha que P seja irredutível e aperiódica, com distribuição estacionária π . Então existe uma constante $\alpha \in (0, 1)$ e $C > 0$ tal que*

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

O significado desse teorema é claro: não importando a distribuição inicial, depois de uma quantidade suficiente de passos, a probabilidade de estar em y se torna muito próxima de $\pi(y)$.

Dizemos que uma distribuição de probabilidade π satisfaz a equação de balanço detalhado para uma cadeia de Markov P se, $\forall x, y \in \Omega$ se

$$\pi(x)P(x, y) = \pi(y)P(y, x). \quad (1.22)$$

Proposição 1.20. *Se π satisfaz a equação de balanço detalhado (1.22) para uma cadeia P , então ela é estacionária para essa cadeia.*

Demonstração. Temos apenas que somar sobre todo $y \in \Omega$ em ambos os lados da Equação (1.22) e usando a estacionaridade de P temos

$$\sum_{y \in \Omega} \pi(x)P(x, y) = \sum_{y \in \Omega} \pi(y)P(y, x) = \pi(x). \quad (1.23)$$

□

Essa propriedade geralmente é útil para verificar se uma distribuição de probabilidade é estacionária. Além disso, iterando a Equação (1.22)

$$\pi(x_0)P(x_0, x_1)\pi(x_1)P(x_1, x_2) \dots P(x_{n-1}, x_n) = \pi(x_n)P(x_n, x_{n-1}), \dots, P(x_1, x_0).$$

Isso significa que, a partir da medida estacionária, a distribuição das variáveis (X_0, \dots, X_t) e de (X_t, \dots, X_0) é a mesma, ou seja,

$$P_\pi(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = P_\pi(X_n = x_n, \dots, X_1 = x_1, X_0 = x_0).$$

Definição 1.21. *Se uma cadeia (X_t) satisfaz a Equação (1.22) e têm distribuição inicial estacionária, então a distribuição de (X_0, X_1, \dots, X_n) é a mesma que a distribuição de $(X_n, X_{n-1}, \dots, X_0)$. Por esta razão, uma cadeia que satisfaça (1.22) é chamada reversível.*

Uma propriedade interessante da distância de separação é

Lema 1.22. A separação $s^*(n)$ é uma função submultiplicativa, ou seja,

$$s^*(m+n) \leq s^*(m)s^*(n); \quad \text{para todo } m, n \geq 1$$

Em particular, $s^*(n)$ é decrescente em n .

Demonstração. Seja $\alpha(n) := \frac{1}{s^*(n)}$ e escrevamos o vetor linha π na forma

$$\pi = [\pi(1) \pi(2) \cdots \pi(|\Omega|)].$$

Definamos Π como sendo a matriz $|\Omega| \times |\Omega|$ cujas $|\Omega|$ linhas são cópias do vetor π

$$\Pi_{|\Omega| \times |\Omega|} = \begin{pmatrix} \pi(1) & \pi(2) & \cdots & \pi(|\Omega|) \\ \pi(1) & \pi(2) & \cdots & \pi(|\Omega|) \\ \vdots & \vdots & \ddots & \vdots \\ \pi(1) & \pi(2) & \cdots & \pi(|\Omega|) \end{pmatrix}$$

e defina, para cada n a matriz

$$V^n := \alpha(n)P^n + (1 - \alpha(n))\Pi. \quad (1.24)$$

Ou seja, V^n é a matriz cujas entradas satisfazem

$$P^n(x, y) = (1 - s^*(n))\pi(y) + s^*(n)V^n(x, y).$$

Note que V^n é estocástica.

É simples verificar que $M\Pi = \Pi$ para qualquer matriz estocástica M e que $\Pi M = \Pi$ para qualquer matriz M tal que $\pi M = \pi$. Logo

$$\begin{aligned} \Pi V^n &= \alpha(n)\Pi P^n + (1 - \alpha(n))\Pi \Pi \\ &= \alpha(n)\Pi + (1 - \alpha(n))\Pi = \Pi \end{aligned}$$

o que mostra que

$$\pi V^n = \pi.$$

Agora temos que

$$\begin{aligned} P^{m+n} &= [(1 - s^*(m))\Pi + s^*(m)V^m][(1 - s^*(n))\Pi + s^*(n)V^n] \\ &\stackrel{(*)}{=} [1 - s^*(m)s^*(n)]\Pi + s^*(m)s^*(n)V^m V^n, \end{aligned}$$

onde em $\stackrel{(*)}{=}$ usamos o fato que $\Pi V_n = V_m \Pi = \Pi$. Em outras palavras, temos que

$$P^{m+n}(x, y) = [1 - s^*(m)s^*(n)]\pi(y) + s^*(m)s^*(n)(V^m V^n)(x, y).$$

Usando a Equação (1.18) concluimos que $s_x(m+n) \leq s^*(m)s^*(n)$. Maximizando sobre $x \in \Omega$, nos fornece o resultado desejado. \square

O próximo resultado relaciona a separação associada as cadeias cujas as matrizes de transição não são muito diferentes.

Lema 1.23. Sejam P_1, P_2 matrizes de transição em Ω , e sejam s_1^* e s_2^* as respectivas separações máximas definidas em (1.19). Suponha que

- (i) existe $c > 0$ tal que $P_2(x, y) \geq cP_1(x, y)$; para todo $x, y \in \Omega$;
(ii) $P_1P_2 = P_2P_1$.

Então

$$s_2^*(n) \leq \sum_{b=0}^n s_1^*(b) \binom{n}{b} c^b (1-c)^{n-b},$$

para todo $n \geq 1$.

Demonstração. Por (i) podemos escrever que

$$P_2 = cP_1 + (1-c)V, \quad (1.25)$$

para alguma matriz de transição V .

Como consequência de (ii), temos que $P_1V = VP_1$. Com isso, utilizando o Binômio de Newton para expandir a potência da Equação (1.25) temos

$$P_2^n = \sum_{b=0}^n \binom{n}{b} c^b (1-c)^{n-b} P_1^b V^{n-b}. \quad (1.26)$$

A comutatividade entre as matrizes P_1 e P_2 implica que essas matrizes têm a mesma distribuição estacionária π . De fato, supondo que $\pi P_1 = \pi$, vale que

$$\pi P_2 P_1 = \pi P_1 P_2 = \pi P_2.$$

Logo π e πP_2 são ambos autovetores (à esquerda) de P_1 com autovalor um. Como o autoespaço de P_1 associado ao autovalor 1 é unidimensional, πP_2 deve ser um múltiplo de π . Em outras palavras, π é um autovetor de P_2 assim como de P_1 . Consequentemente, a matriz V e suas potências também têm distribuição estacionária π .

Para uma matriz estocástica Q cuja distribuição estacionária seja π , vamos denotar

$$s^*(Q) := \max_{x \in \Omega} \max_{y \in \Omega} \left(1 - \frac{Q(x, y)}{\pi(y)}\right).$$

Note que, se V_1 e V_2 são duas matrizes estocásticas com a mesma distribuição estacionária, então

$$s^*(V_1 V_2) \leq s^*(V_1),$$

em palavras, a multiplicação diminui a distância de separação até a distribuição estacionária. Isso pode ser demonstrado com uma simples adaptação da prova do Lemma 1.22. Então

$$s^*(P_1^b V^{n-b}) \leq s^*(P_1^b) = s_1^*(b). \quad (1.27)$$

Note também que, se Q_1, Q_2, \dots, Q_n são matrizes estocásticas com a mesma distribuição estacionária e q_1, q_2, \dots, q_n são tais que $q_i \geq 0$ e $q_1 + q_2 + \dots + q_n = 1$ então

$$\begin{aligned} & 1 - \frac{q_1 Q_1(x, y) + q_2 Q_2(x, y) + \dots + q_n Q_n(x, y)}{\pi(y)} \\ &= q_1 \left(1 - \frac{Q_1(x, y)}{\pi(y)}\right) + q_2 \left(1 - \frac{Q_2(x, y)}{\pi(y)}\right) + \dots + q_n \left(1 - \frac{Q_n(x, y)}{\pi(y)}\right). \end{aligned}$$

Maximizando sobre x e y em Ω nos fornece

$$s^*(q_1Q_1(x, y) + q_2Q_2(x, y) + \cdots + q_nQ_n(x, y)) \leq q_1s^*(Q_1) + q_2s^*(Q_2) + \cdots + q_ns^*(Q_n). \quad (1.28)$$

Usando as Equações (1.27) e (1.28) em (1.26) iteradamente termina a demonstração. \square

O resultado acima é interessante porque nos diz que, grosso modo, se n passos bastam para fazer s_1^* pequeno, então $\frac{n}{c}$ passos são suficientes para tornar o s_2^* pequeno.

Capítulo 2

Cotas superiores para convergência

Existem vários argumentos disponíveis para estimar as taxas de convergência de cadeias de Markov. Nesse capítulo vamos falar do método de Tempo Estacionário Forte, que foram introduzidos em [3].

2.1 Tempo Estacionário Forte

Definição 2.1. *Seja $\{X_n\}_{n \geq 0}$ uma cadeia de Markov com distribuição de probabilidade estacionária π . Um tempo estacionário forte T é um tempo de parada tal que*

$$P(X_k = x, T = k) = P(T = k)\pi(x), \quad (2.1)$$

para todo $x \in \Omega$ e todo inteiro positivo k .

Lema 2.2. *São equivalentes*

- (i) $P(X_k = x|T = k) = \pi(x)$;
- (ii) $P(X_k = x|T \leq k) = \pi(x)$;
- (iii) T é um tempo estacionário forte.

Ou seja, o item (iii) diz que X_T têm distribuição π e é independente de T .

Demonstração. Que os itens (i) e (iii) são equivalentes segue diretamente das definições de tempo estacionário forte e da definição (1.1) de probabilidade condicional.

Mostremos então que os itens (i) e (ii) são equivalentes.

Primeiro suponha que (i) vale. Então

$$P(X_k = x|T \leq k) = \frac{P(X_k = x, T \leq k)}{P(T \leq k)} = \frac{1}{P(T \leq k)} \sum_{j=0}^k P(X_k = x, T = j).$$

Assim, usando (i) temos que:

$$\sum_{j=0}^k P(X_k = x, T = j) = \sum_{j=0}^k P(X_k = x|T = j)P(T = j) = \sum_{j=0}^k \pi(x)P(T = j).$$

Logo temos que

$$\begin{aligned} P(X_k = x | T \leq k) &= \frac{1}{P(T \leq k)} \sum_{j=0}^k \pi(j) P(T = j) \\ &= \frac{\pi(x)}{P(T \leq k)} \sum_{j=0}^k P(T = j) = \pi(x). \end{aligned}$$

Reciprocamente, suponhamos que (ii) vale. Então

$$\begin{aligned} P(X_k = x, T = k) &= P(X_k = x, T \leq k) - P(X_k = x, T \leq k - 1) \\ &= \pi(x)P(T \leq k) - \pi(x)P(T \leq k - 1) \\ &= \pi(x)P(T = k), \end{aligned}$$

onde na segunda igualdade usamos os itens (i) e (ii) descritos no Lema 2.2. Dividindo os dois lados por $P(T = k)$ temos o resultado. \square

Para compreender melhor a definição de tempo estacionário forte vamos explorar o seguinte exemplo:

Exemplo 2.3. (*Embaralhamento “de cima ao acaso”*). Vamos considerar uma pilha contendo n cartas que será embaralhada repetidamente removendo-se a carta do topo e inserindo-a em uma posição aleatória escolhida uniformemente nas n possíveis posições. Cada repetição do embaralhamento é feita independentemente das demais. Vamos nos atentar em seguir a posição da carta que começa na posição inferior na pilha original, vamos denominá-la de N . Vamos mostrar, por indução, que em um tempo t existem k cartas embaixo de N e cada uma das $k!$ possíveis orientações estão igualmente distribuídas.

No tempo $t = 0$, não há cartas abaixo de N , sendo a chance de inserir a carta acima de N maior do que a de inseri-la abaixo. A carta N continua na posição final até o primeiro tempo em que uma carta é inserida abaixo dela, que leva cerca de n embaralhamentos. Seja essa carta chamada de k_1 .

Como o procedimento continua, eventualmente uma segunda carta k_2 é inserida abaixo da carta final original (isso leva cerca de mais $\frac{n}{2}$ embaralhas). Nota-se que, nesse momento, as duas cartas abaixo de N têm a mesma probabilidade de serem encontradas na ordem (k_1, k_2) ou (k_2, k_1) . Similarmente, no primeiro tempo em que uma terceira carta k_3 é inserida abaixo de N , cada uma das seis possíveis ordens das três cartas k_1, k_2 e k_3 são igualmente prováveis.

Suponhamos agora que no instante t existam k cartas sob N e cada uma das $k!$ ordenações possíveis é igualmente provável.

Para o instante $t + 1$ existem dois casos:

- *Caso 1:* Uma carta é colocada sob N , então como a carta foi inserida aleatoriamente e no tempo t , as cartas embaixo desta eram equiprováveis, todas as $(k+1)!$ ordenações das $k + 1$ cartas são agora igualmente prováveis.
- *Caso 2:* A carta é colocada acima de N então, a distribuição das cartas sob N permanece inalterada, ou seja, permanecem distribuídas uniformemente.

Agora consideramos o primeiro tempo τ_{top} em que a carta N chega ao topo, ou seja, há $n - 1$ cartas abaixo dela. Por indução, vimos que todos os $(n - 1)!$ arranjos das cartas abaixo de N são igualmente prováveis.

Se pararmos de embaralhar precisamente um embaralhamento após este tempo, então a ordem das cartas neste momento é exatamente uniforme em todos os arranjos possíveis. De fato, quando N é inserida aleatoriamente, todas n possíveis localizações para a carta são igualmente distribuídas, e então, cada um dos $n!$ arranjos da pilha são equiprováveis.

Ademais, no tempo $t = \tau_{top} + 1$ as cartas abaixo de N estão uniformemente distribuídas. Portanto, a ordem das cartas em τ_{top} é uniforme e τ_{top} é independente de $X_{\tau_{top}}$.

Logo, $t = \tau_{top} + 1$ é um tempo estacionário forte.

Quando a carta inferior original (N) está na posição j do fundo, o tempo de espera para que uma nova carta seja inserida abaixo dela é de cerca de $\frac{N}{j}$. Assim, o tempo de espera para que a carta N chegue ao topo e seja inserida é de aproximadamente

$$n + \frac{n}{2} + \frac{n}{3} + \dots + \frac{n}{n} \simeq n \log n.$$

Agora vamos enunciar um resultado muito importante que relaciona distância de separação e o tempo estacionário forte.

Proposição 2.4. (a) Se T é um tempo estacionário forte para $\{X_n\}$ então $s_x(n) \leq P(T > n)$, para todo $n \geq 0$.

(b) Por outro lado, existe um tempo estacionário forte T tal que $s_x(n) = P(T > n)$.

Demonstração. (a) Como T é tempo estacionário forte, temos que pelo Lema 2.2

$$P(X_n = y | T \leq n) = \pi(y).$$

Assim,

$$P(X_n = y) \geq P(X_n = y, T \leq n) = \pi(y)P(T \leq n) = \pi(y)[1 - P(T > n)].$$

Usando que $P(X_n = y) = \pi_n(y)$, temos que

$$\frac{\pi_n(y)}{\pi(y)} \geq 1 - P(T > n),$$

ou seja, em particular $s_x(n) \leq P(T > n)$.

(b) Fixando $x \in \Omega$ e seja

$$a_n = \min_y \frac{P^n(x, y)}{\pi(y)} = \min_y \frac{P(X_n = y)}{\pi(y)} = 1 - s_x(n).$$

Note que a_n é não decrescente, pois $s_x(n)$ é decrescente. Seja K o menor inteiro tal que $a_K > 0$.

Definiremos T indutivamente e mostraremos que é um tempo estacionário forte. O primeiro passo é

$$P(T < K) = 0,$$

$$P(T = K | X_K = y) = \frac{a_K \pi(y)}{P(X_K = y)}.$$

Isso implica que $P(T = K, X_K = y) = a_K \pi(y)$ como na Equação (2.1). Podemos mostrar por indução que para $n > K$ temos

$$P(T = n, X_n = y) = \pi(y)(a_n - a_{n-1}). \quad (2.2)$$

Para isso, seja

$$P(T = n | X_n = y, T > n - 1) = \frac{a_n - a_{n-1}}{\frac{\pi_n(y)}{\pi(y)} - a_{n-1}}. \quad (2.3)$$

A quantidade no lado direito de (2.3) faz sentido pois está em $[0, 1]$ já que a_n é não decrescente e $a_n \leq \frac{P^n(x, y)}{\pi(y)}$.

Supomos que (2.2) vale para todos os inteiros menores que n , assim

$$\begin{aligned} P(X_n = y, T = n) &= P(T = n | X_n = y, T > n - 1) \cdot P(X_n = y, T > n - 1) \\ &= \frac{a_n - a_{n-1}}{\frac{\pi_n(y)}{\pi(y)} - a_{n-1}} \cdot (P(X_n = y) - P(X_n = y, T \leq n - 1)). \end{aligned} \quad (2.4)$$

Então

$$P(X_n = y, T \leq n - 1) = \pi(y)a_{n-1}. \quad (2.5)$$

Usando a Equação (2.5) em (2.4) temos a (2.2) que implica que T é um tempos estacionário forte.

Além disso, a (2.2) implica que $P(T = n) = a_n - a_{n-1}$. Portanto, consequentemente temos que $P(T \leq n) = a_n$. Logo

$$a_n = P(T \leq n) = 1 - P(T > n) = 1 - s_x(n) \Rightarrow s_x(n) = P(T > n).$$

□

2.2 Passeio Aleatório no Hipercubo: Tempo Estacionário Forte

Consideramos aqui o hipercubo $(\mathbb{Z}/2\mathbb{Z})^N$ como sendo o grafo com vértices $\{0, 1\}^N$ e cujas arestas ligam os pares de vértices que diferem em exatamente uma coordenada. Vamos estudar o passeio aleatório neste grafo.

Nesse tipo de passeio a partir de um dado vértice (x_1, x_2, \dots, x_N) escolhe-se uma coordenada $j \in \{1, 2, \dots, N\}$ uniformemente ao acaso e define-se o novo estado como sendo $(x_1, x_2, \dots, x_{j-1}, 1 - x_j, x_{j+1}, \dots, x_N)$. Ou seja, se a coordenada escolhida vale 0 então ela se tornará 1 e vice-versa.

Infelizmente, o passeio aleatório simples no hipercubo é periódico, uma vez que cada movimento inverte a paridade do número de coordenadas que são iguais a 1. Para

evitar esse problema, estudaremos uma versão preguiçosa do passeio aleatório no hiper-cubo. Nessa versão, em cada passo o passeio permanece em sua posição atual com probabilidade positiva e se move segundo a regra anterior com a propabilidade complementar. A escolha mais comum consiste em tomar a probabilidade de que o passeio não se mova igual $\frac{1}{2}$. Abaixo consideraremos uma versão na qual a probabilidade de o passeio não se mover vale $\frac{1}{N+1}$. A probabilidade de transição de entre os vértices $x = (x_1, x_2, \dots, x_N)$ e $y = (y_1, y_2, \dots, y_N)$ é dada por

$$P(x, y) = \begin{cases} \frac{1}{N+1}, & \text{se } \sum_s |x_s - y_s| \in \{0, 1\}; \\ 0, & \text{caso contrário.} \end{cases} \quad (2.6)$$

Vamos construir um tempo estacionário forte para esse processo. Para tanto será útil construir esse processo de uma maneira particular. Definimos então $\{\theta_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias i.i.d. (idependentes e ideticamente distribuídas) no conjunto $\{0, 1, \dots, N\}$, isto é, para cada $j \in \{0, 1, \dots, N\}$,

$$P(\theta_n = j) = \frac{1}{N+1}.$$

Através dessa sequência podemos então definir recursivamente o passeio aleatório

$$\mathbf{X}(n) = (X_i(n); 1 \leq i \leq N)$$

fazendo $\mathbf{X}(0) = 0$ e, condicionado no valor de $\mathbf{X}(n-1)$, definindo

$$\mathbf{X}(n) = \mathbf{X}(n-1) + \sum_{k=1}^N e_k \mathbb{1}_{(\theta_n=k)} \pmod{2},$$

onde e_k é o k -ésimo vetor canônico, isto é, o vetor cujas entradas são nulas, exceto a k -ésima entrada que vale 1.

Note que o passeio não se move no tempo n se $\theta_n = 0$, o que acontece com probabilidade $1/(N+1)$, como queríamos. Ao contrário, se $\theta_n = k \neq 0$ então a k -ésima entrada têm o seu valor alternado. Segue daí que $\mathbf{X}(n)$ é uma cadeia de Markov com transição dada por (2.6).

Vamos agora definir recursivamente uma segunda sequência

$$\mathbf{C}(n) = (C_i(n); 1 \leq i \leq N)$$

que também toma valores em $\{0, 1\}^N$. A construção também será feita de forma recursiva. Definimos assim $\mathbf{C}(0) = (0, \dots, 0)$ e, para cada $n \geq 1$, o valor de $\mathbf{C}(n)$ será dado em termos de $\mathbf{C}(n-1)$, θ_n e outras sequências adicionais de variáveis aleatórias independentes. A construção, será feita de forma que as entradas sejam crescentes, isto é, uma vez que $C_i(n) = 1$ então teremos $C_i(n+m) = 1$ para todo $m \geq 0$.

Antes de passarmos à construção precisa do processo $\mathbf{C}(n)$, vamos somente mencionar intuitivamente que ele pode ser entendido como um processo de “verificação de coordenadas”. Uma coordenada i será declarada verificada quando o valor de C_i passar de 0 para 1. Se $\theta_n = i$, tentamos verificar a coordenada i no tempo n , caso ela ainda não tenha sido verificada. Para isso uma moeda honesta (ζ_n) independente do restante é lançada. Se o resultado for “cara” ($\zeta_n = 1$), de fato declaramos a coordenada i verificada.

Se o resultado for “coroa” escolhamos uma coordenada (ξ_n) diferente de i uniformemente dentre as que ainda não foram verificadas e a declaramos verificada.

De maneira precisa, a regra para a evolução de $\mathbf{C}(n)$ será dada pelo seguinte algoritmo:

1. Enquanto $\sum_{k=1}^N C_k(n-1) \leq N-2$ repita:
 - (a) Se $\theta_n = 0$ então $\mathbf{C}(n) = \mathbf{C}(n-1)$;
 - (b) Se $\theta_n = j \neq 0$ e $C_j(n-1) = 1$ então $\mathbf{C}(n) = \mathbf{C}(n-1)$;
 - (c) Se $\theta_n = j \neq 0$ e $C_j(n-1) = 0$ então defina ζ_n variável aleatória com distribuição Bernoulli(1/2) e ξ_n variável aleatória uniforme em $\{k: C_k(n-1) = 0\}$, ambas independentes de todo o resto e defina

$$\mathbf{C}(n) = \mathbf{C}(n-1) + e_j \mathbb{1}_{(\zeta_n=1)} + \sum_{k \neq j} e_k \mathbb{1}_{(\zeta_n=0, \xi_n=k)}.$$

2. Enquanto $\sum_{k=1}^N C_k(n-1) = N-1$ fixe $\{j\} = \{k \in \{1, \dots, n\}: C_k(n-1) = 0\}$ e repita:
 - (a) Se $\theta_n \notin \{0, j\}$ então $\mathbf{C}(n) = \mathbf{C}(n-1)$;
 - (b) Se $\theta_n \in \{0, j\}$ então $\mathbf{C}(n) = \mathbf{C}(n-1) + e_j$.
3. Uma vez que $\mathbf{C}(n) = (1, \dots, 1)$, pare.

Note que o algoritmo acima necessariamente para, já que, a cada iteração n há uma probabilidade positiva e uniformemente longe de 0 de se acrescentar uma coordenada 1 ao vetor $\mathbf{C}(n)$. Definimos então:

$$T = \min\{n; \mathbf{C}(n) = (1, 1, \dots, 1)\}.$$

Na interpretação intuitiva do processo $\mathbf{C}(n)$ dada acima, pensamos em T como o primeiro tempo no qual todas as coordenadas foram verificadas.

O fato central aqui é que T é um tempo estacionário forte. Para demonstrar esse fato, começamos mostrando, que para cada n , se

$$\mathcal{B}_n = \{i \in \{0, 1, \dots, N\}: C_i(n) = 1\}$$

então cada suconjunto $B \subseteq \{1, 2, \dots, N\}$, condicionado em $\mathcal{B}_n = B$, o vetor $\{X_i(n); i \in B\}$ é uniformemente distribuído em $\{0, 1\}^B$. Isso pode ser feito por indução em n .

Iniciamos com o caso $n = 1$. Vamos supor, por simplicidade que $\theta_1 \neq 0$. Caso contrário, aguarde o primeiro instante de tempo n_o no qual $\theta_{n_o} = 1$ e reinicie a contagem do tempo a partir daí.

Então temos que $B_1 = B$ onde $B = \{j\}$ para algum $j \in \{1, \dots, N\}$ e temos que

$$X_j(1) = \mathbb{1}_{\{\xi_1=1\}} = \xi_1$$

é simplesmente uma variável aleatória de Bernoulli, ou seja, uniformemente distribuída no intervalo $\{0, 1\}$.

Agora, suponha o resultado válido para n . Separamos então o argumento em dois casos dependendo se \mathcal{B}_{n+1} coincide com \mathcal{B}_n ou não.

1. Se $\mathcal{B}_{n+1} = \mathcal{B}_n = B$, e $|B| \leq N - 2$, então temos as seguintes possibilidades:

(a) $\theta_{n+1} = 0$.

Neste caso, $\mathbf{X}(n+1) = \mathbf{X}(n)$, logo a hipótese de indução garante que $\{\mathbf{X}_i(n+1) : i \in B\}$ tem a distribuição uniforme em $\{0, 1\}^B$.

(b) $\theta_{n+1} = j \in B$.

Neste caso, $\mathbf{X}(n+1) = \mathbf{X}(n) + e_j \pmod{2}$, ou seja, a j -ésima coordenada de \mathbf{X} alterna enquanto as demais coordenadas permanecem fixas. Como $\{X_j(n) : j \in B\}$ possui a distribuição uniforme em $\{0, 1\}^B$ que é invariante pela operação de alternar uma coordenada em B , segue que $\mathbf{X}(n+1)$ também tem a distribuição uniforme em $\{0, 1\}^B$.

2. Se $\mathcal{B}_{n+1} = \mathcal{B}_n \cup \{j\} = B \cup \{j\}$ com $|B| \leq N - 2$ e $j \in \{1, \dots, N\} \setminus B$, então a variável $X_j(n+1)$ é obtida alternando-se ou mantendo-se o valor de $X_j(n)$ a depender se $\zeta_{n+1} = 1$ ou $\zeta_{n+1} = 0$, respectivamente. Mais precisamente, temos que

$$X_j(n+1) = (1 - X_j(n)) \mathbb{1}_{\{\zeta_{n+1}=1\}} + X_j(n) \mathbb{1}_{\{\zeta_{n+1}=0\}}$$

cuja distribuição é Bernoulli(1/2). Como as demais coordenadas não variam, $\{X_j(n+1) : j \in B\}$ é igual a $\{X_j(n) : j \in B\}$ que, pela a hipótese de indução é uniforme em $\{0, 1\}^B$ e independente de $X_j(n+1)$. Logo $\{X_j(n+1) : j \in B \cup \{j\}\}$ é uniformemente distribuído em $\{0, 1\}^{B \cup \{j\}}$.

3. Se $\mathcal{B}_{n+1} = \mathcal{B}_n = B$, e $|B| = N - 1$ então necessariamente $\theta_{n+1} \in B$ (porque, caso contrário, a última coordenada faltante seria incorporada a B e então teríamos $\mathcal{B}_{n+1} \neq \mathcal{B}_n = B$). Segue daí que $\mathbf{X}(n+1) = \mathbf{X}(n)$, logo a hipótese de indução garante que $\{\mathbf{X}_i(n+1) : i \in B\}$ tem a distribuição uniforme em $\{0, 1\}^B$.

4. Se $\mathcal{B}_{n+1} = \mathcal{B}_n \cup \{j\} = B \cup \{j\}$ com $|B| = N - 1$ e $j \in \{1, \dots, N\} \setminus B$, então temos que $\theta_{n+1} \in \{0, j\}$. A probabilidade condicional de que θ_{n+1} tenha cada um desses valores é 1/2. Segue daí que a distribuição condicional de $X_j(n+1)$ é Bernoulli(1/2), já que ela é obtida alternando o valor de $X_j(n)$ se $\theta_{n+1} = j$ e mantendo esse valor inalterado se $\theta_{n+1} = 0$. Como as demais coordenadas não variam, $\{X_j(n+1) : j \in B\}$ é igual a $\{X_j(n) : j \in B\}$ que, pela a hipótese de indução é uniforme em $\{0, 1\}^B$ e independente de $X_j(n+1)$. Logo $\{X_j(n+1) : j \in B \cup \{j\}\}$ é uniformemente distribuído em $\{0, 1\}^{B \cup \{j\}}$.

Para concluir que T é um tempo estacionário forte, basta notar que, no evento $\{T = t\}$ temos $\mathcal{B}_t = \{1, \dots, N\}$, logo $\mathbf{X}(T)$ é uniformemente distribuído em $\{1, \dots, N\}$, independentemente do valor de T .

Para estimar a distribuição de T , podemos escrever:

$$T = W_N + W_{N-1} + \dots + W_1,$$

onde W_k é o número mínimo de etapas durante as quais haviam k coordenadas ainda não verificadas. Então os W_k são independentes com distribuição geométrica e probabilidade de sucesso igual a

$$\begin{cases} P(W_1 = m) = \frac{2}{N+1} \left(1 - \frac{2}{N+1}\right)^{m-1} & \text{para todo } m \geq 1; \\ P(W_k = m) = \frac{k}{N+1} \left(1 - \frac{k}{N+1}\right)^{m-1} & \text{para todo } m \geq 1 \text{ e } k \geq 2. \end{cases}$$

Observação 2.5. *Observa-se que a variável aleatória T têm distribuição muito similar à variável aleatória do problema do colecionador (veja, por exemplo, [10, Seção 2.2, pág. 22]). Chamaremos tal variável aleatória de \tilde{T} .*

De fato, podemos considerar o problema clássico do colecionador com um álbum contendo $N+1$ figurinhas indexadas por $\{0, 1, \dots, N\}$, mas que já começa com a figurinha 0 presente. Então a distribuição do número mínimo de etapas durante as quais haviam k figurinhas faltantes é a mesma de W_k acima para $k = 2, \dots, N$. A única diferença é quando $k = 1$, caso no qual temos que substituir a distribuição acima por

$$P(\tilde{W}_1 = m) = \frac{1}{N+1} \left(1 - \frac{1}{N+1}\right)^{m-1} \quad \text{para todo } m \geq 1.$$

Assim, \tilde{T} domina estocasticamente T , significando que:

$$P(T > t) \leq P(\tilde{T} > t), \quad (2.7)$$

para todo $t > 0$.

Vamos agora cotar o valor esperado e a variância de T . Pela linearidade da esperança temos,

$$\begin{aligned} E(T) &= E(W_N) + E(W_{N-1}) + \dots + E(W_1) \\ &= \frac{N+1}{N} + \frac{N+1}{N-1} + \dots + \frac{N+1}{2} \\ &= \frac{N+1}{2} + \sum_{k=2}^N \frac{N+1}{k} \\ &= \frac{N+1}{2} + (N+1) \sum_{k=2}^N \frac{1}{k} \\ &\leq (N+1) + (N+1) \sum_{k=2}^N \frac{1}{k} \\ &\stackrel{(*)}{\leq} (N+1) + (N+1) \log(N) \\ &= (N+1)(1 + \log(N)), \end{aligned}$$

onde na desigualdade marcada com (*), usamos o fato que $\log N - 1 \leq \sum_{k=2}^N \frac{1}{k} \leq \log N$.

Como as W_k são independentes então a variância é linear. Assim, podemos calcular:

$$\begin{aligned} \text{Var}(T) &= \left(\frac{N+1}{2}\right)^2 \left(1 - \frac{2}{N+1}\right) + \sum_{k=2}^N \left[\left(\frac{N+1}{k}\right)^2 \left(1 - \frac{k}{N+1}\right)\right] \\ &= \frac{N^2 - 1}{4} + (N+1)^2 \sum_{k=2}^N \frac{1}{k^2} - (N+1) \sum_{k=2}^N \frac{1}{k} \\ &\leq \frac{N^2 - 1}{4} + (N+1)^2 \sum_{k=2}^N \frac{1}{k^2} \\ &\leq (N+1)^2. \end{aligned}$$

Pela desigualdade de Chebyshev, para $c > 1$, temos:

$$\begin{aligned} P(T \geq (N+1)(\log(N) + c)) &\leq P\left(|T - (N+1)(\log(N) + 1)| \geq (c-1)(N+1)\right) \\ &\leq \frac{(N+1)^2}{(c-1)^2(N+1)^2} \\ &= \frac{1}{(c-1)^2}. \end{aligned}$$

Logo para um passeio aleatório no hipercubo, obtemos da Proposição 2.4(a) uma cota superior para $s_x(n)$:

$$s_x((N+1)(\log(N) + c)) \leq P(T \geq (N+1)(\log(N) + c)) \leq \frac{1}{(c-1)^2}.$$

O resultado anterior pode ser melhorado ainda mais, comparando, para N grande, $P(T > (N+1)(\log(N) + c))$ e e^{-c} .

Para isso, voltemos à comparação com o problema do colecionador. Por (2.7), temos que

$$P(T > \lceil (N+1)(\log(N) + c) \rceil) \leq P(\tilde{T} > \lceil (N+1)(\log(N) + c) \rceil). \quad (2.8)$$

Seja A_i o evento que a i -ésima figurinha não aparece durante as primeiras $N(\log(N) + c)$ etapas. E

$$P(\tilde{T} > \lceil (N+1)(\log(N) + c) \rceil) = P\left(\bigcup_{i=0}^N A_i\right) \leq \sum_{i=0}^N P(A_i).$$

Como cada etapa tem probabilidade $1 - \frac{1}{N+1}$ de não retirar a figurinha i e as etapas são independentes temos:

$$\begin{aligned} \sum_{i=0}^N P(A_i) &\leq \sum_{i=0}^N \left(1 - \frac{1}{N+1}\right)^{\lceil (N+1)(\log(N) + c) \rceil} \\ &= (N+1) \left(1 - \frac{1}{N+1}\right)^{\lceil (N+1)(\log(N) + c) \rceil} \\ &\leq (N+1) \exp\left(\frac{-(N+1)(\log(N) + c)}{N+1}\right) = \frac{N+1}{N} e^{-c}. \end{aligned}$$

Com isso temos que

$$P(\tilde{T} > (N+1)(\log(N) + c)) \leq 2e^{-c},$$

Então, pela Equação (2.8), a cota é bem próxima de

$$P(T > (N+1)(\log(N) + c)) \leq 2e^{-c}.$$

E, portanto temos

$$s_x((N+1)(\log(N) + c)) \leq 2e^{-c}.$$

Como pela relação (1.20), a s_x (1.19) é uma cota superior para d (1.11) então

$$d((N + 1)(\log(N) + c)) \leq 2e^{-c}.$$

Portanto, como a distribuição estacionária do passeio no hipercubo é a medida uniforme sobre todos os vértices, como vimos nessa seção, então para estar suficientemente perto dessa medida, próximo ao "equilíbrio", o tempo necessário de evolução da cadeia é de cerca de $(N + 1) \log N$ etapas. Se esperarmos menos que este tempo, ainda seria possível reconhecer de qual posição a cadeia se iniciou e a distribuição uniforme não será bem aproximada, enquanto aguardar um tempo maior mais não é útil.

Capítulo 3

Acoplamento

Acoplamento é uma técnica padrão, amplamente utilizada em diversos ramos da probabilidade. Acoplar consiste em construir simultaneamente dois processos ou, mais geralmente, duas medidas de probabilidade em um mesmo espaço. De maneira informal, busca-se encontrar a *melhor* forma de fazê-lo. Por exemplo, para estudar a velocidade de convergência ao equilíbrio de uma cadeia de Markov, construímos simultaneamente duas trajetórias e estimamos o tempo que elas levam para se encontrar. Esse tempo depende da lei conjunta dos dois caminhos percorridos, então busca-se encontrar a construção conjunta de duas trajetórias que minimiza, em certo sentido, esse tempo. Para saber mais sobre acoplamento e suas aplicações em probabilidade veja [15].

3.1 Acoplamento e Distância de Variação Total

Iniciamos essa seção com a definição precisa do que é um acoplamento.

Definição 3.1. *Um acoplamento de duas distribuições de probabilidade μ e ν é um par de variáveis aleatórias X, Y definidas no mesmo espaço de probabilidade, de modo que as distribuições marginais de X e de Y sejam μ e ν , respectivamente, ou seja,*

$$P(X = x) = \mu(x) \quad e \quad P(Y = y) = \nu(y).$$

Em geral, existem várias maneiras de acoplar duas distribuições de probabilidade.

Exemplo 3.2. *Suponha que $\mu = \nu$. Um acoplamento de μ e ν consiste em tomar X e Y como variáveis aleatórias independentes com distribuição μ . Outro possível acoplamento consiste em tomar $X = Y$.*

Intuitivamente falando, os acoplamentos do exemplo anterior são triviais e não devem ser *bons o suficiente* para o propósito de estudar a distribuição de probabilidade evolvida.

Quaisquer duas distribuições têm um acoplamento independente. No entanto, quando essas distribuições não são idênticas, não é possível que as variáveis X e Y tenham sempre o mesmo valor. Então quão perto pode um acoplamento chegar de ter X e Y idênticos?

Definição 3.3. *Dizemos que um acoplamento (X, Y) de μ e ν é ótimo se*

$$\| \mu - \nu \|_{TV} = \max_{x, y \in \Omega} P_{x, y}(X \neq Y)$$

onde $P_{x,y}$ representa a probabilidade no espaço onde ambas as distribuições estão definidas.

O próximo resultado, mostra que a distância de variação total fornece a resposta e motiva a procura por um *acoplamento ótimo* além de proporcionar um método para obter cotas superiores para $d(n)$.

Proposição 3.4. *Sejam μ e ν duas distribuições de probabilidade em Ω . Então*

$$\|\mu - \nu\|_{TV} = \inf \{P(X \neq Y) : (X, Y) \text{ é um acoplamento de } \mu \text{ e } \nu\}. \quad (3.1)$$

Demonstração. Seja (X, Y) um acoplamento de μ e ν . Para qualquer $A \subseteq \Omega$ temos:

$$\begin{aligned} \mu(A) - \nu(A) &= P(X \in A) - P(Y \in A) \\ &= P(X \in A, X = Y) + P(X \in A, X \neq Y) \\ &\quad - P(Y \in A, X = Y) - P(Y \in A, X \neq Y) \\ &= P(X \in A, X \neq Y) - P(Y \in A, X \neq Y) \leq P(X \neq Y). \end{aligned}$$

Trocando-se o papel de μ e ν vale que $\nu(A) - \mu(A) \leq P(X \neq Y)$. Portanto,

$$|\mu(A) - \nu(A)| \leq P(X \neq Y).$$

Maximizando sobre A obtemos

$$\|\mu - \nu\|_{TV} \leq \inf \{P(X \neq Y) : (X, Y) \text{ é um acoplamento de } \mu \text{ e } \nu\}.$$

Para mostrar a direção oposta, construiremos um acoplamento para o qual $P(X \neq Y)$ assume exatamente o valor da distância de variação total.

Seja $p = \sum_{x \in \Omega} \mu(x) \wedge \nu(x)$, onde $\mu(x) \wedge \nu(x) = \min\{\mu(x), \nu(x)\}$. Lançamos uma moeda com probabilidade de cara igual a p . Se a moeda der cara, então escolhemos o valor Z de acordo com a distribuição de probabilidade

$$\gamma_{III}(x) = \frac{\mu(x) \wedge \nu(x)}{p} \quad (3.2)$$

e definimos $X = Y = Z$. Se a moeda der coroa, então escolhemos X de acordo com a distribuição

$$\gamma_I(x) = \frac{\mu(x) - \nu(x)}{1 - p} \mathbb{1}_{(\mu(x) > \nu(x))}, \quad (3.3)$$

e Y de acordo com

$$\gamma_{II}(x) = \frac{\nu(x) - \mu(x)}{1 - p} \mathbb{1}_{(\nu(x) > \mu(x))}. \quad (3.4)$$

Note que

$$P(X = x) = p \cdot \frac{\mu(x) \wedge \nu(x)}{p} + (1 - p) \cdot \frac{\mu(x) - \nu(x)}{1 - p} \mathbb{1}_{(\mu(x) > \nu(x))} = \mu(x),$$

$$P(Y = x) = p \cdot \frac{\mu(x) \wedge \nu(x)}{p} + (1 - p) \cdot \frac{\nu(x) - \mu(x)}{1 - p} \mathbb{1}_{(\nu(x) > \mu(x))} = \nu(x).$$

Logo (X, Y) é um acoplamento de μ e ν .

Note que, caso a moeda seja coroa, $X \neq Y$, pois γ_I e γ_{II} são positivos em subconjuntos disjuntos de Ω . Assim $X = Y$ se, e somente se, o lançamento da moeda for cara. Usando a Equação (1.10) concluímos que

$$P(X \neq Y) = 1 - p = 1 - \sum_{x \in \Omega} \mu(x) \wedge \nu(x) = \sum_{\substack{x \in \Omega \\ \mu(x) > \nu(x)}} [\mu(x) - \nu(x)] = \|\mu - \nu\|_{TV}.$$

□

Até agora definimos o que é um acoplamento de variáveis aleatórias. Podemos também acoplar cadeias de Markov inteiras para obter informações sobre a convergência para o equilíbrio.

Definição 3.5. *Um acoplamento de duas cadeias de Markov com matriz de transição P e pontos iniciais x_o e y_o é uma sequência $(X_t, Y_t)_{t=0}^{\infty}$ tal que $(X_t)_{t=0}^{\infty}$ e $(Y_t)_{t=0}^{\infty}$ são cadeias de Markov com matriz de transição P , com $X_0 = x_o$, $Y_0 = y_o$ quase certamente.*

Se $(X_t, Y_t)_{t=0}^{\infty}$ é um acoplamento de Cadeias de Markov X_t e Y_t iniciando nos estados x_o e y_o , a notação P_{x_o, y_o} representa a probabilidade no espaço onde ambas as cadeias estão definidas.

Qualquer acoplamento (X_t, Y_t) de cadeias de Markov com a matriz de transição P pode ser modificado para que as duas cadeias marginais evoluam juntas após sua primeira visita simultânea a um dado estado. Assim, vamos supor a partir de agora que, para todo acoplamento vale

$$\text{se } X_s = Y_s \text{ então } X_t = Y_t \quad \forall t \geq s. \quad (3.5)$$

Vamos definir τ_{coup} como o primeiro tempo em que as cadeias X_t e Y_t se encontram, ou seja,

$$\tau_{coup} = \min\{t; X_t = Y_t\}. \quad (3.6)$$

A importância de τ_{coup} reside na seguinte:

Teorema 3.6. *Seja (X_t, Y_t) um acoplamento de cadeias de Markov com a matriz de transição P . Então*

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

Demonstração. Seja $X_0 = x$ e $Y_0 = y$, observe que $P^t(x, z) = P_{x,y}\{X_t = z\}$ e $P^t(y, z) = P_{x,y}\{Y_t = z\}$ o que implica que (X_t, Y_t) é um acoplamento de $P^t(X, \cdot)$ com $P^t(y, \cdot)$. Pela Proposição 3.4 temos que

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

Como $P_{x,y}\{X_t \neq Y_t\} = P_{x,y}\{\tau_{coup} > t\}$ temos que

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

□

Corolário 3.7. *Suponha que para cada par de estados $x, y \in \Omega$ existe um acoplamento $(X_t, Y_t)_{t \geq 0}$ tal que $X_0 = x$ e $Y_0 = y$. Então*

$$d^*(t) \leq \max_{x, y \in \Omega} P_{x, y} \{ \tau_{coup} > t \} ; \forall t \geq 1, \quad (3.7)$$

onde τ_{coup} é definido em (3.6). Por outro lado, existe um acoplamento tal que a (3.7) vale com igualdade.

Demonstração. Pela definição de distribuição estacionária 1.6 e usando o Lema 1.17

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

Tomando o máximo também sob todos os valores de x , temos

$$d^*(t) \leq \max_{x, y \in \Omega} \| (P^t(x, A) - P^t(y, A)) \|_{TV} . \quad (3.8)$$

Usando a Equação (3.8) e o Teorema 3.6 temos

$$d^*(t) \leq \max_{x, y \in \Omega} P_{x, y} \{ \tau_{coup} > t \} ; \text{ para todo } t \geq 1.$$

A igualdade da Equação (3.7) foi mostrada em [9, Teorema 4, pág 100] em uma extensa demonstração e em [12] de maneira mais simplificada. □

A Proposição 2.4 é intimamente análoga ao Corolário 3.7, a técnica de tempo uniforme forte se ajusta com a distância de separação exatamente da mesma maneira que a técnica de acoplamento se ajusta com a distância de variação total. A próxima proposição mostrará que o Tempo Estacionário Forte é um caso particular do Acoplamento.

Proposição 3.8. *Seja T um Tempo Estacionário Forte para uma Cadeia de Markov (X_n) com estado inicial i_0 . Então existe um acoplamento com tempo de acomodamento $\tau_{coup} = T$.*

Demonstração. Sejam T e (X_n) como no enunciado. Vamos construir uma nova cadeia (Y_n) da seguinte forma: para cada m , no evento $\{T = m\}$, definimos $Y_n^{(m)} = X_n$ para todo $n \geq m$. Condicionando em $\{T = m\}$, o processo futuro da cadeia $\{Y_n^{(m)} : n \geq m\}$ têm distribuição estacionária (pois T é um tempo estacionário forte). Portanto, essa cadeia pode ser “extendida para trás”, ou seja, podemos definir $\{Y_n^{(m)} : 0 \leq n < m\}$ como sendo a cadeia estacionária. Definimos assim

$$Y_n = Y_n^{(m)}, \text{ em } \{T = m\}, \forall m.$$

Então (Y_n) é a cadeia de Markov estacionária e $Y_n = X_n$ se $n \geq T$. Isso nos fornece o acoplamento procurado. \square

Já conhecendo acoplamentos, podemos mostrar uma propriedade importantes de $\bar{d}(t)$ definida na Equação (1.13).

Lema 3.9. $\bar{d}(t)$ é submultiplicativa, ou seja, $\bar{d}(s+t) \leq \bar{d}(s)\bar{d}(t)$ para todo $t, s \geq 0$.

Demonstração. Fixemos x e y e seja (X, Y) o acoplamento ótimo de $P^s(x, \cdot)$ e de $P^s(y, \cdot)$, ou seja,

$$\| P^s(x, \cdot) - P^s(y, \cdot) \|_{TV} = P(X \neq Y).$$

Pela Proposição 1.16, temos que

$$\| P^s(x, \cdot) - P^s(y, \cdot) \|_{TV} = \frac{1}{2} \sum_z | P^{s+t}(x, z) - P^{s+t}(y, z) |. \quad (3.9)$$

Como $P^{s+t} = P^s P^t$, ou seja, o produto das matrizes e a distribuição de X_s é dada por $P^s(x, \cdot)$ então

$$P^{s+t}(x, w) = \sum_z P^s(x, z) P^t(z, w) = \sum_z P(X_s = z) P^t(z, w) = E(P^t(X_s, w)).$$

Analogamente, temos que $P^{s+t}(y, w) = E(P^t(Y_s, w))$, combinando-as e usando a linearidade da esperança

$$P^{s+t}(x, w) - P^{s+t}(y, w) = E(P^t(X_s, w) - P^t(Y_s, w)).$$

Substituindo na Equação (3.9), temos

$$\begin{aligned} \| P^s(x, \cdot) - P^s(y, \cdot) \|_{TV} &= \frac{1}{2} \sum_{w \in \Omega} | E(P^t(X_s, w) - P^t(Y_s, w)) | \\ &\leq E \left(\frac{1}{2} \sum_{w \in \Omega} | (P^t(X_s, w) - P^t(Y_s, w)) | \right) \\ &= E \left(\| (P^t(X_s, w) - P^t(Y_s, w)) \|_{TV} \right) \\ &= E \left(\| (P^t(X_s, w) - P^t(Y_s, w)) \|_{TV} \mathbb{1}_{(X_s \neq Y_s)} \right) \\ &\leq E(\bar{d}(t) \mathbb{1}_{(X_s \neq Y_s)}) \\ &= \bar{d}(t) P(X_s \neq Y_s) = \bar{d}(t) \bar{d}(s). \end{aligned}$$

Maximizando sobre x e y completamos a prova. \square

3.2 Passeio Aleatório no Hipercubo: Acoplamento

Vamos agora mostrar como usar acoplamento para obter cotas para a distância de variação total para o passeio aleatório no hipercubo N -dimensional. Começamos com um acoplamento bastante simples que envolve dois passeios começando de posições diferentes

no hipercubo. Depois apresentaremos uma construção mais complicada que envolve urnas de Ehrenfest.

Por conveniência, nessa seção vamos considerar uma versão do passeio aleatório preguiçoso ligeiramente diferente da considerada em (2.6): a cada vez que o partícula fica onde está com probabilidade $\frac{1}{2}$ e se move para um vizinho aleatório mais próximo com probabilidade $\frac{1}{2N}$. Isso dá a matriz de transição entre dois estados $\mathbf{x} = (x_1, \dots, x_N)$ e $\mathbf{y} = (y_1, \dots, y_N)$:

$$P(\mathbf{x}, \mathbf{y}) = \begin{cases} \frac{1}{2}, & \mathbf{x} = \mathbf{y}; \\ \frac{1}{2N}, & \sum_s |x_s - \bar{x}_s| = 1; \\ 0, & \text{caso contrário.} \end{cases} \quad (3.10)$$

Proposição 3.10. *Para o passeio aleatório no hipercubo definido como em (3.10) vale que $\forall c > 0$*

$$d(N \log N + cN) \leq e^{-c}.$$

Demonstração. Iniciamos com uma sequência de variáveis aleatórias independentes $\{\theta_t\}_t$ uniformemente distribuídas em $\{1, \dots, N\}$. Definamos também uma segunda sequência de variáveis aleatórias independentes ζ_n com distribuição Bernoulli(1/2).

Tomemos dois pontos \mathbf{x} e \mathbf{y} no hipercubo. A partir desses pontos geramos dois processos $\mathbf{X} = (\mathbf{X}_t)$ e $\mathbf{Y} = (\mathbf{Y}_t)$ no hipercubo da seguinte maneira: no tempo t , atualizamos simultaneamente a coordenada θ_t no dois processos, substituindo ambos os valores por ζ_t .

Cada um dos processos \mathbf{X} e \mathbf{Y} tem a distribuição de um passeio no hipercubo dado como em (3.10). Além disso, note que uma vez que uma coordenada tenha sido atualizada, os valores desta coordenada nos dois passeios será sempre a mesma. Assim um tempo de acoplamento τ_{coup} será dado pelo primeiro momento em que todas as coordenadas tenham sido atualizadas. Esse tempo tem então a mesma distribuição da variável aleatória do problema do colecionador \tilde{T} .

Consideramos A_i o evento em que a i -ésima coordenada não aparece durante as primeiras $N(\log(N) + c)$ etapas. Temos que para todo $i \in \{1, 2, \dots, N\}$

$$P(A_i) = \left(1 - \frac{1}{N}\right)^{\lceil N(\log(N)+c) \rceil} \leq e^{-\lceil \log(N)+c \rceil}.$$

Então

$$P(\tau_{coup} > \lceil N(\log(N) + c) \rceil) = P\left(\bigcup_{i=1}^N A_i\right) \leq \sum_{i=1}^N P(A_i) \leq N e^{-\lceil \log(N)+c \rceil} = e^{-c}.$$

Do Corolário 3.7 e da equação acima temos que

$$d(N(\log(N) + c)) = P(\tau_{coup} > \lceil N(\log(N) + c) \rceil) \leq e^{-c}.$$

□

Veremos que a ordem da cota superior para a distância de variação total $N \log N$ pode ser melhorada por um fator constante de $\frac{1}{2}$ como será mostrado mais adiante nessa seção.

Consideramos agora um exemplo, a urna de Ehrenfest, que à primeira vista parece bem diferente do passeio aleatório no hipercubo, mas que nos ajudará a entender o seu comportamento.

Exemplo 3.11. (*Urna de Ehrenfest*) Imagine que temos duas urnas, Urna I e Urna II, dentro das quais são distribuídas N bolas. Em cada passo do processo escolhemos ao acaso e uniformemente umas das N bolas e a transferimos da urna atual para a outra urna. Se chamarmos U_t o número de bolas na Urna I no tempo t , então U_t é a cadeia de Markov no espaço de estados $\Omega := \{0, 1, \dots, N\}$ cuja matriz de transição é

$$P(j, k) = \begin{cases} \frac{N-j}{N}, & \text{se } k = j + 1; \\ \frac{j}{N}, & \text{se } k = j - 1; \\ 0, & \text{caso contrário.} \end{cases} \quad (3.11)$$

Se deixarmos uma probabilidade de $\frac{1}{2}$ para as bolas pararem em cada etapa ($P(j, j) = \frac{1}{2}$), obteremos uma versão preguiçosa da cadeia, ou seja,

$$Q(j, k) = \begin{cases} \frac{N-j}{2N}, & \text{se } k = j + 1; \\ \frac{j}{2N}, & \text{se } k = j - 1; \\ \frac{1}{2}, & \text{caso contrário.} \end{cases} \quad (3.12)$$

A distribuição estacionária para ambas as cadeias (preguiçosa Q e não preguiçosa P), é uma distribuição binomial de parâmetros N e $\frac{1}{2}$. De fato, seja $\pi(k) = \binom{N}{k} \left(\frac{1}{2}\right)^N$, $j \in \Omega$ e usando a (1.7).

Para a cadeia P temos

$$\begin{aligned} \pi P(y) &= \pi(y-1)P(y-1, y) + \pi(y+1, y)P(y+1, y) \\ &= \binom{N}{y+1} \left(\frac{1}{2}\right)^N \left[\frac{y-1}{N}\right] + \binom{N}{y-1} \left(\frac{1}{2}\right)^N \left[\frac{N-y+1}{N}\right] \\ &= \left(\frac{1}{2}\right)^N \left[\binom{N-1}{y-1} + \binom{N-1}{y} \right] \\ &= \left(\frac{1}{2}\right)^N \binom{N}{y} \\ &= \pi(y) \end{aligned} \quad (3.13)$$

Já para a cadeia Q temos

$$\begin{aligned} \pi Q(y) &= \pi(y-1)Q(y-1, y) + \pi(y)Q(y, y) + \pi(y+1, y)P(y+1, y) \\ &= \frac{1}{2}\pi(y-1)P(y-1, y) + \frac{1}{2}\pi(y) + \frac{1}{2}\pi(y+1, y)P(y+1, y) \\ &= \pi(y). \end{aligned} \quad (3.14)$$

Logo a distribuição invariante corresponde a colocar cada bola independentemente na Urna I e Urna II, com igual probabilidade. Então podemos nos perguntar qual é o elo entre essa cadeia, chamada de processo de urna Ehrenfest, e o passeio aleatório no hipercubo?

Para um vértice do hipercubo, podemos definir seu peso de Hamming $W(x)$ como um vetor $x := (x_1, \dots, x_N) \in \{0, 1\}^N$ que é o número de 1's que aparecem no vetor x :

$$W(x) = \sum_{j=1}^N x_j. \quad (3.15)$$

Se X_t é o passeio aleatório (não preguiçoso), então podemos estudar sua projeção (em um sentido que é definido precisamente em [10, Seção 2.3.1, pág. 24]). Seja

$$W_t = W(X_t), \quad (3.16)$$

o peso de Hamming da posição do passeio no tempo t .

Quando $W_t = j$, o peso é incrementado por um valor unitário se uma das $N - j$ coordenadas 0 é selecionada, enquanto diminui uma unidade se uma das j coordenadas 1 for escolhida. Conclui-se que W_t é novamente uma cadeia de Markov e que suas probabilidades de transição também são descritas por (3.11). Neste paralelo, a j -ésima coordenada do passeio aleatório X_t pode ser pensada como a j -ésima bola do processo das urnas: se o j -ésimo bit é 1, então a bola está na Urna I, enquanto se o bit é 0 a bola está na Urna II.

O lema a seguir fornece outra conexão importante entre os dois modelos. Isso nos permitirá lidar com a urna de Ehrenfest em vez do passeio aleatório mais complicado (preguiçoso) no hipercubo, a fim de encontrar a cota superior para a distância de variação total. Assim, W_t é a versão preguiçosa da cadeia da Ehrenfest. Escrevemos π_W para a distribuição estacionária de W_t . Antes introduziremos uma definição que será essencial para a prova desse Lema.

Definição 3.12. Uma cadeia de Markov é transitiva se para cada par $(x, y) \in (\Omega \times \Omega)$ existe uma bijeção $\phi = \phi_{(x,y)} : \Omega \mapsto \Omega$ tal que

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

Lema 3.13. Seja X_t o passeio aleatório simples no hipercubo e seja $W_t := \sum_{j=1}^N X_t(j)$ o peso de Hamming da cadeia para cada passo (correspondendo a cadeia da urna Ehrenfest). Então

$$\| P_1(X_t \in \cdot) - \pi \|_{TV} = \| P_N(W_t \in \cdot) - \pi_W \|_{TV}, \quad (3.17)$$

onde P_1 diz que estamos iniciando o passeio aleatório pelo vértice com todas as coordenadas 1 e P_N diz que estamos iniciando a cadeia das urnas com todas as bolas na Urna I.

Observação 3.14. Uma consequência da propriedade de transitividade de uma cadeia é que, para calcular a distância de variação total da cadeia a partir de sua distribuição estacionária, podemos deixar a cadeia começar em qualquer ponto de Ω .

Demonstração. Defina $\Omega_W := \{x \in \Omega : W(x) = w\}$. Por simetria, ambas funções $x \mapsto P_1\{X_t = x\}$ e $x \mapsto \pi(x)$ são funções constantes em Ω_W . Então

$$\begin{aligned} \sum_{\substack{x \in \Omega, \\ W(x)=w}} |P_1(X_t = x) - \pi(x)| &= \left| \sum_{\substack{x \in \Omega, \\ W(x)=w}} P_1(X_t = x) - \pi(x) \right| \\ &= |P_1(W_t = w) - \pi_W(w)|, \end{aligned} \quad (3.18)$$

onde usamos na primeira igualdade o fato de que todos os termos na primeira soma serem iguais. Somando todos os possíveis $w \in \{0, 1, \dots, N\}$ e dividindo por 2 obtemos a Equação (3.17). \square

Proposição 3.15. *Para o passeio aleatório preguiçoso no hipercubo N -dimensional existe uma constante $c > 0$ tal que*

$$d\left(\frac{1}{2}N \log N + \alpha N\right) \leq \frac{c}{\sqrt{\alpha}}. \quad (3.19)$$

Demonstração. Como a cadeia é transitiva, pelo Lema 3.13, será suficiente limitar o lado direito da Equação (3.17), para isso, vamos usar o método do acoplamento.

Vamos construir duas cópias do modelo da urna de Ehrenfest preguiçosa, W_t e Z_t , partindo de diferentes pontos de Ω_W , $W_0 = w$ e $Z_0 = z$ com $z \geq w$, sem perda de generalidade. Em cada etapa, lançamos uma moeda justa para decidir qual das duas cópias executará um movimento de acordo com as probabilidades descritas na matriz (3.11). Dessa forma, olhando separadamente para W_t e Z_t , veremos duas cópias da versão preguiçosa da urna de Ehrenfest. Além disso, uma vez que as duas cadeias se encontrem, podemos forçá-las a ficarem juntas para sempre.

Denotaremos agora a diferença do número de bolas na Urna I nas duas cadeias, por

$$D_t = Z_t - W_t, \quad (3.20)$$

tal que $D_t \geq 0$ para todo t por construção.

Seja τ o primeiro tempo em que as duas cópias se encontram

$$\tau := \min\{t \geq 0 : Z_t = W_t\}. \quad (3.21)$$

Suponha $(Z_t, W_t) = (z_t, w_t)$, enquanto $\tau > t$

$$D_{t+1} - D_t = \begin{cases} 1, & \text{com probabilidade } \left(\frac{1}{2}\right)\left(1 - \frac{z_t}{N}\right) + \left(\frac{1}{2}\right)\frac{w_t}{N}; \\ -1, & \text{com probabilidade } \left(\frac{1}{2}\right)\left(\frac{z_t}{N}\right) + \left(\frac{1}{2}\right)\left(1 - \frac{w_t}{N}\right). \end{cases} \quad (3.22)$$

Vemos então que no evento $\{\tau > t\}$

$$\begin{aligned} E_{z,w}[D_{t+1} - D_t | Z_t = z_t, W_t = w_t] &= 1 \cdot \left(\frac{1}{2}\left(1 - \frac{Z_t}{N}\right) + \frac{1}{2}\frac{W_t}{N}\right) + \\ &\quad (-1) \cdot \left(\frac{1}{2}\frac{Z_t}{N} + \frac{1}{2}\left(1 - \frac{W_t}{N}\right)\right) \\ &= -\frac{z_t - w_t}{N} \\ &= -\frac{d_t}{N}. \end{aligned} \quad (3.23)$$

Seja $\mathbf{Z}_t = (Z_1, \dots, Z_t)$ e $\mathbf{W}_t = (W_1, \dots, W_t)$. A partir disso, o fato de $\mathbb{1}_{(\tau > t)}$ depender apenas da história da cadeia até o tempo t e da Propriedade de Markov (Definição 1.4), temos

$$\mathbb{1}_{(\tau>t)} E_{z,w}[(D_{t+1} - D_t) | \mathbf{Z}_t, \mathbf{W}_t] = E_{z,w}[\mathbb{1}_{(\tau>t)}(D_{t+1} - D_t) | \mathbf{Z}_t, \mathbf{W}_t]. \quad (3.24)$$

Combinando as Equações (3.23) e (3.24)

$$E_{z,w}[\mathbb{1}_{(\tau>t)}(D_{t+1}) | \mathbf{Z}_t, \mathbf{W}_t] = \left(1 - \frac{1}{N}\right) \mathbb{1}_{(\tau>t)} D_t.$$

Tomando a esperança sobre todos os caminhos possíveis de Z e W

$$E_{z,w}[\mathbb{1}_{(\tau>t)} D_{t+1}] = \left(1 - \frac{1}{N}\right) E_{z,w}[\mathbb{1}_{(\tau>t)} D_t]. \quad (3.25)$$

Como $\mathbb{1}_{(\tau>t+1)} \leq \mathbb{1}_{(\tau>t)}$ temos que

$$E_{z,w}[\mathbb{1}_{(\tau>t+1)} D_{t+1}] \leq \left(1 - \frac{1}{N}\right) E_{z,w}[\mathbb{1}_{(\tau>t)} D_t]. \quad (3.26)$$

Por indução,

$$E_{z,w}[\mathbb{1}_{(\tau>t)} D_t] \leq \left(1 - \frac{1}{N}\right)^t (z - w) \leq N e^{-\frac{t}{N}}. \quad (3.27)$$

Agora note que se $\tau > t$, o processo (D_t) têm pelo menos a mesma probabilidade de se mover para baixo do que para cima. Assim, até o tempo τ , o processo (D_t) pode ser acoplado a um passeio aleatório simples (S_t) . Podemos supor $S_0 = z - w$ e forçar S_t a dominar D_t no sentido de que sempre que S_t dá um passo para a esquerda, obrigamos D_t a fazer o mesmo; dessa forma para todo $t \leq \tau$, $S_0 = D_0$ e $S_t \geq D_t$.

Seja $\tilde{\tau} := \min\{t \geq 0 : S_t = 0\}$. Então, pelo que dissemos, $\tau \leq \tilde{\tau}$ e pelo [10, Teorema 2.26 e Corolário 2.28, pág. 35 e 36], existe $c > 0$ tal que para $z - w \geq 0$

$$P_{z-w}(\tau > u) \leq P_{z-w}(\tilde{\tau} > u) \leq \frac{c(z-w)}{\sqrt{u}}. \quad (3.28)$$

Portanto,

$$P_{z,w}(\tau > s + u | D_0, D_1, \dots, D_s) = \mathbb{1}_{(\tau>s)} P_{D_s}(\tau > u) \leq \mathbb{1}_{(\tau>s)} \frac{c D_s}{\sqrt{u}}. \quad (3.29)$$

Tomando a esperança e aplicando (3.27)

$$P_{z,w}(\tau > s + u) = \frac{c N e^{-s/N}}{\sqrt{u}}. \quad (3.30)$$

Escolhendo $s = \frac{1}{2} N \log N$ e $u = \alpha N$ obtemos

$$P_{z,w}(\tau > \frac{1}{2} N \log(N) + \alpha N) \leq \frac{c}{\sqrt{\alpha}}. \quad (3.31)$$

Pelo Corolário 3.7 temos

$$d\left(\frac{1}{2} N \log N + \alpha N\right) \leq \frac{c}{\sqrt{\alpha}}.$$

□

Assim, um pouco mais de $\frac{1}{2} N \log N$ passos são suficientes para que o passeio aleatório de Ehrenfest atinja o equilíbrio. Portanto, como a distribuição estacionária da cadeia é a medida uniforme sobre todos os vértices do hipercubo então para estar suficientemente perto dessa medida temos que evoluir a cadeia por cerca de $\frac{1}{2} N \log N$ etapas.

Capítulo 4

Propriedades Simétricas

Supomos até agora que as matrizes de transição $P(x, y)$, com $x, y \in \Omega$, são ir-reduzíveis e aperiódicas. Além disso, vários exemplos, como o passeio no hipercubo ou no n -ciclo, compartilham propriedades de simetria importantes. Nessa seção provaremos resultados teóricos que valem sob as hipóteses de simetria. Este capítulo segue [8] e [10].

4.1 Definições e Consequências Elementares

Um primeiro resultado importante e muito útil está relacionado a simetria da matriz de transição.

Proposição 4.1. *Se P é uma matriz de transição simétrica, ou seja, $P(x, y) = P(y, x)$ então a distribuição uniforme é a estacionária para P .*

Demonstração. Seja $|\Omega| = n$ e $\pi(y) = \frac{1}{n}$ para todo $y \in \Omega$, então

$$\sum_y \pi(y)P(y, x) = \frac{1}{n} \sum_j P(x, y) = \frac{1}{n} = \pi(x).$$

□

Definição 4.2. *P é duplamente estocástica (DE) se $\sum_i P(x, y) = 1$ para todo $y \in \Omega$.*

Esta é precisamente a condição sob a qual a distribuição estacionária π é a distribuição uniforme $U(i) = \frac{1}{|\Omega|} \forall i \in \Omega$. Mais que isso, vamos mostrar que são equivalentes.

Proposição 4.3. *P é duplamente estocástica se, e somente se, a distribuição uniforme é a estacionária para P .*

Demonstração. Suponha que $|\Omega| = n$ e denote $\pi(i) = 1/n$ para todo $i \in \Omega$. Então π é a distribuição uniforme em Ω . Denote $\mathbf{1} = (1, 1, \dots, 1)$. Então:

$$P \text{ é DE} \Leftrightarrow \mathbf{1}P = \mathbf{1} \Leftrightarrow \frac{1}{n}\mathbf{1}P = \frac{1}{n}\mathbf{1} \Leftrightarrow \pi P = \pi.$$

□

Definição 4.4. *P é semi-simétrica se existe uma bijeção $\xi : \Omega \rightarrow \Omega$ tal que $P(x, y) = P(\xi(y), \xi(x))$ para todo $x, y \in \Omega$.*

Podemos ver facilmente que uma matriz semi-simétrica é também DE. De fato, supomos que P é semi-simétrica, então existe uma bijeção tal que $\psi : \Omega \mapsto \Omega$ tal que $P(x, y) = P(\psi(y), \psi(x)) \forall x, y \in \Omega$ então

$$\sum_x^{|\Omega|} P(x, y) = \sum_x^{|\Omega|} P(\psi(y), \psi(x)) \stackrel{(1)}{=} \sum_k^{|\Omega|} P(\psi(y), k) \stackrel{(2)}{=} 1,$$

onde em (1) usamos que a soma percorre todos os i's, pois ψ é a bijeção e (2) usamos o fato da matriz ser estocástica.

Definimos o que seria uma cadeia transitiva em 3.12. Informalmente, significa que a cadeia “parece a mesma” de qualquer ponto no espaço de estados, ou seja, todos os estados iniciais são equivalentes. Dessa forma $d^*(n)$ e $s_x(n)$ não dependem dos estados iniciais, então $d(n) = d^*(n)$ e $s_x(n) = s^*(n)$.

4.2 Consequências das Condições Simétricas

Nesta seção vamos mostrar algumas consequências elementares das condições simétricas.

Iniciamos fornecendo a primeira propriedade elementar do espectro de P .

Proposição 4.5. *Seja uma matriz de transição P de uma Cadeia de Markov finita. Então 1 é um autovalor de P . Além disso, se λ é outro autovalor de P então $|\lambda| \leq 1$.*

Demonstração. Vamos computar o produto da matriz P pelo vetor $\mathbf{1} = (1, 1, \dots, 1)$

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & \cdots & a_{n,n} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} a_{1,1} + a_{1,2} + \cdots + a_{1,n} \\ a_{2,1} + a_{2,2} + \cdots + a_{2,n} \\ \vdots \\ a_{n,1} + a_{n,2} + \cdots + a_{n,n} \end{pmatrix} = 1 \cdot \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Aqui a segunda igualdade segue da Definição de matriz estocástica (1.5). Além disso, este cálculo mostra que 1 é um autovalor de P e o vetor $\mathbf{1} = (1, 1, \dots, 1)$ é o seu autovetor associado.

Supomos agora que $Pv = \lambda v$ com $v \neq 0$. Escolhemos $x \in \Omega$ tal que $|v(x)| \geq |v(y)| \forall y \in \Omega$ (Ω finito). Então

$$|\lambda v(x)| = \left| \sum_{y \in \Omega} P(x, y)v(y) \right| \leq \sum_{y \in \Omega} P(x, y)|v(y)| = |v(x)| \sum_{y \in \Omega} P(x, y) = |v(x)|,$$

e portanto $|\lambda| \leq 1$. □

Uma matriz P simétrica e reversível (Definição 1.21), além da suposição de já ser irredutível e aperiódica, têm todos seus autovalores reais e satisfaz uma ordem decrescente para seus autovalores:

$$1 = \lambda_1 > \lambda_2 \geq \dots \geq \lambda_{|\Omega|} > -1. \quad (4.1)$$

Podemos limitar a variação total $d^*(n)$ em termo dos autovalores da matriz.

Proposição 4.6. *Se P é simétrica então*

$$d^*(n) \leq \frac{1}{2} |\Omega|^{\frac{1}{2}} \left\{ \sum_{k \geq 2} \lambda_k^{4n} \right\}^{\frac{1}{4}}.$$

Demonstração. Pela teoria dos espectros das matrizes de transição temos

$$P^n(x, x) = \sum_k a_x \lambda_k^n, \quad \text{onde} \quad \sum_k a_{x,k}^2 = 1 \quad e \quad a_{x,1} = \frac{1}{|\Omega|}$$

(veja [10, Lema 12.11, pág. 160] e [6, Capítulo 9, pág 71].)

Logo

$$(P^n(x, x))^2 = P^{2n}(x, x) = \sum_k a_{x,k} \lambda_k^{2n}.$$

Como $a_{x,1} = \frac{1}{|\Omega|}$ e $\lambda_1 = 1$ podemos escrever que

$$P^{2n}(x, x) = \sum_{k \geq 2} a_{x,k} \lambda_k^{2n} + \frac{1}{|\Omega|} \lambda_1^{2n} = \sum_{k \geq 2} a_{x,k} \lambda_k^{2n} + \frac{1}{|\Omega|}.$$

Assim

$$\begin{aligned} P^{2n}(x, x) - \frac{1}{|\Omega|} &= \sum_{k \geq 2} a_{x,k} \lambda_k^{2n} \\ &\leq \left(\sum_{k \geq 2} a_{x,k}^2 \right)^{\frac{1}{2}} \left(\sum_{k \geq 2} \lambda_k^{4n} \right)^{\frac{1}{2}} \\ &\leq \left\{ \sum_{k \geq 2} \lambda_k^{2n} \right\}^{\frac{1}{2}}, \end{aligned} \tag{4.2}$$

onde na primeira desigualdade usamos Cauchy-Schwarz e na última $\sum_{k \geq 2} a_{x,k}^2 < 1$. Agora, novamente por Cauchy-Schwarz

$$\begin{aligned} \left\{ \sum_y \left| P^n(x, y) - \frac{1}{|\Omega|} \right| \right\}^2 &\leq \sum_y \left\{ P^n(x, y) - \frac{1}{|\Omega|} \right\}^2 \\ &= |\Omega| \sum_y \left\{ P^n(x, y) - \frac{1}{|\Omega|} \right\}^2 \\ &= |\Omega| \sum_y \left\{ (P^n(x, y))^2 - 2P^n(x, y) \frac{1}{|\Omega|} + \frac{1}{|\Omega|^2} \right\} \\ &= |\Omega| \sum_y (P^n(x, y))^2 - 2 + 1 \\ &= |\Omega| \sum_y (P^n(x, y))^2 - 1 \\ &\stackrel{(*)}{=} |\Omega| P^{2n}(x, y) - 1 \\ &\leq |\Omega| \left\{ \sum_{k \geq 2} \lambda_k^{4n} \right\}^{\frac{1}{2}}, \end{aligned} \tag{4.3}$$

onde em ^(*) usamos o fato de P ser simétrica e a última desigualdade usamos a Equação (4.2).

Como P é simétrica sua distribuição estacionária é uniforme, pela Proposição 4.1. Então considerando a definição de $d^*(n)$ em (1.12), temos

$$\begin{aligned} d^*(n) &= \max_x \| P^n(x, \cdot) - \pi(\cdot) \|_{TV} = \frac{1}{2} \sum_{y \in \Omega} | P^n(x, y) - \pi(y) | \\ &= \frac{1}{2} \sum_{y \in \Omega} | P^n(x, y) - \frac{1}{|\Omega|} | . \end{aligned}$$

Assim usando (4.3) e a Proposição 1.16 temos:

$$d^*(n) = \frac{1}{2} \sum_{y \in \Omega} | P^n(x, y) - \frac{1}{|\Omega|} | \leq \frac{1}{2} |\Omega|^{\frac{1}{2}} \left(\sum_{k \geq 2} \lambda_k^{4n} \right)^{\frac{1}{2}} .$$

□

Observação 4.7. Observe que a sequência de desigualdades em (4.3) implica que

$$P^{2n}(x, x) \geq \frac{1}{|\Omega|}$$

para qualquer cadeia de Markov de estado finito com uma matriz de probabilidade de transição simétrica.

4.3 Passeios Aleatórios em Grupos

No capítulo 2 consideramos o exemplo 2.3 do embaralhamento de n cartas: pegue a carta do topo e insira-a uniformemente ao acaso no baralho, esse processo produzirá uma mistura no baralho. Esses movimentos são melhores tratados como um passeio aleatório no grupo S_n de $n!$ possíveis permutações das cartas, que provaremos, na Proposição 4.9, ter distribuição estacionária uniforme. Uma distribuição de probabilidade μ no grupo simétrico descreve um mecanismo para embaralhar cartas: aplique a permutação $\sigma \in S_n$ ao baralho com probabilidade $\mu(\sigma)$. Embaralhar repetidamente o baralho usando esse mecanismo é equivalente a executar o passeio aleatório no grupo com distribuição de incremento μ .

Dada uma distribuição de probabilidade Q em um grupo finito G , definimos o passeio aleatório em G com distribuição de incremento Q da seguinte forma: é uma cadeia de Markov com espaço de estado G e que se move multiplicando o estado atual à esquerda por um elemento aleatório de G selecionado de acordo com Q . Equivalentemente, a matriz de transição P desta cadeia têm entradas

$$P(g, hg) = Q(h); \quad \forall g, h \in G. \quad (4.4)$$

Outra maneira de interpretar o passeio aleatório em um grupo G é a seguinte: dado algum esquema para escolher aleatoriamente um elemento de G , seja $Q(g)$ a probabilidade de g ser escolhido. Então $Q(g)$ é uma distribuição de probabilidade onde $Q(g) \geq 0$ e

$\sum_{g \in G} Q(g) = 1$ como na Definição 1.1. Escolhas sucessivas feitas independentemente através do mesmo esquema geram o produto

$$\begin{aligned} X_0 &= \text{identidade} = id \\ X_1 &= \xi_1 \\ X_2 &= \xi_2 X_1 = \xi_2 \xi_1 \\ X_3 &= \xi_3 X_2 = \xi_3 \xi_2 \xi_1 \\ &\vdots \\ X_n &= \xi_n X_{n-1} = \xi_n \xi_{n-1} \dots \xi_1, \end{aligned}$$

onde (ξ_i) são independentes com distribuição Q . Então X_n é a Cadeia de Markov em $\Omega = G$ com

$$P(g_1, g_2) = Q(g_2 g_1^{-1}) \quad \forall g_1, g_2 \in G, \quad (4.5)$$

onde as Equações (4.4) e (4.5) são equivalentes.

Podemos interpretar X_k como sendo a posição no tempo k de uma partícula que se move aleatoriamente sobre os vértices de G . Vamos mostrar que a distribuição de X_k pode ser expressa em termos do que é conhecido como a convolução da distribuição Q .

Por exemplo, para calcular a distribuição de X_2 olhamos para $P(X_2 = g)$ com $g \in G$. Começamos com $X_0 = id$. Vamos então calcular $P(X_1 = h)$ para algum $h \in G$:

$$\begin{aligned} P(X_1 = h) &= P(X_1 = h | X_0 = id) \\ &= P(id, h.id) = P(id, h) = Q(h). \end{aligned}$$

Podemos agora calcular a distribuição de X_2

$$\begin{aligned} P(X_2 = h) &= \sum_{h \in G} P(X_1 = h | X_0 = id) P(X_2 = g | X_1 = h) \\ &= \sum_{h \in G} Q(h) P(X_2 = g | X_1 = h) \\ &= \sum_{h \in G} Q(h) P(h, g) \\ &= \sum_{h \in G} Q(h) P[(gh^{-1})^{-1}g, g] \\ &= \sum_{h \in G} Q(h) P[(gh^{-1})hg^{-1}g, (gh^{-1})g] \\ &= \sum_{h \in G} Q(h) P[(g, (gh^{-1})g] \\ &= \sum_{h \in G} Q(h) Q(gh^{-1}) \end{aligned}$$

ou seja, $P(X_2 = g) = \sum_{h \in G} Q(h) Q(gh^{-1}) = Q \star Q(g)$, onde $Q(h)Q(gh^{-1})$ é a chance do elemento h ser escolhido antes e gh^{-1} ser escolhido depois. Similarmente $P(X_k = g) = Q^{k\star}$ onde $Q^{k\star}$ é a convolução repetida

$$Q^{k\star} = Q \star Q^{k-1} = \sum_{h \in G} Q(h) Q^{k-1\star}(gh^{-1}). \quad (4.6)$$

Para o caso do passeio aleatório em grupo finito, as suposições de irreduzibilidade e aperiódicidade correspondem a suposição de Q não ser suportada em nenhuma classe de qualquer subgrupo próprio de G . Então sob esta restrição, um resultado fundamental é que as convoluções repetidas convergem para a distribuição uniforme:

$$Q^{k*} \mapsto U(g) = \frac{1}{|G|}, \text{ quando } k \mapsto \infty. \quad (4.7)$$

Esse resultado mostra que eventualmente a distância de variação tende a zero exponencialmente rápido, uma prova curta pode ser vista em [8, Teorema 3, pág 73] ou [2, Teorema 3, pág 340].

Para o caso do exemplo 2.3, como citamos no início dessa seção, as possíveis transições nas permutações do baralho são um passeio aleatório no grupo S_n , irreduzível e aperiódico.

Como a partir de qualquer ordenação do baralho, qualquer outra ordenação pode eventualmente ser alcançada garantimos a irreduzibilidade. De uma maneira mais formal, conforme discutido na [10, Proposição 2.13, pág 28], desde que o suporte de μ gere todo o S_n , a cadeia resultante é irreduzível. Já a periodicidade ocorre em passeios aleatórios em grupos quando todo o suporte da distribuição de incremento cai em uma única coclasse de algum subgrupo. No entanto, quando a distribuição μ em um grupo G satisfaz $\mu(id) > 0$, então o passeio aleatório com distribuição de incremento μ é aperiódico. De fato, seja $g \in G$, como $\mu(id) = P(g, idg) = P(g, g) > 0$, temos $1 \in \{t : P^t(g, g) > 0\}$ e assim $\text{mdc}\{t : P^t(g, g) > 0\} = 1$.

Lembramos aqui o conceito de coclasse.

Definição 4.8. *Se $H \subseteq G$ são grupos (com a mesma operação) dizemos que H é subgrupo de G . Dado um subgrupo H de G , e $a \in G$, definimos a coclasse esquerda (equivalente para direita) por*

$$aH = \{ah, h \in H\}.$$

Claramente, qualquer passeio aleatório em um grupo é transitivo, pensando G agindo sobre si mesmo e definindo $\phi_{(x,y)}(g) = gx^{-1}y$. O passeio é também semi-simétrico (Definição 4.4) sob $g \mapsto g^{-1}$.

Proposição 4.9. *Seja P a matriz de transição de um passeio aleatório em um grupo finito G e seja U a distribuição de probabilidade uniforme em G . Então U é a distribuição estacionária para P .*

Demonstração. Seja Q a distribuição de incremento para o passeio aleatório. Para qualquer $g \in G$

$$\sum_{h \in G} Q(h)P(h, g) = \frac{1}{|G|} \sum_{k \in G} P(k^{-1}g, g) = \frac{1}{|G|} \sum_{k \in G} Q(k) = \frac{1}{|G|} = U(g).$$

Para primeira igualdade, usamos $k = gh^{-1}$. □

Sejam P e Q sejam probabilidades em um grupo finito G . Assim $\sum_{k \in G} P(k) = 1$ e $P(k) \geq 0$. Pela convolução $P \star Q$ queremos dizer a probabilidade

$$P \star Q(k) = \sum_{t \in G} P(kt^{-1})Q(t),$$

ou seja, "primeiro escolha t de Q , então, independentemente, escolha k de P e forme o produto kt ". A ordem das duas funções na convolução é essencial, pois em geral $P \star Q \neq Q \star P$. De fato, o produto da convolução é comutativo se, e somente se, o grupo G for abeliano.

Seja $s, r \in G$ então $\delta_s \star \delta_r = \delta_{sr}$ e $\delta_r \star \delta_s = \delta_{rs}$ tal que

$$\delta_t(g) = \begin{cases} 1, & g = t; \\ 0, & g \neq t. \end{cases} \quad (4.8)$$

Como s, r são elementos arbitrários de G temos que comutatividade do grupo é uma condição necessária para a comutatividade da convolução. O outro lado, é fácil verificar.

A distribuição uniforme em G é $U(k) = \frac{1}{|G|}$ para todo $k \in G$ e $|G|$ denota a cardinalidade de G . Observe que $U \star U = U$, mas isso não caracteriza U - a distribuição uniforme em qualquer subgrupo satisfaz isso também. No entanto, $U \star P = U$ para qualquer distribuição P e isso caracteriza a distribuição uniforme U .

Definição 4.10. Q é uma distribuição de probabilidade simétrica, em um grupo G , se $Q(g) = Q(g^{-1})$ para todo $g \in G$.

O exemplo 2.3 não envolve distribuições simétricas. Já o passeio aleatório nos inteiros $\mathbb{Z}_n \pmod n$, com n ímpar (similar a figura 1.12), têm distribuição dada por

$$Q(-1) = Q(1) = \frac{1}{2}.$$

Proposição 4.11. O passeio aleatório em um grupo finito G , com distribuição de incremento Q é reversível se Q é simétrica.

Demonstração. Seja U a distribuição de probabilidade uniforme em G . Para qualquer $g, h \in G$ temos que

$$U(g)P(g, h) = \frac{Q(hg^{-1})}{|G|} \quad (4.9)$$

$$U(h)P(h, g) = \frac{Q(gh^{-1})}{|G|}, \quad (4.10)$$

onde (4.9) e (4.10) são iguais se $Q(hg^{-1}) = Q((hg^{-1})^{-1})$. \square

Definiremos agora, algumas condições de simetria.

Definição 4.12. Seja $s, t \in G$. Dizemos que s e t são conjugados se existe $g \in G$ tal que $gsg^{-1} = t$.

Isso define uma relação de equivalência em G , suas classes de equivalência são simplesmente chamadas de classes de conjugação, uma noção que é muito importante na teoria dos grupos.

A seguir, introduziremos uma série de conceitos que serão úteis para o estudo de probabilidades em grupos finitos que são "bem comportadas" com respeito às classes de conjugação.

Definição 4.13. *Uma distribuição de probabilidade Q é constante na classe de conjugação se*

$$Q(g) = Q(hgh^{-1}); \quad \forall g, h \in G.$$

Um exemplo é o passeio gerado por transposições aleatórias:

Exemplo 4.14. *(Transposições Aleatórias) Imagine n cartas seguidas em uma mesa. As cartas começam em ordem, a carta 1 à esquerda, a carta 2 ao lado, ..., e a carta n à direita. Os pares de cartas são transpostos aleatoriamente da seguinte forma: a mão esquerda toca uma carta aleatória e a mão direita toca uma carta aleatória (então esquerda = direita com probabilidade $\frac{1}{n}$, pois é possível que as duas cartas sejam as mesmas). As duas cartas tocadas são trocadas. Um modelo matemático para este processo é a seguinte distribuição de probabilidade no grupo simétrico (S_n) :*

$$\begin{aligned} Q(id) &= \frac{1}{n} \\ Q((i, j)) &= \frac{2}{n^2} \\ Q(\pi) &= 0 \text{ caso contrário} \end{aligned}$$

para (i, j) uma transposição e π um elemento de $S_n - \{id, (i, j)\}$.

A transposição repetida k vezes equivale a convoluir repetidamente Q consigo mesmo, ou seja, Q^{k*} . É sabido que $\frac{1}{2}n \log n$ embaralhadas são necessárias para fazer o baralho estar bem misturado (próximo ao aleatório). Pra ver isso, veja [8, Seção 3D, pág 36].

Definição 4.15. *Seja K um subgrupo de G . Q é uma distribuição K -biinvariante se*

$$Q(g) = Q(k_1 g k_2); \quad \forall g \in G, k_1, k_2 \in K.$$

Definição 4.16. *O par (G, K) é um par de Gelfand se a convolução de distribuições K -biinvariante é comutativa.*

Exemplo 4.17. *(S_n, S_{n-1}) é um par de Gelfand. S_n age em $\{1, 2, \dots, n\}$ e $S_{n-1} \simeq \text{stab}(\{1\})$, onde $\text{stab}(\{1\}) = \{g \in G; k.1 = 1\}$. Veja [7, exemplo 4.3.2, pág 123].*

Seja $Z = \frac{G}{K}$ o espaço quociente. Então consideramos o espaço

$$L(Z) = Q : Z \mapsto \mathbb{C}. \tag{4.11}$$

Dada Q uma distribuição K -biinvariante em um par de Gelfand (G, K) , o passeio aleatório associado é a Cadeia de Markov em Z com matriz de transição:

$$P(g_1 K, g_2 K) = Q(g_1^{-1} g_2 K). \tag{4.12}$$

Diaconis [8] usou pares de Gelfand para determinar a taxa de convergência para a distribuição estacionária de cadeias finitas de Markov. Mais precisamente, dada uma cadeia de Markov que é invariante sob a ação de um grupo G , seu operador de transição pode ser expresso como um operador de convolução cujo núcleo pode ser escrito, pelo menos em teoria, como uma “série de Fourier” onde as exponenciais clássicas $\exp(inx)$ são

substituídos pelas representações irredutíveis do grupo G . Mais detalhes também podem ser vistos em [7].

O próximo resultado usará a condição de transitividade para mostrar uma espécie de recíproca da Equação (1.21). Para isso, lembramos que π_n é a distribuição de X_n resultante de algum estágio inicial x e $s_x(n)$ e $d^*(n)$ são as distâncias de separação e variação entre π_n e a distribuição uniforme U (pois estamos em um grupo), o que por transitividade não depende de x . Assim sob condições de simetria $d^*(n)$ e $s_x(n)$ não dependem da posição inicial, então $d(n) = d^*(n)$ e $s_x(n) = s^*(n)$.

Para $0 < \epsilon < \frac{1}{4}$ definimos

$$\phi(\epsilon) = 1 - (1 - 2\epsilon^{\frac{1}{2}})(1 - \epsilon^{\frac{1}{2}})^2. \quad (4.13)$$

Observe que $\phi(\epsilon)$ diminui a medida que ϵ diminui e $\phi(\epsilon) \approx 4\epsilon^{1/2}$ quando $\epsilon \rightarrow 0$.

Grosso modo, se a distância de variação torna-se pequena após k etapas, a separação torna-se pequena após no máximo $2k$ etapas.

Proposição 4.18. *Se P é transitiva e semi-simétrica em um conjunto G , (em particular, para qualquer passeio aleatório em um grupo G),*

$$d(2k) \leq s_x(2k) \leq \phi(2d(k)), \quad k \geq 1,$$

desde que $d(k) < \frac{1}{8}$.

Para provar essa proposição precisamos do próximo Lema.

Lema 4.19. *Seja (q_1, \dots, q_J) uma distribuição de probabilidade, e suponha $\sum_i |q_i - \frac{1}{J}| \leq \epsilon < \frac{1}{4}$, onde I é o conjunto dos índices i tais que $|q_i - \frac{1}{J}| > \frac{\alpha\epsilon}{J}$. Então para qualquer permutação $\pi \in S_J$ (grupo simétrico),*

$$\sum_i q_i q_{\pi(i)} \geq -\frac{(1 - \phi(\epsilon))}{J}.$$

Demonstração. Fixe $\frac{1}{\epsilon} > \alpha > 2$. Então

$$\sum_{i \in I} |q_i - \frac{1}{J}| \geq \frac{\alpha\epsilon |I|}{J}.$$

Usando a hipótese temos que $|I| \leq \frac{J}{\alpha}$

Agora consideramos a soma $\sum_i q_i q_{\pi(i)}$. Para no mínimo $J - \frac{2J}{\alpha}$ termos, ambos i e $\pi(i)$ estão fora de I e então ambos q_i e $q_{\pi(i)}$ são maiores que $\frac{1}{J} - \frac{\alpha\epsilon}{J}$.

Logo,

$$\begin{aligned} \sum_i q_i q_{\pi(i)} &\geq \left(J - \frac{2J}{\alpha}\right) \left(\frac{1}{J} - \frac{\alpha\epsilon}{J}\right)^2 = \left(J - \frac{2J}{\alpha} \left(\frac{1}{J^2}\right)\right) (1 - \alpha\epsilon)^2 \\ &= \left(J - \frac{2J}{\alpha}\right) \left(\frac{1}{J}\right) \left(\frac{1}{J}\right) (1 - \alpha\epsilon)^2 \\ &= \frac{1}{J} \left(1 - \frac{2}{J}\right) (1 - \alpha\epsilon)^2. \end{aligned}$$

Tomando $\alpha = \epsilon^{-\frac{1}{2}}$ temos

$$\frac{1}{J} \left(1 - \frac{2}{\epsilon^{-\frac{1}{2}}}\right) \left(1 - \epsilon^{\frac{1}{2}}\right)^2 = \frac{1}{J} \left(1 - 2\epsilon^{\frac{1}{2}}\right) \left(1 - \epsilon^{\frac{1}{2}}\right)^2$$

e o resultado segue. \square

Demonstração. (Proposição 4.18)

Provaremos apenas a segunda desigualdade pois a primeira é consequência do Lema 1.17 e da Equação (1.21).

Supomos que $d(n) < \frac{1}{8}$. Então a hipótese do lema 4.19 é satisfeita pela probabilidade $q_i = Q^{k*}(i) = P(i_0, i)$, $i_0, i \in G$, com $\epsilon = 2d(k)$ e $J = |G|$. Agora para qualquer $g \in G$,

$$\begin{aligned} Q^{2k*}(g) &= \sum_i P^k(i_0, i) P^k(i, g) \\ &= \sum_i P^k(i_0, i) P^k(\epsilon(g), \epsilon(i)) \text{ por alguma bijeção } \epsilon : G \mapsto G \\ &= \sum_i P^k(i_0, i) P^k(i_0, H_g \epsilon(i)) \text{ por alguma bijeção } H_g : G \mapsto G \\ &\geq \frac{(1 - \phi(2d(k)))}{|G|} \text{ pelo lema 4.19.} \end{aligned}$$

E por (1.21) para $s_x(2k)$ temos que $s_x(2k) \leq \phi(2d(k))$. \square

O próximo resultado é um lema de comparação para passeios aleatórios em grupos.

Lema 4.20. *Sejam Q_1, Q_2 distribuições no grupo G satisfazendo:*

$$Q_1 \star Q_2 = Q_2 \star Q_1. \quad (4.14)$$

Sejam $s_x^1(n), s_x^2(n)$ as distâncias de separação para os passeio aleatórios associados.

- a) *Suponha que exista $c > 0$ tal que $Q_2(g) \geq cQ_1(g)$, $g \in G$. Então*

$$s_x^2(n) \leq \sum_{b=0}^n s_x^1(b) \binom{n}{b} c^b (1-c)^{n-b}, \quad n \geq 1.$$

- b) *Seja $Q = Q_1 * Q_2$ e seja $s_x(n)$ a distância de separação associada.*

Então

$$s_x(n) \leq s_x^1(n), \quad n \geq 1.$$

Demonstração. A parte a) é uma especialização do lema 1.23. Podemos escrever $Q_2 = cQ_1 + (1-c)Q_0$, para alguma distribuição Q_0 em G . Considerando variáveis aleatórias independentes

$$\xi_1^0, \xi_2^0, \dots; \xi_1^1, \xi_2^1, \dots; J_0, J_1, \dots,$$

onde ξ_i^1 têm distribuição Q_1 e ξ_i^0 têm distribuição Q_0 e $P(J_i = 1) = c$ e $P(J_i = 0) = 1 - c$. Definimos

$$\xi_i^2 = \xi_i^{J_i}.$$

Então ξ_i^2 têm distribuição Q_2 e, conseqüentemente

$$\begin{aligned} X_n = \xi_1^2 \star \xi_2^2 \star \dots \star \xi_n^2 \text{ têm distribuição } (Q_2)^{n\star} \\ \text{condicional em } \{J_i = j_i; 1 \leq i \leq n\}, X_n = \xi_1^{j_1} \star \xi_2^{j_2} \star \dots \star \xi_n^{j_n}. \end{aligned} \quad (4.15)$$

A hipótese (4.14) $\xi_{i_1}^2 \xi_{i_2}^1$ têm a mesma distribuição que $\xi_{i_1}^1 \xi_{i_2}^2$. Então sob repetidas tranposições em (4.15) de termos adjacentes,

$$\xi_1^i \star \xi_2^i \star \dots \star \xi_{B_n}^1 \star \xi_{B_n+1}^0 \star \dots \star \xi_n^0 \text{ têm distribuição } (Q_2)^{n\star},$$

onde $B_n = \sum_{i=1}^n \mathbb{1}_{(J_i=1)}$ têm distribuição binomial (n, c) . Logo,

$$(Q_2)^{n\star} = \sum_{b=0}^n P(B_n = b)Q = \sum_{b=0}^n \binom{n}{b} c^b (1-c)^{n-b} Q, \quad (4.16)$$

onde Q é a distribuição de $\xi_1^1 \star \dots \star \xi_b^1 \star \xi_{b+1}^0 \star \dots \star \xi_n^0$.

Como a convolução diminui a distância de separação então a separação $s_x(n)$, associada a $Q = Q_1 \star Q_2$ na parte b), é no máximo $s_x^1(n)$, pois $\sum \lambda_b Q$ é no máximo $\sum \lambda_b s_x(n)$, usando a Equação (1.28). \square

A razão para mencionar esse resultado é que as condições de simetria, vistas nas Definições 4.13 e 4.15 são úteis na verificação de comutatividade (Equação (4.14)). Se Q_1, Q_2 são K -biinvariantes em um par de Gelfand então a Equação (4.14) vale por definição. Se Q é constante na classe de conjugação então

$$Q \star P = P \star Q, \quad \forall P \text{ distribuição em } G. \quad (4.17)$$

Outra classe natural de exemplos são passeios aleatórios em grafos. Para um grafo finito não direcionado $\Gamma = (V, E)$ com conjunto de vértices V e conjunto de arestas E , onde os elementos de E são pares não ordenados de vértices: $E \subseteq \{\{x, y\} : x, y \in V, x \neq y\}$. Podemos pensar em V como um conjunto de pontos, onde dois pontos x e y são unidos por uma linha se, e somente se x, y for um elemento do conjunto de arestas. Quando $x, y \in E$, escrevemos $x \sim y$ e dizemos que y vizinho de x (e vice-versa). O grau $deg(x)$ de um vértice x é o número de vizinhos de x .

Dado um grafo $\Gamma = (V, E)$, podemos definir o passeio aleatório simples em Γ como a cadeia de Markov com espaço de estado V e matriz de transição

$$P(x, y) = \begin{cases} \frac{1}{deg(x)}, & x \sim y; \\ 0, & \text{caso contrário.} \end{cases} \quad (4.18)$$

Isso quer dizer que, quando a cadeia está no vértice x , ela examina todos os vizinhos de x , escolhe um uniformemente ao acaso e se move para o vértice escolhido.

Aqui, P é irredutível e aperiódico se, e somente se, Γ é conexo (se existir pelo menos um caminho entre cada par de vértices do grafo) e não 2-colorível (seu número cromático é menor ou igual a 2). E P é duplamente estocástica se, e somente se, Γ é

regular ($\deg(x) = m, \forall x \in V$). Com essas condições P é automaticamente simétrica. Essas e outras definições, sobre o estudo de grafos, podem ser melhor vistas em [10].

Assim para qualquer vértice $y \in V$,

$$\sum_{x \in V} \deg(x) P(x, y) = \sum_{x \sim y} \frac{\deg(x)}{\deg(x)} = \deg(y). \quad (4.19)$$

Para obter uma probabilidade, simplesmente normalizamos por

$$\sum_{y \in V} \deg(y) = 2 | E |,$$

e assim concluímos que a medida de de probabilidade é $\forall y \in V$

$$\pi(y) = \frac{\deg(y)}{2 | E |}.$$

O hipercubo N -dimensional, introduzido na Seção 2.2, é um grafo cujos vértices são as N -uplas binárias $\{0, 1\}^N$. Como o hipercubo um grafo N -regular (os graus de todos os vértices são iguais a N), a distribuição estacionária do passeio simples e do preguiçoso são uniformes em $\{0, 1\}^N$.

Para encerrar esse capítulo vamos considerar G um grupo e S um subconjunto de G que não está contido em nenhuma coclasse de nenhum subgrupo de G (pois, caso contrário seria possível encontrar elementos de G que não são gerados a partir dos elemento de S). Seja Q a distribuição uniforme em $S : Q(g) = \frac{1}{|S|}, g \in S$.

Nesse cenário, mostramos que há uma construção de tempos estacionários fortes. Seja $k \geq 1$ e $g \in G$, definimos

$$B_g^k = \{(s_1, s_2, \dots, s_k) \in S^k / s_1 s_2 \cdots s_k = g\}, \quad (4.20)$$

$$b_k = \min_{g \in G} | B_g^k |. \quad (4.21)$$

Como $Q^{k*}(g) = \frac{|B_g^k|}{|S|^k}$, temos que a definição de distância de separação em (1.18) leva a

$$\begin{aligned} s_x(k) &= \max_g \left(1 - \frac{Q^{k*}(g)}{\pi(g)} \right) = 1 - \min_g \left(\frac{Q^{k*}(g)}{1/|G|} \right) \\ &= 1 - \min_g \left(\frac{|B_g^k| / |S|^k}{1/|G|} \right) = 1 - \frac{|G| b_k}{|S|^k}. \end{aligned} \quad (4.22)$$

Seja $\hat{B}_g^k \subset B_g^k$ tal que $|\hat{B}_g^k| = b_k$. Escrevemos o passeio aleatório associado com a distribuição Q como

$$X_n = \xi_n \xi_{n-1} \cdots \xi_1.$$

Como $b_k > 0$, podemos definir um tempo aleatório T tomando valores em $\{k, 2k, \dots\}$ por

$$T = k \min\{j \geq 1 : (\xi_{jk}, \xi_{j(k-1)}, \dots, \xi_{(j-1)k+1}) \in \hat{B}_g^k \text{ para algum } g \in G\}. \quad (4.23)$$

Temos que

$$P(T = k, X_k = g) = \frac{|\hat{B}_g^k|}{|S|^k} = \frac{b_k}{|S|^k} = \frac{1 - s_x(k)}{|G|} \quad (4.24)$$

e então somando a Equação (4.24) sobre $g \in G$ obtemos

$$P(T = k) = 1 - s_x(k). \quad (4.25)$$

Em particular, temos que

$$P(X_k = g|T = k) = \frac{P(X_k = g, T = k)}{P(T = k)} = \frac{1}{|G|}.$$

Como T é um tempo de parada (cf. Definição 1.1), temos que o evento $\{T > (j-1)k\}$ depende apenas de $X_0, \dots, X_{(j-1)k}$. Vamos utilizar indução em j para mostrar que

$$P(T = jk, X_{jk} = g) = \frac{s_x(k)^{j-1}(1 - s_x(k))}{|G|}. \quad (4.26)$$

para todo $j \geq 1$.

O caso $j = 1$ segue da Equação (4.24). Suponha agora que a Equação (4.26) seja válida para j . Então

$$\begin{aligned} P(T = (j+1)k, X_{(j+1)k} = g) &= P(X_{(j+1)k} = g, T = (j+1)k, T > jk) \\ &= E(P(X_{(j+1)k} = g, T = (j+1)k, T > jk | X_0, \dots, X_{jk})) \\ &= E(\mathbb{1}_{\{T > jk\}} P(X_{(j+1)k} = g, T = (j+1)k | X_0, \dots, X_{jk})) \\ &= P(T = k, X_k = g) P(T > jk) \\ &= \frac{1 - s_x(k)}{|G|} s_x(k)^j, \end{aligned}$$

onde E representa a esperança associada com a probabilidade P . Isso prova (4.26).

Somando sobre $g \in G$ em (4.24), concluímos imediatamente que, T/k têm a distribuição geométrica com parâmetro $1 - s_x(k)$, isto é:

$$P(T = jk) = s_x(k)^{j-1}(1 - s_x(k)). \quad (4.27)$$

Proposição 4.21. *Supondo as Equações (4.20) e (4.21) definindo T como em (4.23) para algum $k \geq 1$ tal que $b_k > 0$ (o que equivale a $s_x(k) < 1$), então temos que T é um tempo estacionário forte para o qual*

$$P(T > jk) = s_x^j(k), \quad j \geq 1. \quad (4.28)$$

Demonstração. A afirmação que T é um tempo estacionário forte segue dividindo (4.26) por (4.27).

Agora precisamos mostrar que vale (4.28). Pela proposição 2.4 b) sabemos que existe T tal que $s_x(k) = P(T > k)$.

Então

$$\begin{aligned} P(T > jk) &= P(T = (j+1)k) + P(T = (j+2)k) + \dots \\ &= s_x^j(k)(1 - s_x(k)) + s_x^{j+1}(k)(1 - s_x(k)) + \dots \\ &= \frac{s_x^j(k)(1 - s_x(k))}{1 - s_x(k)} \\ &= s_x^j(k), \end{aligned}$$

onde na penúltima igualdade usamos a fórmula da soma dos termos de uma P.G. com razão $0 < s_x^j(k) < 1$. \square

Vamos supor agora que o conjunto S de (4.20) satisfaça

$$S = S^{-1} \text{ ou seja, se } s \in S \Rightarrow s^{-1} \in S \text{ e } S \text{ contém a identidade.} \quad (4.29)$$

Definimos o “diâmetro” de um elemento do grupo G como

$$\Delta_g = \min\{n : \text{cada } g \in G \text{ é da forma } g = h_1 h_2 \dots h_n \text{ para algum } h_i \in S\}. \quad (4.30)$$

Proposição 4.22. *Sob as condições (4.20), (4.21) e (4.29),*

$$d(n) \leq \frac{1}{2} |G|^{\frac{3}{4}} \left\{ 1 - \frac{1}{17\Delta_g^2 |S|} \right\}^n, \quad n \geq 1.$$

Demonstração. Seja $S_0 = S - \{\text{identidade}\}$, Q_0 a distribuição uniforme em S_0 e P e P_0 as matrizes de transição referentes a Q e Q_0 . Como podemos escrever $P = Q(g)I + (1 - Q(g))P_0$, ou seja, uma combinação convexa, temos

$$P = \frac{1}{|S|} I + \frac{|S| - 1}{|S|} P_0. \quad (4.31)$$

Seja λ_2 o segundo maior autovalor de P_0 e seja $\tau_e = \frac{-1}{\log(\lambda_2)}$ o que implica que $\lambda^n = \exp\left(-\frac{n}{\tau_e}\right)$.

Para formalizar essa ideia, utilizaremos um resultado de [4, Lema 2.4, pág 86], adaptando-o para o caso em questão. Seja $\Gamma = (V, E)$ como definido na seção 4.3 e $\epsilon > 0$ o maior número tal que $A \subset V$ com $|A| \leq \frac{1}{2} |V|$, o conjunto $F = \{w \notin A; (w, v) \text{ é uma aresta para algum } v \in A \text{ tal que } |F| \geq \epsilon |A|\}$. Então para o caso de um grafo r -regula (todos vértices com grau igual a r), ou seja um grupo G , temos que

$$\tau_e \leq r(2 + 4\epsilon^{-2}), \quad (4.32)$$

veja em [1, Proposição 3.1, pág. 39].

O parâmetro ϵ é uma espécie de medida do quão conectado é o grafo, o que não é tão óbvio de ser estimado. No entanto, para o caso onde G é um grupo $S, S_0 \subset G$ satisfazendo a Equação (4.29) e como diâmetro definido na Equação (4.30) temos que, pelo [1, Lema 3.1, pág. 39]

$$\epsilon \geq \frac{1}{2\Delta_g}, \quad (4.33)$$

Neste caso temos que

$$\tau_e \leq 17\Delta_g^2 |S_0|. \quad (4.34)$$

De fato, fixando $A \subset G$ com $|A| \leq \frac{1}{2} |G|$, seja $Ag = \{ag; a \in A\}$. Então

$$\frac{1}{|G|} = \sum_{g \in G} |A \cup Ag| = \frac{|A|^2}{|G|},$$

e então existe um $\hat{g} \in G$ tal que $|A \cup A\hat{g}| \leq \frac{|A|^2}{|G|}$ que implica que

$$|A\hat{g} - A| \geq |A| - \frac{|A|^2}{|G|} \geq \frac{1}{2} |A|. \quad (4.35)$$

Se $g = h_1 h_2 \dots h_d$ com $h_i \in S_0$ e $d \leq \Delta_g$ podemos escrever $g_i = h_1 h_2 \dots h_i$ então

$$|Ag - A| \leq \sum_{i=1}^d |Ag_i - Ag_{i-1}| = \sum_{i=1}^d |Ah_i - A|. \quad (4.36)$$

Por (4.35) e (4.36), existe algum $h_0 \in S_0$ tal que $|Ah_0 - A| \geq \frac{1}{2} |A| \Delta_g$. Então (4.33) segue considerando o conjunto de arestas $(a, ah_0), a \in A$.

Assim pela expansão de Taylor, temos que $\log x \approx x - 1$ então

$$\log \lambda_2 \approx \lambda_2 - 1 \Rightarrow -\log \lambda_2 \approx 1 - \lambda_2.$$

Portanto

$$\begin{aligned} \tau_e &= \frac{1}{1 - \lambda_2} \leq 17\Delta_g^2 |S_0| \\ 1 &\leq 17\Delta_g^2 |S_0| (1 - \lambda_2) \\ \frac{1}{17\Delta_g^2 |S_0|} &\leq 1 - \lambda_2 \\ \lambda_2 &\leq 1 - \frac{1}{17\Delta_g^2 |S_0|}. \end{aligned}$$

Segue que (4.36) que todos os autovalores $\lambda_k, k \geq 2$ de P satisfaz

$$|\lambda_k| \leq 1 - \frac{1}{17\Delta_g^2 |S|},$$

e o resultado segue da Proposição 4.6. □

Capítulo 5

Representações de Grupos

A teoria das representações estuda operadores lineares que são dados por elementos de grupo que atuam em espaços vetoriais. Além disso, fornece uma ferramenta poderosa para a obtenção de cotas na distância de separação e variação total, onde os detalhes podem ser vistos em [8], [7] e [13].

5.1 Definições Básicas e Resultados da Teoria da Representação

Nessa seção forneceremos a definição de uma representação de um grupo finito e formas de estudá-la com vários conceitos e teoremas notáveis.

Seja G um grupo finito e V um espaço vetorial finito sobre \mathbb{C} . Denotamos por $GL(V)$ o grupo linear de V consistindo de todos os mapas lineares invertíveis $A : V \mapsto V$.

Uma representação linear de G em V é um homomorfismo $\rho : G \mapsto GL(V)$. Em outras palavras, uma representação ρ de G em algum espaço de dimensão finita n atribui a cada $s \in G$ uma matriz $n \times n$ invertível $\rho(s)$ de tal forma que a matriz atribuída ao produto de dois elementos é o produto das matrizes atribuídas a cada elemento:

$$\rho(st) = \rho(s)\rho(t), \quad \forall s, t \in G. \quad (5.1)$$

Nesse texto denotaremos uma representação por (ρ, V) ou apenas ρ . Observe que (5.1) implica que:

$$\rho(id) = I \quad \text{e} \quad \rho(s^{-1}) = \rho(s)^{-1}.$$

Seja $\dim V = n$, como $GL(V)$ é isomorfo a $GL(n, \mathbb{C})$ o grupo de todas as matrizes invertíveis $n \times n$, podemos considerar uma representação de G como um homomorfismo de grupo $\rho : G \mapsto GL(n, \mathbb{C})$. Então n é a dimensão ou grau de ρ e será denotada por d_ρ .

Se W é um subespaço de V G -invariante ($\rho(g)W \subseteq W$ para todo $g \in G$) então ρ restrito a W é uma subrepresentação de (ρ, V) . Claramente, toda representação é uma subrepresentação de si mesma.

Definição 5.1. *Se a representação ρ não admite subrepresentação não trivial (o subespaço $W = 0$ e o subespaço $W = V$ são subrepresentações triviais), então ρ é chamado de irredutível.*

Observação 5.2. Qualquer representação unidimensional é claramente irredutível.

Exemplo 5.3. Seja $G = S_n$ o grupo simétrico de grau n , que é o grupo de permutações de n elementos. Nesse grupo temos três tipos de representações.

- (i) A representação trivial (mapa que atua como a identidade $\rho_1(g) = 1, \forall g \in G$) é unidimensional, ou seja, $V = \mathbb{C}$;
- (ii) A representação sinal (ou alternada). Também é unidimensional (irredutível), onde atribui a cada elemento de grupo a identidade 1 se g for uma permutação par (ou seja, g é o produto de um número par de transposições, equivalentemente $g \in A_n$, o grupo alternado), e o oposto da identidade, -1 caso contrário, isto é, se g for uma permutação ímpar. Em outras palavras, denotando pelo sinal (g) o sinal de uma permutação g , temos $\rho_2(g) = \text{sinal}(g)I_n$. Como sinal é um homomorfismo $G \mapsto \frac{S_n}{A_n} \cong C_2$ ou seja, $\text{sinal}(g_1g_2) = \text{sinal}(g_1)\text{sinal}(g_2)$ então ρ_2 é de fato uma representação;
- (iii) A representação de permutação é n -dimensional: fixemos uma base ortonormal $\{e_1, e_2, \dots, e_n\}$ de um espaço vetorial n -dimensional V e defina $\rho_3(g)e_j = e_{g(j)}$ para todo $g \in S_n$. Com relação à base V , $\rho_3(g)$ é a chamada matriz de permutação, ou seja, uma matriz obtida permutando as linhas da matriz identidade. Essa é uma representação não irredutível, pois, por exemplo, o hiperplano $\sum_{i=1}^n x_i = 0$ é um $n - 1$ -dimensional subespaço S_n -invariante.

Definição 5.4. Sejam (ρ, V) e (σ, W) duas representações do mesmo grupo G . Então ρ e σ são equivalentes se existe $T : V \mapsto W$ bijeção linear tal que $\forall g \in G, \sigma(g)T = T\rho(g)$. Em imagem, temos que o diagrama comuta.

$$\begin{array}{ccc} \rho(g) : V & \longrightarrow & V \\ T \downarrow & & \downarrow T \\ \sigma(g) : W & \longrightarrow & W \end{array}$$

Suponha que o espaço vetorial V seja dotado de um produto escalar $\langle \cdot | \cdot \rangle$ então podemos definir a noção de uma representação unitária.

Definição 5.5. Uma representação (ρ, V) é unitária se preserva o produto interno, ou seja, $\langle \rho(g)v | \rho(g)v' \rangle = \langle v | v' \rangle$ para todo $g \in G$ e $v, v' \in V$.

Vale ressaltar que, dada uma representação arbitrária (ρ, V) é sempre possível dotar V com um produto interno para que ρ se torne uma representação unitária. Dado um produto interno arbitrário $\langle \cdot | \cdot \rangle$ para V defina, $\forall v, w \in V$

$$(v|w) = \sum_{g \in G} \langle \rho(g)v | \rho(g)w \rangle.$$

Proposição 5.6. Seja V um espaço vetorial munido dos produtos internos: $(\cdot | \cdot)$ e $\langle \cdot | \cdot \rangle$. A representação $(\rho, V_{(|)})$ é equivalente a $(\rho, V_{\langle | \rangle})$. Em particular, toda representação de G é equivalente a uma representação unitária.

A prova é simples e pode ser vista em [7, Proposição 3.3.1, pág 84].

Sejam $\rho : G \mapsto GL(V)$ e $\sigma : G \mapsto GL(W)$ duas representações de um grupo G . Seja $f : V \mapsto W$ uma transformação linear. Diz-se que f comuta ρ e σ se $f\rho(g) = \sigma(g)f$ para todo $g \in G$.

O seguinte resultado relaciona a noção de redutibilidade de uma representação com a existência de comutadores.

Lema 5.7. (*Lema de Schur*) *Sejam (ρ, V) e (σ, W) duas representações irredutíveis de um grupo G . Se f comuta ρ e σ ($f\rho(g) = \sigma(g)f$), então*

- (1) f é zero ou é um isomorfismo
- (2) Se $V = W$ e $\rho = \sigma$ então f é um múltiplo escalar da matriz identidade ($f = \lambda I$).

A prova desse Lema pode ser vista em [13, Proposição 4, pág. 13].

Agora vamos descrever uma informação [14, Corolário 4.1.8, pág 29] útil das representações irredutíveis para grupos abelianos, consequência do Lema de Schur.

Corolário 5.8. *Um grupo G é abeliano se, e somente se, todas as representações irredutíveis de G têm grau um.*

Demonstração. Seja $\rho : G \mapsto GL(V)$ uma representação irredutível. Fixe por um momento $h \in G$. Então, para todo $g \in G$, temos que

$$\rho(h)\rho(g) = \rho(hg) = \rho(gh) = \rho(g)\rho(h).$$

Agora, pelo Lema de Schur 5.7 $\exists \lambda_h \in \mathbb{C}$ tal que $\rho(h) = \lambda_h I$.

Logo, temos que $\forall v \in V : \rho(h)v = \lambda_h v$.

Como vale para todo $v \in V$ isso implica que todo subespaço $W \subset V$ é G -invariante. Portanto (ρ, W) é uma subrepresentação de (ρ, V) . Como supomos, por hipótese, que ρ é irredutível então $W = 0$ ou $W = V$.

Supomos, por absurdo, que $\dim V > 1$. Se tomarmos o span de um elemento da base então teríamos um subespaço de v que deveria ser invariante. Pela irredutibilidade de ρ , isso contradiz a hipótese.

Portanto como $V \neq 0$ então $\dim V = 1$.

Por outro lado, o número de classes de conjugação é igual ao número de representações irredutíveis, veja [8, Teorema 7 e Teorema 8, pág 15]. Se $\sum_{\rho} d_{\rho}^2 = |G|$ e cada uma é unidimensional, deve haver tantos elementos nos grupos quanto como existem classes de conjugação, então G é abeliano. □

Observação 5.9. *Uma representação de grau um de G é apenas um homomorfismo de grupo $\rho : G \rightarrow \mathbb{C}^{\times}$, onde \mathbb{C}^{\times} é o grupo multiplicativo de \mathbb{C} .*

Lema 5.10. *Seja G um grupo finito, $\rho : G \mapsto GL(V)$ então para todo $g \in G$, todos os autovalores de $\rho(g)$ são raízes da unidade.*

Demonstração. Seja $\rho(g) = A, g \in G$. Note que $\rho(id) = I$ e portanto, os autovalores da matriz identidade são 1. Como G é finito, então qualquer elemento $g \in G$ têm ordem finita, ou seja, $\exists N \in \mathbb{Z}_{+}^{*}$ tal que $g^N = id$

$$\rho(g)^N = \rho(g^N) = \rho(id) = I \Rightarrow A = I.$$

Seja λ um autovalor de A e \vec{v} seu autovetor correspondente, ou seja,

$$A\vec{v} = \lambda\vec{v}. \quad (5.2)$$

Multiplicando por A à esquerda em (5.2), temos

$$A.A\vec{v} = A^2\vec{v} = A\lambda\vec{v} = \lambda(A\vec{v}) = \lambda(\lambda\vec{v}) = \lambda^2\vec{v}. \quad (5.3)$$

Multiplicando novamente à esquerda temos $A^3\vec{v} = \lambda^3\vec{v}$. Indutivamente obtemos $A^N\vec{v} = \lambda^N\vec{v}$.

Como $A^N = I \Rightarrow \vec{v} = \lambda^N\vec{v} \Rightarrow (\lambda^N - 1)\vec{v} = 0$ e como $\vec{v} \neq 0$ (pois é autovetor) temos

$$\lambda^N = 1 \Rightarrow \lambda \text{ é uma } n\text{-ésima raiz da unidade.}$$

□

Um caso particular desse Lema é quando ρ é uma representação de grau um de G e $|G| = n$. Seja $g \in G$ então, pela Observação 5.9, $(\rho(g))^n = 1 \in \mathbb{C}^\times$. Neste caso, $\rho(g)$ é uma n -ésima raiz da unidade em \mathbb{C} , para todo $g \in G$.

Lema 5.11. *Seja $G = \langle g \rangle$ um grupo cíclico de ordem n . Então G têm exatamente n representações de grau um.*

Demonstração. Se $\rho : G \rightarrow \mathbb{C}^\times$ é uma representação do grau um, então, como vimos no Lema 5.10, $\rho(g)$ é uma n -ésima raiz da unidade em \mathbb{C} . Por outro lado, se $\zeta \in \mathbb{C}$ é uma n -ésima raiz da unidade, então o mapa $\rho : G \rightarrow \mathbb{C}^\times$ definido por $\rho(g^j) = \zeta^j$, para todo j , é obviamente um homomorfismo de grupo bem definido. Então, se ζ_1, \dots, ζ_n são as n -ésimas raízes da unidade, então existem exatamente n representações de G que têm grau um, são elas:

$$\rho_k : G \rightarrow \mathbb{C}^\times, \quad 1 \leq k \leq n, \quad \text{definido por } \rho_k(g^j) = \zeta_k^j, \quad \text{para todo } j.$$

□

Pelo Lema de Schur, o [8, Corolário 3, pág 11] fornece relações de ortogonalidade para as entradas da matriz de ρ unitária. Mas, podemos supor, sem perda de generalidade, sempre que necessário que ρ é unitária (Proposição 5.6), ou seja, $\rho(t)^* = \rho(t^{-1})$. Definimos o produto interno usual em todas as funções $\phi, \psi : G \rightarrow \mathbb{C}$:

$$\langle \phi | \psi \rangle = \frac{1}{|G|} \sum \phi(t)\psi(t)^*.$$

Então $\langle \rho^{(i,j)} | \eta^{(k,l)} \rangle = 0$ se ρ e η são representações unitárias inequivalentes, para qualquer entrada $(i, j), (k, l)$.

$$\langle \rho^{(i,j)} | \rho^{(k,l)} \rangle = \begin{cases} \frac{1}{d_\rho}, & i = k, j = l; \\ 0, & \text{exceto quando } i = k \text{ e } j = l. \end{cases} \quad (5.4)$$

A noção do caracter de uma representação é extraordinariamente útil. Se ρ for uma representação de G , definimos o caracter associado a ρ como a função

$$\chi_\rho : G \mapsto \mathbb{C} \text{ definida por } \chi_\rho(s) = \text{Tr}(\rho(s)).$$

Lembrando que o traço de uma matriz $A_{n \times n}$ é dado por

$$\text{Tr}(A) = \text{Trace}(A) = \sum_i^n a_{ii}$$

ou seja, a soma dos elementos da diagonal principal, que também é a soma dos autovalores de A (contados com suas multiplicidades), e não depende da escolha da base.

Observação 5.12. *Observamos que se ρ é uma representação unidimensional, então coincide com seu caracter: $\rho = \chi_\rho$.*

Proposição 5.13. *Se χ_ρ é o caracter de uma representação (ρ, V) de um grupo G . Então $\forall s, t \in G$,*

- (i) $\chi_\rho(1) = d_\rho$;
- (ii) $\chi_\rho(s^{-1}) = \overline{\chi_\rho(s)}$;
- (iii) $\chi_\rho(t^{-1}st) = \chi_\rho(s)$,

onde denotamos o conjugado do número complexo $z = x + iy$ por $z^* = x - iy$ ou \bar{z} .

Demonstração. Observe que $\rho(1_G) = I$ a identidade (matriz identidade) em V cujo o traço é claramente a dimensão de V , d_ρ , logo temos (i).

Para (ii) observamos que $\rho(s)$ têm ordem finita o que segue que os autovalores $\rho(s)$ são raízes da unidade (esse fato pode ser visto no Lema 5.10). Então

$$\chi_\rho(s)^* = \text{Tr}(\rho(s)^*) = \sum \lambda_i^* = \sum \frac{1}{\lambda_i} = \text{Tr}(\rho(s)^{-1}) = \text{Tr}(\rho(s^{-1})) = \chi_\rho(s^{-1}).$$

Já para mostrar (iii) podemos escrever $\chi_\rho(vu) = \chi_\rho(uv)$, tomando $u = ts$ e $v = t^{-1}$ e o resultado segue do fato de $\text{Tr}(AB) = \text{Tr}(BA)$, onde A e B são matrizes quadradas. \square

Agora que temos os mecanismo básicos da teoria da representação, podemos definir a transformada de Fourier de maneira adequada.

Seja Q uma distribuição de probabilidade em G . A transformada de Fourier de Q na representação ρ é definida por

$$\hat{Q}(\rho) = \sum_{g \in G} Q(g)\rho(g). \quad (5.5)$$

Abordaremos agora uma propriedade adicional da transformada de Fourier. A propriedade de convolução que mostra que a transformada de Fourier da convolução de duas funções é simplesmente o produto das transformadas de Fourier das funções.

Consideramos Q_1, Q_2 distribuições e (ρ, V) uma representação.

$$\begin{aligned}
(\widehat{Q_1 \star Q_2})(\rho) &= \sum_{g \in G} (Q_1 \star Q_2)(g) \rho(g) \\
&= \sum_{g \in G} \left(\sum_{h \in G} Q_1(gh^{-1}) Q_2(h) \right) \rho(g) \\
&= \left[\sum_{g \in G} Q_1(gh^{-1}) \rho(gh^{-1}) \right] \sum_{h \in G} Q_2(h) \rho(h) \\
&= \hat{Q}_1(\rho) \hat{Q}_2(\rho).
\end{aligned} \tag{5.6}$$

Indutivamente vale para $n \leq 2$. Portanto, podemos evitar fazer convolução tomando as transformadas de Fourier! Em muitos casos, isso será muito mais conveniente do que executar diretamente a convolução.

Exemplo 5.14. *Seja a distribuição uniforme U em G , dada por $U(g) = \frac{1}{|G|} \forall g \in G$. Vamos mostrar que a transformada de Fourier é*

$$\hat{U}(g) = \begin{cases} I, & \rho \text{ é a representação trivial;} \\ 0, & \rho \text{ é não-trivial e irredutível.} \end{cases} \tag{5.7}$$

Se $\rho = \rho_1$, como no exemplo 5.3, $\rho_1(g) = I, \forall g \in G$

$$\hat{U}(\rho_1) = \sum_{g \in G} U(g) \rho_1(g) = \sum_{g \in G} U(g) I = I.$$

Se ρ é uma representação irredutível não-trivial

$$\hat{U}(\rho) = \sum_{g \in G} U(g) \rho(g) = \frac{1}{|G|} \sum_{g \in G} \rho(g). \tag{5.8}$$

Vamos supor, por absurdo, que $\hat{U}(\rho) \neq 0$ o que implica que $\hat{U}(\rho)$ têm autovetor $v \neq 0$.

Seja λ é um autovalor de $\hat{U}(\rho)$ e $v \in \mathbb{C}^n$ temos que

$$\lambda v = \hat{U}(\rho)v = \frac{1}{|G|} \sum_{g \in G} \rho(g)v \in \text{span}(v).$$

Então $\rho(g)v \in \text{span}(v)$. Como $\text{span}(v)$ é um subespaço não zero invariante, temos um absurdo pois ρ é irredutível e portanto, $\hat{U}(\rho) = 0$.

Observação 5.15. *Denotaremos por \sum_{ρ} o somatório sob todas as representações irredutíveis não triviais.*

Definimos como

$$\text{Tr}(AA^*) = \text{Trace}(AA^*) = \|A\|^2, \tag{5.9}$$

onde A é uma matriz quadrada e A^* sua matriz conjugada.

Enunciaremos agora, dois resultados que estarão em uso constante até a conclusão desse trabalho, onde as provas podem ser vistas em [8, pág 15] ou [7, pág 47 e 48].

Teorema 5.16. (Teorema da Inversão de Fourier) Seja Q uma distribuição em G então

$$Q(g) = \frac{1}{|G|} \sum_{\hat{G}} d_{\rho} \text{Tr}(\hat{Q}(\rho)\rho(g^{-1})). \quad (5.10)$$

onde \hat{G} é o dual de G , um conjunto completo de representações não-equivalentes (unitárias) pareadas irredutíveis de G (em outras palavras, \hat{G} contém exatamente um elemento pertencente a cada classe de equivalência de representações irredutíveis).

Teorema 5.17. (Teorema de Plancherel) Seja $P, Q : G \mapsto \mathbb{C}$, então

$$\langle P, Q \rangle = \frac{1}{|G|} \sum_{\rho} d_{\rho} | \hat{P}(\rho)\hat{Q}(\rho)^* |. \quad (5.11)$$

Se $P = Q$

$$\| Q \|_2^2 = \frac{1}{|G|} \sum_{\rho} d_{\rho} \| \hat{Q}(\rho) \|_{HS}^2, \quad (5.12)$$

onde $\| Q \|_2^2 = \sqrt{\langle P, Q \rangle}$ e $\| \hat{Q}(\rho) \|_{HS} = \sqrt{\text{Tr}(\hat{Q}(\rho)\hat{Q}(\rho)^*)}$.

A fórmula de inversão de Fourier (5.10) mostra que cada distribuição é determinada exclusivamente por suas transformadas de Fourier com respeito às representações irredutíveis. Já Fórmula de Plancherel (5.11) diz que o produto interno de duas funções é igual ao "produto interno" de suas transformadas.

Uma cota superior útil é dada pelo seguinte Lema.

Lema 5.18. (Lema do Limite Superior) Para uma distribuição Q em um grupo finito G ,

$$\| Q^{n*} - U \|^2 \leq \frac{1}{4} \sum_{\rho} d_{\rho} \| \hat{Q}^n(\rho) \|^2. \quad (5.13)$$

Demonstração. Pela Proposição 1.16 temos que

$$\| P - U \| = \frac{1}{2} \sum_{s \in G} | P(s) - Q(s) |,$$

onde P, Q são distriuições em G . Portanto, a partir disso temos

$$\begin{aligned} 4 \| Q^{n*} - U \|^2 &= \sum_{s \in G} | Q^{n*}(s) - U(s) |^2 \\ &\leq |G| \sum_{s \in G} | Q^{n*}(s) - U(s) |^2 \\ &= \sum_{\rho} d_{\rho} \text{Tr}[(\hat{Q}(\rho))^n (\hat{Q}(\rho))^n] \\ &\leq \sum_{\rho} d_{\rho} \| \hat{Q}^n(\rho) \|^2, \end{aligned} \quad (5.14)$$

onde na primeira desigualdade usamos Cauchy-Schwarz e na segunda igualdade o Teorema de Plancherel 5.17 e $\hat{Q}(\rho) = 1$ se $\rho = \rho_1$ e $\hat{U}(\rho) = 0$ para ρ não trivial. \square

Assim, o problema de delimitar a distância de variação $d(n)$ para o passeio aleatório em G associado a Q se reduz ao problema de estimar autovetores e autovalores de $\hat{Q}(\rho)$. Esta técnica é particularmente útil nas condições de simetria das Definições 4.13 ou 4.15.

O restante dessa seção será usado para obtenção de cotas para distância da variação total em grupos não necessariamente abelianos, e que satisfazem certas condições de simetria, vistas no capítulo 4. A análise de Fourier de funções que são constantes nas classes de conjugação. E a análise de funções bi-invariantes em um par de Gelfand que oferece uma rica teoria e uma coleção de exemplos. Embora este último não sendo utilizado para obter cotas relevantes para o nosso exemplo do hipercubo, eles possuem interesse teórico e aplicações interessantes como, por exemplo, o problema do embaralhamento de cartas.

Proposição 5.19. *Se Q é constante na classe de conjugação (Definição 4.13) então sua transformada de Fourier com respeito as representações irredutíveis é da forma*

$$\hat{Q}(\rho) = \lambda I \quad \text{onde} \quad \lambda = \frac{1}{d_\rho} \sum_{g \in G} Q(g) \chi_\rho(g). \quad (5.15)$$

Demonstração. Observe que $\forall g \in G$

$$\begin{aligned} \rho(g) \hat{Q}(\rho) \rho^{-1}(g) &= \sum_{h \in G} \rho(g) (Q(h) \rho(h)) \rho(g^{-1}) \\ &= \sum_{h \in G} Q(h) \rho(ghg^{-1}) \\ &= \sum_{h \in G} Q(ghg^{-1}) \rho(ghg^{-1}) \\ &= \hat{Q}(\rho), \end{aligned}$$

onde a última igualdade segue de uma mudança de variáveis apropriada.

Então $\hat{Q}(\rho)$ comuta ρ consigo mesmo. Pelo Lema de Shur 5.7, temos que

$$\hat{Q}(\rho) = \lambda I. \quad (5.16)$$

Computando o traço de ambos os lados em (5.16), temos

$$\text{Tr}(\hat{Q}(\rho)) = \sum_{h \in G} Q(h) \chi_\rho(g) = \lambda d_\rho.$$

□

Seja o espaço $L(Z)$ definido em (4.11). Termos correspondentes à representação que não aparecem na decomposição de $L(Z)$ têm transformada de Fourier nula, por causa da Reciprocidade de Frobenius [13, Teorema 13, pág 56], que diz que o número de vezes que uma representação ρ de G aparece na indução de σ (uma representação de K) de G é igual à multiplicidade de σ em ρ restrito a K . Isso implica que uma representação ρ ocorre em $L(Z)$ com multiplicidade correspondente à dimensão do espaço de K vetores fixos de (ρ, V) . Assim se ρ não aparece em $L(Z)$ ($L(Z)$ é a representação trivial de K induzida até G), a representação trivial não ocorre em ρ restrita a K . Veja [8, Lema 5, pág 55] para mais detalhes.

Proposição 5.20. *O par (G, K) é um par de Gelfand se cada representação irredutível de G aparecer em $L(Z)$ com multiplicidade no máximo 1.*

Demonstração. Vamo supor, por absurdo, que (G, K) é um par de Gelfand, mas alguma representação ρ_i têm multiplicidade $j > 1$ em $L(Z)$.

Escolhemos uma base de V_i com as primeiras j coordenadas percorrendo o espaço K -invariante.

Seja A_1 e A_2 matrizes $j \times j$ não comutativas. Definimos duas distribuições Q_1 e Q_2 em G tais que

$$\begin{cases} \hat{Q}_1(\rho) = \hat{Q}_2(\rho) = 0, & \text{se } \rho \neq \rho_i; \\ \hat{Q}_1(\rho_i) = \begin{pmatrix} A_1 & 0 \\ 0 & 0 \end{pmatrix}, \hat{Q}_2(\rho_i) = \begin{pmatrix} A_2 & 0 \\ 0 & 0 \end{pmatrix}. \end{cases} \quad (5.17)$$

Pela fórmula da inversão de Fourier (5.10), essas são funções K -biinvariantes, não nulas e $Q_1 * Q_2 \neq Q_2 * Q_1$. O que é uma contradição pela Definição 4.16. \square

Observação 5.21. *Na verdade, a Proposição 5.20 é uma equivalência, onde a recíproca pode ser vista [8, Teorema 9, pág 54] ou [7, Teorema 4.4.2, pág 125].*

Proposição 5.22. *Seja Q uma distribuição de um passeio aleatório em um par de Gelfand (G, K) , então sua transformada de Fourier com respeito as representações irredutíveis é da forma*

$$\hat{Q}(\rho) = \begin{pmatrix} \alpha & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

onde $\alpha = |K| \sum_{x \in Z} Q(x) s_i(x)$.

Demonstração. Pela Observação 5.21, para um par de Gelfand, seja

$$L(Z) = V_0 \oplus V_1 \oplus \cdots \oplus V_N \quad (5.18)$$

a decomposição em irredutíveis distintos (isso significa ter multiplicidade no máximo 1 o que chamamos de multiplicidade livre). A reciprocidade de Frobenius [13, Teorema 13, pág 56] implica que cada V_j têm um subespaço unidimensional K -invariante $V_j^k := \{v \in V_j : \rho_j(k)v = v\}$. Seja $s_j(x)$ ser uma função K -invariante, ou seja, $s_j \in V_j^k$ normalizada e $s_j(id) = 1$. É chamada de j -ésima função esférica e, estas formam uma base para o espaço $L(Z)$, segundo [7, Proposição 4.7.1, pág. 135]. As funções esféricas foram explicitamente computadas para muitos Pares de Gelfand.

Fixando i , considere o espaço vetorial V_i de (5.18) como uma representação ρ_i de G . Complete s_i a uma base para V_i ortonormal, tomando s_i como o primeiro vetor de base $(\{s_i, v_2^i, \dots, v_d^i\})$. Então para $k \in K$ a matriz $\rho(k)$ é do tipo:

$$\rho(k) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & * & \cdots & * \\ \vdots & \vdots & * & \vdots \\ 0 & * & * & * \end{pmatrix}.$$

Com relação a esta base, a transformada de Fourier de qualquer função K bi-invariante Q em G torna-se

$$\begin{aligned}
\hat{Q}(\rho) &= \sum_{g \in G} Q(g) \rho_i(g) \\
&= \sum_{kt \in G} Q(tk) \rho_i(tk) \\
&= \sum_{k, t \in G} Q(t) \rho_i(tk) \\
&= \sum_{k, t \in G} Q(t) \rho_i(t) \rho_i(k) \\
&= \sum_t Q(t) \rho_i(t) \sum_{k \in K} \rho_i(k),
\end{aligned} \tag{5.19}$$

onde na terceira igualdade usamos a invariância a direita. Pelas relações de ortogonalidade de Schur (Equação (5.4)) temos que a entrada (a, b) com $a \neq b$ da matriz satisfaz:

$$\langle \rho^{(a,b)} | \rho^{(a,b)} \rangle := \frac{1}{|G|} \sum_{g \in G} \rho^{(a,b)}(g) \rho^{(a,b)}(g)^* = \frac{1}{|G|} \sum_{g \in G} \rho^{(a,b)}(g) \rho^{(a,b)}(g^{-1}) = 0.$$

Com isso temos que

$$\rho(k) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & * & \cdots & 0 \\ \vdots & \vdots & * & \vdots \\ 0 & 0 & 0 & * \end{pmatrix}.$$

Mas como (G, K) é um par de Gelfand a dimensão do subespaço K -invariante é 1. Ou seja, se por exemplo, a entrada $(2, 2)$ da matriz for diferente de zero, a segunda coluna também seria um autovetor, logo vetores K -invariantes.

Assim temos que

$$\sum_{k \in K} \rho_i^{(a,b)}(k) = \begin{cases} |K|, & \text{se } a = b = 1; \\ 0, & \text{caso contrário.} \end{cases} \tag{5.20}$$

Logo, $\hat{Q}(\rho_i)$ têm forma $\sum Q(x)(:0) = (:0)$ para K -invariantes à direita e para Q invariante à esquerda temos que $\hat{Q}(\rho_i)$ têm forma

$$\begin{pmatrix} * & \cdots & * & * \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Comparando os dois caso de invariância à direita e à esquerda e pelo [8, Lema 5, pág. 53] temos que para qualquer K -biinvariante distribuição Q existe uma base de V tal que

$$\hat{Q}(\rho_i) = \begin{pmatrix} * & \cdots & \cdots & 0 \\ \vdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

é uma matriz com entradas nulas, exceto na posição $(1, 1)$. Para calcular o valor dessa entrada, aplicamos a Equação (5.20) na (5.19) temos

$$\sum_t Q(t)\rho_i(t) \sum_{k \in K} \rho_i(k) = |K| \sum_t Q(t)\rho_i^{(1,1)}(t).$$

No Lema 5.25 mostraremos que $\rho_i^{(1,1)}(t) = s_{\rho_i}(t)$, concluindo a prova. \square

Nos casos acima, o produto de matrizes é reduzido ao produto de números.

Proposição 5.23. *Suponha que Q é constante na classe de conjugação (Definição 4.13) de G . Para uma representação irredutível ρ seja*

$$\tilde{Q}(\rho) = \frac{1}{d_\rho} \sum_{g \in G} Q(g)\chi_\rho(g), \quad (5.21)$$

onde $\chi_\rho(g) = \text{Tr}(\rho(g))$. Então $\forall g \in G$

$$(a) \quad 1 - |G| Q(g) \leq \sum_\rho d_\rho |\chi_\rho(g)\tilde{Q}(\rho)| \leq \sum_\rho d_\rho^2 |\tilde{Q}(\rho)|;$$

$$(b) \quad \|Q - U\|_{TV}^2 \leq \frac{1}{4} \sum_\rho d_\rho^2 |\tilde{Q}(\rho)|^2.$$

Observação 5.24. *Perceba que λ na Proposição 5.19 é igual a $\tilde{Q}(\rho)$.*

Demonstração. (a) Para a primeira desigualdade observamos, primeiramente, que a Fórmula da inversão de Fourier (5.10) é dada pela soma sob todas representações irredutíveis de G incluindo a representação trivial ρ_1 .

Segundo, se Q é constante na classe de conjugação e ρ é uma representação irredutível, então temos que

$$\hat{Q}(\rho) = \tilde{Q}(\rho)I \quad (5.22)$$

pelo que vimos que na Proposição 5.19.

Logo,

$$\begin{aligned}
1 - |G|^{-1} |Q(g)| &\leq |1 - |G|^{-1} Q(g)| \\
&= |1 - \sum_{\hat{G}} d_\rho \operatorname{Tr}(\hat{Q}(\rho) \rho(g^{-1}))| \\
&= |1 - \sum_{\hat{G}} d_\rho \operatorname{Tr}(\tilde{Q}(\rho) I \rho(g^{-1}))| \\
&= |1 - \sum_{\hat{G}} d_\rho \tilde{Q}(\rho) \chi_\rho(g)| \\
&= 1 - \sum_{\hat{G}} d_\rho |\tilde{Q}(\rho)| |\chi_\rho(g^{-1})| \\
&= 1 - \sum_{\hat{G}} d_\rho |\tilde{Q}(\rho)| |\overline{\chi_\rho(g)}| \\
&= |1 - d_{\rho_1} \tilde{Q}(\rho_1) \chi_{\rho_1}(g) - \sum_{\rho} d_\rho \chi_\rho(g) \tilde{Q}(\rho)|,
\end{aligned}$$

onde ρ_1 denota a representação trivial. Temos que $d_{\rho_1} = 1 = \chi_{\rho_1}(g)$. Além disso,

$$\tilde{Q}(\rho_1) = \sum_{g \in G} Q(g) \chi_{\rho_1}(g) = \sum_{g \in G} Q(g) = 1.$$

Logo segue que

$$\begin{aligned}
1 - |G|^{-1} |Q(g)| &\leq |1 - 1 - \sum_{\rho} d_\rho \tilde{Q}(\rho) \chi_\rho(g)| \\
&\leq \sum_{\rho} |d_\rho \tilde{Q}(\rho) \chi_\rho(g)| \\
&= \sum_{\rho} d_\rho |\tilde{Q}(\rho) \chi_\rho(g)|.
\end{aligned}$$

Para demonstrar a segunda desigualdade, seja α autovalor de $\rho(g)$, então $|\alpha| = 1$ (pelo Lema (5.10)). Assim temos

$$|\chi_\rho(g)| = |\operatorname{Tr}(\rho(g))| = \left| \sum_i \alpha_i \right| \leq \sum_i |\alpha_i| = d_\rho,$$

onde na segunda igualdade usamos o fato de $\operatorname{Tr}(A) = \sum_{i=1}^n \lambda_i$ tal que $A_{n \times n}$ matriz e λ_i são os autovalores de A .

(b) Usaremos o lema 5.18 para $n = 1$

$$\|Q - U\|_{TV}^2 \leq \frac{1}{4} \sum_{\rho} \|\hat{Q}(\rho)\|^2.$$

Como $\|\hat{Q}(\rho)\|^2 = \text{Tr}(\hat{Q}(\rho)\hat{Q}(\rho)^*)$, substituindo temos

$$\begin{aligned} \|Q - U\|_{TV}^2 &\leq \frac{1}{4} \sum_{\rho} d_{\rho} \text{Tr}(\hat{Q}(\rho)\hat{Q}(\rho)^*) \\ &= \frac{1}{4} \sum_{\rho} d_{\rho} \text{Tr}(\tilde{Q}(\rho)I\tilde{Q}(\rho)^*I) \\ &= \frac{1}{4} \sum_{\rho} d_{\rho} \text{Tr}(|\tilde{Q}(\rho)|^2 I) \\ &= \frac{1}{4} \sum_{\rho} d_{\rho}^2 |\tilde{Q}(\rho)|^2, \end{aligned}$$

onde na primeira igualdade usamos (5.16). □

Lema 5.25. *Para uma representação irredutível ρ de G , seja*

$$s_{\rho}(g) = \frac{1}{|K|} \sum_{k \in K} \chi_{\rho}(gk), \quad (5.23)$$

onde s_{ρ} são as funções esféricas mencionadas na Proposição 5.22 e $\chi_{\rho}(g) = \text{Tr}(\rho(g))$. Então $s_{\rho}(g)$ coincide com a primeira entrada da matriz $\rho(g)$.

Demonstração.

$$\begin{aligned} s_{\rho}(g) &= \frac{1}{|K|} \sum_{k \in K} \chi_{\rho}(gk) \\ &= \frac{1}{|K|} \sum_{k \in K} \text{Tr}(\rho(g)\rho(k)) \\ &= \frac{1}{|K|} \sum_{k \in K} \rho^{(1,1)}(g) \\ &= \frac{1}{|K|} |K| \rho^{(1,1)}(g) \\ &= \rho^{(1,1)}(g), \end{aligned} \quad (5.24)$$

onde na terceira igualdade usamos (5.20). □

Proposição 5.26. *Seja Q uma distribuição K -biinvariante no par de Gelfand (G, K) (Definições 4.15 e 4.16). Para uma representação irredutível ρ de G , seja $s_{\rho}(g)$ como definido acima na Equação (5.23). Definimos*

$$\bar{Q}(\rho) = \sum_{g \in G} Q(g)s_{\rho}(g). \quad (5.25)$$

Então para todo $g \in G$,

$$(a) \quad 1 - |G| Q(g) \leq \sum_{\rho} d_{\rho} |\bar{Q}(\rho)s_{\rho}(g)| \leq \sum_{\rho} d_{\rho} |\bar{Q}(\rho)|$$

$$(b) \quad \|Q - U\|_{TV}^2 \leq \frac{1}{4} \sum_{\rho} d_{\rho} |\bar{Q}(\rho)|^2.$$

Observação 5.27. Perceba que na Proposição 5.22 $\alpha = \bar{Q}$. De fato,

$$\begin{aligned} \bar{Q}(\rho) &= \sum_{g \in G} Q(g) s_{\rho}(g) \\ &= \sum_{z \in Z} \sum_{k \in K} Q(zk) s_{\rho_i}(zk) \\ &= \sum_{z \in Z} |K| Q(z) s_{\rho_i}(z) \\ &= |K| \sum_{z \in Z} Q(z) s_{\rho_i}(z). \end{aligned}$$

Visto que Q e s_{ρ} são funções K -biinvariantes.

Demonstração. (a) Para mostrar a primeira desigualdade, temos que se Q é K -biinvariante no par de Gelfand (G, K) então

$$\hat{Q}(\rho) = \bar{Q}(\rho)$$

pelo que vimos que na Proposição 5.22.

Logo, usando a Fórmula da Inversão de Fourier (5.10),

$$\begin{aligned} 1 - |G| |Q(g)| &\leq |1 - |G| |Q(g)| \\ &= |1 - \sum_{\hat{G}} d_{\rho} \text{Tr}(\hat{Q}(\rho) \rho(g^{-1}))| \\ &= |1 - \sum_{\hat{G}} d_{\rho} \text{Tr}(\bar{Q}(\rho) \rho(g^{-1}))| \\ &= |1 - \sum_{\hat{G}} d_{\rho} \bar{Q}(\rho) \rho^{(1,1)}(g)^{-1}| \\ &= |1 - d_{\rho_1} \bar{Q}(\rho_1) \rho_1^{(1,1)}(g)^{-1} - \sum_{\rho} d_{\rho} \bar{Q}(\rho) \rho^{(1,1)}(g^{-1})|. \end{aligned}$$

Como $d_{\rho_1} = 1 = \rho_1^{(1,1)}(g)^{-1}$, temos

$$\begin{aligned} s_{\rho_1}(g) &= \frac{1}{|K|} \sum_{k \in K} \chi_{\rho_1}(gk) \\ &= \frac{1}{|K|} \sum_{k \in K} \text{Tr}(I) \\ &= \frac{1}{K} \sum_{k \in K} d_{\rho_1} \\ &= \frac{1}{|K|} \sum_{k \in K} 1 = 1. \end{aligned}$$

Assim,

$$\bar{Q}(\rho_1) = \sum_{g \in G} Q(g) s_{\rho_1}(g) = 1. \quad (5.26)$$

pela Equação (5.1).

Retomando a desigualdade temos que

$$\begin{aligned} |1 - |G| Q(g)| &\leq |1 - 1 - \sum_{\rho} d_{\rho} \bar{Q}(\rho) \rho^{(1,1)}(g)^{-1}| \\ &\leq \sum_{\rho} |d_{\rho} \bar{Q}(\rho) \rho^{(1,1)}(g)^{-1}| \\ &= \sum_{\rho} d_{\rho} |\bar{Q}(\rho) \rho^{(1,1)}(g)^{-1}| \\ &= \sum_{\rho} d_{\rho} |\bar{Q}(\rho) \rho^{(1,1)}(g)| \\ &= \sum_{\rho} d_{\rho} |\bar{Q}(\rho) s_{\rho}(g)|, \end{aligned}$$

onde na última igualdade usamos a Equação (5.24).

Para demonstrar a segunda desigualdade, como podemos assumir que as representações irredutíveis são sempre unitárias, pela Proposição 5.6, isso implica que as colunas da matriz são ortonormais. Ou seja, denotando por v_1 a primeira coluna da matriz $\rho(g)$ então $\|v_1\| = 1$, assim

$$\rho^{(1,1)}(g) \leq \|v_1\| = 1.$$

Pelo Lema 5.25 temos que

$$\rho^{(1,1)}(g) = s_{\rho}(g) \leq 1 \quad \forall g \in G.$$

Assim temos que $\sum_{\rho_i} d_{\rho} |\bar{Q}(\rho) s_{\rho}(g)| \leq \sum_{\rho} d_{\rho} |\bar{Q}(\rho)|$.

(b) Usaremos o lema 5.18 para $n = 1$

$$\|Q - U\|_{TV}^2 \leq \frac{1}{4} \sum_{\rho} \|\hat{Q}(\rho)\|^2.$$

Como $\|\hat{Q}(\rho)\|^2 = \text{Tr}(\hat{Q}(\rho)\hat{Q}(\rho)^*)$, substituindo temos

$$\begin{aligned} \|Q - U\|_{TV}^2 &\leq \frac{1}{4} \sum_{\rho} d_{\rho} \text{Tr}(\hat{Q}(\rho)\hat{Q}(\rho)^*) \\ &= \frac{1}{4} \sum_{\rho} d_{\rho} \text{Tr}(\bar{Q}(\rho)\bar{Q}(\rho)^*) \\ &= \frac{1}{4} \sum_{\rho} d_{\rho} \text{Tr}(|\bar{Q}(\rho)|^2) \\ &= \frac{1}{4} \sum_{\rho} d_{\rho}^2 |\bar{Q}(\rho)|^2, \end{aligned}$$

onde na última igualdade que $\text{Tr}(\bar{Q}(\rho)) = \bar{Q}(\rho)$ pois $\bar{Q}(\rho)$ é da forma mostrada na Proposição 5.22. \square

Os limites finais em cada (a) da Proposição 5.23 e 5.26 não envolvem g e, portanto, é um limite na separação $s_x(n)$. Aplicar os limites a Q^{n^*} fornece limites em $s_x(n)$ e $d(n)$.

5.2 Análise Harmônica no Hipercubo

Retomando o exemplo do hipercubo, seja $H_N = \{(x_1, x_2, \dots, x_N) : x_1, x_2, \dots, x_N \in \{0, 1\}\}$, o passeio aleatório preguiçoso, que pode ser visto como um passeio em $(\mathbb{Z}/2\mathbb{Z})^N$, que é o produto direto de N cópias do grupo de dois elementos, $(\mathbb{Z}/2\mathbb{Z}) = \{0, 1\}$.

Como $(\mathbb{Z}/2\mathbb{Z})$ é um grupo cíclico e, portanto abeliano, temos que para cada ρ , existe $y \in H_N$ tal que o mapa $x \mapsto \rho(x) = (-1)^{x \cdot y}$ é uma representação unidimensional (Corolário 5.8), onde

$$x \cdot y = \sum_u x_u y_u \pmod{2} \quad (5.27)$$

para $x, y \in H_N$.

O passeio aleatório preguiçoso em H_N é baseado na distribuição de probabilidade

$$Q(x) = \begin{cases} \frac{1}{N+1}, & \text{se } |x| \leq 1; \\ 0, & \text{se } |x| \geq 2, \end{cases} \quad (5.28)$$

onde $x = (x_1, x_2, \dots, x_N) \in H_N$ e $|x| = x_1 + x_2 + \dots + x_N$ (ou seja, o número de 1's em x).

Podemos calcular a transformada de Fourier de Q definido na Equação (5.28)

$$\begin{aligned} \hat{Q}(\rho) &= \sum_{x \in H_N} Q(x) \rho(x) \\ &= \sum_{x \in H_N} Q(x) (-1)^{x \cdot y} \\ &= \sum_{x: |x| \leq 1} \frac{1}{N+1} (-1)^{x \cdot y} \\ &= \frac{1}{N+1} \sum_{x: |x| \leq 1} (-1)^{x \cdot y}. \end{aligned} \quad (5.29)$$

Se $|x| = 1$, então x é do tipo $e_u = (0, \dots, 0, \overset{u}{1}, 0, \dots, 0)$. Se $y = (y_1, y_2, \dots, y_N)$ então

$e_u \cdot y = y_u$. Logo,

$$\begin{aligned}
\hat{Q}(\rho) &= \frac{1}{N+1} \left[1 + \sum_{u=1}^N (-1)^{e_u \cdot y} \right] \\
&= \frac{1}{N+1} \left[1 + \sum_{u=1}^N (-1)^{y_u} \right] \\
&= \frac{1}{N+1} \left[1 + \sum_{u=1}^N (-1) \mathbb{1}_{(y_u=1)} + \sum_{u=1}^N \mathbb{1}_{(y_u=0)} \right] \\
&= \frac{1}{N+1} \left[1 - |\rho| + (N - |\rho|) \right] \\
&= \frac{1}{N+1} \left[(1+N) - 2|\rho| \right] \\
&= 1 - \frac{2|\rho|}{N+1}.
\end{aligned} \tag{5.30}$$

Por (5.28) e (5.29) $Q(x)$ e $\hat{Q}(\rho)$ são funções de $|x|$ e $|\rho|$, respectivamente.

Como vimos na Equação (5.6) a transformada de Fourier converte convolução em produto, logo $(\hat{Q}^N) = (\hat{Q})^N$.

Aplicando a fórmula da inversão (5.10) para $Q^{k\star}$ temos

$$\begin{aligned}
Q^{k\star} &= \frac{1}{2^N} \sum_y d_y \text{Tr}(\hat{Q}^k \rho(x^{-1})) \\
&= \frac{1}{2^N} \sum_y (-1)^{x \cdot y} (\hat{Q}^k) \\
&= \frac{1}{2^N} \sum_y (-1)^{x \cdot y} \left(1 - \frac{2|\rho|}{N+1} \right)^k,
\end{aligned} \tag{5.31}$$

onde na segunda igualdade usamos o fato que o caracter de uma representação unidimensional ser a própria representação, ou seja, $\text{Tr}(\rho(x)) = \rho(x)$.

Seja $L(H_N) = \{f : H_N \mapsto \mathbb{C}\}$ o espaço das funções complexas em H_N . Seja V_k o subespaço de $L(H_N)$ gerado pelas representações com $|x| = k$

$$V_k = \langle y : |y| = k \rangle,$$

onde a $\dim V_k = \binom{N}{k}$ que é o número de 1's que podemos escolher nas N coordenadas.

Proposição 5.28. *Seja Q como em (5.28). Para $k = \frac{1}{4}(N+1)(\log N + c)$ temos que*

$$\|Q^{k\star} - U\|_{TV}^2 \leq \frac{1}{2}(e^{e^{-c}} - 1). \tag{5.32}$$

Demonstração. Pela Equação (5.22) temos que

$$((\tilde{Q})^{k\star}(\rho)) = \left(1 - \frac{2|\rho|}{N+1} \right)^k$$

Então, pela parte b) da proposição 5.23

$$\begin{aligned}
\| Q^{k^*} - U \|_{TV}^2 &\leq \frac{1}{4} \sum_{\rho \neq 0} \left(1 - \frac{2|\rho|}{N+1}\right)^{2k} \\
&= \frac{1}{4} \sum_{j=1}^N \dim V_j \left(1 - \frac{2j}{N+1}\right)^{2k} \\
&= \frac{1}{4} \sum_{j=1}^N \binom{N}{j} \left(1 - \frac{2j}{N+1}\right)^{2k} \\
&\leq \frac{1}{2} \sum_{j=1}^{\lfloor \frac{N+1}{2} \rfloor} \binom{N}{j} \left(1 - \frac{2j}{N+1}\right)^{2k}.
\end{aligned}$$

Agora usando $\binom{N}{j} = \frac{N!}{(N-j)!j!} \leq \frac{N!}{j!}$ e $1 - x \leq e^{-x}$ consequentemente

$$\left(1 - \frac{2j}{N+1}\right)^{2k} \leq e^{-\frac{4kj}{N+1}}$$

quando $j \leq \frac{N+1}{2}$, pois se j satisfaz essa condição então $1 - \frac{2j}{N+1} > 0$
Portanto, temos

$$\| Q^{k^*} - U \|_{TV}^2 \leq \frac{1}{2} \sum_{j=1}^{\lfloor \frac{N+1}{2} \rfloor} \frac{N!}{j!} e^{-\frac{4kj}{N+1}}. \quad (5.33)$$

Agora, se $k = \frac{1}{4}(N+1)(\log N + c)$ temos que

$$e^{-\frac{4kj}{N+1}} = e^{-j \log N - jc} = \exp\left(-\frac{jc}{Nj}\right)$$

e aplicando à (5.33) temos

$$\begin{aligned}
\| Q^{k^*} - U \|_{TV}^2 &\leq \frac{1}{2} \sum_{j=1}^{\lfloor \frac{N+1}{2} \rfloor} \frac{N!}{j!} e^{-j \log N - jc} \\
&= \frac{1}{2} \sum_{j=1}^{\lfloor \frac{N+1}{2} \rfloor} \frac{1}{j!} e^{-jc} \\
&\leq \frac{1}{2} \sum_{j=1}^{\infty} \frac{1}{j!} (e^{-c})^j \\
&= \frac{1}{2} (e^{e^{-c}} - 1).
\end{aligned}$$

□

Como $\frac{1}{2}(e^{e^{-c}} - 1) \leq \frac{1}{2}e^{e^{-c}}$, tomando a raiz quadrada dos dois lados,

$$d(k) = \| Q^{k^*} - U \|_{TV} \leq \frac{1}{\sqrt{2}} \exp\left(\frac{1}{2}e^{-c}\right). \quad (5.34)$$

Capítulo 6

Conclusões

Cadeias de Markov irredutíveis, aperiódicas e com estado finito têm distribuições estacionárias únicas. Além disso, para qualquer cadeia desse tipo, as probabilidades de transição convergem para a distribuição estacionária. Em várias aplicações é desejável saber quantos passos são necessários para que as probabilidades de transição se tornem próximas das estacionárias. Nesse trabalho, apresentamos três das técnicas para estudar este problema: Tempo estacionário forte, acoplamento e análise Fourier. Concentramos nosso interesse na cadeia de Markov para o passeio aleatório no hipercubo.

Tempos estacionários fortes são aliados à distância de separação, como vimos no capítulo 2, e diz que o tempo necessário para diminuir a distância de separação, que é uma cota superior para a distância de variação total, é assintoticamente limitado acima por $N \log N$ para passeio no hipercubo, que é exatamente o dobro $\left(\frac{1}{2}N \log N\right)$ da cota superior obtida usando acoplamento (capítulo 3), onde para melhorar o tempo para a cota superior encontrada fizemos uma comparação com o problema da Urna de Ehrenfest.

A análise de Fourier e suas generalizações para a teoria de representação de grupos foram usadas para limitar as taxas de convergência através do Lema do Limite Superior 5.18, com ele vimos que é preciso metade do tempo necessário alcançado pelo acoplamento e a quarta parte do tempo da técnica do Tempo estacionário forte, ou seja, é preciso convoluir a cadeia por cerca de $\left(\frac{1}{4}N \log N\right)$ para estar próxima ao equilíbrio. Assim essa técnica fornece melhores resultados em relação aos métodos probabilísticos. Neste método analítico (capítulo 5) vimos que a transição de um estado para o seguinte é determinado por uma medida P , onde $Q \star P(g) = \sum_{h \in G} Q(h)P(h^{-1}g)$ representa a probabilidade de permanecer no elemento g após uma etapa, onde Q pode representar a probabilidade inicial.

Quando G é abeliano, o Lema de Schur (5.7) implica $d_\rho = 1, \forall \rho \in \hat{G}$. Este é o caso do passeio aleatório no hipercubo. O cálculo das n -ésimas potências das matrizes $\hat{P}(\rho)$ é em teoria mais fácil do que o cálculo das n -ésimas potências da matriz $P(h^{-1}g)$, pois, para determinar essas matrizes, é necessário um conhecimento detalhado das representações irredutíveis de G . Ou seja, isso permite estudar a distribuição convoluída de $Q^{k\star}$ sem mencionar o passeio aleatório X_k .

A computação pode ser realmente mais fácil se P satisfizer certas condições de simetria e as matrizes $\hat{P}(\rho)$ podem reduzir-se a matrizes com apenas entradas na diagonal principal (constante na classe de conjugação) ou uma entrada não trivial que está na diagonal (Pares de Gelfand).

Referências Bibliográficas

- [1] David Aldous. On the markov chain simulation method for uniform combinatorial distributions and simulated annealing. Probability in the Engineering and Informational Sciences, 1(1):33–46, 1987.
- [2] David Aldous and Persi Diaconis. Shuffling cards and stopping times. The American Mathematical Monthly, 93(5):333–348, 1986.
- [3] David Aldous and Persi Diaconis. Strong uniform times and finite random walks. Advances in Applied Mathematics, 8(1):69–97, 1987.
- [4] Noga Alon. Eigenvalues and expanders. Combinatorica, 6(2):83–96, June 1986. Copyright: Copyright 2008 Elsevier B.V., All rights reserved.
- [5] Gely P. Basharin, Amy N. Langville, and Valeriy A. Naumov. The life and work of a.a. markov. Linear Algebra and its Applications, 386:3–26, 2004. Special Issue on the Conference on the Numerical Solution of Markov Chains 2003.
- [6] Ehrhard Behrends. Introduction to Markov Chains with Special Emphasis on Rapid Mixing. 01 2000.
- [7] Tullio Ceccherini-Silberstein, Fabio Scarabotti, and Filippo Tolli. Harmonic Analysis on Finite Groups: Representation Theory, Gelfand Pairs and Markov Chains. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2008.
- [8] Persi Diaconis. Group representations in probability and statistics. Lecture Notes-Monograph Series, 11:i–192, 1988.
- [9] David Griffeath. A maximal coupling for markov chains. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 31:95–106, 1975.
- [10] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. Markov chains and mixing times. American Mathematical Society, Providence, RI, 2009. With a chapter by James G. Propp and David B. Wilson.
- [11] J. R. Norris. Markov Chains. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1997.
- [12] Jim Pitman. On coupling of markov chains. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 35:315–322, 1976.
- [13] Jean Pierre. Serre. Linear representations of finite groups / Jean-Pierre Serre ; translated from the French by Leonard L. Scott. Springer-Verlag New York, 1977.

- [14] Benjamin Steinberg. Representation theory of finite groups. An introductory approach. 01 2011.
- [15] H. Thorisson. Coupling, Stationarity, and Regeneration. Probability and Its Applications. Springer New York, 2000.