



UNIVERSIDADE FEDERAL DA BAHIA - UFBA
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA - IME
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA - PGMAT
DISSERTAÇÃO DE MESTRADO



CADEIAS DE MARKOV E TEMPO DE MISTURA

JOEZITO COSTA DOS SANTOS JÚNIOR

Salvador-Bahia

2023

Cadeias de Markov e tempo de mistura

Joezito Costa dos Santos Junior

Dissertação apresentada ao Colegiado do Curso de Pós-graduação em Matemática da Universidade Federal da Bahia, como requisito parcial para obtenção do Título de Mestre em Matemática.

Banca examinadora

Tertuliano Franco

Prof. Dr. Tertuliano Franco Santos Franco (UFBA)

Otávio de Macedo Menezes

Prof. Dr. Otávio de Macedo Menezes (UFBA)

Renato Soares

Prof. Dr. Renato Soares dos Santos (UFMG)

S237

Santos Júnior, Jozito Costa dos

Cadeias de Markov e tempo de mistura / Jozito Costa dos Santos
Júnior – Salvador 2023

72 f.

Orientador: Prof. Dr. Dirk Erhard

Dissertação (Mestrado) – Universidade Federal da Bahia, Instituto
de Matemática e Estatística, 2023.

1. Probabilidade. 2. Matemática. 3. Cadeias de Markov. I. Erhard,
Dirk. II. Universidade Federal da Bahia. III. Título.

CDU : 519.2

Agradecimentos

Agradeço, primeiramente, a Deus, pois sem Ele não teria chegado até aqui. Foi Ele quem construiu e pavimentou a estrada que percorri. Deus foi o responsável por esse milagre, pois mesmo sem merecer, Ele permanece fiel. Thanks God!

Meu muito obrigado, Professor Dirk, pelo empenho, dedicação, profissionalismo e pela grande paciência durante esse período de orientação. Encontrou um aluno meio perdido, buscando Álgebra e tendo a Probabilidade como última opção, mas que me fez gostar bastante da sua área, que agora chamo de minha. Agradeço também aos Professores Benigno e Tertuliano, que foram bastante sensíveis à minha situação e me possibilitaram cursar suas disciplinas.

Agradeço à minha esposa, Suellen, que disponibilizou seu tempo para assumir algumas tarefas minhas, permitindo-me concentrar apenas nos estudos. No entanto, não apenas por isso, mas também por dedicar suas orações em prol do nosso futuro e de nossa família. Meu amor, você é mais do que especial para mim. Te amo!

Agradeço a Deus pelo meu filho Joe Lucca, minha alegria e escape nos dias difíceis. Em várias aulas ele me "ajudando" na concentração e querendo aparecer nas chamadas e fotos da turma.

Agradeço aos meus pais Jozito e Maria da Luz que sonharam esse sonho junto comigo. Meus irmãos Joseval e Juliana, essa vitória é de vocês também! Ah, não posso esquecer minha sogra Wânia e minhas cunhadas Lais e Suenny!

Agradeço aos meus amigos Carol, Ênio, Janara, Mirele e Sara, que me auxiliaram nessa caminhada. Obrigado pela disponibilidade em ajudar, tanto nos estudos quanto no aprendizado do LaTeX.

Agradeço aos meus amigos da Igreja Boas Novas, que estiveram orando por mim e me incentivando. Agradeço também as meus colegas de profissão dos Colégios Polivalente, Emab e Edith.

Por fim, agradeço a Jorge Fonseca, que me deu a primeira oportunidade de lecionar em suas turmas no Curso Aprovado. Minha trajetória como professor teve início lá. Também expresso minha gratidão ao mestre Cristiano Mascarenhas, que sempre insistiu na minha vida universitária. Quando eu pensava em desistir, lá estava ele para elevar minha moral.

"Sem sacrificio, não há vitória"
Optimus Prime

Resumo

Neste trabalho fazemos o estudo de ferramentas que permitem modelar o problema do Caminho de Rede e do Embaralhamento de Cartas por Transposições Adjacentes. Utilizamos a técnica de Acoplamento entre cadeias de Markov para estudarmos sua taxa de convergência, isto é, o tempo necessário para que a distribuição da cadeia esteja bem perto da distribuição estacionária.

Palavras-chave: Cadeias de Markov; Tempos de mistura; Distância de variação total; Acoplamento; Caminhos de rede; Embaralhamento de cartas.

Abstract

In this work, we undertake the study of tools that enable the modeling of the Network Path and the Shuffle of Cards by Adjacent Transpositions problems. We employ the technique of coupling between Markov chains to investigate their convergence rate, i.e., the time required for the chain's distribution to be close to the stationary distribution.

Keywords: Markov Chains; Mixing times; Total Variation Distance; Coupling; Lattice path; Card shuffling.

Conteúdo

Introdução	1
1 Cadeia de Markov	3
1.1 Definições e exemplos	3
1.2 Distribuição estacionária	7
1.3 Distância de variação total	12
1.4 Acoplamento entre distribuições de probabilidade	16
1.5 Tempo de mistura	26
1.6 Teorema da Convergência	26
1.7 Acoplamento de Cadeias de Markov	29
2 Embaralhamento Top to random	37
2.1 O modelo	37
2.2 Coleta de cupons	38
3 Caminho de rede e Embaralhamento de cartas por transposições adjacentes	43
3.1 Cadeia de Markov no Caminho de rede	44
3.1.1 Propriedade de contração	44
3.1.2 Cota Superior	49
3.1.3 Cota Inferior	53
3.2 Embaralhamento de cartas por transposições adjacentes	58
Referências	62

Introdução

A Cadeia de Markov é um conceito cunhado pelo matemático russo Andrei Andreievich Markov, que estudou a alternância de vogais e consoantes no poema *Onegin* de Pushkin no início do século XX. Ele desenvolveu um modelo probabilístico no qual os resultados sucessivos dependiam de todos os predecessores apenas por meio de predecessores imediatos. O modelo lhe deu uma boa estimativa das frequências relativas das vogais na poesia. Na mesma época, o matemático francês Henri Poincaré estava trabalhando em sequências de variáveis aleatórias que eram na verdade cadeias de Markov. Pesquisas têm mostrado uma grande aplicação e utilidade das Cadeias de Markov em áreas como Física, Química, Medicina, Música, reconhecimento de fala através de sistemas de reconhecimento de voz, sistemas de informação e em aplicações da internet como o PageRank de uma página na web.

Para uma compreensão adequada deste texto, é essencial que o leitor possua conhecimento prévio em conceitos fundamentais e alguns resultados de Probabilidade. Entre eles, destacamos a Esperança Condicional, que foi usada em parte das demonstrações. Organizamos este trabalho em três capítulos, com o objetivo de facilitar a compreensão dos temas discutidos.

No Capítulo 1, nos concentramos na exposição teórica essencial para alcançarmos os resultados descritos nos capítulos subsequentes. Nesse capítulo, estabelecemos as definições fundamentais e conceitos relacionados às cadeias de Markov, bem como duas demonstrações do Teorema da Convergência, sendo uma delas conduzida por meio do acoplamento de cadeias de Markov.

No Capítulo 2, apresentamos o exemplo do Embaralhamento Top to random presente no livro [1]. Também apresentamos o exemplo da Coleta de Cupons, o qual foi utilizado para determinar estimativas para o tempo de mistura, isto é, o tempo no qual consideramos as cartas bem embaralhadas.

Por fim, no Capítulo 3, fundamentamo-nos no artigo de Wilson [2], no qual apresentamos as cadeias de Markov no Caminho de Rede e o processo de Embaralhamento

de cartas por transposições adjacentes. Obtivemos estimativas para o tempo de mistura de cada uma dessas cadeias. Normalmente, a análise do Embaralhamento de cartas por transposições adjacentes é realizada utilizando grupos de permutações. No entanto, conforme sugerido pelo artigo, desmembramos o problema e o analisamos por meio do Caminho de Rede.

Capítulo 1

Cadeia de Markov

Nesse capítulo serão introduzidos conceitos sobre cadeias de Markov, com o objetivo de fornecer ao leitor ferramentas básicas para a leitura do trabalho. Esse capítulo tem como referência o livro [1].

1.1 Definições e exemplos

Seja $\Omega = \{x_1, x_2, \dots, x_n\}$ um conjunto finito. Ω será chamado de espaço de estados, onde cada $x \in \Omega$ é chamado de estado.

Definição 1.1. Uma função $\mu : \Omega \rightarrow (0, 1)$ é uma distribuição de probabilidade em um conjunto finito Ω , se:

$$i) \mu(x) \geq 0,$$

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

Desta forma dado um conjunto $B \subset \Omega$, a probabilidade de B é dada por: $\mu(B) = \sum_{x \in B} \mu(x)$.

Definição 1.2. Uma matriz quadrada P é **estocástica** quando suas entradas são reais não negativas e a soma em cada linha é sempre igual a 1.

Exemplo 1.3. Matrizes estocásticas de ordens 2 e 3.

$$P_1 = \begin{pmatrix} 0,3 & 0,7 \\ 0,6 & 0,4 \end{pmatrix},$$

$$P_2 = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}, \text{ onde } 0 \leq p \leq 1 \text{ e } 0 \leq q \leq 1,$$

$$P_3 = \begin{pmatrix} 0,1 & 0,3 & 0,6 \\ 0 & 0,5 & 0,5 \\ 1 & 0 & 0 \end{pmatrix}.$$

Note que naturalmente podemos associar a cada linha de uma matriz estocástica a uma distribuição $\mu \in \Omega$. O que nos motiva a definir Cadeias de Markov utilizando matrizes, pois uma Cadeia de Markov finita pode ser entendida como um processo que se move entre os elementos de um conjunto finito Ω da seguinte maneira: quando $x \in \Omega$, a próxima posição é escolhida de acordo com a distribuição de probabilidade fixa $P(x, \cdot)$.

Definição 1.4. *Uma sequência de variáveis aleatórias (X_0, X_1, \dots) é uma Cadeia de Markov com espaço de estados Ω e matriz de transição P se, para todo $x, y \in \Omega, t \geq 1$, e todos os eventos $H_{t-1} = \bigcap_{s=0}^{t-1} \{X_s = x_s\}$ satisfazendo $P(H_{t-1} \cap \{X_t = x\}) > 0$, nós temos: $P\{X_{t+1} = y | H_{t-1} \cap \{X_t = x\}\} = P\{X_{t+1} = y | X_t = x\} = P(x, y)$.*

Essa última equação é conhecida como Propriedade de Markov, isto significa que a probabilidade condicional de proceder do estado x para o estado y é a mesma, não importando a sequência de estados anteriores a $x : x_0, x_1, \dots, x_{t-1}$. É exatamente por isso que a matriz quadrada P de ordem $|\Omega|$ é suficiente para descrever as transições.

$$P = \begin{pmatrix} P(x_0, x_0) & P(x_0, x_1) & \cdots & P(x_0, x_n) \\ P(x_1, x_0) & P(x_1, x_1) & \cdots & P(x_1, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ P(x_i, x_0) & P(x_i, x_1) & \cdots & P(x_i, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ P(x_n, x_0) & P(x_n, x_1) & \cdots & P(x_n, x_n) \end{pmatrix}$$

Note que a x_i -linha de P é a distribuição $P(x_i, \cdot)$. Então P é estocástica, isto é, todas as suas entradas são não negativas e $\sum_{y \in \Omega} P(x, y) = 1$, para todo $x \in \Omega$.

Exemplo 1.5. *Um certo sapo vive em um lago com dois lírios leste(l) e oeste(o). Há muito tempo, ele encontrou duas moedas no fundo do lago e levou uma para cada lírio. Todas as manhãs, o sapo decide se deve pular jogando a moeda do lírio atual. Se a moeda cair cara, o sapo pula para o outro lírio. Se a moeda cair coroa, ele permanece onde está. Sejam $\Omega = (l, o)$ o espaço de estados e (X_0, X_1, \dots) a sequência de lírios ocupados pelo sapo no Domingo, na Segunda, ... Suponha que a moeda do leste tem probabilidade p de cair cara, e a do oeste, tenha q de cair esse mesmo resultado, com $p, q \neq 0$.*

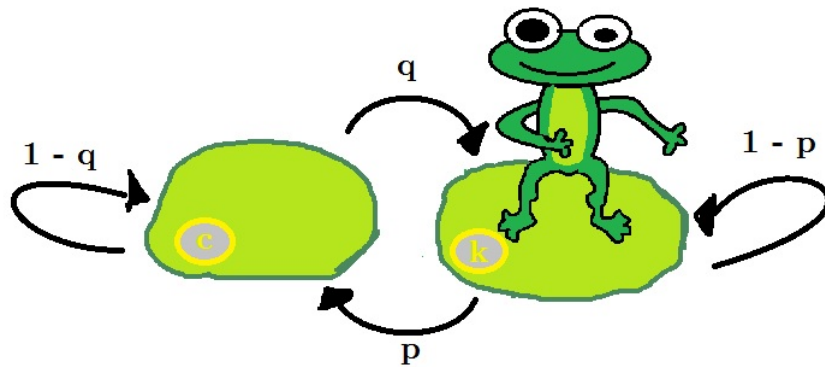


Figura 1.1: diagrama ilustrado para a cadeia de Markov do exemplo (1.5)

As regras do sapo para pular implicam que, se definirmos

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

então (X_0, X_1, \dots) é uma Cadeia de Markov com matriz de transição P . Observe que a primeira linha de P é a distribuição condicional de X_{t+1} quando $X_t = l$, enquanto a segunda linha é a distribuição condicional de X_{t+1} quando $X_t = o$.

Suponha que o sapo passe o domingo no bloco leste. Quando ele acorda na segunda-feira, ele tem probabilidade p de se mudar para o bloco oeste e probabilidade $1 - p$ de ficar no lírio leste. Isto é

$$P\{X_1 = l, X_0 = l\} = 1 - p \text{ e } P\{X_1 = o, X_0 = l\} = p.$$

O que acontece terça-feira? Considerando as duas possibilidades para X_1 , vemos que

$$P\{X_2 = l, X_0 = l\} = (1 - p)(1 - p) + pq \text{ e } P\{X_2 = o, X_0 = l\} = (1 - p)p + p(1 - q).$$

Embora pudéssemos continuar escrevendo fórmulas como essas, podemos armazenar nossas informações de distribuição em um vetor linha

$$\mu_t := (P\{X_t = l, X_0 = l\}, P\{X_t = o, X_0 = l\}).$$

Nossa suposição de que o sapo começa no bloco leste pode agora ser escrita como

$$\mu_0 = (1, 0), \quad \mu_1 = \mu_0 P, \quad \mu_2 = \mu_1 P = \mu_0 P^2, \dots$$

Desta forma multiplicar por P à direita atualiza a distribuição em outra etapa:

$$\mu_t = \mu_{t-1} P, \text{ para todo } t \geq 1.$$

De fato, para qualquer distribuição μ_0 ,

$$\mu_t = \mu_0 P^t, \text{ para todo } t \geq 0.$$

Escrevendo $\mu_0 = (P(X_0 = l), P(X_0 = o)) = (1, 0)$, daí podemos obter

$$\begin{aligned}\mu_1 &= \mu_0 P = (1 - p, p) \\ \mu_2 &= \mu_1 P = ((1 - p)^2 + pq, (1 - p)p + p(1 - q)).\end{aligned}$$

Definição 1.6. Dizemos que uma Cadeia de Markov com matriz de transição P é irreduzível se para quaisquer dois estados $x, y \in \Omega$ existe um inteiro t tal que $P^t(x, y) > 0$, onde P^t é a matriz de transição para t passos.

Ser irreduzível significa que sempre é possível irmos, após um certo número de t passos, do estado x para o estado y . Dita de outra forma, para quaisquer $x, y \in \Omega$, existe uma sequência de saltos com probabilidade positiva que sai de x e chega em y , isto é, existem $x = x_0, x_1, \dots, x_n = y$ tais que

$$P(x_0, x_1)P(x_1, x_2) \cdots P(x_{n-1}, x_n) > 0.$$

No exemplo (1.5) temos que $P(x, y) > 0, \forall x, y \in \Omega$, logo essa cadeia de Markov é irreduzível.

Definição 1.7. O conjunto $\tau(x) := \{t \geq 1 : P^t(x, x) > 0\}$ é o conjunto de tempos de retorno de x , e indica os tempos t tais que é possível sair do estado x , e retornar a ele próprio exatamente um tempo t depois. Chamamos de período do estado x o MDC dos tempos de retorno de x , ou seja, $MDC(\tau(x))$.

Definição 1.8. Diz-se que uma cadeia de Markov P com espaço de estados Ω é aperiódica se, para todo $x \in \Omega$, x possui período 1. Caso contrário, diz-se que P é periódica.

Proposição 1.9. Se P é uma cadeia de Markov irreduzível e aperiódica, com espaço de estados Ω , então existe um inteiro não-negativo r tal que $P^r(x, y) > 0$, para todo $x, y \in \Omega$.

Essa proposição será importante na demonstração do Teorema da Convergência. A demonstração de 1.9 pode ser encontrada no livro [1].

Exemplo 1.10. Tome uma Cadeia de Markov com espaço de estados $\{1, 2, 3, 4, 5\}$ e seja P sua matriz de transição:

$$P = \begin{pmatrix} 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1/4 & 1/2 & 1/4 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 2/5 & 3/5 \\ 0 & 0 & 0 & 1/5 & 4/5 \end{pmatrix} = \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix}$$

$$\text{onde } P_1 = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/4 & 1/2 & 1/4 \\ 1/2 & 0 & 1/2 \end{pmatrix} \text{ e } P_2 = \begin{pmatrix} 2/5 & 3/5 \\ 1/5 & 4/5 \end{pmatrix}.$$

Note que se começarmos com estado inicial x no conjunto $\{1, 2, 3\}$, teremos que os próximos estados estarão nesse conjunto. Da maneira semelhante se começarmos com estado inicial y no conjunto $\{4, 5\}$. Isto significa que, para qualquer inteiro t , temos: $P^t(x, y) = P^t(y, x) = 0$ com $x \in \{1, 2, 3\}$ e $y \in \{4, 5\}$. Portanto essa matriz é redutível.

1.2 Distribuição estacionária

Analisando o problema do Exemplo (1.5) temos que $\mu_t = \mu_{t-1}P$, para todo $t \geq 1$. Queremos saber como a distribuição μ_t se comporta a longo prazo, isto é, quando $t \rightarrow \infty$. Se essa distribuição converge para uma distribuição π , então deve satisfazer $\pi = \pi P$, e então, queremos saber quais condições uma Cadeia de Markov deve satisfazer para que possamos garantir a existência de π e sua unicidade.

Definição 1.11. (*Distribuição estacionária*) Dada uma Cadeia de Markov (X_0, X_1, \dots) com matriz de transição P , uma distribuição de probabilidade π em Ω é dita estacionária, ou invariante, se $\pi = \pi P$, ou de forma equivalente, se $\pi(y) = \sum_{x \in \Omega} \pi(x)P(x, y)$, $\forall x, y \in \Omega$.

Retomando o Exemplo (1.5), seja $\pi = (\pi(l), \pi(o))$, onde $\pi(l) + \pi(o) = 1$, então para encontrar essa distribuição, é suficiente resolver o sistema de equações:

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

que nos dá como solução: $\pi(l) = \frac{q}{p+q}$ e $\pi(o) = \frac{p}{p+q}$.

Exemplo 1.12. Um grafo $G = (V, E)$ não ordenado consiste em um conjunto de vértices V e um conjunto de arestas E . Onde E é formado por pares de vértices, ou seja, $E \subseteq \{(x, y); x, y \in V, x \neq y\}$. Quando $(x, y) \in E$ escreveremos $x \sim y$, isto significa que x é vizinho de y , isto é, x e y formam uma aresta. O número de vizinhos de x , o grau de x , será denotado por $\text{deg}(x)$. Então, dado um grafo $G = (V, E)$, definimos o passeio aleatório simples em G como sendo a cadeia de Markov com espaço de estado V e matriz de transição:

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

Ou seja, 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.

Seja π a distribuição de probabilidade estacionária, notemos que para todo vértice $y \in V$ temos:

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

E como $\sum_{y \in V} \deg(y) = 2|E|$, então π terá coordenadas $\pi(y) = \frac{\deg(y)}{2|E|}, \forall y \in V$.

De fato: $\sum_{y \in V} \pi(y) = 1$ e

$$\sum_{x \in V} \pi(x)P(x, y) = \sum_{x \sim y} \left(\frac{\deg(x)}{2|E|} \cdot \frac{1}{\deg(x)} \right) = \sum_{x \sim y} \frac{1}{2|E|} = \pi(y), \forall y \in V.$$

Exemplo 1.13. Considere uma cadeia de Markov que tenha como matriz de transição uma matriz simétrica P . Se P é simétrica, ela é igual à sua transposta, isto nos diz que $P(x, y) = P(y, x)$. Afirmamos que a distribuição uniforme é a distribuição estacionária dessa cadeia.

De fato, seja $|\Omega| = n$, então $\pi(x) = \frac{1}{n}, \forall x \in \Omega$, já que

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

Exemplo 1.14. Uma probabilidade μ em Ω é dita reversível, com respeito a matriz estocástica P , se

$$\pi(x)P(x, y) = \pi(y)P(y, x), \forall x, y \in \Omega.$$

Afirmamos que se μ é reversível, então μ é invariante.

De fato, seja μ reversível, temos:

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

Exemplo 1.15. Passeio aleatório em Grupos. Dada uma distribuição de probabilidade μ em um grupo (G, \cdot) , definimos a caminhada aleatória em G com distribuição de incremento μ da seguinte forma: é uma cadeia de Markov com espaço de estados G , que se move multiplicando o estado atual à esquerda por um elemento aleatório de G selecionado de acordo com μ . De forma equivalente, a matriz de transição P dessa cadeia tem entradas

$$P(g, hg) = \mu(h)$$

para todo $g, h \in G$.

Proposição 1.16. *Seja P a matriz de transição de uma caminhada aleatória em um grupo finito G e seja U a distribuição de probabilidade uniforme em G . Então, U é uma distribuição estacionária para P .*

Demonstração. Considere μ como a distribuição de incremento da caminhada aleatória. Para qualquer $g \in G$,

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

Para a primeira igualdade, reindexamos definindo $k = gh^{-1}$. \square

Definição 1.17. *(Tempos de chegada e de retorno) Seja (X_0, X_1, \dots) uma Cadeia de Markov com espaço de estados finitos Ω e matriz de transição P . Para $x \in \Omega$, definimos o **tempo de chegada (hitting time)** para x como*

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

a primeira vez que a cadeia visita o estado x . Para situações onde isso acontece apenas em um tempo positivo, definimos também

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

*Quando $X_0 = x$ chama-se τ_x^+ de **primeiro tempo de retorno (first return time)**.*

Lema 1.18. *Para quaisquer estados x e y de uma Cadeia de Markov irredutível,*

$$E_x(\tau_y^+) = E(\tau_y^+ | X_0 = x) < \infty.$$

Demonstração. A definição de irredutibilidade implica que existe um inteiro $r > 0$ e um real $\varepsilon > 0$ tal que para quaisquer estados $z, w \in \Omega$, existe um $j \leq r$ com $P^j(z, w) > \varepsilon$. Assim, para qualquer valor de X_t , a probabilidade de alcançar o estado y em um tempo entre t e $t + r$ é pelo menos ε . Portanto, para $k > 0$ temos

$$P_x(\tau_y^+ > kr) = \sum_{z_0, \dots, z_{(k-1)r} \neq y} P_x(\tau_y^+ > kr, H_{(z-2)r}, \{X_{(k-1)r} = z_{(k-1)r}\}), \quad (1.1)$$

onde $H_{(z-2)r} = \bigcap_{s=0}^{(z-2)r} \{X_s = z_s\}$. Denote a probabilidade dentro do somatório de W , desenvolvendo-a temos:

$$\begin{aligned} W &= P_x(\tau_y^+ > kr, H_{(z-2)r}, \{X_{(k-1)r} = z_{(k-1)r}\}) \\ &= P_x(\tau_y^+ > kr | H_{(z-2)r}, \{X_{(k-1)r} = z_{(k-1)r}\}) P_x(H_{(z-2)r}, \{X_{(k-1)r} = z_{(k-1)r}\}) \\ &= P_x(\tau_y^+ > kr | X_{(k-1)r} = z_{(k-1)r}) P_x(H_{(z-2)r}, \{X_{(k-1)r} = z_{(k-1)r}\}) \\ &= P_{z_{(k-1)r}}(\tau_y^+ > r) P_x(H_{(z-2)r}, \{X_{(k-1)r} = z_{(k-1)r}\}) \\ &\leq (1 - \varepsilon) P_x(H_{(z-2)r}, \{X_{(k-1)r} = z_{(k-1)r}\}). \end{aligned}$$

Juntando esse resultado com (1.1), temos

$$\begin{aligned}
P_x(\tau_y^+ > kr) &\leq \sum_{z_0, \dots, z_{(k-1)r} \neq y} (1 - \varepsilon) P_x(H_{(z-2)r}, \{X_{(k-1)r} = z_{(k-1)r}\}) \\
&= (1 - \varepsilon) \sum_{z_0, \dots, z_{(k-1)r} \neq y} P_x(H_{(z-2)r}, \{X_{(k-1)r} = z_{(k-1)r}\}) \\
&= (1 - \varepsilon) \sum_{z_0, \dots, z_{(k-1)r} \neq y} P\left(\bigcup_{z_0, \dots, z_{(k-1)r} \neq y} \{X_0 = z_0, \dots, X_{(k-1)r} = z_{(k-1)r}\}\right) \\
&= (1 - \varepsilon) P(\{\tau_y^+ > (k-1)r\})
\end{aligned} \tag{1.2}$$

Repetindo (1.2), obteremos

$$P_x(\tau_y^+ > kr) \leq (1 - \varepsilon)^k \tag{1.3}$$

Lembre-se de que quando Y é uma variável aleatória de valor inteiro não negativo, temos

$$\sum_{t \geq 0} P(Y > t) = \sum_{t \geq 0} E[\mathbf{1}_{Y > t}] = E\left[\sum_{t \geq 0} \mathbf{1}_{Y > t}\right] = E\left[\sum_{t \geq 0}^{Y-1} \mathbf{1}_{Y > t}\right] = E[Y].$$

Como $P_x(\tau_y^+ > t)$ é uma função decrescente de t , (1.3) é suficiente para juntar todos os termos da expressão correspondente para $E_x(\tau_y^+)$:

$$E_x(\tau_y^+) = \sum_{t \geq 0} P_x(\tau_y^+ > t) \leq \sum_{k \geq 0} r P_x(\tau_y^+ > kr) \leq \sum_{k \geq 0} r(1 - \varepsilon)^k \leq \infty.$$

□

De posse desse resultado agora podemos demonstrar a seguinte proposição que garante a existência e unicidade da distribuição estacionária.

Proposição 1.19. *Seja P a matriz de transição de uma Cadeia de Markov irredutível, então:*

(i) *Existe uma única distribuição de probabilidade $\pi \in \Omega$ tal que $\pi = \pi P$ e $\pi(x) > 0$ para todo $x \in \Omega$.*

(ii) $\pi(x) = \frac{1}{E_x(\tau_x^+)}.$

Demonstração. Seja $z \in \Omega$ um estado arbitrário da cadeia de Markov irredutível. Examinaremos de perto o tempo que a cadeia gasta, em média, em cada estado entre as visitas a z . Portanto, defina

$$\tilde{\pi}_z(y) := E_z(\text{número de visitas a } y \text{ antes de retornar a } z)$$

para cada estado y . Então podemos escrever

$$\tilde{\pi}_z(y) = E_z\left[\sum_{t=0}^{\infty} \mathbf{1}_{(X_t=y, \tau_z^+ > t)}\right] = \sum_{t=0}^{\infty} E_z[\mathbf{1}_{(X_t=y, \tau_z^+ > t)}] = \sum_{t=0}^{\infty} P_z\{X_t = y, \tau_z^+ > t\}.$$

Sabemos que $E_z \tau_z^+ = \sum_{t=0}^{\infty} P_z(\tau_z^+ > t)$, que implica $\tilde{\pi}_z(y) \leq E_z \tau_z^+$. Portanto o Lema (1.18) garante que $\tilde{\pi}_z(y) < \infty$ para todo $y \in \Omega$. Partindo da definição, mostraremos que $\tilde{\pi}_z$ é estacionária:

$$(\tilde{\pi}_z P)(y) = \sum_{x \in \Omega} \tilde{\pi}_z(x) P(x, y) = \sum_{x \in \Omega} \sum_{t=0}^{\infty} P_z\{X_t = x, \tau_z^+ > t\} P(x, y). \quad (1.4)$$

Como o evento $\{\tau_z^+ \geq t + 1\} = \{\tau_z^+ > t\}$ é determinado apenas por X_0, \dots, X_t , ele é independente do evento $X_{t+1} = y$, quando condicionado em $X_t = x$, o que nos dá

$$\begin{aligned} & P_z\{X_t = x, X_{t+1} = y, \tau_z^+ \geq t + 1\} \\ &= P_z\{X_t = x, \tau_z^+ \geq t + 1\} \frac{P_z\{X_t = x, X_{t+1} = y, \tau_z^+ \geq t + 1\}}{P_z\{X_t = x, \tau_z^+ \geq t + 1\}} \\ &= P_z\{X_t = x, \tau_z^+ \geq t + 1\} P_z\{X_{t+1} = y | X_t = x, \tau_z^+ \geq t + 1\} \\ &= P_z\{X_t = x, \tau_z^+ \geq t + 1\} P_z\{X_{t+1} = y | X_t = x\}. \end{aligned} \quad (1.5)$$

Dai,

$$P_z\{X_t = x, X_{t+1} = y, \tau_z^+ \geq t + 1\} = P_z\{X_t = x, \tau_z^+ \geq t + 1\} P(x, y). \quad (1.6)$$

Invertendo a ordem da soma em (1.4) e usando a identidade (1.6), temos que

$$\begin{aligned} (\tilde{\pi}_z P)(y) &= \sum_{t=0}^{\infty} \sum_{x \in \Omega} P_z\{X_t = x, X_{t+1} = y, \tau_z^+ \geq t + 1\} \\ &= \sum_{t=0}^{\infty} P_z\left\{ \bigcup_{x \in \Omega} [X_t = x] \cap [X_{t+1} = y, \tau_z^+ \geq t + 1] \right\} \\ &= \sum_{t=0}^{\infty} P_z\{X_{t+1} = y, \tau_z^+ \geq t + 1\} \\ &= \sum_{t=1}^{\infty} P_z\{X_t = y, \tau_z^+ \geq t\} \\ &= \sum_{t=0}^{\infty} P_z\{X_t = y, \tau_z^+ > t\} - P_z\{X_0 = y, \tau_z^+ > 0\} + \sum_{t=1}^{\infty} P_z\{X_t = y, \tau_z^+ = t\} \\ &= \tilde{\pi}_z(y) - P_z\{X_0 = y\} + P_z\{X_{\tau_z^+} = y\} \\ &= \tilde{\pi}_z(y). \end{aligned}$$

No último passo foi levado em conta que $P_z\{X_0 = y\} = P_z\{X_{\tau_z^+} = y\}$. De fato, se $y = z$, temos $P_z\{X_0 = y\} = P_z\{X_0 = z\} = 1$ e $P_z\{X_{\tau_z^+} = y\} = P_z\{X_{\tau_z^+} = z\} = 1$, pela definição de τ_z^+ . Se $y \neq z$, temos $P_z\{X_0 = y\} = 0$ e $P_z\{X_{\tau_z^+} = y\} = 0$.

Portanto $(\tilde{\pi}_z P)(y) = \tilde{\pi}_z(y)$ para todo $y \in \Omega$. Isso prova que $\tilde{\pi}_z$ é estacionário. Finalmente, para obter uma medida de probabilidade, normalizamos por $\sum_{x \in \Omega} \tilde{\pi}_z(x) = E_z(\tau_z^+)$:

$$\pi_z(x) = \frac{\tilde{\pi}_z(x)}{\sum_{x \in \Omega} \tilde{\pi}_z(x)} = \frac{\tilde{\pi}_z(x)}{E_z(\tau_z^+)}, \text{ que satisfaz } \pi_z = \pi_z P, \forall z \in \Omega.$$

Se $x = z$, temos $\tilde{\pi}_z(z) = 1$, pois é o número de visitas a z antes de retornar a z , temos:

$$\pi_z(z) = \frac{1}{E_z(\tau_z^+)}.$$

Se $x \neq z$, tomando $s \in \Omega$, daí:

$$\pi_s(x) = \frac{\tilde{\pi}_s(x)}{E_s(\tau_s^+)} \Rightarrow \pi_s(s) = \frac{1}{E_s(\tau_s^+)}.$$

Pela unicidade (assuma, pois provaremos logo em seguida), $\pi_z = \pi_s$, $\forall s, z \in \Omega$, daí $\pi_z(x) = \pi_s(x)$, $\forall s, x, z \in \Omega$ (isto significa que π_z não depende da escolha de z). Então tomando $x = s$ temos que:

$$\pi_z(s) = \pi_s(s) = \frac{1}{E_s(\tau_s^+)}, \text{ ou ainda, } \pi(s) = \frac{1}{E_s(\tau_s^+)}.$$

Para mostrar a unicidade, suponha que existam duas medidas estacionárias μ_1 e $\mu_2 \in \Omega$. Se $\frac{\mu_1(x)}{\mu_2(x)}$ for constante, então existe $C \in \mathbb{R}_+^*$ tal que,

$$\frac{\mu_1(x)}{\mu_2(x)} = C \Rightarrow C = C \sum_{x \in \Omega} \mu_2(x) = \sum_{x \in \Omega} C \mu_2(x) = \sum_{x \in \Omega} \mu_1(x) = 1,$$

para todo $x \in \Omega$. Logo $\mu_1 = \mu_2$.

Suponha agora que $\frac{\mu_1(x)}{\mu_2(x)}$ não seja constante, então existe um estado $s \in \Omega$ tal que $\frac{\mu_1(s)}{\mu_2(s)} \leq \frac{\mu_1(x)}{\mu_2(x)}$, para todo $x \in \Omega$. Também existe ao menos um $j \in \Omega$ tal que $\frac{\mu_1(s)}{\mu_2(s)} < \frac{\mu_1(j)}{\mu_2(j)}$. Como $\pi(x) > 0$ para todo $x \in \Omega$, lembrando que P é irredutível, temos:

$$\begin{aligned} \mu_1(s) &= \sum_{y \in \Omega} \mu_1(y) P(y, s) \\ &= \sum_{y \in \Omega} \frac{\mu_1(y)}{\mu_2(y)} \mu_2(y) P(y, s) \\ &> \sum_{y \in \Omega} \frac{\mu_1(s)}{\mu_2(s)} \mu_2(y) P(y, s) \\ &= \frac{\mu_1(s)}{\mu_2(s)} \sum_{y \in \Omega} \mu_2(y) P(y, s) = \mu_1(s), \end{aligned}$$

gerando um absurdo! Portanto o quociente entre as distribuições é constante. \square

1.3 Distância de variação total

Agora estamos prontos para discutir o comportamento de longo prazo de cadeias de Markov finitas. Como estamos interessados em quantificar a velocidade de convergência de famílias de cadeias de Markov, precisamos escolher uma métrica apropriada para medir a distância entre as distribuições.

Definição 1.20. A distância de variação total entre duas distribuições de probabilidade μ e ν em Ω é definida por

$$\|\mu - \nu\|_{TV} = \max_{A \subset \Omega} |\mu(A) - \nu(A)|. \quad (1.7)$$

Esta definição é explicitamente probabilística: a distância entre μ e ν é a diferença máxima entre as probabilidades atribuídas a um único evento pelas duas distribuições.

No Exemplo (1.5), sua matriz de transição é $\begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$ e sua distribuição estacionária é $\pi = \left(\frac{q}{p+q}, \frac{p}{p+q}\right)$. Suponha que o sapo comece na almofada leste, isto é, $\pi = (1, 0)$ e defina $\Delta_t = \mu_t(l) - \pi(l)$. Como existem apenas dois estados, existem apenas quatro eventos possíveis com $A \subset \Omega$, esses quatro eventos são $\emptyset, \{l\}, \{o\}$ e Ω .

$$\|\mu - \pi\|_{TV} = \Delta_t = |P^t(l, l) - \pi(l)| = |\pi(o) - P^t(l, o)|.$$

Como nem sempre é fácil calcular a distância de variação total pela definição, agora daremos três caracterizações alternativas extremamente úteis.

Proposição 1.21. Sejam μ e ν duas distribuições de probabilidade em Ω . Então

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|. \quad (1.8)$$

Demonstração. Seja $B = \{x : \mu(x) \geq \nu(x)\}$ e seja $A \subset \Omega$ um evento qualquer. Escrevendo A como uma união disjunta $(A \cap B) \cup (A \cap B^c)$, temos:

$$\begin{aligned} \mu(A) - \nu(A) &= \sum_{x \in A} \mu(x) - \sum_{x \in A} \nu(x) \\ &= \left(\sum_{x \in A \cap B} \mu(x) + \sum_{x \in A \cap B^c} \mu(x) \right) - \left(\sum_{x \in A \cap B} \nu(x) + \sum_{x \in A \cap B^c} \nu(x) \right) \\ &= \left(\sum_{x \in A \cap B} \mu(x) - \sum_{x \in A \cap B} \nu(x) \right) + \left(\sum_{x \in A \cap B^c} \mu(x) - \sum_{x \in A \cap B^c} \nu(x) \right). \end{aligned}$$

Observe que se $x \in A \cap B^c$, então $\mu(x) < \nu(x) \Rightarrow \sum_{x \in A \cap B^c} \mu(x) - \sum_{x \in A \cap B^c} \nu(x) < 0$, daí temos que:

$$\mu(A) - \nu(A) \leq \mu(A \cap B) - \nu(A \cap B) \quad (1.9)$$

De maneira semelhante, escrevendo $B = (A \cap B) \cup (A^c \cap B)$, temos:

$$\begin{aligned} \mu(B) - \nu(B) &= \sum_{x \in B} \mu(x) - \sum_{x \in B} \nu(x) \\ &= \left(\sum_{x \in A \cap B} \mu(x) + \sum_{x \in A^c \cap B} \mu(x) \right) - \left(\sum_{x \in A \cap B} \nu(x) + \sum_{x \in A^c \cap B} \nu(x) \right) \\ &= \left(\sum_{x \in A \cap B} \mu(x) - \sum_{x \in A \cap B} \nu(x) \right) + \left(\sum_{x \in A^c \cap B} \mu(x) - \sum_{x \in A^c \cap B} \nu(x) \right). \end{aligned}$$

Agora note que se $x \in A^c \cap B$, então $\mu(x) \geq \nu(x) \Rightarrow \sum_{x \in A^c \cap B} \mu(x) - \sum_{x \in A^c \cap B} \nu(x) \geq 0$, daí temos que:

$$\mu(B) - \nu(B) \geq \mu(A \cap B) - \nu(A \cap B). \quad (1.10)$$

Então pelas equações (1.9) e (1.10) temos que:

$$\mu(A) - \nu(A) \leq \mu(B) - \nu(B). \quad (1.11)$$

Como tomamos A qualquer, podemos escrever $\mu(A^c) - \nu(A^c) \leq \mu(B) - \nu(B)$, por outro lado temos:

$$\mu(A^c) - \nu(A^c) = 1 - \mu(A) - (1 - \nu(A)) = -(\mu(A) - \nu(A)),$$

com isso podemos reescrever a desigualdade (1.11) como $-(\mu(A) - \nu(A)) \leq \mu(B) - \nu(B)$, que é equivalente a $-(\mu(B) - \nu(B)) \leq \mu(A) - \nu(A)$. Portanto,

$$|\mu(A) - \nu(A)| \leq \mu(B) - \nu(B). \quad (1.12)$$

Também temos que $\mu(B) - \nu(B) = 1 - \mu(B^c) - (1 - \nu(B^c)) = \nu(B^c) - \mu(B^c)$, então podemos escrever a desigualdade (1.12) como:

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

Quando tomamos $A = B$ (ou $A = B^c$), então $|\mu(A) - \nu(A)|$ é igual ao seu maximal, então:

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

□

A prova da Proposição 1.21 também mostra que:

$$\|\mu - \nu\|_{TV} = \sum_{\substack{x \in \Omega \\ \mu(x) \geq \nu(x)}} [\mu(x) - \nu(x)]. \quad (1.13)$$

Demonstração. De fato, pela Proposição 1.21 temos:

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

□

Podemos ainda dizer que a distância de variação total é uma métrica no conjunto de todas as distribuições de probabilidade em Ω . De fato, quaisquer que sejam as distribuições μ , ν e η em Ω , temos:

- i) $\|\mu - \nu\|_{TV} = \max_{A \subset \Omega} |\mu(A) - \nu(A)| \geq 0$;
- ii) $\|\mu - \nu\|_{TV} = \max_{A \subset \Omega} |\mu(A) - \nu(A)| = \max_{A \subset \Omega} |\nu(A) - \mu(A)| = \|\nu - \mu\|_{TV}$;
- iii) Usando a Proposição 1.21 juntamente com a desigualdade triangular temos:

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

Proposição 1.22. *Sejam μ e ν duas distribuições de probabilidade em Ω . Então a distância de variação total entre elas satisfazem:*

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sup \left\{ \sum_{x \in \Omega} f(x)\mu(x) - \sum_{x \in \Omega} f(x)\nu(x); \max_{x \in \Omega} |f(x)| \leq 1 \right\} \quad (1.14)$$

Demonstração. Quando f satisfaz $\max_{x \in \Omega} |f(x)| \leq 1$, então nós temos:

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

o que mostra que o lado direito de (1.9) não é maior que $\|\mu - \nu\|_{TV}$. Agora defina

$$f^*(x) = \begin{cases} 1, & \text{se } \mu(x) \leq \nu(x), \\ -1, & \text{se } \mu(x) > \nu(x). \end{cases}$$

Então:

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

E como

$$\frac{1}{2} \sup \left\{ \sum_{x \in \Omega} f(x)\mu(x) - \sum_{x \in \Omega} f(x)\nu(x); \max_{x \in \Omega} |f(x)| \leq 1 \right\} \geq \frac{1}{2} \left[\sum_{x \in \Omega} f^*(x)\mu(x) - \sum_{x \in \Omega} f^*(x)\nu(x) \right],$$

segue que

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

Com isso completamos nossa demonstração. \square

1.4 Acoplamento entre distribuições de probabilidade

Definição 1.23. Um acoplamento de duas distribuições de probabilidade μ e ν é um par de variáveis aleatórias (X, Y) , definido em um único espaço de probabilidade (Ω, \mathcal{F}, P) , tal que a distribuição marginal de X é μ e a distribuição marginal de Y é ν . Ou seja, um acoplamento (X, Y) satisfaz $P\{X = x\} = \mu(x)$ e $P\{Y = y\} = \nu(y)$, com x e y em Ω .

O acoplamento é uma técnica geral e poderosa; pode ser aplicado em vários caminhos. Vamos estudar essa técnica de forma introdutória, mostrando a estreita conexão entre acoplamentos de duas variáveis aleatórias e a distância de variação total entre essas variáveis.

Dado um acoplamento (X, Y) de μ e ν , se q é a distribuição conjunta de (X, Y) em $\Omega \times \Omega$, isto significa que $q(x, y) = P\{X = x, Y = y\}$, então q satisfaz:

$$\sum_{y \in \Omega} q(x, y) = \sum_{y \in \Omega} P\{X = x, Y = y\} = P\{X = x\} = \mu(x)$$

e

$$\sum_{x \in \Omega} q(x, y) = \sum_{x \in \Omega} P\{X = x, Y = y\} = P\{Y = y\} = \nu(y).$$

Por outro lado, dado uma distribuição de probabilidade q no espaço produto $\Omega \times \Omega$ que satisfaz:

$$\sum_{y \in \Omega} q(x, y) = \mu(x) \text{ e } \sum_{x \in \Omega} q(x, y) = \nu(y).$$

Então existirá um par de variáveis aleatórias (X, Y) tendo q como sua distribuição conjunta. E conseqüentemente este par (X, Y) é um acoplamento de μ e ν .

Exemplo 1.24. *Considere o lançamento de duas moedas honestas e associe o valor 1 à face cara e 0 à face coroa. E sejam μ e ν duas distribuições de probabilidade definidas em $\{0, 1\}$. Veja duas maneiras que podemos acoplar μ e ν :*

i) Uma maneira de acoplar μ e ν é definir (X, Y) como um par de resultados de lançamentos de moedas honestas e independentes, tal que $P\{X = x, Y = y\} = \frac{1}{4}$, para todo $x, y \in \{0, 1\}$. A probabilidade conjunta q_1 em $\{0, 1\} \times \{0, 1\}$ é dada por:

$$q_1(x, y) = \frac{1}{4}, \quad \forall (x, y) \in \{0, 1\}^2.$$

ii) Outra maneira de acoplar μ e ν é tomar X como cara ou coroa e fazer $Y = X$. Nesse caso temos $P\{X = Y = 0\} = P\{X = Y = 1\} = \frac{1}{2}$, e $P\{X \neq Y\} = 0$. Então a probabilidade conjunta q_2 é dada por:

$$q_2(x, y) = \begin{cases} \frac{1}{2}, & \text{se } (x, y) = (0, 0), (x, y) = (1, 1) \\ 0, & \text{se } (x, y) = (0, 1), (x, y) = (1, 0). \end{cases}$$

Exemplo 1.25. *Imagine agora o lançamento de um dado honesto de 6 faces numerados de 1 a 6 e de uma moeda honesta. Associe a cada face o valor do seu número e a moeda o valor 1 à face cara e 2 à face coroa. Desta forma, considere μ e ν duas distribuições em $\Omega = \{1, 2, 3, 4, 5, 6\}$. Podemos acoplar μ e ν definindo o par (X, Y) com X os resultados*

do lançamento do dado e Y os do lançamento da moeda, daí: $P\{X = x, Y = y\} = \frac{1}{12}$, para todo $x \in \Omega$ e $y \in \{1, 2\}$ e $P\{X = x, Y = y\} = 0$, para todo $x \in \Omega$ e $y \in \{3, 4, 5, 6\}$. Logo, a probabilidade conjunta é:

$$q(x, y) = \begin{cases} \frac{1}{12}, & \forall (x, y) \in \Omega \times \{1, 2\}. \\ 0, & \text{caso contrário.} \end{cases}$$

Quaisquer duas distribuições μ e ν têm um acoplamento independente. No entanto, quando μ e ν não são idênticos, não será possível que X e Y tenham sempre o mesmo valor. Quão perto um acoplamento pode chegar de ter X e Y idênticos? A distância de variação nos dá a resposta. A seguinte proposição nos permite limitar inferiormente $P\{X \neq Y\}$.

Proposição 1.26. *Sejam μ e ν duas distribuições de probabilidade em Ω . Então:*

$$\|\mu - \nu\|_{TV} = \inf\{P\{X \neq Y\}; (X, Y) \text{ é um acoplamento de } \mu \text{ e } \nu\}. \quad (1.15)$$

Mostraremos, de fato, que existe um acoplamento (X, Y) que atinge o ínfimo em (1.15). Chamaremos esse acoplamento de ótimo.

Demonstração. Primeiro notamos que para qualquer acoplamento (X, Y) de μ e ν e qualquer evento $A \subset \Omega$, temos:

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

De maneira semelhante podemos obter: $\nu(A) - \mu(A) \leq P\{Y \neq X\} = P\{X \neq Y\}$. Daí segue que:

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

Vamos mostrar a igualdade, para isto vamos construir um acoplamento tal que $P\{X \neq Y\}$ seja exatamente igual a $\|\mu - \nu\|_{TV}$. Faremos isso forçando X e Y a serem iguais sempre que

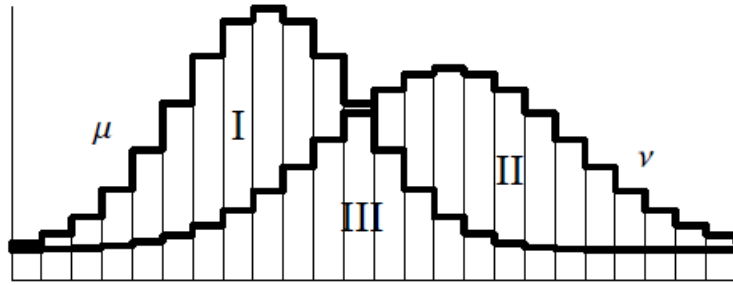


Figura 1.2: Como cada uma das regiões I e II tem área $\|\mu - \nu\|_{TV}$ e μ e ν são medidas de probabilidades, a região III tem área $1 - \|\mu - \nu\|_{TV}$. A figura foi retirado do livro [1].

possível. Considere a Figura (1.2), a região III, limitada por $\mu(x) \wedge \nu(x) = \min\{\mu(x), \nu(x)\}$ pode ser visto como a sobreposição entre as duas distribuições. Informalmente, nosso acoplamento procede escolhendo um ponto na união das regiões I, II e III. Sempre que “pousarmos” na região III, tomamos $X = Y$. Caso contrário, aceitamos que X deve estar na região I e Y deve estar na II.

Uma vez que essas regiões tem suporte disjunto, X e Y não pode ser igual. Mais formalmente usamos o seguinte procedimento para gerar X e Y . Seja

$$p = \sum_{x \in \Omega} \mu(x) \wedge \nu(x). \quad (1.16)$$

escreva

$$\sum_{x \in \Omega} \mu(x) \wedge \nu(x) = \sum_{\substack{x \in \Omega \\ \mu(x) \leq \nu(x)}} \mu(x) + \sum_{\substack{x \in \Omega \\ \mu(x) > \nu(x)}} \nu(x)$$

adicionando e subtraindo $\sum_{\substack{x \in \Omega \\ \mu(x) > \nu(x)}} \mu(x)$ no lado direito da Equação (1.16) temos:

$$\begin{aligned} \sum_{x \in \Omega} \mu(x) \wedge \nu(x) &= \sum_{\substack{x \in \Omega \\ \mu(x) \leq \nu(x)}} \mu(x) + \sum_{\substack{x \in \Omega \\ \mu(x) > \nu(x)}} \nu(x) - \sum_{\substack{x \in \Omega \\ \mu(x) > \nu(x)}} \mu(x) + \sum_{\substack{x \in \Omega \\ \mu(x) > \nu(x)}} \mu(x) \\ &= 1 - \sum_{\substack{x \in \Omega \\ \mu(x) > \nu(x)}} [\mu(x) - \nu(x)]. \end{aligned}$$

Obtemos assim:

$$\sum_{z \in \Omega} \mu(z) \wedge \nu(z) = 1 - \|\mu - \nu\|_{TV} = p, \quad (1.17)$$

onde nós usamos (1.13) para a primeira igualdade. Agora, jogue uma moeda com probabilidade de dar cara igual a p .

i) Se a moeda der cara, então escolha uma valor Z de acordo com a distribuição de probabilidade:

$$\gamma_{III}(x) = \frac{\mu(x) \wedge \nu(x)}{p} \text{ e defina } X = Y = Z.$$

ii) Se a moeda der coroa, escolha X de acordo com a distribuição de probabilidade:

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

e independentemente escolha Y de acordo com a distribuição de probabilidade:

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

Note que por

$$\|\mu(x) - \nu(x)\|_{TV} = \sum_{\substack{x \in \Omega \\ \mu(x) \geq \nu(x)}} [\mu(x) - \nu(x)] \Rightarrow \sum_{\substack{x \in \Omega \\ \mu(x) > \nu(x)}} \frac{[\mu(x) - \nu(x)]}{\|\mu(x) - \nu(x)\|_{TV}} = 1$$

podemos concluir sem muito esforço que γ_I e γ_{II} são distribuições de probabilidades. Ainda temos que:

$$p\gamma_{III} + (1-p)\gamma_I = \mu(x) \wedge \nu(x) + \mu(x) - \nu(x) = \nu(x) + \mu(x) - \nu(x) = \mu(x)$$

e

$$p\gamma_{III} + (1-p)\gamma_{II} = \mu(x) \wedge \nu(x) + \nu(x) - \mu(x) = \mu(x) + \nu(x) - \mu(x) = \nu(x).$$

De modo que a distribuição de X é μ e a distribuição de Y é ν . Observe que no caso da moeda dar coroa, $X \neq Y$, já que γ_I e γ_{II} são subconjuntos disjuntos. Assim, $X = Y$ se, e somente se o lançamento da moeda der cara. Portanto, chegamos ao seguinte resultado $P\{X \neq Y\} = \|\mu - \nu\|_{TV}$. \square

Agora estamos prontos para provar que cadeias de Markov irreduzíveis e aperiódicas convergem às suas distribuições estacionárias. Mas antes disso, vamos examinar algumas definições cruciais que serão úteis para restringir o tempo necessário para que duas distribuições atinjam uma distância específica determinada pela variação total. Nosso foco reside em limitar a distância máxima entre $P^t(x_0, \cdot)$ e π , independentemente do valor de x_0 pertencente a Ω .

Definição 1.27. *Sejam $x, y \in \Omega$ definimos:*

$$d(t) := \max_{x \in \Omega} \|P^t(x, \cdot) - \pi(\cdot)\|_{TV} \quad (1.18)$$

e

$$\bar{d}(t) := \max_{x, y \in \Omega} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \quad (1.19)$$

Proposição 1.28. *Seja $d(t)$ e $\bar{d}(t)$ definidas em (1.18) e (1.19), respectivamente, então:*

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

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

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

Para mostrar que $d(t) \leq \bar{d}(t)$, observe primeiro que, como π é estacionária, temos $\pi(A) = \sum_{y \in \Omega} \pi(y)P^t(y, A)$ para qualquer conjunto A (Essa é a definição de estacionariedade se A for um conjunto unitário $\{x\}$). Para obter isso para A arbitrário, basta somar sobre os elementos em A e ainda $P^t(x, A) = \sum_{y \in \Omega} \pi(y)P^t(x, A)$. Desta forma, temos:

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

onde a última desigualdade é proveniente do fato de que a média ponderada de um conjunto de números nunca é maior do que o seu elemento máximo. \square

Proposição 1.29. *Deixe \mathcal{P} denotar a coleção de todas as distribuições de probabilidade em Ω . E seja π uma distribuição estacionária para a cadeia de Markov P , então*

$$d(t) = \sup_{\mu \in \mathcal{P}} \|\mu P^t - \pi\|_{TV} \quad (1.20)$$

e

$$\bar{d}(t) = \sup_{\mu, \nu \in \mathcal{P}} \|\mu P^t - \nu P^t\|_{TV}. \quad (1.21)$$

Demonstração. 1º demonstramos (1.20). Tome de modo arbitrário uma distribuição de probabilidade $\mu \in \Omega$. Usando a proposição 1.21, temos:

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

tomando o supremo, temos:

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

Por outro lado, pela definição de supremo temos que:

$$d(t) = \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{TV} \leq \sup_{\mu \in \mathcal{P}} \|\mu P^t - \pi\|_{TV},$$

com isso demonstramos a igualdade.

Agora demonstramos 1.21. Tome arbitrariamente μ e ν , duas distribuições de probabili-

dade em Ω . Então temos:

$$\begin{aligned}
\|\mu P^t - \nu P^t\|_{TV} &= \frac{1}{2} \sum_{z \in \Omega} |\mu P^t(z) - \nu P^t(z)| \\
&= \frac{1}{2} \sum_{z \in \Omega} \left| \sum_{y \in \Omega} \nu(y) \mu P^t(z) - \sum_{y \in \Omega} \nu(y) P^t(y, z) \right| \\
&\leq \frac{1}{2} \sum_{z \in \Omega} \sum_{y \in \Omega} \nu(y) |\mu P^t(z) - P^t(y, z)| \\
&= \sum_{y \in \Omega} \nu(y) \frac{1}{2} \sum_{z \in \Omega} |\mu P^t(z) - P^t(y, z)| \\
&= \sum_{y \in \Omega} \nu(y) \|\mu P^t - P^t(y, \cdot)\|_{TV} \leq \max_{y \in \Omega} \|\mu P^t - P^t(y, \cdot)\|_{TV} \\
&= \max_{y \in \Omega} \frac{1}{2} \sum_{w \in \Omega} \|\mu P^t(w) - P^t(y, w)\|_{TV} \\
&= \max_{y \in \Omega} \frac{1}{2} \sum_{w \in \Omega} \left| \sum_{x \in \Omega} \mu(x) P^t(x, w) - \sum_{x \in \Omega} \mu(x) P^t(y, w) \right| \\
&\leq \max_{y \in \Omega} \frac{1}{2} \sum_{w \in \Omega} \sum_{x \in \Omega} \mu(x) |P^t(x, w) - P^t(y, w)| \\
&= \max_{y \in \Omega} \sum_{x \in \Omega} \mu(x) \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \\
&\leq \max_{y \in \Omega} \max_{x \in \Omega} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} = \bar{d}(t).
\end{aligned}$$

Então, tomando o supremo sobre μ e ν temos:

$$\sup_{\mu, \nu \in \mathcal{P}} \|\mu P^t - \nu P^t\|_{TV} \leq \bar{d}(t).$$

Como em 1.20, a demonstração de $\sup_{\mu, \nu \in \mathcal{P}} \|\mu P^t - \nu P^t\|_{TV} \geq \bar{d}(t)$ é óbvia. \square

Lema 1.30. *Seja P a matriz de transição de uma cadeia de Markov com espaço de estados em Ω e sejam μ e ν duas distribuições de probabilidade em Ω . Então, podemos afirmar que:*

$$\|\mu P - \nu P\|_{TV} \leq \|\mu - \nu\|_{TV}.$$

(Isso mostra, em particular, que $\|\mu P^{t+1} - \pi\|_{TV} \leq \|\mu P^t - \pi\|_{TV}$, ou seja, avançar a cadeia só pode movê-la para mais perto da estacionariedade)

Demonstração.

$$\begin{aligned}
\|\mu P - \nu P\|_{TV} &= \frac{1}{2} \sum_{w \in \Omega} |\mu P(w) - \nu P(w)| \\
&= \frac{1}{2} \sum_{w \in \Omega} \left| \sum_{z \in \Omega} \mu(z) P(z, w) - \sum_{z \in \Omega} \nu(z) P(z, w) \right| \\
&= \frac{1}{2} \sum_{w \in \Omega} \left| \sum_{z \in \Omega} P(z, w) (\mu(z) - \nu(z)) \right| \\
&\leq \frac{1}{2} \sum_{w \in \Omega} \sum_{z \in \Omega} P(z, w) |\mu(z) - \nu(z)| \\
&= \frac{1}{2} \sum_{z \in \Omega} \sum_{w \in \Omega} P(z, w) |\mu(z) - \nu(z)| \\
&= \frac{1}{2} \sum_{z \in \Omega} (|\mu(z) - \nu(z)| \sum_{w \in \Omega} P(z, w)) \\
&= \frac{1}{2} \sum_{z \in \Omega} |\mu(z) - \nu(z)| = \|\mu - \nu\|_{TV}.
\end{aligned}$$

□

Corolário 1.31. *Nós temos que d e \bar{d} são não crescentes, isto é:*

$$d(t+1) \leq d(t)$$

e

$$\bar{d}(t+1) \leq \bar{d}(t).$$

Demonstração. Considere P irredutível e aperiódica com distribuição estacionária $\pi \in \Omega$, isto nos diz que $\pi = P^t \pi$. Para a primeira desigualdade seja μ uma distribuição qualquer em Ω , temos que:

$$\|\mu P^{t+1} - \pi\|_{TV} = \|(\mu P^t)P - \pi P\|_{TV} \leq \|\mu P^t - \pi\|_{TV},$$

tomando o supremo sobre μ e usando 1.18, chegamos na desigualdade desejada. Para a segunda desigualdade, sejam μ e ν distribuições genéricas em Ω , temos que:

$$\|\mu P^{t+1} - \nu P^{t+1}\|_{TV} = \|(\mu P^t)P - (\nu P^t)P\|_{TV} \leq \|\mu P^t - \nu P^t\|_{TV},$$

tomando o supremo sobre μ e ν e, usando 1.19, chegamos na desigualdade desejada. □

Proposição 1.32. *A função \bar{d} é sub-multiplicativa: $\bar{d}(s+t) \leq \bar{d}(s)\bar{d}(t)$.*

Demonstração. Fixe $x, y \in \Omega$, e deixe (X_s, Y_s) ser o acoplamento ótimo de $P^s(x, \cdot)$ e $P^s(y, \cdot)$, cuja existência é garantida por (1.15). Portanto,

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

Dado que P^{s+t} é o produto matricial de P^t e P^s e a distribuição de X^s é $P^s(x, \cdot)$, temos

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

Combinando isso com a identidade similar $P^{s+t}(y, w) = \mathbb{E}(P^t(Y_s, w))$ permite-nos escrever

$$P^{s+t}(x, w) - P^{s+t}(y, w) = \mathbb{E}(P^t(X_s, w)) - \mathbb{E}(P^t(Y_s, w)) = \mathbb{E}(P^t(X_s, w) - P^t(Y_s, w)). \quad (1.22)$$

Combinar as esperança é possível, uma vez que X_s e Y_s são definidos juntos no mesmo espaço de probabilidade. Somando 1.22 sobre $w \in \Omega$ e aplicando a proposição 1.21, temos que:

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

Dessa forma, vemos que a quantidade dentro da esperança é $\|P^t(X_s, \cdot) - P^t(Y_s, \cdot)\|_{TV}$ e, é exatamente 0 sempre que $X_s = Y_s$. Além disso, essa distância é limitada por $\bar{d}(t)$. Isso mostra que,

$$\begin{aligned} \|P^{s+t}(x, \cdot) - P^{s+t}(y, \cdot)\|_{TV} &\leq \mathbb{E}(\mathbf{1}_{X_s \neq Y_s} \bar{d}(t)) \\ &= P\{X_s \neq Y_s\} \bar{d}(t) \\ &= \|P^s(x, \cdot) - P^s(y, \cdot)\|_{TV} \bar{d}(t). \end{aligned}$$

Finalmente, maximizando sobre x, y completamos a prova. □

Corolário 1.33. *Quando c e t são inteiros não negativos quaisquer, vale a desigualdade:*

$$d(ct) \leq \bar{d}(ct) \leq \bar{d}(t)^c. \quad (1.23)$$

Demonstração. É uma simples aplicação das proposições (1.28) e (1.32) □

1.5 Tempo de mistura

É útil introduzir um parâmetro que mede o tempo requerido por uma Cadeia de Markov para que a distância até a estacionaridade seja pequena.

Definição 1.34. *Seja $\varepsilon \in [0, 1]$ definimos o tempo de mistura (mixing time) por:*

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

e

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

Observe que a escolha de $\varepsilon \in [0, 1]$ é necessária pois $0 \leq d(t) \leq 1$. Mais ainda, vamos mostrar que a escolha de ε deve ser menor que $1/2$. De fato, usando a definição de tempo de mistura, (1.33) e (1.32) temos que:

$$d(lt_{mix}(\varepsilon)) \leq \bar{d}(lt_{mix}(\varepsilon)) \leq \bar{d}(t^{mix}(\varepsilon))^l \leq (2d(t_{mix}(\varepsilon)))^l \leq (2\varepsilon)^l,$$

daí

$$d(lt_{mix}(\varepsilon)) \leq (2\varepsilon)^l. \quad (1.24)$$

Se $1/2 \leq \varepsilon \leq 1$ a desigualdade (1.24) é satisfeita trivialmente. Para isso não ocorrer, basta escolher $\varepsilon \in [0, 1/2)$. Em particular ao escolhermos $\varepsilon = 1/4$ e substituirmos em (1.24) temos:

$$d(lt_{mix}) \leq (2)^{-l}. \quad (1.25)$$

Como l é inteiro não negativo, então podemos escolher, convenientemente, $l = \lceil \log_2 \varepsilon^{-1} \rceil > 0$. Substituindo esse valor em (1.25), temos:

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

Pela definição de tempo de mistura, segue que:

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

1.6 Teorema da Convergência

O Teorema da Convergência é amplamente reconhecido como um resultado altamente relevante na teoria de cadeias de Markov. Quando consideramos uma cadeia de

Markov com matriz de transição P , se P for irredutível, pela Proposição 1.19, então há uma única distribuição estacionária. Adicionando a hipótese da cadeia ser aperiódica, o Teorema da Convergência nos fornece uma estimativa do tempo necessário para que a distribuição da cadeia convirja para a distribuição estacionária.

Teorema 1.35. *(Teorema da Convergência) Suponha que P seja irredutível e aperiódica, com distribuição estacionária π . Então existe uma constante $\alpha \in (0, 1)$ e $C > 0$ tal que*

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

Vamos provar 1.35 de duas formas, onde a segunda forma será feita após aprofundarmos um pouco na teoria de acoplamento em Cadeias de Markov. A forma da prova apresentada aqui decompõe a cadeia em uma mistura de amostragem repetida e independente da distribuição estacionária e outra cadeia de Markov. Para isso demonstraremos antes dois lemas que serão uteis.

Lema 1.36. *Se M é uma matriz estocástica e Π uma matriz de $|\Omega|$ linhas, em que cada linha é o vetor estacionário π , então $M\Pi = \Pi$.*

Demonstração. Com efeito, sejam $x, y \in \Omega$, temos:

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

□

Lema 1.37. *Se $\pi M = \pi$, então $\Pi M = \Pi$.*

Demonstração. A demonstração é imediata, pois todas as linhas de Π são formadas pelo vetor π . □

Agora vamos à primeira demonstração do Teorema da Convergência 1.35

Demonstração. Seja P irredutível e aperiódica, pela proposição 1.9, existe um inteiro positivo r tal que P^r tem somente entradas positivas. E pela Proposição 1.19, $\pi(y) > 0$ para todo $y \in \Omega$, então defina Π a matriz com $|\Omega|$ linhas, em que cada linha é o vetor π . Para algum $\delta > 0$ suficientemente pequeno, temos que

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

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

$$P^r = (1 - \theta)\Pi + \theta Q \tag{1.26}$$

define uma matriz estocástica Q . De fato isolando Q em (1.26) temos:

$$Q = \frac{P^r - (1 - \theta)\Pi}{\theta},$$

daí temos que

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

e também

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

Para todo $x \in \Omega$. Agora provemos, por indução sobre k , que

$$P^{rk} = (1 - \theta^k)\Pi + \theta^k Q^k, \quad (1.27)$$

para $k \in \mathbb{N} \setminus \{0\}$.

Se $k = 1$, temos $P^r = (1 - \theta)\Pi + \theta Q$, o que é verdade por nossa hipótese (1.26). Assumindo que (1.27) é válido para $k = n$, segue que

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

Distribuindo e expandindo P^r no segundo membro e usando (1.26), temos

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

Usando que $\Pi P^r = \Pi$ e $Q^n \Pi = \Pi$, pelos lemas 1.37 e 1.36, temos que

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

Isso estabelece (1.27) para $k = n + 1$, portanto ela vale para todo $k \in \mathbb{N} \setminus \{0\}$. Multiplicando (1.27) por P^j :

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

Usando que $\Pi P^j = \Pi$, pelo lema 1.37, segue que

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

ou

$$P^{rk+j} - \Pi = \theta^k(Q^k P^j - \Pi). \quad (1.28)$$

Para completar a prova somamos os valores absolutos dos elementos na linha x_0 em ambos os lados da equação (1.28) e dividindo por 2. Portanto, para qualquer $x_0 \in \Omega$, temos:

$$\frac{1}{2} \sum_{x \in \Omega} |P^{rk+j}(x_0, x) - \Pi(x_0, x)| = \frac{\theta^k}{2} \sum_{x \in \Omega} |Q^k P^j(x_0, x) - \Pi(x_0, x)|,$$

usando a proposição 1.21 e lembrando que $\Pi(x_0, \cdot) = \pi$, segue que

$$\|P^{rk+j}(x_0, \cdot) - \pi(x_0, \cdot)\|_{TV} = \theta^k \|Q^k P^j(x_0, \cdot) - \pi(x_0, \cdot)\|_{TV} \leq \theta^k. \quad (1.29)$$

Como $\frac{\theta^k}{\theta^{rk+j}} > 0$, tome C como uma cota superior de $\frac{\theta^k}{\theta^{rk+j}}$. Desta forma, obtemos de (1.29)

$$\max_{x \in \Omega} \|P^{rk+j}(x_0, \cdot) - \pi\|_{TV} \leq \theta^k = C\theta^{rk+j}.$$

Em particular, dado $t \in \mathbb{N}$ com $t \geq r$, existem $j, k \in \mathbb{N}$, tal que $0 \leq j < r$ e $t = rk + j$. Além disso, tome $\alpha = \theta = 1 - \delta$, daí $\alpha \in [0, 1]$. Portanto, obtemos

$$\max_{x \in \Omega} \|P^t(x_0, \cdot) - \pi\|_{TV} \leq \theta^k = C\alpha^t.$$

□

1.7 Acoplamento de Cadeias de Markov

Já definimos em 1.23 os acoplamentos entre duas distribuições. E vimos que são úteis porque permitem que a comparação entre distribuições seja reduzida a uma comparação entre variáveis aleatórias. A Proposição 1.15 caracteriza $\|\mu - \nu\|_{TV}$ como o mínimo, entre todos os acoplamentos (X, Y) de μ e ν , da probabilidade de X e Y serem diferentes. Isso fornece um método eficaz para obter limites superiores para a distância. Agora extrairemos mais informações acoplando não apenas pares de distribuições, mas trajetórias inteiras de cadeias de Markov. Vamos analisar o seguinte exemplo para fixar melhor essa idéia.

Exemplo 1.38. *Um passeio aleatório simples no segmento $\{0, 1, \dots, n\}$ é uma cadeia de Markov que se move para cima ou para baixo a cada passo com igual probabilidade. Se o passeio tentar sair do intervalo quando estiver em um ponto de fronteira, ele permanece*

parado. É intuitivamente claro que $P^t(x, n) \leq P^t(y, n)$ sempre que $x \leq y$, pois isso significa que a probabilidade de estar no valor “topo” n após t etapas não diminui à medida que aumentamos a altura da posição inicial.

Uma prova simples utiliza um acoplamento das distribuições $P^t(x, \cdot)$ e $P^t(y, \cdot)$. Sejam $\Delta_1, \Delta_2, \dots$ uma sequência de variáveis aleatórias independente identicamente distribuídas com valores $\{-1, +1\}$ e média zero, de modo que cada Δ_i seja igualmente provável de ser $+1$ ou -1 . Vamos definir dois passeios aleatórios em $\{0, 1, \dots, n\}$: o passeio (X_t) começa em x , enquanto o passeio (Y_t) começa em y .

Usaremos a mesma regra para os movimentos em ambas as cadeias (X_t) e (Y_t) : se $\Delta_t = +1$, mova a cadeia para cima, se possível, e se $\Delta_t = -1$, mova a cadeia para baixo, se possível. Portanto, as cadeias se movem em sincronia, embora comecem em alturas diferentes. Contudo, uma vez que as duas cadeias se encontram (necessariamente em 0 ou n), elas permanecem juntas a partir desse momento. Claramente, a distribuição de X_t é $P^t(x, \cdot)$, e a distribuição de Y_t é $P^t(y, \cdot)$. É importante ressaltar que X_t e Y_t são definidos no mesmo espaço de probabilidade subjacente, pois ambas usam a sequência (Δ_t) para determinar seus movimentos.

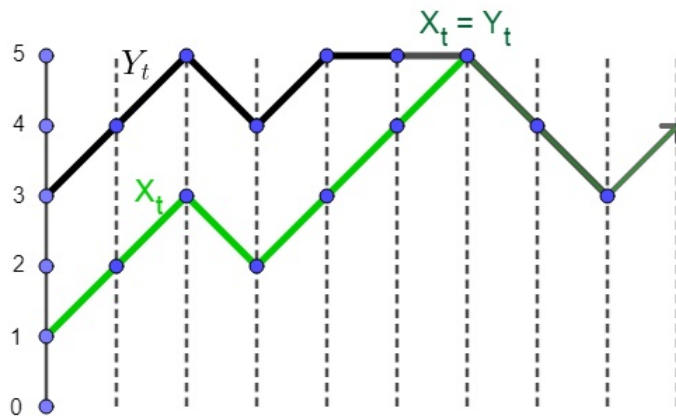


Figura 1.3: Acoplamento dos passeios aleatórios X_t e Y_t em $\{0, 1, 2, 3, 4, 5\}$. As Cadeias permanecem juntas depois de se encontrarem.

Então, se $x \leq y$, então $X_t \leq Y_t$ para todo t . Em particular, se $X_t = n$, o estado “topo”, então devemos ter que $Y_t = n$ também. A partir disso, podemos concluir que

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

Vamos apresentar agora a definição de acoplamento de cadeias de Markov e usaremos esta técnica para encontrar limitantes para o tempo de mistura.

Definição 1.39. Um acoplamento de cadeias de Markov com matriz de transição P é um processo $(X_t, Y_t)_{t=0}^{\infty}$ com a propriedade de que tanto (X_t) quanto (Y_t) são cadeias

de Markov com matriz de transição P , embora as duas cadeias possam ter distribuições iniciais diferentes.

Qualquer acoplamento de cadeias de Markov com matriz de transição P pode ser modificado para que as duas cadeias permaneçam juntas o tempo todo após sua primeira visita simultânea a um único estado - mais precisamente, de forma que:

$$\text{se } X_s = Y_s, \text{ então } X_t = Y_t, \text{ para } t \geq s. \quad (1.30)$$

Observação: Para construir um acoplamento que satisfaça a condição (1.30), basta executar as cadeias de acordo com o acoplamento original até que se encontrem; em seguida, execute-as juntas.

Notação: Se (X_t) e (Y_t) são cadeias de Markov acopladas com $X_0 = x$ e $Y_0 = y$, escrevemos $P_{x,y}$ para a probabilidade no espaço em que (X_t) e (Y_t) estão definidos.

Teorema 1.40. *Seja P uma matriz de transição irredutível, com espaço de estados Ω , e seja π sua distribuição estacionária. Seja $\{(X_t, Y_t)\}$ um acoplamento satisfazendo (1.30) onde $X_0 = x$ e $Y_0 = y$. Seja τ_{couple} o primeiro tempo que as cadeias se encontram:*

$$\tau_{couple} = \min\{t : X_t = Y_t\}.$$

Então

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

Demonstração. Note que $P^t(x, z) = P_{x,y}\{X_t = z\}$ e $P^t(y, z) = P_{x,y}\{Y_t = z\}$. Consequentemente, (X_t, Y_t) é um acoplamento de $P^t(x, \cdot)$ com $P^t(y, \cdot)$, então pela Proposição 1.15 temos que:

$$\|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \leq P_{x,y}\{X_t \neq Y_t\}. \quad (1.31)$$

Como $X_t \neq Y_t$ se, e somente se $\tau_{couple} > t$, e por (1.30), $P_{x,y}\{X_t \neq Y_t\} = P_{x,y}\{\tau_{couple} > t\}$, logo de (1.31) obtemos

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

□

Corolário 1.41. *Suponha que para cada par de estados $x, y \in \Omega$, exista um acoplamento (X_t, Y_t) com $X_0 = x$ e $Y_0 = y$. Para cada acoplamento desses, seja τ_{couple} o primeiro momento em que as cadeias se encontram. Então,*

$$d(t) \leq \max_{x,y \in \Omega} P_{x,y}\{\tau_{couple} > t\}.$$

Demonstração. Basta usar a primeira desigualdade do Lema 1.28, e o Teorema 1.40, lembrando que $\bar{d}(t) = \max_{x,y \in \Omega} \|P^t(x, \cdot) - P^t(y, \cdot)\|$, obtemos

$$d(t) \leq \bar{d}(t) = \max_{x,y \in \Omega} \|P^t(x, \cdot) - P^t(y, \cdot)\| \leq \max_{x,y \in \Omega} P_{x,y} \{\tau_{couple} > t\}$$

a desigualdade desejada. \square

Este resultado é crucial para estimar o tempo de mistura. Ele indica que, ao criar um acoplamento que satisfaça a condição (1.30), o Corolário 1.41 nos assegura que a distância total variacional entre a distribuição da cadeia após t passos e a distribuição estacionária é, no máximo, a probabilidade de as cadeias não se encontrarem antes do instante t . Essa probabilidade é avaliada maximizando sobre todos os possíveis estados iniciais. Em outras palavras, a medida em que as cadeias estão distantes (de acordo com a distância total variacional) após t passos é, no máximo, a probabilidade de que elas ainda não tenham se encontrado até esse momento, considerando o pior caso entre todos os pares de estados iniciais possíveis. Essa relação é útil para entender como a distância entre as cadeias de Markov está relacionada à probabilidade de encontro entre elas.

Demonstração do teorema da convergência 1.35, agora usando a técnica de acoplamento de cadeias de Markov.

Demonstração. Considere X_t e Y_t duas cadeias de Markov com matriz de transição P , irreduzível e aperiódica, e com distribuições μ e ν , respectivamente. Considere (X_t, Y_t) um acoplamento que satisfaça (1.30). Daí $P^t(x, \cdot) = \mu P^t$ e $P^t(y, \cdot) = \nu P^t$. Então pelo Teorema 1.40, e tomando $\mu = \delta_x$ e $\nu = \pi$, onde π é a distribuição estacionária, temos:

$$\|P^t - \pi\|_{VT} \leq P_x(\tau_{couple} > t). \quad (1.32)$$

A idéia aqui é mostrar que $P(\tau_{couple} < \infty) = 1$ e que $P(\tau_{couple} > t) = 0$ converge rapidamente para 0. Para isso, como P é aperiódica e irreduzível, pela Proposição 1.9 temos que existe r tal que $\varepsilon = \min_{x,y \in \Omega} P^r(x, y) > 0$. Fixando um $x_0 \in \Omega$ tome $J_r = \{X_r \neq x_0, Y_r \neq x_0\}$, então temos que:

$$P(J_r) \leq 1 - \varepsilon.$$

De fato, podemos estimar

$$P(J_r) \leq P(X_r \neq x_0) = 1 - P(X_r = x_0) \leq 1 - \varepsilon. \quad (1.33)$$

Após r passos temos:

$$\begin{aligned} P(J_{2r}|J_r) &\leq (1 - \varepsilon) \\ P(J_{2r}) &\leq (1 - \varepsilon)^2 \end{aligned}$$

Procedendo indutivamente, temos para k inteiro:

$$P(\tau_{couple} > kr) = P(J_{kr}) \leq (1 - \varepsilon)^k.$$

Portanto quando $k \rightarrow \infty$, $P(\tau_{couple} > kr) \rightarrow 0$. Ou ainda $P(\tau_{couple} < \infty) = 1$. Pela divisão euclidiana escrevemos $t = kr + n$, com $n < r$ e $k, b \in \mathbb{Z}_+$. Portanto, temos:

$$\begin{aligned} P(\tau_{couple} > t) &= P(\tau_{couple} > kr + b) \\ &\leq P(\tau_{couple} > kr, \tau_{couple} > b) \\ &= P(\tau_{couple} > kr) \\ &= (1 - \varepsilon)^k. \end{aligned}$$

Substituindo k por $\frac{t-n}{r}$, segue que

$$\begin{aligned} P(\tau_{couple} > t) &\leq (1 - \varepsilon)^{\frac{t-n}{r}} \\ &\leq (1 - \varepsilon)^{\frac{t}{r}} (1 - \varepsilon)^{\frac{-n}{r}}. \end{aligned}$$

Tome $\alpha = (1 - \varepsilon)^{\frac{1}{r}} \in (0, 1)$. Como $n < r$, $(1 - \varepsilon)^{-1}$ é uma cota superior para $(1 - \varepsilon)^{\frac{-n}{r}}$. Portanto, escolhamos $C = (1 - \varepsilon)^{-1}$. Pela desigualdade (1.32), o resultado segue

$$\|P^t - \pi\|_{VT} \leq C\alpha^t.$$

□

Definição 1.42. Tempo de Parada (Stopping Time). Dada uma sequência $(X_t)_{t=0}^{\infty}$ de variáveis aleatórias em Ω , uma variável aleatória τ com valores em $\{0, 1, 2, 3, \dots, \infty\}$ é um tempo de parada para (X_t) se, para cada $t \in \{0, 1, 2, \dots\}$, existe um conjunto $B_t \subset \Omega^{t+1}$ tal que

$$\{\tau = t\} = \{(X_0, X_1, \dots, X_t) \in B_t\}.$$

Em outras palavras, um tempo aleatório τ é um tempo de parada se, e somente se, a função indicadora $\mathbf{1}_{\tau=t}$ for uma função do vetor (X_0, X_1, \dots, X_t) .

Exemplo 1.43. Tempo de Chegada (Hitting Time). Fixe $A \subseteq \Omega$. Definimos

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

ou seja, τ_A é a primeira vez que a sequência (X_t) está em A . Equivalentemente,

$$\{\tau_A = t\} = \{X_0 \notin A, X_1 \notin A, \dots, X_{t-1} \notin A, X_t \in A\}.$$

Portanto, τ_A é um tempo de parada.

Definição 1.44. Tempo Estacionário Forte. Um tempo estacionário forte para uma cadeia de Markov (X_t) com distribuição estacionária π é um tempo de parada aleatório τ , possivelmente dependendo da posição inicial x , tal que

$$P_x\{X_\tau = y | \tau = t\} = \pi(y). \quad (1.34)$$

Em outras palavras, para um dado τ , a variável aleatória X_τ tem uma distribuição π e é independente de τ sob a medida de probabilidade usual P_x . A equação (1.34) pode ser reescrita como:

$$P_x\{X_\tau = y, \tau = t\} = \pi(y)P_x\{\tau = t\}.$$

Lema 1.45. Seja (X_t) uma cadeia de Markov irredutível com distribuição estacionária π . Se τ é um tempo estacionário forte para (X_t) , então para todo $t \geq 0$,

$$P_x\{\tau \leq t, X_t = y\} = P\{\tau \leq t\}\pi(y). \quad (1.35)$$

Demonstração. Seja Z_1, Z_2, \dots seja a sequência *i.i.d.* usada no representante de mapeamento aleatório de (X_t) . Para qualquer $s \leq t$,

$$P_x\{\tau = s, X_t = y\} = \sum_{z \in \Omega} P_x\{X_t = y | \tau = s, X_s = z\} P_x\{\tau = s, X_s = z\}. \quad (1.36)$$

Como τ é um tempo de parada para (Z_t) , o evento $\{\tau = s\}$ é equivalente a $\{(Z_1, \dots, Z_s) \in B\}$ para algum $B \subset \Omega^s$. Também, para os inteiros $r, s \geq 0$, existe uma função $\tilde{f}_r : \Omega^{r+1} \rightarrow \Omega$, tal que

$$X_{r+s} = \tilde{f}_r(X_s, Z_{s+1}, \dots, Z_{s+r}).$$

Como os vetores aleatórios (Z_1, \dots, Z_s) e $(Z_{s+1}, \dots, Z_{s+r})$ são independentes,

$$\begin{aligned} P_x\{X_t = y | \tau = s, X_s = z\} &= P_x\{\tilde{f}_{t-s}(z, Z_{s+1}, \dots, Z_t) = y | (X_1, \dots, X_s \in B, X_s = z)\} \\ &= P^{t-s}(z, y). \end{aligned}$$

Portanto, usando a definição 1.44 juntamente com a igualdade acima, (1.36) pode ser reescrita como

$$P_x\{\tau = s, X_t = y\} = \sum_{z \in \Omega} P^{t-s}(z, y)\pi(z)P_x\{\tau = s\} = \left(\sum_{z \in \Omega} P^{t-s}(z, y)\pi(z) \right) P_x\{\tau = s\}. \quad (1.37)$$

Por hipótese, π é distribuição estacionária da cadeia, então satisfaz $\pi = \pi P^{t-s}$ ou de forma equivalente $\pi(y) = \sum_{z \in \Omega} \pi(z)P^{t-s}(z, y)$, então a igualdade (1.37) fica assim:

$$P_x\{\tau = s, X_t = y\} = \pi(y)P_x\{\tau = s\}.$$

Finalmente temos:

$$P_x\{\tau \leq t, X_t = y\} = \sum_{s \leq t} P_x\{\tau = s, X_t = y\} = \sum_{s \leq t} \pi(y) P_x\{\tau = s\} = \pi(y) P_x\{\tau \leq t\}.$$

□

A proposição que conecta o tempo estacionário forte com $d(t)$ é a seguinte:

Proposição 1.46. *Se τ é um tempo estacionário forte, então*

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

Entretanto, antes vamos provar dois lemas que nos auxiliarão na prova da Proposição 1.46. Vamos introduzir um parâmetro $s_x(t)$, chamado **distância de separação** e definida por:

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

Também definimos

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

Provemos a seguinte relação entre $s_x(t)$ e tempos estacionários fortes:

Lema 1.47. *Se τ é um tempo estacionário forte, então*

$$s_x(t) \leq P_x\{\tau > t\},$$

e portanto $s(t) \leq \max_{x \in \Omega} P_x\{\tau > t\}$.

Demonstração. Fixe $x \in \Omega$. Observe que para qualquer $y \in \Omega$, temos

$$1 - \frac{P^t(x, y)}{\pi(y)} = 1 - \frac{P_x\{X_t = y\}}{\pi(y)} \leq 1 - \frac{P_x\{X_t = y, \tau \leq t\}}{\pi(y)}$$

usando o Lema 1.45, temos

$$1 - \frac{P^t(x, y)}{\pi(y)} \leq 1 - \frac{\pi(y) P_x\{\tau \leq t\}}{\pi(y)} = 1 - P_x\{\tau \leq t\} = P_x\{\tau > t\}$$

que garante o resultado. □

Lema 1.48. *A distância de separação $s_x(t)$ satisfaz*

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

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

Demonstração. Usando resultado equivalente a (1.13), temos que:

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

□

Portanto a Proposição 1.46 segue do resultado dos dois lemas anteriores.

Capítulo 2

Embaralhamento Top to random

2.1 O modelo

Considere o seguinte método (lento) de embaralhar um baralho com n cartas: pegue a carta do topo e insira-a uniformemente e aleatoriamente no baralho. Este processo eventualmente misturará o baralho - as disposições sucessivas do baralho formam uma caminhada aleatória no grupo S_n das $n!$ permutações possíveis das cartas, o que, de acordo com a Proposição 1.16, possui uma distribuição estacionária uniforme. Quanto tempo devemos embaralhar usando esse método até que a disposição do baralho esteja próxima do aleatório? Seja τ_{top} o momento um passo após a primeira ocasião em que a carta original de baixo chega ao topo do baralho. Mostramos agora que a disposição das cartas no momento τ_{top} é distribuída uniformemente no conjunto S_n de todas as permutações de $\{1, \dots, n\}$ e, além disso, esse elemento aleatório de S_n é independente do tempo τ_{top} .

Proposição 2.1. *Seja (X_t) a caminhada aleatória em S_n correspondente ao embaralhamento Top-to-random em n cartas. Dado no tempo t que existem k cartas abaixo da carta original de baixo, cada uma das $k!$ ordenações possíveis dessas cartas tem igual probabilidade. Portanto, se τ_{top} é um embaralhamento após a primeira vez que a carta original de baixo vai para o topo do baralho, então a distribuição de $X_{\tau_{top}}$ é uniforme sobre S_n , e o tempo τ_{top} é independente de $X_{\tau_{top}}$.*

Demonstração. Provemos por indução e chame de J a carta originalmente de baixo. Quando $t = 0$, não há cartas abaixo de J , e a afirmação é trivialmente válida. Agora suponha que a afirmação seja verdadeira no tempo t . Existem duas possibilidades no tempo $t + 1$: ou uma carta é colocada abaixo de J , ou não. No segundo caso, as cartas abaixo de J permanecem em ordem aleatória. No primeiro caso, dado que a carta é colocada

abaixo de J , cada uma das $k + 1$ possíveis posições para a carta tem igual probabilidade, e assim cada uma das $(k + 1)!$ ordenações é equiprovável.

Temos ainda que, no tempo $t = \tau_{top} - 1$ as cartas que estão abaixo de J estão distribuídas uniformemente. Portanto, a ordem das cartas em τ_{top} é uniforme e τ_{top} é independente de $X_{\tau_{top}}$ \square

Se pararmos de embaralhar precisamente um embaralhamento após a carta original de baixo subir para o topo do baralho pela primeira vez, então a ordem das cartas neste momento é exatamente uniforme sobre todas as disposições possíveis. Ou seja, $X_{\tau_{top}}$ tem exatamente a distribuição estacionária da cadeia. Agora seja A o conjunto de disposições em que a carta J está no topo. Então, $\tau_A = \tau_{top} + 1$ é um tempo de parada.

2.2 Coleta de cupons

Vamos introduzir agora um exemplo que será útil para analisar melhor o embaralhamento Top-to-random.

Exemplo 2.2. Coleta de Cupons. *Uma empresa emite n tipos diferentes de cupons. Um colecionador deseja um conjunto completo. Supomos que cada cupom que ele adquire tem igual probabilidade de ser de cada um dos n tipos. Quantos cupons ele deve obter para que sua coleção contenha todos os n tipos? Pode não ser óbvio por que isso é uma cadeia de Markov. Seja X_t o número de tipos diferentes representados entre os primeiros t cupons do colecionador. Claramente, $X_0 = 0$. Quando o colecionador possui cupons de k tipos diferentes, faltam $n - k$ tipos. Das n possibilidades para seu próximo cupom, apenas $n - k$ irão expandir sua coleção. Portanto*

$$P\{X_{t+1} = k + 1 | X_t = k\} = \frac{n - k}{n}$$

e

$$P\{X_{t+1} = k | X_t = k\} = \frac{k}{n}.$$

Cada trajetória dessa cadeia é não decrescente. Uma vez que a cadeia chega ao estado n (correspondente a uma coleção completa), ela é absorvida lá. Estamos interessados no número de passos necessários para alcançar o estado absorvente.

Proposição 2.3. *Considere um colecionador tentando completar um conjunto completo de cupons. Suponha que cada novo cupom seja escolhido de forma uniforme e independente do conjunto de n tipos possíveis, e seja τ o número (aleatório) de cupons coletados quando*

o conjunto contém pela primeira vez todos os tipos. Então,

$$E(\tau) = n \sum_{k=1}^n \frac{1}{k}.$$

Demonstração. A esperança $E(\tau)$ pode ser calculada escrevendo τ como uma soma de variáveis aleatórias geométricas. Seja τ_k o número total de cupons acumulados quando a coleção contém pela primeira vez k cupons distintos. Então,

$$\tau = \tau_n = \tau_1 + (\tau_2 - \tau_1) + \cdots + (\tau_n - \tau_{n-1}).$$

Além disso, $\tau_k - \tau_{k-1}$ é uma variável aleatória geométrica com probabilidade de sucesso $(n - k + 1)/n$, após coletar τ_{k-1} cupons, faltam $n - k + 1$ tipos na coleção. Cada cupom subsequente sorteado tem a mesma probabilidade $(n - k + 1)/n$ de ser um tipo ainda não coletado, até que um novo tipo seja finalmente sorteado. Assim, $E(\tau_k - \tau_{k-1}) = n/(n - k + 1)$ e

$$E(\tau) = \sum_{k=1}^n E(\tau_k - \tau_{k-1}) = n \sum_{k=1}^n \frac{1}{n - k + 1} = n \sum_{k=1}^n \frac{1}{k}.$$

□

A seguinte proposição nos afirma que é improvável que τ seja muito maior que seu valor esperado.

Proposição 2.4. *Seja τ uma variável aleatória de coleta de cupons, como na Proposição 2.3. Para qualquer $c > 0$,*

$$P\{\tau > \lceil n \log n + cn \rceil\} \leq e^{-c}$$

Demonstração. Seja A_i o evento de o i -ésimo tipo não aparece entre os primeiros $\lceil n \log n + cn \rceil$ cupons sorteados. Observe primeiro que

$$P\{\tau > \lceil n \log n + cn \rceil\} = P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i).$$

Uma vez que cada tentativa tem probabilidade $1 - n^{-1}$ de não sortear o cupom i e as tentativas são independentes, o lado direito acima está limitado superiormente por:

$$\sum_{i=1}^n \left(1 - \frac{1}{n}\right)^{\lceil n \log n + cn \rceil} = n \left(1 - \frac{1}{n}\right)^{\lceil n \log n + cn \rceil} \quad (2.1)$$

$$= n \exp \log \left(1 - \frac{1}{n}\right)^{\lceil n \log n + cn \rceil} \quad (2.2)$$

$$= n \exp \lceil n \log n + cn \rceil \log \left(1 - \frac{1}{n}\right). \quad (2.3)$$

Agora, usando uma propriedade conhecida de logaritmos que diz que: $\log(1+x) \leq x$, sempre que $|x| < 1$, temos que $\log\left(1 - \frac{1}{n}\right) \leq -\frac{1}{n}$. Além disso, como $\lceil n \log n + cn \rceil \geq n \log n + cn$, então

$$\lceil n \log n + cn \rceil \log\left(1 - \frac{1}{n}\right) \leq \left(-\frac{n \log n + cn}{n}\right) = -\log n - c. \quad (2.4)$$

Então combinando (2.3) e (2.4), temos que:

$$\sum_{i=1}^n \left(1 - \frac{1}{n}\right)^{\lceil n \log n + cn \rceil} \leq n \exp(-\log n - c) = e^{-c}.$$

□

Voltando agora ao embaralhamento top-to-random, Considere o movimento da carta original de baixo. Quando há k cartas abaixo dela, a chance de que ela suba uma carta permanece k/n até que um embaralhamento coloque a carta do topo embaixo dela. Assim, a distribuição de τ_{top} é a mesma que o tempo do coletor de cupons.

Proposição 2.5. *Para o Embaralhamento Top-to-random, $d(n \log n + \alpha n) \leq e^{-\alpha}$, para todo n .*

Demonstração. Veja, como τ_{top} é um tempo estacionário forte e pela Proposição 2.4 temos que para todo $\alpha > 0$ temos:

$$P\{\tau_{top} > \lceil n \log n + \alpha n \rceil\} \leq e^{-\alpha},$$

e pela Proposição 1.46, temos o resultado desejado. □

Corolário 2.6. $t_{mix}(\varepsilon) \leq n \log n + \log(\varepsilon^{-1})n$, para todo n .

Demonstração. Basta tomar $\alpha = \log \varepsilon^{-1}$ na Proposição 2.5. □

A seguinte proposição traz uma cota inferior para o tempo de mistura.

Proposição 2.7. *Seja (X_t) a cadeia Top-to-Random em n cartas. Para qualquer $\varepsilon > 0$, existe uma constante α_0 tal que $\alpha > \alpha_0$ implica que para n suficientemente grande,*

$$d_n(n \log n - \alpha n) \geq 1 - \varepsilon.$$

Em particular, existe uma constante α_1 tal que para n suficientemente grande,

$$t_{mix} \geq n \log n - \alpha_1 n.$$

Demonstração. Defina o evento A_j como

$$A_j = \{\text{As } j \text{ cartas originais da parte inferior estão em sua ordem relativa original}\}.$$

Deixe id denotar a permutação identidade. Nós iremos cotar $\|P^t(id, \cdot) - \pi\|_{TV}$ por baixo. Seja τ_j o tempo necessário para que a carta inicialmente na posição j da parte inferior, a contar do fundo, chegue ao topo. E seja $\tau_{j,i}$ o tempo que a carta inicialmente na posição j , a contar do fundo, leva subir da posição i para a posição $i + 1$. Então

$$\tau_j = \sum_{i=j}^{n-1} \tau_{j,i}.$$

As variáveis $\{\tau_{j,i}\}_{i=j}^{n-1}$ são independentes e $\tau_{j,i}$ tem distribuição geométrica com parâmetro $p = \frac{i}{n}$. Portanto, $E(\tau_{j,i}) = \frac{n}{i}$ e $Var(\tau_{j,i}) = \frac{1 - \frac{i}{n}}{\frac{i^2}{n^2}} = \frac{n-i}{n} \cdot \frac{n^2}{i^2} < \frac{n^2}{i^2}$. Por outro lado, sabemos que $\log(1+x) \leq x$, para $x \in [0, 1]$. Daí $\sum_{i=1}^n 1/i \geq \log 2 + \log(3/2) + \dots + \log(n + 1/n) = \log(n+1) > \log n$. Por isso,

$$E(\tau_j) = \sum_{i=j}^{n-1} \frac{n}{i} = \sum_{i=j}^n \frac{n}{i} - 1 \geq n \left(\sum_{i=1}^n \frac{1}{i} - \sum_{i=1}^j \frac{1}{i} \right) - n \geq n(\log n - \log j - 1)$$

e

$$Var(\tau_j) \leq n^2 \sum_{i=j}^{\infty} \frac{1}{i(i-1)} = n^2 \sum_{i=j}^{\infty} \left(\frac{1}{i-1} - \frac{1}{i} \right) \leq \frac{n^2}{j-1}.$$

Suponha que $\alpha \geq \log j + 2$. Então,

$$\begin{aligned} P\{\tau_j < n \log n - \alpha n\} &\leq P\{\tau_j - E(\tau_j) < n \log n - \alpha n - n(\log n - \log j - 1)\} \\ &= P\{\tau_j - E(\tau_j) < -\alpha n + n \log j + n\} \\ &= P\{\tau_j - E(\tau_j) < -n(\alpha - \log j - 1)\} \\ &\leq P\{|\tau_j - E(\tau_j)| > n(\alpha - \log j - 1)\}. \end{aligned}$$

Usando a desigualdade de Chebyshev nesta última desigualdade temos

$$\begin{aligned} P\{\tau_j < n \log n - \alpha n\} &\leq \frac{Var \tau_j}{(n(\alpha - \log j - 1))^2} \\ &\leq \frac{\frac{n^2}{j-1}}{n^2(\alpha - \log j - 1)^2} \\ &= \frac{1}{(j-1)(\alpha - \log j - 1)^2} \\ &\leq \frac{1}{j-1}. \end{aligned}$$

Defina $t_n(\alpha) = n \log n - \alpha n$. Se $\tau_j \geq t_n$, então as j cartas originalmente contadas do fundo do baralho estão em sua ordem relativa original no tempo $t_n(\alpha)$, então para $\alpha > \log j + 2$,

$$P^{t_n(\alpha)}(id, A_j) \geq P\{\tau_j \geq t_n(\alpha)\} \geq 1 - \frac{1}{j-1}.$$

Por outro lado, para a distribuição estacionária uniforme,

$$\pi(A_j) = \frac{1}{j!} \leq \frac{1}{j-1}.$$

Portanto, para $\alpha \geq \log j + 2$,

$$\begin{aligned} d(t_n(\alpha)) &\geq \|P^{t_n(\alpha)}(id, \cdot) - \pi\|_{TV} \\ &\geq P^{t_n(\alpha)}(id, A_j) - \pi(A_j) \\ &\geq 1 - \frac{2}{j-1}. \end{aligned}$$

Suponha que $n \geq e^{\alpha-2}$. Então, defina $j = \lceil e^{\alpha-2} \rceil$. Defina também

$$g(\alpha) = 1 - \frac{2}{\lceil e^{\alpha-2} \rceil - 1}.$$

Como $d(t_n(\alpha)) > g(\alpha)$,

$$\liminf_{n \rightarrow \infty} d(t_n(\alpha)) \geq g(\alpha),$$

e quando $\alpha \rightarrow \infty$, $g(\alpha) \rightarrow 1$. □

Capítulo 3

Caminho de rede e Embaralhamento de cartas por transposições adjacentes

Este capítulo tem como objetivo estudar o tempo de mistura em dois exemplos de cadeias de Markov: o embaralhamento de cartas por transposições adjacentes e o modelo de caminho de rede. O estudo desses exemplos nos permitirá compreender como as propriedades das cadeias de Markov podem ser aplicadas em contextos práticos, como no embaralhamento de cartas, e como o modelo do caminho de rede pode ser uma ferramenta útil para analisar o tempo de mistura nesses casos.

O embaralhamento de cartas por transposições adjacentes é um problema clássico em teoria das permutações e tem aplicações em algoritmos de ordenação e criptografia. Ao estudarmos o tempo de mistura neste contexto, buscamos entender quantas operações de embaralhamento são necessárias para que as cartas estejam suficientemente embaralhadas.

Por sua vez, o modelo de caminho de rede é uma abstração comum em diversos campos, como ciência da computação e física estatística. Veremos que este modelo pode ser utilizado como uma ferramenta analítica poderosa para encontrar o tempo de mistura em processos de embaralhamento, proporcionando insights valiosos sobre a dinâmica desses sistemas.

Ao combinar a análise do embaralhamento de cartas com a abordagem do modelo de caminho de rede, esperamos obter uma compreensão mais profunda do tempo de mistura em cadeias de Markov, contribuindo assim para a teoria e aplicação desses modelos em contextos práticos. Esse capítulo tem como referência o artigo de Wilson [2].

3.1 Cadeia de Markov no Caminho de rede

Considere um retângulo $a \times b$ composto por caixas 1×1 giradas 45° , de modo que os lados de comprimento a estejam orientados noroeste/sudeste. Um Caminho de rede (veja Figura (3.2)) é uma travessia do canto mais à esquerda até o canto mais à direita do retângulo, viajando ao longo das bordas das caixas 1×1 , de modo que cada movimento seja para cima e para a direita (nordeste) ou para baixo e para a direita (sudeste). Esses caminhos de rede podem ser codificados como strings (sequência de caracteres) de comprimento $a + b$ consistindo em a 0's (movimentos para baixo) e b 1's (movimentos para cima). Existem $n!/(a!b!)$ tais caminhos de rede, onde por conveniência deixamos $n = a + b$.

Considere a seguinte cadeia de Markov para gerar aleatoriamente um caminho de rede entre os cantos opostos do retângulo $a \times b$. Dado um caminho, a cadeia de Markov escolhe aleatoriamente uma das $(n - 1)$ colunas internas (assumimos $n \geq 2$) e então decide aleatoriamente se deve tentar empurrar o caminho para cima naquele ponto ou tentar empurrá-lo para baixo. Se empurrar o caminho para cima (ou para baixo) resultar em um caminho inválido, a cadeia de Markov simplesmente fica ociosa durante essa etapa. Veja Figura (3.1) para melhor entendimento.

Deixando um pouco mais claro: Cada passo da cadeia de Markov é uma transformação de sua trajetória, como vimos nem toda transformação é válida. Em outras palavras, cada caminho de rede é um estado dessa cadeia, e a matriz que representa essa cadeia tem ordem $\frac{n!}{a!b!}$. Usando como exemplo a cadeia de Markov definida no retângulo 4×5 temos $\frac{9!}{4!5!} = 126$ trajetórias, logo teríamos que ter uma matriz de ordem 126. Contudo, dado um caminho qualquer no retângulo $a \times b$, é sempre possível, após um certo número de passos, chegar a qualquer outro caminho no retângulo. Assim, a nossa cadeia de Markov no Caminho de rede é irredutível, e, pela Proposição 1.19, existe uma (e única) distribuição estacionária para essa cadeia.

3.1.1 Propriedade de contração

Estabeleceremos uma cota inferior para o tempo de mistura calculando o deslocamento de um caminho quando não foram dados passos suficientes e mostramos que isso geralmente difere do deslocamento de um caminho aleatório escolhido uniformemente ao acaso. E também estabeleceremos uma cota superior para o tempo de mistura mostrando que, após um número suficiente de passos, partindo dos caminhos superior e inferior, o

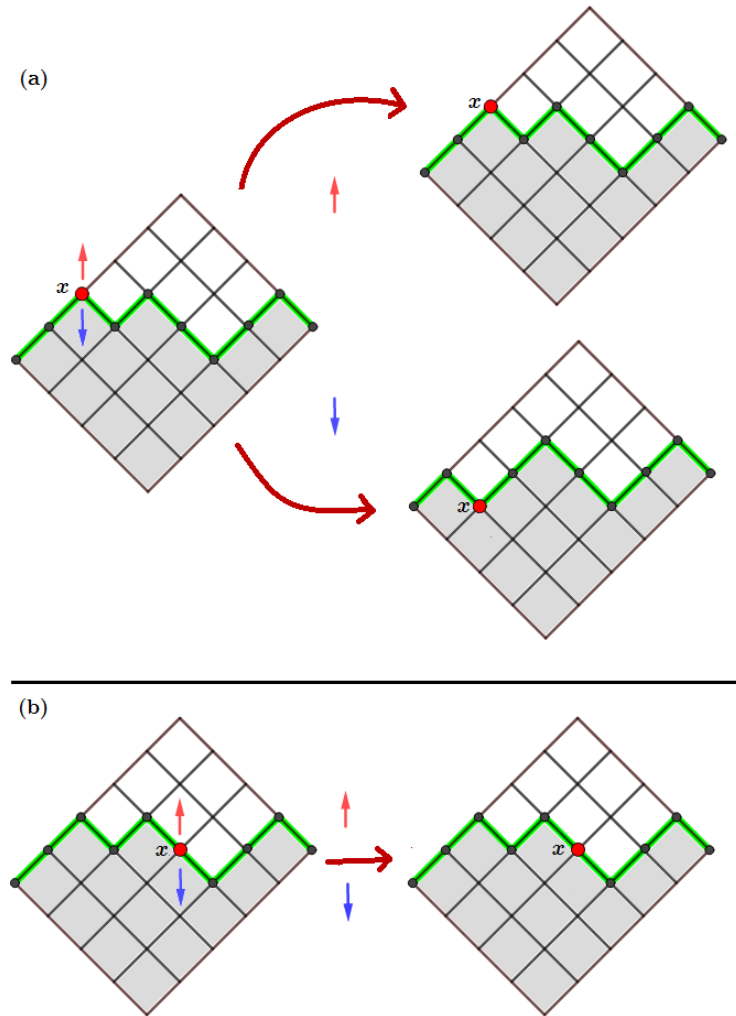


Figura 3.1: Escolhido o ponto x temos duas opções: subir ou descer. Em (a) o caminho de subida não é válido, então permanece no mesmo lugar. Já o movimento de descida é válido. Em (b) ambos os movimentos são proibidos, logo a trajetória permanece a mesma.

espaço esperado entre eles é tão pequeno que os caminhos quase com certeza se fundiram no mesmo caminho.

Será útil mais tarde usar coordenadas horizontais que variam de $-n/2$ a $n/2$. Deixe $h(x)$ denotar a altura do caminho na posição x em relação à linha que conecta os cantos opostos da caixa. Ou seja, $h(x)$ é o número de movimentos para cima à esquerda da posição x menos o número esperado de tais movimentos para cima. Assim, $h(-n/2) = h(n/2) = 0$ e $h(x) = h(x-1) + a/n$ se houve um movimento para cima entre $x-1$ e x , ou então $h(x) = h(x-1) - b/n$ se houve um movimento para baixo. Para o exemplo na Figura (3.2), as alturas mudam de $+4/9 = a/(a+b)$ para movimentos para cima e $-5/9 = -b/(a+b)$ para movimentos para baixo, e são as seguintes:

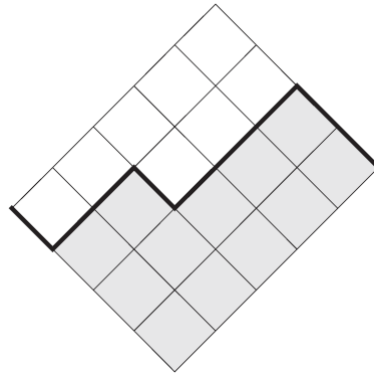


Figura 3.2: Um caminho de rede através do retângulo 4×5 , com as caixas 1×1 por baixo dele sombreadas. Aqui, $a = 4$, $b = 5$ e $n = 9$, e a codificação do caminho de rede é 011011100.

x	$-9/2$	$-7/2$	$-5/2$	$-3/2$	$-1/2$	$1/2$	$3/2$	$5/2$	$7/2$	$9/2$
$h(x)$	0	$-5/9$	$-1/9$	$3/9$	$-2/9$	$2/9$	$6/9$	$10/9$	$5/9$	0

A função de deslocamento de h que achamos útil é

$$\Phi(h) = \sum_{x=-n/2}^{n/2} h(x) \cos \frac{\beta x}{n}, \quad (3.1)$$

onde $0 \leq \beta \leq \pi$. Esta função pesa os desvios da expectativa mais fortemente perto do meio do caminho do que perto dos seus pontos finais. Dados dois caminhos de rede com funções de altura \check{h} e \hat{h} , onde $\check{h} \leq \hat{h}$ para todo x , defina a função de lacuna como $\hat{h} - \check{h}$ e a lacuna como

$$\Phi(\hat{h}, \check{h}) = \Phi(\hat{h} - \check{h}) = \Phi(\hat{h}) - \Phi(\check{h}).$$

Observe que, como $0 \leq \beta \leq \pi$, a lacuna é estritamente positiva quando os caminhos \check{h} e \hat{h} são diferentes e é igual a 0 caso contrário. Após a cadeia de Markov ter alcançado o equilíbrio, de modo que cada caminho tenha a mesma probabilidade, $E[h(x)] = 0$, portanto, o deslocamento esperado é $E[\Phi(h(x))] = 0$.

Lema 3.1. *Seja a função de deslocamento Φ definida por (3.1). Suponha que h seja uma função de altura (portanto, $\Phi(h)$ é o deslocamento) e $\beta = \pi$ ou $h = \hat{h} - \check{h}$ seja uma função de lacuna (então $\Phi(h)$ é uma lacuna) e $0 \leq \beta \leq \pi$. Seja h' a função de altura ou lacuna após um passo da cadeia de Markov. Então*

$$E[\Phi(h') - \Phi(h)|h] \leq \frac{-1 + \cos(\beta/n)}{n-1} \Phi(h),$$

com igualdade quando $\beta = \pi$. O coeficiente do lado direito é limitado por

$$-\frac{\beta^2}{2n^2(n-1)} \leq \frac{-1 + \cos(\beta/n)}{n-1} \leq -\frac{\beta^2}{2n^3}. \quad (3.2)$$

Demonstração. Suponha que escolhamos um local x , lancemos uma moeda e ajustemos a altura de acordo. Nesse caso, o valor esperado da nova altura em x é simplesmente $[h(x+1) + h(x-1)]/2$. Assuma que escolhemos cada local (exceto $-n/2$ e $n/2$) com probabilidade $1/p$, onde $p = n - 1$ é o número de posições que podem ser escolhidas. Então, seja denotando a variável h atualizada no próximo passo,

$$E[h'(x)|h] = \frac{p-1}{p}h(x) + \frac{1}{p} \frac{h(x+1) + h(x-1)}{2},$$

quando $-n/2 < x < n/2$, considerando que $E[h'(\pm n/2)] = 0$, de modo que

$$\begin{aligned} E[\Phi(h')|h] &= E\left[\sum_{x=-n/2+1}^{n/2-1} h'(x) \cos \frac{\beta x}{n} | h\right] \\ &= \sum_{x=-n/2+1}^{n/2-1} E\left[h'(x) \cos \frac{\beta x}{n} | h\right] \\ &= \sum_{x=-n/2+1}^{n/2-1} \cos \frac{\beta x}{n} E[h'(x)|h] \\ &= \sum_{x=-n/2+1}^{n/2-1} \cos \frac{\beta x}{n} \left(\frac{p-1}{p}h(x) + \frac{1}{p} \frac{h(x+1) + h(x-1)}{2} \right) \\ &= \sum_{x=-n/2+1}^{n/2-1} \left(\cos \frac{\beta x}{n} \cdot \frac{p-1}{p} \cdot h(x) + \cos \frac{\beta x}{n} \cdot \frac{1}{p} \cdot \frac{h(x+1) + h(x-1)}{2} \right) \\ &= \frac{p-1}{p} \sum_{x=-n/2+1}^{n/2-1} h(x) \cos \frac{\beta x}{n} + \frac{1}{2p} \sum_{x=-n/2+1}^{n/2-1} (h(x+1) + h(x-1)) \cos \frac{\beta x}{n} \\ &= \frac{p-1}{p} \Phi(h) + \frac{1}{2p} \sum_{x=-n/2+1}^{n/2-1} (h(x+1) + h(x-1)) \cos \frac{\beta x}{n}. \end{aligned}$$

Como $E[\Phi(h)|h] = \Phi(h)$ temos que:

$$E[\Phi(h') - \Phi(h)|h] = E[\Phi(h')|h] - E[\Phi(h)|h] = \frac{-1}{p} \Phi(h) + \frac{1}{2p} \sum_{x=-n/2+1}^{n/2-1} (h(x+1) + h(x-1)) \cos \frac{\beta x}{n},$$

reescrevendo temos que

$$\begin{aligned}
E[\Phi(h') - \Phi(h)|h] &= \frac{-1}{p}\Phi(h) + \frac{1}{2p} \sum_{x=-n/2+1}^{n/2-1} \sum_{\substack{y=-n/2 \\ |x-y|=1}}^{n/2} h(y) \cos \frac{\beta x}{n} \\
&= \frac{-1}{p}\Phi(h) + \frac{1}{2p} \sum_{x=-n/2+1}^{n/2-1} \sum_{\substack{y=-n/2+1 \\ |x-y|=1}}^{n/2-1} h(y) \cos \frac{\beta x}{n} \\
&= \frac{-1}{p}\Phi(h) + \frac{1}{2p} \sum_{y=-n/2+1}^{n/2-1} h(y) \sum_{\substack{x=-n/2+1 \\ |x-y|=1}}^{n/2-1} \cos \frac{\beta x}{n} \\
&\leq \frac{-1}{p}\Phi(h) + \frac{1}{2p} \sum_{y=-n/2+1}^{n/2-1} h(y) \sum_{\substack{x=-n/2 \\ |x-y|=1}}^{n/2} \cos \frac{\beta x}{n}.
\end{aligned}$$

Essa desigualdade é justificada pelo fato de que, para $x = \pm n/2$, temos $\cos(\pm\beta/2) \geq 0$, pois $\beta \leq \pi$, e também porque assumimos que $h = \hat{h} - \check{h} \geq 0$. Calculando o somatório onde x está variando, segue que

$$E[\Phi(h') - \Phi(h)|h] \leq \frac{-1}{p}\Phi(h) + \frac{1}{2p} \sum_{y=-n/2+1}^{n/2-1} h(y) \left[\cos \frac{\beta(y+1)}{n} + \cos \frac{\beta(y-1)}{n} \right]. \quad (3.3)$$

Pela identidade trigonométrica $\cos(\alpha + \theta) + \cos(\alpha - \theta) = 2 \cos(\alpha) \cos(\theta)$, segue que $\cos \frac{\beta(y+1)}{n} + \cos \frac{\beta(y-1)}{n} = 2 \cos \frac{\beta y}{n} \cos \frac{\beta}{n}$. Daí, a inequação (3.3) pode ser reescrita da seguinte forma

$$\begin{aligned}
E[\Phi(h') - \Phi(h)|h] &\leq \frac{-1}{p}\Phi(h) + \frac{1}{2p} \sum_{y=-n/2+1}^{n/2-1} h(y) 2 \cos \frac{\beta y}{n} \cos \frac{\beta}{n} \\
&= \frac{-1}{p}\Phi(h) + \frac{1}{p} \cos \frac{\beta}{n} \sum_{y=-n/2+1}^{n/2-1} h(y) \cos \frac{\beta y}{n} \\
&= \frac{-1}{p}\Phi(h) + \frac{1}{p} \cos \frac{\beta}{n} \Phi(h) \\
&= \frac{-1 + \cos(\beta/n)}{p} \Phi(h) \\
&= \frac{-1 + \cos(\beta/n)}{n-1} \Phi(h).
\end{aligned}$$

A igualdade ocorre quando $\beta = \pi$.

Para limitar superiormente o lado direito de (3.2), usamos a desigualdade $\cos(x) \leq 1 - x^2/2 + x^4/24 = 1 + (x^2/2)(-1 + x^2/12)$ (que pode ser obtida tomando como base a Série de Maclaurin $\cos(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{2n!} x^{2n}$, que vale para todo $x \in \mathbb{R}$). Então, tomando $x = \beta/n$,

$$\frac{-1 + \cos(\beta/n)}{n-1} \leq \frac{1}{n-1} \cdot \frac{(\beta/n)^2}{2} \left(-1 + \frac{(\beta/n)^2}{12} \right) = -\frac{\beta^2}{2n^2} \left[1 - (\beta^2/12)/n^2 \right] / (n-1). \quad (3.4)$$

Se $\beta = 0$ obtemos a igualdade. Como $\beta \leq \pi$ implica que $\beta^2 \leq \pi^2 \Rightarrow \beta^2/12 \leq \pi^2/12$, logo $\beta^2/12 < 1 < n \Rightarrow -\beta^2/12 > -n \Rightarrow n - (\beta^2/12)/n > n - 1 \Rightarrow n \left[1 - (\beta^2/12)/n^2 \right] > n - 1$, agora dividindo ambos os lados desta última desigualdade por $n(n - 1)$, temos:

$$\left[1 - (\beta^2/12)/n^2 \right] / (n - 1) > 1/n \Rightarrow -\frac{\beta^2}{2n^2} \left[1 - (\beta^2/12)/n^2 \right] / (n - 1) < -\frac{\beta^2}{2n^3}. \quad (3.5)$$

O resultado segue de (3.4) e de (3.5). A cota inferior de (3.2) é mais fácil. Basta usar a desigualdade $\cos(x) \geq 1 - \frac{x^2}{2}$ que o resultado é imediato. \square

3.1.2 Cota Superior

Teorema 3.2. *Quando n é grande, após*

$$\frac{2 + o(1)}{\pi^2} n^3 \log \frac{ab}{\varepsilon}$$

etapas, a distância de variação da estacionariedade é ε e a probabilidade de que os dois caminhos extremos se uniram é $1 - \varepsilon$. (O termo $o(1)$ é uma função apenas de n .)

O que esse teorema nos diz é que se $t = \frac{2 + o(1)}{\pi^2} n^3 \log \frac{ab}{\varepsilon}$, então $d(t) \leq \varepsilon$, isso implica, pela Definição 1.34, que $t_{mix} \leq \frac{2 + o(1)}{\pi^2} n^3 \log \frac{ab}{\varepsilon}$.

Demonstração. A ideia aqui é criar um acoplamento específico entre duas trajetórias extremas (estado inicial). O objetivo é achar um tempo t tal que $P[\tau_{couple} > t] \leq \varepsilon$. Usando o Corolário (1.41), isso implica que $d(t) \leq \varepsilon$. Após isso vamos minimizar sobre todos os tempos t satisfazendo $P[\tau_{couple} > t] \leq \varepsilon$.

Para obter a cota superior, consideramos um par de caminhos acoplados \hat{h}_t e \check{h}_t tais que \hat{h}_0 é o caminho mais alto e \check{h}_0 é o caminho mais baixo. As sequências \hat{h}_t e \check{h}_t são geradas pela cadeia de Markov usando “os mesmos movimentos aleatórios”, de modo que \hat{h}_{t+1} e \check{h}_{t+1} são obtidos de \hat{h}_t e \check{h}_t , respectivamente, por ordenação ou desordenação (mesma decisão aleatória tomada em ambos os casos) no mesmo local aleatório x .

Como o acoplamento foi definido, é claro que $\hat{h}_t \geq \check{h}_t$ (o que também pode ser provado por indução sobre t). Seja $\Phi(t) = \Phi(\hat{h}_t - \check{h}_t)$; $\Phi(t) = 0$ se, e somente se $\hat{h}_t = \check{h}_t$.

Vamos mostrar por indução sobre t que

$$E[\Phi(t)] \leq \Phi_0 \left(1 - \frac{1 - \cos(\beta/n)}{n - 1} \right)^t \quad (3.6)$$

Suponha que a cadeia tenha início com Φ_0 , logo escrevemos $E[\Phi_1] = E[\Phi_1 | \Phi_0] = E[\Phi_1 | \Phi_0] -$

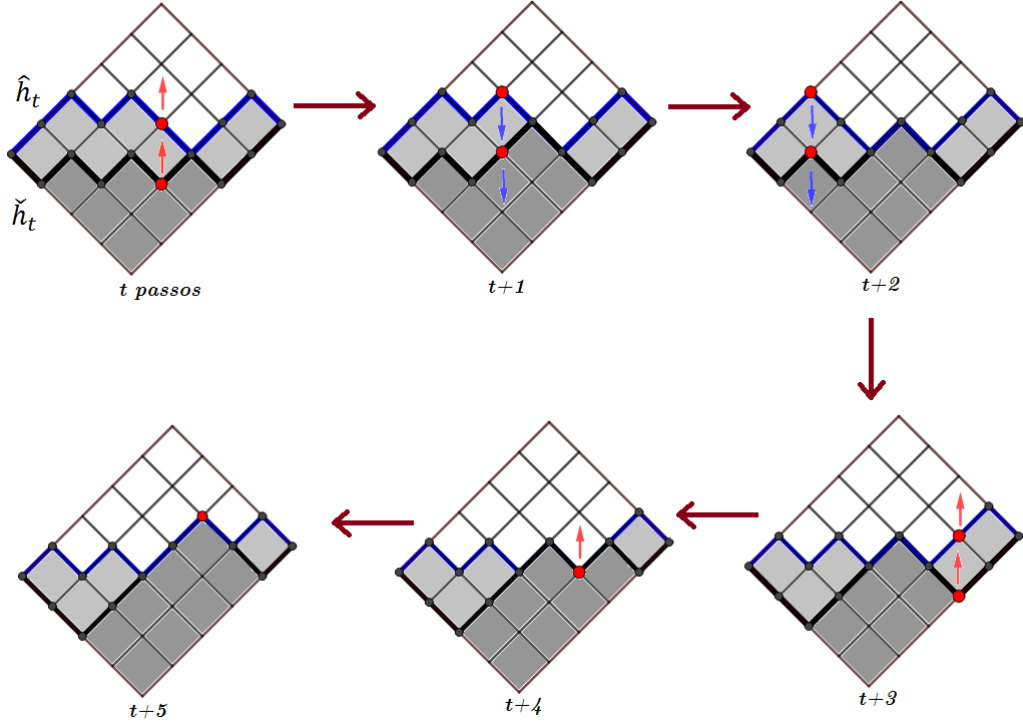


Figura 3.3: Um exemplo de um par de caminhos acoplados, sendo \hat{h}_t a trajetória azul e \check{h}_t a preta. Observe que, embora representando apenas 5 passos dessas cadeias, a trajetória azul estará sempre acima da preta até que elas coincidam.

$E[\Phi_0|\Phi_0] + E[\Phi_0|\Phi_0] = E[\Phi_1 - \Phi_0|\Phi_0] + \Phi_0$. Então, pelo Lema 3.1, temos

$$E[\Phi_1] \leq \Phi_0 + \frac{-1 + \cos(\beta/n)}{n-1} \Phi(0) = \Phi_0 \left(1 - \frac{1 - \cos(\beta/n)}{n-1} \right).$$

Tomando como válida nossa hipótese de indução para t e usando o Lema 3.1, temos que

$$\begin{aligned} E[\Phi_{t+1}] &= \sum_y y P(\Phi_{t+1} = y) \\ &= \sum_y y \sum_z P(\Phi_{t+1} = y, \Phi_t = z) \\ &= \sum_y y \sum_z P(\Phi_{t+1} = y | \Phi_t = z) P(\Phi_t = z) \\ &= \sum_z P(\Phi_t = z) \sum_y y P(\Phi_{t+1} = y | \Phi_t = z) \\ &= \sum_z P(\Phi_t = z) E[\Phi_{t+1} | \Phi_t = z] \\ &\leq \sum_z P(\Phi_t = z) \left(1 - \frac{1 - \cos(\beta/n)}{n-1} \right) z \end{aligned}$$

$$\begin{aligned}
&= \left(1 - \frac{1 - \cos(\beta/n)}{n-1}\right) \sum_z z P(\Phi_t = z) \\
&= \left(1 - \frac{1 - \cos(\beta/n)}{n-1}\right) E[\Phi_t] \\
&\leq \Phi_0 \left(1 - \frac{1 - \cos(\beta/n)}{n-1}\right)^{t+1}.
\end{aligned}$$

Além disso, usando a parte direita da desigualdade do Lema 3.2, juntamente com o fato de que $|\frac{\beta^2}{2n^3}| < 1$, temos que

$$\left(1 - \frac{1 - \cos(\beta/n)}{n-1}\right)^t \leq \left(1 - \frac{\beta^2}{2n^3}\right)^t \leq e^{-\frac{\beta^2 t}{2n^3}}.$$

Conseqüentemente

$$E[\Phi(t)] \leq \Phi_0 \left(1 - \frac{1 - \cos(\beta/n)}{n-1}\right)^t \leq \Phi_0 \exp\left(-\frac{\beta^2 t}{2n^3}\right).$$

No entanto, usando a desigualdade de Markov, onde Φ_{\min} é o menor valor positivo assumido por Φ , temos que $E[\Phi_t] \geq P[\Phi_t \geq \Phi_{\min}] \Phi_{\min} = P[\Phi_t > 0] \Phi_{\min}$. Assim, após $t \geq (2/\beta^2)n^3 \log(\Phi_0/(\Phi_{\min}\varepsilon))$ passos, temos

$$\begin{aligned}
P[\Phi_t > 0] &\leq E[\Phi_t]/\Phi_{\min} \leq \Phi_0 \exp\left[-\frac{\beta^2(2/\beta^2)n^3 \log(\Phi_0/(\Phi_{\min}\varepsilon))}{2n^3}\right] / \Phi_{\min} \\
&= \Phi_0 \exp[-\log(\Phi_0/(\Phi_{\min}\varepsilon))] / \Phi_{\min} \\
&= \Phi_0(\Phi_{\min}\varepsilon/\Phi_0) / \Phi_{\min} = \varepsilon.
\end{aligned}$$

Portanto para valores de $t \geq (2/\beta^2)n^3 \log(\Phi_0/(\Phi_{\min}\varepsilon))$, $d(t) \leq \varepsilon$. Além disso $1 - \varepsilon \leq P[\Phi_t = 0] \leq 1$, isto significa que a probabilidade de que os dois caminhos extremos se uniram é de pelo menos $1 - \varepsilon$.

Vamos agora encontrar uma estimativa para t . Pela função de deslocamento dada em (3.1) temos que

$$\Phi_0 = \Phi(\hat{h}_0 - \check{h}_0) = \Phi(\hat{h}_0) - \Phi(\check{h}_0) = \sum_{x=-n/2}^{n/2} (\hat{h}_0 - \check{h}_0) \cos \frac{\beta x}{n}.$$

Considerando que $\cos \frac{\beta x}{n} \leq 1$, temos que $\Phi_0 \leq \sum_{x=-n/2}^{n/2} (\hat{h}_0 - \check{h}_0)$. Como \hat{h}_0 é o caminho mais alto, então ele deve percorrer, no retângulo $a \times b$, $n = b + a$ passos, sendo b passos no sentido nordeste e em seguida a passos no sentido sudeste, daí $\hat{h}_0(-n/2 + b) = b \cdot a/n$ e

$\hat{h}_0(-n/2+b+a) = b \cdot a/n - a \cdot b/n = 0$. Tomando $k \in \{0, 1, \dots, b\}$, $\hat{h}_0(-n/2+k) = \cdot ka/n$ e para $k \in \{b+1, b+2, \dots, b+a\}$, $\hat{h}_0(-n/2+k) = \cdot ba/n - (k-b) \cdot b/n$. Calculando o somatório

$$\begin{aligned}
\sum_{x=-n/2}^{n/2} \hat{h}_0 &= \sum_{k=0}^b \frac{ka}{n} + \sum_{k=b+1}^{b+a} \left[\frac{ba}{n} - \frac{(k-b)b}{n} \right] \\
&= \frac{a}{n} \sum_{k=0}^b k + \frac{ba^2}{n} - \frac{b}{n} \sum_{k=b+1}^{b+a} (k-b) \\
&= \frac{a}{n} \cdot \frac{b(b+1)}{2} + \frac{ba^2}{n} - \frac{b}{n} \cdot \frac{(1+a)a}{2} \\
&= ab \left[\frac{b+1}{2n} + \frac{a}{n} - \frac{1+a}{2n} \right] \\
&= \frac{ab}{2}.
\end{aligned}$$

De modo semelhante \check{h}_0 é o caminho mais baixo, logo deve percorrer a passos no sentido sudeste e em seguida b passos no sentido nordeste, daí $\check{h}_0(-n/2+a) = -a \cdot b/n$ e $\check{h}_0(-n/2+a+b) = -a \cdot b/n + b \cdot a/n = 0$. Tomando $k \in \{0, 1, \dots, a\}$, $\check{h}_0(-n/2+k) = -k \cdot a/n$ e para $k \in \{a+1, a+2, \dots, a+b\}$, $\check{h}_0(-n/2+k) = -b \cdot a/n + (k-a) \cdot a/n$. Daí, pode-se facilmente ver que $\sum_{x=-n/2}^{n/2} \check{h}_0 = -\sum_{x=-n/2}^{n/2} \hat{h}_0 = -\frac{ab}{2}$. Portanto, $\Phi_0 \leq ab$. Por outro lado, temos $\Phi_{\min} \neq 0$, quando minimizamos $h_t \neq 0$. Como $h_t(x) = \hat{h}_t(x) - \check{h}_t(x)$, desejamos que

$$h_t(x) = \begin{cases} 0, & \text{se } x \neq n/2 - 1 \\ 1, & \text{se } x = n/2 - 1. \end{cases}$$

Daí,

$$\Phi_{\min} = \Phi(h_t) = \sum_{x=-n/2}^{n/2} h_t(x) \cos \frac{\beta x}{n} = \cos \frac{\beta(n/2-1)}{n} = \cos(\beta/2 - \beta/n) > \cos(\beta/2).$$

Por outro lado sabemos que $\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1$, usando que $\sin(\pi - \beta)/2 = \cos(\beta/2)$, isso implica que $\Phi_{\min} > \cos(\beta/2) \approx (\pi - \beta)/2$. Aqui usamos o símbolo \approx para denotar o fato de que $\frac{\cos(\beta/2)}{(\pi - \beta)/2} \approx 1$, quando $\beta \rightarrow \pi$. Agora vamos minimizar a expressão $K = (2/\beta^2)n^3 \log(\Phi_0/(\Phi_{\min}\varepsilon))$, para obter uma cota melhor para nosso tempo de mistura.

$$K = (2/\beta^2)n^3 \log(\Phi_0/(\Phi_{\min}\varepsilon)) \approx (2/\beta^2)(\log ab/\varepsilon - \log(\pi - \beta)/2)n^3.$$

Desejamos que $\Phi_{\min} \rightarrow 0$, tendo isso em vista, escolhamos convenientemente $\beta = \pi -$

$1/\log n$, já que para n grande, $\beta - \pi \rightarrow 0$. Então

$$\begin{aligned} K &\approx (2/(\pi - 1/\log n)^2) [\log ab/\varepsilon - \log 1/(2 \log n)] n^3 \\ &= (2/(\pi - 1/\log n)^2) [\log ab/\varepsilon + \log 2 + \log(\log n)] n^3 \\ &= (2/(\pi - 1/\log n)^2) \left[1 + \frac{\log 2 + \log(\log n)}{\log ab/\varepsilon} \right] n^3 \log ab/\varepsilon. \end{aligned}$$

Contudo, para n grande temos que $2/(\pi - 1/\log n)^2 \approx 2/\pi^2$ e também podemos assumir que $ab/\varepsilon \geq n$, implica $\log ab/\varepsilon \geq \log n$, segue que

$$\begin{aligned} K &\approx \frac{2}{\pi^2} \left[1 + \frac{\log 2 + \log(\log n)}{\log n} \right] n^3 \log ab/\varepsilon \\ &\approx \left[\frac{2}{\pi^2} + \frac{\log 2 + \log(\log n)}{\log n} \frac{2}{\pi^2} \right] n^3 \log ab/\varepsilon \\ &\approx \left[\frac{2}{\pi^2} + \Theta(\log(\log n)/\log n) \right] n^3 \log ab/\varepsilon \\ &= \frac{2 + o(1)}{\pi^2} n^3 \log \frac{ab}{\varepsilon}. \end{aligned}$$

O símbolo Θ significa que, dadas duas funções f e g que sejam não negativas para valores grandes de x , escrevemos $g(x) = \Theta(f(x))$, se existem constantes positivas c e d , tais que $cg(x) \leq f(x) \leq dg(x)$ para x suficientemente grande. \square

3.1.3 Cota Inferior

Vamos obter uma cota inferior para o tempo de mistura quando o retângulo não é muito estreito. Para a demonstração do Teorema 3.4 usaremos o seguinte lema técnico:

Lema 3.3. *Se uma função Φ no espaço de estados de uma cadeia de Markov satisfaz $E[\Phi(X_{t+1})|X_t] = (1 - \gamma)\Phi(X_t)$, e $E[(\Delta\Phi)^2|X_t] \leq R$, onde $\Delta\Phi = \Phi(X_{t+1}) - \Phi(X_t)$ e Φ_{\max} é o valor máximo assumido por Φ , então, quando o número de etapas da cadeia de Markov t é limitado por*

$$t \leq \frac{\log \Phi_{\max} + \frac{1}{2} \log(\gamma\varepsilon/(4R))}{-\log(1 - \gamma)}$$

e $0 < \gamma \leq 2 - \sqrt{2} \approx 0,58$ (ou então $0 < \gamma \leq 1$ e t é ímpar), então a distância de variação da estacionariedade é pelo menos $1 - \varepsilon$.

Demonstração. Nesta demonstração, se não for dito diferente, X_0 sempre seja um ponto inicial determinístico da cadeia $(X_t)_{t \in \mathbb{N}}$. Assim, para qualquer t temos que $E[\Phi_t] = E[\Phi_t|X_0]$, onde $\Phi_t = \Phi(X_t)$. Note que $E[\Phi_t|X_0] = \Phi_0(1 - \gamma)^t$. De fato, pela nossa hipótese $E[\Phi_1|X_0] = \Phi_0(1 - \gamma)$. Supomos agora que $E[\Phi_t|X_0] = \Phi_0(1 - \gamma)^t$ seja válida para algum $t > 1$. Então

$$E[\Phi_{t+1}|X_0] = E[E[\Phi_{t+1}|X_t]|X_0] = E[(1 - \gamma)\Phi_t|X_0] = (1 - \gamma)E[\Phi_t|X_0] = (1 - \gamma)^{t+1}\Phi_0. \quad (3.7)$$

Pela suposição sobre γ , em equilíbrio $E[\Phi] = 0$, e escrevemos $\Phi = \Phi_\infty$. De fato, em equilíbrio temos que $E[\Phi_t] = E[\Phi_{t+1}] = (1 - \gamma)E[\Phi_t]$. Se $\gamma = 1$, então $E[\Phi_t] = E[\Phi_{t+1}] = 0$. Se $\gamma \neq 1$, então $\gamma E[\Phi_t] = 0$, como $\gamma \neq 0$, logo $E[\Phi_t] = 0$.

Com Δ_t denotando $\Phi_{t+1} - \Phi_t$, nós temos:

$$\begin{aligned}\Phi_{t+1}^2 &= \Phi_t^2 + 2\Phi_t\Delta_t + \Delta_t^2, \text{ então} \\ E[\Phi_{t+1}^2|X_t] &= E[(\Phi_t^2 + 2\Phi_t\Delta_t + \Delta_t^2)|X_t] \\ &= E[\Phi_t^2|X_t] + 2\Phi_t[E[\Phi_{t+1} - \Phi_t]] + E[\Delta_t^2|X_t] \\ &\leq \Phi_t^2 + 2\Phi_t[(1 - \gamma)\Phi_t - \Phi_t] + R \\ &= (1 - 2\gamma)\Phi_t^2 + R.\end{aligned}$$

Afirmamos agora que $E[\Phi_t^2] \leq (1 - 2\gamma)^t\Phi_0^2 + \frac{R}{2\gamma}$ para todo $t \geq 2$. De fato, pelo resultado anterior, $E[\Phi_1^2] = E[\Phi_1^2|X_0] \leq (1 - 2\gamma)\Phi_0^2 + R$. Dai, para $t = 2$

$$\begin{aligned}E[\Phi_2^2] &= E[E[\Phi_2^2|X_1]] \leq E[(1 - 2\gamma)\Phi_1^2 + R], \text{ então} \\ E[\Phi_2^2] &\leq (1 - 2\gamma)E[\Phi_1^2] + R.\end{aligned}$$

Isso implica que

$$\begin{aligned}E[\Phi_2^2] &\leq (1 - 2\gamma)[(1 - 2\gamma)\Phi_0^2 + R] + R \\ E[\Phi_2^2] &\leq (1 - 2\gamma)^2\Phi_0^2 + (2 - 2\gamma)R \leq (1 - 2\gamma)^2\Phi_0^2 + \frac{R}{2\gamma}.\end{aligned}\tag{3.8}$$

Continuando por indução obtemos que

$$\begin{aligned}E[\Phi_{t+1}^2] &= E[E(\Phi_{t+1}^2|X_t)] \leq E[(1 - 2\gamma)\Phi_t^2 + R] \\ &= (1 - 2\gamma)E[\Phi_t^2] + R \\ &\leq (1 - 2\gamma) \left[(1 - 2\gamma)^t\Phi_0^2 + \frac{R}{2\gamma} \right] + R \\ &= (1 - 2\gamma)^{t+1}\Phi_0^2 + \frac{R}{2\gamma}.\end{aligned}$$

Calculamos agora a variância de Φ_t , (lembre que $E[\Phi_t] = E[\Phi_t|\Phi_0]$), temos:

$$Var[\Phi_t] = E[\Phi_t^2] - E[\Phi_t]^2 \leq (1 - 2\gamma)^t\Phi_0^2 + \frac{R}{2\gamma} - \Phi_0^2(1 - \gamma)^{2t},\tag{3.9}$$

$$Var[\Phi_t] \leq \Phi_0^2 [(1 - 2\gamma)^t - (1 - \gamma)^{2t}] + \frac{R}{2\gamma}.\tag{3.10}$$

Vamos analisar o lado direito da desigualdade (3.10). Como $(1 - \gamma)^2 = 1 - 2\gamma + \gamma^2 \geq 1 - 2\gamma$, então quando t é ímpar, $(1 - \gamma)^{2t} \geq (1 - 2\gamma)^t$, isso implica que $Var[\Phi_t] \leq \frac{R}{2\gamma}$. No entanto, quando t é par, nós precisamos ter $(1 - \gamma)^{2t} \geq |1 - 2\gamma|$, que é satisfeito quando $\gamma \leq 2 - \sqrt{2}$

ou $\gamma \geq 2 + \sqrt{2}$, daí $Var[\Phi_t] \leq \frac{R}{2\gamma}$.

Da desigualdade de Chebychev temos:

$$P \left[|\Phi_t - E(\Phi_t)| \geq \sqrt{R/(\gamma\varepsilon)} \right] \leq \frac{\varepsilon}{2}. \quad (3.11)$$

Isso implica que

$$P \left[|\Phi_t - E(\Phi_t)| \leq \sqrt{R/(\gamma\varepsilon)} \right] \geq 1 - \frac{\varepsilon}{2}. \quad (3.12)$$

Suponha agora que $E[\Phi_t] \geq \sqrt{4R/(\gamma\varepsilon)}$. Neste caso temos a seguinte inclusão de eventos:

$$\{|\Phi_t - E[\Phi_t]| \leq \sqrt{R/(\gamma\varepsilon)}\} \subseteq \{\Phi_t \geq \sqrt{R/(\gamma\varepsilon)}\}. \quad (3.13)$$

Daí, por (3.12)

$$P \left[\Phi_t \geq \sqrt{R/(\gamma\varepsilon)} \right] \geq 1 - \frac{\varepsilon}{2}. \quad (3.14)$$

Por outro lado, usando que $E[\Phi_\infty] = 0$,

$$P \left[\Phi_\infty \geq \sqrt{R/(\gamma\varepsilon)} \right] \leq \frac{\varepsilon}{2}. \quad (3.15)$$

Daí, denotando com $d(t)$ a variação total entre Φ_t e Φ_∞

$$\begin{aligned} d(t) &= \max_A |P(\Phi_t \in A) - P(\Phi_\infty \in A)| \\ &\geq |P(\Phi_t \geq \sqrt{R/(\gamma\varepsilon)}) - P(\Phi_\infty \geq \sqrt{R/(\gamma\varepsilon)})| \\ &\geq 1 - \varepsilon. \end{aligned} \quad (3.16)$$

E esse tempo é limitado por

$$\begin{aligned} E[\Phi_t] = \Phi_0(1 - \gamma)^t &\geq \sqrt{4R/(\gamma\varepsilon)}, \text{ vamos isolar } t \\ (1 - \gamma)^t &\geq \Phi_0^{-1} \sqrt{4R/(\gamma\varepsilon)} \\ (1 - \gamma)^{-t} &\leq \Phi_0 \sqrt{\gamma\varepsilon/(4R)} \\ -t \log(1 - \gamma) &\leq \log(\Phi_0 \sqrt{\gamma\varepsilon/(4R)}), \text{ note que } -\log(1 - \gamma) > 0, \text{ daí} \\ t &\leq \frac{\log(\Phi_0 \sqrt{\gamma\varepsilon/(4R)})}{-\log(1 - \gamma)}. \end{aligned}$$

Agora vamos escrever Φ_{\max} como o estado inicial que maximiza Φ_0 , e o resultado segue. \square

Teorema 3.4. *Se $\min(a, b) \gg 1$, então depois de*

$$\frac{1 - o(1)}{\pi^2} n^3 \log \min(a, b)$$

passos, a distância de variação da estacionariedade é $1 - o(1)$.

Observação: $\min(a, b) \gg 1$ significa que o menor valor entre a e b é muito maior do que 1. Em nosso contexto, isso implica que a e b crescem para o infinito com n .

Demonstração. Começemos mostrando que a função Φ satisfaz a propriedade de contração do Lema 3.3. De fato, tomando $\beta = \pi$, da definição da função Φ temos

$$\Delta\Phi = \sum_{x=-n/2}^{n/2} [h'(x) - h(x)] \cos \frac{\pi x}{n}.$$

Lembre que h' é obtido de h , ou mantendo a mesma trajetória, ou empurrando um ponto da trajetória para cima ou para baixo. Se $h' = h$ suponha que seja num ponto k qualquer no intervalo $[-n/2, n/2]$, nesse caso teremos que

$$\Delta\Phi = [h'(k) - h(k)] \cos \frac{\pi k}{n} \Rightarrow |\Delta\Phi| \leq 1 \Rightarrow (\Delta\Phi)^2 \leq 1,$$

isso implica que $0 \leq E[(\Delta\Phi)^2|h] \leq 1$.

Por outro lado, pelo Lema 3.1 temos que

$$E[\Phi(h') - \Phi(h)|h] = \frac{-1 + \cos(\pi/n)}{n-1} \Phi(h)$$

$$E[\Phi(h')|h] = \left(1 - \frac{1 - \cos(\pi/n)}{n-1}\right) \Phi(h).$$

Tomando $\gamma = \frac{1 - \cos(\pi/n)}{n-1} = -\frac{-1 + \cos(\pi/n)}{n-1} \geq \frac{\pi^2}{2n^3}$. Além disso $\frac{\pi^2}{2n^3}$ é uma boa aproximação para γ . De fato, da expansão de Taylor do \cos em torno de $x_0 = 0$, temos que:

$$\cos(x) = \cos 0 + x \cos' 0 + \frac{x^2}{2} \cos'' 0 + O(x^3), \text{ substituindo } x = \pi/n$$

$$\cos(\pi/n) = 1 - \frac{\pi^2}{2n^2} + O(\pi^3/n^3).$$

Substituindo esse valor em γ obtemos:

$$\gamma = \frac{1 - 1 + \pi^2/2n^2 - O(\pi^3/n^3)}{n-1} = \frac{\pi^2}{2n^2(n-1)} - O(\pi^3/n^4).$$

Agora, note que $\gamma \cdot \frac{2n^3}{\pi^2} = \frac{n}{n-1} - O(2/n) \rightarrow 1$, quando $n \rightarrow \infty$. Portanto $\gamma \approx \frac{\pi^2}{2n^3}$. A restrição em γ é satisfeita quando $n \geq 3$.

Para obter uma cota R , observe que qualquer caminho h pode ter no máximo $2 \min(a, b)$ pontos extremos locais (pontos que podem ser empurrados para baixo ou para cima no próximo passo da cadeia). Daí como temos um total de $n-1$ pontos que podem ser escolhidos, e desses pontos extremos locais temos probabilidade ser empurrado de $1/2$, então $P[\Delta\Phi \neq 0] = \frac{\min(a, b)}{n-1}$. Mas $|\Delta\Phi| \leq 1$, então $\max_h E[(\Delta\Phi(h))^2] \leq \frac{\min(a, b)}{n-1} \equiv R$.

O caminho máximo maximiza Φ , dando $\Phi_0 = \Theta(ab)$, lembre que $\Phi_0 \leq ab$ quando \hat{h}_0 e \check{h}_0 são extremais. Então, pelo Lema 3.3, $d(t) \geq 1 - \varepsilon$, quando

$$t \leq \frac{\log ab + \frac{1}{2} \log(\gamma\varepsilon/4R)}{-\log(1 - \gamma)}. \quad (3.17)$$

Considere agora as seguinte simplificações, considerando quando n é grande:

- (i) $\gamma/R \approx \frac{\pi^2}{2n^3} \cdot \frac{n-1}{\min(a,b)} \approx \frac{\pi^2}{2n^2 \min(a,b)}$.
(ii) Simplificando o numerador de (3.17):

$$\begin{aligned} \log ab + \frac{1}{2} \log(\gamma\varepsilon/4R) &\approx \log ab + \frac{1}{2} \log \left(\frac{\pi^2}{2n^2 \min(a,b)} \cdot \frac{\varepsilon}{4} \right) \\ &\approx \log ab - \log n + \frac{1}{2} \log \left(\frac{\pi^2 \varepsilon}{8} \right) - \frac{1}{2} \log \min(a,b). \end{aligned}$$

Por hipótese $\min(a,b) \gg 1$. Então, sem perda de generalidade, suponha que $a = \min(a,b)$, daí $\log ab = \log a(n-a) = \log \min(a,b) + \log(n-a)$. Com isso temos que:

$$\begin{aligned} \log ab + \frac{1}{2} \log(\gamma\varepsilon/4R) &\approx \log \min(a,b) + \log(n-a) - \log n + \frac{1}{2} \log \left(\frac{\pi^2 \varepsilon}{8} \right) - \frac{1}{2} \log \min(a,b) \\ &\approx \frac{1}{2} \log \min(a,b) + \log \left(1 - \frac{a}{n} \right) + \frac{1}{2} \log \left(\frac{\pi^2 \varepsilon}{8} \right). \end{aligned}$$

Como $a = \min(a,b) \Rightarrow a \leq n/2 \Rightarrow a/n \leq 1/2 \Rightarrow \log(1 - a/n) \geq \log 1/2$, e está última igualdade implica que $\log 1/2 \leq \log(1 - a/n) \leq 0$, em outras palavras $\log(1 - a/n)$ tem o comportamento de uma constante, daí $\log(1 - a/n) + \frac{1}{2} \log(\pi^2 \varepsilon/8) \in -O(1)$. Logo

$$\log ab + \frac{1}{2} \log(\gamma\varepsilon/4R) \approx \frac{1}{2} \log \min(a,b) - O(1).$$

(iii) Lembrando que $\log(1+x) \leq x$, $\forall x > -1$. Daí $\log(1-\gamma) \leq -\gamma \Rightarrow -\log(1-\gamma) \leq \gamma$. Portanto de (i), (ii) e (iii) a desigualdade (3.17) pode ser reescrita como:

$$\begin{aligned} t &\leq \frac{\log ab + \frac{1}{2} \log(\gamma\varepsilon/4R)}{-\log(1-\gamma)} \approx \frac{\frac{1}{2} \log \min(a,b) - O(1)}{\gamma} \approx \left[\frac{1}{2} \log \min(a,b) - O(1) \right] \frac{2n^3}{\pi^2} \\ t &\leq \left[1/\pi^2 - O(1)/(\pi^2 \log \min(a,b)) \right] n^3 \log \min(a,b) \\ t &\leq \left[1/\pi^2 - o(1) \right] n^3 \log \min(a,b). \end{aligned}$$

O último passo da desigualdade é justificado pelo fato de que

$$\min(a,b) \gg 1 \Rightarrow \lim_{n \rightarrow \infty} \log \min(a,b) = \infty \Rightarrow \lim_{n \rightarrow \infty} O(1)/(\pi^2 \log \min(a,b)) = 0,$$

fazendo com que $O(1)/(\pi^2 \log \min(a,b)) \in o(1)$. □

3.2 Embaralhamento de cartas por transposições adjacentes

O embaralhamento de cartas por transposições adjacentes é um método de misturar um baralho de cartas trocando a posição de cartas adjacentes. Nesse processo, você pega duas cartas consecutivas e as troca de lugar. Esse procedimento é repetido várias vezes para garantir uma distribuição aleatória das cartas.

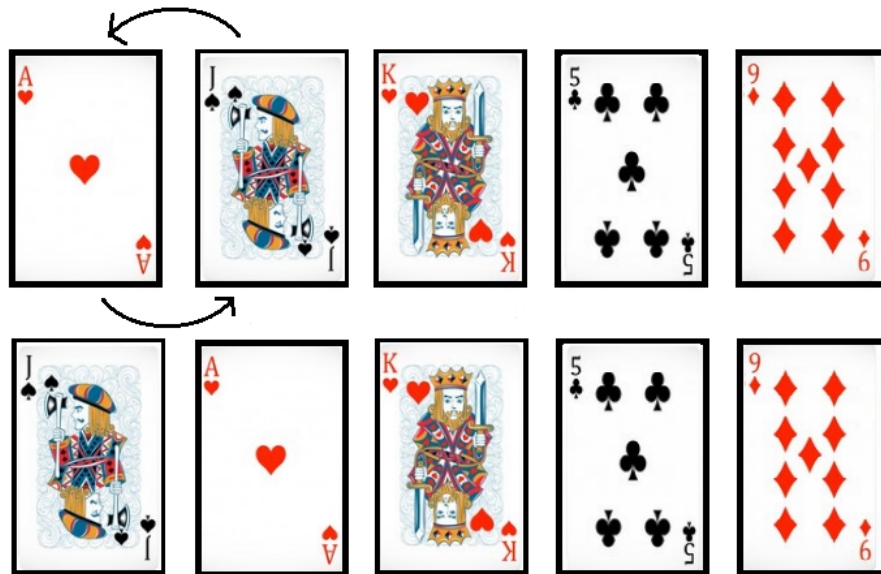


Figura 3.4: Exemplo de permutação com 5 cartas de um baralho.

O embaralhamento de cartas por transposições adjacentes pode ser modelado por uma cadeia de Markov. Consideramos um baralho com n cartas classificadas de 1 a n , do topo ao fundo, onde os estados de nossa cadeia serão as permutações das cartas, totalizando $n!$ estados. Nossa cadeia escolhe ao acaso uma carta $i \in \{1, \dots, n-1\}$ e atualizamos a cadeia da seguinte forma: jogamos uma moeda J , se sair cara trocamos a i -ésima carta com a $(i+1)$ -ésima, se sair coroa, nada acontece. Observe que, dada uma permutação (estado) de n cartas, é sempre possível, após um certo número de embaralhadas (tempo), chegar a qualquer outra permutação (estado). Isso nos diz que nossa cadeia de Markov é irredutível. Desta forma, pela Proposição 1.19, Essa cadeia de Markov tem uma (e única) distribuição estacionária.

A mesma atualização aleatória definida por i e J pode ser aplicada a mais de uma permutação para obter cadeias de Markov acopladas. Uma permutação nos números $1, \dots, n$ tem associadas a ele $(n+1)$ funções de limite, onde a i -ésima função de limite ($0 \leq i \leq n$) é uma sequência de i 1's e $n-i$ 0's, com os 1's nas localizações dos i maiores números da permutação. A permutação pode ser recuperada dessas funções de limite simplesmente somando-as (veja Figura 3.5).

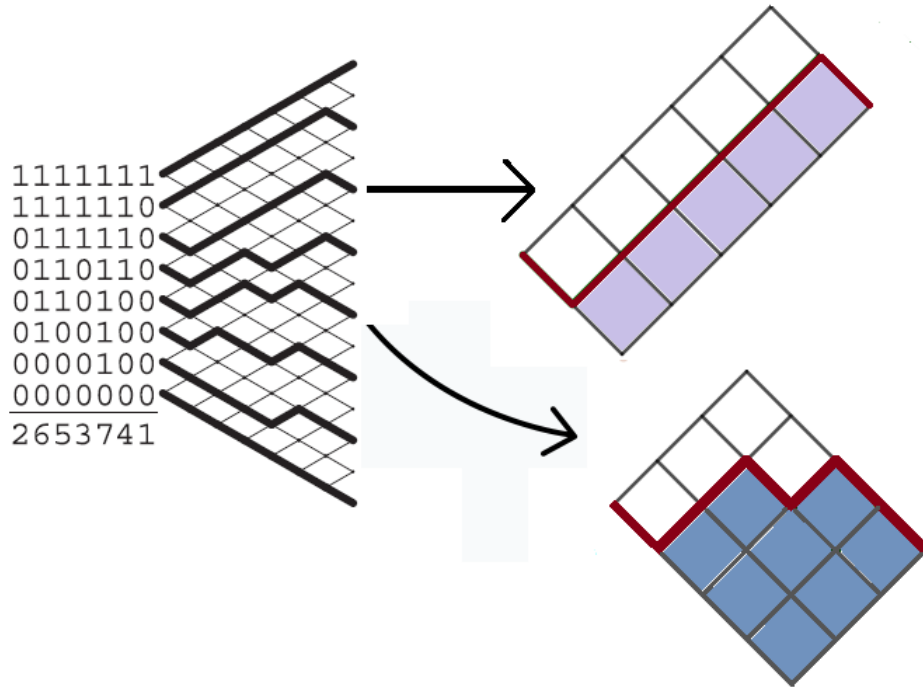


Figura 3.5: A permutação 2653741 e suas 8 funções limite mostradas como caminho de rede. Demos destaque às 3ª e 5ª funções limites.

Também podemos representar as permutações através de uma matriz $(n + 1) \times n$, onde as linhas são as funções de limite. Quando um par adjacente aleatório de números dentro da permutação são transpostos, por exemplo $(i, i + 1)$, nós transpomos as colunas $(i, i + 1)$.

Exemplo 3.5. Considere um baralho com 5 cartas. Seja o estado inicial a permutação $x_0 = 43152$, transpomos o par $(1, 5)$, resultando na permutação $x_1 = 43512$. Em seguida transpomos o par $(4, 3)$ resultando na permutação $x_2 = 34512$. Denote $A[t]$ e $x_{t[f(i)]}$ a representação matricial e a i -ésima função limite do estado t , respectivamente. Na forma matricial temos:

$$A[x_0] = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \rightarrow A[x_1] = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \rightarrow A[x_2] = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Observe que em cada matriz, as i -ésimas funções limite tem a mesma quantidade de 1's, a menos da ordem (e de 0's também). Podemos analisar cada linha dessas matrizes como um caminho de rede. Escolhendo a 4ª função limite de cada estado temos: $x_{0[f(4)]}$ e $x_{1[f(4)]}$ e $x_{2[f(4)]}$

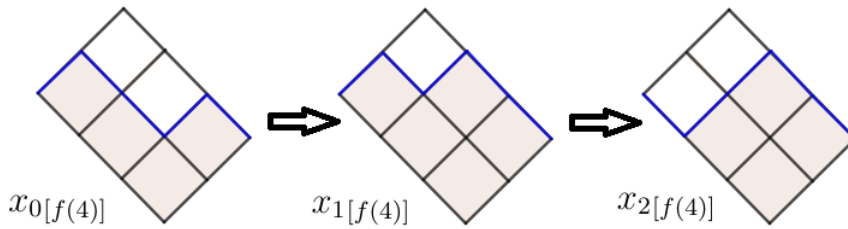


Figura 3.6: Atualização da cadeia iniciada em $x_0 = 43152$ em relação a 4ª função limite, vista em caminhos de rede.

A permutação identidade $12 \dots n$ e seu reverso $n(n-1) \dots 1$ fornecem os caminhos mínimo e máximo para qualquer função limite. A ideia é analisar essas permutações através das funções limite, já que cada função limite representa um caminho de rede. Em outras literaturas, como no capítulo 8 de [1], os embaralhamentos de cartas são analisados estudando o grupo simétrico. Nós seguiremos o método de Wilson [2] utilizando as funções limite. Dessa forma, vamos utilizar os resultados obtidos na cadeia de Markov no Caminho de rede para analisar esse tipo de embaralhamento.

Teorema 3.6. *Após $(1/\pi^2 - o(1))n^3 \log n$ embaralhamentos, a distância de variação da estacionariedade é $1 - o(1)$ e a probabilidade de coalescência é $o(1)$. Após $(6/\pi^2 + o(1))n^3 \log n$ embaralhamentos, a distância de variação da estacionariedade é $o(1)$ e a probabilidade de coalescência é $1 - o(1)$.*

Demonstração. Dada qualquer permutação, queremos que, após um certo tempo t , $x_{t[f(i)]}$ esteja perto da estacionariedade para todos os i 's entre 0 e n , ou seja, queremos que todas as funções limites estejam perto da estacionariedade. Daí, para obter só uma cota inferior, vamos só investigar a $\lfloor n/2 \rfloor$ -ésima função limite $x_{t[f[\lfloor n/2 \rfloor]}$. Pelo Teorema 3.4, após $[1/\pi^2 - o(1)]n^3 \log n$ passos, a distância de variação apenas desta função limite à estacionariedade é pelo menos $1 - o(1)$. Isso nos dá a cota inferior. Para obter o limite superior, tome as permutações $x = 123 \dots n$ e $y = n \dots 321$ e a elas associamos as $n+1$ funções de limite. Nós queremos que todas coincidam. Sejam f_i e g_i as i -ésimas funções de limite de x e y , respectivamente. Considere o evento $A_i = \{f_i \text{ e } g_i \text{ diferem depois de } [\frac{2}{\pi^2} + o(1)]n^3 \log(n^3/\delta) \text{ passos}\}$, para cada $i \in \{1, \dots, n+1\}$. Pelo Teorema 3.2, temos $P(A_i) \leq \frac{\delta}{n+1}$. Daí

$$\begin{aligned}
 P(x \text{ e } y \text{ diferem depois de } [\frac{2}{\pi^2} + o(1)]n^3 \log(n^3/\delta) \text{ passos}) &\leq P(\cup_{i=1}^{n+1} \{P(A_i)\}) \\
 &\leq (n+1)P(A_i) \\
 &\leq (n+1)\frac{\delta}{n+1} = \delta.
 \end{aligned}$$

Tomando $\delta \ll 1$, mas com $\log \delta^{-1} \ll \log n$, daí $\log n^3/\delta \approx \log n^3$, logo:

$$\left[\frac{2}{\pi^2} + o(1) \right] n^3 \log(n^3/\delta) \approx \left[\frac{6}{\pi^2} + o(1) \right] n^3 \log n.$$

□

Referências

- [1] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2008.
- [2] D. B. Wilson, *Mixing Times of Lozenge Tiling and Card Shuffling Markov Chains*, The Annals of Applied Probability 2004, Vol. 14, No. 1, 274 – 325.
- [3] T. Franco. *Princípios de Combinatória e Probabilidade*. IMPA, Coleção Matemática Universitária.
- [4] Feofiloff, Paulo. *Análise de Algoritmos*, https://www.ime.usp.br/~pf/analise_de_algoritmos/aulas/0h.html. Acesso em: 13 de outubro de 2023, às 18h.

Universidade Federal da Bahia - UFBA
Instituto de Matemática e Estatística / Colegiado da Pós-Graduação em
Matemática

Av. Adhemar de Barros, s/n, Campus Universitário de Ondina, Salvador - BA

CEP: 40170-110

<<http://www.pgmat.ufba.br>>