



Rodrigo Marinho de Souza

**Tempos de mistura para passeios aleatórios no
grupo simétrico**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Matemática do Departamento de Matemática do Centro Técnico Científico da PUC-Rio.

Orientador : Prof. Simon Richard Griffiths
Coorientadora: Prof. Ana Patricia Carvalho Gonçalves

Rio de Janeiro
Julho de 2017



Rodrigo Marinho de Souza

**Tempos de mistura para passeios aleatórios no
grupo simétrico**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Matemática do Departamento de Matemática do Centro Técnico Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Simon Richard Griffiths

Orientador

Departamento de Matemática – PUC-Rio

Prof. Ana Patricia Carvalho Gonçalves

Coorientadora

Instituto Técnico Superior – IST

Prof. Augusto Quadros Teixeira

Instituto Nacional de Matemática Pura e Aplicada – IMPA

Prof. Freddy Rolando Hernandez Romero

Universidade Federal Fluminense – UFF

Prof. Milton David Jara Valenzuela

Instituto Nacional de Matemática Pura e Aplicada – IMPA

Prof. Márcio da Silveira Carvalho

Coordenador Setorial do Centro Técnico Científico – PUC-Rio

Rio de Janeiro, 07 de Julho de 2017

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Rodrigo Marinho de Souza

Terminou sua graduação em bacharelado em Matemática na Universidade Federal Fluminense, onde recebeu o Prêmio Láurea Acadêmica da UFF de 2015 pela sua dedicação como discente alcançando o maior coeficiente de rendimento dentre os formandos do ano. Foi monitor das disciplinas Pré-Cálculo e Álgebra I durante a graduação e realizou atividades de iniciação científica na área de probabilidade, estudando Cadeias de Markov e suas aplicações, como bolsista da Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro. Além disso, é técnico em Polímeros formado pelo Instituto Federal de Educação, Ciência e Tecnologia do Rio de Janeiro.

Ficha Catalográfica

De Souza, Rodrigo Marinho

Tempos de mistura para passeios aleatórios no grupo simétrico / Rodrigo Marinho de Souza; orientador: Simon Richard Griffiths; coorientadora: Ana Patricia Carvalho Gonçalves. – 2017.

v., 96 f: il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Matemática.

Inclui bibliografia

1. Matemática – Teses. 2. Combinatória, probabilidade e processos estocásticos – Teses. 3. Cadeias de Markov;. 4. Embaralhamentos;. 5. Tempos de Mistura;. 6. Cutoffs;. 7. Processo de Exclusão Simples..
I. Griffiths, Simon Richard. II. Gonçalves, Ana Patricia Carvalho. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Matemática. IV. Título.

CDD: 620.11

Agradecimentos

Primeiramente, agradeço a Deus, não apenas pela oportunidade de fazer um mestrado, mas por todas as bênçãos que derrama sobre mim, tanto na vida acadêmica quanto na vida pessoal.

Agradeço à minha família, principalmente à minha irmã Rita Marinho pelo apoio incondicional.

Aos meus amigos da PUC-Rio, pelas brincadeiras e pela bagunça, que muitas vezes serviram de escape da rotina exaustiva de estudo. Dentre estes, sinto-me obrigado a citar Edison, Simone e Thiago, cuja amizade levarei para sempre em minha vida.

Agradeço também aos professores e demais funcionários do departamento pela simpatia e dedicação. Em particular, à Creuza Nascimento e Carlos Henrique, e aos professores Carlos Tomei, cujas aulas de álgebra linear jamais esquecerei, e Marcos Craizer, pelo incentivo e constante preocupação.

Ao meu amigo professor Freddy Hernandez pelo apoio, incentivo e acompanhamento desde a graduação. Obrigado por me apresentar esta linda área de pesquisa. Devo-te muito, e espero um dia poder retribuir.

Ofereço minha eterna gratidão aos meus orientadores. Ao Simon Griffiths pela amizade, e por sempre me atender quando precisei de ajuda. À Patricia Gonçalves por dedicar grande parte de seu tempo na melhoria constante desta dissertação e pela confiança no meu trabalho.

Ao professor Hubert Lacoïn, pelo maravilhoso artigo e por me ajudar a entendê-lo melhor.

Por último, porém não menos importante, agradeço à CAPES pelo apoio financeiro.

Resumo

De Souza, Rodrigo Marinho; Griffiths, Simon Richard; Gonçalves, Ana Patricia Carvalho. **Tempos de mistura para passeios aleatórios no grupo simétrico**. Rio de Janeiro, 2017. 96p. Dissertação de Mestrado – Departamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro.

O objetivo desta dissertação é apresentar algumas técnicas e ferramentas para a obtenção de cotas superiores e inferiores para tempos de mistura de cadeias de Markov. Para que isso se torne mais interessante, apresentaremos estes conceitos através de cadeias de Markov que atuam sobre o grupo simétrico, que podem ser vistas como embaralhamentos de cartas. Ademais, usaremos um destes embaralhamentos como toy model para o processo de exclusão simples simétrico, o que nos ajudará a determinar os tempos de mistura do embaralhamento e do famoso sistema de partículas.

Palavras-chave

Cadeias de Markov; Embaralhamentos; Tempos de Mistura; Cut-offs; Processo de Exclusão Simples.

Abstract

De Souza, Rodrigo Marinho; Griffiths, Simon Richard (Advisor); Gonçalves, Ana Patricia Carvalho (Co-Advisor). **Mixing times for random walks on the symmetric group**. Rio de Janeiro, 2017. 96p. Dissertação de Mestrado – Departamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro.

The aim of this dissertation is to introduce some techniques and tools to obtain upper and lower bounds for Markov chains mixing times. To make it more interesting, we introduce these concepts through Markov chains that act on the symmetric group, which can be seen as card shuffles. Furthermore, we use one of these shuffles as a toy model for the symmetric simple exclusion process, which helps us to determine mixing times for the shuffle and for the famous particle system.

Keywords

Markov Chains; Shuffles; Mixing Times; Cutoffs; Simple Exclusion Process.

Sumário

1	Introdução	11
2	Tempos de mistura e distância do equilíbrio	13
2.1	Distância de variação total	13
2.2	Tempos de mistura	14
2.3	Distância de separação	15
2.4	Tempos estacionários fortes	16
2.5	Cota por contagem	18
2.6	O cutoff	19
3	Embaralhamento Top-to-random	22
3.1	Introdução ao modelo	22
3.2	Cota superior	24
3.3	Cota inferior	26
3.4	Cutoff	27
4	O embaralhamento Gilbert-Shannon-Reeds (GSR)	29
4.1	Uma introdução ao modelo	29
4.2	Uma primeira cota superior	32
4.3	Uma primeira cota inferior	33
4.4	Generalizando o modelo: O a-embaralhamento	33
4.5	Encontrando o tempo de mistura	37
5	O embaralhamento de transposições adjacentes visto em tempo contínuo	43
5.1	O embaralhamento de transposições adjacentes	43
5.2	Tirando vantagem da monotonicidade	44
5.2.1	Mapeando permutações em superfícies discretas	45
5.2.2	A construção gráfica	46
5.2.3	Ordenação estocástica e sua preservação	48
5.2.4	Desigualdades de correlação e a desigualdade FKG	50
5.2.5	A desigualdade censuradora	52
5.2.6	Projeção e monotonia	57
5.3	Mais algumas ferramentas	62
5.3.1	A conexão com a equação do calor	62
5.3.2	A cota superior de Wilson	65
5.3.3	Apagando os marcadores e decompondo o procedimento de mistura	65
5.4	A cota superior	67
5.5	A cota inferior	74
5.6	O cutoff	74
6	O embaralhamento de transposições adjacentes visto em tempo discreto	75
6.1	A cota superior	78
6.2	A cota inferior	79
6.3	O cutoff	79

7	Uma maravilhosa aplicação: Encontrando o tempo de mistura para o processo de exclusão simples simétrico	80
7.1	O processo de exclusão simples simétrico	80
7.2	Conectando a exclusão simples e o embaralhamento de transposições adjacentes	82
7.3	A cota superior	83
7.4	A cota inferior	83
7.5	O cutoff	84
	Referências Bibliográficas	85
A	Cadeias de Markov em espaços finitos	87
A.1	Cadeias de Markov em tempo discreto	87
A.2	Cadeias de Markov em tempo contínuo	90
B	Sistemas dinâmicos aleatórios	91
C	Passeios aleatórios em Grupos	93
D	Tempos de mistura analisados	96

Lista de figuras

3.1	Embaralhamento top-to-random em um baralho com 7 cartas.	24
4.1	Vizualização de um embaralhamento cascata.	29
5.1	Exemplo de uma transposição adjacente.	43
7.1	Transição de uma partícula para um sítio vizinho.	81
7.2	Permutação e associada à configuração inicial das partículas na construção gráfica.	82

Lista de tabelas

- D.1 Tabela com todas as cadeias de Markov analisadas e seus respectivos tempos de mistura. 96

1

Introdução

Um passeio aleatório em um grupo é uma cadeia de Markov cujo espaço de estados é o conjunto de elementos desse grupo, e a cada passo, a probabilidade de transição de um elemento para outro é dada por um incremento que caracteriza tal passeio aleatório. Mais especificamente, esta cadeia move-se multiplicando o estado atual, à esquerda, por um elemento aleatório do grupo escolhido de acordo com o incremento.

Um embaralhamento de cartas pode ser visto como um passeio aleatório no grupo simétrico S_n (grupo das permutações). O nosso interesse é saber quantas vezes precisamos embaralhar um baralho para que ele esteja bem embaralhado. Mas o que é estar bem embaralhado?

Um baralho tradicional possui 52 cartas, o que nos dá $52!$ permutações possíveis para ele, e isso é da ordem de 10^{68} que é maior que o número de partículas no sistema solar. Diremos que um baralho está bem embaralhado quando todas essas possíveis permutações forem equiprováveis, o que pode nos intimidar à saber quantas embaralhadas são necessárias e suficientes para chegarmos à uniformidade, o que chamaremos mais à frente de tempo de mistura.

Uma outra questão que podemos nos fazer é: Qual é a melhor maneira de embaralhar cartas? Esse é um problema de muito valor para cassinos e apostadores. Nesta dissertação, analisaremos três embaralhamentos. O primeiro, chamado de embaralhamento Top-to-random, consiste em pegar a carta do topo do baralho e colocá-la de volta ao baralho em qualquer posição escolhida uniformemente ao acaso. Mostraremos que o tempo de mistura para este tipo de embaralhamento é da ordem de $n \log n$, resultado obtido por Aldous e Diaconis em [1].

Em seguida, analisaremos o embaralhamento Gilbert-Shannon-Reeds (riffle shuffle), para o qual Bayer e Diaconis mostraram, em [3], que sete embaralhadas do tipo certo são necessárias e suficientes para embaralhar

bem um baralho padrão com 52 cartas. Este resultado ficou conhecido como teorema das sete embaralhadas. Por último, porém não menos importante, estudaremos o embaralhamento das transposições adjacentes, que Lacoïn em [11], usou como um toy model para o processo de exclusão simples encontrando tempos de mistura tanto para o embaralhamento, quanto para o sistema de partículas.

Ademais, mostraremos que alguns destes passeios aleatórios em S_n apresentam um fenômeno muito interessante conhecido como cutoff, que ocorre quando a distância de variação total, que nos fornecerá a distância do equilíbrio, cai abruptamente de 1 para 0 no tempo de mistura.

É aconselhável que o leitor tenha um prévio conhecimento de álgebra linear, bem como conceitos básicos de probabilidade. E para uma leitura mais prazerosa, recomenda-se o primeiro capítulo de [12].

2

Tempos de mistura e distância do equilíbrio

Neste capítulo, introduziremos algumas definições e técnicas para podermos determinar o tempo que uma cadeia de Markov leva para alcançar o equilíbrio.

2.1

Distância de variação total

Definição 2.1 Definimos a *Distância de Variação Total* entre duas distribuições de probabilidade μ e ν em um espaço de estados Ω como

$$d_{TV}(\mu, \nu) = \|\mu - \nu\|_{TV} := \max_{A \subset \Omega} |\mu(A) - \nu(A)|.$$

Proposição 2.2 Também podemos definir a *Distância de Variação Total* como

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

Prova. De fato, sejam $B = \{x : \mu(x) \geq \nu(x)\}$ e $A \subset \Omega$ um evento qualquer. Então, como $x \in (A \cap B^c) \Rightarrow \mu(x) < \nu(x)$, temos:

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

Do mesmo modo, temos:

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

Assim, podemos ver que os lados direitos das desigualdades acima são iguais, pois quando fazemos a diferença entre eles, obtemos zero. Daí, se $A = B$ (ou $A = B^c$), então $|\mu(A) - \nu(A)| = \nu(B^c) - \mu(B^c)$. Portanto,

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

□

Definição 2.3 Se $\nu^{(1)}, \nu^{(2)}, \dots$ e ν são distribuições de probabilidade em Ω , então dizemos que $\nu^{(t)}$ converge para ν em variação total quando $t \rightarrow \infty$, escrevendo $\nu^{(t)} \xrightarrow{TV} \nu$, se

$$\lim_{t \rightarrow \infty} d_{TV}(\nu^{(t)}, \nu) = 0.$$

2.2

Tempos de mistura

Sabemos que uma cadeia de Markov irredutível e aperiódica com espaço de estados finito Ω e matriz de transição P possui uma única distribuição estacionária π e que dada uma distribuição inicial arbitrária, essa converge para π em variação total (veja o Teorema A.7). Estamos interessados no tempo necessário para que a cadeia atinja o equilíbrio. Dessa forma, definamos a distância entre a cadeia de Markov no tempo t e a distribuição estacionária π por

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

onde $P^t(x, y)$ é a probabilidade de alcançar o estado y em t passos, saindo do estado x .

Definição 2.4 Definimos o tempo de mistura para uma cadeia de Markov por

$$t_{mix}(\epsilon) := \inf \{t : d(t) \leq \epsilon\}$$

e

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

Proposição 2.5 Seja μ uma distribuição de probabilidade em Ω e P uma matriz de transição. Então, $d(t) = \sup_{\mu} \|\mu P^t - \pi\|_{TV}$.

Prova. Seja μ uma distribuição de probabilidade em Ω . Pela Proposição 2.2 e pela desigualdade triangular, temos

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

Como μ foi tomada arbitrariamente,

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

A desigualdade oposta é óbvia. \square

2.3

Distância de separação

É fácil ver que a distância de variação total é uma métrica. Uma outra distância que, apesar de não ser uma métrica, é muito útil, é a distância de separação.

Definição 2.6 Definimos a distância de separação, para cada $x \in \Omega$, por

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

E também definimos

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

Do mesmo modo, o tempo de separação pode ser definido por

$$t_{sep}(\epsilon) := \inf \{t : s(t) \leq \epsilon\}$$

Proposição 2.7 A distância de separação $s_x(t)$ satisfaz $\|P^t(x, \cdot) - \pi\|_{TV} \leq s_x(t)$, e portanto $d(t) \leq s(t)$.

Prova. A demonstração que segue encontra-se na página 80 de [12].

De fato,

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

\square

Nas próximas duas seções, introduziremos algumas ferramentas que utilizaremos mais à frente para encontrar cotas para os tempos de mistura.

2.4

Tempos estacionários fortes

Os tempos estacionários fortes foram introduzidos pela primeira vez por Aldous e Diaconis em [2] (veja também [1]), e são uma ferramenta muito simples e útil para a obtenção de cotas superiores para tempos de mistura. As definições e resultados desta seção foram extraídos de [12].

Definição 2.8 *Dada uma sequência $(X_t)_{t \geq 0}$, de variáveis aleatórias em Ω , dizemos que uma variável aleatória τ em $\{0, 1, 2, \dots, \infty\}$ é um tempo de parada para (X_t) se, para cada $t \geq 0$, 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) .

Toda matriz de transição P tem um sistema dinâmico aleatório associado, ou seja, podemos encontrar uma sequência $(Z_t)_{t \geq 1}$, de variáveis aleatórias i.i.d. e uma função $f : \Omega \times \Lambda \rightarrow \Omega$ tal que a sequência $(X_t)_{t \geq 0}$, definida por

$$X_0 = x, \text{ e } X_t = f(X_{t-1}, Z_t)$$

seja uma cadeia de Markov com matriz de transição P iniciada em x (veja apêndice B).

Definição 2.9 *Dizemos que um tempo aleatório τ é um tempo de parada randomizado para a cadeia de Markov $(X_t)_{t \geq 0}$ se é um tempo de parada para a sequência $(Z_t)_{t \geq 1}$.*

Definição 2.10 *Seja $(X_t)_{t \geq 0}$ uma cadeia de Markov irredutível com distribuição estacionária π . Um tempo estacionário τ para (X_t) é um tempo de parada randomizado, possivelmente dependendo da posição inicial x , tal que a distribuição de X_τ é π , isto é,*

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

Definição 2.11 *Um tempo estacionário forte para uma cadeia de Markov $(X_t)_{t \geq 0}$ com distribuição estacionária π é um tempo de parada randomizado τ , possivelmente dependendo da posição inicial x , tal que*

$$P_x \{\tau = t, X_\tau = y\} = P_x \{\tau = t\} \pi(y). \quad (2.1)$$

Em palavras, X_τ tem distribuição π e é independente de τ .

Lema 2.12 *Seja $(X_t)_{t \geq 0}$ uma cadeia de Markov irredutível com distribuição estacionária π . Se τ é um tempo estacionário forte para $(X_t)_{t \geq 0}$, então para todo $t \geq 0$,*

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

Prova. Seja Z_1, Z_2, \dots a sequência i.i.d. usada no sistema dinâmico aleatório associado a (X_t) . Para qualquer $s \leq t$,

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

Como τ é um tempo de parada para $(Z_t)_{t \geq 0}$, o evento $\{\tau = s\}$ é equivalente a $\{(Z_1, \dots, Z_s) \in B\}$ para algum conjunto $B \subset \Lambda^s$. Além disso, para inteiros $r, s \geq 0$, existe uma função $\tilde{f}_r : \Omega \times \Lambda^r \rightarrow \Omega$ 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_t) são independentes,

$$\begin{aligned} P_x \{X_t = y | \tau = s, X_s = z\} &= P_x \{\tilde{f}_{t-s}(z, Z_{s+1}, \dots, Z_t) = y | (Z_1, \dots, Z_s) \in B, X_s = z\} \\ &= P^{t-s}(z, y). \end{aligned}$$

Portanto,

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

Como π satisfaz $\pi = \pi P^{t-s}$,

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

Agora, somando em $s \leq t$, temos o resultado. □

Proposição 2.13 *Se τ é um tempo estacionário forte, então*

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

Prova. Sejam $x, y \in \Omega$. Então,

$$\begin{aligned} 1 - \frac{P^t(x, y)}{\pi(y)} &= 1 - \frac{P_x \{X_t = y\}}{\pi(y)} \leq 1 - \frac{P_x \{X_t = y, \tau \leq t\}}{\pi(y)} \\ &= 1 - \frac{\pi(y) P_x \{\tau \leq t\}}{\pi(y)} = P_x \{\tau > t\}. \end{aligned}$$

que garante o resultado. \square

Teorema 2.14 *Se τ é um tempo estacionário forte, então*

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

Prova. Sabemos que $d(t) \leq s(t) = \max_{x \in \Omega} s_x(t)$. Então, usando a Proposição 2.13, temos que

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

\square

2.5

Cota por contagem

Definição 2.15 *Seja $(X_t)_{t \geq 0}$ uma cadeia de Markov com matriz de transição P , irreduzível e aperiódica em um espaço de estados Ω , cuja distribuição estacionária π é a uniforme. Definamos $d_{out}(x) := |\{y \in \Omega : P(x, y) > 0\}|^a$ como o número de estados acessíveis em um passo saindo de x , e seja*

$$\Delta := \max_{x \in \Omega} d_{out}(x)$$

o número máximo de estados acessíveis em um passo.

Proposição 2.16 *Se $\Delta^t < (1 - \epsilon)|\Omega|$, então*

$$t_{mix}(\epsilon) \geq \frac{\log(|\Omega|(1 - \epsilon))}{\log \Delta}.$$

Prova. De fato, seja Ω_t^x o conjunto de todos os estados acessíveis em t passos saindo de x . Observemos que $|\Omega_t^x| \leq \Delta^t$.

Como $\Delta^t < (1 - \epsilon)|\Omega|$, temos

$$\|P^t(x, \cdot) - \pi\|_{TV} \geq P^t(x, \Omega_t^x) - \pi(\Omega_t^x) \geq 1 - \frac{\Delta^t}{|\Omega|} > \epsilon.$$

Então,

$$t_{mix}(\epsilon) \geq \frac{\log(|\Omega|(1 - \epsilon))}{\log \Delta}.$$

Em particular,

$$t_{mix} \geq \frac{\log\left(\frac{3}{4}|\Omega|\right)}{\log \Delta}.$$

\square

^aUsamos a notação $|A|$ para denotar a cardinalidade de um conjunto A .

2.6

O cutoff

O cutoff é um fenômeno que ocorre em algumas cadeias de Markov quando a distância de variação total (ou de separação) cai abruptamente de 1 para 0 ao redor do tempo de mistura.

Definição 2.17 Uma sequência de cadeias de Markov tem cutoff se, para todo $0 < \epsilon < 1$,

$$\lim_{n \rightarrow \infty} \frac{t_{mix}^{(n)}(\epsilon)}{t_{mix}^{(n)}(1 - \epsilon)} = 1. \quad (2.2)$$

Teorema 2.18 Sejam $t_{mix}^{(n)}$ e d_n o tempo de mistura e a distância da estacionariedade para a n -ésima cadeia numa sequência de cadeias Markov. A sequência tem cutoff se, e somente se,

$$\lim_{n \rightarrow \infty} d_n(ct_{mix}^{(n)}) = \begin{cases} 1, & \text{se } c < 1; \\ 0, & \text{se } c > 1. \end{cases} \quad (2.3)$$

Prova. Suponhamos que vale (2.3). Seja $0 < \gamma < 1$. Como $\lim_{n \rightarrow \infty} d_n((1 - \gamma)t_{mix}^{(n)}) = 1$, para cada $0 < \epsilon < 1$, existe $n_1 \in \mathbb{N}$ tal que $n \geq n_1$ implica $|d_n((1 - \gamma)t_{mix}^{(n)}) - 1| < \epsilon$. Então, se $n \geq n_1$, $-\epsilon < d_n((1 - \gamma)t_{mix}^{(n)}) - 1$, ou seja,

$$d_n((1 - \gamma)t_{mix}^{(n)}) > 1 - \epsilon.$$

Logo,

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

Analogamente, como $\lim_{n \rightarrow \infty} d_n((1 + \gamma)t_{mix}^{(n)}) = 0$, para cada $0 < \epsilon < 1$, existe $n_2 \in \mathbb{N}$ tal que $n \geq n_2$ implica $|d_n((1 + \gamma)t_{mix}^{(n)})| < \epsilon$. Logo, se $n \geq n_2$, temos

$$d_n((1 + \gamma)t_{mix}^{(n)}) < \epsilon,$$

implicando que $(1 + \gamma)t_{mix}^{(n)} \geq t_{mix}^{(n)}(\epsilon)$. Tomemos $N = \max\{n_1, n_2\}$ e seja $n > N$. Então,

$$\frac{t_{mix}^{(n)}(\epsilon)}{(1 + \gamma)} \leq t_{mix}^{(n)} \leq \frac{t_{mix}^{(n)}(1 - \epsilon)}{1 - \gamma}.$$

Logo, para $0 < \epsilon < \frac{1}{2}$, temos

$$1 \leq \frac{t_{mix}^{(n)}(\epsilon)}{t_{mix}^{(n)}(1 - \epsilon)} \leq \frac{(1 + \gamma)}{1 - \gamma},$$

pois se $0 < \epsilon < \frac{1}{2}$, temos que $t_{mix}^{(n)}(\epsilon) \geq t_{mix}^{(n)}(1 - \epsilon)$.

Agora, fazendo $\gamma \rightarrow 0$ e $n \rightarrow \infty$, temos

$$\lim_{n \rightarrow \infty} \frac{t_{mix}^{(n)}(\epsilon)}{t_{mix}^{(n)}(1 - \epsilon)} = 1, \text{ para todo } 0 < \epsilon < \frac{1}{2}. \quad (2.4)$$

Note que se (2.4) vale, então

$$\lim_{n \rightarrow \infty} \frac{t_{mix}^{(n)}(1 - \epsilon)}{t_{mix}^{(n)}(\epsilon)} = 1. \quad (2.5)$$

Logo, (2.4) também vale para todo $1 > \epsilon > \frac{1}{2}$. Portanto, a sequência tem cutoff.

Reciprocamente, suponhamos que a sequência tenha cutoff. Então, como

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

dado $0 < \gamma < 1$, existe $n_0 \in \mathbb{N}$ tal que se $n \geq n_0$, então

$$\frac{t_{mix}^{(n)}(\epsilon)}{t_{mix}^{(n)}(1 - \epsilon)} \leq 1 + \gamma =: c.$$

Assim, como $c > 1$, $t_{mix}^{(n)}(\epsilon) \leq ct_{mix}^{(n)}(1 - \epsilon)$. Tomando $\epsilon \in (0, 1/2)$, temos $t_{mix}^{(n)}(\epsilon) \leq ct_{mix}^{(n)}(1 - \epsilon) \leq ct_{mix}^{(n)}$. Agora, pela definição de d_n ,

$$d_n(ct_{mix}^{(n)}) \leq d_n(t_{mix}^{(n)}(\epsilon)) = \epsilon,$$

para todo $\epsilon \in (0, 1/2)$. Tomando $\epsilon = \frac{1}{4n}$, para n suficientemente grande, temos que $\epsilon \in (0, 1/2)$ e

$$\lim_{n \rightarrow \infty} d_n(ct_{mix}^{(n)}) \leq \lim_{n \rightarrow \infty} \frac{1}{4n} = 0.$$

Do mesmo modo, como

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

dado $0 < \gamma < 1$, existe $n_0 \in \mathbb{N}$ tal que se $n \geq n_0$, então

$$c := (1 - \gamma) \leq \frac{t_{mix}^{(n)}(\epsilon)}{t_{mix}^{(n)}(1 - \epsilon)}.$$

Assim, como $c < 1$, $ct_{mix}^{(n)}(1 - \epsilon) \leq t_{mix}^{(n)}(\epsilon)$. Tomando $\epsilon = \left(1 - \frac{1}{4n}\right)$, para n suficientemente grande, temos que $ct_{mix}^{(n)}\left(\frac{1}{4n}\right) \leq t_{mix}^{(n)}\left(1 - \frac{1}{4n}\right)$.

Daí,

$$ct_{mix}^{(n)} \leq ct_{mix}^{(n)}\left(\frac{1}{4n}\right) \leq t_{mix}^{(n)}\left(1 - \frac{1}{4n}\right).$$

Logo,

$$d_n(ct_{mix}^{(n)}) \geq \left(1 - \frac{1}{4n}\right).$$

Portanto,

$$\lim_{n \rightarrow \infty} d_n(ct_{mix}^{(n)}) \geq \lim_{n \rightarrow \infty} \left(1 - \frac{1}{4n}\right) = 1.$$

□

3

Embaralhamento Top-to-random

3.1

Introdução ao modelo

Começaremos introduzindo um assunto que, *a priori*, não parece ter relação com embaralhamentos, mas veremos que, no final, ele nos ajudará a entender melhor o problema.

Suponha que uma empresa produz n figurinhas diferentes. Um colecionador deseja um álbum completo. Suponhamos que cada figurinha que ele adquire possui probabilidade igual de ser cada um dos n tipos. Assim nos perguntamos: Quantas figurinhas ele deve obter para completar o seu álbum sem trocar figurinhas com outras pessoas?

Seja X_t o número de diferentes tipos entre as t primeiras figurinhas do colecionador. Claramente $X_0 = 0$. Quando o colecionador tem k tipos diferentes de figurinhas, tem-se $n - k$ tipos faltando. Portanto,

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

e

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

Podemos perceber que toda trajetória desta cadeia é não-decrescente e, uma vez que a cadeia chega no estado n (correspondendo ao álbum completo), ela é absorvida lá.

Estamos interessados no número de passos requeridos para alcançar o estado absorvente.

Proposição 3.1 *Considere um colecionador que deseja obter um álbum completo de figurinhas. Assuma que cada nova figurinha é escolhida uniforme e independentemente do conjunto de n tipos possíveis, e seja τ o número (aleatório) de*

figurinhas colecionadas quando o álbum finalmente contém todos os tipos. Então

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

Prova. A demonstração que segue encontra-se na página 22 de [12].

A esperança $\mathbb{E}(\tau)$ pode ser calculada escrevendo τ como a soma de variáveis aleatórias geométricas. Seja τ_k o número total de figurinhas acumuladas quando o álbum finalmente contém k figurinhas distintas. Então

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

Assim, $\tau_k - \tau_{k-1}$ é uma variável aleatória geométrica com probabilidade de sucesso $(n - k + 1)/n$, uma vez que depois de colecionar τ_{k-1} figurinhas, tem-se $n - k + 1$ tipos faltando para a coleção. E cada figurinha subsequentemente obtida tem a mesma probabilidade $(n - k + 1)/n$ de ser de um tipo ainda não encontrado, até que um novo tipo seja finalmente obtido. Assim, $\mathbb{E}(\tau_k - \tau_{k-1}) = n/(n - k + 1)$ e

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

□

Proposição 3.2 A variável aleatória τ definida acima satisfaz $P(\tau > \lceil n \log n + cn \rceil) \leq e^{-c}$ para cada $c > 0$.

Prova. A demonstração que segue encontra-se na página 23 de [12].

Seja A_i o evento que o i -ésimo tipo não aparece entre as primeiras $\lceil n \log n + cn \rceil$ figurinhas obtidas. Primeiro, observe que

$$P(\tau > \lceil n \log n + cn \rceil) = P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i). \quad (3.1)$$

Como cada ensaio tem probabilidade $1 - n^{-1}$ de não tirar a figurinha i e os ensaios são independentes, o lado direito da desigualdade (3.1) fica limitado superiormente por

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

provando a proposição. □

^aUsamos \approx para denotar uma aproximação assintótica.

Agora, consideremos o seguinte método para embaralhar um baralho com n cartas: Pegue a carta do topo e coloque-a aleatoriamente e uniformemente no baralho, como na Figura 3.1.

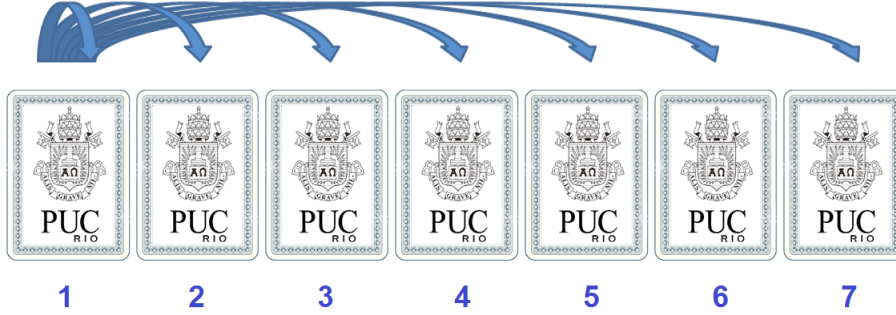


Figura 3.1: Embaralhamento top-to-random em um baralho com 7 cartas.

Esse processo, conhecido como embaralhamento top-to-random, irá misturar o baralho. As possíveis transições nas permutações do baralho são um passeio aleatório no grupo S_n , irreduzível e aperiódico, que sabemos que possui uma distribuição estacionária e esta é a uniforme (veja proposições A.7 e C.2).

A pergunta agora é a seguinte: Quantas vezes devemos embaralhar este baralho usando este método para que o mesmo esteja bem embaralhado?

3.2

Cota superior

Definamos como τ_{top} o tempo após um movimento depois que a carta que estava originalmente no fundo do baralho é movida para o topo deste. Mostraremos que τ_{top} é um tempo estacionário forte.

Proposição 3.3 *Seja $(X_t)_{t \geq 0}$ o passeio aleatório em S_n correspondente ao embaralhamento top-to-random com n cartas. Então, dado que em um tempo t existem k cartas embaixo da que estava originalmente no fundo, cada uma das $k!$ possíveis ordenações dessas cartas estão igualmente distribuídas. Portanto, pela definição de τ_{top} , a distribuição de $X_{\tau_{top}}$ é uniforme em S_n e o tempo τ_{top} é independente de $X_{\tau_{top}}$.*

Prova. A demonstração que segue encontra-se na página 76 de [12].

Demonstraremos esta proposição por indução. Primeiro, quando $t = 0$, não há cartas embaixo da carta originalmente no fundo, e portanto, a afirmação é verdadeira.

Agora, suponhamos que a afirmação seja verdadeira para um tempo t . Queremos mostrar que também é para o tempo $(t + 1)$.

Com efeito, existem duas possibilidades para $t + 1$: Uma carta é colocada embaixo da carta originalmente no fundo ou não. No segundo caso, as cartas embaixo da originalmente no fundo permanecem em arranjo uniforme. No primeiro caso, como uma carta foi colocada aleatoriamente embaixo da que estava originalmente no fundo, as $(k + 1)$ possíveis localizações para a carta são igualmente distribuídas, e então, cada um dos $(k + 1)!$ arranjos são equiprováveis.

Além disso, em $\tau_{top} - 1$ as cartas abaixo da inicialmente no fundo estão uniformemente distribuídas. Portanto, a ordem das cartas em τ_{top} é uniforme em todos os arranjos possíveis e τ_{top} é independente de $X_{\tau_{top}}$.

□

Teorema 3.4 Para o embaralhamento Top-to-Random, $t_{mix} \leq n \log n + O(n)^b$.

Prova. Notemos que quando há k cartas embaixo da que estava originalmente no fundo, a chance desse número aumentar em uma carta é k/n até que uma embaralhada coloque a carta do topo embaixo dela. Então, a distribuição de τ_{top} é a mesma do tempo do colecionador de figurinhas.

Pela Proposição 3.3, vemos que τ_{top} é um tempo estacionário forte e pela Proposição 3.2, temos que para todo $c > 0$,

$$P(\tau_{top} > \lceil n \log n + cn \rceil) \leq e^{-c}.$$

Portanto, pelo Teorema 2.14, temos que

$$d(n \log n + cn) \leq e^{-c}.$$

Consequentemente,

$$t_{mix}(\epsilon) \leq n \log n + \log \left(\frac{1}{\epsilon} \right) n.$$

□

^bDizemos que $f(n) = O(g(n))$ se $\frac{f(n)}{g(n)} \leq C$ quando $n \rightarrow \infty$.

3.3

Cota inferior

Teorema 3.5 *Seja (X_t) a cadeia Top-to-Random em n cartas. Para qualquer $\epsilon > 0$, existe uma constante α_0 tal que $\alpha > \alpha_0$ implica que para n suficientemente grande,*

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

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

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

Prova. A demonstração que segue encontra-se na página 96 de [12].

Seja A_j o evento em que as j cartas originalmente no fundo estão em sua ordem relativa original. Seja τ_j o tempo necessário para a carta inicialmente na posição j , a contar do fundo, alcançar o topo. Então,

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

onde $\tau_{j,i}$ é o tempo que a carta inicialmente na posição j , a contar do fundo, vai da posição i , a contar do fundo, para a posição $i + 1$.

As variáveis $\tau_{j,i}$, $i = j, \dots, n - 1$, são independentes e $\tau_{j,i}$ tem distribuição geométrica com parametro $p = i/n$. Portanto,

$$\mathbb{E}(\tau_{j,i}) = \frac{n}{i} \text{ e } Var(\tau_{j,i}) < \frac{n^2}{i^2},$$

onde $Var(X)$ é a variância da variável aleatória X . Assim,

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

e

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

Agora, usando a desigualdade de Chebyshev, temos

$$\begin{aligned} P\{\tau_j < n \log n - \alpha n\} &\leq P\{\tau_j - \mathbb{E}(\tau_j) < -n(\alpha - \log j - 1)\} \\ &\leq \frac{n^2}{(j-1)n^2(\alpha - \log j - 1)^2} \leq \frac{1}{j-1}, \end{aligned}$$

desde que $\alpha \geq \log j + 2$.

Definamos $t_n(\alpha) = n \log n - \alpha n$. Se $\tau_j \geq t_n(\alpha)$, então as j cartas originalmente contadas a partir do fundo do baralho estão em sua ordem relativa original em tempo $t_n(\alpha)$.

Logo,

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

para $\alpha \geq \log j + 2$.

Por outro lado, para a distribuição estacionária uniforme,

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

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

$$d_n(t_n(\alpha)) \geq \|P^{t_n(\alpha)}(e, \cdot) - \pi\|_{TV} \geq P^{t_n(\alpha)}(e, A_j) - \pi(A_j) > 1 - \frac{2}{j-1}.$$

Tomando $j = e^{\alpha-2}$, e $n \geq e^{\alpha-2}$, temos

$$d_n(t_n(\alpha)) > g(\alpha) := 1 - \frac{2}{e^{\alpha-2} - 1}.$$

Portanto,

$$\liminf d_n(t_n(\alpha)) \geq g(\alpha),$$

onde $g(\alpha) \rightarrow 1$ quando $\alpha \rightarrow \infty$.

□

3.4

Cutoff

Teorema 3.6 *O embaralhamento Top-to-Random tem cutoff.*

Prova. De fato, para o embaralhamento Top-to-Random em um baralho com n cartas, obtemos as cotas

$$d_n(n \log n + \alpha n) \leq e^{-\alpha}$$

e

$$\liminf d_n(n \log n - \alpha n) \geq 1 - 2e^{2-\alpha}.$$

Em particular, a cota superior tende a 0 quando $\alpha \rightarrow \infty$, e a cota inferior tende a 1 quando $\alpha \rightarrow \infty$. Portanto, quando $n \rightarrow \infty$, a função d_n se aproxima da função

degrau

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

Portanto, pelo Teorema 2.18, o embaralhamento *Top-to-Random* tem cutoff. \square

4

O embaralhamento Gilbert-Shannon-Reeds (GSR)

4.1

Uma introdução ao modelo

Neste capítulo, vamos analisar o método mais usado, no mundo, para embaralhar cartas. O método baseia-se em cortar o baralho aproximadamente no meio e intercalar as cartas das duas metades como uma cascata. Por esta razão, este embaralhamento é conhecido como riffle shuffle. Veja a Figura 4.1.

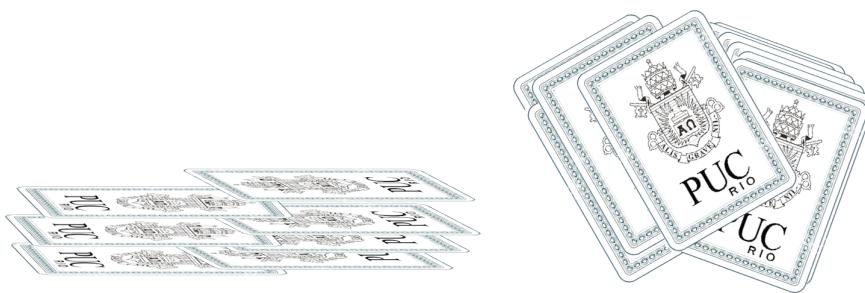


Figura 4.1: Vizualização de um embaralhamento cascata.

Definição 4.1 *Uma sequência levantadora de uma permutação σ é uma sequência maximal de valores consecutivos que ocorre na ordem relativa correta em σ .*

Por exemplo, se num baralho com 10 cartas temos $e = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$ e $\sigma = 7, 1, 8, 2, 3, 9, 4, 10, 5, 6$, então σ tem duas sequências levantadoras $1, 2, 3, 4, 5, 6$ e $7, 8, 9, 10$.

O conceito de sequência levantadora foi definido através da análise de um truque de mágica chamado *Premo*. Não estudaremos o truque, mas este é muito bem explicado em [3].

Definição 4.2 *Um riffle shuffle é uma permutação com uma ou duas sequências levantadoras, tendo apenas uma quando a permutação obtida for a identidade.*

Agora analisaremos o modelo sugerido por Gilbert e Shannon (1955), e Reeds (1981), para um riffle shuffle aleatório (veja [3]). Este modelo é muito similar aos

embaralhamentos feitos por jogadores amadores.

Abaixo listamos três métodos para embaralhar um baralho com n cartas.

1. Seja M uma variável aleatória binomial $(n, 1/2)$. Divida o baralho em duas pilhas, sendo uma delas as M cartas do topo. Existem $\binom{n}{M}$ formas de juntar as duas pilhas com um riffle shuffle preservando a ordem relativa de cada pilha. Escolha um desses arranjos uniformemente ao acaso.
2. Seja M uma variável aleatória binomial $(n, 1/2)$. Divida o baralho em duas pilhas, sendo uma delas as M cartas do topo. Segure as duas pilhas sobre uma mesa e solte as cartas, uma por uma, formando uma única pilha, procedendo da seguinte maneira: Se em um momento particular a pilha da esquerda contém a cartas e a pilha da direita contém b cartas, então deixe cair uma carta da pilha da esquerda com probabilidade $a/(a+b)$ e uma carta da pilha da direita com probabilidade $b/(a+b)$. Repita o procedimento até que todas as cartas tenham caído.
3. Marque as n cartas com n marcadores "0"s ou "1"s escolhidos uniforme e independentemente, de acordo com uma moeda honesta. Uma sequência de cartas marcadas forma uma frase de "0"s e "1"s. Coloque todas as cartas marcadas com 0 no topo do baralho, preservando sua ordem relativa.

Chamaremos por X_t a trajetória da cadeia de Markov gerada pelo embaralhamento GSR.

Lema 4.3 *Os métodos 1 e 2 geram a mesma distribuição Q nas permutações, onde*

$$Q(\sigma) = \begin{cases} (n+1)/2^n, & \text{se } \sigma = e, \\ 1/2^n, & \text{se } \sigma \text{ tem exatamente duas sequências levantadoras,} \\ 0, & \text{caso contrário.} \end{cases}$$

Prova. Vejamos primeiro o método 1.

Se σ tiver duas sequência levantadoras $1, 2, \dots, k$ e $(k+1), (k+2), \dots, n$, então é claro, pois o evento $M = k$ acontece com probabilidade $\binom{n}{k}/2^n$ e somente uma das $\binom{n}{k}$ possíveis permutações é σ . Assim, se $\mathbb{P}_e(\cdot) = \mathbb{P}(\cdot | X_0 = e)$, temos

$$Q(\sigma) = \mathbb{P}_e(X_1 = \sigma, M = k) = \mathbb{P}_e(X_1 = \sigma | M = k) \mathbb{P}(M = k) = \frac{1}{2^n}.$$

No caso de $\sigma = e$ o valor de M pode ser qualquer um em entre 0 e n , por isso obtém-se $n + 1$ vezes o valor calculado acima.

Agora analisaremos o método 2.

Dado o evento $\{M = k\}$, que tem probabilidade $\binom{n}{k} \frac{1}{2^n}$, mostraremos que a probabilidade de observar um certo intercalamento σ de $1, 2, \dots, k$ e $(k+1), \dots, n$ é igual a

$$\frac{k!(n-k)!}{n!} = \frac{1}{\binom{n}{k}}.$$

Com efeito, suponhamos que $\sigma = e$. Então,

$$P(X_1 = \sigma | M = k) = \left(\frac{k}{n} \frac{k-1}{n-1} \cdots \frac{1}{n-k+1} \right) \left(\frac{n-k}{n-k} \frac{n-k-1}{n-k-1} \cdots \frac{1}{1} \right) = \frac{1}{\binom{n}{k}}.$$

Do mesmo modo, a probabilidade de gerar qualquer outra permutação σ com sequências levantadoras $1, 2, \dots, k$ e $(k+1), \dots, n$ é igual, pois apenas a ordem dos termos no numerador muda.

Portanto, o método 2 gera Q .

□

Proposição 4.4 *Seja \hat{Q} a distribuição inversa de Q (veja a Definição C.7). Então o método 3 gera \hat{Q} , isto é, o método 3 é o embaralhamento inverso do embaralhamento GSR.*

Prova. Vamos mostrar que o método 3 leva o estado σ no estado e com probabilidade $Q(\sigma)$, ou seja, $\hat{Q}(\sigma^{-1}) = Q(\sigma)$.

Com efeito, se $\sigma = e$, então $\hat{Q}(\sigma)$ é a probabilidade de todos os "0"s serem colocados antes de todos os "1"s.

Então,

$$\hat{Q}(e) = \frac{n+1}{2^n} = Q(e).$$

Se σ tem mais do que duas sequências levantadoras, então

$$\hat{Q}(\sigma^{-1}) = \hat{P}(\sigma, e) = 0 = Q(\sigma).$$

Finalmente, suponhamos que σ tenha duas sequências levantadoras, ou seja, para algum k , temos as sequências $1, 2, \dots, k$ e $(k+1), \dots, n$. Para que em um passo voltemos de σ para e é necessário que a moeda dê valor 0 a todas as cartas na primeira sequência levantadora e 1 para as demais. Isso ocorre com probabilidade $\frac{1}{2^n}$. Portanto, $\hat{Q}(\sigma^{-1}) = Q(\sigma)$. □

4.2

Uma primeira cota superior

Proposição 4.5 *Seja τ o número de embaralhamentos inversos necessários para que todas as cartas tenham frases (sequências de marcadores) distintas. Então, τ é um tempo estacionário forte.*

Prova. Temos que provar que $\mathbb{P}(X_\tau = \sigma, \tau = t) = \mathbb{P}(\tau = t)\pi(\sigma)$. Com efeito, seja t tal que $P(\tau = t) > 0$. Então,

$$\mathbb{P}(X_\tau = \sigma, \tau = t) = \mathbb{P}(X_\tau = \sigma | \tau = t) \mathbb{P}(\tau = t).$$

O evento $\{\tau = t\}$ garante que as n cartas no tempo t tenham frases distintas, mas não diz nada a respeito da ordem das cartas. Como os marcadores são escolhidos uniforme e independentemente, cada permutação das cartas tem a mesma probabilidade, que é $\frac{1}{n!}$. \square

Teorema 4.6 *Para o embaralhamento GSR em um baralho com n cartas, tem-se $t_{mix} \leq 2 \log_2(4n/3)$ para n suficientemente grande.*

Prova. A demonstração que segue encontra-se na página 108 de [12]. Pela Proposição C.8, Teorema 2.14 e Proposição 4.5, apenas precisamos calcular $P(\tau \leq t)$. Com efeito, se $\tau \leq t$, então no tempo t , todas as frases obtidas pelos marcadores são diferentes. Logo,

$$\begin{aligned} P(\tau \leq t) &= \frac{2^t}{2^t} \frac{2^t - 1}{2^t} \frac{2^t - 2}{2^t} \cdots \frac{2^t - (n-1)}{2^t} \\ &= \prod_{k=0}^{n-1} \left(1 - \frac{k}{2^t}\right). \end{aligned}$$

Então, tomando $t = 2 \log_2(n/c)$ e usando a aproximação de $\log x$ por série de Taylor, temos:

$$\begin{aligned} \log \prod_{k=0}^{n-1} \left(1 - \frac{k}{2^t}\right) &= - \sum_{k=0}^{n-1} \left(\frac{c^2 k}{n^2} + O\left(\frac{k}{n^2}\right)^2 \right) \\ &= - \frac{c^2 n(n-1)}{2n^2} + O\left(\frac{n^3}{n^4}\right) = - \frac{c^2}{2} + O\left(\frac{1}{n}\right). \end{aligned}$$

Assim,

$$\lim_{n \rightarrow \infty} \frac{P(\tau \leq t)}{e^{-c^2/2}} = 1.$$

Tomando $c < \sqrt{2 \log(4/3)} \approx 0,7585$, temos que $t_{mix} \leq 2 \log_2(4n/3)$. \square

4.3

Uma primeira cota inferior

Teorema 4.7 *Para o embaralhamento GSR em um baralho com n cartas, tem-se $t_{mix} \geq \log_2 n + O(1)$*

Prova. Vamos utilizar a cota por contagem. Lembremos que Δ é o número máximo de permutações alcançáveis em um passo. Neste caso, $\Delta \leq 2^n$. Como $|\Omega| = n!$, usando a cota por contagem (Proposição 2.16), temos

$$t_{mix} \geq \frac{\log \left(\frac{3}{4} |\Omega| \right)}{\log \Delta} \geq \frac{\log \left(\frac{3}{4} n! \right)}{n \log 2}.$$

Agora, usando a fórmula de Stirling $\left(n! \geq \sqrt{2\pi} e^{-n} n^{(n+\frac{1}{2})} \right)$, temos

$$\begin{aligned} t_{mix} &\geq \frac{\log \left(\frac{3}{4} \sqrt{2\pi} \right) - n + \left(n + \frac{1}{2} \right) \log n}{n \log 2} \geq \frac{-n + n \log n}{n \log 2} = \frac{\log n}{\log 2} - \frac{1}{\log 2} \\ &= \log_2 n + O(1). \end{aligned}$$

□

4.4

Generalizando o modelo: O a-embaralhamento

Utilizando a primeira cota superior encontrada, temos uma estimativa de que 12 embaralhadas são suficientes para alcançarmos o equilíbrio. Para encontrarmos uma cota melhor, iremos generalizar o nosso modelo, no qual cortamos o baralho em 2 pilhas, para um modelo no qual cortamos o baralho em um número $a > 1$ de pilhas.

Vejamos as seguintes descrições do novo modelo:

1. Descrição Geométrica:

- Coloquemos n pontos, x_1, x_2, \dots, x_n , de maneira uniforme e independente no intervalo $[0, 1]$, de forma que $x_1 < x_2 < \dots < x_n$.
- A aplicação $x \mapsto ax(\text{mod } 1)$, que leva $[0, 1]$ em $[0, 1]$, preserva medida, rearranja os pontos x_i e fornece uma medida em S_n que chamaremos de a-embaralhamento. ($y(\text{mod } 1) = y - \lfloor y \rfloor$).

2. Descrição de Entropia Máxima: Todas as maneiras de cortar o baralho em a pilhas e depois intercalá-las são equiprováveis. Pilhas vazias são permitidas.

3. Descrição Inversa: Todas as maneiras de separar um baralho embaralhado de volta em a pilhas são equiprováveis. O seguinte procedimento gera um a -embaralhamento inverso com probabilidade correta:

- Um baralho com n cartas é colocado virado para baixo.
- Sucessivas cartas são viradas para cima e colocadas em uma de a pilhas uniforme e independentemente.
- Depois de todas as cartas serem distribuídas, as pilhas são sobrepostas da esquerda para a direita e o baralho é virado para baixo.

4. Descrição Sequencial:

- Escolha a inteiros não negativos j_1, j_2, \dots, j_a de acordo com a distribuição multinomial

$$P(j_1, j_2, \dots, j_a) = \frac{n!}{j_1! j_2! \dots j_a!} \frac{1}{a^n}.$$

Assim, $0 \leq j_i \leq n$, $\sum_{i=1}^a j_i = n$ e os j_i têm a mesma distribuição que o número de bolas em cada caixa i se n bolas são colocadas ao acaso em a caixas.

- Dados os j_i , corte as j_1 cartas do topo. Em seguida, corte as próximas j_2 cartas do topo e prossiga assim, obtendo a ou menos pilhas.
- Embaralhe as duas primeiras pilhas usando o embaralhamento GSR. Depois embaralhe a pilha obtida com a terceira pilha, e prossiga assim. Isso é equivalente a intercalar todas as pilhas de uma só vez, onde se há A_i cartas restando em uma pilha, a chance de que a próxima carta saia da pilha i é $A_i / (A_1 + \dots + A_a)$.

Lema 4.8 *As quatro descrições geram a mesma distribuição de incremento (veja a Definição C.1). Além disso, em cada modelo, um a -embaralhamento seguido de um b -embaralhamento é equivalente a um ab -embaralhamento.*

Prova. Dividiremos a demonstração em três afirmações.

Afirmação 1: Cada descrição resulta em um número de cartas em cada pilha multinomial.

De fato, para a descrição sequencial isso segue pela própria definição e é claro para a descrição inversa. Para a descrição geométrica, os tamanhos das pilhas são determinados por quantos pontos são escolhidos no intervalo $[(i-1)/a, (i/a)]$, que é feito de acordo com a distribuição multinomial. Para a descrição de entropia máxima, o número de possíveis intercalamentos à partir de um corte dado é

multinomial, pois estão em correspondência 1 : 1 com as maneiras de dividir um baralho em a pilhas com seus tamanhos correspondentes.

Afirmção 2: Para as descrições de entropia máxima, inversa e sequencial, a regra do produto é válida.

De fato, dados os tamanhos das pilhas, a descrição de entropia máxima diz que todas os possíveis intercalamentos são equiprováveis. Isso também é claramente válido para a descrição inversa.

Para a descrição sequencial, observemos que quando as duas primeiras pilhas de tamanhos j_1, j_2 são embaralhadas, a probabilidade de uma sequência qualquer (dentre as possíveis) cair é

$$\frac{j_1(j_1 - 1) \cdots 1 j_2(j_2 - 1) \cdots 1}{(j_1 + j_2)(j_1 + j_2 - 1) \cdots 1} = \frac{1}{\binom{j_1 + j_2}{j_2}}.$$

Quando esta pilha resultante é embaralhada com a terceira pilha (de tamanho j_3), todas as $\binom{j_1 + j_2 + j_3}{j_3}$ possíveis sequências de cartas são equiprováveis. Prosseguindo indutivamente, todos os possíveis intercalamentos são equiprováveis.

Como vimos, nenhuma informação sobre os estados é retida entre embaralhamentos nestes três modelos, então a regra do produto para uma sequência de embaralhamentos vale em cada modelo uma vez que está estabelecida para um deles.

Assim, basta olharmos para a descrição inversa:

Combinando lexicograficamente os marcadores das pilhas de um a -embaralhamento inverso com os de um b -embaralhamento inverso, temos marcadores das pilhas de um ab -embaralhamento inverso. Para ver isso, seja A_i o evento em que o marcador a_i é escolhido para uma carta em um a -embaralhamento inverso e B_j o evento em que o marcador b_j é escolhido para a mesma carta em um b -embaralhamento inverso. Então,

$$P(A_i, B_j) = P(A_i)P(B_j) = \frac{1}{a} \frac{1}{b} = \frac{1}{ab} = P((AB)_k),$$

onde $(AB)_k$ é o evento em que o marcador $(ab)_k$ é escolhido para aquela carta em um ab -embaralhamento inverso.

Afirmção 3: Para a descrição geométrica, a regra do produto segue da independência de dígitos na base a dos pontos escolhidos uniformemente em $[0, 1]$. De fato, escolher n pontos em $[0, 1]$, marcá-los com os seus dígitos líderes (na base a) e aplicar $x \mapsto ax \pmod{1}$, é o mesmo que escolher n pontos em $[0, 1]$ e marcá-los com inteiros $\{0, 1, \dots, a - 1\}$ escolhidos uniformemente. Assim, como a

descrição geométrica nos permite usar os pontos para embaralhamentos sucessivos, o resultado segue da identidade $b(ax(\text{mod } 1))(\text{mod } 1) = abx(\text{mod } 1)$.

□

Teorema 4.9 *Seja E o evento em que um a -embaralhamento resulta em uma permutação σ com r sequências levantadoras. Então,*

$$P(E) = \frac{\binom{a+n-r}{n}}{a^n}.$$

Prova. Usando a descrição de entropia máxima, essa probabilidade é determinada pelo número de maneiras de cortar um baralho ordenado em a pilhas de maneira que σ seja um intercalamento possível. Como cada pacote mantém sua ordem original quando as cartas são intercaladas juntas, cada sequência levantadora no baralho é uma união de pilhas. Assim, queremos contar o número de maneiras de termos r sequências levantadoras em a pilhas. Faremos essa conta usando o argumento de estrelas e barras usado em combinatória. Devemos colocar $(a-1)$ barras (cortes) entre as cartas, gerando a pilhas, mas $(r-1)$ dessas barras devem estar entre os pares sucessivos de sequências levantadoras. As $(a-r)$ barras restantes podem estar em qualquer posição. Agora, consideremos $(a+n-r)$ objetos, sendo n cartas e $(a-r)$ barras. Então, o número de maneiras de colocarmos as barras entre as cartas é $\binom{a+n-r}{n}$, e este é o número de maneiras de dividirmos o baralho com n cartas em a pilhas de modo que σ seja uma configuração possível. Além disso, o número total de cortes possíveis para a obtenção de a pilhas é a^n , pois podemos associar cada carta a um dos a pacotes. Portanto,

$$P(E) = \frac{\binom{a+n-r}{n}}{a^n}.$$

□

O resultado a seguir não será utilizado em nossos problemas, mas é uma identidade famosa que pode ser provada facilmente usando o que temos até agora.

Corolário 4.10 (*Identidade de Worpitzky's*) *Sejam $A_{n,r}$ os números de Euler. Então,*

$$a^n = \sum_{r=1}^n A_{n,r} \binom{a+n-r}{n}.$$

Prova. Existem $A_{n,r}$ permutações com r sequências levantadoras. Assim, se R é o conjunto de todas as permutações com r sequências levantadoras, então

$$\sum_{r=1}^n \sum_{\sigma \in R} \frac{\binom{a+n-r}{n}}{a^n} = 1.$$

Logo,

$$a^n = \sum_{r=1}^n A_{n,r} \binom{a+n-r}{n}.$$

□

Corolário 4.11 *Se um baralho recebe uma sequência de m embaralhamentos dos tipos a_1, a_2, \dots, a_m , então a probabilidade de que o baralho esteja no arranjo σ é dada por*

$$\frac{\binom{a+n-r}{n}}{a^n},$$

onde $a = a_1 a_2 \cdots a_m$ e r é o número de sequências levantadoras em σ .

Prova. Basta combinarmos o Lema 4.8 com o Teorema 4.9. □

Teorema 4.12 *Se n cartas são embaralhadas m vezes com o embaralhamento GSR, então a probabilidade de que o baralho esteja no arranjo σ é dada por $\frac{\binom{2^m + n - r}{n}}{2^{mn}}$, onde r é o número de sequências levantadoras em σ .*

Prova. Segue imediatamente do Corolário 4.11. □

4.5

Encontrando o tempo de mistura

Proposição 4.13 *Seja $Q^m(r) = \binom{2^m + n - r}{n} / 2^{mn}$ a probabilidade de termos uma permutação com r sequências levantadoras depois de m embaralhadas com distribuição GSR. Seja $r = n/2 + h$, onde $-n/2 + 1 \leq h \leq n/2$ e tomemos $m = \log_2(n^{3/2}c)$ para algum $c > 0$. Então,*

$$Q^m(r) = \frac{1}{n!} \exp \left\{ \frac{1}{c\sqrt{n}} \left(-h + \frac{1}{2} + O_c \left(\frac{h}{n} \right) \right) - \frac{1}{24c^2} - \frac{1}{2} \left(\frac{h}{cn} \right)^2 + O_c \left(\frac{1}{n} \right) \right\}.$$

Prova.

$$\begin{aligned}
Q^m(r) &= \frac{(2^m + n - r)!}{n!(2^m - r)!2^{mn}} \\
&= \frac{1}{n!} \left[\frac{(2^m + n - r) \cdots (2^m + 1 - r)(2^m - r)!}{(2^m - r)!2^{mn}} \right] \\
&= \frac{1}{n!} \left(\frac{2^m + n - r}{2^m} \cdots \frac{2^m + 1 - r}{2^m} \right) \\
&= \frac{1}{n!} \exp \left\{ \sum_{i=0}^{n-1} \log \left(1 + \frac{(n-i) - r}{2^m} \right) \right\} \\
&= \frac{1}{n!} \exp \left\{ \sum_{i=0}^{n-1} \log \left(1 + \frac{(n-i) - r}{cn^{3/2}} \right) \right\} \\
&= \frac{1}{n!} \exp \left\{ \sum_{i=0}^{n-1} \log \left(1 + \frac{(n-i) - n/2 - h}{cn^{3/2}} \right) \right\} \\
&= \frac{1}{n!} \exp \left\{ \sum_{i=0}^{n-1} \log \left(1 + \frac{n/2 - h - i}{cn^{3/2}} \right) \right\}. \tag{4.1}
\end{aligned}$$

Afirmção: $-\frac{1}{2} < \frac{n/2-h-i}{cn^{3/2}} < 1$, para n suficientemente grande.

De fato, para $0 \leq i \leq n-1$, temos

$$n/2 - h - i \leq \frac{n}{2} + \left(\frac{n}{2} - 1\right) - i = n - 1 - i \leq n - 1 \leq n.$$

Daí,

$$\frac{n/2 - h - i}{cn^{3/2}} \leq \frac{n}{cn^{3/2}} = \left(\frac{1}{c}\right) \left(\frac{1}{\sqrt{n}}\right).$$

Por outro lado,

$$n/2 - h - i \geq \frac{n}{2} - \frac{n}{2} - i = -i \geq -n.$$

Daí,

$$\frac{n/2 - h - i}{cn^{3/2}} \geq -\frac{n}{cn^{3/2}} = \left(-\frac{1}{c}\right) \left(\frac{1}{\sqrt{n}}\right).$$

Logo, para n suficientemente grande, $-\frac{1}{2} < \frac{n/2-h-i}{cn^{3/2}} < 1$, para todo $i \in \{0, \dots, n-1\}$.

Veja também que para n suficientemente grande

$$\frac{1}{cn^{3/2}} \sum_{i=0}^{n-1} (n/2 - h - i) = \frac{-h + 1/2}{c\sqrt{n}}, \tag{4.2}$$

$$\frac{1}{2c^2n^3} \sum_{i=0}^{n-1} (n/2 - h - i)^2 = \frac{1}{24c^2} + \frac{1}{2} \left(\frac{h}{cn}\right)^2 + O_c\left(\frac{1}{n}\right), \tag{4.3}$$

$$\frac{1}{3c^3n^{9/2}} \sum_{i=0}^{n-1} (n/2 - h - i)^3 = O_c\left(\frac{h}{n^{3/2}}\right), \tag{4.4}$$

$$\frac{1}{c^4 n^6} \sum_{i=0}^{n-1} (n/2 - h - i)^4 = O_c\left(\frac{1}{n}\right). \quad (4.5)$$

Agora, usando o fato de que

$$x - \frac{x^2}{2} + \frac{x^3}{3} - x^4 \leq \log(1+x) \leq x - \frac{x^2}{2} + \frac{x^3}{3},$$

para $-\frac{1}{2} < x < 1$, e usando (4.2), (4.3), (4.4) e (4.5) em (4.1), temos que

$$\begin{aligned} Q^m(r) &= \frac{1}{n!} \exp \left\{ \frac{-h + 1/2}{c\sqrt{n}} - \left(\frac{1}{24c^2} + \frac{1}{2} \left(\frac{h}{cn} \right)^2 + O_c\left(\frac{1}{n}\right) \right) + O_c\left(\frac{h}{n^{3/2}}\right) + O_c\left(\frac{1}{n}\right) \right\} \\ &= \frac{1}{n!} \exp \left\{ \frac{1}{c\sqrt{n}} \left(-h + \frac{1}{2} + O_c\left(\frac{h}{n}\right) \right) - \frac{1}{24c^2} - \frac{1}{2} \left(\frac{h}{cn} \right)^2 + O_c\left(\frac{1}{n}\right) \right\}. \end{aligned}$$

□

Note que a probabilidade $Q^m(r)$ é monótona decrescente em r para m fixo. A próxima proposição determinará quando ela alcança o ponto $1/n!$.

Proposição 4.14 *Com a notação da Proposição 4.13, seja h^* um inteiro tal que $Q^m(n/2 + h) \geq 1/n! \Leftrightarrow h \leq h^*$. Então, para qualquer c fixo, quando $n \rightarrow \infty$,*

$$h^* = \frac{-\sqrt{n}}{24c} + \frac{1}{12c^3} + B + O_c\left(\frac{1}{\sqrt{n}}\right),$$

onde $-1 \leq B \leq 1$.

Prova. Pela Proposição 4.13, $Q^m(n/2 + h) \geq 1/n!$ se, e somente se,

$$\frac{1}{c\sqrt{n}} \left(-h + \frac{1}{2} + O_c\left(\frac{h}{n}\right) \right) - \frac{1}{24c^2} - \frac{1}{2} \left(\frac{h}{cn} \right)^2 + O_c\left(\frac{1}{n}\right) \geq 0.$$

Assim, basta tomarmos h^* tal que

$$\frac{1}{c\sqrt{n}} \left(-h^* + \frac{1}{2} + O_c\left(\frac{h^*}{n}\right) \right) - \frac{1}{24c^2} - \frac{1}{2} \left(\frac{h^*}{cn} \right)^2 + O_c\left(\frac{1}{n}\right) = 0.$$

□

O próximo resultado é o principal teorema deste capítulo, que combina as estimativas acima para obtermos o comportamento assintótico da variação total. Para isso, precisaremos da Proposição 2 de [16], que diz que se $b_n = x\sqrt{n/12} + n/2$, então

$$\lim_{n \rightarrow \infty} \sqrt{\frac{n}{12}} \frac{1}{n!} A_{n, [b_n]} = \frac{1}{\sqrt{2\pi}} e^{-(x^2/2)} \quad (4.6)$$

e

$$\lim_{n \rightarrow \infty} \frac{1}{n!} \sum_{h=1}^{[b_n]} A_{n,h} = \Phi(x), \quad (4.7)$$

onde $\Phi(x) = \int_{-\infty}^x e^{-t^2/2} dt / \sqrt{2\pi}$.

Teorema 4.15 *Sejam Q^m a distribuição do embaralhamento GSR no grupo simétrico S_n e U a distribuição uniforme. Se $m = \log_2(n^{3/2}c)$, com $0 < c < \infty$ fixo, então, quando $n \rightarrow \infty$,*

$$d(m) = \|Q^m - U\|_{TV} = 1 - 2\Phi\left(\frac{-1}{4c\sqrt{3}}\right) + O_c\left(\frac{1}{n^{1/4}}\right).$$

Prova. Seja $R_{n,h}$ o número de permutações com $n/2 + h$ sequências levantadoras. Pela Proposição 4.14, temos que

$$d(m) = \sum_{\sigma \in S_n} \left(Q^m(\sigma) - \frac{1}{n!}\right)^+ = \sum_{-n/2 < h \leq h^*} R_{n,h} \left(Q^m\left(\frac{n}{2} + h\right) - \frac{1}{n!}\right). \quad (4.8)$$

Tomando $x(n) := \frac{h}{\sqrt{n/12}} =: x_n$ e substituindo em (4.6), obtemos

$$\frac{R_{n,h}}{n!} = \frac{A_{n,n/2+h}}{n!} = \frac{e^{-(1/2)x_n^2}}{\sqrt{2\pi n/12}} \left(1 + o\left(\frac{1}{\sqrt{n}}\right)\right). \quad (4.9)$$

Já substituindo em (4.7), obtemos

$$\begin{aligned} \frac{1}{n!} \sum_{h=-n/2+1}^{h^*} R_{n,h} &= \frac{1}{n!} \sum_{h=-n/2+1}^{h^*} A_{n,n/2+h} \\ &= \frac{1}{n!} \sum_{h=1}^{h^*+n/2} A_{n,h} \\ &= \Phi\left(\frac{h^*}{\sqrt{n/12}}\right) = \Phi\left(\frac{2\sqrt{3}h^*}{\sqrt{n}}\right) \\ &= \Phi\left(\frac{2\sqrt{3}}{\sqrt{n}} \left[-\frac{\sqrt{n}}{24c} + O_c(1)\right]\right) \\ &= \Phi\left(\frac{-\sqrt{3}}{12c} + O_c\left(\frac{1}{\sqrt{n}}\right)\right) \\ &= \Phi\left(\frac{-1}{4c\sqrt{3}} + O_c\left(\frac{1}{\sqrt{n}}\right)\right) \\ &= \Phi\left(\frac{-1}{4c\sqrt{3}}\right) \left(1 + O_c\left(\frac{1}{\sqrt{n}}\right)\right). \end{aligned} \quad (4.10)$$

A soma em (4.8) pode ser quebrada em duas zonas:

$$\text{zona 1: } \left\{ \frac{-10n^{3/4}}{\sqrt{c}} \leq h \leq h^* \right\} = I_1,$$

e

$$\text{zona 2: } \left\{ -\frac{n}{2} < h < \frac{-10n^{3/4}}{\sqrt{c}} \right\} = I_2.$$

Será mostrado que apenas a zona 1 contribui. Antes de realizarmos as contas, vejamos que em I_1

$$x_n = \frac{h}{\sqrt{n/12}} \geq \frac{-10\sqrt{12}n^{3/4}}{n^{1/2}\sqrt{c}} = -20\sqrt{3}n^{1/4}/\sqrt{c}$$

e

$$x_n = \frac{h}{\sqrt{n/12}} \leq \frac{h^*}{\sqrt{n/12}} = \frac{\left(\frac{-\sqrt{n}}{24c} + O_c(1)\right) \sqrt{12}}{\sqrt{n}} = \frac{-\sqrt{12}}{24c} + o(1) = -\frac{1}{2c\sqrt{12}} + o(1).$$

Por (4.9) e pela Proposição 4.13,

$$\begin{aligned} & \sum_{I_1} R_{n,h} Q^m(n/2 + h) \\ &= \frac{e^{\frac{-1}{24c^2}}}{\sqrt{2\pi n/12}} \sum_{I_1} \exp \left\{ -\frac{1}{2} \left(\frac{h}{\sqrt{n/12}} \right)^2 - \frac{h}{c\sqrt{n}} + O_c \left(\frac{1}{n^{1/4}} \right) \right\} \left\{ 1 + o \left(\frac{1}{\sqrt{n}} \right) \right\} \\ &= \frac{e^{\frac{-1}{24c^2}}}{\sqrt{2\pi}} \left(\frac{1}{\sqrt{n/12}} \sum_{I_1} \exp \left\{ -\frac{x_n^2}{2} - \frac{x_n}{2c\sqrt{3}} + O_c \left(\frac{1}{n^{1/4}} \right) \right\} \right) \left\{ 1 + o \left(\frac{1}{\sqrt{n}} \right) \right\} \\ &= \frac{e^{\frac{-1}{24c^2}}}{\sqrt{2\pi}} \left(\frac{1}{\sqrt{n/12}} \sum_{x=\lceil -20\sqrt{3}n^{1/4}/\sqrt{c} \rceil}^{\lfloor -(2c\sqrt{12})^{-1} + o(1) \rfloor} \exp \left\{ -\frac{x^2}{2} - \frac{x}{2c\sqrt{3}} + O_c \left(\frac{1}{n^{1/4}} \right) \right\} \right) \times \\ & \quad \times \left\{ 1 + o \left(\frac{1}{\sqrt{n}} \right) \right\} \\ &= \frac{e^{\frac{-1}{24c^2}}}{\sqrt{2\pi}} \int_{-\infty}^{-(2c\sqrt{12})^{-1}} e^{-x^2/2 - x/(2c\sqrt{3})} dx \left(1 + O_c \left(\frac{1}{n^{1/4}} \right) \right) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-(2c\sqrt{12})^{-1}} e^{-x^2/2 - x/(2c\sqrt{3}) - 1/(24c^2)} dx \left(1 + O_c \left(\frac{1}{n^{1/4}} \right) \right) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-(2c\sqrt{12})^{-1}} e^{-\frac{(x+1/(2c\sqrt{3}))^2}{2}} dx \left(1 + O_c \left(\frac{1}{n^{1/4}} \right) \right) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(4c\sqrt{3})^{-1}} e^{-\frac{y^2}{2}} dy \left(1 + O_c \left(\frac{1}{n^{1/4}} \right) \right) \\ &= \Phi \left(\frac{1}{4c\sqrt{3}} \right) \left(1 + O_c \left(\frac{1}{n^{1/4}} \right) \right) \\ &= \left[1 - \Phi \left(\frac{-1}{4c\sqrt{3}} \right) \right] \left(1 + O_c \left(\frac{1}{n^{1/4}} \right) \right). \end{aligned} \tag{4.11}$$

Na zona 2, $Q^m(n/2 + h) \leq Q^m(1) \leq e^{\sqrt{n}/(2c)}/n!$. A cota de grandes desvios padrões, como no capítulo 16 de [7] aplicada à soma de n variáveis aleatórias

uniformes garante que

$$\begin{aligned}
 \sum_{I_2} R_{n,h} Q^m(n/2 + h) &\leq \sum_{I_2} \frac{R_{n,h} e^{\sqrt{n}/(2c)}}{n!} \\
 &= e^{\sqrt{n}/(2c)} \sum_{I_2} \frac{R_{n,h}}{n!} \\
 &\sim \frac{e^{\sqrt{n}/(2c)}}{10n^{1/4}\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{10\sqrt{12}n^{1/4}}{\sqrt{c}} \right)^2 \right], \quad (4.12)
 \end{aligned}$$

que converge para 0 quando $n \rightarrow \infty$.

O resultado segue por (4.8), (4.10), (4.11) e (4.12). \square

Corolário 4.16 *Se n cartas são embaralhadas m vezes com $m = \frac{3}{2} \log_2 n + \theta$, então para n suficientemente grande,*

$$d(m) = \|Q^m - U\|_{TV} = 1 - 2\Phi \left(\frac{-2^{-\theta}}{4\sqrt{3}} \right) + O \left(\frac{1}{n^{1/4}} \right).$$

Prova. Basta tomarmos $c = 2^\theta$ e o resultado segue imediatamente do Teorema 4.15. \square

Embora a distância de variação total fique menor que $1/4$ apenas depois de oito embaralhadas, este teorema é conhecido como teorema das sete embaralhadas pois é a partir deste número que a distância de variação total fica menor que $1/2$.

5

O embaralhamento de transposições adjacentes visto em tempo contínuo

Até agora, sempre analisamos a mistura dos passeios aleatórios em S_N em tempo discreto. Também faremos isso para o embaralhamento de transposições adjacentes, porém no Capítulo 6. Antes, analisaremos este embaralhamento quando o mesmo é visto em tempo contínuo, o que nos ajudará não apenas a resolver o problema em tempo discreto, mas também nos dará uma bela aplicação em sistemas de partículas que será apresentada no Capítulo 7.

5.1

O embaralhamento de transposições adjacentes

O embaralhamento das transposições adjacentes é uma cadeia de Markov em tempo contínuo em S_N . Suponhamos que tenhamos um baralho com N cartas marcadas de 1 até N do topo para o fundo. Assim, para um vetor de cartas, associamos uma permutação σ dizendo que $\sigma(x) = y$ se a x -ésima posição na pilha está ocupada pela carta y . Nossa cadeia escolhe uma carta uniformemente ao acaso dentre as cartas nas posições 1 até $N - 1$ e troca sua posição com a da carta que está imediatamente abaixo desta. Veja um exemplo na Figura 5.1

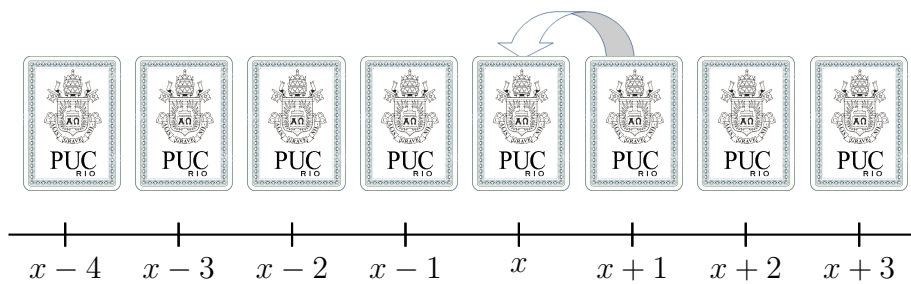


Figura 5.1: Exemplo de uma transposição adjacente.

Mais formalmente, sejam $(\tau_x)_{1 \leq x \leq N-1}$ as transposições para o vizinho mais próximo $(x, x + 1)$.

Note que o conjunto $\{\tau_x : 1 \leq x \leq N-1\}$ é um gerador de S_N no sentido da teoria de grupos.

O gerador \mathcal{L} (veja [13, 14]) do embaralhamento é definido por sua ação nas funções de \mathbb{R}^Ω como:

$$(\mathcal{L}f)(\sigma) := \sum_{x=1}^{N-1} f(\sigma \circ \tau_x) - f(\sigma). \quad (5.1)$$

Sejam $(\sigma_t)_{t \geq 0}$ a trajetória da cadeia de Markov com $\sigma_0 = e$ (a identidade) e P_t a lei de distribuição na marginal σ_t . Dada uma distribuição de probabilidade ν , definimos P_t^ν como a distribuição marginal de σ_t^ν , a cadeia de Markov começando com distribuição inicial ν . Além disso, como $(\tau_x)_{x=1}^{N-1}$ gera o grupo S_N , essa cadeia de Markov é irredutível e a única medida de probabilidade invariante é a uniforme π .

Agora, definamos a distância do equilíbrio dessa cadeia de Markov como

$$d^N(t) := \|P_t - \pi\|_{TV}.$$

Pela simetria de S_N , d^N não depende da condição inicial e pela Proposição 2.5,

$$d^N(t) := \max_{\{\nu \text{ probabilidade em } S_N\}} \|P_t^\nu - \pi\|_{TV}.$$

Finalmente, para $\epsilon \in (0, 1)$, definiremos o ϵ -tempo de mistura como

$$T_{mix}^N(\epsilon) := \inf \{t \geq 0 : d^N(t) \leq \epsilon\}.$$

Algumas vezes, para que alguns conceitos fiquem mais claros, tentaremos explicá-los usando um baralho usual com 52 cartas. Por convenção, nossa identidade estará organizada de modo que as cartas vermelhas estejam sobre as pretas e entre duas cartas de um naipe fixo tenham apenas cartas deste mesmo naipe.

5.2

Tirando vantagem da monotonicidade

Devido à completa simetria de S_N , colocar ordem no conjunto das permutações parece ser uma ideia estranha e muito complicada. Assim, daremos um papel especial para a identidade, a qual fixamos como o elemento maximal em um certo sentido, e isso nos ajudará a quebrar esta simetria.

Antes de definirmos nossa ordem em S_N , vamos introduzir uma aplicação que leva permutações em superfícies discretas.

5.2.1

Mapeando permutações em superfícies discretas

Definição 5.1 À cada $\sigma \in S_N$ associamos a função $\tilde{\sigma} : \{0, \dots, N\}^2 \rightarrow \mathbb{R}$ definida por

$$\tilde{\sigma}(x, y) := \sum_{z=1}^x \mathbf{1}_{\{\sigma(z) \leq y\}} - \frac{xy}{N}.$$

Note que o termo xy/N é subtraído para que $\tilde{\sigma}(x, y)$ tenha média zero com respeito à medida de equilíbrio π .

De fato,

$$\begin{aligned} \mathbb{E}_\pi [\tilde{\sigma}(x, y)] &= \mathbb{E}_\pi \left[\sum_{z=1}^x \mathbf{1}_{\{\sigma(z) \leq y\}} \right] - \frac{xy}{N} \\ &= \left(\sum_{z=1}^x \mathbb{P}_\pi [\sigma(z) \leq y] \right) - \frac{xy}{N} = \left(\sum_{z=1}^x \frac{y}{N} \right) - \frac{xy}{N} = 0. \end{aligned}$$

Proposição 5.2 A aplicação supracitada é injetiva.

Prova. De fato,

$$\begin{aligned} &\tilde{\sigma}(x, y) - \tilde{\sigma}(x, y-1) - \tilde{\sigma}(x-1, y) + \tilde{\sigma}(x-1, y-1) + \frac{1}{N} \\ &= \sum_{z=1}^x \mathbf{1}_{\{\sigma(z) \leq y\}} - \frac{xy}{N} - \sum_{z=1}^x \mathbf{1}_{\{\sigma(z) \leq y-1\}} + \frac{x(y-1)}{N} \\ &\quad - \sum_{z=1}^{x-1} \mathbf{1}_{\{\sigma(z) \leq y\}} + \frac{(x-1)y}{N} + \sum_{z=1}^{x-1} \mathbf{1}_{\{\sigma(z) \leq y-1\}} - \frac{(x-1)(y-1)}{N} + \frac{1}{N} \\ &= \sum_{z=1}^x \left[\mathbf{1}_{\{\sigma(z) \leq y\}} - \mathbf{1}_{\{\sigma(z) \leq y-1\}} \right] - \sum_{z=1}^{x-1} \left[\mathbf{1}_{\{\sigma(z) \leq y\}} - \mathbf{1}_{\{\sigma(z) \leq y-1\}} \right] \\ &= \sum_{z=1}^x \mathbf{1}_{\{\sigma(z)=y\}} - \sum_{z=1}^{x-1} \mathbf{1}_{\{\sigma(z)=y\}} = \mathbf{1}_{\{\sigma(x)=y\}}. \end{aligned}$$

Assim, se $\tilde{\sigma}_1 = \tilde{\sigma}_2$, então para todos $x, y \in \{0, \dots, N\}$, $\mathbf{1}_{\{\sigma_1(x)=y\}} = \mathbf{1}_{\{\sigma_2(x)=y\}}$. Logo, $\sigma_1 = \sigma_2$.

□

Identificaremos o conjunto imagem $\{\tilde{\sigma} : \sigma \in S_N\}$ com S_N desde que isso não gere confusão.

Definição 5.3 Definimos uma relação de ordem natural (parcial) em S_N por

$$\sigma \leq \sigma' \Leftrightarrow \forall x, y, \tilde{\sigma}(x, y) \leq \tilde{\sigma}'(x, y).$$

Proposição 5.4 A identidade, que denotamos por e , é o elemento maximal de (S_N, \geq) , e a permutação σ_{min} definida por

$$\forall x \in \{1, \dots, N\}, \sigma_{min}(x) := N + 1 - x$$

é o elemento minimal.

Prova. Primeiro, notemos que $\tilde{e}(x, y) + \frac{xy}{N} = \sum_{z=1}^x \mathbf{1}_{\{z \leq y\}} = \min \{x, y\}$.

Se $x \leq y$, então

$$\tilde{e}(x, y) + \frac{xy}{N} = x = \sum_{z=1}^x 1 \geq \sum_{z=1}^x \mathbf{1}_{\{\sigma(z) \leq y\}} = \tilde{\sigma}(x, y) + \frac{xy}{N}.$$

Logo, $\tilde{e}(x, y) \geq \tilde{\sigma}(x, y)$. Indiferentemente, se $x > y$, então, como há no máximo $y - 1$ cartas com marcadores menores que y ,

$$\tilde{\sigma}(x, y) + \frac{xy}{N} = \sum_{z=1}^x \mathbf{1}_{\{\sigma(z) \leq y\}} \leq y = \tilde{e}(x, y) + \frac{xy}{N}.$$

Assim, $\tilde{e}(x, y) \geq \tilde{\sigma}(x, y)$, e portanto e é o elemento maximal de (S_N, \geq) . A prova de que σ_{min} é o elemento minimal é análoga. \square

5.2.2

A construção gráfica

Apresentaremos agora uma construção da dinâmica que nos permite construir todas as trajetórias σ_t^ξ , começando com condição inicial $\xi \in S_N$, simultaneamente (um grande acoplamento e tem a propriedade de conservar ordem).

Associamos à cada $x \in \{1, \dots, N - 1\}$ um processo de Poisson independente $(\mathcal{T}^x) = (\mathcal{T}_n^x)_{n \geq 0}$ com intensidade dois.

Em outras palavras, $\mathcal{T}_0^x = 0$ para todo x e

$$(\mathcal{T}_n^x - \mathcal{T}_{n-1}^x)_{x \in \{1, \dots, N-1\}, n \geq 1}$$

é um campo de variáveis aleatórias exponenciais i.i.d. com média $1/2$.

Referimo-nos a $\mathcal{T} = (\mathcal{T}^x)_{1 \leq x \leq N-1}$ como o processo do relógio. Note que o conjunto de valores tomados pelo processo do relógio é quase certamente um subconjunto discreto de \mathbb{R} , e como consequência, a probabilidade de que dois relógios toquem simultaneamente é zero.

Seja $(U_n^x)_{x \in \{1, \dots, N-1\}, n \geq 1}$ um campo de variáveis aleatórias Bernoulli i.i.d. ($U_n^x \in \{0, 1\}$) com parâmetro $1/2$, independentes de \mathcal{T} . Agora, dados \mathcal{T} e U , construiremos a trajetória da cadeia de Markov começando em $\xi \in S_N$.

A trajetória $(\sigma_t^\xi)_{t \geq 0}$ é càdlàg (contínua à direita com limite à esquerda) e é constante em intervalos onde o processo do relógio é silencioso.

Quando um relógio toca, isto é, no tempo $t = \mathcal{T}_n^x (n \geq 1)$, σ_t^ξ é construída atualizando σ_{t-}^ξ como abaixo:

- Se $\sigma_{t-}(x+1) \leq \sigma_{t-}(x)$ e $U_n^x = 1$, ou $\sigma_{t-}(x+1) \geq \sigma_{t-}(x)$ e $U_n^x = 0$, então nós trocamos os valores de $\sigma_{t-}(x)$ e $\sigma_{t-}(x+1)$;
- Em outros casos, não fazemos nada.

Em outras palavras, quando o processo do relógio associado a x toca, nós ordenamos as cartas das posições x e $x+1$ se $U_n^x = 1$, e as desordenamos se $U_n^x = 0$.

É imediato verificar que essa construção nos fornece uma cadeia de Markov com gerador \mathcal{L} descrito em (5.1).

O efeito da atualização em $\tilde{\sigma}$ é o seguinte: Para cada $y \in \{1, \dots, N-1\}$, se $(\tilde{\sigma}_{t-}(z, y))_{z \in \{1, \dots, N-1\}}$ apresenta um mínimo local em $z = x$ e $U_n^x = 1$, então esse é transformado em um máximo local, isto é, $\tilde{\sigma}_t(x, y) = \tilde{\sigma}_{t-}(x, y) + 1$. Do contrário, se apresenta um máximo local em $z = x$ e $U_n^x = 0$, então $\tilde{\sigma}_t(x, y) = \tilde{\sigma}_{t-}(x, y) - 1$. Chamamos esta operação de atualização de σ na coordenada x .

Agora mostraremos que essa construção preserva a ordem.

Proposição 5.5 *Seja $\xi \geq \xi'$ dois elementos de S_N . Com a construção gráfica acima, temos $\sigma_t^\xi \geq \sigma_t^{\xi'}$.*

Prova. A demonstração que segue encontra-se em [11].

A única coisa a se verificar é que a ordem é conservada cada vez que o processo do relógio toca, isto é, para todo (n, x) e $t = \mathcal{T}_n^x$,

$$\sigma_{t-}^\xi \geq \sigma_{t-}^{\xi'} \Rightarrow \sigma_t^\xi \geq \sigma_t^{\xi'}.$$

Suponhamos que $\sigma_{t-}^{\xi} \geq \sigma_{t-}^{\xi'}$. Mostraremos que para todo $y \in \{1, \dots, N-1\}$, $\tilde{\sigma}_t^{\xi}(x, y) \geq \tilde{\sigma}_t^{\xi'}(x, y)$, o que implica o resultado já que as outras coordenadas não são trocadas no tempo t .

Com efeito, fixemos $y \in \{1, \dots, N-1\}$. Notemos que quando $\tilde{\sigma}_{t-}^{\xi}(x, y) > \tilde{\sigma}_{t-}^{\xi'}(x, y)$, não precisamos fazer nada, pois não é possível que $\tilde{\sigma}^{\xi}$ pule para baixo enquanto $\tilde{\sigma}^{\xi'}$ pula para cima.

Por esse motivo, vamos assumir que $\tilde{\sigma}_{t-}^{\xi}(x, y) = \tilde{\sigma}_{t-}^{\xi'}(x, y)$. Se $U_n^x = 1$, apenas temos que verificar que se $\tilde{\sigma}_t^{\xi'}(x, y)$ pula para cima, então $\tilde{\sigma}_t^{\xi}(x, y)$ também o faz. Isso ocorre porque se $\tilde{\sigma}_{t-}^{\xi'}(., y)$ apresenta um mínimo local em x , então $\tilde{\sigma}_{t-}^{\xi}(., y)$ também o faz. Analogamente, se $U_n^x = 0$ e $\tilde{\sigma}_t^{\xi'}(x, y)$ pula para baixo então $\tilde{\sigma}_t^{\xi}(x, y)$ também o faz. \square

5.2.3

Ordenação estocástica e sua preservação

Definição 5.6 *Sejam α, β duas medidas de probabilidade em um conjunto finito e ordenado Ω . Dizemos que α domina β estocasticamente e escrevemos $\alpha \succeq \beta$ se podemos encontrar um acoplamento Π , isto é, uma probabilidade em Ω^2 tal que a primeira marginal tem lei α e a segunda β , e que satisfaz*

$$\Pi \{(\omega_1, \omega_2) \in \Omega^2 : \omega_1 \geq \omega_2\} = 1$$

Definição 5.7 *Dizemos que uma função f em Ω é crescente se $\forall \omega, \omega' \in \Omega, \omega \geq \omega' \Rightarrow f(\omega) \geq f(\omega')$.*

Definição 5.8 *Para um conjunto ordenado Ω , dizemos que um subconjunto A (um evento) é crescente se a função $\mathbf{1}_A$ é crescente ou equivalentemente, se $\forall \omega \in A, \omega \geq \omega' \Rightarrow \omega' \in A$.*

Agora daremos uma caracterização equivalente para a dominação estocástica.

Usaremos a notação $\alpha(f)$ para a esperança de f com respeito à α .

Lema 5.9 *Sejam α e β duas medidas de probabilidade em um conjunto finito e ordenado Ω . As seguintes afirmações são equivalentes:*

- (i) α domina β ;
- (ii) Para toda função crescente f definida em Ω , $\alpha(f) \geq \beta(f)$.

Prova. Seja f uma função crescente definida em Ω . Se $\alpha \succeq \beta$, então existem variáveis aleatórias X, Y com leis α, β , respectivamente, tais que $\mathbb{P}[X > z] \geq \mathbb{P}[Y > z]$, $\forall z \in \mathbb{R}$.

Assim,

$$\mathbb{P}[f(X) > z] = \mathbb{P}[X > f^{-1}(z)] \geq \mathbb{P}[Y > f^{-1}(z)] = \mathbb{P}[f(Y) > z].$$

Se $f(X), f(Y) \geq 0$, então

$$\alpha(f) = \mathbb{E}[f(X)] = \int_0^\infty \mathbb{P}[f(X) > t] dt \geq \int_0^\infty \mathbb{P}[f(Y) > t] dt = \mathbb{E}[f(Y)] = \beta(f).$$

Se não temos $f(X) \geq 0$ e $f(Y) \geq 0$, então

$$\begin{aligned} \alpha(f) &= \mathbb{E}[f(X)] = \int_0^\infty dt - \int_{-\infty}^\infty \mathbb{P}[f(X) \leq t] dt \\ &\geq \int_0^\infty dt - \int_{-\infty}^\infty \mathbb{P}[f(Y) \leq t] dt \\ &= \mathbb{E}[f(Y)] = \beta(f). \end{aligned}$$

Reciprocamente, para cada $z \in \mathbb{R}$, definamos $f_z(X) := \mathbf{1}_{\{X > z\}}$, que é monótona.

Assim,

$$\mathbb{P}[X > z] = \mathbb{E}[f_z(X)] = \alpha(f_z) \geq \beta(f_z) = \mathbb{E}[f_z(Y)] = \mathbb{P}[Y > z].$$

Logo, $\alpha \succeq \beta$.

□

Uma consequência da Proposição 5.5 é que se ν e ν' são medidas de probabilidade em S_N , então

$$\nu \succeq \nu' \Rightarrow \forall t \geq 0, P_t^\nu \succeq P_t^{\nu'}. \quad (5.2)$$

Agora vamos introduzir uma simples ferramenta para produzir acoplamentos estocásticos.

Lema 5.10 *Seja Ω um conjunto finito e $(\omega_t^1)_{t \geq 0}$ e $(\omega_t^2)_{t \geq 0}$ dois processos estocásticos em Ω . Assuma que as distribuições de ω_t^1 e ω_t^2 convergem para duas medidas de probabilidade α e β , respectivamente, quando t tende ao infinito. Se podemos encontrar um acoplamento dos processos tal que para todo $t \geq 0$, $\omega_t^1 \geq \omega_t^2$ quase certamente, então $\alpha \succeq \beta$.*

Prova. A demonstração que segue encontra-se em [11].

Seja Π_t a lei de (ω_t^1, ω_t^2) sob o acoplamento dado pelo lema.

Para todo $t \geq 0$, Π_t é suportado no conjunto

$$\mathcal{D} = \{(\omega^1, \omega^2) \in \Omega^2 : \omega^1 \geq \omega^2\}.$$

Como Π_t mora num espaço compacto (pela topologia induzida pela distância de variação total), ela tem pelo menos um ponto limite que chamaremos de Π e é suportado em \mathcal{D} . A medida Π gera um acoplamento que satisfaz $\alpha \succeq \beta$. \square

5.2.4

Desigualdades de correlação e a desigualdade FKG

Um das importantes ferramentas que usaremos são as desigualdades de correlação, que dizem que condicionar π em um evento crescente faz todos os outros eventos crescentes mais prováveis.

Lema 5.11 *Sejam f e g duas funções reais crescentes de uma variável e X uma variável aleatória com lei \mathbb{P} . Então,*

$$\mathbb{E}[f(X)g(X)] \geq \mathbb{E}[f(X)] \mathbb{E}[g(X)]. \quad (5.3)$$

Prova. Seja X' uma cópia independente de X , que existe pelo Teorema de Kolmogorov (veja o Teorema 2.1.14. na página 51 de [6]). Então, como f e g são crescentes, temos que

$$\begin{aligned} 0 &\leq \mathbb{E}[(f(X) - f(X'))(g(X) - g(X'))] \\ &= \mathbb{E}[f(X)g(X) - f(X)g(X') - f(X')g(X) + f(X')g(X')] \\ &= \mathbb{E}[f(X)g(X)] - \mathbb{E}[f(X)g(X')] - \mathbb{E}[f(X')g(X)] + \mathbb{E}[f(X')g(X')] \\ &= 2\mathbb{E}[f(X)g(X)] - 2\mathbb{E}[f(X)] \mathbb{E}[g(X)]. \end{aligned}$$

Portanto,

$$\mathbb{E}[f(X)g(X)] \geq \mathbb{E}[f(X)] \mathbb{E}[g(X)].$$

\square

A desigualdade (5.3) não é verdadeira em geral para todas as noções de ordem parcial, mas existe uma generalização dela para "lattices distributi-

vos"conhecida por desigualdade FKG (Fortuin-Kasteleyn-Ginibre).

Infelizmente, S_N não é um lattice distributivo. Mais precisamente, se definirmos para σ e σ' em S_N , $\min(\tilde{\sigma}, \tilde{\sigma}')$ e $\max(\tilde{\sigma}, \tilde{\sigma}')$ por

$$\min(\tilde{\sigma}, \tilde{\sigma}')(x, y) := \min(\tilde{\sigma}(x, y), \tilde{\sigma}'(x, y)),$$

e

$$\max(\tilde{\sigma}, \tilde{\sigma}')(x, y) := \max(\tilde{\sigma}(x, y), \tilde{\sigma}'(x, y)),$$

então $\min(\tilde{\sigma}, \tilde{\sigma}')$ e $\max(\tilde{\sigma}, \tilde{\sigma}')$ não são necessariamente imagens de elementos de S_N . Porém, em [11], é feita uma adaptação da prova escrita em [9] para o nosso caso.

Proposição 5.12 (A desigualdade FKG para permutações) *Para qualquer par de funções crescentes f e g definidas em S_N ,*

$$\pi(f(\sigma)g(\sigma)) \geq \pi(f(\sigma))\pi(g(\sigma)).$$

Prova.

Primeiro, vejamos que para qualquer par (A, B) de conjuntos crescentes, temos $\pi(A \cap B) \geq \pi(A)\pi(B)$.

De fato, sejam A e B dois subconjuntos crescentes de S_N . Vamos começar da identidade e iterar duas dinâmicas acopladas σ_t e σ_t^A definidas da seguinte maneira: σ_t é um embaralhamento de transposições adjacentes normal e σ_t^A tem a mesma regra de transição exceto que todas as transições que saem de A são canceladas (isso é chamado de cadeia de Markov refletida).

Acoplamos as duas dinâmicas usando a construção gráfica feita da seção 5.2.2, com ambas as dinâmicas usando o mesmo processo do relógio \mathcal{T} e variáveis de atualização U , a única diferença sendo o fato de que σ_t^A cancela as transições que fazem a cadeia sair de A .

A cadeia de Markov σ_t^A é irredutível. A razão para isso ocorrer é o fato de que para cada $(\sigma, \sigma') \in A^2$, podemos sempre encontrar uma sequência de transições que vão para cima (ordenando vizinhos) que leva σ até e (a identidade) e uma sequência de transições que vão para baixo e que leva e até σ' . A concatenação destas duas sequências nos fornece um caminho de transições de σ até σ' cujos passos estão todos em A (elas são $\geq \sigma$ na primeira parte e $\geq \sigma'$ na segunda parte).

Além disso, $\pi(\cdot|A)$, isto é, a medida uniforme em A , é reversível para σ^A , pois

$$\pi(\sigma|A) \cdot P_t(\sigma, \sigma') = \pi(\sigma'|A) \cdot P_t(\sigma', \sigma), \forall \sigma, \sigma' \in A.$$

Portanto, σ_t^A converge para $\pi(\cdot|A)$.

Como as únicas transições que são canceladas em σ^A são as transições que "descem- correspondendo a desordenação de um par adjacente -, temos como consequência da Proposição 5.5 que

$$\forall t \geq 0, \sigma_t^A \geq \sigma_t.$$

Usando o Lema 5.10, obtemos $\pi(\cdot|A) \succeq \pi$. Logo, $\pi(B|A) \geq \pi(B)$. E portanto, $\pi(A \cap B) \geq \pi(A)\pi(B)$.

Agora, sejam f e g duas funções positivas e crescentes (não há perda de generalidade em assumir positividade pois adicionar uma constante em f ou g não altera a desigualdade proposta) e $x, y \in \mathbb{R}$.

Definimos os seguintes conjuntos crescentes:

$$A_x := \{f(\sigma) \geq x\} \text{ e } B_y := \{g(\sigma) \geq y\}.$$

Como $f = \int_{\mathbb{R}} \mathbf{1}_{A_x} dx$ e $g = \int_{\mathbb{R}} \mathbf{1}_{B_y} dy$,

$$\begin{aligned} \pi(f(\sigma)g(\sigma)) &= \pi\left(\int_{\mathbb{R}^2} \mathbf{1}_{A_x} \mathbf{1}_{B_y} dx \cdot dy\right) = \int_{\mathbb{R}^2} \pi(A_x \cap B_y) dx \cdot dy \\ &\geq \int_{\mathbb{R}^2} \pi(A_x) \pi(B_y) dx \cdot dy = \left(\int_{\mathbb{R}} \pi(A_x) dx\right) \left(\int_{\mathbb{R}} \pi(B_y) dy\right) \\ &= \left(\int_{\mathbb{R}} \pi(\mathbf{1}_{A_x}) dx\right) \left(\int_{\mathbb{R}} \pi(\mathbf{1}_{B_y}) dy\right) = \pi\left(\int_{\mathbb{R}} \mathbf{1}_{A_x} dx\right) \pi\left(\int_{\mathbb{R}} \mathbf{1}_{B_y} dy\right) \\ &= \pi(f(\sigma)) \pi(g(\sigma)). \end{aligned}$$

□

5.2.5

A desigualdade censuradora

A desigualdade censuradora, um resultado estabelecido em [15] para "sistemas monótonos", é uma noção que levemente generaliza a dinâmica de Glauber para sistemas de spins com um espaço de spins totalmente ordenado e diz que cancelar algumas das atualizações dos spins não tem o efeito de atrasar muito a mistura. Infelizmente, o embaralhamento de transposições adjacentes não é um

sistema monótono no sentido de [15]. Todavia, em [11], é feita uma adaptação da prova do resultado supracitado para o nosso setup. Antes de apresentarmos o resultado, vamos introduzir um pouco de terminologia e notação.

Definição 5.13 *Um esquema censurador é uma função càdlàg $\mathcal{C} : \mathbb{R} \rightarrow \mathcal{P}(\{1, \dots, N-1\})$ onde $\mathcal{P}(S)$ é o conjunto das partes de S .*

A dinâmica censurada com esquema \mathcal{C} é a dinâmica obtida usando a construção gráfica descrita na seção 5.2.2 exceto que se \mathcal{T}^x toca em tempo t , a atualização é feita se, e somente se, $x \in \mathcal{C}(t)$.

É natural pensar que a cada vez que o relógio toca, σ_t estará mais perto do equilíbrio e portanto a censura apenas fará a convergência ao equilíbrio mais lenta. A desigualdade censuradora garante que isso é verdade se começarmos com uma medida cuja densidade é uma função crescente.

Dados um esquema censurador \mathcal{C} e uma distribuição de probabilidade ν em S_N , seja $P_t^{\nu, \mathcal{C}}$ a distribuição de σ_t que realizou a dinâmica censurada até tempo t começando com distribuição inicial ν .

Definição 5.14 *Dizemos que uma lei de probabilidade ν em S_N é crescente se $\sigma \mapsto \nu(\sigma)$ é uma função crescente.*

Proposição 5.15 (Desigualdade censuradora) *Se ν é crescente, então para todo $t \geq 0$,*

$$\|P_t^{\nu, \mathcal{C}} - \pi\|_{TV} \geq \|P_t^\nu - \pi\|_{TV}.$$

Prova. Dados $x \in \{1, \dots, N-1\}$ e $\sigma \in S_N$, definimos

$$\sigma_x^\bullet := \{\xi \in S_N; \forall y \notin \{x, x+1\}, \xi(y) = \sigma(y)\}.$$

O conjunto σ_x^\bullet contém dois elementos (sendo σ um deles) $\sigma_x^+ \geq \sigma_x^-$, que são obtidos respectivamente ordenando e desordenando $\sigma(x)$ e $\sigma(x+1)$. Dada ν uma medida de probabilidade em S_N , definimos $\theta_x(\nu)$, a medida “atualizada em x ” por

$$\theta_x(\nu)(\sigma) := \frac{\nu(\sigma_x^\bullet)}{2}.$$

O operador θ_x descreve como a lei de σ_t é mudada quando o processo do relógio toca em x .

Afirmção 1: Se ν é crescente, então $\theta_x(\nu)$ também é, e além disso, $\nu \succeq \theta_x(\nu)$.

De fato, é fácil verificar que se $\sigma \geq \xi$, então $\sigma_x^+ \geq \xi_x^+$ e $\sigma_x^- \geq \xi_x^-$. Logo, se ν é crescente, temos que

$$\nu(\sigma_x^\bullet) = \nu(\sigma_x^+) + \nu(\sigma_x^-) \geq \nu(\xi_x^+) + \nu(\xi_x^-) = \nu(\xi_x^\bullet),$$

e portanto, $\theta_x(\nu)$ é crescente.

Agora, seja g uma função crescente definida no espaço de estados Ω_N com valores em \mathbb{R} , e suponhamos que ν é crescente. Então, $\nu(\sigma_x^+) \geq \nu(\sigma_x^-)$ e $g(\sigma_x^+) \geq g(\sigma_x^-)$. Assim,

$$\begin{aligned} 0 &\leq [g(\sigma_x^+) - g(\sigma_x^-)] [\nu(\sigma_x^+) - \nu(\sigma_x^-)] \\ &= [g(\sigma_x^+)\nu(\sigma_x^+) + g(\sigma_x^-)\nu(\sigma_x^-)] - [g(\sigma_x^+)\nu(\sigma_x^-) + g(\sigma_x^-)\nu(\sigma_x^+)]. \end{aligned}$$

Logo,

$$[g(\sigma_x^+)\nu(\sigma_x^-) + g(\sigma_x^-)\nu(\sigma_x^+)] \leq [g(\sigma_x^+)\nu(\sigma_x^+) + g(\sigma_x^-)\nu(\sigma_x^-)].$$

Portanto,

$$\begin{aligned} &[g(\sigma_x^+)\nu(\sigma_x^+) + g(\sigma_x^-)\nu(\sigma_x^-)] \\ &= \frac{[g(\sigma_x^+)\nu(\sigma_x^+) + g(\sigma_x^-)\nu(\sigma_x^-)] + [g(\sigma_x^+)\nu(\sigma_x^+) + g(\sigma_x^-)\nu(\sigma_x^-)]}{2} \\ &\geq \frac{[g(\sigma_x^+)\nu(\sigma_x^+) + g(\sigma_x^-)\nu(\sigma_x^-)] + [g(\sigma_x^+)\nu(\sigma_x^-) + g(\sigma_x^-)\nu(\sigma_x^+)]}{2} \\ &= \frac{[g(\sigma_x^+) + g(\sigma_x^-)] [\nu(\sigma_x^+) + \nu(\sigma_x^-)]}{2} \\ &= [g(\sigma_x^+) + g(\sigma_x^-)] \theta_x(\nu)(\sigma) \\ &= g(\sigma_x^+)\theta_x(\nu)(\sigma) + g(\sigma_x^-)\theta_x(\nu)(\sigma). \end{aligned}$$

Agora, somando em $\sigma \in S_N$ e dividindo por dois, temos

$$\nu(g) = \sum_{\sigma \in S_N} g(\sigma) \nu(\sigma) \geq \sum_{\sigma \in S_N} g(\sigma) \theta_x(\nu)(g(\sigma)) = \theta_x(\nu)(g).$$

Como g é arbitrária, temos que

$$\nu \succeq \theta_x(\nu).$$

Afirmção 2: Seja ν uma medida de probabilidade crescente em S_N . Então, para todo $t \geq 0$, P_t^ν também é crescente e para qualquer esquema censurador \mathcal{C} , $P_t^{\nu, \mathcal{C}}$ é crescente.

Seja ν uma probabilidade crescente e σ_t^ν a trajetória da cadeia de Markov obtida com a construção gráfica. Por definição, temos $P_t^\nu = \mathbb{P}[\sigma_t^\nu \in \cdot]$. Seja \mathcal{N} o número de atualizações que ocorrem antes do tempo t e $X_1, \dots, X_{\mathcal{N}}$ a sequência de vértices que tocaram no processo do relógio (com repetições). Então, a lei de probabilidade $\mathbb{P}[\sigma_t^\nu \in \cdot | \mathcal{T}]$, sabendo que o processo do relógio é dado por

$$\theta_{X_{\mathcal{N}}} \circ \dots \circ \theta_{X_1}(\nu),$$

é crescente de acordo com a Afirmção 1.

A monotonicidade é então preservada quando fazemos a média com respeito à \mathcal{T} . O resultado é análogo para a dinâmica censurada.

Afirmção 3: As atualizações preservam dominação estocástica no sentido de que se $\nu_1 \succeq \nu_2$, então $\theta_x(\nu_1) \succeq \theta_x(\nu_2)$.

A veracidade da afirmação segue da definição de θ_x .

Afirmção 4: Se ν_1 tem densidade crescente e $\nu_1 \preceq \nu_2$, então

$$\|\nu_1 - \pi\|_{TV} \leq \|\nu_2 - \pi\|_{TV}.$$

De fato, definamos $A := \{\sigma; \nu_1(\sigma) \geq \pi(\sigma) = (n!)^{-1}\}$.

Como ν_1 tem densidade crescente, A é um evento crescente e

$$\|\nu_1 - \pi\|_{TV} = \nu_1(A) - \pi(A) \leq \nu_2(A) - \pi(A) = \|\nu_2 - \pi\|_{TV}.$$

Afirmção 5: Seja ν_0 uma probabilidade crescente em S_N e $k \in \mathbb{N}$. Dados $(x_1, \dots, x_k) \in \{1, \dots, N-1\}^k$ (repetições são permitidas) e $j \in \{1, \dots, k\}$,

denotemos por ν_1 a medida obtida realizando sucessivas atualizações nos sítios x_1, \dots, x_k e ν_2 a medida obtida realizando a mesma sequência de atualizações mas omitindo a atualização em x_j (isto é, $x_1, \dots, x_{j-1}, x_{j+1}, x_{j+2}, \dots, x_k$). Então,

$$\|\nu_1 - \pi\|_{TV} \geq \|\nu_2 - \pi\|_{TV}.$$

Observação: O resultado permanece válido se várias atualizações são omitidas ao invés de uma.

De fato, sem perda de generalidade podemos considerar que $j = 1$ pois a lei obtida depois de realizar as $j - 1$ primeiras atualizações tem densidade crescente.

Seja ν_0' a medida obtida depois de atualizar x_1 . Da Afirmação 1, temos que $\nu_0' \preceq \nu_0$. Como a monotonicidade é preservada pelas atualizações em (x_2, \dots, x_k) , a Afirmação 3 garante que $\nu_2 \preceq \nu_1$. Além disso, pela Afirmação 1, ambas possuem densidades crescentes. Portanto, a Afirmação 4 garante o resultado.

O caso de várias omissões pode ser provado usando o princípio de indução.

Prova da proposição: Finalmente, vamos mostrar a desigualdade censuradora. Na nossa dinâmica, em tempo t , o conjunto das atualizações realizadas é aleatório e dado pelo processo do relógio \mathcal{T} restrito a $[0, t]$, logo a Afirmação 5 não pode ser aplicada diretamente. Todavia, para uma realização fixa de \mathcal{T} , podemos aplicar a Afirmação 5 condicionada a \mathcal{T} . Com efeito, sejam $P_t^{\mathcal{T}} := \mathbb{P}[\sigma_t^\nu \in \cdot | \mathcal{T}]$ a lei de σ obtida depois de fazer as atualizações correspondentes a \mathcal{T} , e $P_t^{\mathcal{T}, \mathcal{C}} := \mathbb{P}^{\mathcal{C}}[\sigma_t^\nu \in \cdot | \mathcal{T}]$ a lei obtida depois de fazer as atualizações permitidas pelo esquema censurador. Ambas medidas de probabilidade são *i)* crescentes, e pela Afirmação 5, *ii)* $P^{\mathcal{T}} \succeq P^{\mathcal{T}, \mathcal{C}}$.

Essas duas propriedades são conservadas quando fazemos a média com respeito a \mathcal{T} . Logo, $P_t^\nu \succeq P_t^{\nu, \mathcal{C}}$ e portanto a Afirmação 4 garante o resultado. \square

A desigualdade censuradora tem sido usada em uma variedade de contextos para cotar tempos de mistura para cadeias de Markov. A estratégia geralmente é encontrar um esquema censurador que nos permita controlar melhor a dinâmica sem reduzir muito sua velocidade (LACOIN, 2016, p. 15).

5.2.6

Projeção e monotonia

Agora mostraremos que se ν é uma medida de probabilidade crescente em S_N , então suas projeções têm densidades crescentes com respeito às projeções da medida invariante π .

Definição 5.16 Definimos o semi-esqueleto de uma permutação $\sigma \in S_N$ por $\hat{\sigma} : \{0, \dots, N\} \times \{0, \dots, K\} \rightarrow \mathbb{R}$ que faz

$$\hat{\sigma}(x, j) := \tilde{\sigma}(x, x_j), \forall 0 \leq j \leq K,$$

onde $x_j := \lceil jN/K \rceil$. Chamaremos de \hat{S}_N o conjunto de semi-esqueletos admissíveis, isto é, a imagem de S_N por $\hat{\sigma}$.

Definição 5.17 Definimos o esqueleto de uma permutação $\sigma \in S_N$ por $\bar{\sigma} : \{0, \dots, K\}^2 \rightarrow \mathbb{R}$ que faz

$$\bar{\sigma}(i, j) := \tilde{\sigma}(x_i, x_j), \forall 0 \leq i, j \leq K.$$

Chamaremos de $\bar{S}_N := \{\bar{\sigma} | \sigma \in S_N\}$ o conjunto de esqueletos admissíveis.

Definição 5.18 Equiparemos \bar{S}_N e \hat{S}_N com a ordem natural, isto é,

$$\bar{\sigma} \geq \bar{\sigma}' \Leftrightarrow \bar{\sigma}(i, j) \geq \bar{\sigma}'(i, j), \forall 0 \leq i, j \leq K,$$

e

$$\hat{\sigma} \geq \hat{\sigma}' \Leftrightarrow \hat{\sigma}(x, j) \geq \hat{\sigma}'(x, j), \forall 0 \leq x \leq N, 0 \leq j \leq K.$$

Notemos que quando $N = 52$ e $K = 2$, o semi-esqueleto codifica as posições das cartas vermelhas no baralho.

De fato, se $N = 52$ e $K = 2$, então $x_1 = 26$. Assim,

$$\hat{\sigma}(x, 1) = \tilde{\sigma}(x, 26) = \sum_{z=1}^x \mathbf{1}_{\{\sigma(z) \leq 26\}} - \frac{x}{2}.$$

Já o esqueleto indica o número de cartas vermelhas na primeira metade do baralho, pois

$$\bar{\sigma}(1, 1) + 13 = \tilde{\sigma}(26, 26) + 13 = \sum_{z=1}^{26} \mathbf{1}_{\{\sigma(z) \leq 26\}}.$$

Definição 5.19 Seja ν uma medida de probabilidade em S_N . Chamaremos por $\bar{\nu}$ a medida imagem em \bar{S}_N de ν pela projeção do esqueleto e por $\hat{\nu}$ a medida imagem do semi-esqueleto. Escreveremos $\bar{\nu}_{i,j}$ para a medida imagem de ν pela projeção $\sigma \mapsto \bar{\sigma}(i, j)$. Em particular $\bar{\pi}$ e $\bar{\pi}_{i,j}$ denotam as projeções da medida invariante.

Proposição 5.20 (Preservação da monotonia pela projeção) As seguintes afirmações são verdadeiras:

- (i) Sejam $\bar{\sigma}^1, \bar{\sigma}^2 \in \bar{S}_N$. Se $\bar{\sigma}^1 \geq \bar{\sigma}^2$, então $\pi(\cdot | \bar{\sigma} = \bar{\sigma}^1) \succeq \pi(\cdot | \bar{\sigma} = \bar{\sigma}^2)$.
- (ii) Dados $i, j \in \{1, \dots, K\}$ e $z_1 \geq z_2$ dois valores admissíveis para $\bar{\sigma}(i, j)$, temos $\pi(\cdot | \bar{\sigma}(i, j) = z_1) \succeq \pi(\cdot | \bar{\sigma}(i, j) = z_2)$.
- (iii) Se ν é uma medida de probabilidade crescente em S_N , então a densidade $\bar{\nu}/\bar{\pi}$ é uma função crescente em \bar{S}_N .
- (iv) Se ν é uma medida de probabilidade crescente em S_N , então $\bar{\nu}_{i,j}/\bar{\pi}_{i,j}$ é uma função crescente no conjunto de valores admissíveis para $\bar{\sigma}(i, j)$.

Prova. Dividiremos a prova em três afirmações.

Afirmção 1: Se $\hat{\sigma}^1, \hat{\sigma}^2 \in \hat{S}_N$ e $\hat{\sigma}^1 \geq \hat{\sigma}^2$, então

$$\pi(\cdot | \hat{\sigma} = \hat{\sigma}^1) \succeq \pi(\cdot | \hat{\sigma} = \hat{\sigma}^2).$$

Notemos que a informação que $\hat{\sigma}$ carrega é exatamente o valor dos conjuntos

$$\sigma^{-1}(\{x_{i-1} + 1, \dots, x_i\}), i \in \{1, \dots, K\}.$$

Para cada $i \in \{0, \dots, K\}$ esse conjunto é dado por

$$\{x \in \{1, \dots, N\} | \hat{\sigma}(x, i+1) - \hat{\sigma}(x-1, i+1) - \hat{\sigma}(x, i) + \hat{\sigma}(x-1, i) > 0\},$$

pois

$$\begin{aligned} & \hat{\sigma}(x, i+1) - \hat{\sigma}(x-1, i+1) - \hat{\sigma}(x, i) + \hat{\sigma}(x-1, i) \\ &= [\hat{\sigma}(x, i+1) - \hat{\sigma}(x-1, i+1)] - [\hat{\sigma}(x, i) - \hat{\sigma}(x-1, i)] \\ &= \left(\mathbf{1}_{\{\sigma(x) \leq x_{i+1}\}} - \frac{x_{i+1}}{N} \right) - \left(\mathbf{1}_{\{\sigma(x) \leq x_i\}} - \frac{x_i}{N} \right) \\ &= \mathbf{1}_{\{\sigma(x) \leq x_{i+1}\}} - \mathbf{1}_{\{\sigma(x) \leq x_i\}} - \underbrace{\frac{x_{i+1} - x_i}{N}}_{<1} \\ & \begin{cases} > 0 & \text{se } x \in \{x_{i-1} + 1, \dots, x_i\}; \\ \leq 0 & \text{caso contrário.} \end{cases} \end{aligned}$$

Porém não temos informação sobre a ordem das cartas cujos marcadores pertencem a $\{x_{i-1} + 1, \dots, x_i\}$. Assim, para cada $\xi \in \hat{S}_N$, definimos a bijeção natural

$$\begin{aligned} \otimes_{i=1}^K S_{\Delta x_i} &\rightarrow \{\sigma \in S_N | \hat{\sigma} = \xi\}, \\ (\sigma_1, \dots, \sigma_K) &\mapsto \sigma_{\xi}^{(\sigma_1, \dots, \sigma_K)}, \end{aligned} \quad (5.4)$$

onde $\Delta x_i := x_i - x_{i-1}$, e a permutação $\sigma_{\xi}^{(\sigma_1, \dots, \sigma_K)}$ é definida como a única em $\{\sigma \in S_N | \hat{\sigma} = \xi\}$ tal que para todo $i \in \{1, \dots, K\}$, as cartas com marcadores $\{x_{i-1}, \dots, x_i\}$ aparecem na ordem especificada por σ_i ,

$$\begin{aligned} \forall a, b \in \{x_{i-1} + 1, \dots, x_i\} \\ \sigma^{-1}(a) \leq \sigma^{-1}(b) \Leftrightarrow \sigma_i^{-1}(a - x_{i-1}) \leq \sigma_i^{-1}(b - x_{i-1}). \end{aligned}$$

A aplicação (5.4), em termos de superfícies discretas, tem a seguinte expressão: Para todo $y \in \{x_{i-1}, \dots, x_i\}$

$$\begin{aligned} \tilde{\sigma}_{\xi}^{(\sigma_1, \dots, \sigma_K)}(x, y) &= \frac{y - x_{i-1}}{\Delta x_i} \xi(x, i) + \frac{x_i - y}{\Delta x_i} \xi(x, i - 1) \\ &\quad + \tilde{\sigma}_i \left(\xi(x, i) - \xi(x, i - 1) + \frac{x \Delta x_i}{N}, y - x_{i-1} \right). \end{aligned}$$

Se $\xi \geq \xi'$ são dois semi-esqueletos admissíveis, então para qualquer $(\sigma_1, \dots, \sigma_K)$,

$$\tilde{\sigma}_{\xi}^{(\sigma_1, \dots, \sigma_K)} \geq \tilde{\sigma}_{\xi'}^{(\sigma_1, \dots, \sigma_K)}.$$

De fato, sejam $x, y \in \{x_{i-1}, \dots, x_i\}$,

$$\begin{aligned} \tilde{\sigma}_{\xi}^{(\sigma_1, \dots, \sigma_K)}(x, y) - \tilde{\sigma}_{\xi'}^{(\sigma_1, \dots, \sigma_K)}(x, y) &= \frac{y - x_{i-1}}{\Delta x_i} [\xi(x, i) - \xi'(x, i)] \\ &\quad + \frac{x_i - y}{\Delta x_i} [\xi(x, i - 1) - \xi'(x, i - 1)] \\ &\quad + \tilde{\sigma}_i \left(\xi(x, i) - \xi(x, i - 1) + \frac{x \Delta x_i}{N}, y - x_{i-1} \right) \\ &\quad - \tilde{\sigma}_i \left(\xi'(x, i) - \xi'(x, i - 1) + \frac{x \Delta x_i}{N}, y - x_{i-1} \right) \\ &\geq 0. \end{aligned}$$

Portanto, a medida uniforme em $\otimes_{i=1}^K S_{\Delta x_i}$ induz um acoplamento monótono que garante a veracidade de nossa afirmação.

Afirmção 2: Se $\bar{\sigma}^1, \bar{\sigma}^2 \in \bar{S}_N$ e $\bar{\sigma}^1 \geq \bar{\sigma}^2$, então $\pi(\cdot | \bar{\xi} = \bar{\sigma}^1) \succeq \pi(\cdot | \bar{\xi} = \bar{\sigma}^2)$.

De fato, sejam $\bar{\sigma}^1 \geq \bar{\sigma}^2$ e para cada $i \in \{1, 2\}$ definamos

$$\hat{S}^i := \{\xi \in \hat{S}_N | \bar{\xi} = \bar{\sigma}_i\}.$$

Mostraremos que \hat{S}^i tem um elemento maximal ξ_{max}^i e que $\xi_{max}^1 \geq \xi_{max}^2$.

Com efeito, seja $\sigma \in S_N$ tal que $\hat{\sigma} \in \hat{S}^1$ e seja σ' a permutação obtida ordenando os elementos de cada intervalo $\{x_{i-1} + 1, \dots, x_i\}$, para todo $i \in \{1, \dots, K\}$, isto é, a única permutação que satisfaz

$$\forall i \in \{1, \dots, K\}, \sigma'(\{x_{i-1} + 1, \dots, x_i\}) = \sigma(\{x_{i-1} + 1, \dots, x_i\}),$$

e

$$\begin{aligned} \forall i \in \{1, \dots, K\}, \forall y, z \in \{x_{i-1} + 1, \dots, x_i\}, \\ y \leq z \Rightarrow \sigma'(y) \leq \sigma'(z). \end{aligned}$$

Assim, para todo $i \in \{1, \dots, K\}$, $j \in \{0, \dots, K\}$ e $x \in \{x_{i-1}, \dots, x_i\}$, temos

$$\hat{\sigma}'(x, j) := \min \left(\frac{N - x_j}{N} (x - x_{i-1}) + \bar{\sigma}(i - 1, j), \frac{x_j}{N} (x_i - x) + \bar{\sigma}(i, j) \right). \quad (5.5)$$

Portanto, $\hat{\sigma}'$ é maximal em \hat{S}^1 . Analogamente, encontramos o elemento maximal em \hat{S}^2 , e a expressão (5.5) garante que $\xi_{max}^1 \geq \xi_{max}^2$.

Sejam $(\xi_t^1)_{t \geq 0}$ e $(\xi_t^2)_{t \geq 0}$ as cadeias de Markov em \hat{S}^i começando em ξ_{max}^1 e ξ_{max}^2 , respectivamente, que seguem a construção gráfica de U e \mathcal{T} , mas ignoram as atualizações em x_i , $i = 1, \dots, K - 1$. Em outras palavras, a censura cancela as atualizações que tiram ξ_t^i de \hat{S}^i .

As cadeias de Markov $(\xi_t^1)_{t \geq 0}$ e $(\xi_t^2)_{t \geq 0}$ são irredutíveis, pois dado $\xi \in \hat{S}^1$, podemos encontrar σ tal que $\hat{\sigma} = \xi$. Assim, partindo de σ , podemos construir um caminho de transições que nos leva a σ' (o elemento maximal descrito em (5.5)) e que não usa nenhum dos τ_{x_i} . Projetando esse caminho com o semi-esqueleto, encontramos um caminho de transições não censuradas de ξ até ξ_{max}^1 . Como $(\xi_t^i)_{t \geq 0}$ são cadeias de Markov refletidas, suas respectivas medidas de equilíbrio são $\hat{\pi}(\cdot | \bar{\xi} = \bar{\sigma}^i)$, $i = 1, 2$, que são as medidas uniformes em \hat{S}^i . A ordem das condições iniciais e a preservação de ordem induzida pela construção gráfica garantem que $\forall t \geq 0, \xi_t^1 \geq \xi_t^2$.

Com esse acoplamento monótono entre os dois processos, o Lema 5.10 garante a veracidade da afirmação.

Afirmção 3: Sejam $\bar{\sigma}^1, \bar{\sigma}^2 \in \bar{S}_N$. Se $\bar{\sigma}^1 \geq \bar{\sigma}^2$, então $\pi(\cdot|\bar{\sigma} = \bar{\sigma}^1) \succeq \pi(\cdot|\bar{\sigma} = \bar{\sigma}^2)$.

Seja f uma função crescente em S_N e definamos \hat{f} em \hat{S}_N por

$$\hat{f}(\xi) = \pi(f(\sigma)|\hat{\sigma} = \xi).$$

Pela Afirmção 1, \hat{f} é uma função crescente em \hat{S}_N . Assim, pela Afirmção 2, se $\bar{\sigma}_1 \geq \bar{\sigma}_2$,

$$\begin{aligned} \pi(f(\sigma)|\bar{\sigma} = \bar{\sigma}_1) &= \hat{\pi}(\hat{f}(\xi)|\bar{\xi} = \bar{\sigma}_1) \\ &\geq \hat{\pi}(\hat{f}(\xi)|\bar{\xi} = \bar{\sigma}_2) \\ &= \pi(f(\sigma)|\bar{\sigma} = \bar{\sigma}_2). \end{aligned}$$

O resultado segue pelo Lema 5.9.

Afirmção 4: Dados $i, j \in \{1, \dots, K\}$ e $z_1 \geq z_2$ dois valores admissíveis para $\bar{\sigma}(i, j)$, temos $\pi(\cdot|\bar{\sigma}(i, j) = z_1) \succeq \pi(\cdot|\bar{\sigma}(i, j) = z_2)$.

De fato, a prova do item **(i)** também é válida se trocarmos a grade $(x_i, x_j)_{i,j=1}^{K-1}$ por uma grade assimétrica $(x_i, y_j)_{i,j=1}^{K-1}$, e em qualquer caso os valores de x_i não possuem nenhum papel. Assim o item **(ii)** corresponde ao caso do item **(i)** em que $K = 2$.

Afirmção 5: Se ν é uma medida de probabilidade crescente em S_N , então a densidade $\bar{\nu}/\bar{\pi}$ é uma função crescente em \bar{S}_N .

Apenas devemos integrar a função ν/π contra a ordenação do item **(i)**. Com efeito, como ν é crescente, ν/π também é. Assim, pela dominação estocástica, se $\bar{\sigma}^1 \geq \bar{\sigma}^2$, temos

$$\begin{aligned} (\bar{\nu}/\bar{\pi})(\bar{\sigma}^1) &= \pi(\nu/\pi|\bar{\sigma} = \bar{\sigma}^1) \\ &\geq \pi(\nu/\pi|\bar{\sigma} = \bar{\sigma}^2) \\ &= (\bar{\nu}/\bar{\pi})(\bar{\sigma}^2). \end{aligned}$$

Afirmção 6: Se ν é uma medida de probabilidade crescente em S_N , então $\bar{\nu}_{i,j}/\bar{\pi}_{i,j}$ é uma função crescente no conjunto de valores admissíveis para $\bar{\sigma}(i, j)$.

Agora, devemos integrar a função ν/π contra a ordenação do item (ii).

□

5.3

Mais algumas ferramentas

5.3.1

A conexão com a equação do calor

Existe uma conexão entre o embaralhamento de transposições adjacentes e difusões. Para vermos isso, basta olharmos o movimento de apenas uma carta, que corresponde a um passeio aleatório simétrico com saltos para o vizinho mais próximo no conjunto $\{1, \dots, N\}$. Também podemos perceber essa conexão quando olhamos para a evolução da esperança de $\tilde{\sigma}_t(x, y)$. Lembremos que $\tilde{\sigma}_t(x, y)$ só pode dar uma passo para baixo quando $\tilde{\sigma}_t(\cdot, y)$ apresenta um máximo local em x , e para cima quando apresenta um mínimo local, cada caso com taxa um. Assim,

$$\begin{aligned} \partial_t \mathbb{E} [\tilde{\sigma}_t(x, y)(t)] &= \mathbb{E} \left[\mathbf{1}_{\{\tilde{\sigma}_t(x, y) > \max(\tilde{\sigma}_t(x-1, y), \tilde{\sigma}_t(x+1, y))\}} \right] \\ &\quad - \mathbb{E} \left[\mathbf{1}_{\{\tilde{\sigma}_t(x, y) < \min(\tilde{\sigma}_t(x-1, y), \tilde{\sigma}_t(x+1, y))\}} \right] \\ &= \mathbb{E} [\tilde{\sigma}_t(x-1, y) + \tilde{\sigma}_t(x+1, y) - 2\tilde{\sigma}_t(x, y)], \end{aligned}$$

Portanto, $f(x, y, t) := \mathbb{E} [\tilde{\sigma}_t(x, y)]$ é a solução da equação do calor discreta unidimensional

$$\begin{cases} \partial_t f = \Delta_x f \text{ em } \{1, \dots, N-1\} \times \mathbb{R}_+, \\ f(0, t) = f(N, t) = 0, \\ f(x, y, 0) = \tilde{\sigma}_0(x, y), \end{cases} \quad (5.6)$$

onde Δ_x denota o Laplaciano discreto sobre a coordenada x , e y está fixado.

Lema 5.21 Para toda $\sigma_0 \in S_N$ e $t \geq 0$, temos

$$\max_{x \in \{0, \dots, N\}} \mathbb{E} [\tilde{\sigma}_t(x, y)] \leq 4 \min(y, N-y) e^{-\lambda_N t}, \quad (5.7)$$

onde

$$\lambda_N := 2 \left(1 - \cos \left(\frac{\pi}{N} \right) \right) = \frac{\pi^2}{N^2} (1 + o(1)).$$

Em particular,

$$\max_{(x, y) \in \{0, \dots, N\}^2} \mathbb{E} [\tilde{\sigma}_t(x, y)] \leq 2N e^{-\lambda_N t}.$$

Além disso, para $\sigma_0 = e$, temos

$$\mathbb{E} [\tilde{\sigma}_t(x, y)] \geq \frac{\min(y, N-y)}{\pi} \sin \left(\frac{\pi x}{N} \right) e^{-\lambda_N t}. \quad (5.8)$$

Prova. A prova que segue encontra-se em [11], Lema 4.1.

Fixemos $y \in \{0, \dots, N\}$. Podemos calcular a solução de (5.6) através da decomposição de Fourier na base das autofunções $(u_i)_{i=1}^{N-1}$ de Δ_x que são dadas por

$$u_i : x \mapsto \sqrt{\frac{2}{N}} \sin\left(\frac{xi\pi}{N}\right),$$

cujos autovalores associados são $-\lambda_{N,i}$, onde

$$\lambda_{N,i} := 2 \left(1 - \cos\left(\frac{i\pi}{N}\right)\right).$$

Portanto,

$$f(x, y, t) = \frac{2}{N} \sum_{i=1}^{N-1} a_i(\tilde{\sigma}_0(\cdot, y)) e^{-\lambda_{N,i}t} \sin\left(\frac{xi\pi}{N}\right), \quad (5.9)$$

onde o coeficiente de Fourier a_i é dado por

$$a_i(\tilde{\sigma}_0(\cdot, y)) := \sum_{x=1}^{N-1} \tilde{\sigma}_0(x, y) \sin\left(\frac{xi\pi}{N}\right).$$

Lembremos que $\tilde{\sigma}_0(x, y) = \sum_{z=1}^x \mathbf{1}_{\{\sigma_0(z) \leq y\}} - \frac{xy}{N}$.

Sabemos que há apenas y cartas com marcadores menores ou igual a y .

Logo, $\sum_{z=1}^x \mathbf{1}_{\{\sigma_0(z) \leq y\}} \leq y$. Portanto,

$$\tilde{\sigma}_0(x, y) \leq y. \quad (5.10)$$

Além disso,

$$\tilde{\sigma}_0(x, y) \leq x - \frac{xy}{N} = x \left(1 - \frac{y}{N}\right) = \left(\frac{x}{N}\right) (N - y) \leq N - y, \quad (5.11)$$

$$\tilde{\sigma}_0(x, y) \geq -\frac{xy}{N} = -\left(\frac{x}{N}\right) y \geq -y, \quad (5.12)$$

e como há apenas $N - y$ cartas com marcadores maiores que y ,

$$\begin{aligned}
 \tilde{\sigma}_0(x, y) &= \sum_{z=1}^x \left(1 - \mathbf{1}_{\{\sigma_0(z) > y\}}\right) - \frac{xy}{N} \\
 &= x - \sum_{z=1}^x \mathbf{1}_{\{\sigma_0(z) > y\}} - \frac{xy}{N} \\
 &\geq x - (N - y) - \frac{xy}{N} = -(N - y) + x \left(1 - \frac{y}{N}\right) \\
 &= -(N - y) + x \left(\frac{N - y}{N}\right) \\
 &\geq -(N - y).
 \end{aligned} \tag{5.13}$$

Assim, por (5.10), (5.11), (5.12) e (5.13), temos que

$$|\tilde{\sigma}_0(x, y)| \leq \min(y, N - y), \forall x \in \{0, \dots, N\},$$

Logo, os coeficientes de Fourier satisfazem

$$|a_i(\tilde{\sigma}(\cdot, y))| \leq \min(y, N - y)N, \forall i \in \{1, \dots, N - 1\}.$$

Ademais, como $\cos\left(\frac{i\pi}{N}\right) \leq \cos\left(\frac{\pi}{N}\right)$ para todo $i \in \{1, \dots, N - 1\}$, temos que $\lambda_{N,i} \geq i\lambda_N$ para todo $i \in \{1, \dots, N - 1\}$, e portanto, por (5.9),

$$|f(x, y, t)| \leq 2 \min(y, N - y) \sum_{i=1}^{N-1} e^{-\lambda_N t} = \frac{2 \min(y, N - y) e^{-\lambda_N t}}{1 - e^{-\lambda_N t}}.$$

Quando $e^{-\lambda_N t} \leq 1/2$, temos (5.7).

Por outro lado, se $e^{-\lambda_N t} \geq 1/2$, como $|\tilde{\sigma}_t(x, y)| \leq \min(y, N - y)$, temos

$$|f(x, t)| \leq \min(y, N - y) < 2 \min(y, N - y) \leq 4 \min(y, N - y) e^{-\lambda_N t}.$$

Portanto (5.7) continua valendo.

Para mostrarmos (5.8), notemos que quando

$$\begin{aligned}
 \tilde{\sigma}_0(x, y) &= \tilde{e}(x, y) = \sum_{z=1}^x \mathbf{1}_{\{z \leq y\}} - \frac{xy}{N} = \min \{x, y\} - \frac{xy}{N} \\
 &= \min \left\{ x - \frac{xy}{N}, y - \frac{xy}{N} \right\} \\
 &= \min \left\{ x \left(1 - \frac{y}{N} \right), \frac{yN - xy}{N} \right\} \\
 &= \min \left\{ x \left(\frac{N - y}{N} \right), (N - x) \frac{y}{N} \right\} \\
 &\geq \min \left(x \frac{\min(y, N - y)}{N}, (N - x) \frac{\min(y, N - y)}{N} \right) \\
 &= \frac{\min(y, N - y)}{\pi} \min \left(\frac{x\pi}{N}, \pi - \frac{x\pi}{N} \right).
 \end{aligned}$$

Como $\sin u \leq \min(u, \pi - u)$ para todo $u \in [0, \pi]$, temos que

$$\forall x \in \{1, \dots, N - 1\}, \quad \tilde{\sigma}_0(x, y) \geq \frac{\min(y, N - y)}{\pi} \sin \left(\frac{x\pi}{N} \right). \quad (5.14)$$

Agora, podemos deduzir (5.8) considerando a solução de (5.6) no tempo t para ambos os lados de (5.14) através da monotonicidade da solução da equação do calor com as condições iniciais dadas. \square

5.3.2

A cota superior de Wilson

Proposição 5.22 Para todo N suficientemente grande e para todo $\epsilon > 0$

$$d^N(t) \leq 10Ne^{-t\lambda_N},$$

onde $\lambda_N := 2(1 - \cos(\pi/N))$.

Prova. A prova encontra-se em [17], Teorema 10. \square

5.3.3

Apagando os marcadores e decompondo o procedimento de mistura

Suponhamos que trocamos os marcadores das cartas de modo que cada carta cujo marcador pertença a $\{x_{i-1} + 1, \dots, x_i\}$, $i = 1, \dots, K$ agora recebe o marcador i . Para visualizarmos o que está acontecendo, tomemos $K = 4$ e $N = 52$. Assim, estamos diferenciando as cartas apenas pelos naipes: copas, ouro, espadas e paus. O baralho remarcado é descrito pelo semi-esqueleto $\hat{\sigma}$.

Intuitivamente, para que $(\sigma_t)_{t \geq 0}$ alcance o equilíbrio é necessário que:

- (i) o semi-esqueleto $(\hat{\sigma}_t)_{t \geq 0}$ esteja próximo da sua distribuição de equilíbrio;
- (ii) condicionando a cada semi-esqueleto, a ordem das cartas com marcadores i esteja perto de estarem uniformemente distribuídas.

Nesta subseção, mostraremos essas afirmações de forma mais rigorosa.

Considere a seguinte transformação das medidas que fará com que cartas cujos marcadores pertencem a $\{x_{i-1} + 1, \dots, x_i\}$ sejam indistinguíveis.

Definição 5.23 *Seja \tilde{S}_N o maior subgrupo de S_N que deixa todos os conjuntos $\{x_{i-1} + 1, \dots, x_i\}$ invariantes.*

É fácil ver que \tilde{S}_N é isomorfo a $\otimes_{i=1}^K S_{\Delta x_i}$.

Definição 5.24 *Seja ν uma medida de probabilidade em S_N . Definimos $\tilde{\nu}$ como*

$$\tilde{\nu}(\sigma) := \frac{1}{\prod_{i=1}^K (\Delta x_i)!} \sum_{\tilde{\sigma} \in \tilde{S}_N} \nu(\tilde{\sigma} \circ \sigma).$$

Notemos que o semi-esqueleto de σ é mantido invariante pela composição à direita por um elemento de \tilde{S}_N . Logo, temos

$$\tilde{\nu}(\sigma) := \frac{1}{|\tilde{S}_N|} \hat{\nu}(\hat{\sigma}). \quad (5.15)$$

Lema 5.25 *Para todas as leis de probabilidade ν em S_N , temos*

$$\|\tilde{\nu} - \pi\|_{TV} = \|\hat{\nu} - \hat{\pi}\|_{TV}.$$

Trivialmente,

$$\|\nu - \pi\|_{TV} \leq \|\hat{\nu} - \hat{\pi}\|_{TV} + \|\nu - \tilde{\nu}\|_{TV}.$$

Prova. Primeiro, vejamos que

$$2\|\tilde{\nu} - \pi\|_{TV} = \sum_{\sigma \in S_N} |\tilde{\nu}(\sigma) - \pi(\sigma)| = \sum_{\xi \in \tilde{S}_N} \sum_{\{\sigma \in S_N | \hat{\sigma} = \xi\}} |\tilde{\nu}(\sigma) - \pi(\sigma)|.$$

Por (5.15), $\tilde{\nu}$ é constante em $\{\sigma \in S_N | \hat{\sigma} = \xi\}$ e portanto

$$\begin{aligned}
 2\|\tilde{\nu} - \pi\|_{TV} &= \sum_{\xi \in \hat{S}_N} \left| \sum_{\{\sigma \in S_N | \hat{\sigma} = \xi\}} \tilde{\nu}(\sigma) - \pi(\sigma) \right| \\
 &= \sum_{\xi \in \hat{S}_N} \left| \sum_{\{\sigma \in S_N | \hat{\sigma} = \xi\}} \nu(\sigma) - \pi(\sigma) \right| \\
 &= \sum_{\xi \in \hat{S}_N} |\hat{\nu}(\xi) - \hat{\pi}(\xi)| \\
 &= 2\|\hat{\nu} - \hat{\pi}\|_{TV}.
 \end{aligned}$$

□

5.4

A cota superior

Nesta seção, mostraremos que o tempo de mistura do embaralhamento de transposições adjacentes em tempo contínuo é menor ou igual a $\frac{1}{2\pi^2} N^2 \log N$.

Para obtermos tal cota superior, necessitaremos de enunciar três proposições que necessitam da seguinte notação:

Sejam $t_1 := \frac{N^2}{2\pi^2}(\delta/3) \log N$, $t_2 := \frac{N^2}{2\pi^2}(1+2\delta/3) \log N$, $t_3 := \frac{N^2}{2\pi^2}(1+\delta) \log N$ e $K := \lceil 1/\delta \rceil$.

Consideraremos a dinâmica σ_t começando em $\sigma_0 = e$ e vamos aderir ao seguinte esquema censurador:

- no período $[0, t_1]$, as atualizações em $x_i, i = 1, \dots, K-1$ são canceladas;
- no período $(t_1, t_2]$, não há censura;
- no período $[t_2, t_3]$, as atualizações em $x_i, i = 1, \dots, K-1$ são canceladas.

Chamaremos por $\nu_t = P_t^C$ a distribuição de σ_t para a dinâmica censurada. Um dos principais ingredientes na nossa prova será o fato de que como a identidade é o elemento maximal, a distribuição inicial é uma probabilidade crescente e, pela Afirmação 2 na prova da Proposição 5.15, ν_t é crescente para todo t .

A primeira proposição diz que depois do tempo t_1 , a distribuição de ν_t não é muito diferente de $\tilde{\nu}_t$.

Proposição 5.26 *Para todo $\epsilon > 0$, δ suficientemente pequeno e N suficientemente grande, temos para todo $t \geq t_1$, $\|\tilde{\nu}_t - \nu_t\|_{TV} \leq \frac{\epsilon}{3}$.*

Prova. Para todo $t \leq t_1$, temos que

$$\sigma_t \in S_N \text{ e } \tilde{\nu}_t = \tilde{\delta}_e,$$

onde $\tilde{\delta}_e$ é a probabilidade uniforme em \tilde{S}_N e δ_e é a massa de Dirac na identidade e .

Para cada $i = 1, \dots, K$, sejam ν_t^i a lei de σ_t restrita a $\{x_{i-1} + 1, \dots, x_i\}$ e π^i a medida de equilíbrio nas permutações de $\{x_{i-1} + 1, \dots, x_i\}$. Usando a Proposição 5.22 para cada dinâmica em $S_{\Delta x_i}$ e o fato de que a distância de variação total entre medidas produto é menor que a soma das distâncias de variação total das marginais, temos

$$\begin{aligned} \|\nu_t - \tilde{\delta}_e\|_{TV} &\leq \sum_{i=1}^K \|\nu_t^i - \pi^i\|_{TV} \leq \sum_{i=1}^K 10\Delta x_i e^{-t\lambda\Delta x_i} \\ &\leq K \left[10 \left(\frac{N}{K} + 1 \right) \exp \left(-2t \left(1 - \cos \left(\frac{\pi}{(N/K + 1)} \right) \right) \right) \right], \end{aligned}$$

pois $\Delta x_i \leq N/K + 1$. Tomando $t = t_1$ e $N \geq 10K$, temos

$$\begin{aligned} \|\nu_t - \tilde{\delta}_e\|_{TV} &\leq (10N + 10K) \exp \left(-\frac{N^2\delta \log N}{3\pi^2} \left(1 - \cos \left(\frac{\pi}{(N/K + 1)} \right) \right) \right) \\ &\leq 11N \exp \left(-\frac{N^2\delta \log N}{3\pi^2} \left(1 - \cos \left(\frac{\pi}{(N/K + 1)} \right) \right) \right) \\ &\leq 11N \exp \left(-(10\delta)^{-1} \log N \right) \\ &\leq \frac{\epsilon}{3}, \end{aligned}$$

desde que N seja suficientemente grande, pois

$$\frac{N^2\delta}{3\pi^2} \left(1 - \cos \left(\frac{\pi}{(N/K + 1)} \right) \right) \geq \frac{1}{10\delta}. \quad (5.16)$$

Notemos que mostrar (5.16) é equivalente a mostrar que

$$\frac{10N^2}{3\pi^2} \left(1 - \cos \left(\frac{\pi}{(N/K + 1)} \right) \right) \geq \frac{1}{\delta^2}.$$

De fato,

$$\cos \left(\frac{\pi}{N/K + 1} \right) = 1 - \frac{\pi^2}{2(N/K + 1)^2} + O(1/N^4).$$

Daí,

$$\left(1 - \cos\left(\frac{\pi}{(N/K + 1)}\right)\right) = \frac{\pi^2}{2(N/K + 1)^2} - O(1/N^4).$$

Logo,

$$\begin{aligned} \frac{10N^2}{3\pi^2} \left(1 - \cos\left(\frac{\pi}{(N/K + 1)}\right)\right) &= \frac{10N^2}{6(N/K + 1)^2} - O(1/N^2) \\ &= \frac{10N^2K^2}{6(N + K)^2} - O(1/N^2) \\ &= \frac{10K^2}{6(K/N + 1)^2} - O(1/N^2) \\ &\geq K^2, \text{ quando } N \text{ é suficientemente grande.} \end{aligned}$$

Como $K \geq 1/\delta$, temos o desejado.

Agora, mostraremos que $\|\nu_t - \tilde{\nu}_t\|_{TV}$ é decrescente. Com efeito, observe-mos que, da Definição 5.24, $\tilde{\nu}_t$ é simplesmente a lei de σ_t para a dinâmica com distribuição inicial $\tilde{\delta}_e$. O resultado segue de um argumento padrão de acoplamento.

O resultado também é válido para $t \geq t_1$ pois com a censura, a dinâmica é apenas o produto de K dinâmicas independentes em $S_{\Delta x_i}$, $i \in \{1, \dots, K\}$. \square

A segunda proposição diz que no tempo t_2 , a lei do esqueleto $\bar{\sigma}_t$ está próxima do equilíbrio.

Proposição 5.27 *Para todo $\epsilon > 0$, $\delta > 0$ e N suficientemente grande, temos*

$$\|\bar{\nu}_{t_2} - \bar{\pi}\|_{TV} \leq \frac{\epsilon}{3}.$$

Prova. Primeiramente, esqueçamos que $K = \lceil \delta^{-1} \rceil$, pois o que faremos vale para todo K finito. Também, por simplicidade, vamos supor que N é par.

Para entendermos melhor a ideia, vamos explicar no caso $K = 2$. Queremos mostrar que começando da distribuição ν_{t_1} , depois de um tempo $\frac{N^2}{2\pi^2}(1+\delta/3) \log N$, a altura $\tilde{\sigma}(N/2, N/2) = \bar{\sigma}(1, 1)$, que chamaremos por $\bar{\sigma}$ desde que não haja confusão, está próxima da sua distribuição de equilíbrio.

Afirmção 1: No equilíbrio, $\bar{\sigma} \approx (\sqrt{N}/4)\mathcal{N}$, onde \mathcal{N} é uma Gaussiana padrão.

De fato, notemos que $\pi\left(\bar{\sigma}(1, 1) = k - \frac{N}{4}\right) = \frac{\binom{N/2}{k} \binom{N/2}{N/2 - k}}{\binom{N}{N/2}}$. Agora

basta usarmos a fórmula de Stirling para deduzirmos um Teorema Central do Limite.

Agora, usando o Lema 5.21 com o tempo t_2 , obtemos para o valor esperado de $\bar{\sigma}$ em t_2 :

$$\begin{aligned} \nu_{t_2}(\bar{\sigma}) &\leq 2Ne^{-\lambda_N(t_2 - t_1)} \\ &= 2N \exp \left\{ - \left(\frac{\pi^2}{N^2} (1 + o(1)) \right) \left(\frac{N^2}{2\pi^2} (1 + \delta/3) \log N \right) \right\} \\ &= 2Ne^{-\left(\frac{1}{2} + \frac{\delta}{6}\right) \log N (1 + o(1))} \\ &= 2N \exp \left\{ \log N^{-\left(\frac{1}{2} + \frac{\delta}{6}\right)} \right\} (1 + o(1)) \\ &= (2N) N^{-\left(\frac{1}{2} + \frac{\delta}{6}\right)} (1 + o(1)) \\ &= 2N^{1/2 - \delta/6} \\ &\leq N^{1/2 - \delta/10} \end{aligned}$$

quando N é suficientemente grande.

Portanto, o valor esperado de $\bar{\sigma}$ em t_2 é muito menor que sua flutuação no equilíbrio.

Porém, isso não é suficiente para concluirmos que ν_{t_2} está próxima do equilíbrio. Da Afirmação 2 na prova da Proposição 5.15, ν_{t_2} tem densidade crescente e pela Proposição 5.20, a densidade $\bar{\nu}_{t_2}/\bar{\pi}$ da distribuição de $\bar{\sigma}$ é crescente. Assim, a seguinte afirmação é suficiente para concluirmos o caso $K = 2$:

Afirmação 2: Existe uma constante C tal que para qualquer N e qualquer medida ν tal que $\bar{\nu}/\bar{\pi}$ é crescente, temos

$$\|\bar{\nu} - \bar{\pi}\|_{TV} \leq \frac{C\bar{\nu}(\bar{\sigma})}{\sqrt{N}}.$$

De fato, seja $\mathcal{A} := \{x \in \{-N/4, N/4 + 1, \dots, N/4\} \mid \bar{\nu}(x) \geq \bar{\pi}(x)\}$.

Pela hipótese de ν , \mathcal{A} é um conjunto crescente. Além disso, pela definição da distância de variação total, temos

$$\|\bar{\nu} - \bar{\pi}\|_{TV} = \bar{\nu}(\mathcal{A}) - \bar{\pi}(\mathcal{A}).$$

Agora, vejamos que

$$\begin{aligned} \bar{\nu}(\bar{\sigma}) &= \bar{\nu}(\bar{\sigma}, \mathcal{A}) + \bar{\nu}(\bar{\sigma}, \mathcal{A}^c) \\ &= \bar{\nu}(\mathcal{A})\bar{\nu}(\bar{\sigma}|\mathcal{A}) + \bar{\nu}(\mathcal{A}^c)\bar{\nu}(\bar{\sigma}|\mathcal{A}^c). \end{aligned} \quad (5.17)$$

Usando a desigualdade de correlação (Lema 5.11) para as funções $f : \bar{\sigma} \mapsto \bar{\sigma}$ e $g : \bar{\sigma} \mapsto \frac{\bar{\nu}}{\bar{\pi}}(\bar{\sigma})$, que é crescente pelo item **(iii)** da Proposição 5.20, temos

$$\begin{aligned} \bar{\nu}(\mathcal{A})\bar{\nu}(\bar{\sigma}|\mathcal{A}) &= \bar{\pi}(\mathcal{A})\bar{\pi}\left(\frac{\bar{\nu}}{\bar{\pi}}(\bar{\sigma})\bar{\sigma}|\mathcal{A}\right) \\ &\geq \bar{\pi}(\mathcal{A})\bar{\pi}\left(\frac{\bar{\nu}}{\bar{\pi}}(\bar{\sigma})|\mathcal{A}\right)\bar{\pi}(\bar{\sigma}|\mathcal{A}) \\ &= \bar{\nu}(\mathcal{A})\bar{\pi}(\bar{\sigma}|\mathcal{A}). \end{aligned} \quad (5.18)$$

Analogamente,

$$\bar{\nu}(\mathcal{A}^c)\bar{\nu}(\bar{\sigma}|\mathcal{A}^c) \geq \bar{\nu}(\mathcal{A}^c)\bar{\pi}(\bar{\sigma}|\mathcal{A}^c). \quad (5.19)$$

Agora, substituindo (5.18) e (5.19) em (5.17) e subtraindo

$$0 = \bar{\pi}(\bar{\sigma}) = \bar{\pi}(\mathcal{A})\bar{\pi}(\bar{\sigma}|\mathcal{A}) + \bar{\pi}(\mathcal{A}^c)\bar{\pi}(\bar{\sigma}|\mathcal{A}^c),$$

obtemos

$$\begin{aligned} \bar{\nu}(\bar{\sigma}) &\geq \bar{\nu}(\mathcal{A})\bar{\pi}(\bar{\sigma}|\mathcal{A}) + \bar{\nu}(\mathcal{A}^c)\bar{\pi}(\bar{\sigma}|\mathcal{A}^c) \\ &\quad - \bar{\pi}(\mathcal{A})\bar{\pi}(\bar{\sigma}|\mathcal{A}) - \bar{\pi}(\mathcal{A}^c)\bar{\pi}(\bar{\sigma}|\mathcal{A}^c) \\ &= [\bar{\nu}(\mathcal{A}) - \bar{\pi}(\mathcal{A})]\bar{\pi}(\bar{\sigma}|\mathcal{A}) + \bar{\nu}(\mathcal{A}^c) - [\bar{\pi}(\mathcal{A}^c)]\bar{\pi}(\bar{\sigma}|\mathcal{A}^c) \\ &= \|\bar{\nu} - \bar{\pi}\|_{TV}\bar{\pi}(\bar{\sigma}|\mathcal{A}) - \|\bar{\nu} - \bar{\pi}\|_{TV}\bar{\pi}(\bar{\sigma}|\mathcal{A}^c) \\ &= \|\bar{\nu} - \bar{\pi}\|_{TV}[\bar{\pi}(\bar{\sigma}|\mathcal{A}) - \bar{\pi}(\bar{\sigma}|\mathcal{A}^c)] \\ &= \|\bar{\nu} - \bar{\pi}\|_{TV}[\bar{\pi}(\bar{\sigma}|\bar{\sigma} \geq x_{\mathcal{A}}) - \bar{\pi}(\bar{\sigma}|\bar{\sigma} < x_{\mathcal{A}})], \end{aligned}$$

onde $x_{\mathcal{A}}$ é um elemento de \mathcal{A} .

Finalmente, da escala Gaussiana, temos

$$\bar{\pi}(\bar{\sigma}|\bar{\sigma} > 0) = -\bar{\pi}(\bar{\sigma}|\bar{\sigma} < 0) \geq c\sqrt{N}.$$

Portanto,

$$\bar{\nu}(\mathcal{A}) \geq c\sqrt{N}\|\bar{\nu} - \bar{\pi}\|_{TV}.$$

Para $K \geq 3$, a ideia é basicamente a mesma. Ponhamos $v(\bar{\sigma}) := \sum_{i,j=1}^{K-1} \bar{\sigma}(i, j)$ como o volume abaixo do gráfico do esqueleto. Similarmente à prova da Afirmação 2, mostra-se que se $\nu(v(\bar{\sigma}))$ é pequena com respeito às suas flutuações no equilíbrio (que são de ordem \sqrt{N}) e ν é crescente, então $\bar{\nu}$ e $\bar{\pi}$ estão próximas uma da outra. Mais formalmente:

Afirmação 3: Seja ν uma medida de probabilidade em S_N cuja densidade com respeito a π é crescente. Para todo $\epsilon > 0$, existe $\eta(K, \epsilon)$ tal que para N suficientemente grande, temos

$$\|\bar{\nu} - \bar{\pi}\|_{TV} \leq \frac{\epsilon}{3}$$

sempre que $\nu(v(\bar{\sigma})) \leq \sqrt{N}\eta$.

Para entender a prova dessa afirmação, veja a prova do Lema 5.5 de [11], onde desta vez, é utilizada a desigualdade FKG para permutações.

Como no caso em que $K = 2$, usando o Lema 5.21, com o tempo t_2 temos

$$\nu_{t_2}[v(\bar{\sigma})] \leq 2N(K-1)^2 e^{-\lambda_N(t_2-t_1)} \leq \sqrt{N}\eta,$$

onde a última desigualdade é válida para qualquer η fixado quando N é suficientemente grande.

Assim, a Afirmação 3 é suficiente para concluir nosso resultado. □

Finalmente, a terceira proposição diz que no tempo t_3 , a distribuição do semi-esqueleto está próxima do equilíbrio.

Proposição 5.28 Para todo $\epsilon > 0$, $\delta > 0$ e N suficientemente grande, temos

$$\|\hat{\nu}_{t_3} - \hat{\pi}\|_{TV} \leq \frac{2\epsilon}{3}.$$

Prova. No período $[t_2, t_3]$, devido à censura, os valores tomados pelos conjuntos $\sigma_t(\{x_{i-1} + 1, \dots, x_i\})$, $i \in \{1, \dots, K\}$, são constantes no tempo. Neste período, as dinâmicas podem ser consideradas como um produto de K embaralhamentos

independentes, e a medida de equilíbrio correspondente condicionada no ponto de partida σ_{t_2} é simplesmente

$$\pi_{\sigma_{t_2}} := \pi(\cdot | \sigma(\{x_{i-1} + 1, \dots, x_i\}) = \sigma_{t_2}(\{x_{i-1} + 1, \dots, x_i\}), \forall i \in \{1, \dots, K\}).$$

Usando a Proposição 5.22 da mesma forma que na prova da Proposição 5.26, temos para qualquer realização de σ_{t_2} ,

$$\begin{aligned} & \left\| \mathbb{P}(\sigma_{t_3} \in \cdot | \sigma_{t_2}) - \pi_{\sigma_{t_2}} \right\|_{TV} \\ & \leq K \left[10 \left(\frac{N}{K} + 1 \right) \exp \left(-2(t_3 - t_2) \left(1 - \cos \left(\frac{\pi}{(N/K + 1)} \right) \right) \right) \right] \\ & = K \left[10 \left(\frac{N}{K} + 1 \right) \exp \left(-2t_1 \left(1 - \cos \left(\frac{\pi}{(N/K + 1)} \right) \right) \right) \right] \\ & \leq \frac{\epsilon}{3}, \end{aligned}$$

desde que N seja escolhido suficientemente grande.

Considerando o *push-forward* das medidas do semi-esqueleto e integrando no evento $\{\bar{\sigma}_{t_2} = \xi\}$, obtemos que para todo $\xi \in \bar{S}_N$,

$$\|\hat{\nu}_{t_3}(\cdot | \bar{\sigma} = \xi) - \hat{\pi}(\cdot | \bar{\sigma} = \xi)\|_{TV} \leq \frac{\epsilon}{3}. \quad (5.20)$$

Observe que a distribuição de $\bar{\sigma}_{t_3}$ é a mesma que a de $\bar{\sigma}_{t_2}$. De fato, com a censura, temos $\bar{\sigma}_{t_3} = \bar{\sigma}_{t_2}$ que, pela Proposição 5.27, está próxima do equilíbrio. Mais formalmente,

$$\begin{aligned} 2\|\hat{\nu}_{t_3} - \hat{\pi}\|_{TV} &= \sum_{\xi \in \bar{S}_N} \sum_{\{\hat{\sigma} \in \hat{S}_N | \bar{\sigma} = \xi\}} |\hat{\nu}_{t_3}(\hat{\sigma}) - \hat{\pi}(\hat{\sigma})| \\ &\leq \sum_{\xi \in \bar{S}_N} \sum_{\{\hat{\sigma} \in \hat{S}_N | \bar{\sigma} = \xi\}} \bar{\nu}_{t_3}(\xi) |\hat{\nu}_{t_3}(\hat{\sigma} | \bar{\sigma} = \xi) - \hat{\pi}(\hat{\sigma} | \bar{\sigma} = \xi)| \\ &\quad + \hat{\pi}(\hat{\sigma} | \bar{\sigma} = \xi) |\bar{\nu}_{t_3}(\xi) - \bar{\pi}(\xi)| \\ &= 2 \left(\|\bar{\nu}_{t_3} - \bar{\pi}\|_{TV} + \sum_{\xi \in \bar{S}_N} \bar{\nu}_{t_3}(\xi) \|\hat{\nu}_{t_3}(\cdot | \bar{\sigma} = \xi) - \hat{\pi}(\cdot | \bar{\sigma} = \xi)\|_{TV} \right) \\ &\leq 4\epsilon/3, \end{aligned}$$

sendo a última desigualdade verdadeira pela Proposição 5.27 e por (5.20). \square

Teorema 5.29 Para todo $\epsilon > 0$, $\delta > 0$ e N suficientemente grande, temos

$$d_N \left((1 + \delta) \frac{N^2}{2\pi^2} \log N \right) \leq \epsilon.$$

Prova. Pela Proposição 5.15 e pelo Lema 5.25, temos

$$\begin{aligned} d_N(t_3) &:= \|P_{t_3} - \pi\|_{TV} \leq \|\nu_{t_3} - \pi\|_{TV} \\ &\leq \|\hat{\nu}_{t_3} - \hat{\pi}\|_{TV} + \|\tilde{\nu}_{t_3} - \nu_{t_3}\|_{TV} \\ &\leq \frac{2\epsilon}{3} + \frac{\epsilon}{3} = \epsilon. \end{aligned}$$

Sendo a última desigualdade verdadeira pelas Proposições 5.26 e 5.28. \square

5.5

A cota inferior

Teorema 5.30 Para o embaralhamento de transposições adjacentes em tempo contínuo, temos que $T_{mix}^N(\epsilon) \geq \frac{1}{2\pi^2} N^2 \log N (1 - o(1))$.

Prova. (A prova de Wilson) Veja a prova do Teorema 6 de [17]. A prova no tempo contínuo é análoga. \square

Veremos no capítulo 7 que há uma outra maneira de encontrarmos essa cota.

5.6

O cutoff

Corolário 5.31 O embaralhamento de transposições adjacentes em tempo contínuo tem cutoff.

Prova. De fato, pelos Teoremas 5.29, 5.30, e 2.18, temos que

$$\frac{2\pi^2 T_{mix}^N(\epsilon)}{N^2 \log N} = 1.$$

Portanto, esta sequência de cadeias de Markov tem cutoff. \square

6

O embaralhamento de transposições adjacentes visto em tempo discreto

Vamos considerar a seguinte maneira de embaralhar um baralho com N cartas: A cada passo, com probabilidade $1/2$ nós trocamos as posições de um par de cartas adjacentes escolhido uniformemente ao acaso dentre as $N - 1$ possíveis escolhas, e com probabilidade $1/2$ não fazemos nada. Este é o mesmo embaralhamento que estudamos no capítulo anterior, porém visto em tempo discreto. Mais uma vez nos perguntamos: Quantas vezes temos que embaralhar o baralho dessa maneira até que ele esteja bem embaralhado?

A cota inferior para o tempo de mistura desta cadeia foi dada por Wilson em [17] (Veja o Teorema 6.4). Obteremos a cota superior através da cota obtida quando o embaralhamento é feito em tempo contínuo.

Sejam $(\sigma_n)_{n \geq 0}$ e $(\sigma'_t)_{t \geq 0}$ as trajetórias das cadeias em tempo discreto e contínuo, respectivamente, e denotemos por \mathbf{P}_n a lei de σ_n e P_t a lei de σ'_t . Vamos obter uma relação entre as distribuições \mathbf{P}_n e P_t .

Proposição 6.1 *Para todo $t \geq 0$, temos*

$$P_t = \sum_{k=0}^{\infty} \frac{(2(N-1)t)^k e^{-2(N-1)t}}{k!} \mathbf{P}_k.$$

Prova. A cadeia em tempo contínuo pode ser descrita da seguinte maneira: Consideramos um processo de Poisson \mathcal{T} com taxa $2(N-1)$ que é independente, isto é, $\mathcal{T}_0 = 0$ e $\mathcal{T}_n - \mathcal{T}_{n-1}$ são variáveis aleatórias exponenciais i.i.d. com média $1/[2(N-1)]$ para todo $n \geq 1$. Agora, para cada $n \geq 0$,

$$\sigma'_t = \sigma_n, \forall t \in [\mathcal{T}_n, \mathcal{T}_{n+1}).$$

Grosso modo, estamos dizendo que para ver o embaralhamento em tempo discreto através do mesmo em tempo contínuo, basta congelarmos o tempo quando os relógios estão silenciosos.

Assim,

$$\begin{aligned}
 P_t &= \mathbb{P}(\sigma'_t \in \cdot) = \sum_{k=0}^{\infty} \mathbb{P}(\sigma'_t \in \cdot, t \in [\mathcal{T}_k, \mathcal{T}_{k+1})) \\
 &= \sum_{k=0}^{\infty} \mathbb{P}(\sigma_k \in \cdot, t \in [\mathcal{T}_k, \mathcal{T}_{k+1})) \\
 &= \sum_{k=0}^{\infty} \mathbb{P}(\sigma_k \in \cdot) \mathbb{P}(t \in [\mathcal{T}_k, \mathcal{T}_{k+1})) \\
 &= \sum_{k=0}^{\infty} \frac{(2(N-1)t)^k e^{-2(N-1)t}}{k!} \mathbf{P}_k.
 \end{aligned}$$

□

Agora, veremos um lema que nos ajudará a encontrar o tempo de mistura para o embaralhamento comparando a distância do equilíbrio deste com a do mesmo feito em tempo contínuo.

Lema 6.2 Para todos $t, n > 0$, temos

$$\|P_t - \pi\|_{TV} \leq \frac{\|\mathbf{P}_n - \pi\|_{TV}}{\sum_{k=0}^n (2(N-1)t)^k e^{-2(N-1)t}/k!}$$

e

$$\|P_t - \pi\|_{TV} \geq \frac{\|\mathbf{P}_n - \pi\|_{TV} - \sum_{k=0}^{n-1} (2(N-1)t)^k e^{-2(N-1)t}/k!}{\sum_{k=n}^{\infty} (2(N-1)t)^k e^{-2(N-1)t}/k!}.$$

Prova. Sejam $t, n > 0$. Pela Afirmação 2 na prova da Proposição 5.15 (cuja prova é análoga em tempo discreto), \mathbf{P}_n e P_t são medidas de probabilidade crescentes. Logo,

$$A_1 := \{\sigma; \mathbf{P}_n(\sigma) \geq \pi(\sigma)\}$$

e

$$A_2 := \{\sigma; P_t(\sigma) \geq \pi(\sigma)\}.$$

são eventos crescentes.

Pela Afirmação 1 na prova da proposição 5.15, para qualquer evento crescente A , $(\mathbf{P}_k(A))_{k \geq 0}$ é uma sequência não crescente que converge para $\pi(A)$.

Assim,

$$\begin{aligned}
 \|P_t - \pi\|_{TV} &\geq P_t(A_1) - \pi(A_1) \\
 &= \sum_{k=0}^{\infty} \left(\frac{(2(N-1)t)^k e^{-2(N-1)t}}{k!} (\mathbf{P}_k(A_1) - \pi(A_1)) \right) \\
 &= \sum_{k=0}^n \left(\frac{(2(N-1)t)^k e^{-2(N-1)t}}{k!} (\mathbf{P}_k(A_1) - \pi(A_1)) \right) \\
 &\quad + \sum_{k=n+1}^{\infty} \left(\frac{(2(N-1)t)^k e^{-2(N-1)t}}{k!} (\mathbf{P}_k(A_1) - \pi(A_1)) \right) \\
 &\geq \sum_{k=0}^n \left(\frac{(2(N-1)t)^k e^{-2(N-1)t}}{k!} \right) (\mathbf{P}_n(A_1) - \pi(A_1)) \\
 &\quad + \sum_{k=n+1}^{\infty} \left(\frac{(2(N-1)t)^k e^{-2(N-1)t}}{k!} (\mathbf{P}_k(A_1) - \pi(A_1)) \right) \\
 &\geq \sum_{k=0}^n \left(\frac{(2(N-1)t)^k e^{-2(N-1)t}}{k!} \right) (\mathbf{P}_n(A_1) - \pi(A_1)) \\
 &= \sum_{k=0}^n \left(\frac{(2(N-1)t)^k e^{-2(N-1)t}}{k!} \right) \|\mathbf{P}_n - \pi\|_{TV}.
 \end{aligned}$$

Além disso,

$$\begin{aligned}
 \|P_t - \pi\|_{TV} &= P_t(A_2) - \pi(A_2) \\
 &= \sum_{k=0}^{\infty} \left(\frac{(2(N-1)t)^k e^{-2(N-1)t}}{k!} (\mathbf{P}_k(A_2) - \pi(A_2)) \right) \\
 &= \sum_{k=0}^{n-1} \left(\frac{(2(N-1)t)^k e^{-2(N-1)t}}{k!} (\mathbf{P}_k(A_2) - \pi(A_2)) \right) \\
 &\quad + \sum_{k=n}^{\infty} \left(\frac{(2(N-1)t)^k e^{-2(N-1)t}}{k!} (\mathbf{P}_k(A_2) - \pi(A_2)) \right) \\
 &\leq \sum_{k=0}^{n-1} \left(\frac{(2(N-1)t)^k e^{-2(N-1)t}}{k!} (\mathbf{P}_k(A_2) - \pi(A_2)) \right) \\
 &\quad + \sum_{k=n}^{\infty} \left(\frac{(2(N-1)t)^k e^{-2(N-1)t}}{k!} \right) (\mathbf{P}_n(A_2) - \pi(A_2)) \\
 &\leq \sum_{k=0}^{n-1} \left(\frac{(2(N-1)t)^k e^{-2(N-1)t}}{k!} \right) \\
 &\quad + \sum_{k=n}^{\infty} \left(\frac{(2(N-1)t)^k e^{-2(N-1)t}}{k!} \right) (\mathbf{P}_n(A_2) - \pi(A_2)) \\
 &= \sum_{k=0}^{n-1} \left(\frac{(2(N-1)t)^k e^{-2(N-1)t}}{k!} \right) \\
 &\quad + \sum_{k=n}^{\infty} \left(\frac{(2(N-1)t)^k e^{-2(N-1)t}}{k!} \right) \|\mathbf{P}_n - \pi\|_{TV}.
 \end{aligned}$$

Portanto, temos o resultado. \square

6.1

A cota superior

Teorema 6.3 *Para o embaralhamento de transposições adjacentes com n cartas, temos que $t_{mix}^{(n)}(\epsilon) \leq \frac{1}{\pi^2} n^3 \log n(1 - o(1))$.*

Prova. Dividiremos a prova em duas afirmações.

Afirmção 1: $\|\mathbf{P}_n - \pi\|_{TV} \leq \|P_{(n-n^{1/3})/(2(N-1))} - \pi\|_{TV} + o(1)$.

De fato, usando o Lema 6.2 com $t = \frac{n-n^{1/3}}{2(N-1)}$, temos que

$$\|\mathbf{P}_n - \pi\|_{TV} \leq \frac{\|P_{(n-n^{1/3})/(2(N-1))} - \pi\|_{TV}}{A_n},$$

onde $A_n = \sum_{k=0}^n (n - n^{1/3})^k e^{-(n-n^{1/3})}/k!$.

Portanto, apenas precisamos mostrar que $\lim_{n \rightarrow \infty} A_n = 1$.

Com efeito,

$$\begin{aligned} A_n &= \sum_{k=0}^n \frac{(n - n^{1/3})^k e^{-(n-n^{1/3})}}{k!} = e^{-(n-n^{1/3})} \sum_{k=0}^n \frac{(n - n^{1/3})^k}{k!} \\ &= e^{-(n-n^{1/3})} \sum_{k=0}^{\infty} \frac{(n - n^{1/3})^k}{k!} - e^{-(n-n^{1/3})} \sum_{k=n+1}^{\infty} \frac{(n - n^{1/3})^k}{k!} \\ &= e^{-(n-n^{1/3})} \sum_{k=0}^{\infty} \frac{(n - n^{1/3})^k}{k!} - o(1) \\ &= e^{-(n-n^{1/3})} e^{(n-n^{1/3})} - o(1) = 1 - o(1). \end{aligned}$$

Afirmção 2: Para todo $\epsilon > 0$, e N suficientemente grande temos

$$\|\mathbf{P}_{N^3 \log N / \pi^2} - \pi\|_{TV} \leq \epsilon.$$

De fato, tome $n = N^3 \log N / \pi^2$. Assim, pela Afirmção 1, temos que

$$\|\mathbf{P}_{N^3 \log N / \pi^2} - \pi\|_{TV} \leq \|P_{(N^3 \log N / \pi^2)(1-o(1))} - \pi\|_{TV}.$$

O resultado segue do Teorema 5.29. \square

6.2

A cota inferior

Teorema 6.4 *Para o embaralhamento de transposições adjacentes com n cartas, temos que $t_{mix}^{(n)}(\epsilon) \geq \frac{1}{\pi^2} n^3 \log n (1 - o(1))$.*

Prova. Veja a prova do Teorema 6 de [17] ou prossiga como fizemos na cota superior utilizando a outra desigualdade do Lema 6.2 com $t = \frac{n+n^{1/3}}{2(N-1)}$. \square

6.3

O cutoff

Corolário 6.5 *O embaralhamento de transposições adjacentes em tempo discreto tem cutoff.*

Prova. De fato, pelos Teoremas 6.3, 6.4, e 2.18, temos que

$$\frac{\pi^2 t_{mix}^{(n)}(\epsilon)}{n^3 \log n} = 1.$$

Portanto, esta sequência de cadeias de Markov tem cutoff. \square

7

Uma maravilhosa aplicação: Encontrando o tempo de mistura para o processo de exclusão simples simétrico

Neste capítulo veremos como obter uma cota superior para o tempo de mistura do processo de exclusão simples simétrico através do embaralhamento de transposições adjacentes e como obter uma cota inferior para embaralhamento usando a exclusão.

7.1

O processo de exclusão simples simétrico

Consideremos um segmento com N sítios e coloquemos $k \in \{1, \dots, N-1\}$ partículas neste segmento, com no máximo uma partícula por sítio. Agora, consideremos a seguinte dinâmica: Cada partícula pula, independentemente, com taxa igual ao número de sítios vazios em sua vizinhança, para um de seus sítios vizinhos escolhido uniformemente entre eles. Equivalentemente, ela pula com taxa um em cada um dos seus vizinhos vazios. Queremos saber quanto tempo devemos esperar para alcançarmos o equilíbrio do sistema de partículas.

Mais formalmente, considere o segmento $[0, N]$ dividido em N intervalos de tamanho um. Identificaremos o intervalo $[x-1, x]$ com $x \in \{1, \dots, N\}$ e chamaremos este intervalo de sítio. Cada um destes sítios tem dois possíveis estados: ou está vazio, ou contém uma partícula.

Quando consideramos o processo de exclusão com k partículas, o espaço de estados é definido por

$$\Omega_{N,k} := \left\{ \gamma \in \{0, 1\}^N : \sum_{x=1}^N \gamma(x) = k \right\}.$$

O processo de exclusão no segmento $[0, N]$ é uma cadeia de Markov em tempo contínuo em $\Omega_{N,k}$, onde cada uma das k partículas pula para o vizinho à

direita e para o vizinho à esquerda com taxa um sempre que estes sítios estiverem vazios. Veja a Figura 7.1.

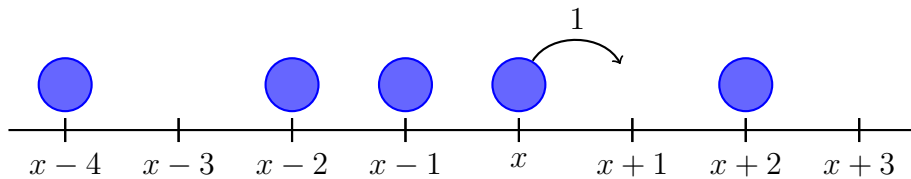


Figura 7.1: Transição de uma partícula para um sítio vizinho.

Agora, note que S_N naturalmente atua sobre $\Omega_{N,k}$.

Para $\sigma \in S_N$, $\gamma \in \Omega_{N,k}$, podemos definir

$$\sigma \bullet \gamma(x) := \gamma(\sigma(x)).$$

Assim, o gerador para o processo de exclusão em $[0, N]$ pode ser escrito como

$$(\mathcal{L}f)(\gamma) := \sum_{x=1}^{N-1} f(\tau_x \bullet \gamma) - f(\gamma),$$

onde τ_x denota a transposição adjacente $(x, x+1)$.

Chamaremos por π_k (ou π quando não houver possibilidade de confusão) a medida de equilíbrio desta cadeia. Escreveremos $(\gamma_t^\xi)_{t \geq 0}$ para a cadeia de Markov começando em $\xi \in \Omega_{N,k}$, P_t^ξ para a lei da marginal γ_t^ξ e definiremos

$$d^{N,k}(t) := \max_{\xi \in \Omega_{N,k}} \|P_t^\xi - \pi\|_{TV} = \max_{\nu \text{ probabilidade em } \Omega_{N,k}} \|P_t^\nu - \pi\|_{TV},$$

$$T_{mix}^{N,k}(\epsilon) := \inf \{t \geq 0 : d^{N,k}(t) \leq \epsilon\},$$

Note que, diferente do embaralhamento de transposições adjacentes, $\|P_t^\xi - \pi\|_{TV}$ depende da condição inicial ξ pois não há simetria.

7.2

Conectando a exclusão simples e o embaralhamento de transposições adjacentes

Existe uma projeção natural do conjunto das permutações no conjunto das configurações das partículas $\sigma \mapsto \gamma_\sigma$. Como na Figura 7.2, à cada carta marcada de 1 até k , associamos o papel de uma partícula e às cartas marcadas de $k + 1$ até N , associamos o papel de sítios vazios. Assim,

$$\gamma_\sigma(x) := \begin{cases} 1 & \text{se } \sigma(x) \leq k; \\ 0 & \text{se } \sigma(x) > k. \end{cases}$$

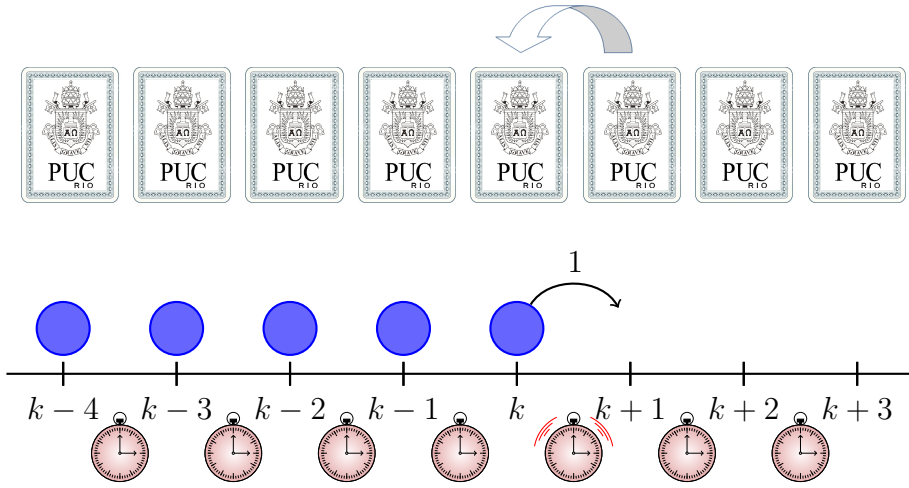


Figura 7.2: Permutação e associada à configuração inicial das partículas na construção gráfica.

Com essa aplicação, o embaralhamento de transposições adjacentes $(\sigma_t)_{t \geq 0}$ é projetado sobre o processo de exclusão, e como a distância só pode ser reduzida com projeção, temos que para todo $k \in \{1, \dots, N-1\}$, para todo $t \geq 0$ e para todo $\epsilon \in (0, 1)$,

$$d^{N,k}(t) \leq d^N(t),$$

e

$$T_{mix}^{N,k}(\epsilon) \leq T_{mix}^N(\epsilon). \quad (7.1)$$

Portanto, encontrando uma cota superior para o tempo de mistura do embaralhamento de transposições adjacentes, também encontramos uma cota superior para o processo de exclusão, e encontrando uma cota inferior para a

exclusão, encontramos uma cota inferior para o embaralhamento.

Nosso interesse está em utilizar a cota superior encontrada para o embaralhamento de transposições adjacentes com o objetivo de obter uma cota superior para o sistema de partículas introduzido. Todavia, por questão de completude, apresentaremos, sem provas, resultados que implicam na existência de cutoff para o processo de exclusão.

7.3

A cota superior

Proposição 7.1 *Para o processo de exclusão simples simétrico temos*

$$T_{mix}^{N,k}(\epsilon) \leq \frac{1}{2\pi^2} N^2 \log N (1 + o(1)).$$

Prova. De fato, basta considerarmos (7.1) e o Teorema 5.29. \square

Em [11], Lacoïn mostra uma cota ainda melhor para o tempo de mistura do processo de exclusão, que coincide com a cota obtida acima quando $k = N/2$. Vejamos tal resultado:

Teorema 7.2 *Para o processo de exclusão simples simétrico temos*

$$T_{mix}^{N,k}(\epsilon) \leq \frac{1}{2\pi^2} N^2 \log \min(k, N - k) (1 + o(1)).$$

Prova. Veja a seção 8 de [11].

Na demonstração desse resultado são usadas algumas das ferramentas apresentadas no capítulo 5, mas não apenas essas. \square

7.4

A cota inferior

Teorema 7.3 *Para o processo de exclusão simples simétrico temos*

$$T_{mix}^{N,k}(\epsilon) \geq \frac{1}{2\pi^2} N^2 \log \min(k, N - k) (1 - o(1)).$$

Prova. Veja a seção 7 de [11]. \square

Observe que o Teorema 5.30 é imediatamente provado usando (7.1) e o Teorema 7.3 com $k = N/2$.

7.5

O cutoff

Corolário 7.4 *O processo de exclusão simples simétrico tem cutoff.*

Prova. De fato, pelos Teoremas 7.2 e 7.3, temos que

$$\frac{2\pi^2 T_{mix}^{N,k}(\epsilon)}{N^2 \log \min(k, N-k)} = 1.$$

Portanto, esta sequência de cadeias de Markov tem cutoff. □

Referências bibliográficas

- [1] ALDOUS, D., DIACONIS, P. **Shuffling cards and stopping times**, The American Mathematical Monthly 93.5 (1986): 333-348.
- [2] ALDOUS, D., DIACONIS, P. **Strong uniform times and finite random walks**, Advances in Applied Mathematics 8.1 (1987): 69-97.
- [3] BAYER, D., DIACONIS, P. **Trailing the dovetail shuffle to its lair**, The Annals of Applied Probability (1992): 294-313.
- [4] BERESTYCKI, N. **Lectures on Mixing Times**, Cambridge University (2014).
- [5] DIACONIS, P. **Group representations in probability and statistics**, Lecture Notes-Monograph Series 11 (1988): i-192.
- [6] DURRETT, R. **Probability: theory and examples**, Cambridge University Press, 2010.
- [7] FELLER, W. **Theory of Probability and its Applications**, vol. 2. (1971).
- [8] HÄGGSTRÖM, O. **Finite Markov chains and algorithmic applications**, Vol. 52. Cambridge University Press, 2002.
- [9] HOLLEY, R. **Remarks on the FKG inequalities**, Communications in Mathematical Physics 36.3 (1974): 227-231.
- [10] JAMES, B. R. **Probabilidade: Um Curso em Nível Intermediário**, Coleção Euclides, Rio de Janeiro. IMPA, 3ª Edição (2011).
- [11] LACOIN, H. **Mixing time and cutoff for the adjacent transposition shuffle and the simple exclusion**, The Annals of Probability 44.2 (2016): 1426-1487.
- [12] LEVIN, D. A., PERES, Y., WILMER, E.L. **Markov chains and mixing times**, American Mathematical Soc., 2009.
- [13] LIGGETT, T.M. **Interacting particle systems**, Vol. 276. Springer Science and Business Media, 2012.

- [14] LIGGETT, T. M. **Continuous time Markov processes: an introduction**, Vol. 113. American Mathematical Soc., 2010.
- [15] PERES, Y., PETER, W. **Can extra updates delay mixing?**, Communications in Mathematical Physics 323.3 (2013): 1007-1016.
- [16] TANNY, S. **A probabilistic interpretation of Eulerian numbers**, Duke Math. J 40.4 (1973): 717-722.
- [17] WILSON, D. B. **Mixing times of lozenge tiling and card shuffling Markov chains**, Annals of Applied Probability (2004): 274-325.

A

Cadeias de Markov em espaços finitos

A.1

Cadeias de Markov em tempo discreto

Definição A.1 *Seja P uma matriz $n \times n$ com elementos $P(w_i, w_j) := P_{i,j}$, $i, j = 1, \dots, n$. Chamamos de Cadeia de Markov homogênea um processo aleatório (X_0, X_1, X_2, \dots) , com espaço de estados finito $\Omega = \{w_1, \dots, w_n\}$, se para todo $t \geq 0$, todo $i, j \in 1, \dots, n$ e todo $i_0, \dots, i_{t-1} \in 1, \dots, n$, temos:*

$$P(X_{t+1} = w_j | X_0 = w_{i_0}, \dots, X_{t-1} = w_{i_{t-1}}, X_t = w_i) = P(X_{t+1} = w_j | X_t = w_i) \\ = P(w_i, w_j).$$

Diremos que P é a matriz de transição da Cadeia de Markov e $P(w_i, w_j)$ são probabilidades de transição, ou seja, $P(w_i, w_j)$ é a probabilidade condicional de chegar ao estado w_j no tempo $t+1$ dado que no tempo t estava-se no estado w_i .

Toda matriz de transição deve seguir as seguintes condições:

- $P(x_i, x_j) \geq 0, \forall i, j \in \{1, \dots, n\}$;
- $\sum_{j=1}^n P(x_i, x_j) = 1, \forall i \in \{1, \dots, n\}$.

Além disso, definimos as distribuições da Cadeia de Markov nos tempos $0, 1, 2, \dots$ por $\mu^{(0)}, \mu^{(1)}, \mu^{(2)}, \dots$ de forma que

$$\mu^{(t)} = (\mu^{(t)}_1, \mu^{(t)}_2, \dots, \mu^{(t)}_n) \\ = (P(X_t = w_1), P(X_t = w_2), \dots, P(X_t = w_n)).$$

Teorema A.2 *Para uma cadeia de Markov (X_0, X_1, \dots) com espaço de estados $\Omega = \{w_1, w_2, \dots, w_n\}$, distribuição inicial $\mu^{(0)}$ e matriz de transição P , temos que para todo t , a distribuição $\mu^{(t)}$ satisfaz em tempo t :*

$$\mu^{(t)} = \mu^{(0)} . P^t$$

Prova. Considere o caso $t = 1$. Temos que para $j = 1, \dots, n$

$$\begin{aligned}\mu^{(1)}_j &= P(X_1 = w_j) = \sum_{i=1}^n P(X_0 = w_i, X_1 = w_j) \\ &= \sum_{i=1}^n P(X_0 = w_i)P(X_1 = w_j | X_0 = w_i) \\ &= \sum_{i=1}^n \mu^{(0)}_i P_{i,j} = (\mu^{(0)}P)_j,\end{aligned}$$

onde $(\mu^{(0)}P)_j$ é o j -ésimo elemento de $\mu^{(0)}P$. Portanto, $\mu^{(1)} = \mu^{(0)}P$.

Agora, suponhamos que para um t fixo, $\mu^{(t)} = \mu^{(0)}P^t$. Mostraremos que $\mu^{(t+1)} = \mu^{(0)}P^{t+1}$. Com efeito,

$$\begin{aligned}\mu^{(t+1)}_j &= P(X_{t+1} = w_j) = \sum_{i=1}^n P(X_t = w_i, X_{t+1} = w_j) \\ &= \sum_{i=1}^n P(X_t = w_i)P(X_{t+1} = w_j | X_t = w_i) \\ &= \sum_{i=1}^n \mu^{(t)}_i P_{i,j} = (\mu^{(t)}P)_j\end{aligned}$$

Então,

$$\mu^{(t+1)} = \mu^{(t)}P = \mu^{(0)}P^tP = \mu^{(0)}P^{t+1}$$

Portanto, $\mu^{(t)} = \mu^{(0)}P^t$ para todo t . \square

Considere uma cadeia de Markov (X_0, X_1, \dots) com espaço de estados $\Omega = \{w_1, w_2, \dots, w_n\}$ e matriz de transição P . Dizemos que um estado w_i se comunica com outro estado w_j , com a notação $w_i \rightarrow w_j$, se a cadeia tem probabilidade positiva de alcançar w_j saindo do estado w_i , ou seja,

$$P(X_{t+s} = w_j | X_t = w_i) > 0 \text{ para algum } s \Rightarrow w_i \rightarrow w_j.$$

Se $w_i \rightarrow w_j$ e $w_j \rightarrow w_i$, então dizemos que w_i e w_j se intercomunicam, e usamos a notação $w_i \leftrightarrow w_j$.

Definição A.3 Dizemos que uma cadeia de Markov (X_0, X_1, \dots) com espaço de estados $\Omega = \{w_1, w_2, \dots, w_n\}$ e matriz de transição P é *irredutível* se $w_i \leftrightarrow w_j, \forall w_i, w_j \in \Omega$. Caso contrário, a cadeia é *redutível*.

Definição A.4 O período $d(w_i)$ de um estado $w_i \in \Omega$ é definido como

$$d(w_i) = \text{mdc}\{t \geq 1 : (P^t)_{i,i} > 0\}$$

Se $d(w_i) = 1$, dizemos que o estado w_i é aperiódico.

Definição A.5 Um cadeia de Markov é aperiódica se todos seus estados são aperiódicos. Caso contrário, a cadeia é periódica.

Definição A.6 Seja (X_0, X_1, \dots) uma cadeia de Markov com espaço de estados $\Omega = \{w_1, w_2, \dots, w_n\}$ e matriz de transição P . Um vetor linha $\pi = (\pi_1, \dots, \pi_n)$ é dito ser uma distribuição estacionária para a Cadeia de Markov, se satisfaz:

- $\pi_i \geq 0$ para $i = 1, \dots, n$, e $\sum_{i=1}^n \pi_i = 1$, e
- $\pi.P = \pi$, ou seja, $\sum_{i=1}^n \pi_i.P_{i,j} = \pi_j$ para $j = 1, \dots, n$.

Teorema A.7 Seja (X_0, X_1, \dots) uma cadeia de Markov irredutível e aperiódica com espaço de estados $\Omega = \{w_1, w_2, \dots, w_n\}$, matriz de transição P e distribuição inicial $\mu^{(0)}$ arbitrária. Então, existe uma única distribuição de probabilidade π em Ω tal que

$$\|\mu^{(t)} - \pi\|_{TV} \rightarrow 0,$$

quando $t \rightarrow \infty$.

Prova. Veja em [8], onde é feita uma belíssima demonstração usando um acoplamento. \square

Definição A.8 Seja (X_0, X_1, \dots) uma cadeia de Markov com espaço de estados $\Omega = \{w_1, w_2, \dots, w_n\}$ e matriz de transição P . Uma distribuição de probabilidade π em Ω é reversível para a cadeia (ou para a matriz de transição P) se para todos $i, j \in \{1, \dots, n\}$ e $t > 0$ temos:

$$\pi_i P_{i,j}^t = \pi_j P_{j,i}^t.$$

A cadeia de Markov é reversível se existe uma distribuição reversível para ela.

Teorema A.9 Seja (X_0, X_1, \dots) uma cadeia de Markov com espaço de estados $\Omega = \{w_1, w_2, \dots, w_n\}$ e matriz de transição P . Se π é uma distribuição reversível para a cadeia, então π também é uma distribuição estacionária para a cadeia.

Prova. A primeira propriedade da definição de distribuição estacionária é imediata, então nos resta mostrar que para qualquer $j \in \{1, \dots, n\}$ temos $\pi_j = \sum_{i=1}^n \pi_i P_{i,j}$. Com efeito,

$$\pi_j = \pi_j \sum_{i=1}^n P_{j,i} = \sum_{i=1}^n \pi_j P_{j,i} = \sum_{i=1}^n \pi_i P_{i,j}.$$

\square

A.2**Cadeias de Markov em tempo contínuo**

As cadeias de Markov apresentadas nesta dissertação são exibidas com suas respectivas construções. Como o tempo agora é contínuo, falar de aperioidicidade não faria muito sentido. Assim, generalizando o que vimos, toda cadeia de Markov em tempo contínuo irredutível possui uma medida invariante π e dada uma distribuição inicial arbitrária, a distribuição da cadeia convergirá para π .

O gerador \mathcal{L} de uma cadeia de Markov, *grosso modo*, explica como são feitas atualizações. Para nossa proposta, devemos saber apenas que

$$\partial_t \mathbb{E}[f(X_t)] = \mathbb{E}[\mathcal{L}f(X_t)].$$

B

Sistemas dinâmicos aleatórios

Definição B.1 Um sistema dinâmico aleatório de uma matriz de transição P num estado de espaços Ω é uma função $f : \Omega \times \Lambda \longrightarrow \Omega$, juntamente com uma variável aleatória Z com valores em Λ que satisfaz

$$P(f(x, Z) = y) = P(x, y).$$

Proposição B.2 Se Z_1, Z_2, \dots é uma sequência de variáveis aleatórias independentes, cada uma tendo a mesma distribuição Z , e X_0 tem distribuição μ , então a sequência (X_0, X_1, \dots) definida por

$$X_n = f(X_{n-1}, Z_n), \text{ para } n \geq 1$$

é uma cadeia de Markov com matriz de transição P e distribuição inicial μ .

Prova. De fato,

$$\begin{aligned} & P(X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_0 = x_0) \\ &= \frac{P(f(X_n, Z_{n+1}) = x_{n+1}, f(X_{n-1}, Z_n) = x_n, \dots, f(X_0, Z_1) = x_1, X_0 = x_0)}{P(f(X_{n-1}, Z_n) = x_n, \dots, f(X_0, Z_1) = x_1, X_0 = x_0)} \\ &= \frac{P(f(x_n, Z_{n+1}) = x_{n+1}, f(x_{n-1}, Z_n) = x_n, \dots, f(x_0, Z_1) = x_1, X_0 = x_0)}{P(f(x_{n-1}, Z_n) = x_n, \dots, f(x_0, Z_1) = x_1, X_0 = x_0)} \\ &= \frac{P(f(x_n, Z_{n+1}) = x_{n+1})P(f(x_{n-1}, Z_n) = x_n) \cdots P(f(x_0, Z_1) = x_1)P(X_0 = x_0)}{P(f(x_{n-1}, Z_n) = x_n) \cdots P(f(x_0, Z_1) = x_1)P(X_0 = x_0)} \\ &= P(f(x_n, Z_{n+1}) = x_{n+1}) = P(x_n, x_{n+1}). \end{aligned}$$

E por outro lado,

$$\begin{aligned} P(X_{n+1} = x_{n+1} | X_n = x_n) &= \frac{P(f(X_n, Z_{n+1}) = x_{n+1}, f(X_{n-1}, Z_n) = x_n)}{P(f(X_{n-1}, Z_n) = x_n)} \\ &= \frac{P(f(x_n, Z_{n+1}) = x_{n+1}, f(x_{n-1}, Z_n) = x_n)}{P(f(x_{n-1}, Z_n) = x_n)} \\ &= \frac{P(f(x_n, Z_{n+1}) = x_{n+1})P(f(x_{n-1}, Z_n) = x_n)}{P(f(x_{n-1}, Z_n) = x_n)} \\ &= P(f(x_n, Z_{n+1}) = x_{n+1}) = P(x_n, x_{n+1}). \end{aligned}$$

□

Proposição B.3 *Toda matriz de transição em um espaço de estados finito tem um sistema dinâmico aleatório associado.*

Prova. Seja P a matriz de transição de uma cadeia de Markov com espaço de estados $\Omega = \{x_1, \dots, x_n\}$. Tomemos $\Lambda = [0, 1]$; Nossas variáveis aleatórias auxiliares Z, Z_1, Z_2, \dots serão uniformemente escolhidas neste intervalo.

Seja $F_{j,k} = \sum_{i=1}^k P(x_j, x_i)$ e definamos

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

Assim,

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

□

C

Passeios aleatórios em Grupos

Definição C.1 Dada uma distribuição de probabilidade Q em um grupo (G, \cdot) , dizemos que um passeio aleatório em G com distribuição de incremento Q é uma cadeia de Markov (veja A.1) com espaço de estados G e que se move multiplicando o estado atual, à esquerda, por um elemento aleatório de G escolhido de acordo com Q . Equivalentemente, a matriz de transição P dessa cadeia tem entradas $P(g, hg) = Q(h)$ para todo $g, h \in G$.

Partindo do princípio que $X_0 = e$, o elemento identidade, temos que $P(X_1 = g) = Q(g)$. Além disso, a distribuição de X_2 é dada pela convolução

$$P(X_2 = g) = Q^*Q(g) = \sum_{h \in G} Q(h)Q(gh^{-1}).$$

Prosseguindo indutivamente, $P(X_k = g) = Q^{k*}(g)$, onde Q^{k*} é a convolução iterada

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

Proposição C.2 Sejam G um grupo finito, P a matriz de transição de um passeio aleatório em G , e U a distribuição de probabilidade uniforme em G . Então, U é uma distribuição estacionária para P .

Prova. Seja Q a distribuição de incremento do passeio aleatório. Para qualquer $g \in G$, temos:

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

Observação: Para a primeira igualdade, mudamos o índice fazendo $k = gh^{-1}$. \square

Definição C.3 Sejam G um grupo, $H \subset G$ um conjunto e $\langle H \rangle$ o menor grupo contendo todos os elementos de H (Lembre-se que qualquer elemento de $\langle H \rangle$ pode ser escrito como um produto de elementos em H e seus inversos). Dizemos que H gera G se $\langle H \rangle = G$.

Proposição C.4 Sejam G um grupo finito e Q uma distribuição de probabilidade em G . O passeio aleatório em G com distribuição de incremento Q é irredutível se, e somente se, $S = \{g \in G : Q(g) > 0\}$ gera G .

Prova. Tomemos $a \in G$, arbitrariamente. Se o passeio aleatório é irredutível, então existe $r > 0$ tal que $P^r(e, a) > 0$.

Daí, deve existir uma sequência $s_1, \dots, s_r \in G$ tal que $a = s_r \cdot s_{r-1} \cdots s_1$ com $s_i \in S$ para todo $i = 1, \dots, r$. Assim, $a \in \langle S \rangle$.

Agora, suponhamos que S gera G e tomemos $a, b \in G$. Sabemos que ba^{-1} pode ser escrito como um produto de elementos em S e seus inversos. Como todo elemento de G tem ordem finita, qualquer inverso aparecendo na expressão para ba^{-1} pode ser reescrito como uma potência positiva do mesmo elemento do grupo.

Considere essa última expressão sendo $ba^{-1} = s_r \cdot s_{r-1} \cdots s_1$, onde $s_i \in S$ para todo $i = 1, \dots, r$. Portanto,

$$\begin{aligned} P^r(a, b) &\geq P(a, s_1 \cdot a) P(s_1 \cdot a, s_2 \cdot s_1 \cdot a) \cdots P(s_{r-1} \cdot s_{r-2} \cdots s_1 \cdot a, (ba^{-1})a) \\ &= Q(s_1) \cdots Q(s_r) > 0. \end{aligned}$$

□

Definição C.5 *Seja G um grupo. Dizemos que uma distribuição de probabilidade Q em G é simétrica se $Q(g) = Q(g^{-1})$ para todo $g \in G$.*

Proposição C.6 *Seja G um grupo finito. Um passeio aleatório em G com distribuição de incremento Q é reversível se, e somente se, Q é simétrica.*

Prova. Seja U a distribuição de probabilidade uniforme em G e suponhamos que Q é simétrica. Para quaisquer $g, h \in G$ temos que

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

Reciprocamente, sejam $g, k \in G$ e tomemos $h = kg$. Assim, $k = hg^{-1}$ e como U é reversível, temos:

$$\begin{aligned} \frac{Q(k)}{|G|} &= \frac{1}{|G|} Q(hg^{-1}) = \frac{1}{|G|} P(g, h) \\ &= \frac{1}{|G|} P(h, g) = \frac{1}{|G|} Q(gh^{-1}) \\ &= \frac{1}{|G|} Q((hg^{-1})^{-1}) \\ &= \frac{1}{|G|} Q(k^{-1}). \end{aligned}$$

□

Definição C.7 Para uma distribuição Q em um grupo G , a distribuição inversa \hat{Q} é definida por $\hat{Q}(g) := Q(g^{-1})$ para todo $g \in G$. Seja P a matriz de transição de um passeio aleatório em G com distribuição de incremento Q . Então, o passeio aleatório com distribuição de incremento \hat{Q} é exatamente o tempo reverso \hat{P} de P .

Notemos que quando $\hat{Q} = Q$, o passeio aleatório em G com distribuição de incremento Q é reversível e $P = \hat{P}$.

Proposição C.8 Seja P a matriz de transição de um passeio aleatório em um grupo G com distribuição de incremento Q e seja \hat{P} a matriz de transição do passeio aleatório em G com distribuição de incremento \hat{Q} . Seja π a distribuição uniforme em G . Então, para qualquer $t \geq 0$

$$\|Q^{t*} - \pi\|_{TV} = \|P^t(e, \cdot) - \pi\|_{TV} = \|\hat{P}^t(e, \cdot) - \pi\|_{TV} = \|\hat{Q}^{t*} - \pi\|_{TV}.$$

Prova. Seja $(X_t) = (e, X_1, X_2, \dots)$ a cadeia de Markov com matriz de transição P e posição inicial e . Podemos escrever X_k como uma palavra $X_k = g_1 g_2 \dots g_k$, onde os elementos aleatórios $g_1, g_2, \dots \in G$ são independentes escolhas dadas pela distribuição Q .

Analogamente, seja (Y_t) a cadeia de Markov com matriz de transição \hat{P} , com incrementos $h_1, h_2, \dots \in G$ escolhidos independentemente sob \hat{Q} .

Para quaisquer elementos fixados $a_1, \dots, a_t \in G$,

$$\mathbb{P}\{g_1 = a_1, \dots, g_t = a_t\} = \mathbb{P}\{h_1 = a_t^{-1}, \dots, h_t = a_1^{-1}\}$$

pela definição de \hat{P} .

Somando em todas as palavras tais que $a_1 \dots a_t = a$, temos

$$P^t(e, a) = \hat{P}^t(e, a^{-1}).$$

Portanto,

$$\sum_{a \in G} \left| P^t(e, a) - \frac{1}{|G|} \right| = \sum_{a \in G} \left| \hat{P}^t(e, a^{-1}) - \frac{1}{|G|} \right| = \sum_{a \in G} \left| \hat{P}^t(e, a) - \frac{1}{|G|} \right|$$

Logo,

$$\|P^t(e, \cdot) - \pi\|_{TV} = \|\hat{P}^t(e, \cdot) - \pi\|_{TV}.$$

□

D

Tempos de mistura analisados

Cadeia de Markov	Tempo de Mistura
Embaralhamento Top-to-random	$n \log n$
Embaralhamento Gilbert-Shannon-Reeds	$\frac{3}{2} \log_2 n$
Embaralhamento de transposições adjacentes em tempo contínuo	$\frac{1}{2\pi^2} n^2 \log n$
Embaralhamento de transposições adjacentes em tempo discreto	$\frac{1}{\pi^2} n^3 \log n$
Processo de exclusão simples simétrico com k partículas	$\frac{1}{2\pi^2} n^2 \log \min(k, n - k)$

Tabela D.1: Tabela com todas as cadeias de Markov analisadas e seus respectivos tempos de mistura.