Éric Santana Oliveira

PERFIL DO CUTOFF DO PROCESSO DE EXCLUSÃO NO GRAFO COMPLETO

Vitória
2024

Éric Santana Oliveira

PERFIL DO CUTOFF DO O PROCESSO DE EXCLUSÃO NO GRAFO COMPLETO

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
PROGRAMA DE PÓS GRADUAÇÃO EM MATEMÁTICA

Dissertação de mestrado apresentada ao PPG-MAT como parte dos requisitos exigidos para a obtenção do título de Mestre em Matemática

Orientador: Prof. Dr. Fábio Júlio da Silva Valentim
Coorientador: Prof. Dr. Milton David Jara Valenzuela

Vitória
2024

Perfil do Cutoff do Processo de Exclusão no Grafo Completo

Éric Santana Oliveira

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

Aprovada em 15 de abril de 2024 por:

____
Prof. Dr. Fábio Júlio da Silva Valentim
Universidade Federal do Espírito Santo
Orientador

____
Prof. Dr. Diogo Manuel Fernandes Bessam
Universidade Federal do Espírito Santo

____
Prof. Dr. Freddy Rolando Hernandez Romero
Universidade Nacional da Colômbia

Universidade Federal do Espírito Santo
Vitória, 15 de abril de 2024

Datas e horários baseados em Brasília, Brasil
Sincronizado com o NTP.br e Observatório Nacional (ON) em 23/04/2024 às 14:25:10

Folha de assinaturas Éric Santana Oliveira

Data e Hora de Criação: 22/04/2024 às 10:23:35

Documentos que originaram esse envelope:
- Folha de assinaturas Éric Santana Oliveira.docx (Documento Microsoft Word) - 1 página(s)

Hashs únicas referente à esse envelope de documentos

[SHA256]: 6c3079e580de3674f361b9c0002edf5c291955fb082eb601f8687d1ac133bed5
[SHA512]: ab392f83bad37eb7471b807a8bccfedbdae3839da5a2b33151bfc6b9a570b8275dd63923fdfc7a0464780cd312646b7246f07d2f56222d03b279575da046791a

Lista de assinaturas solicitadas e associadas à esse envelope

ASSINADO - Fabio Julio Da Silva Valentim (fabio.valentim@ufes.br)
Data/Hora: 22/04/2024 - 10:59:44, IP: 164.163.204.71, Geolocalização: [-20.345651, -40.294809]
[SHA256]: 1e805f8928a5fd0474ef1aefdddf4f8c6e75d5b9ad44e15fb3ad23a93d58ffce

ASSINADO - diogo.bessam@ufes.br
Data/Hora: 22/04/2024 - 11:29:11, IP: 187.36.168.30
[SHA256]: b4e051e337e36b7efe4d82aa0afd6f7f47653d7451fef4f2c2e536afbc1c1019

ASSINADO - freddyrolando@gmail.com
Data/Hora: 23/04/2024 - 14:25:10, IP: 186.154.34.46
[SHA256]: d8ab06f74d54bda9c31f784e0f0eab9d11f5b75c400301209623ab83c91cf8bf

Histórico de eventos registrados neste envelope

23/04/2024 14:25:10 - Envelope finalizado por freddyrolando@gmail.com, IP 186.154.34.46
23/04/2024 14:25:10 - Assinatura realizada por freddyrolando@gmail.com, IP 186.154.34.46
22/04/2024 11:29:11 - Assinatura realizada por diogo.bessam@ufes.br, IP 187.36.168.30
22/04/2024 10:59:44 - Assinatura realizada por fabio.valentim@ufes.br, IP 164.163.204.71
22/04/2024 10:59:39 - Envelope visualizado por fabio.valentim@ufes.br, IP 164.163.204.71
22/04/2024 10:52:24 - Envelope registrado na Blockchain por wesley.a.silva@ufes.br, IP 200.137.65.103
22/04/2024 10:52:23 - Envelope encaminhado para assinaturas por wesley.a.silva@ufes.br, IP 200.137.65.103
22/04/2024 10:23:39 - Envelope criado por wesley.a.silva@ufes.br, IP 200.137.65.103

Ficha catalográfica disponibilizada pelo Sistema Integrado de Bibliotecas - SIBI/UFES e elaborada pelo autor

---

S231p

Santana, Éric, 2000-
Perfil do cutoff do processo de exclusão no grafo completo. / Éric Santana. - 2024.
127 f. : il.

Orientador: Fábio Júlio Valentim.
Coorientador: Milton Jara.
Dissertação (Mestrado em Matemática) - Universidade Federal do Espírito Santo, Centro de Ciências Exatas.

1. Matemática. 2. Probabilidades. 3. Processos de Markov. 4. Convergência. I. Valentim, Fábio Júlio. II. Jara, Milton. III. Universidade Federal do Espírito Santo. Centro de Ciências Exatas. IV. Título.

CDU: 51

Agradecimentos

Agradeço à minha mãe por todo o amor e carinho, e por ter acreditado no meu sonho de estudar matemática. Sua confiança e suporte foram fundamentais para que eu pudesse me dedicar aos estudos.

Agradeço também aos meus avós, Darci e Branca, por terem cuidado de mim quando pequeno (e também quando grande).

Ao meu irmão, Pedro Henrique, agradeço pela parceria e pelas boas risadas ao longo dos anos.

Gostaria de expressar minha gratidão ao Fábio Júlio pelo tempo dedicado, pelos conselhos tanto na graduação (durante o PICME) quanto no mestrado, e por sempre ter sido tão solícito, além de me motivar ao longo de todo o processo.

Agradeço ao Milton Jara por ter me passado o problema e por ter tirado dúvidas sempre que necessário ao longo do caminho.

Ao Professor Juniano Vieira, agradeço por ter sido a primeira pessoa a me mostrar a "matemática de verdade".

Aos professores João Costalonga e Diogo Bessam, sou grato por terem me orientado em estudos dirigidos durante a graduação. Este último, sem dúvida, é o principal *culpado* por eu ter escolhido a probabilidade.

Agradeço aos membros da banca pelo tempo dedicado e pelas sugestões para melhorar o trabalho.

Sou grato também aos colegas da graduação e do mestrado, especialmente ao amigo Yuri Nuyryn pela parceria nos últimos anos, e ao Jomar Júnior pelas conversas e trocas de sugestões nessa reta final da dissertação.

Agradeço a todos os professores que de alguma forma contribuíram para a minha formação.

Por fim, gostaria de agradecer à Fundação de Amparo à Pesquisa do Estado do Espírito Santo (FAPES) pelo apoio financeiro durante o mestrado.

"I was fine, until I was born. And it's been downhill ever since. We live to avoid death, we exist to avoid unexistence."

-Peter Steele.

Resumo

A dissertação aborda o estudo das cadeias de Markov, processos evolutivos com "perda de memória", amplamente aplicados em áreas como biologia, estatística e finanças. A convergência dessas cadeias para uma distribuição estacionária é analisada utilizando a "distância de variação total". Introduzem-se os tempos de mistura, representando o tempo necessário para a convergência.

O conceito de acoplamento entre cadeias de Markov é apresentado, revelando sua utilidade na determinação de cotas para os tempos de mistura. Explora-se o fenômeno de cutoff, um decrescimento abrupto na distância de variação total em sequências de cadeias de Markov, proporcionando uma compreensão detalhada da convergência. O objetivo final é calcular o perfil do cutoff para o processo de exclusão no grafo completo.

Os capítulos abordam construção de cadeias, conceitos técnicos, acoplamentos e tempos de mistura, culminando na análise do fenômeno de cutoff e sua aplicação específica ao processo de exclusão no grafo completo.

Abstract

This dissertation delves into the study of Markov chains, evolutionary processes characterized by “memory loss”, widely applied in diverse fields such as biology, statistics, and finance. The convergence of these chains to a stationary distribution is analyzed using the “total variation distance”. The times of ε-mixing are introduced, representing the time required for convergence. The concept of coupling between Markov chains is presented, demonstrating its utility in determining bounds for mixing times. The phenomenon of cutoff, an abrupt decrease in total variation distance, is explored, providing a detailed understanding of convergence. The ultimate goal is to calculate the cutoff profile for the simple exclusion processes on complete graphs. Chapters cover the construction of chains, technical concepts, couplings, and mixing times, culminating in the analysis of the cutoff phenomenon and its specific application to the exclusion process in the complete graph.

Conteúdo

Introdução

Informalmente cadeias de Markov são processos evolutivos com uma característica particular que chamamos de “perda de memória”, isto é, tentar prever o futuro do processo conhecendo todo o histórico até um certo instante s é equivalente a tentar prever o futuro conhecendo o processo apenas no instante s, e particularmente dado o presente, o passado e o futuro são independentes. Esse tipo de processo tem ampla aplicação em diversas áreas do conhecimento, como biologia, estatística, finanças e física.
Parte da utilidade das cadeias de Markov reside na capacidade de extrair resultados valiosos com base em uma teoria simples. Essa teoria permite analisar a convergência e o comportamento geral dos processos, além de facilitar a implementação em algoritmos computacionais. Em muitas situações, a distribuição (ou lei) de algumas cadeias de Markov converge, ao longo do tempo, para uma distribuição específica chamada de distribuição estacionária ou invariante. Para medir essa convergência, introduzimos a “distância de variação total”, uma métrica que atribui valores de 0 a 1, indicando a proximidade da cadeia à convergência. Inicialmente, exploraremos conceitos fundamentais de cadeias de Markov, referindo-nos principalmente a [LPW] e [Nor98]. Posteriormente, seguindo a abordagem proposta por [Nor98], construiremos uma família de cadeias de Markov em tempo contínuo, e deduziremos algumas de suas propriedades.

Ao estudar uma cadeia de Markov que admite distribuição estacionária, à fim de observar o tempo que uma cadeia leva para se aproximar de sua distribuição estacionária, definimos os tempos de ε-mistura, $t_{mix}(\varepsilon)$, que representam o tempo necessário para que a distância entre uma cadeia e sua distribuição invariante seja menor do que ε.

Introduzimos o conceito de acoplamento entre duas cadeias de Markov $(X_t)_{t\in\mathbb{F}}$ e $(Y_t)_{t\in\mathbb{F}}$[(1)], que consiste em exibir uma terceira cadeia $(U_t)_{t\in\mathbb{F}} = (Z_t, W_t)_{t\in\mathbb{F}}$ de modo que cada uma das coordenadas de $(U_t)_{t\in\mathbb{F}}$ quando observadas individualmente se comportem como $(X_t)_{t\in\mathbb{F}}$ e $(Y_t)_{t\in\mathbb{F}}$ respectivamente. Através de um resultado que obteremos no Capítulo 3, mostraremos que utilizando acoplamentos é possível encontrar cotas para os tempos de mistura de uma dada

(1) $\mathbb{F} = [0, \infty)$ ou $\mathbb{N}$.

cadeia.

Um fenômeno importante observado em algumas sequências de cadeias de Markov é conhecido como cutoff. Primeiramente identificado nos trabalhos de Persi Diaconis, Siavash Shahshahani e David Aldous nos anos 80, o fenômeno em uma sequência de cadeias $(X_t^n; t \geq 0)_{n\in\mathbb{N}}$ consiste no decrescimento abrupto da distância de variação total de $1-\varepsilon$ até ε (para $\varepsilon \in (0, 1/2)$), quando observada na escala do tempo de mistura, de modo que conforme n cresce a distância de variação total na escala dos tempos de mistura t_{mix}^n se aproxima de uma função degrau.
Em alguns casos, é possível determinar a sequência de intervalos onde esse decrescimento ocorre. Nesses casos, chamamos esta sequência de *janela de cutoff*. Tanto esta parte quanto a do estudo de tempos de mistura e acoplamentos, estão baseados no conteúdo de [LPW].

Diaconis, em [Dia96], fez a seguinte afirmação: *"At present writing, proof of a cutoff is a difficult, delicate affair, requiring detailed knowledge of the chain, such as all eigenvalues and eigenvectors."*
Embora existam métodos atuais que não exijam esse conhecimento, demonstrar a ocorrência do cutoff permanece desafiador.

Por fim, em alguns casos, conseguimos apresentar uma descrição ainda mais precisa do cutoff: Se para uma sequência $(X_t^n; t \geq 0)_{n\in\mathbb{N}}$ de cadeias existe uma função real G tal que o limite da distância de variação total na escala da janela é descrito pela função G.
Neste cenário particular de cutoff, obtemos uma compreensão extremamente detalhada sobre a natureza da convergência. No entanto, em contrapartida, calcular o perfil do cutoff para uma sequência de cadeias de Markov se revela uma tarefa extraordinariamente desafiadora. Constatamos que até mesmo para processos simples, são necessárias ferramentas sofisticadas, resultando em perfis que frequentemente se apresentam como funções não tão usuais.

O objetivo final deste texto é calcular o perfil do cutoff para o processo de exclusão no grafo completo (processo definido na Seção 3.3), formalizando os conceitos mencionados e apresentando construções conforme necessário.

0

Preliminares

0.1 Probabilidade

Neste capítulo, exibiremos de maneira sucinta alguns conceitos elementares que serão utilizados no decorrer do texto. Esperamos que o leitor já domine conceitos básicos de teoria da medida como conjuntos mensuráveis, funções mensuráveis, σ-álgebras, espaços de medida, entre outros.

Definição 0.1.1 (Variável aleatória)**.** Fixados dois espaços mensuráveis $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$, uma variável aleatória é uma função $X : \Omega_1 \to \Omega_2$ mensurável.

Definição 0.1.2 (distribuição de probabilidade)**.** Fixada uma medida μ em um espaço $(\Omega, \mathcal{F})$, dizemos que μ é uma distribuição de probabilidade se $\mu(\Omega) = 1$.

Definição 0.1.3 (Espaço de Probabilidade)**.** Um espaço de medida $(\Omega, \mathcal{F}, \mathbb{P})$ é chamado espaço de probabilidade se $\mathbb{P}$ é uma medida de probabilidade.

Definição 0.1.4 (Valor Esperado/Média)**.** Fixada uma variável aleatória X definida em $(\Omega, \mathcal{F}, \mathbb{P})$ assumindo valores em $\mathbb{R}$, definimos o valor esperado de X (Notação $\mathbb{E}_{\mathbb{P}}[X]$) por

$$\mathbb{E}_{\mathbb{P}}[X] = \int_{\Omega} X \, \mathrm{d}\mathbb{P}.$$

Observação 0.1.5. Frequentemente, quando uma única medida é especificada no problema, denotamos o valor esperado de X apenas por $\mathbb{E}[X]$.

Definição 0.1.6 (Variância)**.** Fixada uma variável aleatória X definida em $(\Omega, \mathcal{F}, \mathbb{P})$ assumindo valores em $\mathbb{R}$, definimos a variância de X (Notação $\mathrm{Var}(X)$) por

$$\mathrm{Var}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2.$$

Definição 0.1.7 (Conjuntos Independentes)**.** Fixado um espaço de medida $(\Omega, \mathcal{F}, \mathbb{P})$ e uma família $\mathcal{A} = \{A_\alpha,\ \alpha \in I\}$ de eventos, dizemos que os elementos de $\mathcal{A}$ são independentes se

$$\mathbb{P}\left(\bigcap_{k \in I'} A_k\right) = \prod_{k \in I'} \mathbb{P}(A_k)$$

para qualquer $I' \subseteq I$ finito.

Definição 0.1.8 (Independência entre variáveis aleatórias)**.** Fixadas variáveis aleatórias $X_1, X_2, \ldots, X_n$ tomando valores em $(S_1, \mathcal{S}_1), \ldots, (S_n, \mathcal{S}_n)$ respectivamente. Dizemos que $X_1, X_2, \ldots, X_n$ são independentes se para quaisquer conjuntos $A_1 \in \mathcal{S}_1, \ldots, A_n \in \mathcal{S}_n$ mensuráveis tem-se:

$$\mathbb{P}\left(\bigcap_{k=1}^{n} \{X_k \in A_k\}\right) = \prod_{k=1}^{n} \mathbb{P}\left(X_k \in A_k\right).$$

Definição 0.1.9 (Independência entre variáveis aleatórias (caso geral))**.** Fixada uma família arbitrária $\mathcal{A} = \{X_\alpha\ :\ \alpha \in I\}$ de variáveis aleatórias, dizemos que as variáveis aleatórias de $\mathcal{A}$ são independentes se para qualquer $J \subseteq I$ finito as variáveis em $\{X_k\ :\ k \in J\}$ são independentes.

Nesta última parte da seção apresentaremos algumas distribuições que serão utilizadas ao longo do texto.

Definição 0.1.10 (Distribuição de uma variável aleatória)**.** Fixada uma variável aleatória X em um espaço de medida $(\Omega, \mathcal{F}, \mathbb{P})$, assumindo valores em $\mathbb{R}$[(1)], definimos a distribuição de X como sendo a medida $\mu : \mathcal{B}(\mathbb{R}) \to \mathbb{R}$, definida por

$$\mu(A) = \mathbb{P}(X \in A)$$

para todo $A \in \mathcal{B}(\mathbb{R})$[(2)].

Observação 0.1.11. Neste texto usaremos i.i.d. para nos referirmos variáveis independentes e iguais em distribuição.

Abaixo listaremos alguns exemplos de distribuições que serão utilizadas no texto. Para simplificar a notação, considere que o espaço de probabilidade em questão é $(\Omega, \mathcal{F}, \mathbb{P})$.

Exemplo 0.1.12 (Distribuição de Bernoulli)**.** Dizemos que uma variável aleatória X possui distribuição de Bernoulli com parâmetro $p \in [0, 1]$ (Notação $X \sim \text{Ber}(p)$) se

$$\mathbb{P}(X = 1) = p \quad \mathbb{P}(X = 0) = 1 - p.$$

Ou seja, uma variável com distribuição $\text{Ber}(p)$ modela o lançamento de uma moeda (honesta ou não).

---

(1) Ao longo deste texto, sempre que mencionarmos "assumindo valores em $\mathbb{R}$" queremos dizer que o espaço de chegada em questão é $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

(2) Onde $\mathcal{B}(\mathbb{R})$ é a σ-álgebra de Borel de $\mathbb{R}$

Exemplo 0.1.13 (Distribuição Binomial)**.** Dizemos que uma variável aleatória X possui distribuição Binomial com parâmetros $n \in \mathbb{N}^{*}$[(3)] e $p \in [0,1]$ (Notação $X \sim \mathrm{Bin}(n,p)$) se para qualquer $k \in \{0,1,\ldots,n\}$ tem-se

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Este tipo de cadeia modela a quantidade de sucessos observados em n lançamentos de moeda independentes.

Exemplo 0.1.14 (Distribuição Uniforme em A)**.** Fixado um conjunto finito e não-vazio A, dizemos que uma variável aleatória X possui distribuição Uniforme em A (Notação $X \sim \mathrm{Unif}(A)$) se para qualquer $a \in A$ tem-se

$$\mathbb{P}(X = a) = \frac{1}{|A|},$$

onde $|A|$ denota a quantidade de elementos em A. Este tipo de cadeia modela a escolha de um elemento aleatório de A de maneira uniforme.

Exemplo 0.1.15 (Distribuição de Poisson)**.** Dizemos que uma variável aleatória X possui distribuição de Poisson com parâmetro λ (Notação $X \sim \mathrm{Pois}(\lambda)$) se para qualquer $k \in \mathbb{N}$ tem-se

$$\mathbb{P}(X = k) = \frac{\lambda^k \, \mathrm{e}^{-\lambda}}{k!}.$$

Abaixo listaremos algumas distribuições contínuas que serão usadas no texto. Antes disso faremos algumas observações.

Observação 0.1.16. Observe que pelo fato de a σ-álgebra de Borel de $\mathbb{R}$ ser gerada por intervalos da forma $(-\infty, x)$ com $x \in \mathbb{R}$, para que μ seja a distribuição de X, basta que

$$\mu((-\infty, y)) = \mathbb{P}(X \in (-\infty, y))$$

para todo $y \in \mathbb{R}$.

Definição 0.1.17 (Função densidade)**.** Fixada uma variável aleatória X, dizemos que uma função $f : \mathbb{R} \to \mathbb{R}_{\geq 0}$ é a função densidade de X se para qualquer $x \in \mathbb{R}$ tem-se

$$\mathbb{P}(X = x) = \int_{-\infty}^{x} f(y) \, \mathrm{d}y.$$

Exemplo 0.1.18 (Distribuição Exponencial)**.** Dizemos que uma variável aleatória X possui distribuição exponencial de parâmetro $\lambda > 0$ (Notação $X \sim \mathrm{Exp}(\lambda)$) se X possui função densidade f dada por

$$f(x) = \begin{cases} \lambda \, \mathrm{e}^{-\lambda x}, & \text{se } x \geq 0; \\ 0, & \text{caso contrário.} \end{cases}$$

Observe que particularmente $\mathbb{P}(X \leq x) = 1 - e^{-\lambda x}$.

(3) Conjunto $\mathbb{N} \setminus \{0\}$.

Exemplo 0.1.19 (Distribuição Uniforme)**.** Dizemos que uma variável aleatória X possui distribuição uniforme em um intervalo finito $[a, b)$ (Notação $X \sim \text{Unif([a, b))}$) se X possui função densidade f dada por

$$f(x) = \begin{cases} 0, & \text{se } x < a; \\ \dfrac{1}{b-a}, & \text{se } a \leq x < b; \\ 0, & \text{se } x \geq b. \end{cases}$$

Desta maneira, tem-se

$$\mathbb{P}(X \leq x) = \begin{cases} 0, & \text{se } x \leq a; \\ \dfrac{x-a}{b-a}, & \text{se } a \leq x \leq b; \\ 1, & \text{se } x \geq b. \end{cases}$$

para todo $x \in \mathbb{R}$.

Exemplo 0.1.20 (Distribuição Normal)**.** Dizemos que uma variável aleatória X possui distribuição Normal de parâmetros $\mu \in \mathbb{R}$ e $\sigma^2 \geq 0$ (Notação $X \sim \mathcal{N}(\mu, \sigma^2)$) se X possui função densidade f dada por

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \, \mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

Observação 0.1.21. Geralmente em aplicações costuma-se utilizar a distribuição chamada *Normal Padrão* que consiste na distribuição Normal de parâmetros 0 e 1. Isto é, $\mathcal{N}(0, 1)$ com densidade

$$f(x) = \frac{1}{\sqrt{2\pi}} \, \mathrm{e}^{-\frac{x^2}{2}}.$$

0.2 Mixórdia

Nesta seção, como indicado no título, apresentaremos conceitos variados que serão utilizados ao longo deste texto.

Definição 0.2.1 (Matriz estocástica)**.** Fixada uma matriz $A = (a_{ij} \ : \ i, j \in I)$ com entradas em $\mathbb{R}_{\geq 0}$ [(4)] para algum conjunto de índices I enumerável[(5)]. Dizemos que A é uma matriz estocástica se para qualquer $i \in I$ tem-se

$$\sum_{j \in I} a_{ij} = 1.$$

Através de um cálculo simples obtemos o seguinte resultado.

Proposição 0.2.2. *O produto de duas matrizes estocásticas é uma matriz estocástica.*

[(4)] $\mathbb{R}_{\geq 0} = [0, \infty)$

[(5)] Neste texto a noção de enumerável também abrange conjuntos finitos.

Definição 0.2.3 (Função indicadora)**.** Fixado um conjunto A, definimos a função $\mathbf{1}_A$ denominada função indicadora de A em Ω, por

$$\mathbf{1}_A(x) = \begin{cases} 1, & \text{se } x \in A; \\ 0, & \text{se } x \notin A. \end{cases}$$

Definição 0.2.4 (conjunto das partes)**.** Fixado um conjunto A, o conjunto das partes de A é $\{E \ : \ E \subseteq A\}$, denotado por $\mathcal{P}(A)$.

Observação 0.2.5. Note que fixado um conjunto A, $(A, \mathcal{P}(A))$ é uma σ-álgebra, que rotineiramente chamamos de σ-álgebra das partes A, ou somente σ-álgebra das partes quando não houver risco de ambiguidade.

Abaixo definiremos uma noção de distância que será usada para medir convergência de cadeias de Markov.

Definição 0.2.6 (distância de variação total)**.** Dadas duas distribuições de probabilidade μ e ν definidas em um espaço $(\Omega, \mathcal{F})$, definimos a distância de variação total entre as medidas por

$$d_{TV}(\mu, \nu) := \sup_{E \in \mathcal{F}} |\mu(E) - \nu(E)|.$$

No caso em que o espaço de estados é enumerável, existe uma caracterização da distância acima que é particularmente interessante e será utilizada em todo o texto.

Proposição 0.2.7. *Se μ e ν são distribuições de probabilidade definidas em um espaço enumerável Ω, então*

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

Dem.: Seja $B = \{x \ : \ \mu(x) \geq \nu(x)\}$. Fixado qualquer evento A, temos

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

E de maneira semelhante,

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

Note que

$$\mu(B) - \nu(B) - (\nu(B^c) - \mu(B^c)) = \mu(\Omega) - \nu(\Omega) = 0,$$

o que revela que essas quantidades são iguais.

Portanto:

$$\begin{aligned} d_{TV}(\mu, \nu) &= \sup_{E \in \mathcal{F}} |\mu(E) - \nu(E)| \leq \mu(B) - \nu(B) \\ &\leq \frac{1}{2}\left(\mu(B) - \nu(B) + \nu(B^c) - \nu(B^c)\right) \end{aligned}$$

$$\leq \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)| \, .$$

Note que a desigualdade contrária vem do fato de que a cota superior é alcançada tomando $E = B$ ou B^c.

Portanto, concluímos que

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

■

Note que através da demonstração da proposição acima também obtivemos o seguinte resultado:

Corolário 0.2.8. *Com as condições e notações da Proposição 0.2.7, temos:*

$$d_{TV}(\mu, \nu) = \sum_{x \in B} (\mu(x) - \nu(x)) \, ,$$

onde $B = \{x \in \Omega \mid \mu(x) \geq \nu(x)\}$.

Definição 0.2.9 (Grafo)**.** Um grafo é um par de conjuntos $G = (V, E)$ onde $V \neq \emptyset$ e E é um subconjunto de $\{\{x, y\} \ : \ x, y \in V\}$.
Chamamos os elementos de V de *vértices* e os elementos de E de *arestas*.

Parte da utilidade dos grafos mora no fato de eles serem particularmente úteis para mapear relações entre objetos de maneira visual. O próximo exemplo exibe uma representação típica de um grafo.

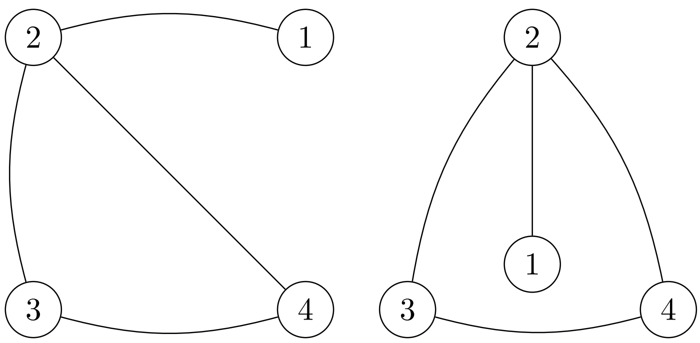


Figura 1: Duas representações distintas de um grafo G com conjunto de vértices $\{1, 2, 3, 4\}$ e conjunto de arestas $\{\{1, 2\}, \{2, 3\}, \{2, 4\}, \{3, 4\}\}$
.

Observação 0.2.10. Por abuso de notação, frequentemente usamos a palavra grafo para nos referirmos à uma representação de um grafo.

Exemplo 0.2.11 (Grafo completo de n vértices)**.** Chamamos de grafo completo de n vértices o grafo com vértices $\{x_1, \ldots, x_n\}$ e conjunto de arestas $\{\{x, y\} \ : \ x, y \in V\}$.

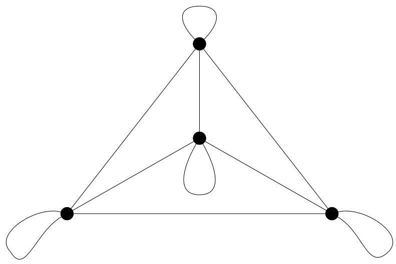


Figura 2: Grafo completo de 4 vértices

Ou seja, o grafo com n vértices formado por todas as arestas possíveis.

Observação 0.2.12. Note que a definição acima permite o que chamamos de loop (arestas da forma $\{x, x\}$. Em outros textos é comum definirmos o grafo completo de n vértices como o grafo com todas as arestas possíveis exceto loops.

Definição 0.2.13 (grau de um vértice)**.** Fixado um grafo $G = (V, E)$, definimos o grau de um vértice x (notação $\deg(x)$) como sendo o número total de arestas incidentes a x (loops são contados duas vezes por incidirem duas vezes).

Definição 0.2.14 (Caminho)**.** Fixado um grafo $G = (V, E)$ um caminho entre dois vértices x e y é uma sequência $a_0, a_1, \ldots, a_{n-1}$ de vértices tais que para qualquer $i = 0, 1, \ldots, n-1$ tem-se $\{a_i, a_{i+1}\} \in E$, onde $a_0 = x$ e $a_n = y$.

Definição 0.2.15 (Grafo Conexo)**.** Dizemos que um grafo é conexo se fixados dois vértices quaisquer existe um caminho entre eles.

O conceito que introduziremos à seguir será bastante útil na construção e na demonstração de algumas propriedades para cadeias de Markov em tempo contínuo.

Definição 0.2.16. Fixada uma matriz quadrada M, definimos a exponencial e^M de M por

$$\mathrm{e}^M := \sum_{k=0} \frac{M^k}{k!}.$$

Observação 0.2.17. Note que qualquer matriz quadrada possui exponencial, visto que

$$\left| \sum_{k=0}^{\infty} \frac{M^k(x,y)}{k!} \right| \leq \sum_{k=0}^{\infty} \left| \frac{M^k(x,y)}{k!} \right| \leq \sum_{k=0}^{\infty} \frac{s^{k-1} \|M\|_\infty^k}{k!} = \frac{\mathrm{e}^{s\|M\|_\infty}}{s},$$

onde s é a dimensão da matriz e $\|M\|_\infty$ é a maior das entradas de M.
Portanto, cada entrada de e^M está bem definida.

Definição 0.2.18 (Big-O e Little-o)**.** Escrevemos $f(x) = o(g(x))$ se

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0 \text{ ou } \lim_{x\to 0} \frac{f(x)}{g(x)} = 0$$

a depender do contexto. Na literatura costuma se dizer que f é “little-o” de g.

Escrevemos $f(x) = O(g(x))$ se

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} < \infty \text{ ou } \lim_{x\to 0} \frac{f(x)}{g(x)} < \infty$$

a depender do contexto. Na literatura costuma se dizer que f é "big-o" de g.

1
Cadeias de Markov em tempo discreto

Neste capítulo usaremos a existência de sequências de variáveis aleatórias i.i.d.[(1)] para construir cadeias de Markov em tempo discreto. Os detalhes referentes à existência estão delineados no Apêndice B.
Ao longo do capítulo, quando não especificado, estaremos supondo que a medida do espaço em questão é $\mathbb{P}$, e adicionalmente, o longo de todo o texto, sempre que o espaço de estados for enumerável, consideraremos que a σ-álgebra do espaço de chegada é a das partes.
Exploramos conceitos-chave, como a matriz de transição, distribuição estacionária, recorrência, e ao final do capítulo obteremos critérios e caracterizações para esses fenômenos. Em especial, obteremos um teorema à respeito da convergência de cadeias de Markov, que desempenhará um papel fundamental nos capítulos subsequentes. Este capítulo fundamenta-se amplamente em duas fontes principais: [LPW] e [Nor98].

1.1 Cadeias de Markov em tempo discreto

Um *processo estocástico em tempo discreto* é uma família $(X_n)_{n\in\mathbb{N}}$[(2)] de variáveis aleatórias definidas em um mesmo espaço $(\Omega, \mathcal{F}, \mathbb{P})$ e assumindo valores em um mesmo conjunto E, $(E, \mathcal{G})$. Chamamos E de *conjunto de estados*.

Abaixo enunciaremos a propriedade fundamental que caracteriza cadeias de Markov.

Definição 1.1.1 (Propriedade de Markov)**.** Dizemos que um processo estocástico em tempo discreto $(X_n)_{n\in\mathbb{N}}$ satisfaz a propriedade de Markov se

$$\mathbb{P}(X_{n+k} = y | X_0 = a_0, X_1 = a_1, \ldots, X_n = a_n) = \mathbb{P}(X_{n+k} = y | X_n = a_n)$$

(1) Independentes e iguais em distribuição.
(2) Neste texto $\mathbb{N}$ representará o conjunto dos naturais incluindo o zero.

para quaisquer $n, k \in \mathbb{N}$ e quasiquer que sejam os estados $a_0, \ldots, a_n, y$, desde que

$$\mathbb{P}(X_0 = a_0, X_1 = a_1, \ldots, X_n = a_n) > 0.$$

Uma interpretação da propriedade de Markov seria a seguinte:
Tentar prever o futuro de um processo que se conhece o histórico até um certo ponto t_0, é equivalente à tentar prever o futuro do processo conhecendo o processo apenas em t_0.

Frequentemente no contexto de cadeias de Markov nos deparamos com a famosa frase: "Dado o presente, o futuro e o passado são independentes". E isso é uma consequência direta da propriedade de Markov:

Fixados estados $a_0, a_1, \ldots, a_{n+k}$, para cada $k \in \mathbb{N}$ denotando por A_k o evento $X_n = a_k$, temos

$$\begin{aligned}\mathbb{P}(A_0, \ldots, A_{n-1}, A_{n+1}, \ldots, A_{n+k}|A_n) &= \mathbb{P}(A_{n+1}, \ldots, A_{n+k}|A_0, \ldots, A_{n-1}, A_n)\mathbb{P}(A_0, \ldots, A_{n-1}|A_n),\\ &= \mathbb{P}(A_{n+1}, \ldots, A_{n+k}|A_n)\mathbb{P}(A_0, \ldots, A_{n-1}|A_n).\end{aligned}$$

Onde a primeira igualdade segue de uma propriedade básica de probabilidade condicional, e a segunda da propriedade de Markov. Com essa motivação em mente, definimos:

Definição 1.1.2 (Cadeia de Markov)**.** Dizemos que um processo estocástico em tempo discreto $(X_n)_{n\in\mathbb{N}}$ é uma cadeia de Markov se este satisfaz a propriedade de Markov.

Abaixo apresentaremos alguns exemplos de cadeias de Markov.

Exemplo 1.1.3. Sejam $\xi_1, \xi_2, \ldots$ variáveis aleatórias independentes com distribuições $\xi_k \sim \text{Ber}\,(k^{-1})$ A sequência de variáveis aleatórias $X_0, X_1, \ldots$ definidas por

$$\begin{cases} X_0 = 0; \\ X_n(\omega) = X_{n-1}(\omega) + \xi_n(\omega). \end{cases}$$

forma uma cadeia de Markov.

De fato, observe que se $\mathbb{P}(X_0 = 0, X_1 = a_1, \ldots, X_n = a_n) > 0$ para estados $a_1, a_2, \ldots, a_n$, então fixado um estado y, temos

$$\begin{aligned}&\mathbb{P}(X_{n+t} = y \,|\, X_0 = 0, X_1 = a_1, \ldots, X_n = a_n)\\ &\quad = \mathbb{P}\left(X_{n+t} - X_n = y - a_n \,|\, X_1 = a_1, X_2 - X_1 = a_2 - a_1, \ldots, X_n - X_{n-1} = a_n - a_{n-1}\right)\\ &\quad = \mathbb{P}\left(\sum_{k=1}^{n+t}\xi_k - \sum_{k=1}^{n}\xi_k = y - a_n \,\middle|\, \xi_1 = a_1, \xi_2 = a_2 - a_1, \ldots, \xi_n = a_n - a_{n-1}\right)\\ &\quad = \mathbb{P}\left(\sum_{k=n+1}^{n+t}\xi_k = y - a_n \,\middle|\, \xi_1 = a_1, \xi_2 = a_2 - a_1, \ldots, \xi_n = a_n - a_{n-1}\right)\end{aligned}$$

Assim, utilizando o fato de $(\xi_k)_{k\in\mathbb{N}}$ ser uma sequência de variáveis independentes, temos

$$\mathbb{P}(X_{n+t} = y \,|\, X_0 = 0, X_1 = a_1, \ldots, X_n = a_n)$$

$$
\begin{aligned}
&= \mathbb{P}\left(\sum_{k=n+1}^{n+t} \xi_k = y - a_n\right) \\
&= \mathbb{P}\left(\sum_{k=n+1}^{n+t} \xi_k = y - a_n, X_n = a_n\right) \mathbb{P}(X_n = a_n)^{-1} \\
&= \mathbb{P}\left(\sum_{k=n+1}^{n+t} \xi_k = y - a_n \,\middle|\, X_n = a_n\right) \\
&= \mathbb{P}\left(\sum_{k=n+1}^{n+t} \xi_k + X_n = y \,\middle|\, X_n = a_n\right) \\
&= \mathbb{P}\left(X_{n+t} = y \mid X_n = a_n\right)
\end{aligned}
$$

revelando que de fato o processo acima é uma cadeia de Markov.
Sua dinâmica é bastante simples, uma heurística consiste em: A caixa começa com 0 bolas, e a cada instante de tempo t, com probabilidade $1/t$ uma bola é colocada na caixa. Observe que com o passar do tempo a expectativa de colocar uma bola na caixa vai ficando cada vez mais baixa, porém é fácil de ver que com probabilidade 1 o número de bolas na caixa converge para ∞! (Basta notar que a probabilidade de após qualquer tempo k nenhuma bola ser colocada na caixa é igual a 0).

Exemplo 1.1.4 (Passeio Aleatório em $\mathbb{Z}$)**.** Sejam $\xi_1, \xi_2, \ldots$ variáveis aleatórias i.i.d. com distribuição unif$(\{-1, 1\})$. Definimos o passeio aleatório em $\mathbb{Z}$ como sendo a cadeia $X_1, X_2, \ldots$ onde

$$
\begin{cases}
X_0 = 0; \\
X_n(\omega) = \sum_{k=1}^{n} \xi_k(\omega).
\end{cases}
$$

Ou seja, o processo consiste em cada instante de tempo escolher com probabilidades iguais se a cadeia "anda" uma unidade para a esquerda ou para a direita na reta.
Uma conta idêntica à do exemplo anterior mostra que de fato o passeio aleatório em $\mathbb{Z}$ é uma cadeia de Markov.

Os exemplos acima ilustram dois tipos bastante distintos de cadeias de Markov. O primeiro deles revela uma mudança de comportamento conforme o tempo avança, enquanto o segundo se comporta da mesma maneira em qualquer tempo escolhido. Chamamos cadeias que assim como no segundo exemplo não se comportam de maneira diferente conforme o tempo avança de homogêneas. Mais precisamente:

Definição 1.1.5 (Cadeia homogênea)**.** Dizemos que uma cadeia de Markov $(X_n)_{n\in\mathbb{N}}$ é homogênea se para quaisquer estados x e y tem se

$$
\mathbb{P}(X_{n+1} = y | X_n = x) = K(x, y),
$$

para algum valor $K(x, y)$ para qualquer $n \in \mathbb{N}$ sempre que $\mathbb{P}(X_n = x) > 0$.

Um aspecto interessante que deriva da definição de cadeias de Markov, é a observação de que as cadeias são caracterizadas pelas probabilidades $\mathbb{P}(X_{n+1}|X_n)$ com $n \in \mathbb{N}$. Em virtude dessa forte ligação, surge a necessidade de dar um nome a esses objetos.

Definição 1.1.6 (Matriz de transição)**.** Dizemos que uma matriz estocástica P é a matriz de transição de uma cadeia de Markov $(X_n)_{n\in\mathbb{N}}$, se para qualquer n e qualquer estado $y \in E$ com $\mathbb{P}(X_n = y) > 0$ tem se

$$P(X_{n+1} = \cdot \mid X_n = y) = P(y, \cdot).$$

Comumente trataremos de cadeias nas quais o estado inicial é regido por uma distribuição de probabilidade. Portanto, dada uma cadeia de Markov $(X_n)_{n\in\mathbb{N}}$, diremos que a cadeia possui distribuição inicial λ se para qualquer estado y tem-se

$$\mathbb{P}(X_0 = y) = \lambda(y).$$

Uma consequência da propriedade de Markov para cadeias de Markov homogêneas é o seguinte fato:

$$\mathbb{P}(X_0 = a_0, X_1 = a_1 ..., X_n = a_n) = \lambda(a_0) \prod_{k=0}^{n-1} P(a_k, a_{k+1})$$

para quaisquer $a_0, a_1, ..., a_n \in E$.
Ou seja, a probabilidade de a cadeia seguir uma determinada trajetória é o produto da probabilidade de iniciar no lugar desejado, e as probabilidades de em cada instante de tempo realizar o salto desejado.

O exemplo abaixo mostrará como construir uma cadeia de Markov através de uma matriz de transição. Isso nos permitirá exibir exemplos de uma maneira mais simples.

Exemplo 1.1.7 (Construção de Cadeias de Markov através de matrizes de transição)**.** Seja $E = \{a_0, a_2, \dots\}$ o espaço de estados desejado, λ uma distribuição em E e P uma matriz de transição também em E. Fixado $k \in \mathbb{N}^*$, defina para cada $n \in \mathbb{N}^*$ o intervalo

$$I_n^k := \left[\sum_{i=1}^{n-1} P(a_k, a_i)\, ,\, \sum_{i=1}^{n} P(a_k, a_i)\right)$$

com medida de Lebesgue $P(a_k, a_n)$.
Agora, definamos a variável aleatória $f_{a_k} := [0, 1) \to E$ por

$$f_{a_k}(x) = \begin{cases} a_1, \text{ se } x \in I_1^k; \\ a_2, \text{ se } x \in I_2^k; \\ a_3, \text{ se } x \in I_3^k; \\ \quad\vdots \end{cases}$$

Ou seja, criamos uma função escolha cuja medida correspondente a cada um dos estados da cadeia é a probabilidade de saltar de a_k para os mesmos. De maneira análoga, construímos uma variável aleatória $\theta : [0, 1) \to E$ de modo que

$$\theta(x) = \begin{cases} a_1, \text{ se } x \in J_1^k; \\ a_2, \text{ se } x \in J_2^k; \\ \quad\vdots \end{cases}$$

onde $J_1, J_2, \ldots$ são intervalos concatenados de medidas $\lambda(a_1), \lambda(a_2), \ldots$ respectivamente.

Definimos $F : [0,1) \times E \to E$ por

$$F(t,x) = f_x(t)$$

para quaisquer $x \in E, t \in [0,1)$.
Agora, seja $(\xi_n)_{n\in\mathbb{N}}$ uma sequência de variáveis aleatórias i.i.d. com $\xi_k \sim \text{unif}\,([0,1))$.
Finalmente, definimos a cadeia $(X_n)_{n\in\mathbb{N}^*}$ indutivamente por

$$\begin{cases} X_0 = \theta(\xi_0(\omega)), \\ X_n(\omega) = F\left(\xi_{n-1}(\omega), X_{n-1}(\omega)\right), \text{ para } n \in \mathbb{N}^*. \end{cases}$$

para cada ω no espaço de definição de $(\xi_n)_{n\in\mathbb{N}}$.

Note que fixados estados $b_0, b_1, \ldots, b_n$ tais que se tenha $\mathbb{P}(X_0 = b_0, X_1 = b_1, \ldots, X_n = b_n) > 0$. Fixado um estado y, temos

$$\begin{aligned} &\mathbb{P}(X_0 = b_0, X_1 = b_1, \ldots, X_n = b_n, X_{n+1} = y) \\ &\quad = \mathbb{P}\big(X_0 = b_0, F(X_0,\xi_1) = b_1, F(X_1,\xi_2) = b_2, \ldots, F(X_{n-1},\xi_n) = b_n, X_{n+1} = y\big) \\ &\quad = \mathbb{P}\big(X_0 = b_0, F(b_0,\xi_1) = b_1, F(b_1,\xi_2) = b_2, \ldots, F(b_{n-1},\xi_n) = b_n, F(X_n,\xi_{n+1}) = y\big) \end{aligned}$$

Portanto, segue da independência das variáveis $\xi_0, \xi_1, \ldots$ que

$$\begin{aligned} &\mathbb{P}(X_0 = b_0, X_1 = b_1, \ldots, X_n = b_n, X_{n+k} = y) \\ &\quad = \mathbb{P}(X_0 = b_0)\mathbb{P}(F(b_0,\xi_1) = b_1)\mathbb{P}(F(b_1,\xi_2) = b_2)\cdots\mathbb{P}(F(b_{n-1},\xi_n) = b_n)\mathbb{P}(F(X_n,\xi_{n+1}) = y) \\ &\quad = \mathbb{P}(X_0 = b_0)\mathbb{P}(f_{b_0}(\xi_1) = b_1)\mathbb{P}(f_{b_1}(\xi_2) = b_2)\cdots\mathbb{P}(f_{b_{n-1}}(\xi_n) = b_n)\mathbb{P}(f_{b_n}(\xi_{n+1}) = y) \\ &\quad = \lambda(a_0)P(b_0,b_1)P(b_1,b_2)\cdots P(b_{n-1},b_n)P(b_n,y). \end{aligned}$$

E portanto, vale

$$\mathbb{P}(X_t = z) = (\lambda P^t)(z)$$

para qualquer estado z e $t \in \mathbb{N}$.
Note que particularmente, para quaisquer estados u e v e qualquer $k \in \mathbb{N}$ tem-se

$$\mathbb{P}(X_{k+1} = v \mid X_k = u) = P(u,v),$$

desde que $\mathbb{P}(X_k = u) > 0$.
Ou seja, de fato as probabilidades de transição conferem. Assim, só nos resta mostrar que de fato o processo $(X_n)_{n\in\mathbb{N}}$ é uma cadeia de Markov.

Note que por consequência dos cálculos acima, temos

$$\begin{aligned} &\mathbb{P}(X_0 = b_0, X_1 = b_1, \ldots, X_n = b_n, X_{n+k} = y) \\ &\qquad = \sum_{b_{n+1},\ldots,b_{n+k-1}} \lambda(a_0)P(b_0,b_1)P(b_1,b_2)\cdots P(b_{n+k-2},b_{n+k-2})P(b_{n-k+1},y) \end{aligned}$$

Portanto,

$$\begin{aligned}
&\mathbb{P}(X_{n+k} = y \mid X_0 = b_0, X_1 = b_1, \ldots, X_n = b_n) \\
&\quad = \sum_{b_{n+1},\ldots,b_{n+k-1} \in E} \frac{\lambda(a_0)P(b_0,b_1)P(b_1,b_2)\cdots P(b_{n+k-2},b_{n+k-2})P(b_{n-k+1},y)}{\lambda(a_0)P(b_0,b_1)P(b_1,b_2)\cdots P(b_{n-1},b_n)} \\
&\quad = \sum_{b_{n+1},\ldots,b_{n+k-1} \in E} P(b_n,b_{n+1})\cdots P(b_{n+k-2},b_{n+k-2})P(b_{n-k+1},y) \\
&\quad = \sum_{b_{n+1},\ldots,b_{n+k-1} \in E} \mathbb{P}(F(b_n,\xi_{n+1}) = b_{n+1})\cdots \mathbb{P}(F(b_{n+k-1},\xi_{n+k}) = b_{n+k-1}) \\
&\quad = \sum_{b_{n+1},\ldots,b_{n+k-1} \in E} \mathbb{P}(F(b_n,\xi_{n+1}) = b_{n+1},\cdots, F(b_{n+k-1},\xi_{n+k}) = y) \\
&\quad = \sum_{b_{n+1},\ldots,b_{n+k-1} \in E} \mathbb{P}(F(b_n,\xi_{n+1}) = b_{n+1},\cdots, F(b_{n+k-1},\xi_{n+k}) = y \mid X_n = b_n) \\
&\quad = \sum_{b_{n+1},\ldots,b_{n+k-1} \in E} \mathbb{P}(F(X_n,\xi_{n+1}) = b_{n+1},\cdots, F(X_{n+k-1},\xi_{n+k}) = b_{n+k-1} \mid X_n = b_n) \\
&\quad = \sum_{b_{n+1},\ldots,b_{n+k-1} \in E} \mathbb{P}(X_{n+1} = b_{n+1},\cdots, X_{n+k-1} = b_{n+k-1}, X_{n+k} = y \mid X_n = b_n) \\
&\quad = \mathbb{P}(X_{n+k} = y \mid X_n = b_n).
\end{aligned}$$

Portanto, de fato o processo $(X_n)_{n\in\mathbb{N}}$ é uma cadeia de Markov homogênea e com matriz de transição P.

Observação 1.1.8. Note que a construção acima pode ser feita para qualquer distribuição inicial ν. Daremos um destaque para as distribuições δ_x com $x \in E$. Por abuso de notação denotaremos por $\mathbb{P}_x$ a medida $\mathbb{P}$ aplicada à cadeia com matriz de transição P e distribuição inicial δ_x, ou seja, uma "cópia" de $(X_t)_{t\in\mathbb{N}}$.

Observação 1.1.9. No exemplo acima, poderíamos obter uma forma de gerar cadeias não-homogêneas através de matrizes de transição. Para isso, ao invês de definirmos as variáveis aleatórias f_{a_k} correspondentes à $P(a_k,\cdot)$, precisaríamos definir variáveis aleatórias $(f^n_{a_k})_{n\in\mathbb{N}^*}$, onde cada $f^n_{a_k}$ corresponderia à $P(a_k,\cdot,n)$.
Como o restante da demonstração é idêntica, consideraremos essa construção de cadeias não-homogêneas.

Observação 1.1.10. Em rigor, para mostrar que o processo definido no exemplo é uma cadeia de Markov seria necessário verificar que cada um dos X_k é mensurável. Este detalhe é feito no Lema B.0.3 do Apêndice B.

Exemplo 1.1.11 (Cadeia de Nascimento-Morte)**.** Seja $E = \mathbb{N}$ ou $\{0, 1, \ldots, n\}$ para algum $n \in \mathbb{N}^*$. Construimos uma matriz de transição atribuindo para cada $x \in E$ probabilidades p_x, q_x de modo que $P(x, x+1) = p_x, P(x, x-1) = q_x$, e $P(x, x) = 1 - (p_x + q_x)$ (Nos casos extremos 0 ou n tem-se $p_n = q_0 = 0$).

Exemplo 1.1.12 (Urna de Pólya)**.** Podemos definir a Urna de Pólya como sendo o processo gerado pela matriz de transição P dada por:

$$P(x,y,n) = \begin{cases} 1 - \dfrac{x}{n+2}, & \text{se } y = x \\ \dfrac{x}{n+2}, & \text{se } y = x+1 \end{cases}$$

e estado inicial $X_0 = 1$.
Basicamente a cadeia consiste no seguinte processo: Inicialmente temos duas urnas com uma bola cada. A cada instante escolhemos uma bola ao acaso de maneira uniforme adicionamos mais uma bola à urna na qual a bola escolhida está. O processo consiste em acompanhar o número de bolas na primeira urna. Uma observação interessante é que a Urna de Pólya é uma cadeia de nascimento-morte com a particularidade de ter $q_x = 0$ para todo $x \in \mathbb{N}$. Além disso, a urna de Pólya é não-homogênea.

Exemplo 1.1.13 (Passeio aleatório em $\mathbb{Z}_k$)**.** Considere a matriz de transição P em $\mathbb{Z}_k$ dada por

$$P(x,y) = \begin{cases} p, & \text{se } y = x+1; \\ q, & \text{se } y = x-1; \\ 1-p-q, & \text{se } y = x. \end{cases}$$

onde $p, q > 0$ com $p + q \leq 1$. Chamamos uma cadeia com matriz de transição P de passeio aleatório em $\mathbb{Z}_k$. O caso mais simples é quando $p = q = 1/2$, que é o passeio simples sem viés.

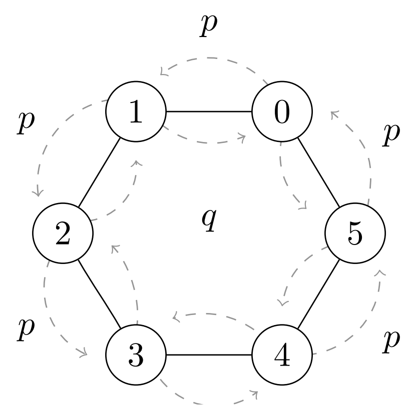


Figura 1.1: Passeio aleatório em $\mathbb{Z}_6$ com probabilidades de transição representadas visualmente.

Exemplo 1.1.14 (Passeio aleatório simples em um grafo)**.** Dado um grafo conexo G, definimos o passeio aleatório simples em G como sendo a cadeia de Markov com matriz de transição P, onde

$$P(x,y) = \begin{cases} 0, & \text{se } x \text{ e } y \text{ não são vizinhos ;} \\ \dfrac{1}{\deg(x)}, & \text{se } x \text{ e } y \text{ são vizinhos.} \end{cases}$$

Ou seja, em cada instante de tempo, escolhe-se um vizinho do estado atual de maneira aleatória e uniforme. A figura abaixo ilustra um passeio aleatório em um grafo.

O exemplo seguinte é um "sub-exemplo" do exemplo acima.

Exemplo 1.1.15 (Passeio aleatório no toro discreto d-dimensional)**.** Fixado $n \in \mathbb{N}^*$, Considere o grafo com conjunto de vértices $\mathbb{Z}_n^d$, e tal que dois vértices $(x_1, \ldots, x_d)$ e $(y_1, \ldots, y_d)$ são vizinhos se e somente se existe $j \in \{1, \ldots, d\}$ de modo que $x_j \equiv y_j \pm 1 \pmod{n}$ e $x_i = y_i$ para qualquer $i \neq j$.
A dinâmica do passeio aleatório no toro d-dimensional consiste em cada instante de tempo escolher um coordenada de maneira aleatoriamente uniforme, e posteriormente escolher se essa coordenada aumenta ou diminui em uma unidade (módulo n) também de maneira aleatoriamente uniforme.
Formalmente, o passeio aleatório no toro d-dimensional é a cadeia de Markov com matriz de transição P dada por

$$P(x,y) = \begin{cases} \dfrac{1}{2d}, \text{ se } y = x \pm e_j \text{ para algum } j \in \{1, \ldots, d\}\,; \\ 0, \text{caso contrário.} \end{cases}$$

Figura 1.2: Representação do toro discreto 2-dimensional $\mathbb{Z}_9^2$.

Uma pergunta interessante a se fazer, é: Fixado $M \in \mathbb{N}$, qual a probabilidade de observarmos $X_M = a$ dado que $X_0 = a_0$? Essa pergunta nos leva a introduzir o conceito de matriz de transição em k passos, que como o nome indica, se trata da matriz que carrega as probabilidades de todas as transições em um intervalo temporal igual a k.

Definição 1.1.16 (Matriz de transição em k passos)**.** Seja $(X_n)_{n\in\mathbb{N}}$ uma cadeia de Markov com espaço de estados E, homogênea e com matriz de transição P. Fixado $k \in \mathbb{N}^*$ chamamos a matriz P^k definida por

$$P^k(x,y) = \mathbb{P}_x(X_k = y).$$

Atráves da matriz de transição em k passos, podemos gerar uma nova cadeia de markov, desta vez com matriz de transição P^k. Por uma certa ótica, esta cadeia é uma versão acelerada da cadeia original.

Observação 1.1.17. Pela observação à respeito da propriedade de Markov, sabemos que

$$P^n(a_0, a_n) = \sum_{a_1, \ldots, a_{n-1} \in E} \prod_{k=0}^{n-1} P(a_k, a_{k+1}).$$

Essa observação é suficiente para obtermos o próximo resultado.

Proposição 1.1.18 (Equações de Chapman-Kolmogorov/Propriedade de Semigrupo)**.** *Seja $(X_n)_{n\in\mathbb{N}}$ uma cadeia de Markov com espaço de estados E, homogênea e com matriz de transição P. A seguinte equação é válida para quaisquer $k, n \in \mathbb{N}^*$ e $x, y \in E$*

$$P^{k+n}(x,y) = \left(P^k P^n\right)(x,y).$$

Já conhecemos as probabilidades de uma cadeia homogênea atingir determinado estado no tempo k quando $k \in \mathbb{N}$. Porém muitas das vezes estamos interessados em expressar a probabilidade da cadeia atingir um estado y no tempo τ quando τ é uma variável aleatória. Nesse contexto, surge o que chamamos de propriedade forte de Markov.
Para este capítulo definiremos o seguinte objeto de uma maneira mais ingênua, e por consequência enunciaremos uma versão mais ingênua da propriedade forte de Markov.

Definição 1.1.19 (Tempo de Parada*)**.** Fixados uma cadeia de Markov X, e uma variável aleatória τ, dizemos que τ é um tempo de parada para X se τ assume valores em $\mathbb{N} \cup \{\infty\}$ e para qualquer $t \in \mathbb{N}$, o evento $\{\tau = t\}$ depende apenas de $X_1, \ldots, X_t$.

Proposição 1.1.20 (Propriedade Forte de Markov*)**.** *Sejam: $(X_n)_{n\in\mathbb{N}}$ uma cadeia de Markov com distribuição inicial λ e matriz de transição P; τ um tempo de parada para $(X_n)_{n\in\mathbb{N}}$. Dado $\tau < \infty$ e $X_\tau = i$ o processo $Y = (Y_k)_{k\in\mathbb{N}}$ definido por $Y_k = X_{\tau+k}$ é uma cadeia de Markov com matriz de transição P. Além disso, dado $\tau < \infty$ e $X_\tau = i$, Y é independente de $X_0, X_1, \ldots, X_\tau$.*

Observação 1.1.21. Em textos mais avançados, é comum definir cadeias de Markov em termos de filtrações, especificando o espaço mensurável em questão, porém ainda é possível estudar a teoria básica sem dar muita ênfase à essas tecnicalidades. Não faremos a demonstração para cadeias em tempo discreto, mas utilizaremos o fato de que cadeias homogêneas satisfazem a propriedade forte de Markov. A demonstração pode ser encontrada em [Nor98].
No capítulo seguinte definiremos os conceitos acima de maneira formal, e faremos a demonstração para cadeias em tempo contínuo. A demonstração para o caso discreto *mutatis mutandis* é análoga à demonstração no caso contínuo.

Exemplo 1.1.22 (Ruína do Apostador)**.** Considere a cadeia de Markov com matriz de transição P em $\{0, 1, \ldots, n\}$ dada por

$$P(x,y) = \begin{cases} \dfrac{1}{2}, & \text{se } x \notin \{0,n\} \text{ e } y \in \{x-1, x+1\}; \\ 1, & \text{se } x \in \{0,n\} \text{ e } y = x; \\ 0, & \text{caso contrário.} \end{cases}$$

Uma interpretação para cadeia é a seguinte: Um apostador em um cassino com uma determinada quantia inicial decide apostar em um jogo de modo que a cada rodada a sua fortuna aumenta em uma unidade ou diminui em uma unidade com probabilidades iguais. Assim, o apostador prossegue até perder tudo ou atingir uma fortuna igual a n, onde o mesmo para de apostar.
Denotando por τ o tempo no qual o apostador para de apostar (que é um tempo de parada), mostraremos que

$$\mathbb{P}_k(X_{\tau=n}) = \frac{k}{n} \quad \text{e} \quad \mathbb{E}_k[\tau] = k(n-k),$$

para qualquer $k \in \{0, 1, \ldots, n\}$.

Por hora assumiremos que $\mathbb{P}(\tau < \infty) = 1$, e posteriormente no capítulo obteremos um resultado que mostra que de fato a igualdade vale.

Observe que

$$\mathbb{P}_0(X_\tau = n) = 0 \quad \text{e} \quad \mathbb{P}_n(X_\tau = n) = 1.$$

Além disso, para $k \in \{1, \ldots, n-1\}$, temos

$$\begin{aligned}
\mathbb{P}_k(X_\tau = n) &= \mathbb{P}_k(X_\tau = n,\, X_1 = k-1) + \mathbb{P}_k(X_\tau = n,\, X_1 = k+1)\\
&= \sum_{s=1}^{\infty} \mathbb{P}_k(X_\tau = n,\, \tau = s,\, X_1 = k-1) + \sum_{s=1}^{\infty} \mathbb{P}_k(X_\tau = n,\, \tau = s,\, X_1 = k+1)\\
&= \sum_{s=1}^{\infty} \mathbb{P}_k\left(X_1 = k+1,\, X_2 \notin \{0,\, n\},\, \ldots,\, X_{s-1} \notin \{0,\, n\},\, X_s = n\right)\\
&\qquad + \sum_{s=1}^{\infty} \mathbb{P}_k\left(X_1 = k-1,\, X_2 \notin \{0,\, n\},\, \ldots,\, X_{s-1} \notin \{0,\, n\},\, X_s = n\right)\\
&= \sum_{s=1}^{\infty} \left[\mathbb{P}_k\left(X_2 \notin \{0,\, n\},\, \ldots,\, X_{s-1} \notin \{0,\, n\},\, X_s = n \mid X_1 = k+1\right) \frac{1}{2}\right]\\
&\qquad + \sum_{s=1}^{\infty} \left[\mathbb{P}_k\left(X_2 \notin \{0,\, n\},\, \ldots,\, X_{s-1} \notin \{0,\, n\},\, X_s = n \mid X_1 = k-1\right) \frac{1}{2}\right].
\end{aligned}$$

Assim, utilizando o Propriedade de Markov obtemos

$$\begin{aligned}
\mathbb{P}_k(X_\tau = n) &= \sum_{s=1}^{\infty} \left[\mathbb{P}_{k+1}\left(X_1 \notin \{0,\, n\},\, \ldots,\, X_{s-2} \notin \{0,\, n\},\, X_{s-1} = n\right) \frac{1}{2}\right]\\
&\qquad + \sum_{s=1}^{\infty} \left[\mathbb{P}_{k+1}\left(X_1 \notin \{0,\, n\},\, \ldots,\, X_{s-2} \notin \{0,\, n\},\, X_{s-1} = n\right) \frac{1}{2}\right]\\
&= \sum_{s=1}^{\infty} \left[\frac{1}{2}\mathbb{P}_{k+1}(\tau = s-1) + \frac{1}{2}\mathbb{P}_{k-1}(\tau = s-1)\right]\\
&= \frac{1}{2}\mathbb{P}_{k+1}(X_\tau = n) + \frac{1}{2}\mathbb{P}_{k-1}(X_\tau = n).
\end{aligned}$$

Observe que desta forma, obtemos o sistema

$$\begin{cases}
\mathbb{P}_n(X_\tau = n) = 1;\\
\mathbb{P}_0(X_\tau = n) = 0;\\
\mathbb{P}_k(X_\tau = n) = \dfrac{1}{2}\mathbb{P}_{k+1}(X_\tau = n) + \dfrac{1}{2}\mathbb{P}_{k-1}(X_\tau = n), \text{ se } k \notin \{0, n\}.
\end{cases}$$

Resolvendo o sistema com as equações e condições de contorno acima obtemos

$$\mathbb{P}_k(\tau = n) = k/n,$$

para todo $k \in \{0, \ldots, n\}$.

Já para o valor esperado, sabemos que

$$\mathbb{E}_0[\tau] = \mathbb{E}_n[\tau] = 0.$$

Portanto, para $k \in \{1, \ldots, n-1\}$, temos

$$\begin{aligned}
\mathbb{E}_k[\tau] &= \sum_{s=1}^{\infty} [\mathbb{P}_k(\tau = s)\, s] \\
&= \sum_{s=1}^{\infty} [\mathbb{P}_k(X_1 = k-1,\, X_2 \notin \{0,n\},\, \ldots,\, X_{s-1} \notin \{0,n\},\, X_s \in \{0,n\})\, s] \\
&\quad + \sum_{s=1}^{\infty} [\mathbb{P}_k(X_1 = k+1,\, X_2 \notin \{0,n\},\, \ldots,\, X_{s-1} \notin \{0,n\},\, X_s \in \{0,n\})\, s],
\end{aligned}$$

onde repetindo os passos da conta anterior obtemos

$$\begin{aligned}
\mathbb{E}_k[\tau] &= \sum_{s=1}^{\infty} \left[\mathbb{P}_{k-1}(\tau = s-1)\frac{s}{2}\right] + \sum_{s=1}^{\infty} \left[\mathbb{P}_{k+1}(\tau = s-1)\frac{s}{2}\right] \\
&= \sum_{s=1}^{\infty} \left[\mathbb{P}_{k-1}(\tau = s-1)\frac{s-1}{2}\right] + \sum_{s=1}^{\infty} \left[\mathbb{P}_{k+1}(\tau = s-1)\frac{s-1}{2}\right] \\
&\quad + \sum_{s=1}^{\infty} \left[\mathbb{P}_{k-1}(\tau = s-1)\frac{1}{2}\right] + \sum_{s=1}^{\infty} \left[\mathbb{P}_{k+1}(\tau = s-1)\frac{1}{2}\right] \\
&= \sum_{s=1}^{\infty} \left[\mathbb{P}_{k-1}(\tau = s)\frac{s}{2}\right] + \sum_{s=1}^{\infty} \left[\mathbb{P}_{k+1}(\tau = s)\frac{s}{2}\right] + \frac{1}{2} + \frac{1}{2} \\
&= 1 + \frac{1}{2}\mathbb{E}_{k-1}[\tau] + \frac{1}{2}\mathbb{E}_{k+1}[\tau].
\end{aligned}$$

Resolvendo o sistema de equações lineares:

$$\begin{cases}
\mathbb{E}_0[\tau] = 0; \\
\mathbb{E}_n[\tau] = 0; \\
\mathbb{E}_k[\tau] = 1 + \frac{1}{2}\mathbb{E}_{k-1}[\tau] + \frac{1}{2}\mathbb{E}_{k+1}[\tau], \text{ se } k \notin \{0,n\}.
\end{cases}$$

Concluímos que

$$\mathbb{E}_k[\tau] = k(n-k)$$

para todo $k \in \{0, \ldots, n\}$.

Abaixo, definiremos alguns tempos de parada que serão fundamentais no estudo de cadeias de Markov em tempo discreto.

Definição 1.1.23 (Tempo de primeira visita)**.** A variável aleatória T_i^k em $(\Omega, \mathcal{F}, \mathbb{P})$ definida por

$$T_i(\omega) = \inf\{t \in \mathbb{N}^* | X_t(\omega) = i\}$$

para todo $\omega \in \Omega$.

Definição 1.1.24 (Tempo de k-ésima visita)**.** Definimos indutivamente as Variáveis aleatórias T_i^k em $(\Omega, \mathcal{F}, \mathbb{P})$ por

$$T_i^k(\omega) = \inf\left\{t \in \mathbb{N}^* | X_t(\omega) = i \ , \ t \geq T_i^{(k-1)}(\omega)\right\}$$

para todo $\omega \in \Omega$.

Observação 1.1.25. Note que a propriedade forte de Markov nos permite por exemplo, calcular as probabilidades de uma cadeia estar em um determinado estado k unidades de tempo depois de visitar um outro estado qualquer pela primeira vez.

Agora, direcionaremos nossos esforços para tentar compreender o comportamento assintótico de uma cadeia em tempo discreto.

Definição 1.1.26 (distribuição estacionária/invariante)**.** Dizemos que uma distribuição π definida em E é invariante com respeito à uma matriz estocástica P, se

$$\pi P = \pi.$$

Observação 1.1.27. Comumente chamamos π de medida estacionária da cadeia (X_n), quando esta última é gerada por P.

Observação 1.1.28. Se μ é uma medida (não necessariamente com $\mu(E) = 1$ ou sequer finita) satisfazendo $\mu P = \mu$, então μ é apenas uma *medida invariante* para P. É importante fazer essa distinção, pois este detalhe será importante em um resultado futuro.

Note que se π é invariante para P, segue das Equações de Chapman-Kolmogorov (Proposição 1.1.18) que $\pi P^k = \pi$ para qualquer $k \in \mathbb{N}$.

Exemplo 1.1.29 (Passeio aleatório simples em $\mathbb{Z}_k$)**.** Verificaremos que a distribuição uniforme em $\mathbb{Z}_k$ é uma medida invariante para o passeio aleatório em $\mathbb{Z}_k$ (definido no Exemplo 1.1.13). Note que:

$$\begin{aligned}\pi P(x) &= \sum_{y \in \mathbb{Z}} \pi(y)P(y,x) = \pi(x-1)P(x-1,x) + \pi(x)P(x,x) + \pi(x+1)P(x+1,x) \\ &= \frac{1}{k}\left[p + (1-p-q) + q\right] = \frac{1}{k} = \pi(x).\end{aligned}$$

Portanto, concluímos que a distribuição uniforme é invariante para P.

Com a finalidade de obter uma condição suficiente e de verificação mais simples para que uma medida seja estacionária para uma cadeia, surgem o que costumamos chamar de condições de equilíbrio.

Definição 1.1.30 (Equilíbrio/Equações de equilíbrio)**.** Dizemos que π e P estão em equilíbrio se as equações

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

chamadas de equações de equilíbrio são válidas para quaisquer estados $x, y \in E$.

Através do próximo resultado obteremos uma maneira mais simples de verificar que uma dada medida é invariante à uma cadeia fixada, que embora não seja uma condição necessária, é suficiente.

Proposição 1.1.31. *Se uma distribuição π satisfaz as equações de equilíbrio para uma matriz de transição P (ambos definidos em E), então π é estacionária com respeito à P.*

Dem.: Fixado um estado $y \in E$, note que

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

onde a segunda igualdade vem do fato de π e P estarem em equilíbrio. ∎

Observação 1.1.32. Em geral, quando as equações de equilíbrio são satisfeitas, na literatura costuma-se dizer que a cadeia é reversível, porém não exploraremos este conceito e suas propriedades neste texto.

Exemplo 1.1.33 (Medida estacionária para passeio aleatório no grafo)**.** Fixado um grafo finito e conexo G, uma distribuição invariante para o passeio aleatório simples em G é π, dada por

$$\pi(x) := \frac{d(x)}{2E(G)}$$

para todo $x \in G$.
De fato, note que se x, y são vizinhos, então

$$\pi(x)P(x,y) = \frac{d(x)}{2E(G)}\frac{1}{d(x)} = \frac{1}{2E(G)} = \frac{d(y)}{2E(G)}\frac{1}{d(y)} = \pi(y)P(y,x).$$

Se x e y não são vizinhos, então é imediato que $\pi(x)P(x,y) = \pi(y)P(y,x)$. Como π satisfaz as equações de equilíbrio para P, concluímos que de fato π é invariante para P.
Particularmente, se o grafo for regular, isto é, o grau de todos os vértices for igual, então a medida uniforme é invariante à cadeia. Em um resultado posterior, mostraremos que de fato essa é a única distribuição invariante para essa cadeia.

Exemplo 1.1.34 (distribuição invariante para Nascimento-Morte)**.** Considere uma cadeia de nascimento-Morte X com probabilidades p_x e q_x como no Exemplo 1.1.11. Definiremos recursivamente uma medida μ que seja invariante para a cadeia.
inicialmente suponha que $\mu(0) = 1$. Supondo que μ satisfaz as condições de equilíbrio, temos

$$\mu(0)p_0 = \mu(1)q_1,$$

e portanto $\mu(1) = \dfrac{p_0}{q_1}$.
Repetindo o processo para $\mu(2)$, temos

$$\mu(1)p_1 = \mu(2)q_2,$$

e portanto $\mu(2) = \dfrac{p_0 p_1}{q_1 q_2}$.
Assim, podemos definir a medida μ por

$$\mu(k) = \prod_{n=0}^{k-1} \frac{p_n}{q_{n+1}},$$

para cada k no espaço de estados.
Segue da construção de μ que a mesma é invariante para a cadeia.
Note que se o espaço de estados E for finito, então a medida π dada por

$$\pi(k) = \frac{\mu(k)}{\sum_{x \in E} \mu(x)}$$

é uma distribuição invariante para a cadeia.
Já para o caso em que o espaço de estados não é finito, para obter uma distribuição invariante, é necessário que a soma dos valores de μ seja finita.

Sabemos verificar se uma certa dada medida é invariante para uma fixada cadeia, porém, ainda não conhecemos nenhum critério que garanta a existência de uma distribuição invariante. O próximo resultado garante a existência de uma medida invariante sob a hipótese de o estado de espaços ser finito.

Proposição 1.1.35 (Existência de distribuição invariante para espaços finitos)**.** *Se X é uma cadeia de Markov homogênea com matriz de transição P em um espaço de estados finito E, então X possui pelo menos uma distribuição invariante.*

Dem.: Para cada $n \in \mathbb{N}^*$ defina $Q_n := n^{-1} \sum_{t=0}^{n} P^t$. Fixada uma distribuição de probabilidade λ qualquer em E, defina para cada $n \in \mathbb{N}^*$ a distribuição $\lambda_n := \lambda Q_n$.
Note que para qualquer $n \in \mathbb{N}^*$, λ_n está na bola unitária, que é um conjunto compacto de $\mathbb{R}^{|E|}$.
Portanto, a sequência $(\lambda_n)_{n \in \mathbb{N}^*}$ admite subsequência $(\lambda_{n_k})_{k \in \mathbb{N}^*}$ convergente.
Mostraremos que $\lim_{n \to \infty} \lambda_{n_k}$ é uma distribuição invariante para a cadeia.
Note que

$$\|\lambda_{n_k}(I - P)\|_1 = \frac{\|\lambda(I - P^{n_k+1})\|_1}{n_k} \leq \frac{2}{n_k},$$

onde $\|\cdot\|_1$ denota a norma da soma. Isso mostra que mostra que

$$\lim_{k \to \infty} \lambda_{n_k} = \lim_{k \to \infty} \lambda_{n_k} P = \left(\lim_{k \to \infty} \lambda_{n_k}\right) P.$$

■

Definição 1.1.36. Fixada uma cadeia de Markov com matriz de transição P, e $i, j \in E$, dizemos que j é acessível à partir de i (Notação: $i \to j$) se existe $k \in \mathbb{N}_+$ tal que $P^k(i, j) > 0$. Se $i \to j$ e $j \to i$, então dizemos que i e j se comunicam. (Notação: $i \leftrightarrow j$).

Observação 1.1.37. Note que a definição acima deixa implícito que dados estados x e y tais que $x \to y$, existem $n \in \mathbb{N}_+$, e estados $x_1, ..., x_{n-1} \in E$ de modo que

$$P^n(x, y) \geq p(x, x_1) p(x_1, x_2) \cdots p(x_{n-1}, y) > 0.$$

Além disso, é imediato que a recíproca é verdadeira.

É possível perceber também que "se comunicar" é uma relação de equivalência, e em alguns textos, suas classes são chamadas de *Classes Comunicadoras.*

Definição 1.1.38 (Período)**.** Fixada uma cadeia de Markov, definimos o período de um estado x por

$$\mathrm{Per}(x) := \mathrm{mdc}\left\{n \in \mathbb{N}_+ \ ; \ p^n(x,x) > 0\right\}.$$

Proposição 1.1.39. *Quaisquer estados $x, y \in E$ espaço de estados na mesma classe comunicadora possuem o mesmo período (Período é uma propriedade de classe).*

Dem.: Fixados $x, y \in E$ em uma mesma classe, sejam $m = \mathrm{Per}(x)$, $n = \mathrm{Per}(y)$, $X := \left\{n \in \mathbb{N}_+ \ ; \ p^n(x,x) > 0\right\}$ e $Y := \left\{n \in \mathbb{N}_+ \ ; \ p^n(y,y) > 0\right\}$.
Como $x \leftrightarrow y$, existem $l, k \in \mathbb{N}$ de modo que $p^l(x,y) > 0$ e $p^k(y,x) > 0$. Note que $\mathrm{Per}(x)$ e $\mathrm{Per}(y)$ dividem $l+k$ (basta concatenar os caminhos). Além disso, para qualquer $w \in X$ segue que $\mathrm{Per}(y)|(l+k+w)$, e consequentemente $\mathrm{Per}(y)|w$. Portanto $\mathrm{Per}(y)|\,\mathrm{Per}(x)$.
Por um argumento análogo segue que $\mathrm{Per}(x)|\,\mathrm{Per}(y)$.
Assim, concluímos que $\mathrm{Per}(y) = \mathrm{Per}(x)$.

■

Definição 1.1.40 (cadeia irredutível)**.** Dizemos que uma cadeia de Markov $(X_n)_{n\in\mathbb{N}}$ com espaço de estados E é irredutível se quaisquer $x, y \in E$ se comunicam.

Tendo em vista a Proposição 1.1.39, o resultado abaixo é imediato.

Corolário 1.1.41. *Se uma cadeia é irredutível, então todos os seus estados possuem o mesmo período.*

Proposição 1.1.42. *Seja X é uma cadeia de Markov irredutível e com matriz de transição P. Se π é uma medida invariante para P então para qualquer x no espaço de estados tem-se $\pi(x) > 0$.*

Dem.: Seja y um estado qualquer. Tome um estado x de modo que $\pi(x) > 0$. Como a cadeia é irredutível, pela Observação 1.1.37 existe $k \in \mathbb{N}^*$ de forma que $P^k(x,y) > 0$.
Portanto,

$$\pi(y) = (\pi P^k)(y) \geq \pi(x) P^k(x,y) > 0.$$

■

Definição 1.1.43 (Cadeia aperiódica)**.** Dizemos que uma cadeia é aperiódica se todos os seus estados possuem período igual a 1. Particularmente, se uma cadeia irredutível possui um estado com período igual a 1, então a cadeia é aperiódica.

Existe um método para transformar uma cadeia qualquer em uma cadeia com uma dinâmica análoga, que seja irredutível. O próximo exemplo mostra o processo.

Exemplo 1.1.44 (Versão lazy de uma cadeia)**.** Seja X uma cadeia gerada pela matriz de transição P. Chamamos a cadeia Y com matriz de transição $\frac{P+\mathrm{Id}}{2}$ de versão lazy de X.
Note que o processo consiste em atribuir $1/2$ de probabilidade de a cadeia permanecer no mesmo estado, e redistribuir o $1/2$ restante de maneira proporcional à P.

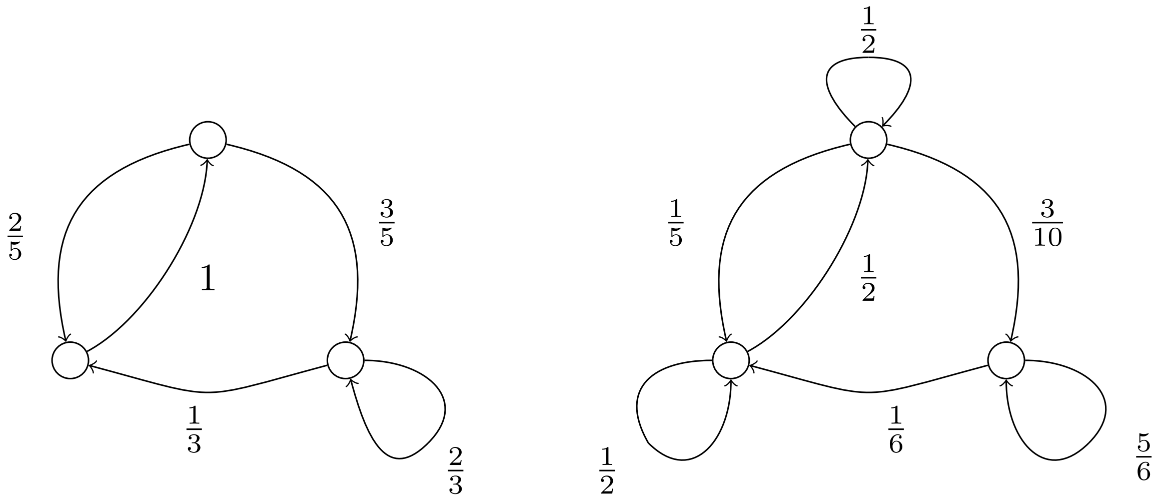


Figura 1.3: Representação em grafo de uma matriz de transição e a sua versão lazy.

Este exemplo acima é bastante útil, tendo em vista que existe um resultado à respeito de convergência (presente no final da seção) que exige que a cadeia em questão seja aperiódica, o que não é um problema se tomarmos por exemplo a versão lazy de uma cadeia.
Adicionalmente, é possível controlar o quão lazy uma cadeia é, isto é, podemos "brincar" com a probabilidade de a cadeia ficar parada, e utilizar valores diferentes de $1/2$.

Exemplo 1.1.45 (Versão α-lazy de uma cadeia)**.** Seja X uma cadeia gerada pela matriz de transição P. Chamamos a cadeia Y com matriz de transição $\alpha P + (1-\alpha)\,\mathrm{Id}$ de versão α-lazy de X.

Proposição 1.1.46. *Se $(X_n)_{n\in\mathbb{N}}$ é uma cadeia irredutível e aperiódica, com espaço de estados E e matriz de transição P, então existe $k_0 \in \mathbb{N}$ tal que para qualquer $k \geq k_0$ tem se $P^k(x,y) > 0$ para quaisquer $x, y \in E$.*

Dem.: Fixado um estado x, seja $X := \{n \in \mathbb{N}^* \;:\; P^n(x,x) > 0\}$.
Pela aperiodicidade da cadeia, sabemos que $\operatorname{mdc} X = 1$, e portanto, pelo Lema de Schur existe $r_x \in \mathbb{N}$ tal que para qualquer $n \geq r_x$ tem-se $P^n(x,x) > 0$.
Fixado um estado y, pela recorrência da cadeia, existe $s \in \mathbb{N}$ tal que $P^s(x,y) > 0$. Logo, para qualquer $n > s + r_x$, temos

$$P^n(x,y) \geq P^{n-s}(x,x)P^s(x,y) > 0.$$

Ou seja, mostramos que para qualquer par (a,b) existe um $t_{a,b} \in \mathbb{N}$ tal que $P^t(a,b) > 0$ sempre que $t > t_{a,b}$. Particularmente, tomando $k_0 := \max\{t_{a,b} \;:\; (a,b) \in E \times E\}$, segue o resultado. ■

Definição 1.1.47 (Recorrência e transiência)**.** Fixada uma cadeia de Markov $(X_n)_{n\in\mathbb{N}}$, e um estado i, dizemos que um estado i é recorrente se

$$\mathbb{P}_i\left(\{k \in \mathbb{N} : X_k = i\} \text{ é não-limitado}\right) = 1.$$

Em contrapartida, dizemos que i é transiente se

$$\mathbb{P}_i\left(\{k \in \mathbb{N} : X_k = i\} \text{ é não-limitado}\right) = 0.$$

Lema 1.1.48. *para qualquer $k \in \mathbb{N}$ a seguinte igualdade é verdadeira*

$$\mathbb{P}_x(V_x > k) = \mathbb{P}_x(T_x < \infty)^k$$

onde $V_x = \sum_{n=0}^{\infty} \mathbf{1}_{\{X_n = x\}}$.

Dem.: Provaremos o resultado utilizando o princípio de indução matemática. Para $k = 0$, o resultado é imediato.
Suponha que a igualdade seja válida para algum $k \geq 0$. Assim:

$$\begin{aligned}
\mathbb{P}_x(V_x > k+1) &= \mathbb{P}_x(T_x^{k+1} < \infty) \\
&= \mathbb{P}_x(T_x^k < \infty,\, T_x^{k+1} - T_x^k < \infty) \\
&= \sum_{s=1}^{\infty} \mathbb{P}_x(T_x^k < \infty,\, T_x^{k+1} - T_x^k = s) \\
&= \sum_{s=1}^{\infty} \mathbb{P}_x(T_x^k < \infty,\, X_{T_x^k+1} \neq x,\, \ldots,\, X_{T_x^k+s-1} \neq x,\, X_{T_x^k+s} = x).
\end{aligned}$$

Utilizando a Propriedade Forte de Markov obtemos

$$\begin{aligned}
\mathbb{P}_x(V_x > k+1) &= \sum_{s=1}^{\infty} \left[\mathbb{P}_x(X_1 \neq x,\, \ldots,\, X_{s-1} \neq x,\, X_s = x)\mathbb{P}_x(T_x^k < \infty)\right] \\
&= \sum_{s=1}^{\infty} \left[\mathbb{P}_x(T_x = s)\mathbb{P}_x(T_x^k < \infty)\right] \\
&= \mathbb{P}_x(T_x < s)\mathbb{P}_x(T_x^k < \infty).
\end{aligned}$$

Finalmente, pela hipótese de indução, concluímos que

$$\mathbb{P}_x(V_x > k+1) = \mathbb{P}_x(T_x < \infty)^{k+1}.$$

Portando, segue do princípio de indução matemática que

$$\mathbb{P}_x(V_x > k) = \mathbb{P}_x(T_x < \infty)^k$$

para todo estado x e qualquer $k \in \mathbb{N}$. ■

Proposição 1.1.49. *Fixada uma cadeia de Markov $(X_n)_{n\in\mathbb{N}}$ com matriz de transição P, e um estado x, as seguintes afirmações são verdadeiras:*

(i) *Se $\mathbb{P}_x(T_x < \infty) = 1$, então x é recorrente e $\sum_{k=0}^{\infty} P^k(x,x) = \infty$;*

(ii) *Se $\mathbb{P}_x(T_x < \infty) < 1$, então x é transiente e $\sum_{k=0}^{\infty} P^k(x,x) < \infty$;*

(iii) *Fixada uma classe comunicadora, seus estados são todos recorrentes ou são todos transientes apenas. Particularmente, se a cadeia for irredutível, os estados são todos transientes ou todos recorrentes apenas.*

Dem.: (i) Note que

$$\mathbb{P}_x(V_x = \infty) = \lim_{r\to\infty} \mathbb{P}_x(V_x > r) = \lim_{r\to\infty} \mathbb{P}_x(T_x < \infty)^r = 1,$$

onde a segunda igualdade segue de Lema 1.1.48. Portanto, pela definição i é recorrente. Agora, temos

$$\sum_{n=0}^{\infty} P^n(x,x) = \sum_{n=0}^{\infty} \mathbb{E}_x\left[\mathbf{1}_{\{X_n=x\}}\right] = \mathbb{E}_x\left[\sum_{n=0}^{\infty} \mathbf{1}_{\{X_n=x\}}\right] = \mathbb{E}_x\left[V_x\right] = \infty,$$

onde a segunda igualdade segue do Teorema de Fubini.

(ii) Observe que

$$\sum_{n=0}^{\infty} P^n(x,x) = \mathbb{E}_x\left[V_x\right] = \sum_{n=0}^{\infty} \mathbb{P}_x(V_x > r) = \sum \mathbb{P}_x(T_x)^r$$

Assim, por Lema 1.1.48 temos

$$\sum_{n=0}^{\infty} P^n(x,x) = \mathbb{P}_x(T_x < \infty)^n = \frac{1}{1 - \mathbb{P}_x(T_x < \infty)} < \infty.$$

Logo, $P_x(V_x = \infty) = 0$, e segue da definição que x é transiente.

(iii) Observe que cada um dos estados é recorrente ou transiente. Fixada uma mesma classe comunicadora C, e $x, y \in C$, segue do fato da cadeia ser irredutível que existem $m, n \in \mathbb{N}^*$ de modo que $P^m(x,y)P^n(y,x) > 0$. Pelas equações de Chapman-Kolmogorov, temos

$$\sum_{k=0}^{\infty} P^k(x,x) \geq \sum_{k=0}^{\infty} P^m(x,y)P^k(y,y)P^n(y,x).$$

Repetindo o argumento, obtemos

$$\sum_{k=0}^{\infty} P^k(y,y) \geq \sum_{k=0}^{\infty} P^r(y,x)P^k(x,x)P^t(x,y)$$

onde $P^r(y,x)P^t(x,y) > 0$. Assim $\sum_{k=0}^{\infty} P^k(y,y)$ é finita se e somente se $\sum_{k=0}^{\infty} P^k(x,x)$ o é. Portanto, em vista dos itens anteriores concluímos a demonstração deste item. ■

Observação 1.1.50. Classicamente, o resultado acima é utilizado para mostrar que o passeio aleatório em $\mathbb{Z}^d$ é recorrente somente para os casos $d = 1, 2$, e transiente para os demais. Não demonstraremos este fato aqui, mas ele é feito em [Nor98], [Dur19] e [GW14].

Proposição 1.1.51. *Seja X uma cadeia irredutível com matriz de transição P e espaço de estados E. A cadeia é recorrente se e somente se tem-se $\mathbb{P}(T_x < \infty) = 1$ para qualquer estado x.*

Dem.: Suponha que todos os estados sejam recorrentes. Fixados estados x, y quaisquer mostraremos que $\mathbb{P}_y(T_x < \infty) = 1$.

Pela irredutibilidade da cadeia, existe $m \in \mathbb{N}^*$ tal que $P^m(y,x) > 0$. Portanto, como a cadeia é recorrente, temos

$$\begin{aligned} 1 &= \mathbb{P}_x(X_n = x \text{ para algum } n > m) \\ &= \sum_{w \in E} \mathbb{P}_x(X_n = x \text{ para algum } n > m \mid X_m = w)\mathbb{P}_x(X_m = w), \end{aligned}$$

segue da propriedade de Markov que

$$\begin{aligned} 1 &= \sum_{w \in E} \mathbb{P}_w(T_x < \infty)P^m(x,w) \\ &\leq [1 - P^m(x,y)] + \mathbb{P}_y(T_x < \infty)P^m(x,y) \end{aligned}$$

De onde segue que $\mathbb{P}_y(T_x < \infty) = 1$.

Agora, observe que pela propriedade de Markov temos

$$\mathbb{P}(T_x < \infty) = \sum_{y \in E} \mathbb{P}(X_0 = y)\mathbb{P}_y(T_x < \infty)$$

usando o fato de $\mathbb{P}_y(T_x < \infty) = 1$ para qualquer y obtemos

$$\mathbb{P}(T_x < \infty) = \sum_{y \in E} \mathbb{P}(X_0 = y) = 1.$$

Reciprocamente, suponha que $\mathbb{P}(T_x < \infty) = 1$ para qualquer estado x. Tome um estado $z \in E$ de modo que tenha-se $\mathbb{P}(X_0 = z) > 0$.
Pela propriedade forte de Markov, temos

$$1 = \mathbb{P}(T_z < \infty) = \sum_{y \in E} \mathbb{P}(X_0 = y)\mathbb{P}_y(T_z < \infty) \leq 1 - \mathbb{P}(X_0 = z) + \mathbb{P}(X_0 = z)\mathbb{P}_z(T_z < \infty),$$

o que nos leva a concluir que $\mathbb{P}_z(T_z < \infty) = 1$, e consequentemente z é recorrente. Como a cadeia é irredutível, concluímos que todos os estados são recorrentes. ■

Fixado um estado x, y, usaremos V_y^x para denotar o número esperado de visitas ao estado y entre duas visitas consecutivas ao estado x. Isto é:

$$V_y^x = \mathbb{E}_x\left[\sum_{n=0}^{T_x} \mathbf{1}_{\{X_n = y\}}\right],$$

onde T_x denota o tempo de primeira visita.

Teorema 1.1.52. *Seja X uma cadeia de Markov irredutível e recorrente com matriz de transição P. As seguintes afirmações são verdadeiras para qualquer x no espaço de estados E.*

(i) $V_x^x = 1$;

(ii) $V^x P = V^x$, *onde* $V^x = \left(V_y^x \,;\, y \in E\right)$;

(iii) $V_y^x \in (0,\infty)$ *para qualquer* $y \in E$.

(iv) *Se* λ *é uma medida tal que* $\lambda P = \lambda$, *então* $\lambda = kV^x$ *para alguma constante* $k \in \mathbb{R}$.

Dem.: (i) Imediata.

(ii) Fixado $n \in \mathbb{N}^*$, note que

$$\mathbb{P}_x\left(X_{n-1} = x\,,\, X_n = y\,,\, n \leq T_x\right) = \mathbb{P}_x\left(X_n = y \,|\, X_{n-1} = x\,,\, n \leq T_x\right)\mathbb{P}_x\left(X_{n-1} = x\,,\, n \leq T_x\right),$$

e pela propriedade de Markov, temos

$$\mathbb{P}_x\left(X_{n-1} = x\,,\, X_n = y\,,\, n \leq T_x\right) = \mathbb{P}_x\left(X_{n-1} = x\,,\, n \leq T_x\right) P(x,y) \tag{1.1}$$

Como a cadeia é recorrente e irredutível, pela Proposição 1.1.51 temos $\mathbb{P}(T_x < \infty) = 1$. Portanto,

$$V_y^x = \mathbb{E}_x\left[\sum_{n=0}^{T_k} \mathbf{1}_{X_n = y}\right] = \mathbb{E}_x\left[\sum_{n=0}^{\infty} \mathbf{1}_{\{X_n = y\}\cap\{n \leq T_x\}}\right]$$

que pelo teorema de Fubini resulta em

$$\begin{aligned} V_y^x &= \sum_{n=1}^{\infty} \mathbb{P}_x\left(X_n = y\,,\, n \leq T_x\right) \\ &= \sum_{z\in E}\sum_{n=1}^{\infty} \mathbb{P}_x\left(X_{n-1} = z\,,\, X_n = y\,,\, n \leq T_x\right) \end{aligned}$$

utilizando (1.1) obtemos

$$\begin{aligned} V_y^x &= \sum_{z\in E}\sum_{n=1}^{\infty} \mathbb{P}_x\left(X_{n-1} = z\,,\, n \leq T_x\right) P(z,y) \\ &= \sum_{z\in E} P(z,y) \sum_{n=0}^{\infty} \mathbb{P}_x\left(X_n = z\,,\, n \leq T_x - 1\right) \\ &= \sum_{z\in E} P(z,y)\mathbb{E}_x\left[\sum_{n=0}^{\infty} \mathbf{1}_{\{X_n = z\}\cap\{n \leq T_x - 1\}}\right] \\ &= \sum_{z\in E} P(z,y)\mathbb{E}_x\left[\sum_{n=0}^{T_x - 1} \mathbf{1}_{\{X_n = z\}}\right] \\ &= \sum_{z\in E} P(z,y) V_z^x = (V^x P)(y). \end{aligned}$$

Portanto, concluímos que $V^x = V^x P$.

(iii) Fixado um estado y, sabemos que existem $k, \overline{k} \in \mathbb{N}^*$ tais que $P^k(x,y)$ e $P^{\overline{k}}(y,x)$ são ambos positivos.

Note que através da conclusão do item anterior é fácil de ver que $V^x P^n = V^x$ para qualquer $n \in \mathbb{N}$. Portanto, temos

$$V_y^x = (V^x P^k)(y) \geq V_x^x P^k(x,y) > 0.$$

Por outro lado, utilizando o mesmo argumento, obtemos

$$V_x^x = \left(V^x P^{\bar{k}}\right)(x) \geq V_y^x P^{\bar{k}}(y,x) > 0.$$

Portanto,

$$\left[P^{\bar{k}}(y,x)\right]^{-1} \geq V_y^x \geq P^k(x,y) > 0,$$

o que conclui a demonstração do item.

(iv) Primeiro mostraremos que $\lambda \geq \lambda(x)V^x$, e em seguida obteremos a igualdade. Pela Proposição 1.1.42, sabemos que $\lambda(x) > 0$ assim, observe que

$$\begin{aligned}\lambda(y) = \sum_{x_1 \in E} \lambda(x_1)P(x_1,y) &= \sum_{x_1 \in E\ \setminus\{x\}} \lambda(x_1)P(x_1,y) + \lambda(x)P(x,y)\\ &= \sum_{x_1 \in E\ \setminus\{x\}} \lambda(x_1)P(x_1,y) + \lambda(x)\mathbb{P}_x(X_1 = y\,,\, T_x \leq 1).\end{aligned}$$

Repetindo o processo para a primeira soma, obtemos:

$$\begin{aligned}\sum_{x_1 \in E} \lambda(x_1)P(x_1,y) &= \sum_{x_1 \in E}\sum_{x_2 \in E\setminus\{x\}} \lambda(x_2)P(x_2,x_1)P(x_1,y)\\ &= \sum_{x_1,x_2 \in E\setminus\{x\}} \lambda(x_2)P(x_2,x_1)P(x_1,y) + \lambda(x)\sum_{x_2 \in E\setminus\{x\}} \lambda(x)P(x,x_2)P(x_2,y)\\ &= \sum_{x_1 \in E}\sum_{x_2 \in E\setminus\{x\}} \lambda(x_2)P(x_2,x_1)P(x_1,y) + \mathbb{P}_x(X_2 = y\,,\, T_x \geq 2)\end{aligned}$$

Assim, indutivamente

$$\lambda(y) = \sum_{a_1,\ldots,a_n \in E\{x\}} \lambda(a_n)P(a_n,a_{n-1})\cdots P(a_k,a_{k-1})P(a_1,y) + \lambda(x)\sum_{k=1}^{\infty}\mathbb{P}_x(X_k = y\,,\, T_x \geq k)$$

e consequentemente,

$$\lambda(y) \geq \lambda(x)\sum_{k=1}^{\infty}\mathbb{P}_x(X_k = y\,,\, T_x \geq k).$$

Agora, Observe que

$$\begin{aligned}\sum_{k=1}^{\infty}\mathbb{P}_x(X_k = y\,,\, T_x \geq k) &= \sum_{k=1}^{\infty}\mathbb{E}_x\left[\mathbf{1}_{\{X_k=y\}}\mathbf{1}_{\{T_x\geq k\}}\right]\\ &= \mathbb{E}_x\left[\sum_{k=1}^{\infty}\mathbf{1}_{\{X_k=y\}}\mathbf{1}_{\{T_x\geq k\}}\right]\end{aligned}$$

$$= \mathbb{E}_x \left[\sum_{k=1}^{T_k - 1} \mathbf{1}_{\{X_k = y\}} \right] = V^x(y).$$

Portanto, de fato temos $\lambda \geq \lambda(x)V^x$. Agora, provaremos a igualdade. Fixado um estado y qualquer, pela irredutibilidade existe $m \in \mathbb{N}^*$ de modo que $P^m(x,y) > 0$. Observe que $\lambda - \lambda(x)V^x$ é uma medida invariante. Pelas equações de Chapman-Kolmogorov, temos

$$0 = \lambda(x) - \lambda(x)V^x(x) = \sum_{w \in E} [(\lambda(w) - \lambda(x)V^x(w))P^m(w,x)] \geq (\lambda(y) - \lambda(x)V^x(y))P^m(y,x).$$

O que revela que $\lambda(y) - \lambda(x)V^x(y)$. Como tomamos $y \in E$ qualquer, segue que $\lambda = \lambda(x)V^x$. ■

Observação 1.1.53. Nas hipóteses do teorema acima, note que se a cadeia X possui uma distribuição invariante π, então sabemos que fixado y, $V^y = k\pi$ para algum $k > 0$. Por consequência

$$\mathbb{E}_y [T_y] = \sum_{x \in E} V^y(x) = k,$$

e portanto

$$\pi = \mathbb{E}_y [T_y]^{-1} V^y.$$

Como isso vale para qualquer $y \in E$, segue que

$$\pi(x) = \frac{1}{\mathbb{E}_x [T_x]}$$

para todo $x \in E$.
Além disso, para quaisquer $x, y \in E$ vale

$$V^y(x) = \frac{\mathbb{E}_y [T_y]}{\mathbb{E}_x [T_x]}.$$

O comentário acima unido com a Proposição 1.1.35 garantem o próximo resultado.

Corolário 1.1.54. *Se X é uma cadeia de Markov irredutível e recorrente, então X possui no máximo uma distribuição invariante. Particularmente, se o espaço for finito, X possui exatamente uma única distribuição invariante.*

O próximo exemplo mostra que a hipótese de que o espaço seja finito não pode ser retirada da Proposição 1.1.35.

Exemplo 1.1.55 (Passeio aleatório em $\mathbb{Z}$)**.** Considere o passeio aleatório simples em $\mathbb{Z}$ (Exemplo 1.1.4). É facilmente verificável que a medida π dada por

$$\pi(x) = 1 \quad \text{para qualquer } x \in \mathbb{Z}$$

satisfaz as equações de equilíbrio com a cadeia.
Se Proposição 1.1.35 fosse válido para espaço de estados não-finito então a cadeia assumiria uma distribuição invariante μ. Assim, pelo item (iii) Teorema 1.1.52 deveria existir constante k de modo que $k\pi = \mu$. Porém, como π não é uma medida finita não existe constante que torne a igualdade verdadeira.
Logo, concluímos que o passeio aleatório simples em $\mathbb{Z}$ não possui distribuição invariante.

Observação 1.1.56. Note que o exemplo acima fornece uma estratégia para mostrar que uma determinada cadeia não possui distribuição invariante. É suficiente obter uma medida invariante não-finita.

Proposição 1.1.57. *Seja X uma cadeia de Markov em tempo discreto irredutível, com espaço de estados E. Se a cadeia possui uma medida invariante λ* ou *existe algum estado $x \in E$ de modo que $\mathbb{E}_x[T_x] < \infty$, então $\mathbb{E}_y[T_y] < \infty$ para todo $y \in E$, e a distribuição π dada por*

$$\pi(y) = \frac{1}{\mathbb{E}_y[T_y]}$$

é invariante para a cadeia.

Dem.: Se $\mathbb{E}_x[T_x] < \infty$, então segue da Proposição 1.1.49 que a cadeia é recorrente. Portanto, pelo Teorema 1.1.52 segue que $V^x P = V^x$.
Note que

$$\infty > \mathbb{E}_x[T_x] \geq \mathbb{E}_x\left[\sum_{y\in E}\sum_{n=0}^{T_k-1}\mathbf{1}_{\{X_n=y\}}\right] = \sum_{y\in E}\mathbb{E}_x\left[\sum_{n=0}^{T_k-1}\mathbf{1}_{\{X_n=y\}}\right] = \sum_{y\in E}V^x(y),$$

onde a primeira igualdade segue do Teorema de Fubini (Teorema A.1.4). Portanto é imediato que $V^x\left(\mathbb{E}_x[T_x]\right)^{-1}$ é uma distribuição invariante para a cadeia.

Agora, se a cadeia possui uma distribuição invariante λ, então o resultado segue da Observação 1.1.53. ■

Finalizaremos o capítulo obtendo resultados à respeito da convergência de cadeias de Markov. Estes resultados serão fundamentais para o estudo de cadeias de Markov em tempo contínuo. Inicialmente obteremos um resultado mais forte válido apenas para cadeias com espaço de estados finito.

Teorema 1.1.58 (Teorema da convergência (Espaço finito))**.** *Seja $(X_n)_{n\in\mathbb{N}}$ uma cadeia de markov com espaço de estados E finito, matriz de transição P. Se a cadeia é irredutível e aperiódica, então existem constantes $\alpha \in (0,1)$ e $C > 0$ tais que*

$$\max_{x\in E}\left\{\mathrm{d_{TV}}\left(P^t(x,\cdot),\pi\right)\right\} \leq C\alpha^t$$

onde π é a (única) distribuição invariante da cadeia.

Dem.: Pelo fato da cadeia ser irredutível e aperiódica, pela Proposição 1.1.46 existe $r \in \mathbb{N}$ tal que $P^r(x,y) > 0$ para quaisquer $x, y \in E$.
Defina a matriz Π como sendo a matriz $|E| \times |E|$ com todas as linhas iguais à distribuição π. Uma consequência dessa definição é o fato de que $\Pi P^k = \Pi$ para qualquer $k \in \mathbb{N}$.

Tomando $\delta \leq \min\{P^r(x,y) \;:\; x,y \in E\}$, temos $P^r(x,y) \geq \delta\pi(y)$ para quaisquer $x, y \in E$. Assim, denotando $1-\delta$ por θ, note que a equação

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

define a matriz estocástica Q. Além disso, como $P^r\Pi = \Pi$, segue que $Q\Pi = \Pi$, e consequentemente, $Q^k\Pi = \Pi$ para qualquer $k \in \mathbb{N}$.

Observe que se $P^{rn} = (1-\theta^n)\Pi + \theta^n Q^n$, então

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

Como $P^r = (1-\theta)\Pi + \theta Q$, segue do princípio de indução matemática que

$$P^{rn} = (1-\theta^n)\Pi + \theta^n Q^n,$$

para qualquer $n \in \mathbb{N}^*$. Rearranjando os termos, e multiplicando ambos os membros por P^j $(0 \leq j < k)$, obtemos

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

Observe que para qualquer $y \in E$, temos

$$\sum_{x\in E} \left|P^{rk+j}(y,x) - \pi(x)\right| = \sum_{x\in E} \left|P^{rk+j}(y,x) - \Pi(y,x)\right| = 2\theta^k \sum_{x\in E} \left|(Q^k P^j)(y,x) - \Pi(y,x)\right| \leq 2\theta^k.$$

Portanto, concluímos que

$$\max_{y\in E} \left\{ \mathrm{d}_{TV}\left(P^{rk+j}(y,\cdot),\pi\right)\right\} \leq \theta^k \leq \frac{1}{\theta}(\theta^{1/r})^{rk+j}.$$

Logo, definindo $\alpha := \theta^{1/r}$ e $C := \theta^{-1}$ obtemos o resultado desejado. ■

Abaixo, enunciaremos um resultado a respeito de convergência para cadeias com espaço de estados não necessariamente finito, porém com uma conclusão ligeiramente mais fraca.

Teorema 1.1.59 (Teorema da convergência (caso geral))**.** *Seja X é uma cadeia com matriz de transição P, irredutível, aperiódica, com espaço de estados E e distribuição inicial λ. Se π é uma distribuição invariante para P, então*

$$\lim_{n\to\infty} \mathbb{P}(X_n = x) = \pi(x)$$

para todo $x \in E$. Particularmente,

$$\lim_{n\to\infty} P^n(y,x) = \pi(x)$$

para quaisquer $x, y \in E$.

Dem.: Teorema 1.8.3 de [Nor98]. ■

2

Cadeias de Markov em tempo contínuo

Neste capítulo, introduziremos alguns conceitos técnicos essenciais, como filtrações, processos adaptados e tempos de parada, que serão fundamentais para a definição formal de cadeias de Markov. Na segunda seção, aproveitaremos a construção de cadeias de Markov em tempo discreto, elaborada no Capítulo 1, para construir uma família particular de cadeias de Markov em tempo contínuo. Na subsequente seção, derivaremos propriedades fundamentais para as cadeias pertencentes a essa família, e ao final, apresentaremos uma versão para tempo contínuo do teorema da convergência, além de discutir detalhes relacionados à recorrência e unicidade de distribuições invariantes para as cadeias dessa família.
A seção inicial é fundamentada em [Dur19], enquanto o restante do capítulo é guiado pelo conteúdo de [Nor98].

2.1 Cadeias de Markov em tempo contínuo

Nesta seção, definiremos os objetos básicos, e posteriormente definiremos formalmente o conceito de cadeias de Markov em tempo contínuo. Para tal usaremos conceitos e resultados oriundos da teoria da medida.

Definição 2.1.1 (Filtração)**.** Dado um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$, dizemos que uma sequência $(\mathcal{F}_i)_{i\in I}$ de sigma álgebras (com $I = \mathbb{R}_{\geq 0}$ ou $I = \mathbb{N}$) é uma Filtração se as seguintes condições são verificadas:

(i) $\mathcal{F}_i$ é uma σ-álgebra contida em $\mathcal{F}$ para todo $i \in I$;

(ii) $\mathcal{F}_i \subseteq \mathcal{F}_j$ sempre que $i, j \in I$ e $i \leq j$.

Observação 2.1.2. Um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ munido de uma filtração $(\mathcal{F}_i)_{i\in I}$ é chamado de espaço de probabilidade filtrado.

Definição 2.1.3 (Processo estocástico em tempo contínuo)**.** Dado um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$, um processo estocástico em tempo contínuo é uma família de variáveis aleatórias $X = (X_t;\, t \in \mathbb{R}_{\geq 0})$ tomando valores em um mesmo espaço mensurável $(\Omega', \mathcal{F}')$.

Note que a definição acima é bastante geral, e agora estando no "mundo contínuo" problemas antes não observados no caso discreto começam a surgir. O próximo exemplo ilustra um deles.

Exemplo 2.1.4. Se (Y_n) é uma cadeia em tempo discreto, dados eventos $A_1, A_2, \ldots$ disjuntos, uma conta bastante geral que frequentemente fazemos é

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} \{Y_n \in A_n\}\right),$$

que pelos resultados de teoria da medida sabemos ser equivalente a

$$\sum_{n\in\mathbb{N}} \mathbb{P}(Y_n \in A_n).$$

Porém, isso não se traduz quando estamos no mundo contínuo, visto que o resultado acima não faz sentido quando se tem uma quantidade não-enumerável de eventos (que a princípio é válida).

Com isso em mente, definiremos um tipo de processo que evita o problema acima.

Definição 2.1.5 (Processo contínuo à direita (Espaço de estados enumerável))**.** Seja $(X_t)_{t\geq 0}$ um processo estocástico em tempo contínuo com espaço de estados E enumerável. Dizemos que $(X_t)_{t\geq 0}$ é um processo contínuo à direita se para qualquer $\omega \in \Omega$ e $t \geq 0$ existe $\varepsilon > 0$ tal que

$$X_s(\omega) = X_t(\omega) \qquad \text{para } t \leq s \leq t + \varepsilon.$$

De fato, isso evita o problema descrito no exemplo, pois neste caso o processo é completamente determinado pelas probabilidades $\mathbb{P}(X_{t_0} = a_0, X_{t_1} = a_1, \ldots, X_{t_n} = a_n)$ para $n \in \mathbb{N}, t_n \geq \ldots \geq t_0 \geq 0$ e $a_0, \ldots, a_n \in I$ (Ver Observação A.1.13). Por conta disso, doravante trataremos apenas do caso em que o espaço de estados é enumerável.

Outro problema que surge, é a possibilidade de se ter uma quantidade infinita de trocas de estado em um período finito de tempo. Encontraremos condições para evitar este problema para uma classe de cadeias em tempo contínuo na próxima seção.

Observação 2.1.6. Dizemos que o processo $X = (X_t; t \in I)$ é adaptado à filtração $(\mathcal{F}_i)_{i\in I}$ se para cada $i \in I$ a variável aleatória X_i é $(\mathcal{F}_i, \mathcal{F}')$-mensurável.

No contexto de filtrações, uma delas se destaca das demais:

Exemplo 2.1.7 (Filtração natural)**.** Fixado um processo estocástico $(X_t)_{t\geq 0}$ chamamos a filtração $(\mathcal{F}_s)_{s\geq 0}$ definida por

$$\mathcal{F}_s := \sigma\left(X_t \;:\; t \in [0, s] \cap I\right)$$

para todo $s \geq 0$, de filtração natural, onde $\sigma(Y)$ denota a sigma álgebra gerada por Y.

Note que o nome realmente faz jus ao objeto, visto que em cada instante de tempo se tem a menor sigma-álgebra que "compreende" o processo.

Exemplo 2.1.8. Qualquer processo estocástico é adaptado à filtração natural.

Exemplo 2.1.9 (Processo não adaptado à uma dada filtração)**.** Considere o espaço de probabilidade $([0,1)\,, \mathcal{B}\,([0,1))\,, \mathbb{P})^{(1)}$, onde $\mathbb{P}$ é a medida de Lebesgue em $[0,1)$. Definindo o processo estocástico em tempo contínuo $(Y_t)_{t\geq 0}$ por

$$X_t(\omega) = \lfloor t\omega \rfloor$$

para cada $t \geq 0$ e $\omega \in \mathbb{R}$. O processo definido acima não é adaptado à filtração $(\mathcal{F}_t)_{t\geq 0}$ dada por

$$\mathcal{F}_0 = \{\emptyset, \mathbb{R}\} \quad \text{e} \quad \mathcal{F}_t = \sigma\left([n, n+1)\ ;\ n \in \mathbb{N},\ \right)$$

para qualquer $t \geq 0$. Porém é adaptado à filtração $(\mathcal{G}_t)_{t\geq 0}$ dada por

$$\mathcal{G}_0 = \{\emptyset, \mathbb{R}\} \quad \text{e} \quad \mathcal{G}_t = \sigma\left(\left[\frac{n}{s}, \frac{n+1}{s}\right)\ ;\ n \in \mathbb{N},\ s \in (0,t)\right),$$

para qualquer $t \geq 0$.
De fato, basta notar que

$$X_t^{-1}(m) = \left[\frac{m}{t}, \frac{m+1}{t}\right) \in \mathcal{G}_t$$

para qualquer $t \geq 0$ e $m \in \mathbb{N}$, e além disso, pela definição trivialmente se tem $\mathcal{G}_t \subseteq \mathcal{G}_s$ sempre que $s \geq t$.

Exemplo 2.1.10. Considere uma sequência de variáveis aleatórias i.i.d. $(\xi_k)_{k\in\mathbb{N}}$ com distribuições unif $(\{1,\ldots,n\})$, definidas em algum espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$.
O processo $(Y_t)_{t\geq 0}$ com espaço de estados $(\{1,\ldots,n\}\,, \mathcal{P}\,(\{1,\ldots,n\}))$ definido em $(\Omega, \mathcal{F}, \mathbb{P})$ para cada t por

$$X_t(\omega) = \xi_{\lfloor t \rfloor}(\omega)$$

é adaptado à filtração

$$\sigma(\xi_0, \xi_1, \ldots, \xi_{\lfloor t \rfloor}).$$

Buscando definir cadeias de Markov, definimos:

Definição 2.1.11 (Propriedade de Markov)**.** Seja $(X_t)_{t\geq 0}$ um processo estocástico definido no espaço $(\Omega, \mathcal{F}, \mathbb{P})$, assumindo valores em $(\Omega', \mathcal{F}')$ e adaptado à filtração $(\mathcal{F}_s)_{s\geq 0}$. Dizemos que $(X_t)_{t\geq 0}$ satisfaz a propriedade de Markov se para quaisquer $A \in \mathcal{F}', t > s \geq 0$ tem-se

$$\mathbb{P}(X_t \in A \mid \mathcal{F}_s) = \mathbb{P}(X_t \in A \mid X_s).$$

Observação 2.1.12. Note que em vista da definição acima, não faz sentido dizer que uma sequência é uma cadeia de Markov (ou não) se a filtração não for especificada.

$^{(1)}\mathcal{B}\,([0,1)) = \mathcal{B}(\mathbb{R}) \cap [0,1)$

Definição 2.1.13 (Cadeia de Markov em tempo contínuo/ Processo de Markov)**.** Dizemos que um processo estocástico em tempo contínuo é um *Processo de Markov* se este satisfaz a propriedade de Markov. Em particular, quando o espaço de estados é enumerável, dizemos que o processo é uma *Cadeia de Markov em Tempo Contínuo.*

Exemplo 2.1.14. O Exemplo 2.1.10 é uma cadeia de Markov. De fato, é facilmente verificável que

$$\mathbb{P}(Y_t \in A \mid \mathcal{F}_s) = \mathbb{P}(\xi_{\lfloor t \rfloor} \in A) = \mathbb{P}(Y_t \in A \mid Y_s),$$

visto que a sequência $(\xi_k)_{k\in\mathbb{N}}$ é i.i.d. .

Assim como no caso discreto, as vezes é interessante calcular probabilidades de certos eventos em tempos aleatórios ao invés de um determinado instante. Por conta disso, assim como no caso discreto, a propriedade forte de Markov é importante.

Definição 2.1.15 (Tempo de parada)**.** Seja $(X_t)_{t\geq 0}$ um processo estocástico definido no espaço $(\Omega, \mathcal{F}, \mathbb{P})$, assumindo valores em $(\Omega', \mathcal{F}')$ e adaptado à filtração $(\mathcal{F}_s)_{s\geq 0}$. Dizemos que uma variável aleatória τ é um tempo de parada para $(X_t)_{t\geq 0}$ se para qualquer $t \geq 0$ tem se

$$\{\tau \leq t\} \in \mathcal{F}_t.$$

Definição 2.1.16 (σ-álgebra em um tempo de parada)**.** Se T é um tempo de parada com respeito à filtração $(\mathcal{F}_s)_{s\geq 0}$, definimos a σ-álgebra $\mathcal{F}_T$, denominada σ-álgebra do Passado de T, por

$$\mathcal{F}_T := \{A \in \mathcal{F} \mid A \cap \{T \leq t\} \in \mathcal{F}_t \text{ para todo } t \geq 0\}.$$

Definição 2.1.17 (Propriedade Forte de Markov)**.** Seja $(X_t)_{t\geq 0}$ um processo estocástico definido no espaço $(\Omega, \mathcal{F}, \mathbb{P})$, assumindo valores em $(\Omega', \mathcal{F}')$ e adaptado à filtração $(\mathcal{F}_s)_{s\geq 0}$. Dizemos que o processo satisfaz a propriedade forte de Markov se para qualquer tempo de parada τ, qualquer $s \geq 0$ e qualquer $A \in \mathcal{F}'$ tem-se

$$\mathbb{P}(X_{T+s} \in A \mid \mathcal{F}_T) = \mathbb{P}(X_{T+s} \in A \mid X_T)$$

quase certamente no conjunto $\{T < \infty\}$.

Definição 2.1.18 (Núcleo de Transição)**.** Seja $(X_t)_{t\geq 0}$ uma cadeia de Markov em tempo contínuo no espaço $(\Omega, \mathcal{F}, \mathbb{P})$, assumindo valores em $(\Omega', \mathcal{F}')$. Dizemos que uma família $(P^{t+s}(A, \cdot, s)\,;\, t, s \geq 0 \text{ e } A \in \mathcal{F}')$ de medidas em $(\Omega, \mathcal{F}')$ é um núcleo de transição de $(X_t)_{t\geq 0}$ se para qualquer $A \in \mathcal{F}'$ e quaisquer $s, t \geq 0$, vale

$$P^{t+s}(A, \cdot\,, s)\mathbb{P}(X_s \in A) = \mathbb{P}(X_s \in A, X_{s+t} = \cdot\,).$$

Observação 2.1.19. Chamamos as medidas $P^{t+s}(A, \cdot, s)$ de *probabilidades de transição.*

Definição 2.1.20 (Cadeia Homogênea/Não-Homogênea)**.** Dizemos que uma cadeia de Markov em tempo contínuo é homogênea se a mesma admite um núcleo de transição tal que

$$P^{t+s}(A, \cdot, s) = P^{t+r}(A, \cdot, r)$$

para quaisquer $r, s \geq 0$. Nestes casos, denotamos as probabilidades de transição omitindo a primeira entrada temporal. Isto é: $P^{t+s}(A, \cdot, s) =: P^t(A, \cdot\,)$.

Exemplo 2.1.21. Uma cadeia tal que $(X_t(\omega))_{t\geq 0}$ é constante para qualquer ω é particularmente homogênea

Exemplo 2.1.22. A cadeia do Exemplo 2.1.10 é não-homogênea.

No próximo capítulo, nos restringiremos ao estudo de cadeias homogêneas e contínuas à direita, e obteremos uma família (não-trivial) de cadeias de Markov em tempo contínuo, assim como uma maneira construtiva de gerar cadeias em tempo contínuo.

2.2 Construção de Cadeias de Markov em tempo contínuo via matrizes geradoras

Começaremos definindo os objetos básicos para construir cadeias.

Definição 2.2.1 (Matriz geradora)**.** Seja I um conjunto enumerável. Chamamos uma matriz $Q = (q_{ij}\ ;\ i,j \in I)$ de matriz geradora se as seguintes condições são verificadas:

(i) $0 \leq -q_{ii} < \infty$;

(ii) $q_{ij} \geq 0$ para todo $i \neq j$;

(iii) $\sum\limits_{j\in I} q_{ij} = 0$ para todo i.

Note que as condições (i) e (ii) acima nos levam a concluir que $\sum_{i\neq j} q_{ij} < \infty$ mesmo quando I não é finito. Além disso, é notório que αQ também é uma matriz geradora para qualquer $\alpha \geq 0$. Esse conceito será importante no futuro.

Observação 2.2.2. Doravante usaremos $q(i)$ para denotar $-q_{ii}$ sempre que a matriz geradora for Q.

Teorema 2.2.3. *Seja Q uma matriz em um conjunto finito I. Definindo $P(t) = \mathrm{e}^{tQ}$ para $t \geq 0$, verificamos as seguintes propriedades:*

(i) *$P(s+t) = P(s)P(t)$ para quaisquer $s,t \geq 0$ (propriedade de semigrupo);*

(ii) *$(P(t)\,;\,t\geq 0)$ é a única solução para a equação conhecida como "Equação progressiva de Kolmogorov"* [(2)]

$$\frac{\mathrm{d}}{\mathrm{d}\,t}P(t) = P(t)Q, \qquad P(0) = Id\,;$$

(iii) *$(P(t)\,;\,t\geq 0)$ é a única solução para a equação conhecida como "Equação regressiva de Kolmogorov"* [(3)]

$$\frac{\mathrm{d}}{\mathrm{d}\,t}P(t) = QP(t), \qquad P(0) = Id\,;$$

(2) Na literatura essa propriedade é conhecida como "Kolmogorov Forward Equation"

(3) Na literatura essa propriedade é conhecida como "Kolmogorov Backward Equation"

(iv) *Para qualquer $k \in \mathbb{N}$ vale:*

$$\frac{\mathrm{d}^k}{\mathrm{d}\, t^k} P(\cdot)(0) = Q^k.$$

Dem.:

(i) Note que

$$\begin{aligned} P(s+t) = e^{(s+t)Q} &= \sum_{k=0}^{\infty} \frac{((s+t)Q)^k}{k!} \\ &= \sum_{k=0}^{\infty} \sum_{n=0}^{k} \frac{k!}{n!(n-k)!k!} (tQ)^n (sQ)^{k-n} \end{aligned}$$

Como a soma converge na norma $\|\cdot\|_\infty$, pelo Teorema de Fubini (Teorema A.1.4), temos:

$$\begin{aligned} P(s+t) &= \sum_{n=0}^{\infty} \sum_{k=n}^{\infty} \frac{(tQ)^n (sQ)^{k-n}}{n!(n-k)!} \\ &= \sum_{n=0}^{\infty} \frac{(tQ)^n}{n!} \sum_{k=n}^{\infty} \frac{(sQ)^{n-k}}{(n-k)!} \\ &= P(t)P(s) \end{aligned}$$

(ii) Por hipótese, sabemos que

$$P(t) = \sum_{k=0}^{\infty} \frac{(tQ)^k}{k!}.$$

Além disso pelo fato de o raio de convergência da função exponencial ser $\mathbb{R}$, através da derivação termo-a-termo obtemos

$$P'(t) = \sum_{k=1}^{\infty} \frac{t^{k-1} Q^k}{(k-1)!} = P(t)Q.$$

Nos resta agora apenas verificar que a solução é de fato única. Seja $M(t)$ uma solução do sistema. Temos

$$\frac{\mathrm{d}}{\mathrm{d}\, t} \left(M(t) e^{-tQ} \right) = M'(t) e^{t(-Q)} + M(t)(-Q) e^{-tQ}$$

assim, segue da hipótese que

$$\frac{\mathrm{d}}{\mathrm{d}\, t} \left(M(t) e^{-tQ} \right) = M(t) Q e^{-tQ} - M(t) Q e^{-tQ} = 0.$$

Portanto, $M(t)(-Q)e^{-tQ}$ é constante. Portanto, concluímos que de fato a solução é única.

(iii) Segue da mesma observação que o item anterior.

(iv) Nos itens anteriores mostramos que

$$P'(t) = P(t)Q.$$

Portanto, indutivamente obtemos

$$\frac{\mathrm{d}^k}{\mathrm{d}\, t^k} P(t) = P(t)Q^k,$$

particularmente, tomando $t = 0$ obtemos o resultado desejado.

■

Observação 2.2.4. A Equação Progressiva de Kolmogorov pode ser reescrita como

$$\frac{\mathrm{d}}{\mathrm{d}t} P^t(x, y) = \sum_{z \in E} P^t(x, z) q(z, y), \qquad P^0(x, y) = \delta_{x,y},$$

para quaisquer $x, y \in E$.
Chamamos as equações acima de Equações Progressivas de Kolmogorov. Em alguns textos encontramos apenas a formulação acima.
De maneira semelhante, definimos as Equações Regressivas de Kolmogorov como

$$\frac{\mathrm{d}}{\mathrm{d}t} P^t(x, y) = \sum_{z \in E} q(x, z) P^t(z, y), \qquad P^0(x, y) = \delta_{x,y},$$

para quaisquer $x, y \in I$ e $t \geq 0$.

Observação 2.2.5 (Equações de Chapman-Kolmogorov)**.** Uma consequência do primeiro item do teorema anterior é o fato de que para quaisquer $x, y \in I$ vale

$$P^{t+s}(x, y) = (P^t P^s)(x, y) = \sum_{w \in I} P^t(x, w) P^s(w, y).$$

Chamamos as equações acima de Equações de Chapman-Kolmogorov.
Particularmente, vale

$$P^{t+s}(x, y) \geq P^t(x, w) P^s(w, y)$$

para qualquer $w \in I$.

O próximo resultado mostra que através de uma matriz geradora podemos obter uma matriz estocástica de maneira relativamente simples.

Proposição 2.2.6. *Uma matriz Q em um conjunto finito I é uma matriz geradora se e somente se $P^t = \mathrm{e}^{tQ}$ é uma matriz estocástica para todo $t \geq 0$.*

Dem.: Pela definição de exponencial de uma matriz e utilizando a aproximação de exponenciais em $\mathbb{R}$, temos

$$P^t = \mathrm{Id} + tQ + O(t^2).$$

Portanto, para $x \neq y$ quaisquer tem-se $q(x,y) \geq 0$ se e somente se $P^t(x,y) \geq 0$ para t suficientemente pequeno.
Pela propriedade de semigrupo para qualquer $n \in \mathbb{N}^*$ tem-se $P^t = (P^{t/n})^n$. Assim, o argumento anterior pode ser reescrito como: Para $x \neq y$ quaisquer tem-se $q(x,y) \geq 0$ se e somente se $P^t(x,y) \geq 0$ para qualquer $t > 0$.

Suponha que Q é uma matriz geradora. Assim, fixado um estado x, para qualquer $n \in \mathbb{N}^*$, temos

$$\sum_{y \in E} q^n(x,y) = \sum_{y \in E} \sum_{z \in E} q^{n-1}(x,z) q(z,y) = \sum_{z \in E} q^{n-1}(x,z) \sum_{y \in E} q(z,y) = \sum_{z \in E} q^{n-1}(x,z) 0 = 0,$$

visto que cada linha de Q soma zero.
Portanto,

$$\sum_{y \in E} P^t(x,y) = \sum_{y \in E} \sum_{k=0} \frac{t^k Q^k(x,y)}{k!} = 1 + \sum_{k=1} \frac{t^k}{k!} \sum_{y \in E} q^k(x,y) = 1 + \sum_{k=1} \frac{t^k}{k!} 0 = 1,$$

o que mostra que P^t é uma matriz estocástica.

Reciprocamente, suponha que P^t é uma matriz estocástica para todo $t \geq 0$. Temos

$$\sum_{y \in E} q(x,y) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \left(\sum_{x \in E} P^t(x,y) \right) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} (1) = 0,$$

o que mostra que Q é de fato uma matriz geradora. ■

Fazendo uma pequena antecipação, este resultado e o anterior nos levam a concluir que cada matriz geradora fornece uma única matriz de transição (no sentido da propriedade de semigrupo), e ainda oferecem uma boa caracterização de matrizes geradoras. Embora o resultado inicialmente seja válido apenas para espaços finitos, obteremos um análogo válido para os casos em que I não é finito.

Definição 2.2.7 (Tempos de salto)**.** Dado um processo $(X_t)_{t \geq 0}$, definimos os tempos de salto $(J_n)_{n \in \mathbb{N}}$ de $(X_t)_{t \geq 0}$ por

$$J_k = \begin{cases} 0, & \text{para } k = 0; \\ \inf\left\{t \geq J_{k-1} \;;\; X_t \neq X_{J_{k-1}}\right\} & \text{para } k \geq 1. \end{cases}$$

Definição 2.2.8 (Tempos de espera)**.** Dado um processo $(X_t)_{t \geq 0}$, definimos os tempos de espera $(S_n)_{n \in \mathbb{N}_+}$ de $(X_t)_{t \geq 0}$ por

$$S_k = \begin{cases} J_k - J_{k-1}, & \text{se } J_{k-1} < \infty; \\ \infty & \text{caso contrário.} \end{cases}$$

Como indicado pela nomenclatura, os tempos de salto funcionarão como instantes em que a cadeia muda de estado, e os tempos de espera serão intervalos de tempo entre saltos consecutivos.

Agora listaremos algumas propriedades da distribuição exponencial que serão úteis para os próximos resultados. As demonstrações podem ser encontradas no Apêndice A.3 (Teorema A.3.4 e Teorema A.3.2).

Teorema 2.2.9. *Uma variável aleatória $T : \Omega \to (0, \infty]$ possui distribuição exponencial se e somente se possui a possui a propriedade da "Perda de Memória". Isto é:*

$$\mathbb{P}(T > s + t \,|\, T > s) = \mathbb{P}(T > t) \quad \text{para quaisquer } s, t \geq 0.$$

Teorema 2.2.10. *Sejam $S_1, S_2, \ldots$ uma sequência de variáveis aleatórias independentes com $S_n \sim \operatorname{Exp}(\lambda_n)$ com $\lambda_n \in (0, \infty)$ para todo $n \in \mathbb{N}_+$. As seguintes propriedades são válidas:*

(i) *Se* $\sum\limits_{n=1}^{\infty} \dfrac{1}{\lambda_n} < \infty$*, então* $\mathbb{P}\left(\sum\limits_{n=1}^{\infty} S_n < \infty\right) = 1.$

(ii) *Se* $\sum\limits_{n=1}^{\infty} \dfrac{1}{\lambda_n} = \infty$*, então* $\mathbb{P}\left(\sum\limits_{n=1}^{\infty} S_n = \infty\right) = 1.$

Definição 2.2.11 (Processo de saltos/Cadeia de saltos)**.** Dado um processo $(X_t)_{t \geq 0}$ em tempo contínuo, chamamos de processo de saltos de $(X_t)_{t \geq 0}$ o processo em tempo discreto $(Y_n)_{n \in \mathbb{N}}$ definido por $Y_n = X_{J_n}$. Se este processo for uma cadeia de Markov, então chamamos de cadeia de saltos.

Moralmente, a cadeia $(Y_n)_{n \in \mathbb{N}}$ é o histórico de sítios visitados pela cadeia $(X_t)_{t \geq 0}$.

Definição 2.2.12 (Matriz de salto)**.** Dada uma matriz geradora Q, definimos a matriz de salto $A = (a_{ij}\,;\, i, j \in I)$ de Q por

$$a_{ij} = \begin{cases} \dfrac{q_{ij}}{q(i)}, & \text{se } q(i) \neq 0 \text{ e } j \neq i; \\ 0, & \text{se } q(i) \neq 0 \text{ e } j = i; \\ 0, & \text{se } q(i) = 0 \text{ e } j \neq i; \\ 1, & \text{se } q(i) = 0 \text{ e } j = i. \end{cases}$$

Note que a Matriz de salto é uma matriz estocástica obtida por uma "espécie" de normalização, que consiste no seguinte processo em cada uma das linhas: Se a linha for nula, acrescente 1 na diagonal principal. Se a linha for não nula, remova o termo da diagonal principal e normalize os termos da linha.

Observação 2.2.13. Doravante, sempre que necessário assumiremos que as entradas da matriz de salto A são $(a_{ij}\,;\, i, j \in I)$.

Na seção anterior apontamos para o possível problema de observarmos uma quantidade infinita de saltos em um intervalo finito de tempo. Em breve, obteremos condições suficientes para que este problema não seja observado. Por hora, definiremos um tipo particular de processos que evitam este problema.

Definição 2.2.14 (Tempo de explosão/processo minimal)**.** Fixado um processo estocástico X em tempo contínuo, com tempos de espera $S_1, S_2, \ldots$ e tempos de salto $J_1, J_2, \ldots$ definimos o primeiro tempo de explosão ζ por

$$\zeta = \sup_n J_n = \sum_{n=1}^{\infty} S_n.$$

Ademais, se $\zeta = \infty$ com probabilidade 1, então dizemos que o processo é minimal.

Note que se $\zeta < \infty$, isso indica que a cadeia realizou uma quantidade infinita de saltos em um tempo finito, em outras palavras, a cadeia "explodiu".

Definição 2.2.15 (Cadeia de Markov em tempo contínuo gerada por Q)**.** Um processo $(X_t)_{t\geq 0}$ minimal e contínuo à direita assumindo valores em I é uma cadeia de Markov com distribuição inicial λ e matriz geradora Q se sua cadeia de saltos $(Y_n)_{n\in\mathbb{N}}$ é uma cadeia de Markov em tempo discreto com distribuição inicial λ e matriz de transição A, e para cada $n \in \mathbb{N}_+$, condicionado à $Y_0, ..., Y_{n-1}$, os seus tempos de espera $S_1, ..., S_n$ são variáveis aleatórias exponenciais independentes com parâmetros $q(Y_0), ..., q(Y_{n-1})$ respectivamente. Em outras palavras

$$X_t = \sum_{n=0} \mathbf{1}_{[J_n, J_{n+1})} Y_n.$$

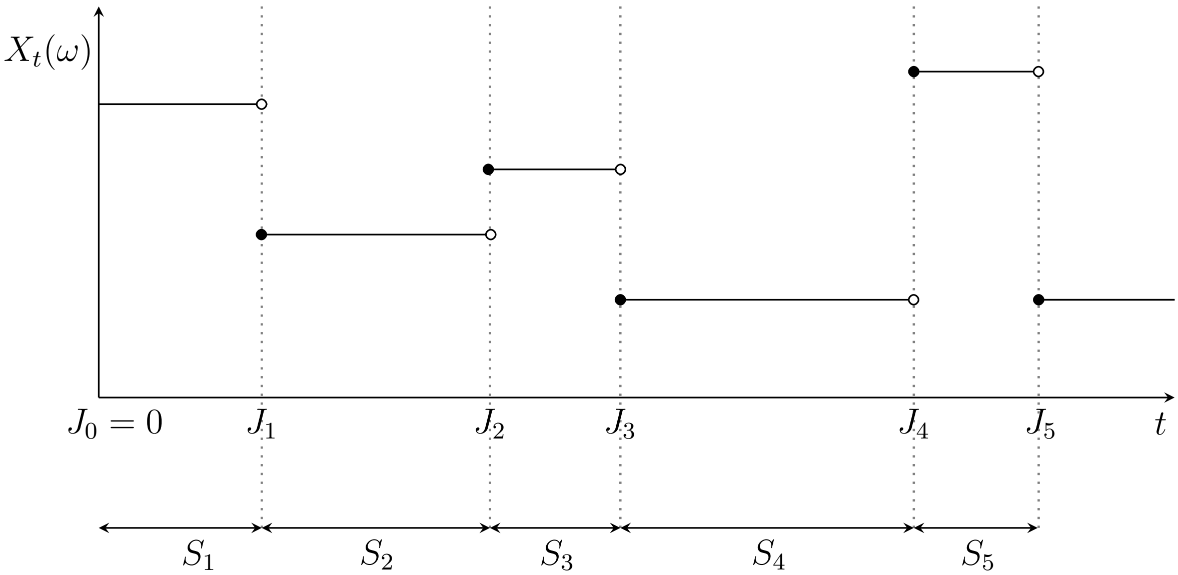


Figura 2.1: Exemplo gráfico de uma cadeia de Markov em tempo contínuo gerada por uma matriz Q.

Observação 2.2.16. Na definição acima ao invés de definir $S_1, S_2, \ldots$ como sendo exponenciais de parâmetros $q(Y_0), q(Y_1), \ldots$, poderíamos definir para cada $k \in \mathbb{N}^*$

$$S_k := \frac{\xi_k}{q(Y_{k-1})},$$

onde $\xi_1, \xi_2, \ldots$ são variáveis aleatórias exponenciais com parâmetro 1. Isso segue do fato de $p\,\mathrm{Exp}(p) \sim \mathrm{Exp}(1)$ para todo $p > 0$.
Essa caracterização será útil no Capítulo 4.

Observação 2.2.17. É possível perceber na definição acima que os números $q(i)$ agem como uma espécie de "taxa" de saída de um determinado estado i, e regulam a quantidade de tempo (em média) em que a cadeia permanece nesse estado antes de realizar uma troca.

Observação 2.2.18. Sabemos que se Q for uma matriz geradora, então para $\alpha > 1$ a matriz αQ também o é. Se X é a cadeia gerada por Q e Y é a cadeia gerada por αQ, então através da definição, é possível perceber que Y possui (em média) tempos de espera menores, e consequentemente (em média) transiciona mais rápido do que X por um fator de α. De maneira semelhante, para $\alpha < 1$, Y transiciona mais rápido do que X por um fator de α.
O argumento acima é formalizado no Exemplo 3.2.6.

Exemplo 2.2.19 (Construção de Cadeias de Markov em tempo contínuo)**.** Considere a sequência $(\xi_k)_{k\in\mathbb{N}}$ de variáveis aleatórias i.i.d. com distribuição comum $\xi_k \sim \mathrm{Exp}(1)$ definidas em um espaço $(\Omega', \mathcal{F}', \mathbb{P}')$. Fixada uma matriz geradora Q, considere também uma cadeia de Markov em tempo discreto $(Y_n)_{n\in\mathbb{N}}$ definida em um espaço $(\Omega, \mathcal{F}, \mathbb{P})$ com espaço de estados E, gerada pela matriz de salto A advinda de Q.

Definamos a sequência $(T_n)_{n\in\mathbb{N}}$ de variáveis aleatórias por

$$T_n = \sum_{k=1}^{n} \frac{\xi_k}{q(Y_k)},$$

para $n \in \mathbb{N}^*$ e $T_0 = 0$.
Assim, para cada $t \geq 0$ definimos a função $X_t : \Omega \times \Omega' \to E$ por

$$X_t(\alpha, \beta) = \begin{cases} Y_0(\alpha), & \text{se } t \in [T_0\,,\, T_1(\beta))\,; \\ Y_1(\alpha), & \text{se } t \in [T_1(\beta)\,,\, T_2(\beta))\,; \\ Y_2(\alpha), & \text{se } t \in [T_2(\beta)\,,\, T_3(\beta))\,; \\ & \vdots \end{cases}$$

Mostraremos que a função X_t é mensurável no espaço $(\Omega \times \Omega', \mathcal{F} \otimes \mathcal{F}')$ (e portanto uma variável aleatória).

Note que

$$\begin{aligned} X_t^{-1}(z) &= \{(\alpha, \beta) \in \Omega \times \Omega' \,:\, t \in [T_n(\beta)\,,\, T_{n+1}(\beta)) \text{ e } Y_n(\alpha) = z \text{ para algum } n \in \mathbb{N}\} \\ &= \bigcup_{n\in\mathbb{N}} \left(\left\{ Y_n^{-1}(z) \right\} \times \left\{ T_n^{-1}([0,t]) \cap T_{n+1}^{-1}((t,\infty)) \right\} \right). \end{aligned}$$

Como as variáveis aleatórias $Y_1, Y_2, \ldots$ e $T_0, T_1, T_2, \ldots$ são mensuráveis, segue que $X_t^{-1}(z)$ é união enumerável de conjuntos mensuráveis, logo é também mensurável. Portanto, segue que cada X_t é mensurável.
Note que pela construção do processo $(X_t)_{t\geq 0}$, o comportamento observado é o mesmo descrito na Definição 2.2.15, portanto fica demonstrada a existência deste tipo de processo.

Observação 2.2.20. Note que fixada uma matriz geradora, podemos construir várias cadeias com condições iniciais diferentes. Daremos ênfase às distribuições iniciais δ_x com $x \in E$.
Por abuso de notação usaremos $\mathbb{P}_x(X_t \in A)$ para denotar as probabilidades de uma "cópia" de $(X_t)_{t\geq 0}$ gerada por Q de maneira independente e com condição inicial δ_x.

Definido o processo, ainda é preciso demonstrar que de fato se trata de uma cadeia de Markov em tempo contínuo segundo a definição da seção anterior, ou seja, que o processo satisfaz a propriedade de Markov. Não faremos a demonstração neste capítulo, mas ela pode ser encontrada no apêndice.

Teorema 2.2.21. *Fixada uma matriz geradora Q, o processo em tempo contínuo da definição 2.2.15 satisfaz a propriedade forte de Markov. Isto é, fixado um tempo de parada T, para qualquer $s \geq 0$ tem-se*

$$\mathbb{P}(X_{T+s} = \xi \mid \mathcal{F}_T) = \mathbb{P}(X_{T+s} = \xi \mid X_T)$$

no conjunto $\{T < \infty\}$. Particularmente,

$$\mathbb{P}(X_{T+s} = \xi, X_T = \alpha, T < \infty) = \mathbb{P}(X_T = \alpha, T < \infty)\mathbb{P}_\alpha(X_s = \xi),$$

o que garante que

$$\mathbb{P}(X_{T+s} = \xi \mid X_T = \alpha, T < \infty) = \mathbb{P}_\alpha(X_s = \xi)$$

se $\mathbb{P}(X_T = \alpha) > 0$.

Dem.: Ver Apêndice B. ■

Note que se tomarmos $T := s$ para algum $s \in \mathbb{R}_{\geq 0}$, obtemos a propriedade de Markov.

Portanto, agora de fato conhecemos uma maneira de gerar cadeias de Markov em tempo contínuo, e mais do que isso, para este tipo de cadeia, podemos por hora não dar muita atenção para a σ- álgebra em questão.

O exemplo abaixo ilustra que o problema de se ter infinitos saltos em um período finito de tempo de fato acontece com algumas cadeias

Exemplo 2.2.22 (Cadeia de Markov não-minimal)**.** Considere a matriz geradora Q definida em $\mathbb{N}^*$ dada por

$$Q(x,y) = \begin{cases} -2^x, & \text{se } y = x \\ 2^x, & \text{se } y = x+1 \\ 0, & \text{se } y \notin \{x, x+1\} \end{cases}$$

A dinâmica da cadeia de Markov em tempo contínuo gerada por Q com estado inicial 0 consiste em cada número natural n esperar um tempo exponencial médio igual a 2^{-n} para transicionar ao seu sucessor. Se $S_1, S_2, \ldots$ são seus tempos de espera, então pelo Teorema 2.2.10, concluímos que

$$\mathbb{P}\left(\sum_{n=1}^{\infty} S_n < \infty\right) = 1.$$

Agora que vimos um exemplo de cadeia em tempo contínuo não-minimal, finalmente obteremos condições suficientes para não seja observada uma quantidade infinita de saltos em um período finito de tempo.

Teorema 2.2.23. *Seja $(X_t)_{t\geq 0}$ uma cadeia de Markov com distribuição inicial λ e matriz geradora Q. Se uma das condições abaixo for verdadeira, então a cadeia não explode.*

(i) $\sup_{i\in\Omega}\{q(i)\} < \infty$ *;*

(ii) Ω *é finito;*

(iii) $X_0 = i$*, e* i *é recorrente para a cadeia de salto.*

Dem.: Definamos as variáveis aleatórias T_n por

$$T_n := q(Y_{n-1})S_n,$$

onde as variáveis $(Y_k)_{k\in\mathbb{N}}$ e $(S_k)_{k\in\mathbb{N}}$ são respectivamente a cadeia de saltos, e os tempos de espera. Pela definição de cadeias de markov em tempo contínuo, as variáveis $T_1, T_2, \ldots$ são independentes, com média igual a 1 e independentes das variáveis aleatórias $Y_0, Y_1, \ldots$.

(i) Por hipótese, sabemos que $\sup_{i\in\Omega}\{q(i)\} < \infty$, portanto, vale

$$\sup_{i\in\Omega}\ \{q(i)\}\,\zeta\ \geq\ \sup_{i\in\Omega}\ \{q(i)\}\sum_{n=1}^{\infty} S_n\ \geq\ \sum_{n=1}^{\infty} q(Y_{n-1})S_n$$

Como

$$\sum_{n=1}\mathbb{E}\left[T_n\right] = \infty,$$

segue do Teorema 2.2.10 que com probabilidade 1, acontece

$$\sum_{n=1}^{\infty} q(Y_{n-1})S_n = +\infty.$$

Consequentemente observa-se $\zeta \geq \infty$, e portanto com probabilidade 1 não acontece explosão.

(ii) Se Ω é finito, particularmente $\sup_{i\in\Omega}\{q(i)\} < \infty$, e consequentemente a demonstração do caso anterior é válida.

(iii) Se o estado i é recorrente, por definição, a cadeia visita i infinitas vezes com probabilidade 1. Portanto, sejam $V_1, V_2, V_3, \ldots$ as visitas ao estado i na cadeia $(Y_k)_{k\in\mathbb{N}}$. Note que

$$q(i)\zeta \geq \sum_{k=1}^{\infty} q(i)S_{V_k+1} = \sum_{k=1}^{\infty}(Y_n)S_{V_k+1} = \sum_{k=1}^{\infty} T_{V_k+1},$$

e portanto de maneira semelhante ao caso (i), concluímos que com probabilidade 1 não ocorre explosão.

■

Uma discussão interessante surge quando é feito o questionamento à respeito de maneiras equivalentes de se definir cadeias de Markov em tempo contínuo. Os próximos resultados apontarão nessa direção, obteremos formulações equivalentes, que em certos contextos proporcionam melhores interpretações acerca do fenômeno estudado. Inicialmente obteremos um resultado válido apenas para cadeias com espaço de estados finito, e em seguida obteremos um resultado equivalente para cadeias em um espaço de estados não-finito.

Teorema 2.2.24. *Seja $(X_t)_{t\geq 0}$ um processo minimal contínuo à direita tomando valores em um conjunto finito E. Seja Q uma matriz geradora em E com matriz de salto A. As seguintes condições são equivalentes:*

(i) *(definição via probabilidades de transição) Definição de cadeia de Markov em tempo contínuo da Definição 2.2.15, com matriz de saltos A, cadeia de saltos $(Y_n)_{n\in\mathbb{N}}$ e distribuição inicial δ_y;*

(ii) *(definição infinitesimal) Para quaisquer $t, h \geq 0$ condicionando em $X_t = y$, X_{t+h} é independente de $(X_s : s \leq t)$, e quando $h \downarrow 0$, uniformemente em t, para qualquer estado z tem-se*
$$\mathbb{P}(X_{t+h} = z \mid X_t = y) = \delta_{yz} + q(y,z)h + o(h);$$

(iii) *(definição via probabilidades de transição) Para todo $n \in \mathbb{N}$, quaisquer $0 \leq t_0 \leq t_1 \leq \cdots \leq t_{n+1}$ e quaisquer estados $y_0, y_1, \ldots, y_{n+1}$ tem-se*
$$\mathbb{P}\left(X_{t_n+1} = y_{n+1} \mid X_{t_0} = y_{t_0}, \ldots, X_{t_n} = y_{t_n}\right) = P^{t_{n+1}-t_n}(y_n, y_{n+1}),$$
onde $(P^t(a,b) : a, b \in E, t \geq 0)$ é a solução da equação progressiva (regressiva) de Kolmogorov:
$$P'(t) = P(t)Q, \quad P(0) = \mathrm{Id}\,.$$

Dem.: (i) $\implies$ (ii) Nesta demonstração abusaremos da notação $o(h)$ e $O(h)$ para simplificar as contas.
Observe que
$$\mathbb{P}_y(X_h = y) \geq \mathbb{P}_y(J_1 > h) = \mathrm{e}^{-q(yh)} = 1 + q(y,y)h + o(h),$$
e para $z \neq y$, temos
$$\begin{aligned}
\mathbb{P}_y(X_h = z) &\geq \mathbb{P}_y(J_1 \leq h, Y_1 = z, S_2 > h)\\
&= (1 - \mathrm{e}^{-q(y)h})a(y,z)\,\mathrm{e}^{-q(z)h}\\
&= (q(y)h + o(h))\frac{q(y,z)}{q(y)}(1 - q(z)h + o(h))\\
&= q(y,z)h(1 - q(z)h + o(h)) + o(h)\\
&= q(y,z)h + o(h).
\end{aligned}$$

Portanto, para qualquer estado z, vale
$$\mathbb{P}_y(X_h = z) \geq \delta_{y,z} + q(y,z)h + o(h),$$

o que implica

$$1 = \sum_{z\in E} \mathbb{P}_y(X_h = z) \geq \sum_{z\in E} (\delta_{y,z} + q(y,z)h) + o(h) = 1 + 0 + o(h).$$

Logo,

$$0 = \sum_{z\in E} (\mathbb{P}_y(X_h = z) - \delta_{y,z} + q(y,z)h) \geq o(h),$$

o que revela que

$$\mathbb{P}_y(X_h = z) = \delta_{y,z} + q(y,z)h + o(h).$$

Observe que por hipótese a cadeia $(X_t)_{t\geq 0}$ satisfaz a propriedade de Markov. Portanto, para quaisquer $t, h \geq 0$ e $z \in E$, temos

$$\mathbb{P}(X_{t+h} = z \mid X_t = y) = \mathbb{P}_y(X_h = z) = \delta_{y,z} + q(y,z)h + o(h).$$

Note que pela propriedade de Markov concluímos também que condicionado a $X_t = y, X_{t+h}$ é independente de $(X_s \, : \, s \leq t)$.

(ii) $\implies$ (iii) Para todo $t \geq 0$ e todo $z \in E$ defina $P^t(x,y) = \mathbb{P}_y(X_t = z)$

$$\begin{aligned}
P^{t+h}(z,y) &= \sum_{w\in E} \mathbb{P}_y(X_t = w)\mathbb{P}(X_{t+h} = z \mid X_t = w) \\
P^{t+h}(z,y) &= \sum_{w\in E} \left[P^t(z,y)(\delta_{w,z} + q(w,z)h + o(h))\right] \\
P^{t+h}(z,y) - P^t(z,y) &= \sum_{w\in E} \left[P^t(z,y)(q(w,z)h + o(h))\right]
\end{aligned}$$

e portanto

$$\begin{aligned}
\frac{P^{t+h}(z,y) - P^t(z,y)}{h} &= \sum_{w\in E} \left[P^t(z,y)(q(w,z) + o(1))\right] \\
&= \sum_{w\in E} \left[P^t(z,y)q(w,z)\right] + o(1).
\end{aligned}$$

Assim, fazendo $h \downarrow 0$, concluímos que $P^t(y,z)$ possui derivada à direita (em t).

Sabemos que

$$P^{t+h}(z,y) - P^t(z,y) = h\left(\sum_{w\in E} \left[P^t(z,y)q(w,z)\right] + o(1)\right),$$

portanto, para $h < t$, tomando $s = t - h$, obtemos

$$P^s(z,y) - P^{s-h}(z,y) = h\left(\sum_{w\in E} \left[P^{s-h}(z,y)q(w,z)\right] + o(1)\right),$$

e fazendo $h \downarrow 0$, concluímos que $P^t(z,y)$ é contínua à direita (em t). Repetindo o processo anterior, deduzimos que $P^t(z,y)$ também possui derivada à esquerda.

Portanto, concluímos que

$$\frac{\mathrm{d}}{\mathrm{d}t}P^t(x,y) = \sum_{z\in E} P^t(x,z)q(z,y) + O(h) \quad e \quad P^0(x,y) = \delta_{x,y},$$

ou seja as equações progressivas de Kolmogorov são satisfeitas, e como o espaço de estados E é finito, P^t é a única solução para a equação progressiva de Kolmogorov. Note que por hipótese, temos

$$\mathbb{P}\left(X_{t_n+1} = y_{n+1} \;\middle|\; X_{t_0} = y_{t_0}, \dots, X_{t_n} = y_{t_n}\right) = P^{t_{n+1}-t_n}(y_n, y_{n+1}),$$

para todo $n \in \mathbb{N}$, quaisquer $0 \leq t_0 \leq t_1 \leq \cdots \leq t_{n+1}$ e quaisquer estados $y_0, y_1, \dots, y_{n+1}$. Portanto, concluímos que

$$\mathbb{P}\left(X_{t_n+1} = y_{n+1} \;\middle|\; X_{t_0} = y_{t_0}, \dots, X_{t_n} = y_{t_n}\right) = P^{t_{n+1}-t_n}(y_{n+1} - t_n).$$

(iii) $\implies$ (i) Através das demonstrações acima, sabemos que uma cadeia Y definida por (i), satisfaz (iii). Portanto, se X é uma cadeia satisfazendo (iii), então as probabilidades de dimensão finita coincidem. Pelo Corolário A.1.12 segue que as probabilidades coincidem no espaço todo. Consequentemente, a cadeia X satisfaz (i). ■

Em alguns textos, define-se uma cadeia de Markov via um objeto que chamamos de operador infinitesimal, que de certa forma coincide com matrizes geradoras.

Definição 2.2.25 (Cadeia de Markov em tempo contínuo via gerador infinitesimal)**.** Fixado Ω enumerável, defina o operador L por

$$Lf(x) = \sum_{y\in\Omega} \alpha_{xy}\left[f(y) - f(x)\right]$$

para toda função $f : \Omega \to \mathbb{R}_+$ limitada. Definimos a cadeia gerada por L como sendo a cadeia gerada pela matriz Q tal que

$$q_{xy} = \begin{cases} \alpha_{xy}, & \text{se } x \neq y; \\ -\displaystyle\sum_{w\in\Omega\setminus\{x\}} \alpha_{xw}, & \text{se } x = y. \end{cases}$$

Os próximos dois teoremas irão generalizar resultados antes obtidos para Q matrizes em um espaço de estados finitos para espaços de estados quaisquer. As demonstrações podem ser encontradas em [Nor98].
O teorema abaixo é uma versão do Teorema 2.2.3 para cadeias em tempo contínuo em um espaço de estados não necessáriamente finito.

Teorema 2.2.26. *Seja Q uma matriz geradora. Então a equação regressiva de Kolmogorov*

$$P'(t) = QP(t), \qquad P(0) = Id$$

possui uma solução não-negativa minimal $(P(t)\,;\, t \geq 0)$. Essa solução forma um semigrupo matricial

$$P(s)P(t) = P(s+t) \qquad \text{para quaisquer } s,t \geq 0,$$

e também é a solução minimal não-negativa da equação progressiva de Kolmogorov

$$P'(t) = P(t)Q, \qquad P(0) = Id.$$

Teorema 2.2.27. *Seja $(X_t)_{t\geq 0}$ processo minimal e contínuo a direita com valores em I (Não necessariamente finito). Seja Q uma matriz geradora em I com matriz de salto A. As seguintes condições são equivalentes:*

(i) *$(X_t)_{t\geq 0}$ é uma cadeia de Markov em tempo contínuo como definida em 2.2.15.*

(ii) *$(X_t)_{t\geq 0}$ é um processo tal que para qualquer $n \in \mathbb{N}$, quaisquer $t_0 \leq t_1 \leq \cdots \leq t_{n+1} \in \mathbb{R}_+$, e estados $a_0, a_1, \ldots, a_{n+1} \in \Omega$, vale*

$$\mathbb{P}(X_{t_{n+1}} = a_{n+1} \,|\, X_{t_0} = a_0, \ldots, X_{t_n} = a_n) = P^{t_{n+1}-t_n}(a_n, a_{n+1}).$$

Note que os dois resultados acima sintetizam uma versão do Teorema 2.2.24 para os casos em que o espaço de estados não necessariamente é finito.

Note que a terceira definição do teorema anterior lembra bastante o caso discreto, onde descrevíamos a cadeia em função das probabilidades de transição. E a semelhança não é mera coincidência, pois fornece uma maneira de gerar cadeias contínuas através de cadeias discretas e vice-versa.

Exemplo 2.2.28 (Cadeia contínua através de cadeia discreta)**.** Seja $Z = (Z_t)_{t\in\mathbb{N}^*}$ uma cadeia de Markov em tempo discreto, com matriz de transição P. Uma maneira de obtermos uma "versão contínua" da cadeia acima consiste no seguinte processo: Defina a matriz M por

$$M(x,y) = \begin{cases} P(x,y), & \text{se } x \neq y; \\ -\displaystyle\sum_{w\in\Omega\setminus\{x\}} P(x,w), & \text{se } x = y. \end{cases}$$

Observe que M é uma matriz geradora. Portanto podemos obter uma cadeia em tempo contínuo através de M.
Fixado $x \in \Omega$, suponha que tanto Z quanto X possuem distribuição inicial δ_x. Assim, supondo que $P(x,x) < 1$ e denotando por $T_1(Z), T_1(X)$ os tempos do primeiro salto das cadeias Z e X respectivamente, temos

$$\mathbb{E}_x\left[T_1(X)\right] = \frac{1}{1-P(x,x)},$$

já que pela definição $T_1(X) \sim Exp\left(-M(x,x)\right)$, e $M(x,x) = P(x,x) - 1$.
Além disso,

$$\mathbb{E}_x\left[T_1(Z)\right] = \sum_{k\in\mathbb{N}^*}\left[(1-P(x,x))P(x,x)^k(k+1)\right] = \frac{1}{1-P(x,x)}.$$

Ou seja, a cadeia de Markov X gerada por M se comporta de maneira semelhante a Z, no sentido que em cada um dos sítios o tempo médio de espera até a troca de estado é o mesmo. Note que o caso $P(x,x) = 1$ é imediato.

Observação 2.2.29. Através de uma matriz geradora, obtemos apenas uma matriz de salto, porém, fixada uma matriz de salto, existe uma infinidade de matrizes geradoras que a tem como matriz de saltos. Por conta disso, fixada uma cadeia em tempo discreto, no exemplo acima poderíamos obter várias versões da cadeia discreta com "tempo distorcido" (Inclusive distorções diferentes em cada um dos sítios), como por exemplo as matrizes da forma αM com $\alpha > 0$.

Exemplo 2.2.30 (Cadeia discreta através de cadeia contínua)**.** Fixada uma cadeia X em tempo contínuo, com núcleo de Transição $P = \{P^t\,;\, t \geq 0\}$, a cadeia Y em tempo discreto gerada por P^1, é uma espécie de discretização de X. Observe que Y mantém uma certa semelhança com a cadeia X, por conta da propriedade de semigrupo da última, no sentido que nos tempos em $\mathbb{N}$, as probabilidades de estar em cada um dos sítios são iguais.
Além disso, assim como no Exemplo 2.2.28 os tempos esperados de salto são os mesmos

Observação 2.2.31. Note que no exemplo acima, poderíamos ter optado por usar P^h com $h \neq 1$. Isso resultaria em uma espécie de discretização da cadeia contínua com uma "distorção temporal".
Essa técnica é conhecida na literatura como "sampling", é possui algumas aplicações em simulações (não só de cadeias de Markov, mas em outros processos). A (provável) principal delas é conhecida como "Markov Chain Monte Carlo" (MCMC) ou "Método de Monte Carlo" para cadeias de Markov.

2.3 Propriedades de cadeias obtidas por matrizes geradoras

Para simplificar a notação, ao longo desta seção a cadeia em questão será $(X_t)_{t\geq 0}$ com espaço de estados E e matriz geradora Q. Quando necessário, suporemos que o espaço de probabilidade em questão é $(\Omega, \mathcal{F}, \mathbb{P})$. O conteúdo desta seção é fortemente baseado em [Nor98].

Buscaremos obter resultados semelhantes ao obtidos para cadeias em tempo discreto. Assim como no caso discreto, fixado um semigrupo de transição P, usamos $x \to y$ (leia-se x leva à y) para dizer que existe $t > 0$ de modo que $P^t(x,y) > 0$, e se $x \to y$ e $y \to x$ então escrevemos $x \leftrightarrow y$ e dizemos que x e y se comunicam.

Teorema 2.3.1. *Fixados estados i e j distintos, as seguintes propriedades são equivalentes:*

(i) $P^t(i,j) > 0$ *para todo* $t > 0$*;*

(ii) $i \to j$*;*

(iii) $i \to j$ *na cadeia de saltos;*

(iv) *Existem estados* $i_1, \ldots, i_{n-1}$ *tais que* $\prod_{k=0}^{n-1} q(i_k, i_{k+1}) > 0$ *, onde* $i_0 = i$*,* $i_n = j$*.*

Dem.: *(i)* $\implies$ *(ii)* Segue imediatamente da definição de $i \to j$.

(ii) $\implies$ *(iii)* Se $i \to j$, então $\mathbb{P}_i(X_t = j$ para algum $t > 0) > 0$, o que implica $\mathbb{P}_i(X_{J_n} = j$ para algum $n \in \mathbb{N}^*) > 0$, e consequentemente em $i \to j$ na cadeia de saltos.

(iii) $\implies$ *(iv)* Pela Observação 1.1.37, sabemos que existem e estados $i_1, \ldots, i_{n-1}$ de modo que $a(i_k, i_{k+1}) > 0$ para qualquer $k \in \{0, 1, \ldots, n-1\}$ onde $A = \{a(x, y) \,:\, x, y \in \Omega\}$ é a matriz de saltos. Segue da definição da Matriz de Saltos que $q(i_k, i_{k+1}) = q(i_k)a(i_k, i_{k+1}) > 0$ para todo $k \in \{0, 1, \ldots, n-1\}$.

(iv) $\implies$ *(i)* Fixados estados $x, y \in \Omega$ tais que $q(x, y) > 0$, note que

$$p^t(x, y) \geq \mathbb{P}_i\left(J_1 \leq t,\, Y_1 = y,\, S_2 > t\right) = (1 - \mathrm{e}^{-q(x)t})a(x, y)\mathrm{e}^{-q(y)t} > 0,$$

onde a igualdade acima segue da independência das variáveis. Portanto, pelas equações de Chapman-Kolmogorov (Observação 2.2.5), segue

$$p^t(i, j) \geq \prod_{k=0}^{n-1} p^{t/n}(i_k, i_{k+1}) > 0 \geq \prod_{k=0}^{n-1}(1 - \mathrm{e}^{-q(i_k)t/n})a(i_k, i_{k+1})\mathrm{e}^{-q(i_{k+1})t/n} > 0.$$

■

Note que o resultado acima revela uma diferença interessante quando comparado ao caso em tempo discreto, pois a noção de periodicidade deixa de existir, visto que em qualquer intervalo de tempo é possível observar uma transição entre estados que se comunicam. Particularmente, se a cadeia de saltos for irredutível, então em qualquer intervalo de tempo é possível observar qualquer transição entre estados quaisquer. Isso simplifica um pouco o estudo.

Definição 2.3.2 (Estados recorrentes e transientes)**.** Fixada uma cadeia de Markov $(X_t)_{t\geq 0}$ em tempo contínuo, dizemos que um estado x é recorrente se

$$\mathbb{P}_x\left(\{t \geq 0 \,:\, X_t = x\} \text{ é ilimitado}\right) = 1.$$

Em contrapartida, dizemos que x é transiente se

$$\mathbb{P}_x\left(\{t \geq 0 \,:\, X_t = x\} \text{ é ilimitado}\right) = 0.$$

Abaixo, obteremos um resultado que nos permite relacionar transiência e recorrencia de cadeias em tempo contínuo com o estudo da recorrência e transiência da cadeia de salto, tornando possível o uso dos resultados obtidos no caso discreto.

Teorema 2.3.3. *As seguintes afirmações são verdadeiras:*

(i) *se x é recorrente para a cadeia de salto $(Y_n)_{n\in\mathbb{N}}$, então x é recorrente para $(X_t)_{t\geq 0}$;*

(ii) *Se x é transiente para a cadeia de salto, então x é transiente para $(X_t)_{t\geq 0}$;*

(iii) *Cada um dos estados é recorrente ou transiente;*

(iv) *Recorrência e transiência são propriedades de classe.*

Dem.: (i) Suponha que x é recorrente para a cadeia de saltos. Se $X_0 = x$, pelo Teorema 2.2.23 que a cadeia não explode. Segue do fato de x ser recorrente para a cadeia de saltos que com probabilidade 1, $x = Y_n = X_{J_n}$ para uma quantidade infinita de índices $n \in \mathbb{N}$. Como a cadeia não explode, particularmente, a sequência $(J_n)_{n\in\mathbb{N}}$ é não-limitada.
Portanto, segue da definição de recorrência que a cadeia $(X_t)_{t\geq 0}$ é recorrente.

(ii) Suponha que x é transiente para a cadeia de saltos. Se $X_0 = x$, então com probabilidade 1,

$$N := \sup\{n \geq 0 \ : \ Y_n = x\} < \infty,$$

e portanto, $\{t \geq : \ X_t = x\} \leq J_{N+1} < \infty$ com probabilidade 1.

(iii) e (iv) seguem dos itens (i),(ii), e da Proposição 1.1.49. ■

Definição 2.3.4. Fixada uma cadeia $(X_t)_{t\geq 0}$, chamamos de primeiro tempo de passagem ao estado x a variável aleatória denotada por T_x e definida por

$$T_x(\omega) = \inf\{t \geq J_1(\omega) \ : \ X_t(\omega) = x\}.$$

para qualquer $\omega \in \Omega$.

Proposição 2.3.5. *As seguintes afirmações são verdadeiras:*

(i) *se* $q_i = 0$ *ou* $\mathbb{P}_i(T_i < \infty) = 1$, *então* i *é recorrente e* $\displaystyle\int_0^\infty p^t(i,i)\,\mathrm{d}t = \infty.$

(ii) *se* $q_i > 0$ *e* $\mathbb{P}_i(T_i < \infty) < 1$, *então* i *é transiente e* $\displaystyle\int_0^\infty p^t(i,i)\,\mathrm{d}t < \infty.$

Dem.: O caso $q(i) = 0$ é imediato, portanto, iremos supor que $q(i) > 0$ e $\mathbb{P}_i(T_i < \infty) = 1$. Definindo $N_x := \min\{n \in \mathbb{N}^* \ : \ Y_n = i\}$, é possível notar que claramente

$$\mathbb{P}_i(N_x < \infty) = \mathbb{P}_i(T_i < \infty),$$

e consequentemente pelo Teorema 2.3.3 e pela Proposição 1.1.49, o estado i é recorrente se e somente se $\mathbb{P}_i(T_i < \infty) = 1$.
Portanto, para finalizar a demonstração basta mostrar que as integrais assumem os valores descritos no enunciado.

Note que:

$$\int_0^\infty p^t(i,i)\,\mathrm{d}t = \int_0^\infty \mathbb{E}_i(\mathbf{1}_{\{X_t=i\}})\,\mathrm{d}t.$$

Segue do Teorema de Fubini (Teorema A.1.4) que

$$\int_0^\infty p^t(i,i)\,\mathrm{d}t = \mathbb{E}_i\left[\int_0^\infty \mathbf{1}_{\{X_t=i\}}\right]\mathrm{d}t$$

$$= \mathbb{E}_i \left[\sum_{n=0}^{\infty} \mathbf{1}_{\{Y_n=i\}} S_n \right] \mathrm{d}t.$$

Novamente, pelo Teorema de Fubini, temos

$$\int_0^{\infty} p^t(i,i)\,\mathrm{d}t = \sum_{n=0}^{\infty} \mathbb{E}_i \left[\mathbf{1}_{\{Y_n=i\}} S_n \right] \mathrm{d}t.$$

Como S_{n+1} e Y_n são independentes, segue que

$$\begin{aligned} \int_0^{\infty} p^t(i,i)\,\mathrm{d}t &= \sum_{n=0}^{\infty} \mathbb{E}_i \left[S_{n+1} | Y_n = i \right] \mathbb{P}_i(Y_n = i) \\ &= \frac{1}{q(i)} \sum_{n=0}^{\infty} a^n(i,i), \end{aligned}$$

onde $A = (a(i,j) \,:\, i,j \in E)$ é a matriz de saltos.

Portanto, segue da Proposição 1.1.49 que

$$\int_0^{\infty} p^t(i,i)\,\mathrm{d}t < \infty \text{ se e somente se } i \text{ é transiente.}$$

■

No estudo de cadeias em tempo discreto, vimos que medidas invariantes desempenham um papel fundamental na compreensão do comportamento de cadeias de Markov. Definiremos uma noção de medidas invariantes para matriz geradora, e em seguida revelaremos a sua relação com a medida invariante da cadeia.

Definição 2.3.6. Fixada uma matriz geradora Q, Dizemos que uma medida λ definida no espaço de estados que indexa Q é invariante com respeito à Q se

$$\lambda Q = 0.$$

Proposição 2.3.7. *Seja Q uma matriz geradora com matriz de salto A. Fixada uma medida λ, as seguintes condições são equivalentes:*

(i) *λ é invariante para Q;*

(ii) *a medida μ definida por $\mu(x) = \lambda(x)q(x)$ é invariante com respeito à matriz de salto ($\mu A = \mu$).*

Dem.: Note que

$$(\mu A)(y) = \sum_{x\in\Omega} \mu(x)a(x,y) = \sum_{x\in\Omega} \lambda(x)q(x)a(x,y) \tag{2.1}$$

$$= \sum_{x\in\Omega} \lambda(x)\left[q(x,y) + \delta_{xy}q(x)\right] = (\lambda Q)(y) + (\mu\,\mathrm{Id})(y), \tag{2.2}$$

o que garante o resultado. ■

Note que se a cadeia de saltos possuir uma única distribuição invariante, então Q também possuirá uma única distribuição invariante, e qualquer medida que seja um múltiplo dessas medidas é também invariante para Q ou A.
Particularmente, se $q(x) = q(y)$ para quaisquer estados $x, y \in E$, isto é, a taxa de saída de qualquer estado é a mesma, então Q e a cadeia de saltos possuem a mesma distribuição invariante. Isso se mostra bastante útil quando o objeto estudado é uma versão contínua de uma cadeia em tempo discreto (Exemplo 2.2.30).

Proposição 2.3.8. *Seja X uma cadeia de Markov em tempo contínuo irredutível, com espaço de estados E e gerada por uma matriz geradora Q. Se Q possui uma distribuição invariante λ* ou *existe algum estado $x \in E$ de modo que $\mathbb{E}_x[T_x] < \infty$, então $\mathbb{E}_y[T_y] < \infty$ para todo $y \in \Omega$, e a medida π dada por*

$$\pi(y) = \frac{1}{q(y)\mathbb{E}_y[T_y]}$$

é a única distribuição invariante para Q.

Dem.: Suponha que $\mathbb{E}_x[T_x] < \infty$. Seja Y_n a cadeia de saltos de X, e sejam também N_x o tempo de primeira passagem do estado x na cadeia de saltos.
Definamos a medida μ^x dada por

$$\mu^x(y) = \mathbb{E}_x\left[\int_0^{T_x} \mathbf{1}_{\{X_s=y\}}\,\mathrm{d}s\right].$$

Note que

$$\mu^x(y) = \mathbb{E}_x\left[\int_0^{T_x} \mathbf{1}_{\{X_s=y\}}\,\mathrm{d}s\right] = \mathbb{E}_x\left[\sum_{n=0}^{\infty} S_{n+1}\mathbf{1}_{\{Y_n=y,n<N_x\}}\right],$$

e pelo Teorema de Fubini (Teorema A.1.4), segue que

$$\mu^x(y) = \sum_{n=0}^{\infty} \mathbb{E}_x\left[S_{n+1}\mathbf{1}_{\{Y_n=y\}\cap\{n<N_x\}}\right].$$

Observe que pela Lema A.2.8, temos

$$\begin{aligned}
&\mathbb{E}_x\left[S_{n+1}\mathbf{1}_{\{Y_n=y\}\cap\{n<N_x\}}\right] \\
&\quad= \mathbb{E}_x\left[S_{n+1}\mathbf{1}_{\{Y_n=y\}\cap\{n<N_x\}}\,|\,Y_n = y\,,\,n < N_x\right]\mathbb{P}_x(Y_n = y\,,\,n < N_x) \\
&\qquad\quad + \mathbb{E}_x\left[S_{n+1}\mathbf{1}_{\{Y_n=y\}\cap\{n<N_x\}}\,|\,Y_n \neq y \text{ ou } n > N_x\right]\mathbb{P}_x(\{Y_n \neq y\} \cup \{n > N_x\}) \\
&\quad= \mathbb{E}_x\left[S_{n+1}\mathbf{1}_{\{Y_n=y\}\cap\{n<N_x\}}\,|\,Y_n = x\,,\,n < N_x\right]\mathbb{P}_x(Y_n = y\,,\,n < N_x) + 0 \\
&\quad= \mathbb{E}_x\left[S_{n+1} \mid Y_n = x\right]\mathbb{P}_x(Y_n = y\,,\,n < N_x).
\end{aligned}$$

Portanto,

$$\mu^x(y) = \sum_{n=0}^{\infty} \mathbb{E}_x\left[S_{n+1} \mid Y_n = x\right]\mathbb{P}_x(Y_n = y\,,\,n < N_x)$$

$$= \sum_{n=0}^{\infty} q(y)^{-1} \mathbb{E}_x \left(\mathbf{1}_{\{Y_n=y\}\cap\{n<N_x\}} \right).$$

Novamente, pelo Teorema de Fubini (Teorema A.1.4), temos

$$\begin{aligned} \mu^x(y) &= q(y)^{-1} \mathbb{E}_x \left[\sum_{n=0}^{\infty} \mathbf{1}_{\{Y_n=y\}\cap\{n<N_x\}} \right] \\ &= q(y)^{-1} \mathbb{E}_x \left[\sum_{n=0}^{N_x-1} \mathbf{1}_{\{Y_n=y\}} \right]. \end{aligned}$$

Assim como na notação do Teorema 1.1.52, usaremos $V^x(y)$ para denotar o número esperado de visitas ao estado y entre duas visitas consecutivas de x na cadeia de saltos A.

Portanto, temos

$$\mu^x(y) = \frac{V^x(y)}{q(y)}. \tag{2.3}$$

Suponha que $\mathbb{E}_x [T_x] < \infty$. Neste caso, por consequência, temos $\mathbb{E}_x [N_x] < \infty$. assim, segue do Teorema 1.1.52 que $V^x(\cdot)$ é uma medida invariante para a cadeia de saltos A.

Logo, segue da Proposição 2.3.7 que μ^x é uma medida invariante para Q. Como

$$\sum_{y\in E} \mu^x(y) = \mathbb{E}_x [T_x] < \infty,$$

segue que

$$\frac{\mu^x(\cdot)}{\mathbb{E}_x [T_x]}$$

é uma distribuição invariante para Q. Ou seja, este caso recai no próximo.

Agora, suponha que Q possui uma distribuição invariante λ, mostraremos que λ é a medida descrita no enunciado.

Fixado um estado x qualquer, pela Proposição 2.3.7 a medida ν definida por

$$\nu(y) = \lambda(y)q(y)$$

é invariante para a cadeia de saltos A. Portanto pelo Teorema 1.1.52 segue que $\nu = kV^x$ para alguma constante positiva k.

Note que

$$\mathbb{E}_x [T_x] = \sum_{y\in E} \mu^x(y) = \sum_{y\in E} \frac{V^x(y)}{q(y)} = \sum_{y\in E} \frac{k\nu(y)}{q(y)} = \sum_{y\in E} k\lambda(y) = k < \infty.$$

Como tomamos $x \in E$ qualquer, concluímos que $\mathbb{E}_y [T_y] < \infty$ para qualquer estado $y \in E$, e consequentemente todos os estados são recorrentes.

Repetindo os passos da primeira parte, concluímos que par todo $x \in E$, V^x é uma medida invariante (e finita) para a matriz de salto A. Assim, segue do Teorema 1.1.52 que A possui exatamente uma medida invariante, e consequentemente, pela Proposição 2.3.7 λ é a única distribuição invariante para Q.

Como pela primeira parte da demonstração sabemos que para todo $x \in E$, μ^x é uma medida (finita) invariante para Q, repetindo os passos da Observação 1.1.53, concluímos que π definida para todo $y \in E$ por

$$\pi(y) = \frac{1}{q(y)\mathbb{E}_y\left[T_y\right]}$$

é a única distribuição invariante para Q. ■

Observação 2.3.9. Assim como em uma observação anterior, note que se tivermos $q(y) = q(x)$ para quaisquer estados $x, y \in E$, então a medida π acima também é a distribuição invariante para a cadeia de saltos.

Observação 2.3.10. Note que nas hipóteses do resultado acima, dizer que Q possui uma distribuição invariante λ significa dizer que $\mathbb{E}_x\left[T_x\right] < \infty$, o que particularmente implica em $\mathbb{P}_x(T_x < \infty) = 1$.

Teorema 2.3.11. *Sejam: Q uma matriz geradora irredutível e recorrente; e λ uma medida. As seguintes afirmações são equivalentes*

(i) $\lambda Q = 0$;

(ii) $\lambda P^h = \lambda$ *para todo* $h \in \mathbb{R}_+$.

Dem.: Por ser recorrente segue do Teorema 2.2.23, que a cadeia não explode.

Observe que para para qualquer $s \in [nh, nh + h)$, utilizando as Equações de Chapman-Kolmogorov e a propriedade de Markov, temos

$$\begin{aligned}
\left(P^h\right)^{n+1}(x,x) &= P^{(n+1)h}(x,x) \\
&\geq P^s(x,x)P^{(n+1)h-s}(x,x) \\
&\geq P^s(x,x)\mathbb{P}_x(J_1 > (n+1)h - s) \\
&= P^s(x,x)\,\mathrm{e}^{-q(x)((n+1)h-s)} \\
&\geq P^s(x,x)\,\mathrm{e}^{-q(x)h}
\end{aligned}$$

Assim através da inequação acima e do uso do Teorema da Convergência Monótona, obtemos.

$$\mathrm{e}^{-q(x)h}\int_0^\infty P^t(x,x)\,\mathrm{d}t \leq \int_0^\infty \left(\sum_{n=0}^\infty P^{(n+1)h}\mathbf{1}_{[nh\,,\,(n+1)h)}\right)\mathrm{d}t = \sum_{n=1}^\infty P^{nh}(x,x) = \sum_{n=1}^\infty (P^h)^n(x,x).$$

Em vista Teorema 2.3.3, por Q ser recorrente, a integral acima não é finita. Assim, pelo Proposição 1.1.49, concluímos que P^h é recorrente.

Por ser recorrente e irredutível, segue do Corolário 1.1.54 que cada P^s possui no máximo uma distribuição invariante. Note também, que por Q ser recorrente e irredutível, pelo teorema anterior Q possui no máximo uma distribuição invariante. Portanto, qualquer medida λ satisfazendo um dos itens do enunciado é única a menos de constante.

Fixado um estado $x \in E$, sabemos da demonstração do Teorema 1.1.52 que a medida μ definida por

$$\mu(y) = \int_0^{T_x} \mathbf{1}_{\{X_t=y\}}\,\mathrm{d}t$$

é invariante para Q.

Logo, para finalizar a demonstração do resultado, basta verificar que $\mu P(s) = \mu$ para todo $s \geq 0$.

Pela propriedade forte de Markov, temos

$$\mathbb{E}_x\left[\int_0^s \mathbf{1}_{\{X_t=y\}}\,\mathrm{d}t\right] = \mathbb{E}_x\left[\int_{T_x}^{s+T_x} \mathbf{1}_{\{X_t=y\}}\,\mathrm{d}t\right].$$

Portanto

$$\begin{aligned}\mu(y) &= \mathbb{E}_x\left[\int_s^{s+T_x} \mathbf{1}_{\{X_t=y\}}\,\mathrm{d}t\right]\\ &= \mathbb{E}_x\left[\int_s^{\infty} \mathbf{1}_{\{X_t=y\}}\mathbf{1}_{\{T_x>t\}}\,\mathrm{d}t\right].\end{aligned}$$

Segue do Teorema de Fubini que

$$\begin{aligned}\mu(y) &= \int_s^{\infty} \mathbb{E}_x\left[\mathbf{1}_{\{X_t=y\}}\mathbf{1}_{\{T_x>t\}}\,\mathrm{d}t\right]\\ &= \int_0^{\infty} \mathbb{P}_x\left(X_{s+t}=y\,,\,t<T_x\right)\mathrm{d}t\end{aligned}$$

de onde segue da propriedade de Markov que

$$\begin{aligned}\mu(y) &= \int_0^{\infty} \sum_{z\in E} \mathbb{P}_x(X_t=z\,,\,t<T_i)P^s(z,y)\,\mathrm{d}t\\ &= \sum_{z\in E}\left(\mathbb{E}_x\left[\int_0^{T_y} \mathbf{1}_{\{X_t=y\}}\,\mathrm{d}t\right]P^s(z,y)\right)\\ &= \sum_{z\in E}\mu(k)P^s(z,y)\\ &= (\mu P^s)(y)\end{aligned}$$

o que finaliza a demonstração do resultado. ∎

Corolário 2.3.12. *Para que λ seja uma distribuição invariante para Q é suficiente que se tenha*

$$\lambda(x)Q(x,y) = \lambda(y)Q(y,x)$$

para quaisquer estados $x, y \in E$.

Dem.: Note que por hipótese temos:

$$0 = \sum_{x \in E} Q(y,x) = \sum_{x \in E} [\lambda(y)Q(y,x)] = \sum_{x \in E} [\lambda(x)Q(x,y)] = (\lambda Q)(y),$$

que em vista do teorema anterior garante o resultado.

■

Teorema 2.3.13 (Convergência ao equilíbrio)**.** *Seja Q uma matriz geradora irredutível com semi-grupo de transição $P(t)$, tal que sua distribuição invariante seja λ. Nesse caso para quaisquer x, y no espaço de estados E, temos*

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

Particularmente, se E for finito, vale

$$\lim_{t \to \infty} \mathrm{d}_{TV}(Lei(X(t)), \lambda) = 0,$$

onde $(X(t))_{t \geq 0}$ é a cadeia gerada por Q.

Dem.: Como por hipótese Q possui uma distribuição invariante λ, segue da Observação 2.3.10 que Q é recorrente. Portanto, segue do Teorema 2.3.11 que $\lambda P^s = \lambda$ para qualquer $s \geq 0$. Além disso, pelo Teorema 2.3.1 sabemos que P^s possui todas as entradas positivas, o que garante que P^s é irredutível para todo $s \geq 0$.

Com as informações acima em mente, observe que pelo Teorema da Convergência em tempo discreto (Teorema 1.1.59), e pela propriedade de semigrupo, temos

$$\lim_{n \to \infty} P^{nh}(x,y) = \lim_{n \to \infty} (P^h)^n(x,y) = \lambda(y)$$

para quaisquer $x, y \in E$.

Além disso, observe que se $t \in [nh\,,\,(n+1)h)$, então

$$\begin{aligned}
\left|P^t(x,y) - P^{nh}(x,y)\right| &= \left|\sum_{z \in E} P^{t-hn}(x,z)P^{hn}(y,z) - P^{hn}(x,y)\right| \\
&= \left|\sum_{z \in E} P^{t-hn}(x,z)P^{hn}(y,z) - \sum_{z \in E} P^{t-hn}(x,z)P^{hn}(x,y)\right| \\
&= \left|\sum_{z \in E \setminus \{x\}} \left(P^{t-hn}(x,z)\left[P^{hn}(y,z) - P^{hn}(x,y)\right]\right)\right| \\
&\leq \sum_{z \in E \setminus \{x\}} P^{t-hn}(x,z) \\
&= 1 - P^{t-hn}(x,x)
\end{aligned}$$

$$
\begin{aligned}
&\leq \mathbb{P}_x(J_1 \leq t - hn) \\
&= 1 - \mathrm{e}^{q(x)(t-hn)} \\
&\leq 1 - \mathrm{e}^{q(x)h}.
\end{aligned}
$$

Assim, dado $\varepsilon > 0$ qualquer e fixados $x, y \in E$ quaisquer, tome $h > 0$ de modo que se tenha

$$1 - \mathrm{e}^{-q(x)h} \leq \frac{\varepsilon}{2}.$$

Tome também $N \in \mathbb{N}$ de modo que se tenha

$$\left|P^{nh}(x,y) - \lambda(y)\right| \leq \frac{\varepsilon}{2}$$

para qualquer $n > N$.

Portanto, se $t \geq Nh$ e $t \in [nh\,,\,(n+1)h)$, temos

$$\left|P^{t}(x,y) - \lambda(y)\right| = \left|P^{t}(x,y) - P^{nh}(x,y)\right| + \left|P^{nh}(x,y) - \lambda(y)\right| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

Como tomamos $\varepsilon > 0$ qualquer, concluímos que

$$\lim_{t\to\infty} P^t(x,y) = \lambda(y)$$

para quaisquer $x, y \in E$. ■

Finalizaremos o capítulo enunciando um resultado que indica que o tempo médio que a cadeia passa em cada um dos estados é inversamente proporcional ao tempo médio de retorno dos mesmos.

Teorema 2.3.14 (Teorema Ergódico)**.** *Seja Q uma matriz geradora irredutível e ν uma distribuição qualquer no espaço de estados E. Se $(X_t)_{t\geq 0}$ é uma cadeia de Markov em tempo contínuo com distribuição inicial ν, então*

$$\mathbb{P}\left(\lim_{t\to\infty} \frac{1}{t}\int_0^t \mathbf{1}_{\{X_s = x\}}\,\mathrm{d}s = \frac{1}{q(x)\mathbb{E}_x\left[T_x\right]}\right) = 1.$$

3
Acoplamentos, Tempos de Mistura e o Fenômeno Cutoff

Nas duas primeiras seções deste capítulo, exploraremos acoplamentos e tempos de mistura em cadeias de Markov. Ao fixarmos duas matrizes de transição, desenvolveremos uma abordagem para construir acoplamentos entre variáveis aleatórias associadas a essas matrizes. Além disso, examinaremos como criar acoplamentos entre duas cadeias em tempo contínuo, tendo uma matriz geradora que satisfaça condições específicas. Exibiremos alguns exemplos de acoplamentos explicando as suas dinâmicas.

Introduziremos o Teorema da Dualidade, um resultado que estabelece uma conexão entre acoplamentos e tempos de mistura. Esse resultado nos permitirá estabelecer uma cota superior para o tempo de mistura usando um tempo de parada associado ao acoplamento. Finalizando a segunda seção, exploraremos aplicações práticas desse resultado, calculando cotas para o tempo de mistura em algumas cadeias de Markov específicas.

Já na última seção, estudaremos um fenômeno observado em algumas sequências de cadeias de Markov, conhecido como *cutoff*. Esse fenômeno, observado em uma sequência $(X_t^n; t \geq 0)_{n \in \mathbb{N}}$ de cadeias de Markov, caracteriza-se pelo decrescimento abrupto da distância de variação total de $1 - \varepsilon$ até ε (para $\varepsilon \in (0, 1/2)$) quando analisado na escala do tempo de mistura, de modo que conforme n cresce a distância de variação total na escala dos tempos de mistura t_{mix}^n se aproxima de uma função degrau.

Apresentaremos uma condição necessária para a ocorrência do cutoff, juntamente com uma caracterização especialmente útil desse fenômeno. Introduziremos os conceitos de "janela do cutoff" e "perfil do cutoff", que, ao serem identificados em uma sequência, proporcionarão uma descrição extremamente precisa do cutoff.
Ao longo deste capítulo utilizaremos $\mathbb{F}$ para representar $\mathbb{N}$ ou $\mathbb{R}_{\geq 0}$.

Na última parte do capítulo, destacaremos alguns exemplos de cutoff e definiremos o processo de exclusão, que será o foco de estudo no capítulo subsequente.
Este capítulo é fundamentado no conteúdo de [LPW].

3.1 Distribuições estacionárias, acoplamentos e distância de variação total

Definição 3.1.1 (Acoplamento)**.** Dadas duas distribuições μ e ν definidas em $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$ respectivamente, dizemos que a medida γ definida em $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ (onde $\mathcal{F}_1 \otimes \mathcal{F}_2$ é a menor sigma álgebra contendo $\mathcal{F}_1 \times \mathcal{F}_2$) é um acoplamento de μ e ν se

$$\gamma(A \times \Omega_2) = \mu(A) \text{ e } \gamma(\Omega_1 \times B) = \nu(B)$$

para quaisquer $A \in \Omega_1, B \in \Omega_2$.
Para duas variáveis aleatórias X e Y definidas em $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ e $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ respectivamente, dizemos que uma variável aleatória (Z, W) definida em algum espaço $\left(\overline{\Omega}, \overline{\mathcal{F}}\right)$ unida a uma medida $\overline{\mathbb{P}}$ é um acoplamento de X e Y se

$$\overline{\mathbb{P}}(\pi_1(Z,W) \in A) = \mathbb{P}_1(X^{-1}(A)) \quad \text{e} \quad \overline{\mathbb{P}}(\pi_2(Z,W) \in B) = \mathbb{P}_2(Y^{-1}(B))$$

para quaisquer A e B mensuráveis nos espaços de chegada de X e Y respectivamente, onde π_1 e π_2 são as projeções na primeira e segunda coordenada respectivamente.

Observação 3.1.2. É possível ver na definição acima que acoplamentos entre distribuições estão intimamente ligados com acoplamentos de variáveis aleatórias, por conta disso é comum encontrar definições ligeiramente diferentes, mas ainda equivalentes. Portanto, se X e Y forem variáveis aleatórias tais que X possui distribuição μ e Y possui distribuição ν, e γ for um acoplamento de μ e ν, por vezes diremos que γ é um acoplamento de X e Y; Assim como diremos que (X', Y') é um acoplamento de μ e ν caso seja um acoplamento de X e Y.

Note que $\gamma := \overline{\mathbb{P}} \circ (Z, W)^{-1}$ é um acoplamento de μ e ν.

Usaremos esta ideia para acoplar cadeias, e posteriormente obter cotas no tempo que a cadeia leva para convergência.

Definição 3.1.3 (Acoplamento de Cadeias de Markov)**.** Sejam $(X_t)_{t\in\mathbb{F}}$ e $(Y_t)_{t\in\mathbb{F}}$ com $\mathbb{F} = \mathbb{R}_{\geq 0}$ ou $\mathbb{N}$, duas cadeias de Markov definidas no mesmo espaço Ω. Dizemos que $(W_t)_{t\in\mathbb{F}}$ é um acoplamento de $(X_t)_{t\in\mathbb{F}}$ e $(Y_t)_{t\in\mathbb{F}}$ se para cada $t \in \mathbb{F}$ W_t é um acoplamento de X_t e Y_t.

Definido o que vem a ser um acoplamento de cadeias de Markov, percebemos que construir acoplamentos nem sempre é uma tarefa fácil. Com a finalidade de facilitar a obtenção de acoplamentos a observação abaixo exibirá uma maneira de gerarmos um acoplamento apenas exibindo uma matriz de transição que satisfaça uma certa condição. Isso nos permitirá citar exemplos e obter resultados de uma maneira mais simples.

Definição 3.1.4. Se P é uma matriz de transição em um espaço de estados E enumerável, fixado $A \subseteq E$, definimos

$$P(x, A) := \sum_{y \in A} P(x, y)$$

para todo $x \in E$.

Observação 3.1.5 (Acoplamento via matriz de transição)**.** Sejam $(X_t)_{t\in\mathbb{N}}$ e $(Y_t)_{t\in\mathbb{N}}$ cadeias de Markov definidas em um mesmo espaço de probabilidade, com espaço de estados E enumerável, distribuições iniciais λ_1 e λ_2, e matrizes de transição P_1 e P_2 respectivamente. Adicionalmente, seja $\overline{P}$ uma matriz de transição em $E\times E$ tal que se tenha

$$\overline{P}((a,b)\,,\,\{x\}\times E)=P_1(a,x)\ \text{ e }\ \overline{P}((c,d)\,,\,E\times\{y\})=P_2(d,y)$$

para quaisquer estados $a,b,c,d,x,y\in E$,.
Pela construção de cadeias discretas feita no segundo capítulo, existe uma cadeia de Markov $(Z_t)_{t\in\mathbb{N}}=(X'_t,Y'_t)_{t\in\mathbb{N}}$ definida em um espaço $(\Omega',\mathcal{F}',\overline{\mathbb{P}})$, assumindo valores em $(E\times E,\mathcal{P}(E\times E))$ e com matriz de transição $\overline{P}$. Note que para qualquer $t\in\mathbb{N}^*$, temos

$$\overline{\mathbb{P}}(X'_t\in A)=\overline{\mathbb{P}}((X'_t,Y'_t)\in A\times E),$$

segue da probabilidade total que

$$\overline{\mathbb{P}}(X'_t\in A)=\sum_{x_{t-1},y_{t-1}\in E}\overline{\mathbb{P}}\big((X'_t,Y'_t)\in A\times E,\,(X'_{t-1},Y'_{t-1})=(x_{t-1},y_{t-1})\big).$$

Logo, segue da propriedade de Markov que

$$\overline{\mathbb{P}}(X'_t\in A)=\sum_{x_{t-1},y_{t-1}\in E}\Big[\overline{P}\left((x_{t-1},y_{t-1}),\,A\times E\right)\overline{\mathbb{P}}\left((X'_{t-1},Y'_{t-1})=(x_{t-1},y_{t-1})\right)\Big].$$

Portanto, indutivamente, temos

$$\begin{aligned}\overline{\mathbb{P}}(X'_t\in A)&=\sum_{x_0,y_0,x_1,y_1,\ldots,x_{t-1},y_{t-1}\in E}\Big[\lambda_1(x_0)\lambda_2(y_0)\overline{P}\left((x_0,y_0),(x_1,y_1)\right)\ldots\overline{P}\left((x_{t-1},y_{t-1}),A\times E\right)\Big]\\&=\sum_{x,y\in E}\lambda_1(x)\lambda_2(y)\overline{P}^t\left((x,y),\,A\times E\right).\end{aligned}$$

Assim, segue da definição da cadeia que

$$\overline{\mathbb{P}}(X'_t\in A)=\sum_{x,y\in E}\lambda_1(x)\lambda_2(y)P_1^t\left(x,A\right)=\sum_{x\in E}\lambda_1(x)P_1^t\left(x,A\right),$$

o que nos leva a concluir que

$$\overline{\mathbb{P}}(X'_t\in A)=(\lambda_1P_1^t)(A)=\mathbb{P}(X_t\in A).$$

Uma verificação semelhante mostra que $\overline{\mathbb{P}}(Y'_t\in A)=\mathbb{P}(Y_t\in A)$.
Portanto, segue da definição de acoplamentos que $(X'_t,Y'_t)_{t\in\mathbb{N}}$ unido a $\overline{\mathbb{P}}$ forma um acoplamento com a matriz de transição desejada.

Note que a definição fornece bastante liberdade para um acoplamento, e como é de se esperar, fixadas duas cadeias existem vários tipos de acoplamentos distintos. Abaixo exibiremos alguns exemplos de acoplamentos diferentes em cadeias em tempo discreto, e em seguida daremos exemplos para cadeias em tempo contínuo.

Neste primeiro exemplo, apresentaremos uma aplicação simples de acoplamentos.

Exemplo 3.1.6. Sejam $(X_t)_{t\in\mathbb{N}}$ e $(Y_t)_{t\in\mathbb{N}}$ duas cadeias de nascimento-morte em $\mathbb{N}$ independentes, com a mesma matriz de transição, e com distribuições iniciais δ_x e δ_y respectivamente, onde $x \geq y$.
Supondo que as probabilidades de subir uma unidade sejam iguais a p, e as de descer sejam iguais a q em qualquer um dos estados (exceto 0, nele a probabilidade de permanecer em 0 é $1-p$, fixado um n, ao fazer a pergunta: Qual das cadeias tem a maior probabilidade de estar acima de n em um fixado tempo s, é de certa forma intuitivo pensar que $(X_t)_{t\in\mathbb{N}}$ é a cadeia com a maior probabilidade. Abaixo utilizaremos um acoplamento para provar que de fato a verdade está do lado da intuição desta vez.
Utilizando a mesma notação do Exemplo 1.1.11, considere a matriz de transição H, dada por

$$H\left((a,b),(z,w)\right) = \begin{cases} p, & \text{se } (z,w) = (a+1,b+1); \\ q, & \text{se } (z,w) = (a-1,b-1); \\ P(a,z)P(b,w), & \text{caso contrário.} \end{cases}$$

A cadeia $(Z_t, W_t)_{t\in\mathbb{N}}$ gerada pela matriz de transição H, e distribuição inicial $\delta_x \times \delta_y$ unida a uma medida de probabilidade $\overline{\mathbb{P}}$ (em algum espaço de medida) é um acoplamento de $(X_t)_{t\in\mathbb{N}}$ e $(X_t)_{t\in\mathbb{N}}$.
Observe que como $X_0 \geq Y_0$, temos $X_t \geq Y_t$ para qualquer $t \in \mathbb{N}$. Portanto, pela definição de acoplamentos, temos

$$\mathbb{P}(X_t \geq n) = \sum_{k=n}^{\infty} \mathbb{P}(X_t = k) = \sum_{k=n}^{\infty} \overline{\mathbb{P}}(X_t = k) \geq \sum_{k=n}^{\infty} \overline{\mathbb{P}}(Y_t = k) = \sum_{k=n}^{\infty} \mathbb{P}(Y_t = k) = \mathbb{P}(Y_t \geq n).$$

No exemplo acima, é possível fazer a mesma conta para cadeias de nascimento-morte em $\{0, 1, \ldots, n\}$, porém isso exige alguns ajustes nas probabilidades de transição do estado n. Outra particularidade do exemplo acima, é o fato de que se ambas começam no mesmo estado, então elas andam juntas para sempre. É sempre possível definir um acoplamento entre duas cópias de uma mesma cadeia de modo que esse fenômeno aconteça.

O próximo exemplo talvez seja o mais simples dos acoplamentos, onde as coordenadas do acoplamento andam de maneira independente.

Exemplo 3.1.7 (Acoplamento independente)**.** Fixadas duas cadeias $(X_t)_{t\in\mathbb{N}}$ e $(Y_t)_{t\in\mathbb{N}}$ com espaço de estados E enumerável, chamamos de Acoplamento Independente a sequência de variáveis aleatórias $(X_t, Y_t)_{t\in\mathbb{N}}$ unidas à medida $\overline{\mathbb{P}}$ definida por

$$\overline{\mathbb{P}}((X_t, Y_t) = (x, y)) := \mathbb{P}(X_t = x)\mathbb{P}(Y_t = y),$$

para quaisquer $x, y \in \Omega$.

O exemplo abaixo ilustra um tipo de acoplamento entre duas cópias de uma mesma cadeia de modo que os valores de uma das cadeias sempre sejam maiores do que os da outra. Isto é, de certa forma o acoplamento preserva ordem.

Exemplo 3.1.8 (Acoplamento dominante)**.** Considere duas cadeias de Nascimento-Morte (Exemplo 1.1.11) $(X_t)_{t\in\mathbb{F}}$ e $(Y_t)_{t\in\mathbb{N}}$ independentes e iguais em distribuição. Suponha que se

tenha $p_x \leq 0.5$ e $q_x \leq 0.5$ para qualquer estado x. Adicionalmente, suponha que $p_x \geq p_y$ e $q_x \leq q_y$ sempre que $y \geq x$, isto é, quanto mais alto for o valor da cadeia em um dado instante, maior será a probabilidade de no instante seguinte a cadeia assumir um valor menor, e em contrapartida, quanto mais baixo for o valor de uma cadeia maior será a probabilidade de a cadeia assumir um valor maior no instante seguinte.
Chamamos de Acoplamento Dominante o acoplamento $(X_t', Y_t')_{t\in\mathbb{N}}$ com matriz de transição P dada por

$$P((x,y),(w,z)) \begin{cases} p_x, & \text{se } x \geq y \text{ e } (w,z) = (x+1, y+1); \\ p_y - p_x. & \text{se } x \geq y \text{ e } (w,z) = (x, y+1); \\ 1 - p_y - q_x, & \text{se } x \geq y \text{ e } (w,z) = (x,y); \\ q_x - q_y, & \text{se } x \geq y \text{ e } (w,z) = (x-1, y); \\ q_y, & \text{se } x \geq y \text{ e } (w,z) = (x-1, y-1); \\ p_y, & \text{se } x < y \text{ e } (w,z) = (x+1, y+1); \\ p_x - p_y. & \text{se } x < y \text{ e } (w,z) = (x+1, y); \\ 1 - p_x - q_y, & \text{se } x < y \text{ e } (w,z) = (x,y); \\ q_y - q_x, & \text{se } x < y \text{ e } (w,z) = (x, y-1); \\ q_x, & \text{se } x < y \text{ e } (w,z) = (x-1, y-1); \\ 0, & \text{caso contrário.} \end{cases}$$

Perceba que se $X_0(\omega) \geq Y_0(\omega)$, então $X_t(\omega) \geq Y_t(\omega)$ para qualquer $t \geq 0$. Semelhantemente, se $X_0(\omega) \leq Y_0(\omega)$, então $X_t(\omega) \leq Y_t(\omega)$ para qualquer $t \geq 0$. Isso justifica o nome do acoplamento. Ademais, se $X_t'(\omega) = Y_t'(\omega)$ para algum t, então $X_s'(\omega) = Y_s'(\omega)$ para qualquer $s \geq t$.

O exemplo abaixo consiste em uma mistura dos acoplamentos acima.

Exemplo 3.1.9. Nas hipóteses do exemplo anterior, considere a matriz de transição G definida por

$$G((x,y),(z,w)) = \begin{cases} p_x, & \text{se } x \geq y \text{ e } (w,z) = (x+1, y+1); \\ p_y - p_x. & \text{se } x \geq y \text{ e } (w,z) = (x, y+1); \\ 1 - p_y - q_x, & \text{se } x \geq y \text{ e } (w,z) = (x,y); \\ q_x - q_y, & \text{se } x \geq y \text{ e } (w,z) = (x-1, y); \\ q_y, & \text{se } x \geq y \text{ e } (w,z) = (x-1, y-1); \\ P(x,w)P(y,z), & \text{se } y > x; \\ 0, & \text{caso contrário.} \end{cases}$$

A dinâmica da cadeia gerada por G consiste no seguinte:
Se $X_0(\omega) \geq Y_0(\omega)$, então $X_t(\omega) \geq Y_t(\omega)$ para qualquer $t \geq 0$;
Se $X_0(\omega) < Y_0(\omega)$, então as cadeias se movem de maneira independente até que aconteça $X_s(\omega) \geq Y_s(\omega)$, e dali a cadeia segue $X_t(\omega) \geq Y_t(\omega)$ para qualquer $t \geq s$.
Ou seja, a cadeia gerada por G é uma mistura do acoplamento independente com o acoplamento dominante.

O próximo exemplo, válido apenas para cadeias lazy, exibe um acoplamento no qual somente uma das coordenas se move a cada instante de tempo.

Exemplo 3.1.10 (Mexe-Não-Mexe)**.** Fixadas duas matrizes de transição P_1 e P_2 em um espaço enumerável E, sejam $(X_t)_{t\in\mathbb{N}}$ e $(Y_t)_{t\in\mathbb{N}}$ de Markov em tempo discreto com matrizes de transição $(P_1+\mathrm{Id})/2$ e $(P_2+\mathrm{Id})/2$ respectivamente, ou seja, versões lazy de cadeias geradas por P_1 e P_2. Considere a matriz de transição M dada por

$$M((x,y),(w,z)) = \begin{cases} \dfrac{P_1(x,w)}{2}, & \text{se } z=y; \\ \dfrac{P_2(y,z)}{2}, & \text{se } x=w; \\ 0, & \text{caso contrário.} \end{cases}$$

Uma interpretação do acoplamento acima é a seguinte: A cada instante de tempo, jogamos uma moeda (honesta), e se o resultado for cara, a cadeia transiciona de (x,y) para (z,y) de a cordo com a matriz de transição P_1, e se o resultado for coroa, a cadeia transiciona de (x,y) para (z,y) de a cordo com a matriz de transição P_2.

Nas hipóteses do exemplo acima, é possível criar um acoplamento que mistura a Mexe-Não-Mexe com o acoplamento no qual as cadeias andam juntas. Para isso, é necessário impor a condição $P_1 = P_2$. O exemplo abaixo apresenta o processo.

Exemplo 3.1.11. Nas hipóteses do exemplo anterior, com a condição adicional $P_1 = P_2$, considere a matriz de transição Ç dada por

$$Ç\left((x,y),(w,z)\right) = \begin{cases} \dfrac{P_1(x,w)}{2}, & \text{se } z=y \text{ e } x\neq y; \\ \dfrac{P_1(y,z)}{2}, & \text{se } x=w \text{ e } x\neq y; \\ P_1(x,w), & \text{se } x=y \text{ e } w=z; \\ 0, & \text{caso contrário.} \end{cases}$$

A dinâmica de uma cadeia gerada por Ç consiste nas cadeias andarem seguindo o mexe-não-mexe até que as coordenadas se encontrem pela primeira vez. Após esse primeiro encontro as coordenadas seguem juntas seguindo a dinâmica de $P_1 = P_2$.

Nesta próxima parte trataremos de acoplamentos em tempo contínuo.
Assim como no caso em tempo discreto, podemos definir acoplamentos em tempo contínuo de uma maneira mais simples. A próxima proposição fornece uma maneira relativamente simples para obter acoplamentos para cadeias em tempo contínuo.

Proposição 3.1.12. *Sejam $(X_t)_{t\geq 0}$ e $(Y_t)_{t\geq 0}$ cadeias de Markov em tempo contínuo definidas em um espaço de probabilidade $(\Omega,\mathcal{F},\mathbb{P})$, com espaço de estados E finito, e com matrizes geradoras M e N respectivamente.*
Se Q é uma matriz geradora no espaço de estados $E\times E$, de modo que tenhamos

$$\sum_{b\in E} Q\left((x,y),(a,b)\right) = M(x,a) \quad e \quad \sum_{b\in E} Q\left((x,y),(b,a)\right) = N(y,a)$$

para quaisquer $x,y,a\in E$, então qualquer cadeia gerada por Q com condição inicial dada por um acoplamento de $\mathbb{P}(X_0=\cdot)$ e $\mathbb{P}(Y_0=\cdot)$ é um acoplamento de $(X_t)_{t\geq 0}$ e $(Y_t)_{t\geq 0}$.

Dem.: Seja $(X'_t, Y'_t)_{t\geq 0}$ uma cadeia gerada por Q de modo que sua condição inicial é um acoplamento de $\mathbb{P}(X_0 = \cdot)$ e $\mathbb{P}(Y_0 = \cdot)$.
Pelo Teorema 2.2.24 sabemos que

$$\overline{P}^t\left((x,y),(a,b)\right) = \mathrm{e}^{tQ}\left((x,y),(a,b)\right) = \sum_{k=0}^{\infty} \frac{t^k}{k!} Q^k\left((x,y),(a,b)\right). \tag{3.1}$$

Inicialmente, mostraremos que

$$\sum_{b\in E} Q^k\left((x,y),(a,b)\right) = M^k(x,a) \tag{3.2}$$

para quaisquer $a, x, y \in E$ e todo $k \in \mathbb{N}^*$.
Para $k = 1$ o resultado segue da hipótese.
Suponha que o resultado seja válido para qualquer $s \leq n \in \mathbb{N}^*$ tal que $s \leq n$. Neste caso, temos

$$\begin{aligned}\sum_{b\in E} Q^{n+1}\left((x,y),(a,b)\right) &= \sum_{b\in E}\sum_{z,w\in E} Q^n\left((x,y),(z,w)\right) Q\left((z,w),(a,b)\right)\\ &= \sum_{z,w\in E} Q^n\left((x,y),(z,w)\right) \sum_{b\in E} Q\left((z,w),(a,b)\right)\end{aligned}$$

como por hipótese o resultado é válido para qualquer $s \leq n \in \mathbb{N}^*$, temos

$$\begin{aligned}\sum_{b\in E} Q^{n+1}\left((x,y),(a,b)\right) &= \sum_{z,w\in E} Q^n\left((x,y),(z,w)\right) M(z,a)\\ &= \sum_{z\in E} M^n(x,z)M(z,a)\\ &= M^{n+1}(x,a).\end{aligned}$$

Portanto, segue do princípio de indução que (3.2) é válida para qualquer $k \in \mathbb{N}^*$.

Provada a igualdade, agora mostraremos que $\overline{\mathbb{P}}(X'_t = a) = \mathbb{P}(X_t = a)$ para qualquer $a \in E$.
Fixado $a \in E$ qualquer, temos:

$$\begin{aligned}\overline{\mathbb{P}}(X'_t = a) &= \sum_{b\in E}\sum_{x,y\in E} \overline{\mathbb{P}}(Z_t = (a,b) \mid Z_0 = (x,y))\overline{\mathbb{P}}(Z_0 = (x,y))\\ &= \sum_{b\in E}\sum_{x,y\in E} \overline{P}^t\left((x,y),(a,b)\right)\lambda(x,y)\end{aligned}$$

por (3.2), temos

$$\begin{aligned}\overline{\mathbb{P}}(X'_t = a) &= \sum_{b\in E}\sum_{x,y\in E}\sum_{k=0}^{\infty} \frac{t^k}{k!} Q^k\left((x,y),(a,b)\right)\lambda(x,y)\\ &= \sum_{x,y\in E}\sum_{k=0}^{\infty}\sum_{b\in E} \frac{t^k}{k!} Q^k\left((x,y),(a,b)\right)\lambda(x,y)\end{aligned}$$

de onde segue de (3.1) que

$$\begin{aligned}
\overline{\mathbb{P}}(X_t' = a) &= \sum_{x,y\in E} \sum_{k=0}^{\infty} \frac{t^k}{k!} M^k(x,a)\lambda(x,y) \\
&= \sum_{x\in E} \sum_{k=0}^{\infty} \frac{t^k}{k!} M^k(x,a)\mathbb{P}(X_0 = x) \\
&= \sum_{x\in E} \mathbb{P}(X_0 = x) \sum_{k=0}^{\infty} \frac{t^k}{k!} M^k(x,a) \\
&= \sum_{x\in E} \mathbb{P}(X_0 = x) P^t(x,a) \\
&= \mathbb{P}(X_t = a).
\end{aligned}$$

Um desenvolvimento análogo mostra que $\overline{\mathbb{P}}(Y_t' = a) = \mathbb{P}(Y_t = a)$ para qualquer $a \in E$, portanto, concluímos que de fato $(Z_t)_{t\geq 0}$ é um acoplamento de $(X_t)_{t\geq 0}$ e $(Y_t)_{t\geq 0}$. ■

Nas hipóteses da proposição anterior, note que resgatando a Definição 2.2.25 podemos definir um acoplamento entre duas cadeias de Markov em tempo contínuo por um gerador L tal que

$$Lf(x,y) = \sum_{(a,b)\in E\times E} \alpha_{(x,y),(a,b)} \left[f(a,b) - f(x,y) \right].$$

Para isso, basta exigir que se tenha

$$\sum_{b\in E} \alpha_{(x,y),(a,b)} = M(x,a) \quad \text{e} \quad \sum_{b\in E} \alpha_{(x,y),(b,a)} = N(x,a).$$

para quaisquer $a, x, y \in E$.

Semelhantemente à acoplamentos de cadeias em tempo discreto, para cadeias em tempo contínuo também existem vários acoplamentos diferentes. Exibiremos alguns deles nos próximos exemplos.

Exemplo 3.1.13 (Acoplamento independente)**.** Sejam $(X_t)_{t\geq 0}$ e $(Y_t)_{t\geq 0}$ cadeias de Markov em tempo contínuo definidas em um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$, com espaço de estados E finito, e com matrizes geradoras M e N respectivamente. Definimos o acoplamento independente de $(X_t)_{t\geq 0}$ e $(Y_t)_{t\geq 0}$ como sendo a variável aleatória gerada por L, onde

$$Lf(x,y) = \sum_{z,w\in E} m(z)n(w) \left[f(z,w) - f(x,y) \right]$$

para qualquer função $f : E \times E \to \mathbb{R}$, onde $m(u) = -M(u,u)$ e $n(u) = -n(u,u)$ para qualquer $u \in E$.
A ideia do acoplamento é manter as coordenadas se movendo de acordo com $(X_t)_{t\geq 0}$ e $(X_t)_{t\geq 0}$ de maneira independente (assim como no caso em tempo discreto).

Exemplo 3.1.14 (Acoplamento dominante)**.** Sejam $(X_t)_{t\geq 0}$ e $(Y_t)_{t\geq 0}$ cadeias de Nascimento-Morte definidas no espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$, com espaço de estados $\{0, 1, \ldots, n\}$. Suponha que ambas sejam sejam advindas da matriz geradora M, onde

$$M(x,y) = \begin{cases} p_x, & \text{se } y = x+1; \\ q_x, & \text{se } y = x-1; \\ -q_x - p_x, & \text{se } y = x; \\ 0, & \text{se } y \notin \{x+1, x, x-1\}. \end{cases}$$

Suponha adicionalmente que $p_x \geq p_y$ e $q_y \geq q_x$ sempre que $x \leq y$, e que $q_n, p_0, p_x, q_x > 0$ para qualquer $x \in \{1, \ldots, n-1\}$.

Definamos a matriz geradora Q em $E \times E$ por

$$Q\left((x,y),(z,w)\right) \begin{cases} -q_x - p_y, & \text{se } x \geq y \text{ e } (x,y) = (z,w) \\ p_x\,, & \text{se } x \geq y \text{ e } (z,w) = (x+1, y+1); \\ p_y - p_x\,, & \text{se } x \geq y \text{ e } (z,w) = (x, y+1); \\ q_x - q_y\,, & \text{se } x \geq y \text{ e } (z,w) = (x-1, y); \\ q_y\,, & \text{se } x \geq y \text{ e } (z,w) = (x-1, y-1); \\ p_y\,, & \text{se } x < y \text{ e } (z,w) = (x+1, y+1); \\ p_x - p_y\,, & \text{se } x < y \text{ e } (z,w) = (x+1, y); \\ q_y - q_x\,, & \text{se } x < y \text{ e } (z,w) = (x, y-1); \\ q_x\,, & \text{se } x < y \text{ e } (z,w) = (x-1, y-1); \\ -q_y - p_x, & \text{se } x < y \text{ e } (x,y) = (z,w); \\ 0\,, & \text{caso contrário.} \end{cases}$$

É facilmente verificável que Q satisfaz as hipóteses do teorema, portanto, tomando um acoplamento das distribuições iniciais de $(X_t)_{t\geq 0}$ e $(Y_t)_{t\geq 0}$, juntamente com Q gera-se um acoplamento $((X'_t, Y'_t)_{t\geq 0}$ de $(X_t)_{t\geq 0}$ e $(Y_t)_{t\geq 0}$ com a particularidade de que se $X'_s(\omega) = Y_s(\omega)$ para algum $s \geq 0$, então $X'_t(\omega) = Y_t(\omega)$ para todo $t \geq s$. Isso justifica o nome do acoplamento.

Observação 3.1.15. No caso do acoplamento acima, um fenômeno interessante acontece, tem se $\overline{\mathbb{P}}(\tau_{(x,x)} < \infty) = 1$ para qualquer $x \in E$, onde $\tau_{(x,x)}$ é o tempo de chegada ao estado (x,x), definido por

$$\tau_{(x,x)}(\omega) = \min\left\{t \geq 0 \;:\; (X'_t, Y'_t)_{t\geq 0} = (x,x)\right\}.$$

Para isso, basta a condição de que p_y e q_x sejam sempre positivos (com exceção de q_0 e p_n). De fato, note que pela definição da cadeia e o fato de o espaço de estados ser finito, algum estado (a,b) com $a \geq b$ é recorrente. Como $(a,b) \to (y,y)$ e (y,y) não leva a (a,b) se $a > b$, segue que nenhum par (a,b) com $a > b$ é recorrente. Portanto, existe algum estado $(y,y) \in E \times E$ recorrente.

Note que $(x,x) \leftrightarrow (y,y)$ para qualquer $X \in E$, assim, (x,x) e (y,y) pertencem à mesma classe comunicadora, e consequentemente (x,x) também é recorrente. Portanto, repetindo a demonstração de Proposição 1.1.51 (levando em consideração a distribuição inicial do acoplamento), concluímos que de fato para qualquer $x \in E$ tem-se

$$\overline{\mathbb{P}}(\tau_{(x,x)} < \infty) = 1.$$

O próximo teorema relaciona distribuições com os seus possíveis acoplamentos.

Teorema 3.1.16 (Teorema da Dualidade)**.** *Sejam μ e ν distribuições de probabilidade definidas em $(\Omega, \mathcal{F})$. Se Ω é enumerável, então*

$$\mathrm{d_{TV}}(\mu,\nu) \leq \inf\left\{\mathbb{P}\left\{X \neq Y\right\} : (X,Y) \textit{ unido a } \mathbb{P} \textit{ é um acoplamento de } \mu \textit{ e } \nu\right\}.$$

Dem.: Fixado um evento A, e um acoplamento qualquer como no enunciado, note que

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

Repetindo o processo com $\nu(A) - \mu(A)$ obtemos

$$|\mu(A) - \nu(A)| \leq \mathbb{P}(X \neq Y),$$

e consequentemente

$$\mathrm{d}_{TV}(\mu,\nu) = \sup_{A\in\mathcal{F}}\left\{|\mu(A) - \nu(A)|\right\} \leq \mathbb{P}(X \neq Y).$$

Como a desigualdade acima vale para um acoplamento qualquer, concluímos que de fato

$$\mathrm{d}_{TV}(\mu,\nu) \leq \inf\left\{\mathbb{P}\left\{X \neq Y\right\} : (X,Y) \text{ unido a } \mathbb{P} \text{ é um acoplamento de } \mu \text{ e } \nu\right\}.$$

■

Observação 3.1.17. Em [LPW] é demonstrada uma versão mais forte do resultado acima, onde além de termos a desigualdade, existe um certo acoplamento que alcança a igualdade.

Neste próximo resultado obteremos uma cota superior para a distância de variação total da cadeia, e para isso, consideraremos apenas acoplamentos (Z_t, W_t) (tempo discreto ou contínuo) que satisfaçam a condição:

$$Z_t(\omega) = W_t(\omega) \implies Z_s(\omega) = W_s(\omega) \text{ para qualquer } s \geq t. \tag{3.3}$$

Teorema 3.1.18. *Sejam $(X_t)_{t\in\mathbb{F}}$ e $(Y_t)_{t\in\mathbb{F}}$ (com $\mathbb{F} = \mathbb{R}_{\geq 0}$ ou $\mathbb{N}$) duas cadeias definidas em um mesmo espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ com espaço de estados E finito, matriz/semigrupo de transição P, e distribuições iniciais δ_x e δ_y respectivamente para $x, y \in E$ fixados.*
Suponha que exista um acoplamento $(Z_t, W_t)_{t\in\mathbb{F}}$ em um espaço $(\Omega', \mathcal{F}', \overline{\mathbb{P}})$ tal que a condição de Equação (3.3) seja satisfeita. Se τ_{ac} é um tempo de parada definido por

$$\tau_{ac}(\omega) := \inf\left\{t \;:\; Z_t(\omega) = W_t(\omega)\right\},$$

então

$$\mathrm{d}_{TV}\left(P^t(x,\cdot), P^t(y,\cdot)\right) \leq \overline{\mathbb{P}}(\tau_{ac} > t).$$

Dem.: Observe que pelo Teorema 3.1.16, temos

$$\mathrm{d}_{TV}\left(P^t(x,\cdot), P^t(y,\cdot)\right) \leq \overline{\mathbb{P}}(W_t \neq Z_t),$$

e como o acoplamento satisfaz (3.3), segue que

$$\overline{\mathbb{P}}(W_t \neq Z_t) = \overline{\mathbb{P}}(\tau_{ac} > t).$$

Portanto, concluímos que

$$\mathrm{d}_{TV}\left(P^t(x,\cdot), P^t(y,\cdot)\right) \leq \overline{\mathbb{P}}(\tau_{ac} > t).$$

■

Observação 3.1.19. Chamamos os tempos τ_{ac} (ou $\tau_{x,y}$ nos exemplos anteriores) de *tempos de acoplamento.*

3.2 Tempos de mistura

Todas as definições aqui listadas valem tanto para cadeias em tempo contínuo quanto discreto a menos que explicitado o contrário.

Fixada um cadeia de Markov em tempo discreto ou contínuo $(X_t)_t$ com espaço de estados E finito, matriz/semigrupo de transição P e distribuição invariante π, denotaremos por $d(t)$ o número

$$d(t) := \max_{x \in E} \left\{ \mathrm{d}_{TV}(P^t(x,\cdot), \pi(\cdot)) \right\},$$

ou seja, a maior das distâncias para a medida invariante no tempo t dentre todos os estados iniciais possíveis. De maneira semelhante, denotaremos por $\overline{d}(t)$ o número

$$\overline{d}(t) := \max_{x,y \in E} \left\{ \mathrm{d}_{TV}(P^t(x,\cdot), P^t(y,\cdot)) \right\}.$$

No restante do capítulo sempre que utilizarmos $d(t)$ e $\overline{d}(t)$, será em relação à cadeia do enunciado, proposição, entre outros.

Proposição 3.2.1. *Se (X_t) é uma cadeia de Markov (em tempo contínuo ou discreto) com espaço de estados I finito, semi-grupo de transição (P^t) e medida invariante π então*

$$d(t) = \sup_{\mu} \left\{ \mathrm{d}_{\mathrm{TV}}(\mu P^t, \pi) \right\}$$

onde μ varia entre todas as distribuições de probabilidade definidas em I.

Dem.: Fixado um estado x, observe que $P^t(x,\cdot) = \delta_x P^t$. Portanto, pela definição de $d(t)$, temos

$$d(t) \leq \sup_{\mu} \left\{ \mathrm{d}_{\mathrm{TV}}(\mu P^t, \pi) \right\}.$$

Por outro lado, suponha que $d(t) = \mathrm{d}_{TV}(P^t(x,\cdot),\pi(\cdot))$. Fixada uma distribuição μ qualquer, temos

$$\begin{aligned}
\mathrm{d}_{TV}(\mu P^t,\pi) &= \frac{1}{2}\sum_{y\in I}\left|(\mu P^t)(y)-\pi(y)\right| \\
&= \frac{1}{2}\sum_{y\in I}\left|\sum_{z\in I}\left[\mu(z)P^t(z,y)\right]-\pi(y)\right| \\
&= \frac{1}{2}\sum_{y\in I}\left|\sum_{z\in I}\left[P^t(z,y)-\pi(y)\right]\mu(z)\right| \\
&\leq \frac{1}{2}\sum_{y\in I}\sum_{z\in I}\left(\mu(z)\left|P^t(z,y)-\pi(y)\right|\right) \\
&= \sum_{z\in I}\Big(\mu(z)\,\mathrm{d}_{TV}(P^t(z,\cdot),\pi(\cdot))\Big) \\
&\leq \mathrm{d}_{TV}(P^t(x,\cdot),\pi(\cdot)).
\end{aligned}$$

■

Observação 3.2.2. Note que o resultado acima revela que a "pior" é atingida escolhendo-se começar no "pior" dos estados iniciais, o que é de certa forma bastante intuitivo.

Observação 3.2.3. Fixado um conjunto A, note que

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

E portanto

$$d(t) = \max_{x\in E}\left\{\mathrm{d}_{TV}(P^t(x,\cdot),\pi(\cdot))\right\} \leq \bar{d}(t).$$

Proposição 3.2.4. *Seja $(X_t)_t$ uma cadeia de Markov (tempo discreto ou contínuo) com espaço de estados E enumerável, matriz/semigrupo de transição P, distribuição inicial μ e medida invariante π. A distância de variação total entre a distribuição da cadeia e a sua medida invariante é decrescente conforme o tempo avança, isto é, para todo $t \geq s$ tem-se*

$$\mathrm{d}_{TV}(\mu P^s(\cdot),\pi(\cdot)) \leq \mathrm{d}_{TV}(\mu P^t(\cdot),\pi(\cdot)).$$

Em particular,

$$\mathrm{d}_{TV}(P^s(x,\cdot),\pi(\cdot)) \leq \mathrm{d}_{TV}(P^t(x,\cdot),\pi(\cdot))$$

para qualquer $x \in E$.

Dem.: Fixados $t \geq s$, pela Proposição 0.2.7, temos

$$\begin{aligned}
\mathrm{d}_{TV}(\mu P^t(\cdot), \pi(\cdot)) &= \frac{1}{2}\sum_{y\in E}\left|(\mu P^t)(y) - \pi(y)\right| \\
&= \frac{1}{2}\sum_{y\in E}\left|\sum_{x\in E}\left[\mu(x)P^t(x,y) - \pi(y)\right]\right| \\
&= \frac{1}{2}\sum_{y\in E}\left|\sum_{x\in E}\left[\mu(x)\left(P^t(x,y) - \pi(y)\right)\right]\right| \\
&= \frac{1}{2}\sum_{y\in E}\left|\sum_{x\in E}\left[\mu(x)\left(P^sP^{t-s}(x,y) - \pi P^{t-s}(y)\right)\right]\right| \\
&= \frac{1}{2}\sum_{y\in E}\left|\sum_{x\in E}\left(\mu(x)\sum_{z\in E}\left[P^sP^{t-s}(x,y) - \pi P^{t-s}(y)\right]\right)\right| \\
&= \frac{1}{2}\sum_{y\in E}\left|\sum_{x\in E}\mu(x)\sum_{z\in E}\left[P^{t-s}(z,y)\left(P^s(x,z) - \pi(z)\right)\right]\right| \\
&\leq \frac{1}{2}\sum_{z\in E}\sum_{y\in E}P^{t-s}(z,y)\left|\sum_{x\in E}\left[\mu(x)\left(P^s(x,z) - \pi(z)\right]\right)\right| \\
&\leq \frac{1}{2}\sum_{z\in E}\left|\sum_{x\in E}\left[\mu(x)\left(P^s(x,z) - \pi(z)\right)\right]\right| \\
&\leq \frac{1}{2}\sum_{z\in E}\left|\mu P^s(z) - \pi(z)\right| = \mathrm{d}_{TV}(\mu P^s(\cdot), \pi(\cdot))
\end{aligned}$$

O que demonstra a primeira parte do teorema.
Note que a segunda parte segue do caso em que a distribuição inicial é δ_x na demonstração acima. ■

O objeto que definiremos a seguir é fundamental no estudo da convergência de cadeias de Markov. Ele servirá para medir o tempo necessário para que a distância de variação total entre uma cadeia e sua distribuição invariante estejam ε-próximas

Definição 3.2.5 (Tempos de mistura)**.** Fixada uma cadeia de Markov (em tempo contínuo ou discreto) $(X_t)_t$, definimos os tempos de ε-mistura $t_{mix}(\varepsilon)$ por

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

Particularmente, definimos

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

Embora pareça arbitrário, a escolha de $t_{mix} := t_{mix}(1/4)$ fornece a desigualdade $t_{mix}(\varepsilon) \leq \log(\varepsilon^{-1})$. Mais detalhes podem ser encontrados em [LPW].

Exemplo 3.2.6. Seja $(X_t)_{t\geq 0}$ uma cadeia de Markov em tempo contínuo com espaço de estados I finito e matriz geradora Q. Suponha que π seja medida invariante de $(X_t)_{t\geq 0}$.
Se $(Y_t)_{t\geq 0}$ é a cadeia com matriz geradora αQ ($\alpha > 0$), com matriz de transição $\overline{P}^t$, então

$$\mathrm{d_{TV}}\left(P^t(x,\cdot)\,,\,\pi\right) = \mathrm{d_{TV}}\left(\mathrm{e}^{tQ}(x,\cdot)\,,\,\pi\right) = \mathrm{d_{TV}}\left(\mathrm{e}^{t/\alpha(\alpha Q)}(x,\cdot)\,,\,\pi\right) = \mathrm{d_{TV}}\left(\overline{P}^{t/\alpha}(x,\cdot),\pi\right).$$

Particularmente, vale

$$t_{mix}(\varepsilon) = \bar{t}_{mix}\left(\frac{\varepsilon}{\alpha}\right)$$

para qualquer $\varepsilon > 0$, onde $\bar{t}_{mix}(s)$ são os tempos de mistura de $(Y_t)_{t\geq 0}$.

Portando, gerar uma cadeia através de um múltiplo de uma matriz geradora acelera (ou retarda) o processo por um valor inversamente proporcional à constante.

Para o próximo resultado, utilizaremos a variável aleatória $\tau_{x,y}$ definida por

$$\tau_{x,y} := \min\{t \;:\; X_t = Y_t\},$$

dado que $X_0 = x$ e $Y_0 = y$.

Proposição 3.2.7. *Sejam $(X_t)_t$ e $(X_t)_t$ duas cadeias de Markov em tempo discreto ou contínuo definidas no mesmo espaço, com a mesma matriz de transição (ou matriz geradora no caso contínuo), espaço de estados E finito.*
Se para cada par $x,y \in E$, $(Z_t^{x,y})_t$ unido a $\overline{\mathbb{P}}_{x,y}$ é um acoplamento com condição inicial $\delta_{(x,y)}$, de maneira que a condição descrita em (3.3) (Após coincidirem pela primeira vez, as cadeias permanecem juntas) seja satisfeita, então

$$\mathrm{d}(t) \leq \max_{x,y\in E}\overline{\mathbb{P}}_{x,y}\{\tau_{x,y} > t\}.$$

Particularmente,

$$t_{mix} \leq 4\max_{x,y\in E}\left\{\mathbb{E}_{\overline{\mathbb{P}}_{x,y}}\left[\tau_{x,y}\right]\right\}.$$

Dem.: A primeira parte vem do Teorema 3.1.18 juntamente com a Observação 3.2.3. Usando a Desigualdade de Markov (Teorema A.1.5), temos

$$d(t) \leq \max_{x,y\in E}\overline{\mathbb{P}}_{x,y}\{\tau_{x,y} > t\} \leq \frac{\max_{x,y\in E}\left\{\mathbb{E}_{\overline{\mathbb{P}}_{x,y}}\left[\tau_{x,y}\right]\right\}}{t}.$$

Tomando $t = 4\max_{x,y\in E}\left\{\mathbb{E}_{\overline{\mathbb{P}}_{x,y}}\left[\tau_{x,y}\right]\right\}$, obtemos

$$d\left(4\max_{x,y\in E}\left\{\mathbb{E}_{\overline{\mathbb{P}}_{x,y}}\left[\tau_{x,y}\right]\right\}\right) \leq \frac{\max_{x,y\in E}\left\{\mathbb{E}_{\overline{\mathbb{P}}_{x,y}}\left[\tau_{x,y}\right]\right\}}{4\max_{x,y\in E}\left\{\mathbb{E}_{\overline{\mathbb{P}}_{x,y}}\left[\tau_{x,y}\right]\right\}} = \frac{1}{4},$$

e portanto

$$t_{mix} \leq 4\max_{x,y\in E}\left\{\mathbb{E}_{\overline{\mathbb{P}}_{x,y}}\left[\tau_{x,y}\right]\right\}.$$

■

Abaixo, apresentaremos um exemplo que fará uso de um acoplamento para estimar o tempo de mistura de uma cadeia usando a proposição acima.

Exemplo 3.2.8 (Cota para o passeio aleatório em $\mathbb{Z}_n$)**.** Fixados $p \in (0,1)$ e $q := 1-p$, considere o passeio aleatório lazy no círculo $\mathbb{Z}_n$ com viés $p-q$, isto é, a cadeia $(X_t)_{t\in\mathbb{N}}$ com matriz de transição P dada por

$$P(x,y) = \begin{cases} \dfrac{p}{2}, & \text{se } y = x+1; \\ \dfrac{q}{2}, & \text{se } y = x-1; \\ \dfrac{1}{2}, & \text{se } y = x; \\ 0, & \text{caso contrário.} \end{cases}$$

Uma simples verificação das condições de equilíbrio (Definição 1.1.30), mostra que a distribuição invariante é a uniforme.
Considere a matriz de transição Q dada por

$$Q((x,y),(a,b)) = \begin{cases} \dfrac{p}{2}, & \text{se } (a,b) = (x+1,y) \text{ e } x \neq y; \\ \dfrac{q}{2}, & \text{se } (a,b) = (x-1,y) \text{ e } x \neq y; \\ \dfrac{p}{2}, & \text{se } (a,b) = (x,y+1) \text{ e } x \neq y; \\ \dfrac{q}{2}, & \text{se } (a,b) = (x,y-1) \text{ e } x \neq y; \\ \dfrac{p}{2}, & \text{se } (a,b) = (x+1,y+1) \text{ e } x = y; \\ \dfrac{q}{2}, & \text{se } (a,b) = (x-1,y-1) \text{ e } x = y; \\ \dfrac{1}{2}, & \text{se } (a,b) = (x,y) \text{ e } x = y; \\ 0,, & \text{caso contrário.} \end{cases}$$

Uma heurística para a uma cadeia com matriz de transição Q consiste no seguinte: Quando as coordenadas estão diferentes, sorteamos uma coordenada e escolhida a coordenada com probabilidade p somamos 1 e com probabilidade q somamos -1 à coordenada sorteada. Já quando as coordenas estão juntas, elas permanecem juntas seguindo as probabilidades de P.
Para cada par $(x,y) \in \mathbb{Z}_n \times \mathbb{Z}_n$, considere o acoplamento $(Z^{x,y}, W^{x,y})_{t\in\mathbb{N}}$ com matriz de transição Q, distribuição inicial $\delta_{(x,y)}$ num espaço com medida $\mathbb{P}_{x,y}$.
Note que se definirmos para cada par (x,y) a variável aleatória $D_t^{x,y}$ como sendo a "distância anti-horária" de x, y (quantidade de arestas no caminho anti-horário de x até y) , então D_t é o problema da Ruína do apostador. Ademais, note que $\inf\{t : D_t^{x,y} \in \{0,n\}\} = \tau^{x,y}$, onde $\tau^{x,y}$ é o tempo de acoplamento de $(Z^{x,y}, W^{x,y})_{t\in\mathbb{N}}$.
Portanto, pelo Exemplo 1.1.22 (Ruína do Apostador) , concluímos que

$$\mathbb{E}_{\mathbb{P}_{x,y}}(\tau_{x,y}) = k(n-k),$$

se a distância anti-horária de x e y é igual a k. Particularmente,

$$\mathbb{E}_{\mathbb{P}_{x,y}}(\tau_{x,y}) \leq \frac{n^2}{4},$$

para quaisquer $x, y \in \mathbb{Z}_n$.

Assim, utilizando o Teorema 3.1.18 e a Desigualdade de Markov concluímos que

$$d(t) \leq \max_{x,y\in\mathbb{Z}_n} \mathbb{P}_{x,y}(\tau_{x,y} > t) \leq \frac{\max_{x,y\in\mathbb{Z}_n} \mathbb{E}_{\mathbb{P}_{x,y}}[\tau_{x,y}]}{t} \leq \frac{n^2}{4t}.$$

Exemplo 3.2.9 (Cota para o passeio lazy no toro discreto)**.** Considere a versão lazy do passeio aleatório no toro discreto d-dimensional (Exemplo 1.1.15).

Considere a matriz de transição S dada por

$$S\left((x,y),(z,w)\right) = \begin{cases} \dfrac{1}{4d}, \text{ se } x_j = y_j \text{ para algum } j \text{ e } (a,b) = (x,y) \pm (e_j, e_j); \\ \dfrac{1}{4d}, \text{ se } x_j \neq y_j \text{ para todo } j \text{ e } (a,b) = (x,y) \pm (0, e_j); \\ \dfrac{1}{4d}, \text{ se } x_j \neq y_j \text{ para todo } j \text{ e } (a,b) = (x,y) \pm (e_j, 0); \\ \dfrac{1}{2}, \text{ se } (x,y) = (a,b); \\ 0, \text{caso contrário.} \end{cases}$$

Note que a cadeia é irredutível, e por ser finita possui um estado recorrente. Assim, todos os seus estados são recorrentes, e pela Proposição 1.1.51, segue que os tempos $\tau_{x,y}$ são finitos com probabilidade 1.

Assim, fixado um par x, y, defina para cada $i \in \{1, 2, \ldots, d\}$ a variável aleatória D_i por

$$D_i(\omega) := \inf\left\{t \in \mathbb{N} \ : \ \pi_i(Z_t^{x,y}(\omega)) = \pi_i((W_t^{x,y}(\omega))\right\},$$

onde π_i é a projeção na i-ésima coordenada. Pelo comentário anterior, sabemos que cada um dos D_i é finito com probabilidade 1.

Observe que cada uma das cadeias D_i é um passeio aleatório em $\mathbb{Z}_n$ com a particularidade de que ao invés de realizar um movimento por unidade de tempo, tem-se um movimento a cada d unidades de tempo (em média).

Portanto, se $X_1, X_2, \ldots X_n$ são variáveis i.i.d. com distribuição comum $\mathrm{Ber}\left(1/d\right)$ e independentes das variáveis D_i (sem perda de generalidade podemos supor que estão definidas no mesmo espaço em que as variáveis D_i), então segue do exemplo anterior que

$$\mathbb{E}\left[\sum_{n=1}^{D_i} X_n\right] \leq \frac{n^2}{4}.$$

Note que

$$\mathbb{E}\left[\sum_{n=1}^{D_i} X_n\right] = \mathbb{E}\left[\sum_{n=1}^{\infty} X_n \mathbf{1}_{D_i \geq n}\right] = \sum_{n=1}^{\infty} \mathbb{E}\left[X_n \mathbf{1}_{D_i \geq n}\right] = \sum_{n=1}^{\infty} \left(\mathbb{E}\left[X_n\right] \mathbb{P}(D_i \geq n)\right) = \mathbb{E}\left[X_n\right] \mathbb{E}\left[D_i\right],$$

e portanto para cada i, temos

$$\mathbb{E}\left[D_i\right] \leq \frac{n^2 d}{4}.$$

Observe que

$$\mathbb{E}\left[\tau_{x,y}\right] \leq \sum_{i=1}^{d} \mathbb{E}\left[D_i\right] \leq \frac{n^2 d^2}{4}.$$

Com isso, segue do Teorema 3.1.18 e da Desigualdade de Markov que

$$d(t) \leq \max_{x,y \in \mathbb{Z}_n^d} \left\{\mathbb{P}_{x,y}(\tau_{x,y} > t)\right\} \leq \frac{\max_{x,y \in \mathbb{Z}_n^d} \left\{\mathbb{E}_{\mathbb{P}_{x,y}}\left[\tau_{x,y}\right]\right\}}{t} \leq \frac{n^2 d^2}{4t}.$$

3.3 Cutoff, Pré-Cutoff e Janela de Cutoff

Definição 3.3.1 (Cutoff)**.** Seja $((X_t^n)_{t \in \mathbb{F}}\ ;\ n \in \mathbb{N})$ uma sequência de cadeias de Markov com $\mathbb{F} = \mathbb{N}$ ou $\mathbb{R}_{\geq 0}$. Denotando para cada $n \in \mathbb{N}$ os tempos de mistura da cadeia $(X_t^n)_t$ por $t_{mix}^n(\varepsilon)$, dizemos que essa sequência de cadeias apresenta *cutoff* se para qualquer $\varepsilon \in (0,1)$ tem se

$$\lim_{n \to \infty} \frac{t_{mix}^n(\varepsilon)}{t_{mix}^n(1-\varepsilon)} = 1. \tag{3.4}$$

Uma vaga interpretação do fenômeno acima é que para n suficientemente grande, os tempos que as cadeias gastam para ficar ε-próximas das suas medidas invariantes diferem pouco dos tempos que as cadeias gastam para ficar $(1-\varepsilon)$-próximas, ou seja, o intervalo entre estes dois eventos é bem curto (proporcionalmente ao tempo que o primeiro deles leva para acontecer). A próxima proposição fornece uma interpretação mais clara acerca do fenômeno.
Para o restante do capítulo, denotaremos por $d_n(t)$ a distância entre a n-ésima cadeia e sua medida invariante no tempo t.

Proposição 3.3.2. *Suponha que a sequência $((X_t^n)_{t \in \mathbb{F}}\ ;\ n \in \mathbb{N})$ de cadeias de Markov é tal que para todo $n \in \mathbb{N}$ os tempos de mistura da cadeia (X_t^n) são $t_{mix}^n(\varepsilon)$. A sequência apresenta cutoff se e somente se*

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

Dem.: Suponha que $((X_t^n)_{t \in \mathbb{F}}\ ;\ n \in \mathbb{N})$ apresenta cutoff. Note que para $0 < \varepsilon \leq 1/4$, vale

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

o que implica

$$1 = \lim_{n\to\infty} \frac{t^n_{mix}(\varepsilon)}{t^n_{mix}(1-\varepsilon)} \geq \lim_{n\to\infty} \frac{t^n_{mix}(\varepsilon)}{t^n_{mix}} \geq 1 \quad \text{e} \quad 1 \geq \lim_{n\to\infty} \frac{t^n_{mix}(1-\varepsilon)}{t^n_{mix}} \geq \lim_{n\to\infty} \frac{t^n_{mix}(1-\varepsilon)}{t^n_{mix}(\varepsilon)} = 1$$

e portanto

$$\lim_{n\to\infty} \frac{t^n_{mix}(1-\varepsilon)}{t^n_{mix}} = \lim_{n\to\infty} \frac{t^n_{mix}(\varepsilon)}{t^n_{mix}} = 1, \tag{3.5}$$

para qualquer $\varepsilon \leq 1/4$.
Assim, fixado $c > 1$, existe $n_0 \in \mathbb{N}$ tal que para qualquer $n > n_0$ tem-se

$$t^n_{mix}(\varepsilon) \leq c t^n_{mix} \qquad \text{e} \qquad t^n_{mix}(1-\varepsilon) \geq \frac{1}{c} t^n_{mix},$$

o que implica

$$\varepsilon \leq d_n(c t^n_{mix}) \qquad \text{e} \qquad 1 - \varepsilon \geq d_n\left(\frac{1}{c} t^n_{mix}\right).$$

Portanto,

$$\varepsilon \geq \lim_{n\to\infty} d_n(c t^n_{mix}) \qquad \text{e} \qquad 1 - \varepsilon \leq \lim_{n\to\infty} d_n\left(\frac{1}{c} t^n_{mix}\right),$$

para qualquer $\varepsilon \leq 1/4$, o que nos leva a concluir que

$$\lim_{n\to\infty} d_n(c t^n_{mix}) = 0 \qquad \text{e} \qquad \lim_{n\to\infty} d_n\left(\frac{1}{c} t^n_{mix}\right) = 1.$$

Recíprocamente, suponha que

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

Fixados $c > 1$ e $\varepsilon \in (0,1)$ quaisquer, existe $n_0 \in \mathbb{N}$ tal que para qualquer $n \geq n_0$ tem se

$$d_n(c t^n_{mix}) \leq \min\{\varepsilon, 1-\varepsilon\} \quad \text{e} \quad d_n\left(\frac{1}{c} t^n_{mix}\right) \geq \max\{\varepsilon, 1-\varepsilon\},$$

o que implica

$$c t^n_{mix} \geq \max\{t^n_{mix}(\varepsilon), t^n_{mix}(1-\varepsilon)\} \quad \text{e} \quad \frac{1}{c} t^n_{mix} \leq \min\{t^n_{mix}(\varepsilon), t^n_{mix}(1-\varepsilon)\}.$$

Portanto, concluímos que

$$\frac{c t^n_{mix}}{c^{-1} t^n_{mix}} \geq \frac{t^n_{mix}(\varepsilon)}{t^n_{mix}(1-\varepsilon)} \geq \frac{c^{-1} t^n_{mix}}{c t^n_{mix}}$$

o que garante que

$$c^2 \geq \lim_{n\to\infty} \frac{t^n_{mix}(\varepsilon)}{t^n_{mix}(1-\varepsilon)} \geq \frac{1}{c^2}.$$

Como a desigualdade acima vale para qualquer $c > 1$, fazendo $c \downarrow 1$ concluímos que

$$\lim_{n\to\infty} \frac{t^n_{mix}(\varepsilon)}{t^n_{mix}(1-\varepsilon)} = 1$$

■

A proposição acima pode ser interpretada da seguinte forma: Conforme n cresce, na escala de t^n_{mix}, d_n se aproxima de uma função degrau, ou seja, o decrescimento acontece de forma abrupta.

A Figura 3.1 ilustra o comportamento típico de uma sequência de cadeias que apresenta cutoff. Interprete a Figura 3.1 como sendo as mudanças no comportamento à medida que n cresce.

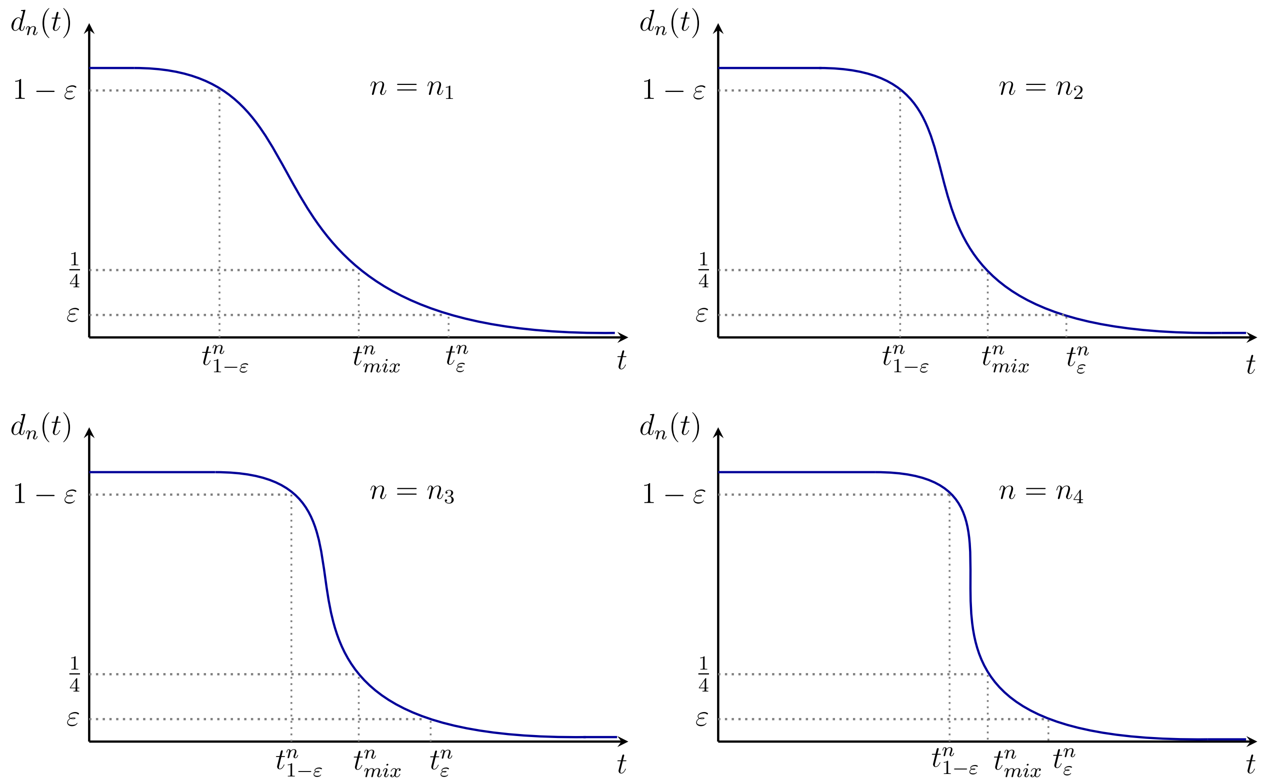


Figura 3.1: Representação gráfica do fenômeno cutoff, onde $n_1 < n_2 < n_3 < n_4$.

Abaixo, definiremos um fenômeno um pouco mais abrangente que o cutoff, e posteriormente mostraremos que tal fenômeno é uma condição necessária para que aconteça cutoff.

Definição 3.3.3 (Pré-cutoff)**.** Uma sequência de cadeias apresenta pré-cutoff se satisfaz a condição

$$\sup_{0<\varepsilon<1/2} \limsup_{n\to\infty} \frac{t^n_{mix}(\varepsilon)}{t^n_{mix}(1-\varepsilon)} < \infty.$$

Note que se uma cadeia apresenta cutoff, então para qualquer $\varepsilon \in (0,1)$ tem se

$$\lim_{n\to\infty} \frac{t^n_{mix}(\varepsilon)}{t^n_{mix}(1-\varepsilon)} = 1$$

o que garante que o a cadeia também apresenta pré-cutoff.

Sabendo que pré-cutoff é uma condição necessária para que ocorra cutoff, é normal nos perguntarmos se pré-cutoff é uma condição suficiente. Em [BHP17], é mostrado que não. Um contra-exemplo apresentado por David J. Aldous (exibido em [BHP17]) consiste na sequência de passeios aleatórios nos grafos descritos na Figura 3.2 com probabilidades de transição também exibidas na Figura 3.2. O comportamento da distância de variação total para a distribuição invariante é mostrado na Figura 3.3.

Figura 3.2: Dinâmica da sequência de cadeias que apresenta pré-cutoff mas não cutoff

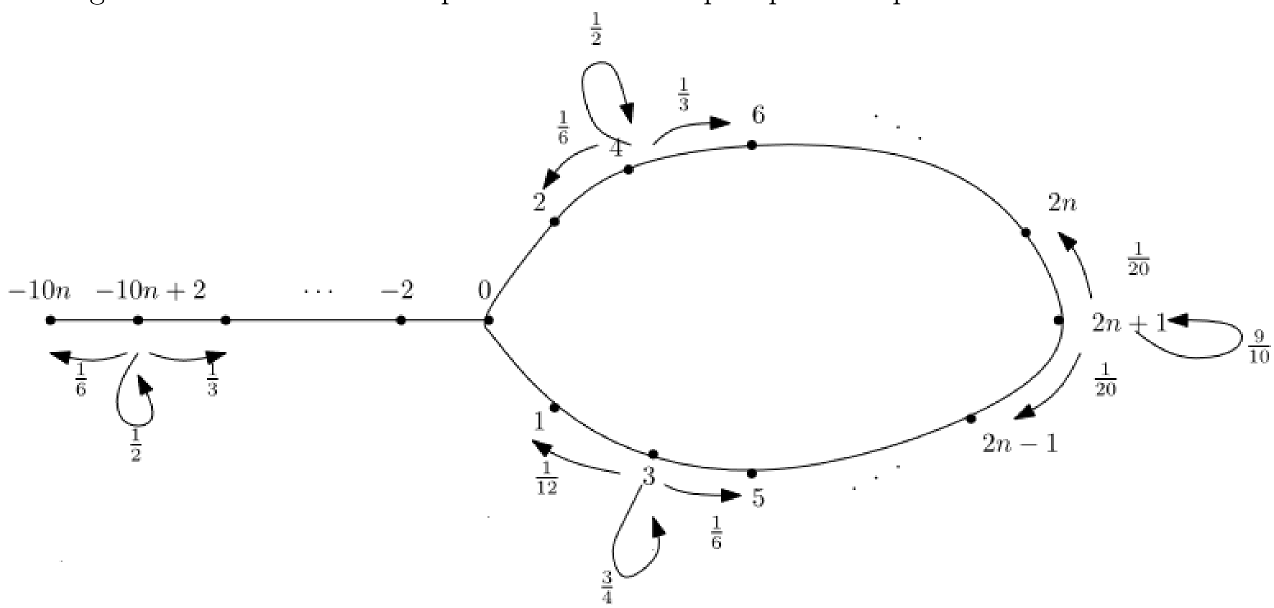
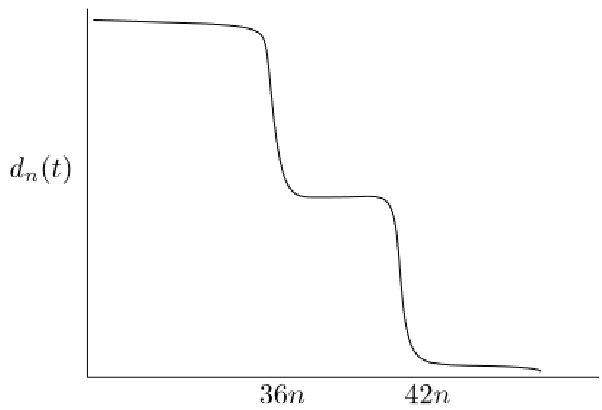


Figura 3.3: Comportamento da distância de variação total da n-esima cadeia



Observação 3.3.4. As figuras acima foram retiradas de [BHP17].

Definição 3.3.5 (Janela do Cutoff)**.** Suponha que a sequência $((X_t^n)_{t\in\mathbb{F}}\,;\,n\in\mathbb{N})$ de cadeias de Markov é tal que para todo $n\in\mathbb{N}$ os tempos de mistura da cadeia (X_t^n) são $t_{mix}^n(\varepsilon)$. Dizemos que a sequência de cadeias de Markov apresenta cutoff com janela de tamanho $O(\omega_n)$ se $\omega_n = o\left(t_{mix}^n\right)$ e

$$\lim_{\alpha\to-\infty}\liminf_{n\to\infty} d_n(t_{mix}^n+\alpha\omega_n)=1,$$
$$\lim_{\alpha\to\infty}\limsup_{n\to\infty} d_n(t_{mix}^n+\alpha\omega_n)=0.$$

Note que a janela do cutoff basicamente especifica um intervalo menor do que t^n_{mix} (em escala) no qual o decrescimento abrupto da distância para a distribuição invariante acontece. Com essa interpretação em mente, nos abrimos para uma outra possibilidade, a de estudarmos o decrescimento abrupto na escala da janela (ao invés de na escala de t^n_{mix}).

Definição 3.3.6 (Cutoff com perfil)**.** Suponha que a sequência $((X^n_t)_{t\in\mathbb{F}}\ ;\ n \in \mathbb{N})$ de cadeias de Markov é tal que para todo $n \in \mathbb{N}$ os tempos de mistura da cadeia (X^n_t) são $t^n_{mix}(\varepsilon)$, e adicionalmente suponha que a sequência apresenta cutoff com janela de tamanho $O(\omega_n)$. Dizemos que a sequência apresenta cutoff com perfil G se para todo $b \in \mathbb{R}$ tem-se

$$\lim_{n\to\infty} d_n(t^n_{mix} + b\omega_n) = G(b)$$

para $G : \mathbb{R} \to \mathbb{R}$, onde $d_n(t)$ é a distância à medida estacionária no tempo t.

Em essência perfil do cutoff se trata de uma função que descreve o comportamento assintótico do cutoff quando olhado na escala da janela transladando a origem para t^n_{mix}.

Abaixo exibiremos um exemplo de sequência de cadeias que apresenta cutoff.

Exemplo 3.3.7 (Passeio aleatório em $\{0, 1, \dots, n\}$ com viés β)**.** Fixado $p \in [0, 1]$, chamamos de passeio aleatório (lazy) em $\{0, 1, \dots, n\}$ com viés $\beta = p - 1/2$ a cadeia de Markov em tempo discreto com matriz de transição P dada por

$$P(x, y) = \begin{cases} p/2, & \text{se } y = x + 1 \text{ e } x \neq 0; \\ q/2, & \text{se } y = x - 1 \text{ e } y \neq n; \\ 1/2, & \text{se } x = 0 \text{ e } y = 1; \\ 1/2, & \text{se } x = n \text{ e } y = n - 1; \\ 1/2, & \text{se } y = x; \\ 0, & \text{caso contrário.} \end{cases}$$

Ou seja, quando em um vértice interno, saltamos para a esquerda com probabilidade $q/2$, para a esquerda com probabilidade $p/2$ e ficamos parados com probabilidade $1/2$. Quando em um dos vértices extremos a cadeia fica parada com probabilidade $1/2$, e se move para o vértice vizinho com probabilidade $1/2$.

Abaixo enunciaremos um resultado à respeito de uma sequência cujo perfil do cutoff é conhecido. O resultado é obtido na íntegra no capítulo 18 de [LPW].

Teorema 3.3.8. *A sequência de passeios aleatórios (X^n_t) com viés $\beta = p - 1/2$ em $\{0, 1, 2, \dots, n\}$ apresenta cutoff em $\beta^{-1}n$ com uma janela de tamanho $O(\sqrt{n})$. Mais precisamente, existe uma constante $c(\beta) > 0$ tal que para qualquer $\alpha \in \mathbb{R}$,*

$$\lim_{n\to\infty} d_n\left(\frac{n}{\beta} + \alpha\sqrt{n}\right) = \Phi(-c(\beta)\alpha),$$

onde Φ é distribuição normal padrão.

Nesta última parte do capítulo, definiremos um tipo de cadeia novo que será o tema central no próximo capítulo. Inicialmente definiremos em tempo discreto, e em seguida trataremos da sua versão contínua.

Começaremos apresentando uma heurística do processo. Fixado $n \in \mathbb{N}$, seja H um grafo conexo com n vértices. Inicialmente escolhe-se uma quantidade $k \in \{1, \ldots, n-1\}$ de vértices de H, e atribuímos o valor 1 para os vértices escolhidos e 0 para os demais. A cada instante de tempo, escolhe-se uma aresta e uniformemente ao acaso, e troca-se os valores dos vértices incidentes à aresta. Frequentemente, dizemos que os sítios que tem 1 associado possuem uma partícula, e o processo é interpretado como as partículas trocando de posição.

A figura abaixo ilustra um exemplo do processo de maneira visual.

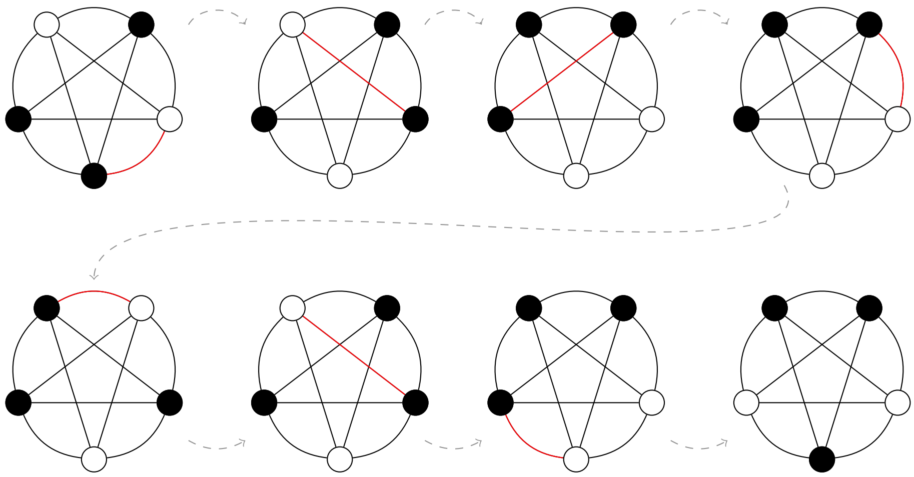


Figura 3.4: Exemplo visual do processo de exclusão em um grafo.

Abaixo segue a definição formal do processo

Definição 3.3.9 (configuração/espaço de configurações)**.** Fixado um grafo H conexo com conjunto de vértices V, chamamos o conjunto $\{0,1\}^V$ de espaço de configurações.
Um elemento η de $\{0,1\}^V$ é chamado de configuração.

Ou seja, na linguagem da heurística antes apresentada, uma configuração é uma certa disposição de partículas nos vértices do grafo.

Fixada uma configuração η e um vértice z, denotamos por η_z o valor atribuído ao vértice z na configuração η. Ademais, fixados vértices x, y do grafo, denotamos por $\eta^{x,y}$ a configuração obtida através da troca dos valores nos vértices x e y, mais precisamente, para cada z vale

$$\eta_z^{x,y} = \begin{cases} \eta_y, & \text{se } z = x; \\ \eta_x, & \text{se } z = y; \\ \eta_z, & \text{se } z \notin \{x, y\}. \end{cases}$$

Fixadas as notações, agora finalmente definiremos formalmente o processo.

Definição 3.3.10 (Processo de exclusão simples)**.** Fixado um grafo H finito e conexo com conjunto de arestas E, chamamos de processo de exclusão simples no grafo H o processo definido pela matriz de transição P em $\{0,1\}^V$ dada por

$$P(\eta,\zeta) = \begin{cases} \dfrac{1}{|E|}, \text{ se } \zeta = \eta^{x,y},\ \eta \neq \zeta \text{ e } \{x,y\} \in E; \\ 1 - \dfrac{|\{\{x,y\} \in E\,:\,\eta^{x,y} \neq \eta\}|}{|E|}, \text{ se } \zeta = \eta; \\ 0, \text{ caso contrário.} \end{cases}$$

Observação 3.3.11. Em alguns casos definimos o processo acima com a matriz de transição S dada por

$$S(\eta,\zeta) = \begin{cases} \dfrac{2}{2|U| + |L|}, \text{ se } \zeta = \eta^{x,y},\ \eta \neq \zeta \text{ e } \{x,y\} \in E; \\ 1 - \dfrac{2|\{\{x,y\} \in E\,:\,\eta^{x,y} \neq \eta\}|}{2|U| + |L|}, \text{ se } \zeta = \eta \\ 0, \text{ caso contrário.} \end{cases}$$

onde L é o conjunto dos loops do grafo e $U = E \setminus L$. No fundo, se trata de uma versão acelerada da cadeia anterior, com uma particularidade.

Esta versão é particularmente útil para o processo de exclusão no grafo completo (incluindo loops), pois nesse caso, as probabilidades de transição são simplesmente $2/n^2$. Em [LL11] é utilizada esta versão.

Observação 3.3.12. Do ponto de vista matemático, não existe nada de especial nas versões citadas acima. Em rigor, poderíamos gerar infinitas versões da cadeia, pois no fundo a única coisa que muda é o quão lazy a cadeia é. Sua dinâmica permanece a mesma.

já para a versão contínua, definimos por:

Definição 3.3.13 (Processo de exclusão simples (tempo contínuo))**.** Fixado um grafo H finito e conexo com conjunto de arestas E, chamamos de processo de exclusão simples no grafo H o processo definido pelo gerador L tal que

$$Lf(\eta) = k \sum_{(x,y)\in E} (f(\eta^{x,y}) - f(\eta))$$

com $k > 0$.

Note que na versão em tempo contínuo, a constante k dita a velocidade (em média) em que acontecem trocas. Para $k = 1$, é observada uma quantidade de trocas igual (em média) ao caso discreto. Um outro ponto importante a se observar é que no caso em tempo contínuo poderíamos fornecer taxas diferentes em cada uma das configurações, ou seja, poderíamos levar intervalos de tempo diferentes (em média) para sair de configurações diferentes.

Exemplo 3.3.14 (Processo de Exclusão no círculo ($\mathbb{Z}_{2n}$))**.** É o processo de exclusão definido no grafo $\mathbb{Z}_{2n}$.

Através das equações de equilíbrio (Definição 1.1.30) é facilmente verificável que a distribuição uniforme é uma distribuição invariante à cadeia, e pelo Corolário 1.1.54 sabemos que é a única. Hubert Lacoin mostrou em [Lac16] que para o processo de exclusão no círculo ($\mathbb{Z}_{2n}$) com configurações iniciais contendo n partículas, vale

$$\lim_{n\to\infty} d_n\left(\frac{n^2}{2\pi^2}\log(n)+\frac{n^2}{\pi^2}s\right)=\operatorname{erf}\left(\frac{\sqrt(2)}{\pi}\mathrm{e}^{\text{-s}}\right),$$

ou seja, a sequência apresenta cutoff com perfil erf $\left(\frac{\sqrt(2)}{\pi}\mathrm{e}^{-(\cdot)}\right)$ e janela n^2/π^2, onde erf é a Função Erro de Gauss dada por

$$\operatorname{erf}(t):=\frac{2}{\sqrt{\pi}}\int_0^t \mathrm{e}^{-u^2}\,\mathrm{d}u.$$

4
Perfil do Cutoff do Processo de Exclusão no grafo completo

4.1 Processo de exclusão simples no grafo completo

Este capítulo terá como objetivo calcular o perfil do cutoff para a sequência de processos de exclusão simples no grafo completo.

Resgatando a definição do capítulo anterior, redefinimos o processo de exclusão no Grafo completo de n vértices por o processo definido pela matriz de transição P em $\{0,1\}^V$ dada por

$$P(\eta,\zeta) = \begin{cases} \dfrac{1}{n^2}, & \text{se } \zeta = \eta^{x,y},\ \eta \neq \zeta \text{ e } \{x,y\} \in E; \\ 1 - \dfrac{|\{\{x,y\} \in E \,:\, \eta^{x,y} \neq \eta\}|}{n^2}, & \text{se } \zeta = \eta; \\ 0, & \text{caso contrário.} \end{cases}$$

Observe que na definição acima estamos considerando o grafo completo do n vértices incluindo os loops.

Observação 4.1.1. Neste capítulo, especialmente usaremos "cutoff para o processo de exclusão" para nos referirmos a um eventual cutoff para uma sequência de processos de exclusão.

O próximo resultado que enunciaremos, retirado de [LL11] se deve a H. Lacoin e R. Leblond.

Teorema 4.1.2. *Para cada $n \in \mathbb{N}^*$ considere o processo de exclusão no grafo completo com n vértices e k_n partículas iniciais.*

(a) *Se $\lim_{n\to\infty} k_n^2/n = \infty$, então para cada $\varepsilon > 0$ existe $\beta > 0$ tal que para qualquer n valem*

$$d_n\left(\frac{1}{2}n\log(n) + \beta n\right) \leq \varepsilon \qquad e \qquad d_n\left(\frac{1}{2}n\log(n) - \beta n\right) \geq 1 - \varepsilon;$$

(b) *Se* $\lim_{n\to\infty} k_n^2/n = 0$*, então para cada* $\varepsilon > 0$ *existe* $\beta > 0$ *tal que para qualquer* n *valem*

$$d_n\left(n\log(k_n) + \beta n\right) \leq \varepsilon \qquad e \qquad d_n\left(n\log(k_n) - \beta n\right) \geq 1 - \varepsilon;$$

(c) *Se* $\lim_{n\to\infty} k_n^2/n \in (0, \infty)$*, então valem as duas afirmações acima.*

Observação 4.1.3. No artigo o tempo de mistura aparece dividido por um fator de 2. Esta diferença acontece pelo fato de termos escolhido uma versão com taxas de salto divididas por 2 em relação à versão do artigo, tendo em vista o Exemplo 3.2.6.

Utilizando o resultado acima, buscaremos obter o perfil do cutoff para o processo de exclusão no grafo completo.

Para o restante do capítulo fixaremos algumas notações.
Denotaremos o conjunto $\{1, \ldots, n\}$ por Λ_n, usaremos Ω_n para denotar o conjunto $\{0,1\}^{\Lambda_n}$ e denotaremos por $\Omega_{n,k}$ o conjunto das configurações com k partículas, isto é,

$$\Omega_{n,k} = \left\{\sigma \in \Omega_n \ : \ \sum_{x\in\Lambda_n} \sigma_x = k\right\}.$$

Recordando, o processo de exclusão simples no grafo completo de n vértices em tempo contínuo é dado pelo gerador M_n tal que

$$M_n f(\sigma) := \frac{1}{n^2} \sum_{x,y\in\Lambda_n} \left(f(\sigma^{x,y}) - f(\sigma)\right)$$

para todo $\sigma \in \Omega_n$ e para toda $f : \Omega_n \to \mathbb{R}$.

Particularmente, nesta seção trabalharemos com a versão n vezes "acelerada" do processo definido acima, isto é, o processo definido pelo gerador L_n dado por

$$L_n f(\sigma) := \frac{1}{n} \sum_{x,y\in\Lambda_n} \left(f(\sigma^{x,y}) - f(\sigma)\right)$$

para todo $\sigma \in \Omega_n$.

Note que a cadeia gerada por L_n não é irredutível, visto que fixada uma configuração $\sigma \in \Omega_n$, só observamos transições entre configurações com o mesmo número de partículas. Para contornar este problema, fixada uma configuração inicial σ com k partículas, a cadeia gerada por L_n é irredutível se o espaço de estados é $\Omega_{n,k}$. Com isso em mente, usaremos $d_{n,k}$ para denotar a distância até a medida invariante para o processo de exclusão no grafo completo com espaço de estados $\Omega_{n,k}$.

Proposição 4.1.4. *Se* $(Z_t)_{t\geq 0}$ *é o processo de exclusão no grafo completo de* n *vértices com estado inicial* $\alpha \in \Omega_{n,k}$*, então* $\mu_{n,k} = \mathrm{unif}(\Omega_{n,k})$ *é a única distribuição invariante para* X_t*.*

Dem.: Uma simples verificação das equações de equilíbrio (Definição 1.1.30), e o Corolário 1.1.54 garantem o resultado. ■

Observação 4.1.5. Fixados $n \geq k \geq 0$, note que que pela simetria do grafo completo, a distância à distribuição invariante em qualquer instante de tempo é a mesma para qualquer $\alpha \in \Omega_{n,k}$. Ou seja, fixado um $\Omega_{n,k}$ a distribuição inicial do processo não influencia na convergência à distribuição estacionária.

Ao final da seção provaremos o seguinte teorema.

Teorema 4.1.6. *Seja $(k_n)_{n\in\mathbb{N}}$ uma sequência de inteiros positivos com $\lim_{n\to\infty} k_n = \infty$, e tal que $2k_n \leq n$ para todo $n \in \mathbb{N}$.*
Se $\lim_{n\to\infty} k_n^2/n = 0$, então existe $\gamma \in \mathbb{R}$ tal que

$$\lim_{n\to\infty} d_{n,k_n}(\log(k_n) + \gamma + t) = \mathbb{P}(\zeta > t),$$

onde $\zeta := \sum_{x=1}^{\infty} \frac{\xi_x - 1}{x}$ e $(\xi_x)_{x\in\mathbb{N}}$ é uma sequência de variáveis aleatórias i.i.d. com lei comum Exp(1).

Assim como foi feita a observação em [LL11], o cálculo de d_{n,k_n} pode ser reduzido ao cálculo da distância de variação total entre uma cadeia mais simples e a sua distribuição invariante. Usaremos esse caminho para obter o resultado.

Suponha que $\alpha \in \Omega_{n,k_n}$. Defina $A_n := \{x \in \Lambda_n \ : \ \alpha_x = 1\}$, ou seja, os sítios que possuem partícula na configuração α. Note que a cardinalidade de A_n é k_n.

Fixado $n \in \mathbb{N}^*$ denotaremos por $(Z^n(t))_{t\geq 0}$ o processo de exclusão no grafo completo de n vértices com estado inicial α.

Para cada $n \in \mathbb{N}$ definamos a cadeia $(W_t^n \ ; \ t \geq 0)$ por

$$W_t^n := \sum_{x\in A_n} Z_x^n(t)$$

para cada $t \geq 0$.
Ou seja, a cadeia $(W_t^n \ ; \ t \geq 0)$ conta quantas partículas estão no conjunto dos sítios que na configuração inicial possuíam partículas.

Por ser uma projeção[(1)] de uma cadeia de Markov, $(W_t^n)_{t\geq 0}$ é também uma cadeia de Markov. Além disso, através da análise das probabilidades de aumento ou diminuição de partículas no conjunto observado, é possível ver que $(W_t^n)_{t\geq 0}$ possui gerador $\mathbb{L}_{n,k}$ dado por

$$\mathbb{L}_{n,k} f(x) := \frac{(k_n - x)^2}{n}(f(x+1) - f(x)) + \frac{x(n - 2k_n + x)}{n}(f(x-1) - f(x))$$

para $f : \{0, 1, \ldots, k_n\} \to \mathbb{R}$ e $x \in \{0, 1, \ldots, k_n\}$.

Observação 4.1.7. A fim de simplificar a notação, no restante da seção omitiremos a dependência em n de k_n, W^n e Z^n.

(1) Ver seção 2.3 de [LPW]

No próximo resultado obteremos a distribuição invariante para a cadeia $(W_t^n)_{t\geq 0}$.

Proposição 4.1.8. *Se $W_0 = k$, então a distribuição invariante $\nu_{n,k}$ da cadeia é dada por*

$$\nu_{n,k}(x) := \frac{\binom{k}{x}\binom{n-k}{k-x}}{\binom{n}{k}}$$

para qualquer $x \in \{0, 1, \dots, k\}$.

Dem.: Note que pelo Corolário 2.3.12, basta verificarmos que $\nu_{n,k}(x,y)Q(x) = \nu_{n,k}(y,x)Q(y)$, onde Q é a matriz associada ao gerador $\mathbb{L}_{n,k}$. A verificação é simples, e por isso não a faremos aqui. ■

Observação 4.1.9. Na proposição acima, a distribuição uniforme de $(W_t)_{t\geq 0}$ é de certa forma *natural*, visto que a medida associada à cada valor x é simplesmente a razão entre o número de configurações com x partículas dentro do conjunto observado e $k - x$ partículas fora dele, e o número total de configurações.

Agora, finalmente mostraremos que a distância entre a cadeia $(W_t)_{t\geq 0}$ e $\nu_{n,k}$ é igual a $d_{n,k}$.

Proposição 4.1.10. *Para todo $t \geq 0$ tem se*

$$d_{n,k}(t) = d_{TV}\left(\mathbb{P}_\alpha(W_t = \cdot), \nu_{n,k}(\cdot)\right),$$

onde α é um elemento qualquer de $\Omega_{n,k}$.

Dem.: Definindo para cada l o conjunto $S_l := \left\{\sigma \in \Omega_{n,k} \ : \ \sum_{x\in A_n} \sigma_x = l\right\}$, temos

$$\begin{aligned}
d_{n,k}(t) &= \frac{1}{2}\sum_{\sigma\in\Omega_{n,k}} |\mathbb{P}(X(t) = \sigma) - \mu_{n,k}(\sigma)| \\
&= \frac{1}{2}\sum_{\sigma\in\Omega_{n,k}} \left|\mathbb{P}(X(t) = \sigma) - \frac{1}{\binom{n}{k}}\right| \\
&= \frac{1}{2}\sum_{l=1}^{k}\sum_{\sigma\in S_l} \left|\mathbb{P}(X(t) = \sigma) - \frac{1}{\binom{n}{k}}\right|
\end{aligned}$$

note que se $W_s = x$, então por termos $W_0 = k$ segue que X_s está uniformemente distribuído nas configurações com exatamente x partículas em A_n. Assim,

$$d_{n,k}(t) = \frac{1}{2}\sum_{l=1}^{k}\left|\sum_{\sigma\in S_l}\mathbb{P}(X(t) = \sigma) - \sum_{\sigma\in S_l}\frac{1}{\binom{n}{k}}\right|$$

Como S_l possui $\binom{k}{l}\binom{n-k}{k-l}$ elementos e $\sum_{\sigma\in S_l}\mathbb{P}(X(t) = \sigma) = \mathbb{P}(W_t = l)$, temos

$$\begin{aligned}
d_{n,k}(t) &= \frac{1}{2}\sum_{l=1}^{k}\left|\mathbb{P}(W_t = l) - \binom{k}{l}\binom{n-k}{k-l}\frac{1}{\binom{n}{k}}\right| \\
&= \mathrm{d}_{TV}(\mathbb{P}_\alpha(W_t = \cdot), \nu_{n,k}).
\end{aligned}$$

■

Portanto, reduzimos o problema inicial a analisar a convergência de $(W_t)_{t\geq 0}$.

4.2 Demonstração do Teorema 4.1.6

Trataremos do caso em que

$$\lim_{n\to\infty} k_n = +\infty \quad \text{e} \quad \lim_{n\to\infty} \frac{k_n^2}{n} = 0.$$

Sabemos que

$$\mathbb{L}_{n,k} f(x) := \frac{(k-x)^2}{n}(f(x+1)-f(x)) + \frac{x(n-2k+x)}{n}(f(x-1)-f(x))$$

para qualquer f limitada. Portanto, tomando f como a função identidade, obtemos

$$\begin{aligned}
\mathbb{L}_{n,k} W_t &= \mathbb{L}_{n,k} f(W_t) \\
&= \frac{(k-W_t)^2}{n}(W_t+1-W_t) + \frac{W_t(n-2k+W_t)}{n}(W_t-1-W_t) \\
&= \frac{(k-W_t)^2}{n} - \frac{W_t(n-2k+W_t)}{n} \\
&= \frac{k^2}{n} - W_t.
\end{aligned}$$

Segue do Lema A.3.6 que o processo M dado por

$$M(t) = W_t - W_0 - \int_0^t \mathbb{L}_{n,k} f(W_s)\,\mathrm{d}s$$

é um Martingal[(2)]. Assim, segue do fato de $M(0)=0$ e do Lema A.2.8 que

$$\begin{aligned}
\mathbb{E}[M(t)] &= \mathbb{E}[M(t) \mid M(0)=0]\,\mathbb{P}[M(0)=0] + \mathbb{E}[M(t) \mid M(0)\neq 0]\,\mathbb{P}[M(0)\neq 0] \\
&= \mathbb{E}[M(t) \mid M(0)=0].
\end{aligned}$$

Como $M(t)$ é um martingal, concluímos que $\mathbb{E}[M(t)] = 0$.

Logo,

$$\begin{aligned}
0 = \mathbb{E}[M(t)] &= \mathbb{E}[W_t] - W_0 - \mathbb{E}\left[\int_0^t \mathbb{L}_{n,k} f(W_s)\,\mathrm{d}s\right] \\
&= \mathbb{E}[W_t] - W_0 - \mathbb{E}\left[\int_0^t \left(\frac{k^2}{n} - W_s\right)\mathrm{d}s\right],
\end{aligned}$$

através do Teorema de Fubini (Teorema A.1.4), obtemos

$$0 = \mathbb{E}[W_t] - W_0 - \int_0^t \mathbb{E}\left[\frac{k^2}{n} - W_s\right]\mathrm{d}s.$$

---

(2) Certo tipo de processo $(X_t)_{t\geq 0}$ tal que particularmente vale $\mathbb{E}[X_{t+s}|X_s] = X_s$

E portanto,
$$\frac{\mathrm{d}\,\mathbb{E}\left[W_t\right]}{\mathrm{d}t} = \mathbb{E}\left[\frac{k^2}{n} - W_t\right] = \frac{k^2}{n} - \mathbb{E}\left[W_t\right],$$
o que unido ao fato de $W_0 = k$ nos leva a concluir que
$$\mathbb{E}\left[W_t\right] = ke^{-t} + \frac{k^2}{n}(1 - e^{-t}).$$
Note que pelo fato de cada uma das cadeias $(W_t)_{t\geq 0}$ ser irredutível e aperiódica, pelo Teorema da Convergência (Teorema 1.1.58) W_∞ está bem definida e tem-se $W_\infty \sim \nu_{n,k}$.
Pelo Teorema da Convergência Dominada (Teorema A.1.2) temos:
$$\mathbb{E}\left[W_\infty\right] = \lim_{t\to\infty} \mathbb{E}\left[W_t\right] = \frac{k^2}{n}.$$
Assim, utilizando a desigualdade de Markov segue que
$$\mathbb{P}(W_\infty \geq 1) \leq \mathbb{E}\left[W_\infty\right] = \frac{k^2}{n}.$$

Logo, temos
$$\begin{aligned} \mathrm{d}_{TV}(\nu_{n,k}, \delta_0) &= 1 - \nu_{n,k}(0) \\ &= 1 - \mathbb{P}(W_\infty = 0) \\ &= \mathbb{P}(W_\infty \geq 1) \\ &\leq \frac{k^2}{n}. \end{aligned} \tag{4.1}$$

Utilizando a Proposição 4.1.10, a Equação (4.1) e a desigualdade triangular na distância de variação total, segue que
$$|d_{n,k}(t) - d_{TV}(\mathrm{Lei}(W_t), \delta_0)| \leq d_{TV}(\nu_{n,k_n}, \delta_0) \leq \frac{k^2}{n},$$
e particularmente
$$d_{TV}(Lei(W_t), \delta_0) - \frac{k^2}{n} \leq d_{n,k}(t). \tag{4.2}$$

Observe que
$$d_{TV}(Lei(W_t), \delta_0) = 1 - \mathbb{P}(W_t = 0) = \mathbb{P}(W_t \geq 1). \tag{4.3}$$

Definamos a variável aleatória H_0 por
$$H_0 := \inf\left\{t \geq 0\,;\ W_t = 0\right\}.$$

Note que $H_0 > t \Longrightarrow W_t \geq 1$. Assim, utilizando (4.3) segue que
$$\mathbb{P}(H_0 > t) \leq \mathbb{P}(W_t \geq 1) = d_{TV}(Lei(W_t), \delta_0). \tag{4.4}$$

Considere as cadeias de Markov em tempo contínuo $Y^1, Y^2, \dots, Y^k$ independentes, geradas por $\mathbb{L}_{n,k}$, e com condições iniciais $1, 2, \dots, k$ respectivamente.
Para cada par $(x, y) \in \{0, 1, \dots, k\} \times \{0, 1, \dots, k\}$, defina um acoplamento $((Z_t^{x,y}, W_t^{x,y})_{t\geq 0}, \mathbb{P}^{x,y})$ dominante (como no Exemplo 3.1.14) entre Y^x e Y^y.
Definindo para cada par x, y a variável aleatória

$$\tau_{x,y} := \inf\left\{t \geq 0 \ : \ Z_t^{x,y} = W_t^{x,y}\right\},$$

segue do Teorema 3.1.16 que

$$d_{n,k}(t) \leq \sup_{x,y\in\{0,\dots,k\}} \mathbb{P}^{x,y}(\tau_{x,y} > t) \tag{4.5}$$

Se $x > y$, note que devido a natureza do acoplamento $(Z_t^{x,y}, W_t^{x,y})_{t\geq 0}, \mathbb{P}^{x,y}$ a variável aleatória

$$H_{x,y} := \inf\left\{t \geq 0 \ : \ Z_t^{x,y} = 0\right\}$$

é tal que

$$\mathbb{P}^{x,y}(H_{x,y} \geq \tau_{x,y}) = 1.$$

Particularmente,

$$\mathbb{P}^{x,y}(H_{x,y} > t) \geq \mathbb{P}^{x,y}(\tau_{x,y} > t)$$

para todo $t \geq 0$.

Observe que o caso $y > x$ é simétrico ao caso anterior, no sentido que podemos definir $H_{x,y}$ de forma semelhante, e que $H_{x,y} \sim H_{y,x}$.

Assim, pela Equação (4.5), concluímos que

$$d_{n,k}(t) \leq \sup_{x,y\in\{1,\dots,k\}} \mathbb{P}^{x,y}(H_{x,y} > t).$$

Observe que $H_{k,y}$ e $H_{x,k}$ possuem a mesma distribuição de H_0 para quaisquer $x, y \in \{0, 1, \dots, k\}$. Além disso, se $x > y$ e $z > w$, então claramente $H_{x,y} \geq H_{z,w}$ sempre que $x \geq z$. Vale também o caso simétrico, para $x < y$ e $z < w$ vale $H_{x,y} \leq H_{z,w}$ sempre que $y \geq w$. Portanto, H_0 domina estocasticamente as variáveis aleatórias $H_{x,y}$.

Assim, segue que

$$d_{n,k}(t) \leq \sup_{x,y\in\{1,\dots,k\}} \mathbb{P}^{x,y}(H_{x,y} > t) \leq \mathbb{P}(H_0 > t). \tag{4.6}$$

Unindo (4.2), (4.4) e (4.6), concluímos que

$$\mathbb{P}(H_0 > t) - \frac{k^2}{n} \leq \ d_{k,n}(t) \ \leq \mathbb{P}(H_0 > t). \tag{4.7}$$

Portanto, para provar o resultado podemos concentrar nossos esforços em descrever o comportamento assintótico da probabilidade $\mathbb{P}(H_0 > t)$ (conforme $n \to \infty$). Com isso em mente, definamos para cada $x \in \{0, 1, \dots \ k\}$, a variável aleatória

$$H_x := \inf\left\{t \geq 0 \ ; \ W_t = x\right\}$$

e para cada $x \neq 0$ defina $H_{x,x-1} := H_{x-1} - H_x$.
Observe que pela propriedade forte de Markov as variáveis aleatórias acima definidas são independentes. Além disso, observe também que

$$H_0 = \sum_{x=1}^{k} H_{x,x-1}.$$

Nos próximos passos usaremos a interpretação da cadeia $(W_t)_{t\geq 0}$ feita na Observação 2.2.16.

Para cada x definamos

$$\lambda_x = \frac{(k-x)^2}{n} + \frac{x(n-2k+x)}{n},$$
$$p_x = \frac{(k-x)^2}{n\lambda_x} \quad \text{e} \quad q_x = \frac{x(n-2k+x)}{n\lambda_x}.$$

Quando no estado x a cadeia espera um tempo exponencial de parâmetro λ_x (Equivalente a esperar E/λ_x, onde E é um tempo exponencial de parâmetro 1, pela Observação 2.2.16) e após isso com probabilidade p_x a cadeia passa para o estado $x+1$ e com probabilidade q_x a cadeia passa para o estado $x-1$.

Sejam $\xi_1, \ldots, \xi_k$ variáveis aleatórias i.i.d. com lei comum $\text{Exp}(1)$. Condicionado no primeiro tempo de salto depois de chegar em x, temos

$$H_{x,x-1} = \frac{\xi_x}{\lambda_x} + \theta_x(H^1_{x+1,x} + H^2_{x,x-1}),$$

onde $\theta_x \sim \text{Bern}(p_x)$, $H^1_{x+1,x}$ e $H^2_{x,x-1}$ são cópias de $H_{x+1,x}$ e $H_{x,x-1}$ de modo que $\xi_x, \theta_x, H^1_{x+1,x}$ e $H^2_{x,x-1}$ sejam todas independentes entre si. A heurística por trás do argumento mora no fato de que ao chegar em x o tempo de espera até o próximo salto é ξ_x/λ_x, e ao saltar com probabilidade p_x a cadeia transiciona para $x-1$, e com probabilidade q_x a cadeia transiciona para $x+1$, tornando necessário esperar $H^1_{x+1,x-1} + H^2_{x,x-1}$ para transicionar para $x-1$.

Para cada $x \in \{1, \ldots, k\}$, defina

$$M_x := H_{x,x-1} - \frac{\xi_x}{\lambda_x}$$

Note que

$$\begin{aligned}
\mathbb{E}\left[M_x\right] &= \mathbb{E}\left[\theta_x(H^1_{x+1,x} + H^2_{x,x-1})\right] \\
&= \mathbb{E}\left[\theta_x(M_{x+1} + M_x) + \theta_x\left(\frac{\xi_{x+1}}{\lambda_{x+1}} + \frac{\xi_x}{\lambda_x}\right)\right]
\end{aligned}$$

Usando a linearidade do valor esperado, e o fato de θ_x ser independente das demais variáveis aleatórias, segue que

$$\mathbb{E}\left[M_x\right] = p_x\left(\frac{1}{\lambda_{x+1}} + \mathbb{E}\left[M_{x+1}\right] + \frac{1}{\lambda_x} + \mathbb{E}\left[M_x\right]\right),$$

e portanto

$$\mathbb{E}\left[M_x\right] = \frac{p_x}{q_x}\left(\frac{1}{\lambda_{x+1}} + \frac{1}{\lambda_x} + \mathbb{E}\left[M_{x+1}\right]\right).$$

Neste momento, provaremos que as variáveis M_x ficam pequenas para n suficientemente grande, e assim trocaremos o estudo das variáveis $H_{x,x-1}$ pelo estudo de $\frac{\xi_x}{\lambda_x}$.

Proposição 4.2.1. *Existe $n_0 \in \mathbb{N}$ tal que para $n > n_0$ vale*

$$\mathbb{E}\left[M_x\right] \leq \frac{16k^2}{nx^2}$$

para todo $x \in \{1, \ldots, k\}$.

Dem.: Tome $n_0 \in \mathbb{N}$ de modo que para todo $n > n_0$ tenha-se $n \geq 4k^2$ (e particularmente $n \geq 4k$).

Assim, fixado $n > n_0$, observe que para qualquer $x \in \{1, \ldots, k-1\}$ temos:

$$\lambda_{x+1} = \frac{(k-x-1)^2}{n} + \frac{(x+1)(n-2k+x+1)}{n} = \lambda_x + \frac{n-4k+4x+2}{n} \geq \lambda_x. \tag{4.8}$$

Note também, que

$$\lambda_x \geq \frac{x(n-2k+x)}{n} \geq \frac{x(n-2k)}{n} \geq \frac{x(n-n/2)}{n} \geq \frac{x}{2}, \tag{4.9}$$

e

$$\frac{p_x}{q_x} = \frac{(k-x)^2}{x(n-2k+x)} \leq \frac{k^2}{x(n-2k)} \leq \frac{k^2}{x(n-n/2)} \leq \frac{2k^2}{nx}. \tag{4.10}$$

Sabendo que

$$\mathbb{E}\left[M_k\right] = 0 \leq \frac{4p_k}{\lambda_k q_k},$$

suponhamos que

$$\mathbb{E}\left[M_{x+1}\right] \leq \frac{4p_{x+1}}{\lambda_{x+1}q_{x+1}}$$

para algum $x \in \{1, \ldots, k-1\}$.

Neste caso, temos:

$$\begin{aligned} 4 &\geq 2 + \frac{2}{(x+1)} \\ &\geq 2 + \frac{4k^2}{n}\frac{2}{(x+1)} \\ &= 2 + 4\frac{2k^2}{n(x+1)}. \end{aligned}$$

Assim, utilizando (4.8) e (4.10) segue que

$$4 \geq \frac{\lambda_x}{\lambda_{x+1}} + \frac{\lambda_x}{\lambda_x} + \frac{4\lambda_x p_{x+1}}{\lambda_{x+1} q_{x+1}},$$

que é suficiente para garantir que

$$\begin{aligned} \frac{4p_x}{\lambda_x q_x} &\geq \frac{p_x}{q_x}\left(\frac{1}{\lambda_{x+1}} + \frac{1}{\lambda_x} + \frac{4p_{x+1}}{\lambda_{x+1} q_{x+1}}\right) \\ &\geq \frac{p_x}{q_x}\left(\frac{1}{\lambda_{x+1}} + \frac{1}{\lambda_x} + \mathbb{E}\left[M_{x+1}\right]\right) \\ &= \mathbb{E}\left[M_x\right]. \end{aligned}$$

Portanto, pelo princípio de indução matemática concluímos que

$$\frac{4p_x}{\lambda_x q_x} \geq \mathbb{E}\left[M_x\right]$$

para qualquer $x \in \{1, \ldots, k\}$.
Assim, segue de (4.9) e (4.10) que

$$\frac{16k^2}{nx^2} \geq \mathbb{E}\left[M_x\right]$$

para qualquer $x \in \{1, \ldots, k\}$. ∎

Observação 4.2.2. Doravante, consideraremos $n > n_0$ da proposição anterior.

Observe que

$$\mathbb{E}\left[\left|H_0 - \sum_{x=1}^{k} \frac{\xi_x}{\lambda_x}\right|\right] \leq \sum_{x=1}^{k} \mathbb{E}\left[M_x\right] \leq \sum_{x=1}^{\infty} \frac{16k^2}{nx^2} \leq \frac{32k^2}{n}. \tag{4.11}$$

ou seja, conforme avançamos no tempo, espera-se que H_0 esteja muito próxima da soma das variáveis $\dfrac{\xi_x}{\lambda_x}$.

Assim como trocamos o estudo de $H_{x,x-1}$ por $\frac{\xi_x}{\lambda_x}$, agora compararemos $\dfrac{1}{\lambda_x}$ e $\dfrac{1}{x}$ a fim de posteriormente trocar o estudo de $\dfrac{\xi_x}{\lambda_x}$ por $\dfrac{\xi_x}{x}$.

Observe que pelo fato de $n \geq 4k^2$, e $k \geq x$, temos

$$|\lambda_x - x| = \left|\frac{k^2 - 4kx + 2x^2}{n}\right| \leq \frac{k^2}{n} \leq \frac{1}{4}.$$

Portanto, utilizando a desigualdade

$$\left|\frac{1}{a} - \frac{1}{b}\right| \leq \frac{|a-b|}{b^2} \frac{1}{1 - \frac{|a-b|}{b}}$$

válida sempre que $|a-b| \leq b$, obtemos

$$\left|\frac{1}{\lambda_x} - \frac{1}{x}\right| \leq \frac{|\lambda_x - x|}{x^2} \frac{1}{1 - \frac{|\lambda_x - x|}{x}} \leq \frac{k^2}{nx^2} \frac{1}{1 - \frac{1}{4x}} \leq \frac{k^2}{nx^2} \frac{1}{1 - \frac{1}{4}} \leq \frac{4k^2}{3nx^2}.$$

Assim, pela desigualdade triangular, e pela linearidade do valor esperado, segue que

$$\mathbb{E}\left[\left|\sum_{x=1}^{k} \left(\frac{\xi_x}{\lambda_x} - \frac{\xi_x}{x}\right)\right|\right] \leq \sum_{x=1}^{k} \mathbb{E}\left[\xi_x \left|\frac{1}{\lambda_x} - \frac{1}{x}\right|\right] \leq \sum_{x=1}^{k} \mathbb{E}\left[\xi_x \frac{4k^2}{3nx^2}\right] \leq \sum_{x=0}^{\infty} \frac{4k^2}{3nx^2} \leq \frac{8k^2}{3n} \tag{4.12}$$

Novamente, segue da desigualdade triangular que

$$\mathbb{E}\left[\left|H_0 - \sum_{x=1}^{k} \frac{\xi_x}{x}\right|\right] \leq \mathbb{E}\left[\left|H_0 - \sum_{x=1}^{k} \frac{\xi_x}{\lambda_x}\right|\right] + \mathbb{E}\left[\left|\sum_{x=1}^{k} \left(\frac{\xi_x}{\lambda_x} - \frac{\xi_x}{x}\right)\right|\right]$$

de onde, utilizando (4.11) e (4.12), concluímos que

$$\mathbb{E}\left[\left|H_0 - \sum_{x=1}^{k} \frac{\xi_x}{x}\right|\right] \leq \frac{104k^2}{3n}.$$

Assim, finalmente reduzimos o estudo do comportamento de $d_{k,n}(t)$ ao estudo da soma

$$\sum_{x=1}^{k} \frac{\xi_x}{x}.$$

Note que

$$\sum_{x=0}^{\infty} \operatorname{Var}\left(\frac{\xi_x - 1}{x}\right) = \sum_{x=0}^{\infty} \operatorname{Var}\left(\frac{\xi_x}{x}\right) = \sum_{x=1}^{k} \frac{1}{x^2} \leq 2$$

e

$$\sum_{x=1}^{\infty} \mathbb{E}\left[\frac{\xi_x - 1}{x}\right] = 0.$$

Portanto, pelo Teorema das duas séries de Kolmogorov (Teorema A.1.8) segue que

$$\zeta := \lim_{k \to \infty} \sum_{x=0}^{k} \frac{\xi_x - 1}{x}$$

é uma variável aleatória bem definida (a soma converge quase certamente).

Para o próximo passo utilizaremos o Lema A.3.5 (enunciado abaixo), cuja demonstração se encontra no apêndice.

Lema 4.2.3. *Existe $\gamma \in (0, \infty)$ tal que para qualquer $y \in \mathbb{N}_+$ vale*

$$\left|\sum_{x=1}^{k} \frac{1}{x} - \log y - \gamma\right| \leq \frac{3}{2k}.$$

Note que
$$\sum_{x=1}^{k}\frac{\xi_x}{x}-\log k-\gamma=\sum_{x=1}^{k}\frac{\xi_x-1}{x}+\sum_{x=1}^{k}\frac{1}{x}-\log k-\gamma.$$
Portanto, pelo lema acima, segue que

$$\lim_{k\to\infty}\left(\sum_{x=1}^{k}\frac{\xi_x}{x}-\log k-\gamma\right)=\zeta,$$

o que nos leva a
$$\lim_{n\to\infty}\mathbb{P}\left(\sum_{x=1}^{k}\frac{\xi_x}{x}>\log k+\gamma+t\right)=\mathbb{P}(\zeta>t),$$

e consequentemente
$$\lim_{n\to\infty}\mathbb{P}\left(H_0>\log k+\gamma+t\right)=\mathbb{P}(\zeta>t).$$

Portanto, pela desigualdade obtida em (4.7), segue que

$$\lim_{n\to\infty}d_{n,k}\left(\log k+\gamma+t\right)=\mathbb{P}(\zeta>t)$$

que é o resultado desejado.

A
Ferramentas da teoria da medida e probabilidade

A.1 Resultados Básicos

As demonstrações e mais detalhes importantes podem ser encontrados em [Bar95, Dur19]. Para o restante da seção, considere um espaço de medida $(\Omega, \mathcal{F}, \mu)$.

Teorema A.1.1 (Teorema da Convergência Monótona). *Seja $(f_n)_{n\in\mathbb{N}}$ uma sequência monótona não-decrescente de funções μ-integráveis assumindo valores reais positivos. Se $(f_n)_{n\in\mathbb{N}}$ converge pontualmente para uma função f, então f é integrável e*

$$\int f \,\mathrm{d}\mu = \lim_{n\to\infty} \int f_n \,\mathrm{d}\mu.$$

Teorema A.1.2 (Teorema da Convergência Dominada). *Seja $(f_n)_{n\in\mathbb{N}}$ uma sequência de funções μ-integráveis assumindo valores reais, e convergindo μ-q.t.p. para uma função mensurável f. Se existe uma função integrável g tal que $|f_n| \leq g$ para todo $n \in \mathbb{N}$, então f é integrável e*

$$\int f \,\mathrm{d}\mu = \lim_{n\to\infty} \int f_n \,\mathrm{d}\mu.$$

Teorema A.1.3. *Se X é uma variável aleatória contínua asssumindo valores reais, com função densidade f_X, e g é uma função mensurável, então*

$$\mathbb{E}\left[g(X)\right] = \int_{-\infty}^{\infty} g(x) f_X(x) \,\mathrm{d}x.$$

Teorema A.1.4 (Teorema de Fubini). *Sejam $(\Omega, \mathcal{F}, \mu)$ e $(A, \mathcal{G}, \nu)$ espaços σ-finitos. Se F é uma função mensurável em $(\Omega \times A, \sigma(\mathcal{F} \times \mathcal{G}), \mu \otimes \nu)$, e $\mu \otimes \nu$- integrável, então as funções definidas $\mu \otimes \nu$ - q.t.p. por*

$$f(x) = \int_A F_x \,\mathrm{d}\nu \quad \textit{e} \quad g(y) = \int_\Omega F^y \,\mathrm{d}\mu$$

possuem integrais finitas e tem-se

$$\int_\Omega \left[F \int_A \mathrm{d}\nu \right] \mathrm{d}\mu = \int_{\Omega \times A} F \, \mathrm{d}(\mu \otimes \nu) = \int_A \left[F \int_\Omega \mathrm{d}\mu \right] \mathrm{d}\nu.$$

Teorema A.1.5 (Desigualdade de Markov)**.** *Se X é uma variável aleatória não negativa, então para todo $t > 0$ tem-se*

$$\mathbb{P}(X > t) \leq \frac{\mathbb{E}[X]}{t}$$

Teorema A.1.6 (Teorema da Extensão de Carathéodory)**.** *Se μ é uma medida σ-finita em uma álgebra $\mathcal{A}$, então μ possui uma única extensão para $\sigma(\mathcal{A})$.*

Dem.: Consultar [Dur19] ou [Bar95]. ■

Teorema A.1.7 (Desigualdade de Chebyshev)**.** *Se X é uma variável aleatória com média μ, então para qualquer $a \in \mathbb{R}$ tem-se*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{\mathbb{E}[X^2] - \mu^2}{a^2}.$$

Teorema A.1.8 (Teorema das duas séries de Kolmogorov)**.** *Seja $(X_n)_{n \in \mathbb{N}}$ uma sequência de variáveis aleatórias independentes de modo que para cada $n \in \mathbb{N}$ se tenha $\mathbb{E}[X_n] = \mu_n$ e $\mathrm{Var}[X_n] = \sigma_n^2$. Se*

$$\sum_{n=0}^{\infty} \mu_n < \infty \quad e \quad \sum_{n=0}^{\infty} \sigma_n^2 < \infty,$$

então a série

$$\sum_{n=0}^{\infty} X_n$$

converge quase certamente em $\mathbb{R}$.

Definição A.1.9 (π-sistema)**.** Se Ω é um conjunto não vazio, dizemos que uma coleção $\mathcal{A}$ de subconjuntos de Ω é um π-sistema se $\mathcal{A}$ é fechado para interseções finitas.

Definição A.1.10 (λ-sistema)**.** Se Ω é um conjunto não vazio, dizemos que uma coleção $\mathcal{A}$ de subconjuntos de Ω é um λ-sistema se as seguintes condições são satisfeitas:

(i) $\Omega \in \mathcal{A}$;

(ii) Se $A, B \in \mathcal{A}$ e $A \subseteq B$, então $B \setminus A \in \mathcal{A}$;

(iii) Se $A_1, A_2, \dots \in \mathcal{A}$ e $A_n \uparrow A$, então $A \in \mathcal{A}$.

Teorema A.1.11 (Teorema λ-π de Dynkin)**.** *Se $\mathcal{A}$ é um π-sistema e $\mathcal{B}$ é um λ-sistema ambos de Ω, e $\mathcal{A} \subseteq \mathcal{B}$, então $\sigma(\mathcal{A}) \subseteq \mathcal{B}$.*

Dem.: Ver Apêndice A de [Dur19]. ■

Corolário A.1.12. *Sejam μ_1 e μ_2 são duas medidas definidas em um mesmo espaço $(\Omega, \mathcal{F})$ tais que*

$$\mu_1(A) = \mu_2(A)$$

para qualquer A em um π-sistema $\mathcal{A}$. Se $\sigma(\mathcal{A}) = \mathcal{F}$, então $\mu_1 = \mu_2$.

Observação A.1.13 (Processos contínuos à direita são definidos pelas distribuições finito-dimensionais)**.** Note que a σ-álgebra $\mathcal{J}$ gerada pelos conjuntos da forma

$$\{X_{t_0} = a_0, X_{t_1} = a_1, \dots, X_{t_n} = a_n\}$$

com $n \in \mathbb{N}$ e estados $a_0, a_1, \dots a_n$ e $t_1, t_2, \dots, t_n \in \mathbb{R}_{\geq 0}$ é igual a $\sigma(X_t : t \geq 0)$. De fato: Como os conjuntos da forma $\{X_t = a\}$ pertencem a $\mathcal{J}$, particularmente cada X_t é $\mathcal{J}$-mensurável, o que garante que $\sigma(X_t : t \geq 0) \subseteq \mathcal{J}$.

Em contrapartida, observe que

$$\{X_{t_0} = a_0, X_{t_1} = a_1, \dots, X_{t_n} = a_n\} = \bigcap_{k=0}^{n} X_{t_k}^{-1}(a_k),$$

mostrando que $\{X_{t_0} = a_0, X_{t_1} = a_1, \dots, X_{t_n} = a_n\} \in \sigma(X_t : t \geq 0)$, e consequentemente $\mathcal{J} \subseteq \sigma(X_t : t \geq 0)$.

Como a coleção $\mathcal{A}$ dos conjuntos da forma $\{X_{t_0} = a_0, X_{t_1} = a_1, \dots, X_{t_n} = a_n\}$ é um π-sistema, pelo Corolário A.1.12 uma medida em $\sigma(X_t : t \geq 0)$ é unicamente determinada pelos seus valores em $\mathcal{A}$.

Porém, por ser um processo contínuo à direita, sabemos que para cada $t \in \mathbb{R}_{\geq 0}$ vale

$$\{X_t = a\} = \bigcup_{q \in \mathbb{Q}_{\geq 0}}^{\infty} \bigcap_{p \in \mathbb{Q} \cap [t, t+q]} \{X_p = a\},$$

onde $\mathbb{Q}_{\geq 0} := \{y \in \mathbb{Q} \mid y \geq 0\}$. Portanto, particularmente a coleção $\mathcal{B}$ dos conjuntos da forma

$$\{X_{t_0} = a_0, X_{t_1} = a_1, \dots, X_{t_n} = a_n\}$$

com $n \in \mathbb{N}$, estados $a_0, a_1, \dots a_n$ e $t_0, \dots, t_n \in \mathbb{Q}_{\geq 0}$ (que também é um π-sistema), é tal que $\sigma(\mathcal{B}) = \sigma(\mathcal{A})$. Assim, pela argumentação anterior, uma medida em $\sigma(X_t : t \geq 0)$ é unicamente determinada pelos seus valores em $\mathcal{B}$.

A.2 Esperança condicional

Definição A.2.1 (Esperança condicional)**.** Seja $(\Omega, \mathcal{F}, \mathbb{P}$ um espaço de probabilidade e $\mathcal{G}$ uma σ-álgebra contida em $\mathcal{F}$. Fixada uma variável aleatória X $\mathcal{G}$-mensurável e com $\mathbb{E}[|X|] < \infty$, definimos a esperança condicional de X dado $\mathcal{G}$ (Notação $\mathbb{E}[X \mid \mathcal{G}]$) como sendo qualquer variável aleatória Y que satisfaça as seguintes condições:

(i) Y é $\mathcal{G}$-mensurável;

(ii) Ppara qualquer $A \in \mathcal{G}$, tem-se $\mathbb{E}[\mathbf{1}_A Y] = \mathbb{E}[\mathbf{1}_A X]$.

Observação A.2.2. Recorde que pela definição de valor esperado, vale

$$\mathbb{E}[\mathbf{1}_A X] = \int X \mathbf{1}_A \, \mathrm{d}\mathbb{P} = \int_A X \, \mathrm{d}\mathbb{P}.$$

Definição A.2.3 (Esperança condicional dada uma variável aleatória)**.** Dadas X, Y duas variáveis aleatórias definidas um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P}$, definimos a esperança condicional de X dado Y (Notação $\mathbb{E}[X \mid Y]$) por

$$\mathbb{E}[X \mid Y] := \mathbb{E}[X \mid \sigma(Y)].$$

Definição A.2.4 (Esperança condicional dado um evento)**.** Dada uma variável aleatória X definida no espaço $(\Omega, \mathcal{F}, \mathbb{P})$ e um evento $A \in \mathcal{F}$ com $\mathbb{P}(A) > 0$, definimos a esperança condicional de X dado A (Notação $\mathbb{E}[X \mid A]$) por

$$\mathbb{E}[X \mid A] := \frac{\mathbb{E}[X\mathbf{1}_A]}{\mathbb{P}(A)}.$$

Lema A.2.5 (Lei da Esperança Iterada)**.** *Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade e $\mathcal{G}$ uma σ-álgebra contida em $\mathcal{F}$. Fixada uma variável aleatória X $\mathcal{G}$-mensurável, então*

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}[X].$$

Dem.: Consequência direta do item (ii) da Definição A.2.1 tomando $A = \Omega$. ■

Definição A.2.6 (Probabilidade Condicional dada uma σ-álgebra)**.** Fixados um conjunto A e uma σ-álgebra $\mathcal{G}$ num mesmo espaço de probabilidade, definimos a probabilidade de A dado $\mathcal{F}$ por

$$\mathbb{P}(A \mid \mathcal{G}) = \mathbb{E}[\mathbf{1}_A \mid \mathcal{G}].$$

Lema A.2.7. *Fixado um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$, suponha que $A_1, A_2, \ldots$ seja uma partição de Ω. Se $\mathbb{P}(A_i) > 0$ para todo índice i, e $\mathcal{G} = \sigma(A_1, A_2, \ldots)$, então*

$$\mathbb{E}[X \mid \mathcal{G}] = \sum_{i=1}^{\infty} \frac{\mathbb{E}[\mathbf{1}_{A_i} X]}{\mathbb{P}(A_i)} \mathbf{1}_{A_i}.$$

Lema A.2.8. *Fixado um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$, suponha que $A_1, A_2, \ldots$ seja uma partição de Ω. Se X é uma variável aleatória definida neste espaço, então*

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} \left(\mathbb{E}[X\mathbf{1}_{A_i}] \, \mathbb{P}(A_i)\right).$$

Teorema A.2.9. *Sejam X, Y são variáveis aleatórias definidas em um mesmo espaço de probabilidade, e seja $\mathcal{F}$ uma sub-σ-álgebra. Se $X \in \mathcal{F}$, $\mathbb{E}[|Y|] < \infty$ e $\mathbb{E}[|XY|] < \infty$, então*

$$\mathbb{E}[XY \mid \mathcal{F}] = X\mathbb{E}[Y \mid \mathcal{F}].$$

Particularmente,

$$\mathbb{E}[\mathbb{E}[XY \mid \mathcal{F}]] = \mathbb{E}[X\mathbb{E}[Y \mid \mathcal{F}]].$$

Proposição A.2.10. *Sejam X, Y são variáveis aleatórias independentes, e $\phi : \mathbb{R}^2 \to \mathbb{R}$ uma função tal que $\mathbb{E}[\phi(X, Y)] < \infty$. Se $g(x) := \mathbb{E}[\phi(x, Y)]$, então $\mathbb{E}[\phi(X, Y) \mid X] = g(X)$.*

Dem.: Ver Exemplo 4.17 de [Dur19]. ■

Proposição A.2.11. *Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes, e seja $\phi : \mathbb{R}^n \to \mathbb{R}$ uma função tal que $\mathbb{E}[\phi(X_1, \ldots, X_n))] < \infty$. Se $g(y_1, \ldots, y_{n-1}) := \mathbb{E}[\phi(y_1, \ldots, y_{n-1}, X_n)]$, então $\mathbb{E}[\phi(X_1, \ldots, X_n) \mid \sigma(X_1, \ldots X_{n-1})] = g(X_n)$.*

Dem.: Ver [Dur19]. ■

A.3 Lemas úteis

Lema A.3.1 (Lema de Schur)**.** *Se $S \subseteq \mathbb{N}$ possui* $\mathrm{mdc}(S) = g_s$ *então existe inteiro m_s tal que para todo $m \geq m_s$ o produto mg_s pode ser escrito como combinação linear de elementos de S com coeficientes inteiros não-negativos.*

Dem.: Lema 1.30 de [LPW]. ■

Teorema A.3.2 (Propriedade da Perda de Memória)**.** *Uma variável aleatória $T : \Omega \to (0, \infty]$ possui distribuição exponencial se e somente se*

$$\mathbb{P}(T > s + t \,|\, T > s) = \mathbb{P}(T > t) \quad \textit{para quaisquer } s, t \geq 0.$$

Dem.: Suponha que $T \sim \mathrm{Exp}(\lambda)$ para algum $\lambda > 0$. Assim

$$\mathbb{P}(T > s + t \,|\, T > s) = \frac{\mathbb{P}(T > s + t, T > s)}{\mathbb{P}(T > s)} = \frac{\mathbb{P}(T > s + t)}{\mathbb{P}(T > s)} = \frac{\mathrm{e}^{-\lambda(s+t)}}{\mathrm{e}^{-\lambda(s)}} = \mathrm{e}^{-\lambda(t)} = \mathbb{P}(T > t).$$

Reciprocamente, suponha que T satisfação a condição do enunciado.
Como $T > 0$, existe $n \in \mathbb{N}^*$ tal que $\mathbb{P}(T > 1/n) > 0$. Definindo a função $f : \mathbb{R}_{\geq 0} \to [0, 1)$ por

$$f(t) = \mathbb{P}(T > t),$$

segue da hipótese que

$$f(1) = \mathbb{P}(T > 1) = \mathbb{P}\left(T > \sum_{k=1}^{n} \frac{1}{n}\right) = \mathbb{P}\left(T > \frac{1}{n}\right)^n > 0.$$

Assim, definimos implicitamente λ pela equação $\mathrm{e}^{-\lambda}$. Observe que

$$f\left(\frac{m}{n}\right) = \mathbb{P}\left(T > \frac{m}{n}\right) = \mathbb{P}\left(T > \sum_{k=1}^{m} \frac{1}{n}\right) = \mathbb{P}\left(T > \frac{1}{n}\right)^m = \left[\mathbb{P}\left(T > \frac{1}{n}\right)^n\right]^{m/n} = f(1)^{m/n},$$

para quaisquer $m, n \in \mathbb{N}^*$, o que nos leva a concluir que

$$f(q) = \mathrm{e}^{-\lambda q}$$

para qualquer $q \in \mathbb{Q}_{\geq 0}$. Assim, segue do fato de f ser não-crescente, e da continuidade de $\mathbb{P}$ que

$$f(x) = \mathrm{e}^{-\lambda x}$$

para todo $x \in \mathbb{R}_{\geq 0}$, e consequentemente $T \sim \mathrm{Exp}(\lambda)$. ■

Teorema A.3.3 (Propriedade da Perda de Memória II)**.** *Se Z é uma variável aleatória com distribuição* $\mathrm{Exp}(\lambda)$*, então para qualquer função mensurável f tem-se*

$$\mathbb{E}\left[f(Z - t)\mathbf{1}_{\{Z > t\}}\right] = \mathrm{e}^{-\lambda t}\,\mathbb{E}\left[f(Z)\right].$$

Dem.: Através do Teorema A.1.3, e

$$\begin{aligned}
\mathbb{E}\left[f(Z-t)\mathbf{1}_{\{Z>t\}}\right] &= \int f(Z-t)\mathbf{1}_{\{Z>t\}}\,\mathrm{d}\mathbb{P} \\
&= \int_0^\infty \lambda\,\mathrm{e}^{-\lambda x}\, f(x-t)\mathbf{1}_{\{x>t\}}\,\mathrm{d}x \\
&= \mathrm{e}^{-\lambda t}\int_0^\infty \lambda\,\mathrm{e}^{-\lambda(x-t)}\, f(x-t)\mathbf{1}_{\{x>t\}}\,\mathrm{d}x \\
&= \mathrm{e}^{-\lambda t}\int_t^\infty \lambda\,\mathrm{e}^{-\lambda(x-t)}\, f(x-t)\,\mathrm{d}x \\
&= \mathrm{e}^{-\lambda t}\int_t^\infty \lambda\,\mathrm{e}^{-\lambda(y)}\, f(y)\,\mathrm{d}y \\
&= \mathrm{e}^{-\lambda t}\,\mathbb{E}\left[f(Z)\right]
\end{aligned}$$

■

Teorema A.3.4. *Sejam $S_1, S_2, \dots$ uma sequência de variáveis aleatórias independentes com $S_n \sim \operatorname{Exp}(\lambda_n)$ com $\lambda_n \in (0,\infty)$ para todo $n \in \mathbb{N}_+$. As seguintes propriedades são válidas:*

(i) *Se $\sum_{n=1}^{\infty} \frac{1}{\lambda_n} < \infty$, então $\mathbb{P}\left(\sum_{n=1}^{\infty} S_n < \infty\right) = 1$.*

(ii) *Se $\sum_{n=1}^{\infty} \frac{1}{\lambda_n} = \infty$, então $\mathbb{P}\left(\sum_{n=1}^{\infty} S_n = \infty\right) = 1$.*

Dem.: Se $\sum_{n=1}^{\infty} \frac{1}{\lambda_n} < \infty$, então pelo Teorema da Convergência Dominada temos

$$\mathbb{E}\left(\sum_{n=1}^{\infty} S_n\right) = \sum_{n=1}^{\infty} \mathbb{E}\left[S_n\right] = \sum_{n=1}^{\infty} \frac{1}{\lambda_n} \leq \infty,$$

o que garante que

$$\mathbb{P}\left(\sum_{n=1}^{\infty} S_n < \infty\right) = 1$$

Em contrapartida, se $\sum_{n=1}^{\infty} \frac{1}{\lambda_n} = \infty$, então pela independência das variáveis aleatórias , temos:

$$\mathbb{E}\left[\exp\left(-\sum_{n=1}^{k} S_n\right)\right] = \prod_{n=1}^{k} \mathbb{E}\left(\mathrm{e}^{-S_n}\right) = \prod_{n=1}^{k} \mathrm{e}^{-\frac{1}{\lambda_n}} = \exp\left(\sum_{n=1}^{\infty} \frac{1}{\lambda_n}\right).$$

Portanto, pelo Teorema da Convergência Dominada, segue que

$$\mathbb{E}\left[\exp\left(-\sum_{n=1}^{k} S_n\right)\right] = 0.$$

Fixado $M > 0$ qualquer , note que

$$\begin{aligned}
0 &= \mathbb{E}\left[\exp\left(-\sum_{n=1}^{\infty} S_n\right)\right] \\
&\geq \mathbb{E}\left[\exp\left(-\sum_{n=1}^{\infty} S_n\right) \mathbf{1}_{\left\{\sum_{n=1}^{\infty} S_n < M\right\}}\right] \\
&\geq \mathbb{E}\left[\exp\left(-M\right) \mathbf{1}_{\left\{\sum_{n=1}^{\infty} S_n < M\right\}}\right] \\
&= \mathrm{e}^{-M}\, \mathbb{P}\left(\sum_{n=1}^{\infty} S_n < M\right),
\end{aligned}$$

o que nos mostra que

$$\mathbb{P}\left(\sum_{n=1}^{\infty} S_n < M\right) = 0.$$

Como tomamos $M > 0$ qualquer, segue que

$$\mathbb{P}\left(\sum_{n=1}^{\infty} S_n < \infty\right) = 0.$$

■

Lema A.3.5. *Existe $\gamma \in (0, \infty)$ tal que para qualquer $y \in \mathbb{N}_+$ vale*

$$\left|\sum_{x=1}^{k} \frac{1}{x} - \log k - \gamma\right| \leq \frac{3}{2k}.$$

Dem.: Note que pela fórmula de Taylor, para cada $x \in \mathbb{N}^*$ temos

$$\log\left(1 + \frac{1}{x}\right) = \sum_{p=1}^{\infty} \frac{(-1)^{p+1}}{p x^p}.$$

Assim, particularmente

$$0 \leq \frac{1}{x} - \log\left(1 + \frac{1}{x}\right) \leq \frac{1}{2x^2},$$

e como

$$\sum_{x=1}^{\infty} \frac{1}{2x^2} < \infty,$$

temos

$$\gamma := \frac{1}{x} - \log\left(1 + \frac{1}{x}\right) < \infty.$$

Observe que para cada $k \in \mathbb{N}^*$ vale

$$\log(k+1) = \sum_{x=1}^{k} \log\left(\frac{x+1}{x}\right) = \sum_{x=1}^{k} \left[\frac{1}{x} - \left(\frac{1}{x} - \log\left(\frac{x+1}{x}\right)\right)\right],$$

que nos leva a concluir que

$$\begin{aligned}\left|\sum_{x=1}^{k}\frac{1}{x}-\log y-\gamma\right| &\leq \left|\sum_{x=1}^{k}\left(\frac{1}{x}\right)-\log k-\sum_{x=1}^{k}\left(\frac{1}{x}-\log\left(1+\frac{1}{x}\right)\right)\right|+\sum_{x=k+1}^{\infty}\left(\frac{1}{x}-\log\left(1+\frac{1}{x}\right)\right)\\ &\leq |\log(k+1)-\log k|+\sum_{x=k+1}^{\infty}\frac{1}{2x^2}\\ &\leq \left|\log\left(1+\frac{1}{k}\right)\right|+\int_k^{\infty}\frac{1}{2t^2}\,\mathrm{d}t\\ &\leq \frac{1}{k}+\frac{1}{2k}\\ &= \frac{2}{3k}.\end{aligned}$$

■

Para o próximo resultado considere uma cadeia de Markov $(X_t)_{t\geq 0}$ em tempo contínuo como as da Definição 2.2.15 com espaço de estados E.

Lema A.3.6. *Seja $F : \mathbb{R}_+ \times E \to \mathbb{R}$ uma função limitada, de modo que para cada $x \in E$ $F(\cdot, x)$ é de classe C^2 e existe constante $C > 0$ tal que*

$$\sup_{(s,x)}\left|(\partial_s^1 F)(s,x)\right| + \sup_{(s,x)}\left|(\partial_s^2 F)(s,x)\right| \leq C.$$

O processo $(M^F(t))_{t\geq 0}$ definido por

$$M^F(t) := F(t, X_t) - F(0, X_0) - \int_0^t (\partial_s + L)F(s, X_s)\,\mathrm{d}s$$

é um martingal com respeito à $(\mathcal{F}_t : t \geq 0)$, onde $(\mathcal{F}_t : t \geq 0)$ é a filtração natural $\mathcal{F}_t = \sigma(X_s : t \geq s \geq 0)$.

Dem.: Lema 5.1 do apêndice 1 de [KL10].

■

B
Construção de sequências de variáveis aleatórias i.i.d. e de cadeias de Markov

Antes de enunciar o primeiro fixaremos uma notação. Denotaremos por $\mathcal{R}^{\mathbb{N}}$ a σ-álgebra de subconjuntos de $\mathbb{R}^{\mathbb{N}}$ gerada pela coleção $\mathcal{A}$ dada por

$$\mathcal{A} := \Big\{ (a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_n, b_n] \times \mathbb{R} \times \mathbb{R} \times \cdots \Big| \\ n \in \mathbb{N}^* \text{ e } (a_1, b_1], (a_2, b_2], \ldots, (a_n, b_n] \text{ são intervalos em } \mathbb{R} \Big\}.$$

Observação B.0.1. É fácil de ver que para qualquer $n \in \mathbb{N}^*$ tem-se $M \times \mathbb{R} \times \mathbb{R} \times \cdots \in \mathcal{R}^{\mathbb{N}}$ para todo $M \in \mathcal{B}(\mathbb{R}^n)$ visto que $\mathcal{B}(\mathbb{R}^n)$ é gerado por retângulos n-dimensionais. Este detalhe será importante no futuro.

Teorema B.0.2 (Teorema da Extensão de Kolmogorov)**.** *Suponha que para cada $n \in \mathbb{N}^*$ se tenha uma medida μ_n definida em $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Se para qualquer $n \in \mathbb{N}^*$, e quaisquer intervalos $(a_1, b_1], (a_2, b_2], \cdots, (a_n, b_n]$ tem se*

$$\mu_{n+1}\left((a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_n, b_n] \times \mathbb{R}\right) = \mu_n\left((a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_n, b_n]\right),$$

então existe uma medida $\mathbb{P}$ em $(\mathbb{R}^{\mathbb{N}}, \mathcal{R}^N)$ de modo que

$$\mathbb{P}\left((a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_n, b_n] \times \mathbb{R} \times \mathbb{R} \times \cdots\right) = \mu_n\left((a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_n, b_n]\right)$$

para qualquer $n \in \mathbb{N}^*$, e quaisquer intervalos $(a_1, b_1], (a_2, b_2], \cdots, (a_n, b_n]$.

Dem.: Consultar apêndice 3 de [Dur19]. ■

Dada uma sequência $X_1, X_2, \ldots$ de variáveis aleatórias reais, denotemos por μ_t a distribuição de X_t. Como cada μ_t é uma medida em $\mathbb{R}$, para cada n podemos definir a medida produto

$$\nu_n = \mu_1 \otimes \cdots \otimes \mu_n,$$

onde

$$\nu_n(A_1 \times \cdots \times A_n) = \prod_{k=1}^{n} \mu_k(A_k)$$

para quaisquer $A_1, \ldots, A_n \in \mathcal{B}(R)$.
Assim, pelo Teorema da Extensão de Kolmogorov, existe uma medida $\mathbb{P}$ em $(\mathbb{R}^{\mathbb{N}}, \mathcal{R}^{\mathbb{N}})$ de modo que

$$\mathbb{P}\left((a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_n, b_n] \times \mathbb{R} \times \mathbb{R} \times \cdots\right) = \mu_n\left((a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_n, b_n]\right)$$

para qualquer $n \in \mathbb{N}^*$, e quaisquer intervalos $(a_1, b_1], (a_2, b_2], \cdots, (a_n, b_n]$.
Definamos a sequência $(Y_n)_{n\in\mathbb{N}^*}$ de variáveis aleatórias em $(\mathbb{R}^{\mathbb{N}}, \mathcal{R}^{\mathbb{N}}, \mathbb{P}$ por

$$Y_n((\omega_1, \omega_2, \ldots)) = \omega_n$$

para todo n, e para todo $(\omega_1, \omega_2, \ldots) \in \mathbb{R}^{\mathbb{N}}$.
Observe que para cada $n \in \mathbb{N}^*$ temos

$$\begin{aligned}
\mathbb{P}(Y_1 \in I_1, \ldots, Y_n \in I_n) &= \mathbb{P}(I_1 \times \cdots \times I_n \times \mathbb{R} \times \cdots) \\
&= \prod_{k=1}^{n} \mu_k(I_k)
\end{aligned}$$

para quaisquer intervalos $I_1, \ldots, I_k$ de $\mathbb{R}$. Assim, particularmente as variáveis aleatórias $Y_1, Y_2, \ldots$ é independente. Além disso, como $\mathbb{P}(Y_k \in I_k) = \mu_k(I_k)$ para qualquer intervalo I_k, e $\mathcal{B}(\mathbb{R})$ é gerada pelos intervalos, segue de Corolário A.1.12 que X_k e Y_k possuem a mesma distribuição.

Lema B.0.3. *As variáveis aleatórias X_k $(k \in \mathbb{N})$ definidas no Exemplo 1.1.7 são mensuráveis.*

Dem.: Note que X_0 é claramente mensurável, uma vez que é composição de funções mensuráveis. Suponha que X_n seja mensurável para algum $n \in \mathbb{N}$.

Fixado um estado z, observe que

$$X_{n+1}^{-1}(z) = \{\omega \mid F(\xi_n(\omega), X_n(\omega))\} = \bigcup_{a\in E} \left\{X_n^{-1}(a) \cap (f_a \circ \xi_n)^{-1}(z)\right\}.$$

Assim, pelo fato de X_n, f_a e ξ_n serem mensuráveis, por consequência X_{n+1} também o é.

Portanto, segue do princípio de indução matemática que X_t é mensurável para todo $t \in \mathbb{N}$.

■

B.1 Construção de Cadeias de Markov em tempo contínuo

Nesta seção, mostraremos que o processo definido em Definição 2.2.15 é de fato uma cadeia de Markov em tempo contínuo. Recordando a definição:

Definição B.1.1 (Cadeia de Markov em tempo contínuo gerada por Q)**.** Um processo $(X_t)_{t\geq 0}$ minimal e contínuo à direita assumindo valores em I é uma cadeia de Markov com distribuição inicial λ e matriz geradora Q se sua cadeia de saltos $(Y_n)_{n\in\mathbb{N}}$ é uma cadeia de Markov em tempo discreto com distribuição inicial λ e matriz de transição A, e para cada $n \in \mathbb{N}_+$, condicionado à $Y_0, ..., Y_{n-1}$, os seus tempos de espera $S_1, ..., S_n$ são variáveis aleatórias exponenciais independentes com parâmetros $q(Y_0), ..., q(Y_{n-1})$ respectivamente. Em outras palavras

$$X_t = \sum_{n=0} \mathbf{1}_{[J_n, J_{n+1})} Y_n.$$

Observação B.1.2. Na definição acima ao invés de definir $S_1, S_2, \ldots$ como sendo exponenciais de parâmetros $q(Y_0), q(Y_1), \ldots$, poderíamos definir para cada $k \in \mathbb{N}^*$

$$S_k := \frac{\xi_k}{q(Y_{k-1})},$$

onde $\xi_1, \xi_2, \ldots$ são variáveis aleatórias exponenciais com parâmetro 1. Isso segue do fato de $p\,\mathrm{Exp}(p) \sim \mathrm{Exp}(1)$ para todo $p > 0$.

Usaremos a mudança descrita na observação acima.

Observação B.1.3. Na ocasião da definição não especificamos uma filtração, portanto, consideraremos o processo definido no espaço $(\Omega, \mathcal{F}, \mathbb{P})$, onde $\mathcal{F} := \sigma(X_t : t \geq 0)$ com a filtração natural $(\mathcal{F}_t)_{t\geq 0}$, onde $\mathcal{F}_s := \sigma(X_t : 0 \leq s \leq t)$.

Proposição B.1.4. *O processo estocástico em tempo contínuo descrito na Definição 2.2.15 satisfaz a propriedade de Markov. Isto é*

$$\mathbb{P}(X_{t+s} = \xi \mid \mathcal{F}_t) = \mathbb{P}(X_{t+s} = \xi \mid X_t),$$

para quaisquer $t, s \geq 0$ e ξ no espaço de estados E. Particularmente,

$$\mathbb{P}(X_{t+s} = \xi, X_t = \alpha) = \mathbb{P}_\alpha(X_s = \xi)\mathbb{P}(X_t = \alpha)$$

o que garante

$$\mathbb{P}(X_{t+s} = \xi \mid X_t = \alpha) = \mathbb{P}_\alpha(X_s = \xi)$$

quando $\mathbb{P}(X_t = \alpha) > 0$.

Dem.: Fixados $t, s \geq 0$ e $\xi \in E$, definamos a função $P^s(X_t, \xi)$ por

$$P^s(X_t, \xi)(\omega) := \mathbb{P}_{X_t(\omega)}(X_s = \xi).$$

Pela definição de probabilidade condicional, provar a igualdade do enunciado é equivalente a mostrar que

$$\mathbb{E}\left[\mathbf{1}\{X_{t+s} = \xi\} \mid \mathcal{F}_t\right] = \mathbb{E}\left[\mathbf{1}\{X_{t+s} = \xi\} \mid X_t\right]. \tag{B.1}$$

Inicialmente mostraremos que

$$\mathbb{E}\left[\mathbf{1}\{X_{t+s} = \xi\} \mid \mathcal{F}_t\right] = P^s(X_t, \xi), \tag{B.2}$$

e posteriormente mostraremos que isso implica (B.1).

Como pela definição de esperança condicional temos

$$\mathbb{P}(X_{t+s} = \xi \mid \mathcal{F}_t) = \mathbb{E}\left[\mathbf{1}\left\{X_{t+s} = \xi\right\} \mid \mathcal{F}_t\right],$$

para mostrar a igualdade principal é suficiente mostrar que se tem

$$\int \mathbf{1}_{\{A\}}\mathbf{1}\{X_{t+s} = \xi\}\, \mathrm{d}\mathbb{P} = \int \mathbf{1}_{\{A\}} P^s(X_t, \xi)\, \mathrm{d}\mathbb{P} \tag{B.3}$$

para qualquer $A \in \mathcal{F}_t$.

Note que os conjuntos da forma $\{S_1 > a_1, \ldots, S_l > a_l, Y_0 = y_0, \ldots, Y_l = y_l, X_l = \zeta\}$ com $l \in \mathbb{N}$, $a_1, \ldots a_k \geq 0$ e $y_0, y_1, \ldots, y_l \in E$ de modo que $J_k < t$, formam um π-sistema Ç.
Além disso, observe que $\mathcal{F}_t \subseteq \sigma(\text{Ç})$. Portanto, pelo Teorema λ-π de Dynkin (Teorema A.1.11) e pela Observação A.1.13, basta verificar que a igualdade (B.2) vale para qualquer conjunto com a forma descrita acima.

Assim, fixado um conjunto $A = \{S_1 > a_1, \ldots, S_l > a_l, Y_0 = y_0, \ldots, Y_l = y_l, X_l = \zeta\}$, observe que

$$\mathbb{P}\left(A, \{X_{t+s} = \xi\}\right) = \sum_{k \geq 0} \mathbb{P}\left(A, \{X_{t+s} = \xi\}, J_k \leq t < J_{k+1}\right).$$

Analisaremos a parcela $\mathbb{P}\left(A, \{X_{t+s} = \xi\}, J_k \leq t < J_{k+1}\right)$ da soma. Observe que pela construção do processo, o valor de X_{t+s} pode ser escrito como

$$X_{t+s} = \sum_{n=0} \mathbf{1}\{t + s \in [J_n, J_{n+1})\} Y_n.$$

Porém como sabemos que $s \in [J_k, J_k + 1)$, e os tempos de espera são independentes, podemos escrever

$$X_{t+s} = \begin{cases} Y_k, & \text{se } s \in \Big[0\ ,\ S_{k+1} - (t - J_k)\Big); \\ Y_{k+1}, & \text{se } s \in \Big[\widetilde{S}_{k+1}\ ,\ \widetilde{S}_{k+1} + S_{k+2}\Big); \\ Y_{k+2}, & \text{se } s \in \Big[\widetilde{S}_{k+1} + S_{k+2}\ ,\ \widetilde{S}_{k+1} + S_{k+2} + S_{k+2}\Big); \\ & \qquad \vdots \end{cases},$$

onde $\widetilde{S}_{k+1} = S_{k+1} - (t - J_k)$.

Ou seja, podemos escrever

$$X_{t+s} = F\left(s, \{Y_k, Y_{k+1}, \ldots\}, \{J_{k+1} - (t - J_k), J_{k+2}, \ldots\}\right),$$

onde F é a função descrita acima. Para simplificar a notação, usaremos $f(W)$ para denotar

$$F\left(s, \{Y_k, Y_{k+1}, \ldots\}, \{W, J_{k+2}, \ldots\}\right),$$

para qualquer variável aleatória W. Além disso, note também que

$$\mathbb{P}\left(A,\{X_{t+s}=\xi\},J_k\le t<J_{k+1}\right)=\mathbb{P}\left(B,\{Y_k=\zeta\},\{X_{t+s}=\xi\},J_k\le t<J_{k+1}\right),$$

onde $B=\{X_{J_1}=y_1,X_{J_2}=y_2,\dots,X_{J_l}=y_l\}$, em outras palavras, B é o conjunto A sem a condição $X_t=\zeta$.

Defina a sigma álgebra $\mathcal{H}_k=\sigma\left(Y_0,Y_1,\dots,Y_k,S_0,S_1,\dots S_k\right)$. Observe que como $t\in[J_k,J_{k+1})$ e $t\ge J_l$, tem-se obrigatoriamente $k\ge l$, assim, B é $\mathcal{H}_k$-mensurável. Logo, temos

$$\begin{aligned}
&\mathbb{P}\left(A,\{X_{t+s}=\xi\},J_k\le t<J_{k+1}\right)\\
&\qquad=\mathbb{E}\left[\mathbf{1}_B\mathbf{1}_{\{Y_k=\zeta\}}\mathbf{1}\{X_{t+s}=\xi\}\mathbf{1}\left\{J_k\le t<J_{k+1}\right\}\right]\\
&\qquad=\mathbb{E}\left[\mathbf{1}_B\mathbf{1}_{\{Y_k=\zeta\}}\mathbf{1}\{f(S_{k+1}-(t-J_k))=\xi\}\mathbf{1}_{\{J_k\le t<J_{k+1}\}}\right]\\
&\qquad=\mathbb{E}\left[\mathbf{1}_B\mathbf{1}_{\{Y_k=\zeta\}}\mathbf{1}\{f(S_{k+1}-(t-J_k))=\xi\}\mathbf{1}_{\{J_k\le t\}}\mathbf{1}_{\{t-J_k<S_{k+1}\}}\right]\\
&\qquad=\mathbb{E}\left[\mathbf{1}_B\mathbf{1}_{\{Y_k=\zeta\}}\mathbf{1}\left\{\overline{f}(S_{k+1}-(t-J_k))=\xi\right\}\mathbf{1}_{\{J_k\le t\}}\mathbf{1}_{\{t-J_k<S_{k+1}\}}\right],
\end{aligned}\tag{B.4}$$

onde $\overline{f}(W)=F\left(s,\{\zeta,Y_{k+1},\dots\},\{W,J_{k+2},\dots\}\right)$, em outras palavras é a função f com a informação de que $Y_k=\zeta$.

Assim, utilizando o Lema A.2.5 temos

$$\begin{aligned}
&\mathbb{P}\left(A,\{X_{t+s}=\xi\},J_k\le t<J_{k+1}\right)\\
&\qquad=\mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_B\mathbf{1}_{\{Y_k=\zeta\}}\mathbf{1}\left\{\overline{f}(S_{k+1}-(t-J_k))=\xi\right\}\mathbf{1}_{\{J_k\le t\}}\mathbf{1}_{\{t-J_k<S_{k+1}\}}\mid\mathcal{H}_k\right]\right]
\end{aligned}$$

que através do Teorema A.2.9 nos fornece

$$\begin{aligned}
&\mathbb{P}\left(A,\{X_{t+s}=\xi\},J_k\le t<J_{k+1}\right)\\
&\qquad=\mathbb{E}\left[\mathbf{1}_B\mathbf{1}_{\{Y_k=\zeta\}}\mathbf{1}_{\{J_k\le t\}}\mathbb{E}\left[\mathbf{1}\left\{\overline{f}(S_{k+1}-(t-J_k))=\xi\right\}\mathbf{1}_{\{t-J_k<S_{k+1}\}}\mid\mathcal{H}_k\right]\right].
\end{aligned}\tag{B.5}$$

Segue de Proposição A.2.11 que

$$\mathbb{E}\left[\mathbf{1}\left\{\overline{f}(S_{k+1}-(t-J_k))=\xi\right\}\mathbf{1}_{\{t-J_k<S_{k+1}\}}\mid\mathcal{H}_k\right]=\mathbb{E}\left[\mathbf{1}\left\{\overline{f}(S_{k+1}-(t-s))=\xi\right\}\mathbf{1}_{\{t-s<S_{k+1}\}}\right]\Big|_{s=J_k}.$$

e Propriedade da Perda de Memória II implica em

$$\mathbb{E}\left[\mathbf{1}\left\{\overline{f}(S_{k+1}-(t-J_k))=\xi\right\}\mathbf{1}_{\{t-J_k<S_{k+1}\}}\mid\mathcal{H}_k\right]=\mathrm{e}^{-q(\zeta)(t-J_k)}\,\mathbb{E}\left[\mathbf{1}\left\{\overline{f}(\hat{S}_{k+1})=\xi\right\}\right],\tag{B.6}$$

onde $\hat{S}_{k+1}$ é uma variável aleatória com a mesma distribuição de S_{k+1} e independente das variáveis $S_0,S_1,\dots$

Assim, substituindo (B.6) em (B.5), obtemos

$$\begin{aligned}
&\mathbb{P}\left(A,\{X_{t+s}=\xi\},J_k\le t<J_{k+1}\right)\\
&\qquad=\mathbb{E}\left[\mathbf{1}_B\mathbf{1}_{\{Y_k=\zeta\}}\mathbf{1}_{\{J_k\le t\}}\,\mathrm{e}^{-q(\zeta)(t-J_k)}\,\mathbb{E}\left[\mathbf{1}\left\{\overline{f}(\hat{S}_{k+1})=\xi\right\}\right]\right]\\
&\qquad=\mathbb{E}\left[\mathbf{1}_B\mathbf{1}_{\{Y_k=\zeta\}}\mathbf{1}_{\{J_k\le t\}}\,\mathrm{e}^{-q(\zeta)(t-J_k)}\right]\mathbb{E}\left[\mathbf{1}\left\{\overline{f}(\hat{S}_{k+1})=\xi\right\}\right]
\end{aligned}$$

$$= \mathbb{E}\left[\mathbf{1}_B \mathbf{1}_{\{Y_k=\zeta\}} \mathbf{1}_{\{J_k \leq t\}} \, \mathrm{e}^{-q(\zeta)(t-J_k)}\right] P^s(\zeta, \xi) \tag{B.7}$$

Note que repetindo os passos de (B.4) até (B.6) com a função $G \equiv 1$ ao invés de f obtemos a igualdade

$$\mathbb{E}\left[\mathbf{1}_B \mathbf{1}_{\{Y_k=\zeta\}} \mathbf{1}_{\{J_k \leq t\}} \, \mathrm{e}^{-q(\zeta)(t-J_k)}\right] = \mathbb{E}\left[\mathbf{1}_B \mathbf{1}_{\{Y_k=\zeta\}} \mathbf{1}_{\{J_k \leq t < J_{k+1}\}}\right] \tag{B.8}$$

Portanto, substituindo (B.8) em (B.7) obtemos

$$\begin{aligned}
\mathbb{P}\left(A, \{X_{t+s} = \xi\}, J_k \leq t < J_{k+1}\right) &= \mathbb{E}\left[\mathbf{1}_B \mathbf{1}_{\{Y_k=\zeta\}} \mathbf{1}_{\{J_k \leq t < J_{k+1}\}}\right] P^s(\zeta, \xi) \\
&= \mathbb{E}\left[\mathbf{1}_B \mathbf{1}_{\{Y_k=\zeta\}} \mathbf{1}_{\{J_k \leq t < J_{k+1}\}} P^s(\zeta, \xi)\right] \\
&= \mathbb{E}\left[\mathbf{1}_A \mathbf{1}_{\{J_k \leq t < J_{k+1}\}} P^s(\zeta, \xi)\right] \\
&= \mathbb{E}\left[\mathbf{1}_A \mathbf{1}_{\{J_k \leq t < J_{k+1}\}} P^s(X_t, \xi)\right].
\end{aligned}$$

Somando em $k \in \mathbb{N}$ obtemos

$$\mathbb{P}\left(A, \{X_{t+s} = \xi\}\right) = \mathbb{E}\left[\mathbf{1}_A P^s(X_t, \xi)\right], \tag{B.9}$$

que através da discussão inicial nos permite concluir que

$$\mathbb{E}\left[\mathbf{1}\{X_{t+s} = \xi\} \mid \mathcal{F}_t\right] = P^s(X_t, \xi)$$

Observe que a pré-imagem de qualquer conjunto mensurável por $P^s(X_t, \xi)$ é um conjunto da forma

$$\bigcup_{i \in I} \{X_t = y_k\},$$

com $I \subseteq \mathbb{N}$ e $y_i \in E$ para todo $i \in I$. Portanto, $P^s(X_t, \xi)$ é $\sigma(X_t)$-mensurável.

Sabemos que

$$\int \mathbf{1}_A \mathbf{1}\{X_{t+s} = \xi\} \, \mathrm{d}\mathbb{P} = \int \mathbf{1}_A P^s(X_t, \xi) \, \mathrm{d}\mathbb{P}$$

para qualquer $A \in \mathcal{F}_t$. Como $\sigma(X_t) \subseteq \mathcal{F}_t$, particularmente vale

$$\int \mathbf{1}_B \mathbf{1}\{X_{t+s} = \xi\} \, \mathrm{d}\mathbb{P} = \int \mathbf{1}_B P^s(X_t, \xi) \, \mathrm{d}\mathbb{P}$$

para qualquer $B \in \mathcal{F}_t$.

Portanto, segue da definição de probabilidade condicional que

$$\mathbb{E}\left[\mathbf{1}\{X_{t+s} = \xi\} \mid \sigma(X_t)\right] = P^s(X_t, \xi) = \mathbb{E}\left[\mathbf{1}\{X_{t+s} = \xi\} \mid \mathcal{F}_t\right],$$

o que finalmente nos permite concluir que o processo $(X_t)_{t \geq 0}$ é uma cadeia de Markov.

Já para a segunda parte, tomando $A = \{X_t = \zeta\}$, segue da igualdade acima que

$$\mathbb{E}\left[\mathbf{1}_{\{X_t=\zeta\}} \mathbf{1}\{X_{t+s} = \xi\}\right] = \mathbb{E}\left[\mathbf{1}_{\{X_t=\zeta\}} P^s(X_t, \xi)\right]$$

o que implica

$$\mathbb{E}\left[\mathbf{1}_{\{X_t=\zeta\}} \mathbf{1}\{X_{t+s} = \xi\}\right] = \mathbb{E}\left[\mathbf{1}_{\{X_t=\zeta\}} P^s(\zeta, \xi)\right] = \mathbb{P}(X_t = \zeta) P^s(\zeta, \xi),$$

que é suficiente para concluir que

$$\mathbb{P}(X_{t+s} = \xi, X_t = \zeta) = \mathbb{P}(X_t = \zeta)P^s(\zeta, \xi).$$

■

Agora direcionaremos os nosso esforços para provar que o processo também satisfaz a propriedade forte de Markov.

Proposição B.1.5. *se $S \leq T$ são dois tempos de parada para uma dada cadeia com filtração $(\mathcal{F}_t)_{t\geq 0}$ então $F_S \subseteq F_T$.*

Dem.: Fixado $A \in \mathcal{F}_S$, segue da definição que $A \cap \{S \leq t\} \in \mathcal{F}_t$ para todo $t \geq 0$. Além disso, como T é um tempo de parada para a cadeia, por definição $\{T \leq t\} \in \mathcal{F}_t$ para todo $t \geq 0$. Note que segue do fato de $S \leq T$ que

$$\{T \leq t\} \subseteq \{S \leq t\}.$$

Assim, vale

$$\{T \leq t\} \cap A \cap \{S \leq t\} = \{T \leq t\} \cap A,$$

e como $\mathcal{F}_t$ é uma σ-álgebra, segue que

$$\{T \leq t\} \in \mathcal{F}_t$$

para todo $t \geq 0$. Logo, concluímos que $A \in \mathcal{F}_T$.

■

Observação B.1.6. Para o próximo resultado, utilizaremos o fato de $(X_t)_{t\geq 0}$ ser contínuo à direita sem fazer menções.

Lema B.1.7. *Se S, T são tempos de parada para $(X_t)_{t\geq 0}$, então X_T, $\{S < T\}$, $\{S > T\}$, $\{S \leq T\}$ e $\{S \geq T\}$ são $\mathcal{F}_T$-mensuráveis.*

Dem.: Note que fixado $t \geq 0$, temos

$$\begin{aligned}\{S > T\} \cap \{T \leq t\} &= \bigcup_{s \in \mathbb{Q} \cap [0,t]} (\{S > s\} \cap \{T \leq s\}) \\ &= \bigcup_{s \in \mathbb{Q} \cap [0,t]} (\{S \leq s\}^c \cap \{T \leq s\}).\end{aligned}$$

Como cada $(\{S \leq s\}^c \cap \{T \leq s\})$ pertence a $\mathcal{F}_s$, segue da definição de σ-álgebra em tempo de parada que $\{S > T\} \in \mathcal{F}_T$. Pelo fato de $\mathcal{F}_T$ ser uma σ-álgebra, particularmente

$$\{S \leq T\} = \{S > T\}^c \in \mathcal{F}_T.$$

De maneira semelhante, fixado $t \geq 0$, temos

$$\{S < T\} \cap \{T \leq t\} = \bigcup_{s \in \mathbb{Q} \cap [0,t]} (\{S \leq s\} \cap \{t \geq T\} \cap \{T > s\})$$

$$= \bigcup_{s \in \mathbb{Q}\cap[0,t]} \left(\{S \leq s\} \cap \{t \geq T\} \cap \{T \leq s\}^c\right).$$

Como por hipótese $\{S \leq s\} \cap \{T \leq s\}$ pertence a $\mathcal{F}_s$ e $s \leq t$, segue que

$$(\{S \leq s\} \cap \{t \geq T\} \cap \{T \leq s\}^c) \in \mathcal{F}_t.$$

Portanto, segue da definição de σ-álgebra em tempo de parada que $\{S < T\} \in \mathcal{F}_T$, e consequentemente $\{S \geq T\} \in \mathcal{F}_T$.

Já para a segunda parte, para mostrar que X_T é $\mathcal{F}_T$-mensurável, basta mostrar que $\{X_T = y\} \in \mathcal{F}_T$ para todo $y \in E$.
Fixado $y \in E$, observe que para todo $m \in \mathbb{N}$, existe $k_m \in \mathbb{N}$ tal que $T \in [(k_m - 1)2^{-m}, k_m 2^{-m})$.
Particularmente pelo fato do processo ser contínuo à direita, sob a condição $T < t$, existe um $m \in \mathbb{N}$ tal que todo $l \geq m$ natural admite $k_l \in \mathbb{N}$ com $k_l 2^{-l} \leq t$ de modo que

$$X_T = X_{k_l 2^{-l}} = y \quad \text{e} \quad T \in \left[\frac{k_l - 1}{2^l}, \frac{k_l}{2^l}\right).$$

Note que reciprocamente, se algum $\omega \in \Omega$ satisfaz as condições acima, então $\omega \in \{X_T = y, T < t\}$.
Portanto, fixado $t \geq 0$, temos

$$\begin{aligned}
&\big\{X_T = y, T \leq t\big\} \\
&\quad = \{X_t = y, T = t\} \bigcup \{X_T = y, T < t\} \\
&\quad = \{X_t = y, T = t\} \bigcup \left(\bigcup_{m=0}^{\infty} \bigcap_{l=m}^{\infty} \bigcup_{k=1}^{\lfloor 2^{-l} t \rfloor} \left\{ X_{k2^{-l}} = y,\, T \in \left[\frac{k-1}{2^m}, \frac{k}{2^m}\right) \right\} \right) \\
&\quad = \{X_t = y, T = t\} \bigcup \left(\bigcup_{m=0}^{\infty} \bigcap_{l=m}^{\infty} \bigcup_{k=1}^{\lfloor 2^{-l} t \rfloor} \left(\{X_{k2^{-l}} = y\} \cap \left\{ T \leq \frac{k}{2^m} \right\} \cap \left\{ T < \frac{k-1}{2^m} \right\}^c \right) \right).
\end{aligned}$$

Como $\{X_t = y\}$, $\{T = t\}$, $\{X_{k2^{-l}} = y\}$, $\left\{T \leq \frac{k}{2^m}\right\}$, $\left\{T < \frac{k-1}{2^m}\right\}^c$ pertencem à $\mathcal{F}_t$, concluímos que $\big\{X_T = y, T \leq t\big\} \in \mathcal{F}_t$.
Assim, segue da definição de σ-álgebra em tempos de parada que X_T é $\mathcal{F}_T$-mensurável. ■

Proposição B.1.8. *O processo definido em Definição 2.2.15 satisfaz a propriedade forte de Markov. Isto é, fixado um tempo de parada T, para qualquer $s \geq 0$ tem-se*

$$\mathbb{P}(X_{T+s} = \xi \mid \mathcal{F}_T) = \mathbb{P}(X_{T+s} = \xi \mid X_T)$$

no conjunto $\{T < \infty\}$. Particularmente,

$$\mathbb{P}(X_{T+s} = \xi, X_T = \alpha, T < \infty) = \mathbb{P}(X_T = \alpha, T < \infty)\mathbb{P}_\alpha(X_s = \xi),$$

o que garante que

$$\mathbb{P}(X_{T+s} = \xi \mid X_T = \alpha, T < \infty) = \mathbb{P}_\alpha(X_s = \xi)$$

se $\mathbb{P}(X_T = \alpha) > 0$.

Dem.: Adotaremos a mesma estratégia da demonstração anterior. Primeiro mostraremos que

$$\mathbb{P}(X_{T+s} = \xi \mid \mathcal{F}_T) = P^s(X_T, \xi),$$

e posteriormente mostraremos que isso implica na desigualdade do enunciado.

Procuramos mostrar que para todo $A \in \mathcal{F}_T$, tem-se

$$\mathbb{E}\left[\mathbf{1}_A \mathbf{1}\left\{X_{T+s} = \xi\right\} \mathbf{1}_{T<\infty}\right] = \mathbb{E}\left[\mathbf{1}_A P^s(X_T, \xi)\mathbf{1}_{T<\infty}\right].$$

Defina para cada n os tempos de parada T_n por

$$T_n = \sum_{k=0}^{\infty} \frac{k+1}{2^n} \mathbf{1}_{\{T \in I(n,k)\}},$$

onde $I(n,k) = \left(\frac{k}{2^n}, \frac{(k+1)}{2^n}\right]$.

Observe que $\{T < \infty\} = \{T_n < \infty\}$, $T_n \geq T$ e além disso, $T_n \downarrow T$.

Fixado $n \in \mathbb{N}^*$ note que

$$\begin{aligned}
\mathbb{E}\left[\mathbf{1}_A \mathbf{1}\{X_{T_n+s} = \xi\}\mathbf{1}_{T<\infty}\right] &= \mathbb{E}\left[\mathbf{1}_A \mathbf{1}\{X_{T_n+s} = \xi\}\mathbf{1}_{T_n<\infty}\right] \\
&= \sum_{k=0}^{\infty} \mathbb{E}\left[\mathbf{1}_A \mathbf{1}\left\{T_n = (k+1)2^{-n}\right\}\mathbf{1}\{X_{T_n+s} = \xi\}\right] \\
&= \sum_{k=0}^{\infty} \mathbb{E}\left[\mathbf{1}_A \mathbf{1}\left\{T_n = (k+1)2^{-n}\right\}\mathbf{1}\left\{X_{(k+1)2^{-n}+s} = \xi\right\}\right]
\end{aligned}$$

Note que como $T \leq T_n$ temos $\mathcal{F}_T \subseteq \mathcal{T}_n$, e como por hipótese A é $\mathcal{F}_T$-mensurável, segue que A é $\mathcal{F}_{T_n}$-mensurável. Portanto, segue da definição que $A \cap \{T_n = (k+1)2^{-n}\} \in \mathcal{F}_{(k+1)2^{-n}}$.

Assim, retomando as contas, pelo Lema A.2.5 temos

$$\begin{aligned}
\mathbb{E}\left[\mathbf{1}_A \mathbf{1}\{X_{T_n+s} = \xi\}\mathbf{1}_{T<\infty}\right] &= \sum_{k=0}^{\infty} \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_A \mathbf{1}\left\{T_n = \frac{(k+1)}{2^n}\right\}\mathbf{1}\left\{X_{(k+1)2^{-n}+s} = \xi\right\} \mid \mathcal{F}_{(k+1)/2^n}\right]\right] \\
&= \sum_{k=0}^{\infty} \mathbb{E}\left[\mathbf{1}_A \mathbf{1}\left\{T_n = \frac{(k+1)}{2^n}\right\}\mathbb{E}\left[\mathbf{1}\left\{X_{(k+1)/2^n+s} = \xi\right\} \mid \mathcal{F}_{(k+1)/2^n}\right]\right].
\end{aligned}$$

Usando a propriedade de Markov, obtemos

$$\begin{aligned}
\mathbb{E}\left[\mathbf{1}_A \mathbf{1}\{X_{T_n+s} = \xi\}\mathbf{1}_{T<\infty}\right] &= \sum_{k=0}^{\infty} \mathbb{E}\left[\mathbf{1}_A \mathbf{1}\left\{T_n = \frac{(k+1)}{2^n}\right\}P^s\left(X_{(k+1)/2^n}, \xi\right)\right] \\
&= \mathbb{E}\left[\mathbf{1}_A P^s(X_{T_n}, \xi)\mathbf{1}_{\{T_n<\infty\}}\right] \\
&= \mathbb{E}\left[\mathbf{1}_A P^s(X_{T_n}, \xi)\mathbf{1}_{\{T<\infty\}}\right]
\end{aligned}$$

Observe que como o processo é contínuo à direita e $T_n \downarrow T$ temos

$$\lim_{n\to\infty} \mathbf{1}\{X_{T_n+s} = \xi\} = \mathbf{1}\{X_{T+s} = \xi\}.$$

Portanto, pelo Teorema da Convergência Dominada, segue que

$$\mathbb{E}\left[\mathbf{1}_A \mathbf{1}\{X_{T+s} = \xi\}\mathbf{1}_{T<\infty}\right] = \lim_{n\to\infty} \mathbb{E}\left[\mathbf{1}_A \mathbf{1}\{X_{T_n+s} = \xi\}\mathbf{1}_{T<\infty}\right] = \lim_{n\to\infty} \mathbb{E}\left[\mathbf{1}_A P^s(X_{T_n}, \xi)\mathbf{1}_{\{T<\infty\}}\right]$$

Como o espaço de estados é enumerável, $P^s(\cdot, \xi)$ é trivialmente contínua, o que nos leva a concluir que

$$\lim_{n\to\infty} P^s(X_{T_n}, \xi) = P^s(X_T, \xi).$$

Portanto, novamente pelo Teorema da Convergência Dominada temos

$$\lim_{n\to\infty} \mathbb{E}\left[\mathbf{1}_A P^s(X_{T_n}, \xi)\mathbf{1}_{\{T<\infty\}}\right] = \mathbb{E}\left[\mathbf{1}_A P^s(X_T, \xi)\mathbf{1}_{\{T<\infty\}}\right].$$

Assim,

$$\mathbb{E}\left[\mathbf{1}_A \mathbf{1}\left\{X_{T+s} = \xi\right\}\mathbf{1}_{T<\infty}\right] = \mathbb{E}\left[\mathbf{1}_A P^s(X_T, \xi)\mathbf{1}_{\{T<\infty\}}\right],$$

e como tomamos $A \in \mathcal{F}_T$ qualquer, segue que

$$\mathbb{P}(X_{T+s} = \xi \mid \mathcal{F}_T) = P^s(X_T, \xi)$$

em $\{T < \infty\}$.

Repetindo o argumento da demonstração da propriedade de Markov, concluímos que

$$\mathbb{P}(X_{T+s} = \xi \mid X_T) = P^s(X_T, \xi),$$

e consequentemente

$$\mathbb{P}(X_{T+s} = \xi \mid X_T) = P^s(X_T, \xi) = \mathbb{P}(X_{T+s} = \xi \mid \mathcal{F}_T)$$

em $\{T < \infty\}$.

Já para a segunda parte, tomando $A = \{X_T = \zeta\}$, segue da igualdade acima que

$$\mathbb{E}\left[\mathbf{1}_{\{T<\infty\}}\mathbf{1}_{\{X_T=\zeta\}}\mathbf{1}\{X_{T+s} = \xi\}\right] = \mathbb{E}\left[\mathbf{1}_{\{T<\infty\}}\mathbf{1}_{\{X_T=\zeta\}}P^s(X_T, \xi)\right]$$

o que implica

$$\mathbb{E}\left[\mathbf{1}_{\{T<\infty\}}\mathbf{1}_{\{X_T=\zeta\}}\mathbf{1}\{X_{T+s} = \xi\}\right] = \mathbb{E}\left[\mathbf{1}_{\{T<\infty\}}\mathbf{1}_{\{X_T=\zeta\}}P^s(\zeta, \xi)\right] = \mathbb{P}(X_T = \zeta, T < \infty)P^s(\zeta, \xi),$$

que é suficiente para o resultado final. ∎

Bibliografia

[Bar95] R.G. Bartle. *The Elements of Integration and Lebesgue Measure*. A Wiley-interscience publication. Wiley, 1995.

[BHP17] Riddhipratim Basu, Jonathan Hermon, and Yuval Peres. Characterization of cutoff for reversible markov chains. *The Annals of Probability*, 45(3), May 2017.

[Dia96] P Diaconis. The cutoff phenomenon in finite markov chains. *Proceedings of the National Academy of Sciences*, 93(4):1659–1664, 1996.

[Dur19] R. Durrett. *Probability: Theory and Examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2019.

[GW14] G. Grimmett and D. Welsh. *Probability: An Introduction*. OUP Oxford, 2014.

[KL10] Claude Kipnis and Claudio Landim. *Scaling limits of interacting particle systems*. Grundlehren der mathematischen Wissenschaften. Springer, Berlin, Germany, October 2010.

[Lac16] Hubert Lacoin. The cutoff profile for the simple exclusion process on the circle. *The Annals of Probability*, 44(5), September 2016.

[LL11] Hubert Lacoin and Remi Leblond. Cutoff phenomenon for the simple exclusion process on the complete graph, 2011.

[LPW] D.A. Levin, Y. Peres, and E.L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society.

[Nor98] J.R. Norris. *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.