UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
DEPARTAMENTO DE MATEMÁTICA

VITOR SCHROEDER DOS ANJOS

UM ESTUDO SOBRE MODELOS MARKOVIANOS

SÃO CARLOS
2023

VITOR SCHROEDER DOS ANJOS

UM ESTUDO SOBRE MODELOS MARKOVIANOS

Monografia apresentada ao Curso de Bacharelado em Matemática da Universidade Federal de São Carlos.

Orientador: Prof. Dr. Márcio Luis Lanfredi Viola

SÃO CARLOS
2023

RESUMO

Cadeias de Markov possuem grande aplicabilidade em modelagem e, devido à sua importância, são bastante estudadas na literatura. Diante disto, o objetivo deste trabalho é estudar cadeias de Markov a tempo discreto de ordem um, com espaço de estados enumerável, apresentando os principais conceitos e propriedades, dando especial atenção à interpretação. Particularmente, definimos e exploramos vários aspectos associados à distribuição estacionária.

Palavras-chave: Cadeias de Markov. Distribuição Estacionaria. Modelagem.

ABSTRACT

Markov chains have a range of applications within modeling, and due to their importance, are greatly studied in the literature. Herein, the objective of this paper is to study first-order Markov chains, presenting the principal concepts and properties related to them, with special attention given to their interpretation. In particular, we will define and explore several aspects associated with the stationary distribution.

Keywords: Markov Chain. Stationary Distribution. Modeling.

LISTA DE FIGURAS

Figura 2.1 – Representação das trajetórias. 7

Figura 2.2 – Representação de uma função de densidade de $Z(t)$ em t_1, t_2, t_3. Fonte: Extraído de (Morettin , 2008, p. 26). 8

SUMÁRIO

1 INTRODUÇÃO 6

2 PROCESSO ESTOCÁSTICO 7

3 CADEIAS DE MARKOV DE ORDEM FIXA 9

4 CLASSIFICAÇÃO DE ESTADOS 15

5 DISTRIBUIÇÃO ESTACIONÁRIA 22

6 CONSIDERAÇÕES FINAIS 30

REFERÊNCIAS 31

1 INTRODUÇÃO

Na área de Processos Estocásticos, Cadeia de Markov é um modelo amplamente estudado e utilizado em modelagem, principalmente na Estatística, com aplicações em fenômenos físicos, biológicos, econômicos e sociais (DURRETT, 1999).

Em particular, árvores de contextos (LEONARDI, 2007; RISSANEN, 1983; BUHLMANN; WYNER, 1999) têm sido estudadas e utilizadas na modelagem de vários problemas práticos como, por exemplo, na análise de dados linguísticos (GALVES et al., 2002), na classificação de proteínas (LEONARDI, 2007; LEONARDI, 2006) e na identificação de genes (BUHLMANN; WYNER, 1999).

Os modelos Markovianos modelam uma sequência de variáveis aleatórias, especificando como é a dependência de uma variável em relação às demais, oferecendo uma abordagem probabilística para modelar sistemas dinâmicos que evoluem ao longo do tempo.

Por exemplo, considere uma variável aleatória X que represente o valor da ação de uma determinada empresa e que o índice t indique o tempo. Então, X_{t_1} é o valor da ação no tempo t_1, X_{t_2} é o valor da ação no tempo t_2, X_{t_3} é o valor da ação no tempo t_3, e assim por diante. Note que, em cada índice t, a variável representa a mesma "característica" (valor da ação) e, na linguagem probabilística, isto é formalizado por "variáveis aleatórias definidas em um mesmo espaço de probabilidade". Quando impomos que a sequência de variáveis aleatórias $\{X_{t_1}, X_{t_2}, ...\}$ é uma cadeia de Markov, estamos especificando como é a evolução temporal do valor da ação, definindo quantos tempos anteriores a t são importantes para a determinação da probabilidade de X_t dado as observações anteriores a t.

O modelo Markoviano mais simples considera a ordem igual a um. Neste caso, o valor da ação no tempo t depende apenas do valor da ação no tempo $t-1$. No entanto, o alcance da dependência pode ser um tempo atrás (X_t depende de X_{t-1}), dois tempos atrás (X_t depende apenas de X_{t-1} e X_{t-2}), e assim por diante.

As cadeias de Markov completas de ordem k estendem os processos Markovianos de ordem um e cadeia de Markov de memória variável (VLMC) (BUHLMANN; WYNER, 1999) estendem as cadeia de Markov de ordem k.

Há, ainda, extensões dos modelos VLMC como os modelos de partição mínima (GARCíA; GONZáLEZ-LóPEZ, 2011).

No entanto, o objetivo deste trabalho é estudar cadeias de Markov a tempo discreto de ordem um, com espaço de estados enumerável, apresentando os principais conceitos e propriedades, dando especial atenção à interpretação. Particularmente, definimos e exploramos vários aspectos associados à distribuição estacionária.

2 PROCESSO ESTOCÁSTICO

Definição 2.1. (Processo Estocástico) Um processo estocástico $\{X_t : t \in T\}$ é uma sequência de variáveis aleatórias indexadas pelo conjunto de índices T e definidas em um mesmo espaço de probabilidade.

O conjunto de índices, frequentemente, representa o tempo. Neste caso, podemos considerar um processo a tempo discreto ($T = \mathbb{Z}$ ou $T = \mathbb{Z}$) ou a tempo contínuo ($T = \mathbb{R}$).

No contexto de dados espaciais, o conjunto T representa o espaço. Por exemplo, se $T = \mathbb{R}^2$, as variáveis aleatórias são consideradas no plano cartesiano, isto é, a variável aleatória X_t representa a variável na posição $t = (x, y)$, $x, y \in \mathbb{R}$, do plano cartesiano.

Neste trabalho, consideramos que os índices representam o tempo discreto, especificamente, $T = \mathbb{Z}$. Informalmente, um processo estocástico descreve uma história que se desenvolve de forma aleatória ao longo do tempo.

Neste sentido, para cada tempo $t \in \mathbb{Z}$, X_t é uma variável aleatória e, assim, quaisquer valores de X_t podem ser observados no tempo t.

O conjunto formado pelos possíveis valores de X_t é chamado alfabeto e o denotamos por $\mathcal{A}$ finito ou infinito enumerável.

Cada uma das combinações dos possíveis valores da variável aleatória ao longo do tempo forma uma trajetória, conforme ilustra a Figura (2.1).

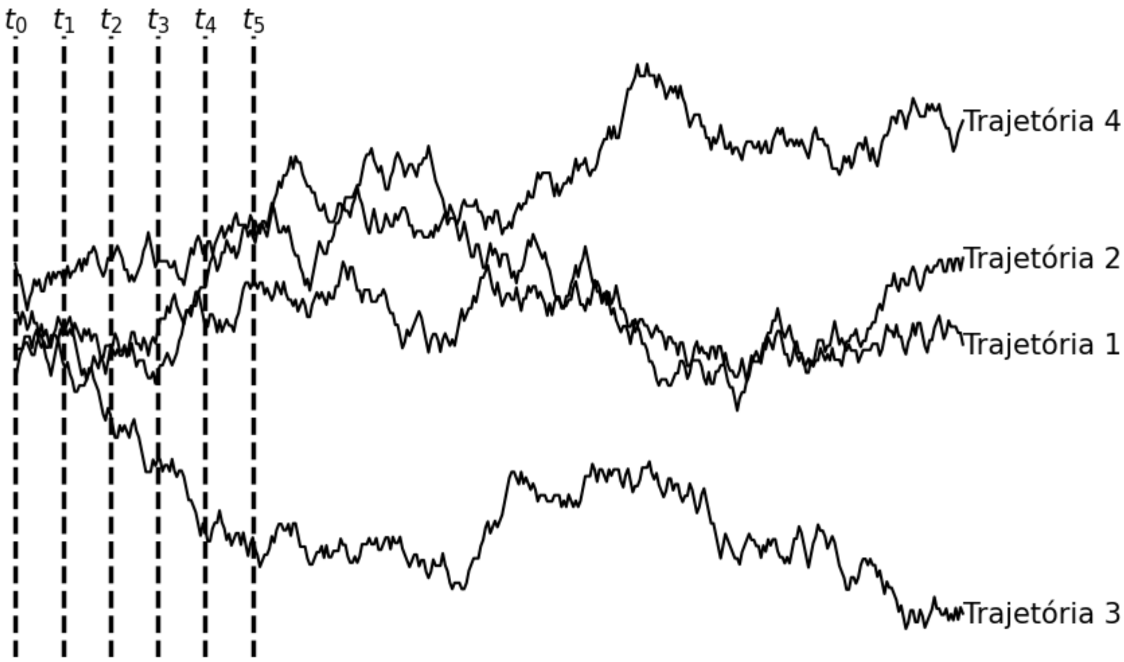


Figura 2.1 – Representação das trajetórias.

Na prática, observamos um único valor da variável aleatória em cada instante de tempo. Neste sentido, podemos entender a trajetória observada é uma das possíveis trajetórias determinada pelo processo estocástico.

Como para cada t, X_t é uma variável aleatória, há uma distribuição de probabilidade associada a cada t, como ilustrado na Figura (2.2).

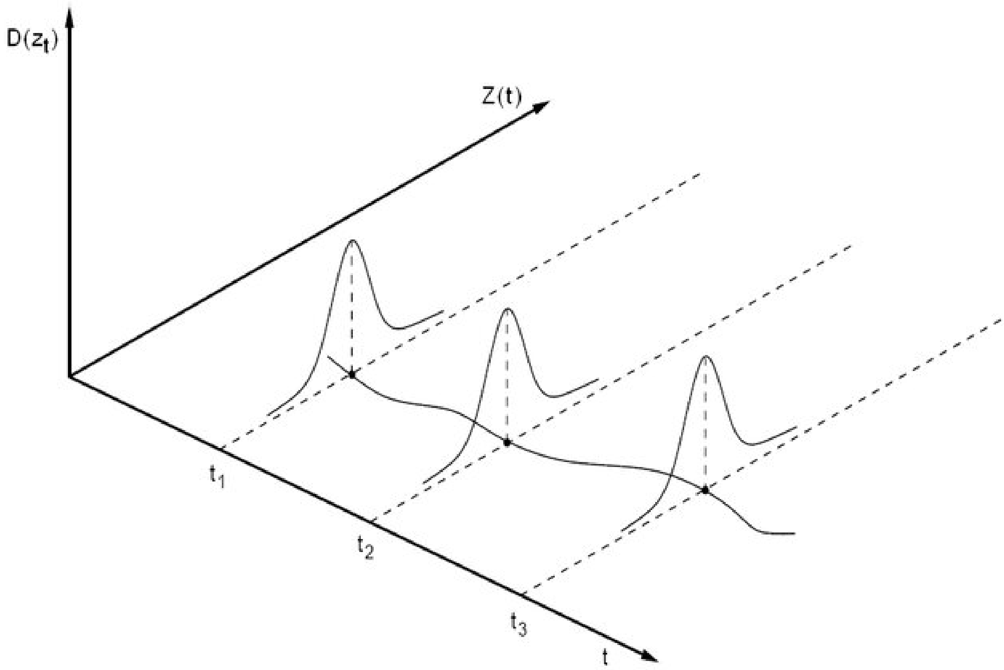


Figura 2.2 – Representação de uma função de densidade de $Z(t)$ em t_1, t_2, t_3.
Fonte: Extraído de (Morettin , 2008, p. 26).

Estamos interessados na determinação da probabilidade condicional $\mathbb{P}(X_t = x_t | X_{t-1} = x_{t-1}, \ldots, X_1 = x_1)$, em que x_i representa um dos possíveis valores de X_i, $i = 1, \ldots, t$.

Uma suposição utilizada é que a probabilidade $\mathbb{P}(X_t = x_t | X_{t-1} = x_{t-1}, \ldots, X_1 = x_1)$ não varia quando o tempo é transladado, ou seja, o processo é homogêneo. Neste trabalho, nos restringimos a processos de Markov homogêneos.

Definição 2.2. (Processo Homogêneo) O processo estocástico $\{X_t : t \in T\}$ é homogêneo se

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_t = x_t) = \mathbb{P}(X_{1+h} = x_1, X_{2+h} = x_2, \ldots, X_{t+h} = x_t),$$

para todo $h \in \mathbb{Z}$ e para quaisquer $x_1, \ldots, x_t \in \mathcal{A}$.

Se o processo for homogêneo, podemos trocar os índices ..., $t-n$, ..., $t-1, t$ pelos índices ..., $-n$, ..., $-1, 0$, $\forall t \in \mathbb{Z}$.

Uma classe muito importante de processos a tempo discreto corresponde aos processos de Markov ou cadeias de Markov. O foco deste trabalho é processos de Markov a tempo discreto de ordem um, cujo espaço de estados pode ser finito ou infinito enumerável.

3 CADEIAS DE MARKOV DE ORDEM FIXA

A ordem de uma Cadeia de Markov está relacionada ao alcance da dependência. Por exemplo, se a cadeia for de ordem um, X_t dependerá apenas de X_{t-1} (tempo imediatamente anterior a t). Em outras palavras, a previsão do próximo passo, conhecendo toda a história passada do processo desde o seu início é tão boa quanto a previsão feita conhecendo-se apenas o valor do processo no presente.

Se a ordem for igual a dois, a dependência de X_t envolverá o processo nos dois tempos imediatamente anteriores a t (X_{t-1} e X_{t-2}).

Definição 3.1. Um processo estocástico discreto $(X_t)_{t\in\mathbb{Z}}$ é um processo de Markov de ordem um, se

$$\mathbb{P}\big(X_t = x_t|X_{t-1} = x_{t-1}, X_{t-2} = x_{t-2}, \dots, X_1 = x_1, X_0 = x_0\big) = \mathbb{P}(X_t = x_t|X_{t-1} = x_{t-1}), \tag{3.1}$$

para todo $x_t \in \mathcal{A}$, $x_{t-1} \in \mathcal{A}, \dots, x_0 \in \mathcal{A}$ e para cada t.

Se o processo estocástico for homogêneo, isto é,

$$\mathbb{P}(X_t = x_t, ..., X_0 = x_0) = \mathbb{P}(X_{t+h} = x_{t+h}, ..., X_h = x_h),$$

para cada h e para quaisquer $x_t, \dots, x_0 \in \mathcal{A}$, a expressão (3.1) é válida para um deslocamento h no tempo.

A propriedade de Markov (3.1) afirma que a probabilidade de uma observação em t, condicionada a todas as observações nos tempos anteriores ($t-1, t-2, ...$) é igual a probabilidade da observação em t, dada a observação no tempo imediatamente anterior, $t-1$, para todo t.

As cadeias de Markov de ordem k, $k \in \mathbb{Z}$, estendem os processos Markovianos de ordem um.

Definição 3.2. Um processo estocástico discreto $(X_t)_{t\in\mathbb{Z}}$ é um processo de Markov de ordem k (ou processo de Markov completo de ordem k) se

$$\begin{aligned}&\mathbb{P}\big(X_t = x_t|X_{t-1} = x_{t-1}, X_{t-2} = x_{t-2}, \dots, X_{t-k} = x_{t-k}, \dots, X_0 = x_0\big)\\ &= \mathbb{P}\big(X_t = x_t|X_{t-1} = x_{t-1}, \dots, X_{t-k} = x_{t-k}\big),\end{aligned}$$

para todo $x_t \in \mathcal{A}$, $x_{t-1} \in \mathcal{A}, ..., x_0 \in \mathcal{A}$ e para cada t.

Vale observar que o número de parâmetros de uma cadeia de Markov cresce exponencialmente com a sua ordem.

O espaço de estados de uma cadeia de Markov é o conjunto $S = \mathcal{A}^k$. Caso a ordem do processo seja igual a um, o valor de X_t depende apenas do valor da variável aleatória em $t-1$,

X_{t-1}, ou seja, em relação à estrutura de dependência, devemos considerar apenas a variável aleatória X_{t-1} para predizer o valor no instante seguinte. Assim, o espaço de estados é formado pelos possíveis valores de X_{t-1}.

Se $\mathcal{A} = \{0, 1\}$ e ordem $k = 1$, os possíveis valores de X_{t-1} são 0 e 1 e, assim, o espaço de estados da referida cadeia de Markov de ordem 1 é $S = \{0, 1\}$ (cardinalidade igual a 2), que coincide com o alfabeto $\mathcal{A}$.

Caso a cadeia de Markov seja de ordem igual a dois, devemos considerar o vetor aleatório (X_{t-2}, X_{t-1}) para que possamos levar em consideração os instantes $t-1$ e $t-2$ na predição do valor do processo no instante seguinte (X_t), ou seja, o valor de X_t depende do valor das variáveis aleatórias em $t-1$ (X_{t-1}) e $t-2$ (X_{t-2}).

Se $\mathcal{A} = \{0, 1\}$ e ordem $k = 2$, os possíveis valores do vetor aleatório (X_{t-2}, X_{t-1}) é o espaço de estados da referida cadeia de Markov de ordem 2, a saber $S = \mathcal{A} \times \mathcal{A} = \{00, 01, 10, 11\}$, que possui cardinalidade igual a 4.

Nesta e na próxima seção, focamos no desenvolvimento teórico relacionado às cadeias de Markov homogêneas de ordem um, que são especificadas pelas probabilidades condicionais $\mathbb{P}(X_t = x_t|X_{t-1} = x_{t-1})$, $x_t, x_{t-1} \in S$, $S = \mathcal{A}$.

Definição 3.3. (Matriz de Transição) Uma matriz de transição P é uma matriz, de ordem $|S| \times |S|$, tal que $\sum_{j \in S} \mathbb{P}(X_t = j|X_{t-1} = i) = 1$, para cada $i \in S$.

De acordo com a Definição 3.3, as linhas e colunas da matriz de transição correspondem aos estados nos tempos $t-1$ e t, respectivamente.

A probabilidade de transição $\mathbb{P}(X_t = j|X_{t-1} = i)$, $i, j \in S$, é denotada por P_{ij} e corresponde ao elemento (i, j) da matriz P.

Uma cadeia de Markov de ordem um pode ser facilmente visualizada como um grafo direcionado, cujos vértices representam os estados e as arestas representam as probabilidades não-nulas de mudança entre estados.

Exemplo 3.1. O seguinte grafo considera o espaço de estados $S = \{a_0, a_1, a_2\}$:

$$
\begin{tikzcd}
 & a_2 \arrow[dl, "\mathbb{P}(a_0|a_2)"'] \\
a_0 \arrow[loop left, "\mathbb{P}(a_0|a_0)"] \arrow[r, "\mathbb{P}(a_1|a_0)"'] & a_1 \arrow[loop right, "\mathbb{P}(a_1|a_1)"] \arrow[u, "\mathbb{P}(a_2|a_1)"']
\end{tikzcd}
$$

A cadeia de Markov representada pelo grafo anterior é representada pela seguinte matriz de transição:

$$\begin{array}{c|ccc} & a_0 & a_1 & a_2 \\ \hline a_0 & \mathbb{P}(a_0|a_0) & \mathbb{P}(a_1|a_0) & \mathbb{P}(a_2|a_0) \\ a_1 & \mathbb{P}(a_0|a_1) & \mathbb{P}(a_1|a_1) & \mathbb{P}(a_2|a_1) \\ a_2 & \mathbb{P}(a_0|a_2) & \mathbb{P}(a_1|a_2) & \mathbb{P}(a_2|a_2) \end{array} \rightarrow P = \begin{bmatrix} \mathbb{P}(a_0|a_0) & \mathbb{P}(a_1|a_0) & 0 \\ 0 & \mathbb{P}(a_1|a_1) & \mathbb{P}(a_2|a_1) \\ \mathbb{P}(a_0|a_2) & 0 & 0 \end{bmatrix},$$

na qual $\sum_{j=1}^{3} \mathbb{P}(a_j|a_i) = 1$, para cada $i = 0, 1, 2$.

Se as transições representadas no grafo anterior forem igualmente prováveis, a matriz de transição é dada por

$$P = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \\ 1 & 0 & 0 \end{bmatrix}.$$

A partir da matriz de transição, é possível calcular probabilidades conjuntas como, por exemplo, $\mathbb{P}(X_t = x_t, ..., X_1 = x_1)$ usando, sucessivamente, a definição de probabilidade condicional.

Exemplo 3.2. Considere a probabilidade $\mathbb{P}(X_t = x_t, ..., X_1 = x_1)$. Usando a definição de probabilidade condicional e a propriedade Markoviana,

$$\begin{aligned} \mathbb{P}(X_t = x_t, ..., X_1 = x_1) &= \mathbb{P}(X_t = x_t|X_{t-1} = x_{t-1}, \dots, X_1 = x_1)\mathbb{P}(X_{t-1} = x_{t-1}, \dots, X_1 = x_1) \\ &= \mathbb{P}(X_t = x_t|X_{t-1} = x_{t-1})\mathbb{P}(X_{t-1} = x_{t-1}, \dots, X_1 = x_1). \end{aligned}$$

Analogamente, para a probabilidade $\mathbb{P}(X_{t-1} = x_{t-1}, ..., X_1 = x_1)$, obtemos

$$\begin{aligned} & \mathbb{P}(X_{t-1} = x_{t-1}, ..., X_1 = x_1) \\ = \; & \mathbb{P}(X_{t-1} = x_{t-1}|X_{t-2} = x_{t-2}, \dots, X_1 = x_1)\mathbb{P}(X_{t-2} = x_{t-2}, \dots, X_1 = x_1) \\ = \; & \mathbb{P}(X_{t-1} = x_{t-1}|X_{t-2} = x_{t-2})\mathbb{P}(X_{t-2} = x_{t-2}, \dots, X_1 = x_1). \end{aligned}$$

Usando este raciocínio, por indução, obtemos

$$\begin{aligned} \mathbb{P}(X_t = x_t, ..., X_1 = x_1) = \mathbb{P}(X_t = x_t|X_{t-1} = x_{t-1}) \ldots \mathbb{P}(X_2 = x_2|X_1 = x_1)\mathbb{P}(X_1 = x_1) \\ = \left(\prod_{i=2}^{t} \mathbb{P}(X_i = x_i|X_{i-1} = x_{i-1}) \right) \mathbb{P}(X_1 = x_1). \end{aligned}$$

Exemplo 3.3. Agora, mostramos como obter a expressão para o cálculo da probabilidade $\mathbb{P}(X_t = x_t, X_{t-2} = x_{t-2})$.

Como

$$\mathbb{P}(X_t = x_t, X_{t-2} = x_{t-2}) = \sum_{x \in S} \mathbb{P}(X_t = x_t, X_{t-1} = x, X_{-2} = x_{t-2})$$

então

$$\mathbb{P}(X_t = x_t, X_{t-2} = x_{t-2}) = \sum_{x \in S} \mathbb{P}(X_t = x_t, X_{t-1} = x, X_{t-2} = x_{t-2})$$
$$= \sum_{x \in S} \mathbb{P}(X_t = x_t | X_{t-1} = x)\mathbb{P}(X_{t-1} = x | X_{t-2} = x_{t-2})\mathbb{P}(X_{t-2} = x_{t-2}).$$

Exemplo 3.4. Seja a probabilidade $\mathbb{P}(X_t = x_t | X_{t-2} = x_{t-2})$.

Pela definição de probabilidade condicional,

$$\mathbb{P}(X_t = x_t | X_{t-2} = x_{t-2}) = \frac{\mathbb{P}(X_t = x_t, X_{t-2} = x_{t-2})}{\mathbb{P}(X_{t-2} = x_{t-2})}.$$

Logo,

$$\mathbb{P}(X_t = x_t | X_{t-2} = x_{t-2}) = \frac{\sum_{x \in S} \mathbb{P}(X_t = x_t, X_{t-1} = x, X_{t-2} = x_{t-2})}{\mathbb{P}(X_{t-2} = x_{t-2})}$$

$$= \frac{\sum_{x \in S} \mathbb{P}(X_t = x_t | X_{t-1} = x)\mathbb{P}(X_{t-1} = x | X_{t-2} = x_{t-2})\mathbb{P}(X_{t-2} = x_{t-2})}{\mathbb{P}(X_{t-2} = x_{t-2})}$$

$$= \sum_{x \in S} \mathbb{P}(X_t = x_t | X_{t-1} = x)\mathbb{P}(X_{t-1} = x | X_{t-2} = x_{t-2}).$$

Note que

$$\mathbb{P}(X_t = j | X_{t-2} = i) = \sum_{l \in S} \mathbb{P}(X_t = j | X_{t-1} = l)\mathbb{P}(X_{t-1} = l | X_{t-2} = i)$$

corresponde ao elemento (i, j), $i, j \in S$, da matriz resultante da multiplicação da matriz de transição P por ela mesma, ou seja, $P^2(i, j)$.

A probabilidade $\mathbb{P}(X_t = j | X_{t-2} = i)$ é denominada probabilidade de transição a dois passos e $P^2 = P * P$ é a matriz de transição a dois passos.

As probabilidades de transição P_{ij} dizem respeito à probabilidade de transição do estado i para o estado j em um único passo da cadeia.

Generalizando, a probabilidade de transição a n passos, denotada por $\mathbb{P}_{ij}^{(n)} = \mathbb{P}(X_t = j | X_{t-n} = i)$, $i, j \in S$, é a probabilidade do processo estar no estado j depois de n passos da cadeia, dado que o processo estava no estado i.

Exemplo 3.5. Considere a seguinte a matriz de transição a um passo e $S = \{1, 2, 3\}$:

$$P = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \\ 1 & 0 & 0 \end{bmatrix}.$$

Logo, a matriz P^2 é

$$P^2 = \begin{bmatrix} 0.25 & 0.5 & 0.25 \\ 0.5 & 0.25 & 0.25 \\ 0.5 & 0.5 & 0 \end{bmatrix}.$$

Então, segue que $\mathbb{P}(X_t = 1|X_{t-1} = 3) = P(3, 1) = 1$, enquanto que $\mathbb{P}(X_t = 1|X_{t-2} = 3) = P^2(3, 1) = 0.5$.

Exemplo 3.6. Considere a seguinte situação: uma empresa aluga carros e possui duas filiais, uma em Campinas (estado 0) e outra em Bauru (estado 1).

Ao alugar um carro dessa empresa, o mesmo pode ser devolvido tanto em Campinas quanto em Bauru. Suponha que $\mathbb{P}(0|0) = 0.8, \mathbb{P}(1|0) = 0.2, \mathbb{P}(0|1) = 0.1, \mathbb{P}(1|1) = 0.9$.

Se ambas as filiais começam com a mesma quantidade de carros em t_0, digamos 1000, a quantidade de carros em Campinas em t_1 é $1000 \cdot 0.8 + 1000 \cdot 0.1 = 900$. Analogamente, a quantidade de carros em Bauru em t_1 é $1000 \cdot 0.2 + 1000 \cdot 0.9 = 1100$. Isto equivale a

$$\begin{bmatrix} 1000 & 1000 \end{bmatrix} * \begin{bmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{bmatrix} = \begin{bmatrix} 900 & 1100 \end{bmatrix}.$$

Se quisermos saber o que acontecerá após dois passos (dois instantes de tempo após t), devemos usar a matriz de transição a dois passos, da seguinte maneira:

$$\begin{bmatrix} 1000 & 1000 \end{bmatrix} * \begin{bmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{bmatrix}^2 = \begin{bmatrix} 1000 & 1000 \end{bmatrix} * \begin{bmatrix} 0.66 & 0.34 \\ 0.17 & 0.83 \end{bmatrix} = \begin{bmatrix} 746.7 & 1253.3 \end{bmatrix}$$

Teorema 3.1. *(Equações de Chapman-Kolmogorov)*

Para uma cadeia de Markov $\{X_t : t \in \mathbb{Z}\}$*, homogênea e com espaço de estados S, temos*

$$P_{ij}^{m+n} = \sum_{l \in \mathcal{A}} P_{il}^{n} P_{lj}^{m}.$$

As Equações de Chapman-Kolmogorov são obtidas usando os conceitos abordados anteriormente.

$$\begin{aligned} P_{ij}^{m+n} &= \mathbb{P}(X_{n+m} = j | X_0 = i) = \sum_{l \in S} \mathbb{P}(X_{m+n} = j | X_n = l, X_0 = i)\mathbb{P}(X_n = l | X_0 = i) \\ &= \sum_{l \in S} \mathbb{P}(X_{m+n} = j | X_n = l)\mathbb{P}(X_n = l | X_0 = i) = \sum_{l \in S} P_{il}^{n} P_{lj}^{m}. \end{aligned}$$

Por exemplo, para $n = 0$ e $m = 2$,

$$\mathbb{P}(X_2 = j | X_0 = i) = \sum_{l \in S} \mathbb{P}(X_2 = j | X_1 = l)\mathbb{P}(X_1 = l | X_0 = i).$$

Vale ressaltar que, pelas Equações de Chapman-Kolmogorov, a matriz de probabilidade de transição a n passos é obtida multiplicando a matriz de probabilidade de transição a um passo, P, n vezes, isto é, P^n.

4 CLASSIFICAÇÃO DE ESTADOS

Definição 4.1. (Acessibilidade) Sejam i e j dois estados de uma cadeia de Markov de ordem um. Dizemos que o estado j é acessível pelo estado i se, quando o processo inicia no estado i, é possível que o processo, eventualmente, chegue ao estado j. Desse modo, o estado j é acessível pelo estado i se, para algum $n \geqslant 0$, $P_{ij}^n > 0$.

Definição 4.2. (Comunicabilidade) Sejam i e j dois estados de uma cadeia de Markov de ordem um. Se os dois estados i e j são acessíveis um pelo outro dizemos que eles são comunicantes. Assim, os dois estados se comunicam.

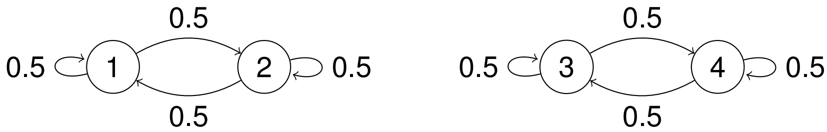
Definição 4.3. (Classe de estados) Se dois estados se comunicam, então eles pertencem a mesma classe de estados.

Definição 4.4. (Irredutibilidade) Uma cadeia de Markov é dita irredutível se ela possui uma única classe de estados, ou seja, todos os estados se comunicam entre si.

Exemplo 4.1. Considere a matriz de transição

$$P = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix}.$$

A representação em grafo desta cadeia é a seguinte:



Percebe-se que a cadeia é redutível, pois possui duas classes de estados.
Neste caso, a matriz de transição para cada classe de estados é a mesma

$$\begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}.$$

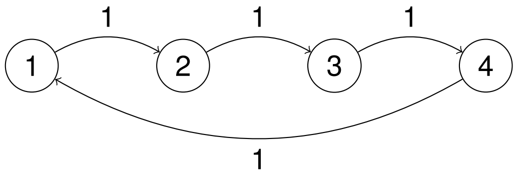
Definição 4.5. (Periodicidade & Aperiodicidade) Em uma cadeia de Markov de ordem um, o período $d(i)$ do estado i é definido por $d(i) = mdc(n \geqslant 1 : P_{ii}^n > 0)$. Se $d(i) = 1$ então o estado i é chamado de aperiódico e se $d(i) > 1$, o estado i é denominado periódico.

Definição 4.6. Uma cadeia de Markov de ordem um é aperiódica se todos os seus estados são aperiódicos.

Exemplo 4.2. (Estados Periódicos) Considere a seguinte cadeia de Markov de ordem um com matriz de transição P e $S = \{1, 2, 3, 4\}$:

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

O grafo correspondente é

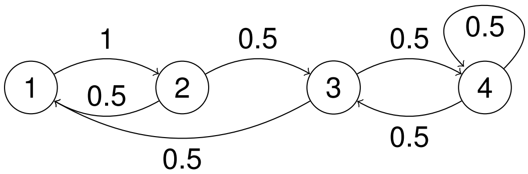


O período de qualquer estado $i \in S$ é $d(i) = 4$, pois $\mathbb{P}(X_n = i|X_0 = i) = 1$, para todo $i \in S$ e para todo $n \in 4\mathbb{N}$, e $\mathbb{P}(X_n = i|X_0 = i) = 0\ \forall n \in (\mathbb{N} - 4\mathbb{N})$.

Exemplo 4.3. (Cadeia Aperiódica) Considere uma cadeia de Markov de ordem um, cujo espaço de estados é $S = \{1, 2, 3, 4\}$ e a matriz de transição é

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix}.$$

A cadeia de Markov é representada pelo seguinte grafo:



- Partindo do estado i=1 temos, por exemplo, os caminhos até retornar ao estado 1: $1 \to 2 \to 1$ e $1 \to 2 \to 3 \to 1$. Logo, $d(1) = mdc(2, 3, ...) = 1$;

- Partindo do estado i=2 temos, por exemplo, os caminhos até retornar ao estado 2: $2 \to 1 \to 2$ e $2 \to 3 \to 1 \to 2$. Logo, $d(2) = mdc(2, 3, ...) = 1$;

- Partindo do estado i=3 temos, por exemplo, os caminhos até retornar ao estado 3: $3 \to 4 \to 3$ e $3 \to 4 \to 4 \to 3$, $3 \to 1 \to 2 \to 3$. Logo, $d(3) = mdc(2, 3, ...) = 1$;

- Partindo do estado i=4 temos, por exemplo, os caminhos até retornar ao estado 4: $4 \to 4$, $4 \to 3 \to 4$, $4 \to 3 \to 1 \to 2 \to 3 \to 4$. Logo, $d(4) = mdc(1, 2, 5, ...) = 1$.

Como todos os estados são aperiódicos, podemos afirmar que essa cadeia de Markov é aperiódica.

Um exemplo de uma cadeia aperiódica e redutível é uma cadeia possuindo a matriz identidade como matriz de transição.

Definição 4.7. (Recorrência e Transiência) Considere uma cadeia de Markov de ordem um.
(i) O estado i é classificado como recorrente se, ao iniciar o processo no estado i, o processo retornará ao estado i com probabilidade 1;
(ii) Dizemos que o estado i é transitório/transiente se, ao iniciar o processo no estado i, existe uma probabilidade positiva do processo não voltar mais ao estado i;
(iii) O estado i é absorvente se entrando nesse estado, o processo nunca sairá desse estado.

Intuitivamente, se o estado y é recorrente, então, iniciando o processo em y, ele retornará ao estado y com probabilidade 1. Se considerarmos esse tempo de retorno como um novo início do processo, então o processo novamente retornará ao estado y, e assim sucessivamente. Logo, o processo retornará infinitas vezes ao estado y.

Definição 4.8. O tempo até o primeiro retorno ao estado i, dado que o processo iniciou-se no estado i é definido como

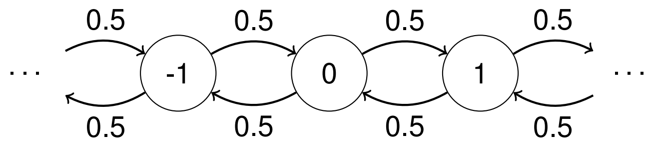
$$T_i = \min\{n \geqslant 1 : X_n = i\}.$$

Em outras palavras, T_y representa o número de transições que uma cadeia leva para retornar a y pela primeira vez, dado que ela iniciou-se no estado y, $y \in S$. Denotamos $\mathbb{E}(T_y|X_0 = y)$ por m_y, que corresponde ao tempo esperado de retorno ao estado y, dado que a cadeia iniciou-se em y.

Exemplo 4.4. Considere $S = \{0, 1\}$ e a seguinte amostra observada de uma cadeia de Markov $\{X_t : t \in \mathbb{N}\}$: $x(0) = 1, x(1) = 0, x(2) = 0, x(3) = 1, x(4) = 0, x(5) = 1$. Neste caso, o valor observado da variável T_1 é igual a 3.

Definição 4.9. (Recorrência Positiva e Nula) Considere uma cadeia de Markov de ordem um e o estado recorrente i. Dizemos que o estado i é recorrente positivo se $\mathbb{E}(T_i|X_0 = i) < \infty$. Além disso, o estado i é recorrente nulo se $\mathbb{E}(T_i|X_0 = i) = \infty$.

Exemplo 4.5. (Passeio Aleatório em $\mathbb{Z}$) Considere uma cadeia de Markov de ordem um, cujo espaço de estados é $S = \mathbb{Z}$, cujas probabilidades de transição são $P_{i,i+1} = p$ e $P_{i,i-1} = 1 - p$, $i \in \{..., -1, 0, 1, ...\}$. Em particular, consideramos o caso em que $p = \frac{1}{2}$, cuja cadeia é representada pelo seguinte diagrama de transição de estados:



Note que a cadeia é irredutível, pois todos os estados se comunicam. Agora, calculamos o tempo esperado até o primeiro retorno ao estado i, dado que o processo iniciou-se no estado i, $i \in S$, ou seja,

$$\mathbb{E}(T_i|X_0 = i) = \sum_{n=1}^{\infty} n\mathbb{P}(T_i = n|X_0 = i) = \sum_{n=1}^{\infty} nf_{ii}^{(n)},$$

em que

$$f_{ii}^{(n)} = \mathbb{P}(T_i = n|X_0 = i) = \mathbb{P}(X_1 \neq i, ..., X_{n-1} \neq i, X_n = i|X_0 = i).$$

Sem perda de generalidade, considere o estado 0. Pelo diagrama de transição de estados, note que o retorno ao estado 0, dado que o processo iniciou-se no estado 0, é possível sempre que o número de transições n for par. Assim, $f_{00}^{(n)} = 0$, para n ímpar.

Agora, obtemos $f_{00}^{(n)}$, para n par. Neste caso, quando andamos n passos, é necessário que tenhamos andado um total de $\frac{n}{2}$ passos para a esquerda e $\frac{n}{2}$ passos para a direita.

Por exemplo, para $n = 2$, há duas possibilidades para retornar ao estado 0 em dois passos, dado que o processo iniciou-se no estado 0, a saber: $0 \to 1 \to 0$ ou $0 \to -1 \to 0$. Neste caso, é necessário que tenhamos andado um total de um passo à esquerda e um passo à direita e, assim, $f_{00}^{(2)} = 0.5 * 0.5 + 0.5 * 0.5 = 0.5$.

Já, para $n = 4$, há seis possibilidades para retornar ao estado 0 em quatro passos, dado que o processo iniciou-se no estado 0, a saber: $0 \to 1 \to 2 \to 1 \to 0$ ou $0 \to -1 \to -2 \to -1 \to 0$ ou $0 \to 1 \to 0 \to -1 \to 0$ ou $0 \to -1 \to 0 \to 1 \to 0$ ou $0 \to 1 \to 0 \to 1 \to 0$ ou $0 \to -1 \to 0 \to -1 \to 0$. Neste caso, é necessário que tenhamos andado um total de dois passos à esquerda e dois passos à direita e, desta forma, $f_{00}^{(2)} = 6 * 0.5 * 0.5 * 0.5 * 0.5 = \binom{4}{2} 0.5^4 = 0.375$. O termo $\binom{4}{2}$ calcula o número de possibilidades em retornar ao estado 0 com duas transições à esquerda e duas à direita, dado que o processo iniciou-se no estado 0.

De modo geral, para n par, $f_{00}^{2k} = \binom{2k}{k} 0.5^{2k}$, $k = 1, 2, ...$. Esta expressão é válida para qualquer estado $i \in S$, ou seja, $f_{ii}^{2k} = \binom{2k}{k} 0.5^{2k}$, $k = 1, 2, ...$, $i \in S$, já que as probabilidades de transição à esquerda são iguais e as probabilidades de transição à direita também são iguais.

Diante disso,

$$\mathbb{E}(T_i|X_0 = i) = \sum_{n=1}^{\infty} n f_{ii}^{(n)} = \sum_{k=1}^{\infty} 2k \binom{2k}{k} 0.5^{2k} = \sum_{k=1}^{\infty} 2k \frac{(2k)!}{k!k!} 0.5^{2k}.$$

A aproximação de Stirling garante que

$$k! \approx k^{k+1/2} e^{-k} \sqrt{2\pi}.$$

Usando a aproximação de Stirling em $\binom{2k}{k}$ obtemos

$$\binom{2k}{k} = \frac{(2k)!}{k! \cdot k!} \approx \frac{(2k)^{2k+1/2} e^{-2k} \sqrt{2\pi}}{(k^{k+1/2} e^{-k} \sqrt{2\pi})(k^{k+1/2} e^{-k} \sqrt{2\pi})} = \frac{(2k)^{2k+1/2}}{(k^{k+1/2})(k^{k+1/2} \sqrt{2\pi})}$$

$$= \frac{k^{2k+1/2} 2^{2k+1/2}}{k^{2k+1} \sqrt{2\pi}} = \frac{\sqrt{2} 2^{2k}}{k^{1/2} \sqrt{2} \sqrt{\pi}} = \frac{2^{2k}}{\sqrt{k\pi}} = \frac{4^k}{\sqrt{k\pi}}.$$

Logo, segue que

$$\begin{aligned} \mathbb{E}(T_i|X_0 = i) &\approx \sum_{k=1}^{\infty} 2k \frac{4^k}{\sqrt{k\pi}} 0.5^{2k} = \sum_{k=1}^{\infty} 2k \frac{4^k}{\sqrt{k\pi}} 4^{-k} \\ &= \sum_{k=1}^{\infty} \frac{2}{\sqrt{\pi}} k^{1/2} = \frac{2}{\sqrt{\pi}} \sum_{k=1}^{\infty} k^{1/2} = \infty. \end{aligned}$$

Portanto, este exemplo corresponde a uma cadeia de Markov irredutível e recorrente, cuja recorrência é nula. É importante notar que não existe recorrência nula para espaço de estados S finito.

A classificação de estados é importante para o estudo da existência e da unicidade da distribuição estacionária de uma cadeia de Markov, ou seja, para caracterizar a cadeia de Markov para a qual é garantida a existência e a unicidade de sua distribuição estacionária. Na Seção 5, abordamos distribuição estacionária e suas propriedades.

Especificamente, a existência do limite:

$$\lim_{n\to\infty} \frac{\sum_{m=1}^{n} P^m(x, y)}{n}, \tag{4.1}$$

para todos os estados $x, y \in S$, é usada para determinar quais cadeias de Markov possuem distribuição estacionária e em qual caso ela é única, lembrando que $P^m(x, y) = P^m_{xy} = \mathbb{P}(X_t = y|X_{t-1} = x)$ é o elemento (x, y) da matriz P^m.

Considere a função indicadora

$$I_y(z) = \begin{cases} 1, & z = y, \\ 0, & z \neq y. \end{cases}$$

Então,

$$\mathbb{E}(I_y(X_n)|X_0 = x) = P(X_n = y|X_0 = x) = P^n(x, y), \tag{4.2}$$

$$N_n(y) = \sum_{m=1}^{n} I_y(X_m), \tag{4.3}$$

e

$$G_n(x, y) = \sum_{m=1}^{n} P^m(x, y). \tag{4.4}$$

A expressão (4.3) representa o número de visitas ao estado y durante $m = 1, \ldots, n$, e $\mathbb{E}(N_n(y)|X_0 = x) = G_n(x, y)$ é o número esperado de tais visitas para uma cadeia iniciando em x, conforme (HOEL; PORT; STONE, 1972), Cap. 2.2, pág. 57.

Além disso, $\dfrac{N_n(y)}{n}$ e $\dfrac{G_n(x, y)}{n}$ representam, respectivamente, a proporção das primeiras n unidades de tempo que a cadeia permanece no estado y e o valor esperado desta proporção para a cadeia iniciando no estado x, conforme (HOEL; PORT; STONE, 1972) no Cap. 2.2, Teo. 1, pág 58.

Observe que a expressão (4.4) é o numerador da fração em (4.5). ou seja podemos reescreve-la como:

$$\lim_{n\to\infty} \frac{\sum_{m=1}^{n} P^m(x, y)}{n} = \lim_{n\to\infty} \frac{G_n(x, y)}{n}. \tag{4.5}$$

Note que, para $x, y \in S$ fixados, $P^m(x, y)$ é a probabilidade do processo estar no estado y depois de m passos, dado que o processo estava no estado x. Logo, $\displaystyle\sum_{m=1}^{n} P^m(x, y)$ corresponde à soma das probabilidades de transição da cadeia estar no estado y dado que estava no estado x a um passo, a dois passos, até a n passos.

A sequência $(P(x, y), P^2(x, y), P^3(x, y), \ldots)$ pode convergir a um limite L finito. Neste caso, quando $n \to \infty$, a fração $\displaystyle\sum_{m=1}^{n} \frac{P^m(x, y)}{n}$, associada ao valor esperado da proporção das primeiras n unidades de tempo que a cadeia permanece no estado y tendo iniciada no estado x, convergirá ao mesmo limite L.

No entanto, a sequência $(P(x, y), P^2(x, y), P^3(x, y), \ldots)$ pode não convergir a um limite. Mesmo assim, quando $n \to \infty$, a fração $\displaystyle\sum_{m=1}^{n} \frac{P^m(x, y)}{n}$ pode convergir.

Teorema 4.1. *Para um estado transiente y,*

- $\lim_{n\to\infty} N_n(y) = N(y) < \infty$ *com probabilidade 1;*
- $\lim_{n\to\infty} G_n(x, y) = G(x, y) < \infty$, *para* $x \in S$.

A demonstração do Teorema 4.1 pode ser consultada em (HOEL; PORT; STONE, 1972), Cap. 1.5, Teo. 1, pág. 19.

5 DISTRIBUIÇÃO ESTACIONÁRIA

Em uma cadeia de Markov homogênea com espaço de estados S e matriz de transição P temos as seguintes distribuições marginais:

- Distribuição inicial: $\alpha_i^0 = P(X_0 = i), i = 1, \dots, |S|$;
- Distribuição marginal no instante n: $\alpha_i^n = P(X_n = i), i = 1, \dots, |S|$.

Considere a representação da distribuição inicial por meio do vetor linha $\alpha^0 = (\alpha_1^0, \dots, \alpha_{|S|}^0)$ e da distribuição marginal no instante n por meio do vetor linha por $\pi^n = (\pi_1^n, \dots, \pi_{|S|}^n)$.

Pelo Teorema da Probabilidade Total obtemos

$$\pi_i^n = \sum_{j \in S} \pi_j^{n-1} P_{ji}, \tag{5.1}$$

para todo $i \in S$.

Matricialmente, a expressão (5.1) é escrita como $\pi^n = \pi^{n-1}\mathbf{P}$.

A longo prazo, isto é, quando $n \to \infty$, se o comportamento probabilístico da cadeia se estabilizar, espera-se que $\pi^n = \pi^{n-1}$. Nesse caso, essa distribuição de equilíbrio π deve satisfazer $\pi = \pi\mathbf{P}$.

Exposto isto, definimos distribuição estacionária.

Definição 5.1. (Distribuição Estacionária) Considere uma cadeia de Markov de ordem um com matriz de transição P. O vetor π é chamado de distribuição estacionária da cadeia de Markov se satisfazer:

(i) $\pi_i \geqslant 0, \forall i \in S$, e $\sum_{i \in S} \pi_i = 1$;

(ii) Equação Global de Balanço: $\pi = \pi P$.

Usando repetitivamente a definição de uma distribuição estacionária π, obtemos uma propriedade interessante (HOEL; PORT; STONE, 1972). De:

$$\begin{aligned}\sum_{x \in S} \pi(x) P^2(x, y) &= \sum_{x \in S} \pi(x) \sum_{z \in S} P(x, z) P(z, y) = \sum_{z \in S} \sum_{x \in S} \pi(x) P(x, z) P(z, y) \\ &= \sum_{z \in S} \pi(z) P(z, y) = \pi(y),\end{aligned}$$

obtemos a relação $\sum_{x \in S} \pi(x) P^2(x, y) = \pi(y)$.

Agora, usando a definição de distribuição estacionária e a Equação de Chapman-Kolmogorov, generalizamos a propriedade supracitada:

$$\sum_{x \in S} \pi(x) P^n(x, y) = \pi(y), \tag{5.2}$$

para $y \in S$.

Somando em $m = 1, \ldots, n$ e dividindo por n ambos os lados de (5.2), obtemos:

$$\sum_{x \in S} \pi(x) \sum_{m=1}^{n} \frac{P^m(x,y)}{n} = \sum_{m=1}^{n} \frac{\pi(y)}{n} \Longrightarrow \sum_{x \in S} \pi(x) \frac{G_n(x,y)}{n} = \pi(y). \tag{5.3}$$

Teorema 5.1. *Seja π uma distribuição estacionária. Se o estado y for transiente ou recorrente nulo então $\pi(y) = 0$.*

Dem: Se y for um estado transiente ou recorrente nulo,

$$\lim_{n \to \infty} \frac{G_n(x,y)}{n} = 0, \tag{5.4}$$

para $y \in S$, conforme visto na seção anterior.

Por (5.3) e pelo Teorema da Convergência Limitada[1]*, tomando $b_n(y) = \dfrac{G_n(x,y)}{n}$ e usando (5.4), obtemos*

$$\pi(y) = \lim_{n \to \infty} \sum_{x \in S} \pi(x) \frac{G_n(x,y)}{n} = 0.$$

Portanto, $\pi(y) = 0$. ■

Teorema 5.2. *Se uma cadeia de Markov for irredutível e recorrente positiva possui uma única distribuição estacionária π dada por $\pi(y) = \dfrac{1}{m_y}$.*

Dem: Como y é um estado recorrente positivo, então

$$\lim_{n \to \infty} \frac{G_n(x,y)}{n} = \frac{1}{m_y} > 0, x, y \in S, \tag{5.5}$$

Já que m_y é um real positivo fixado.

Agora, suponha que π é uma distribuição estacionária. Usando (5.3), o Teorema da Convergência Limitada e (5.5),

$$\pi(y) = \lim_{n \to \infty} \sum_{x \in S} \pi(x) \frac{G_n(x,y)}{n} = \frac{1}{m_y} \sum_{x \in S} \pi(x) = \frac{1}{m_y}.$$

Portanto, se há uma distribuição estacionária π, ela é dada por $\pi(y) = \dfrac{1}{m_y}$.

[1] Seja $a(x), x \in S$ e uma sequência $b_n(x)$ tal que $|b_n(x)| \leqslant 1$, para $x \in S$, e que convirja para $b(x)$, $x \in S$. Então, $\lim\limits_{n \to \infty} \sum\limits_{x \in S} a(x) b_n(x) = \sum\limits_{x \in S} a(x) b(x)$.

Para finalizar, precisamos mostrar que $\pi(y)$, $y \in S$, definida por $\pi(y) = \dfrac{1}{m_y}$, é de fato uma distribuição estacionária, ou seja, precisamos verificar que:

- $\pi(y) = \dfrac{1}{m_y} > 0, y \in S$;
- $\sum_{y \in S} \pi(y) = \sum_{y \in S} \dfrac{1}{m_y} 1$;
- $\pi(y) = \dfrac{1}{m_y}$ *satisfaz a equação global de balanço.*

De fato, $\pi(y) = \dfrac{1}{m_y} > 0, y \in S$.

Agora, verificamos que $\sum_{y \in S} \pi(y) = 1$. *Como* $\sum_{y \in S} P^m(x, y) = 1$, *então, somando em* $m = 1, \dots, n$ *e dividindo por* n *ambos os lados da igualdade, obtemos*

$$\sum_{m=1}^{n} \sum_{y \in S} \frac{P^m(x, y)}{n} = \sum_{m=1}^{n} \frac{1}{n} \Rightarrow \sum_{y \in S} \frac{G_n(x, y)}{n} = 1. \tag{5.6}$$

Se S *for finito, usando (5.5) e (5.6), provamos que*

$$\lim_{n \to \infty} \sum_{y \in S} \frac{G_n(x, y)}{n} = \sum_{y \in S} \frac{1}{m_y} = 1.$$

Portanto, se S *for finito,* $\pi(y), y \in S$ *é uma distribuição de probabilidades.*

Pelo Teorema de Chapman-Kolmogorov (3.1):

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

Logo,

$$\begin{aligned} \sum_{m=1}^{n} \sum_{z \in S} \frac{P^m(x, z) P(z, y)}{n} &= \sum_{m=1}^{n} \frac{P^{m+1}(x, y)}{n} \Rightarrow \sum_{z \in S} \frac{G_n(x, z)}{n} P(z, y) \\ &= \frac{G_{n+1}(x, y)}{n} - \frac{P(z, y)}{n}. \end{aligned} \tag{5.7}$$

Novamente, supomos que S *é finito. Tomando o limite no lado esquerdo de (5.7) obtemos*

$$\lim_{n \to \infty} \sum_{z \in S} \frac{G_n(x, z)}{n} P(z, y) = \sum_{z \in S} \lim_{n \to \infty} \frac{G_n(x, z)}{n} P(z, y) = \sum_{z \in S} \frac{1}{m_z} P(z, y).$$

Tomando o limite no lado direito de (5.7) obtemos

$$\lim_{n \to \infty} \frac{G_{n+1}(x, y)}{n} - \lim_{n \to \infty} \frac{P(z, y)}{n} = \frac{1}{m_y}.$$

Portanto,

$$\sum_{z\in S}\frac{1}{m_z}P(z,y)=\frac{1}{m_y}$$

e, assim, verificamos que a equação global de balanço é satisfeita para $\pi(y)=\dfrac{1}{m_y}$, $y\in S$.

Se S for infinito, a prova é mais complicada, desde que não podemos intercambiar diretamente limite e somatório. A demonstração para este caso pode ser consultada em (HOEL; PORT; STONE, 1972), Cap. 2.8, pág 75. ■

Corolário 5.1. *Seja uma cadeia de Markov irredutível e recorrente positiva com distribuição estacionária* π. *Então, com probabilidade 1,*

$$\lim_{n\to\infty}\frac{N_n(y)}{n}=\pi(y),$$

para $y\in S$.

Teorema 5.3. *Seja uma cadeia de Markov irredutível e recorrente positiva com distribuição estacionária* π. *Se a cadeia for aperiódica,*

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

para $x,y\in S$. *Além disso, se a cadeia é periódica com período* d, *então, para cada par de estados* $x,y\in S$, *existe um número inteiro* r, $0\leqslant r<d$, *tal que* $P^n(x,y)=0$ *ao menos que* $n=md+r$ *para algum número inteiro não-negativo* m, *e*

$$\lim_{m\to\infty}P^{md+r}(x,y)=d\pi(y).$$

A demonstração do Teorema 5.3 pode ser encontrada em (HOEL; PORT; STONE, 1972), Cap. 2.7, teo. 7, pág 73.

Uma observação interessante em relação ao limite $\lim\limits_{n\to\infty}P^n(x,y)$ refere-se à sua velocidade de convergência. Considerando a matriz P diagonalizável, pelo Teorema Espectral, $P=V^{-1}DV$, em que V é uma matriz cujas colunas correspondem aos autovetores de P e D é uma matriz diagonal formada pelos autovalores de P. Logo, $P^n=(V^{-1}DV)^n=\underbrace{(V^{-1}DV)(V^{-1}DV)\dots(V^{-1}DV)}_{n\text{ vezes}}=V^{-1}D^nV$.

Como D é uma matriz diagonal formada pelos autovalores $\lambda_1,\lambda_2,\dots$ de P, D^n é uma matriz diagonal formada por $\lambda_1^n,\lambda_2^n,\dots$.

É possível mostrar que cada matriz P possui um de seus autovalores iguais a um e, no caso de uma cadeia de Markov irredutível, aperiódica e com espaço de estados finito, os demais autovalores satisfazem $|\lambda_{ii}|<1$. Assim, pode-se indexar os autovalores em ordem decrescente de valor absoluto: $\lambda_{11}=1>|\lambda_{22}|\geqslant|\lambda_{33}|\geqslant\dots\geqslant|\lambda_{|S||S|}|$.

Diante disso, à grosso modo, a matriz limite de P^n é geometricamente aproximada na mesma razão que $|\lambda_{22}|^n \to 0$, quando $n \to \infty$.

Exemplo 5.1. Seja uma cadeia de Markov de ordem um, cujo espaço de estados é $S = \{1, 2\}$ e a matriz de transição é dada por:

$$P = \begin{bmatrix} 3/4 & 1/4 \\ 1/2 & 1/2 \end{bmatrix}.$$

Os autovalores de P são obtidos resolvendo-se

$$det\left(\begin{bmatrix} \frac{3}{4} - \lambda & 1/4 \\ 1/2 & \frac{1}{2} - \lambda \end{bmatrix}\right) = 0,$$

que resulta em $\left(\frac{1}{2} - \lambda\right)\left(\frac{3}{4} - \lambda\right) - \frac{1}{8} = 0$, cujas raízes são $\lambda_{11} = 1$ e $\lambda_{22} = \frac{1}{4}$.

Neste caso, os termos de P^n tendem a um valor limite tão rápido quanto $\left(\frac{1}{4}\right)^n \to 0$.

Para o autovalor $\lambda_{11} = 1$, obtém-se o autovetor $(1 \;\; 1)^t$ e, para o autovalor $\lambda_{22} = 1/4$, obtém-se o autovetor $(1/2 \;\; 1)^t$. Assim,

$$P^n = \begin{bmatrix} 1 & 1/2 \\ 1 & -1 \end{bmatrix}^{-1} \begin{bmatrix} 1^n & 0 \\ 0 & \left(\frac{1}{4}\right)^n \end{bmatrix} \begin{bmatrix} 1 & 1/2 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} + \frac{1}{3}\left(\frac{1}{4}\right)^n & \frac{1}{3} - \frac{1}{3}\left(\frac{1}{4}\right)^n \\ \frac{2}{3} - \frac{2}{3}\left(\frac{1}{4}\right)^n & \frac{1}{3} + \frac{2}{3}\left(\frac{1}{4}\right)^n \end{bmatrix}.$$

Portanto,

$$\lim_{n\to\infty} P^n = \begin{bmatrix} 2/3 & 1/3 \\ 2/3 & 1/3 \end{bmatrix}.$$

Ambas as linhas da matriz $\lim\limits_{n\to\infty} P^n$ correspondem à distribuição limite $\pi = (2/3 \;\; 1/3)^t$, ou seja, π satisfaz $\pi = P\pi$.

Exemplo 5.2. (Cadeia de Markov periódica) Considere uma cadeia de Markov, cujo espaço estado é $S = \{0, 1, 2, 3\}$ e a matriz de transição é dada por:

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

Logo,

$$P^2 = \begin{bmatrix} \frac{3}{9} & 0 & \frac{6}{9} & 0 \\ 0 & \frac{7}{9} & 0 & \frac{2}{9} \\ \frac{2}{9} & 0 & \frac{7}{9} & 0 \\ 0 & \frac{6}{9} & 0 & \frac{3}{9} \end{bmatrix}, \quad P^3 = \begin{bmatrix} 0 & \frac{21}{27} & 0 & \frac{6}{27} \\ \frac{7}{27} & 0 & \frac{20}{27} & 0 \\ 0 & \frac{20}{27} & 0 & \frac{7}{27} \\ \frac{6}{27} & 0 & \frac{21}{27} & 0 \end{bmatrix}, \quad P^4 = \begin{bmatrix} \frac{21}{81} & 0 & \frac{60}{81} & 0 \\ 0 & \frac{61}{81} & 0 & \frac{20}{81} \\ \frac{20}{81} & 0 & \frac{61}{81} & 0 \\ 0 & \frac{60}{81} & 0 & \frac{21}{81} \end{bmatrix}$$

Considerando uma aproximação de dois dígitos:

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0.33 & 0 & 0.67 & 0 \\ 0 & 0.67 & 0 & 0.33 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad P^2 = \begin{bmatrix} 0.33 & 0 & 0.67 & 0 \\ 0 & 0.78 & 0 & 0.22 \\ 0.22 & 0 & 0.78 & 0 \\ 0 & 0.67 & 0 & 0.33 \end{bmatrix}$$

$$P^3 = \begin{bmatrix} 0 & 0.78 & 0 & 0.22 \\ 0.26 & 0 & 0.74 & 0 \\ 0 & 0.74 & 0 & 0.26 \\ 0.22 & 0 & 0.78 & 0 \end{bmatrix}, \quad P^4 = \begin{bmatrix} 0.26 & 0 & 0.74 & 0 \\ 0 & 0.75 & 0 & 0.25 \\ 0.25 & 0 & 0.75 & 0 \\ 0 & 0.74 & 0 & 0.26 \end{bmatrix}$$

Para $n >> 1$, P^n tem as seguintes distribuições estacionárias (dependendo da paridade de n):

$$P^{2n} = \begin{bmatrix} \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ 0 & \frac{3}{4} & 0 & \frac{1}{4} \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ 0 & \frac{3}{4} & 0 & \frac{1}{4} \end{bmatrix}, \qquad P^{2n+1} = \begin{bmatrix} 0 & \frac{3}{4} & 0 & \frac{1}{4} \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ 0 & \frac{3}{4} & 0 & \frac{1}{4} \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 \end{bmatrix}$$

Exemplo 5.3. (Cadeia de Markov redutível de recorrência positiva) Considere o espaço de estados $S = \{1, 2, 3\}$ e a matriz de transição a ele associado:

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Note que 3 é um estado absorvente, já que:

$$P(3|3) = 1, \mathbb{P}(1|3) = 0, \mathbb{P}(2|3) = 0$$

Além disso,

$$P^2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} = P,$$

ou seja, $P^2 = P$. Logo, $P^n = P$, $\forall n$.

Observe que $\mathbb{E}(T_3|X_0 = 3) = 1$ pois partindo de 3 sempre se retorna a 3 imediatamente, pois o estado 3 é absorvente.

Para o estado x, $x = 1, 2$

$$\mathbb{E}(T_x|X_0 = x) = \sum_{n=1}^{\infty} nP(T_x = n|X_0 = x) = \sum_{m}^{\infty} P^m(x, y)$$

Neste caso, $\mathbb{E}(T_1|X_0 = 1) = \mathbb{E}(T_2|X_0 = 2)$, pois

$$\mathbb{E}(T_x) = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2^2} + 3 \cdot \frac{1}{2^3} + ...$$

$$= \sum_{i>0}^{\infty} \frac{i}{2^i} = 1 + \sum_{i>1}^{\infty} \frac{i-1}{2^i}$$

$$= 1 + \frac{1}{2} + \sum_{i>2}^{\infty} \frac{i-2}{2^i} = \sum_{i=0}^{\infty} \frac{1}{2^i} = 2 < \infty.$$

Sendo assim, a cadeia é recorrente positivo.

Usando a equação global de balanço, temos que:

$$\begin{bmatrix} \pi_1 & \pi_2 & \pi_3 \end{bmatrix} = \begin{bmatrix} \pi_1 & \pi_2 & \pi_3 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{\pi_1}{2} + \frac{\pi_2}{2}, & \frac{\pi_1}{2} + \frac{\pi_2}{2}, & \pi_3 \end{bmatrix}.$$

Assim, temos o sistema:

$$\begin{cases} \pi_1 = \pi_2, \\ \pi_1 + \pi_2 + \pi_3 = 1, \end{cases}$$

resultando em $2 \cdot \pi_1 + \pi_3 = 1$. Porém, esta equação é satisfeita para infinitas soluções. Neste caso, a distribuição estacionária não é única, ou seja, a unicidade é perdida neste caso, em que a cadeia é redutível (não é irredutível).

Exemplo 5.4. Seja uma cadeia de Markov de ordem um, cujo espaço de estados é $S = \{1, 2\}$ e a matriz de transição é dada por:

$$P = \begin{bmatrix} 0.4 & 0.6 \\ 0.6 & 0.4 \end{bmatrix}.$$

Então,

$$\begin{bmatrix} \pi_1 & \pi_2 \end{bmatrix} = \pi * P = \begin{bmatrix} \pi_1 & \pi_2 \end{bmatrix} * \begin{bmatrix} 0.4 & 0.6 \\ 0.6 & 0.4 \end{bmatrix} \to \pi = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}.$$

Adicionalmente,

$$P^2 = \begin{bmatrix} 0.52 & 0.48 \\ 0.48 & 0.52 \end{bmatrix}, \qquad P^3 = \begin{bmatrix} 0.496 & 0.504 \\ 0.504 & 0.496 \end{bmatrix}, \qquad P^5 = \begin{bmatrix} 0.49984 & 0.50016 \\ 0.50016 & 0.49984 \end{bmatrix}.$$

Note que, $\lim_{n\to\infty} P_{ij}^n = \pi_j$, para cada $j \in S$.

6 CONSIDERAÇÕES FINAIS

Neste trabalho abordamos uma classe especial de processos estocásticos, correspondendo a cadeias de Markov a tempo discreto, com alfabeto enumerável, que possuem grande aplicabilidade em modelagem, além de serem usados em métodos de simulação, como os algoritmos de Monte Carlo acoplados a cadeias de Markov (MCMC).

Em geral, os livros didáticos abordam apenas cadeias de Markov de ordem um. Apesar deste trabalho estar focado neste tipo de cadeia, apresentamos a definição de uma cadeia de Markov de ordem k, que é uma extensão de uma cadeia de Markov de ordem um.

Especificamente, para cadeias de Markov de ordem um, abordamos os principais conceitos e propriedades, dando especial atenção à interpretação.

Um aspecto interessante e importante no estudo de cadeias de Markov é a definição de distribuição estacionária e, consequentemente, para quais cadeias a distribuição estacionária existe e é única. Para isto, foi necessário apresentar a classificação de estados e cadeias como, por exemplo, cadeia irredutível, aperiódica e de recorrência positiva.

Em especial, no estudo de sistemas de filas, o uso de distribuição estacionária é útil e importante.

Como possibilidade de continuidade deste trabalho, sugerimos o estudo de cadeias de Markov que estendem os processos Markovianos de ordem fixa.

REFERÊNCIAS

BUHLMANN, P.; WYNER, A. J. Variable length markov chains. **The Annals of Statistics**, v. 27, n. 2, p. 480–513, 1999. Citado na página 6.

DURRETT, R. **Essentials of stochastic processes**. [S.l.]: Springer, 1999. v. 1. Citado na página 6.

GALVES, A. et al. Sonority as a basis for rhythmic class discrimination. **Paper presented at Speech Prosody, Aix-en-Provence (can be downloaded from www.lpl.univ-aix.fr/sp2002/pdf/galves-etal.pdf)**, 2002. Citado na página 6.

GARCíA, J. E.; GONZáLEZ-LóPEZ, V. A. Minimal markov models. **Fourth Workshop on Information Theoretic Methods in Sciences and Engineering**, p. 25–28, 2011. Citado na página 6.

HOEL, P. G.; PORT, S. C.; STONE, C. J. **Introduction to Stochastic Processes**. [S.l.]: Houghton Mifflin, 1972. Citado 4 vezes nas páginas 20, 21, 22 e 25.

LEONARDI, F. A generalization of the pst algorithm: Modeling the sparse nature of protein sequences. **Bioinformatics**, v. 22, n. 11, p. 1302–1307, 2006. Citado na página 6.

LEONARDI, F. **Cadeias Estocásticas Parcimoniosas com Aplicações à classificação e filogenia das sequências de proteínas**. [S.l.]: Tese de doutorado. Universidade de São Paulo, 2007. Citado na página 6.

RISSANEN, J. A universal data compression system. **IEEE Transactions on Information Theory**, v. 29, n. 5, p. 656–664, 1983. Citado na página 6.