Pedro Lourenço Mendes Júnior

Convergência para a medida estacionária em Cadeias de Markov

Florianópolis

2017

Pedro Lourenço Mendes Júnior

Convergência para a medida estacionária em Cadeias de Markov

Trabalho de Conclusão de Curso apresentado ao Curso de Matemática, do Departamento de Matemática - Centro de Ciências Físicas e Matemáticas da Universidade Federal de Santa Catarina, para obtenção de grau de Licenciado em Matemática.

Universidade Federal de Santa Catarina

Centro de Ciências Físicas e Matemática

Departamento de Matemática

Licenciatura em Matemática

Orientador: Dr. Leandro Batista Morgado

Florianópolis

2017

Pedro Lourenço Mendes Júnior

Convergência para a medida estacionária em Cadeias de Markov

Trabalho de Conclusão de Curso apresentado ao Curso de Matemática, do Departamento de Matemática - Centro de Ciências Físicas e Matemáticas da Universidade Federal de Santa Catarina, para obtenção de grau de Licenciado em Matemática.

Trabalho aprovado. Florianópolis, 2017

Profa Dra. Sonia Palomino Castro
Coordenadora do Curso

Banca Examinadora:

Dr. Leandro Batista Morgado(Orientador)
Universidade Federal de Santa Catarina

Prof. Dr. Fábio Junior Margotti
Universidade Federal de Santa Catarina

Prof. Dr. Leonardo Silveira Borges
Universidade Federal de Santa Catarina

Florianópolis
2017

Agradecimentos

Agradeço ao meu primeiro professor na vida, meu pai, Pedro Lourenço Mendes e a mulher mais importante da minha vida, minha mãe, Valdelina Saes pelo empenho, dedicação e amor incondicional. Agradeço aos meus irmãos, Karina e Cristian, pelo apoio e carinho e aos meus sobrinhos, Guilherme, Gustavo, Enzo e Sofia.

Gostaria de agradecer ao meu orientador, Leandro Batista Morgado, pelo companheirismo, suporte e dedicação durante a produção deste trabalho. Agradeço também a todos os outros professores do curso que ministraram disciplinas ou de alguma outra forma ajudaram na minha formação acadêmica.

Agradeço a todos os funcionários do Centro de Estudos Matemáticos, e um carinho especial pelos professores Erivaldo e Rogério Baiano, por terem me proporcionado uma grande experiência, e aos colegas Gustavo, Tiago, Fábio, Sidinei, Adriano e Dé por todo o suporte durante as minhas dificuldades e colaboração em ensinar e aprender matemática.

Finalmente, agradeço a outra parte fundamental que são os meus amigos. Da matemática ou não, eles foram fundamentais para que eu pudesse continuar minha aventura com leveza e felicidade. Terei sempre um carinho especial pelos colegas André Luiz, Pedro Crizel, Renato Fernandes, Matheus Pimenta, Beli Thomaz, Mateus Patrício, Gustavo Ackermann, Marina Borges e Vanessa da Rocha.

"Não sei o que possa parecer aos olhos do mundo, mas aos meus pareço apenas ter sido como um menino brincando à beira-mar, divertindo-me com o fato de encontrar de vez em quando um seixo mais liso ou uma concha mais bonita que o normal, enquanto o grande oceano da verdade permanece completamente por descobrir à minha frente."

(Isaac Newton)

Resumo

O objetivo deste trabalho é apresentar uma visão geral sobre as Cadeias de Markov, como um processo estocástico capaz de modelar problemas em diversas áreas da matemática aplicada e, além disso, apresentar duas dessas aplicações em áreas distintas. Inicialmente, será discutida uma base teórica de Probabilidade, em matemática discreta, bem como alguns resultados de Álgebra Linear envolvendo autovalores e autovetores. A seguir, apresenta-se a definição de Cadeias de Markov e a classificação delas junto a alguns exemplos. Aborda-se o conceito de medida estacionária e os casos de existência e unicidade de tal medida, além da convergência para a medida através do limite de potências da matriz de transição. Finalmente, será apresentada a aplicação no caso do jogo de tabuleiro Snakes and Ladders e alguns resultados para as probabilidade de acabar o jogo após n rodadas, e a utilização de Cadeias de Markov para estudos de crescimento populacional de espécies de esquilos não-nativos da região.

Palavras-chave: Cadeias de Markov, medida estacionária, convergência, processo estocástico.

Abstract

The goal of this text is to present an overall vision about Markov chains, how a stochastic process is capable of shaping problems in the diverse fields of applied math and, beyond that, presents two of the practical applications in distinct fields. First, it will discuss basic theory in discrete probability and the results of Linear Algebra involving eigenvalues and eigenvectosr. After that, it will be introduced the concept of stationary distribution and the cases of existence and unicity of these distributions. In addition, it was study the convergence through the limiting distribution of matrices powers. Finally, the study presents an application in the case of the Snakes and Ladders table game and the use of Markov Chains to study the growth of a population of non-region native squirrel species.

Keywords: Markov Chains, stachastic process, stationary distribuction, convergence.

Sumário

Introdução

Neste trabalho, vamos apresentar algumas noções básicas de Cadeias de Markov, bem como a aplicação de conceitos de Álgebra Linear na área. Vamos discutir a convergência do respectivo processo para a medida estacionária via potências de matriz, e apresentamos algumas aplicações, tais como o jogo de tabuleiro Snakes and Ladders e na área da Biologia em crescimento populacional.

O trabalho será dividido em três capítulos. Primeiramente, abordaremos aspectos introdutórios de Probabilidade e de Álgebra Linear. Sendo assim, apresentaremos as definições de eventos, espaço amostral, probabilidade de um evento e independência de eventos. Estudaremos probabilidade condicional, conceito fundamental para a compreensão de Cadeias de Markov. Ainda no primeiro capítulo, veremos alguns tópicos de Álgebra Linear, destacando os conceitos de autovalores e autovetores, independência linear, bem como resultados relacionados a esses temas, que usaremos nos capítulos posteriores deste trabalho.

No segundo capítulo, apresentaremos os principais conceitos de Cadeias de Markov, definiremos ainda matriz estocástica, vetor de probabilidade, distribuição inicial da cadeia e veremos que um processo estocástico pode ser compreendido como um modelo matemático probabilístico que evolui ao longo de um tempo. As Cadeias de Markov diferem dos outros processos estocásticos por terem a característica de que a evolução da cadeia nos próximos passos depende apenas do momento atual da cadeia, independente do histórico anterior da mesma. Entende-se esse aspecto como uma espécie de "perda de memória"do processo estocástico. Para uma compreensão melhor, destacaremos alguns exemplos como o Passeio Aleatório no n-Ciclo, e o Problema do Aluguel de Bicicletas.

A nomenclatura Cadeias de Markov deve-se ao matemático russo Andrei Andreyevi Markov (1856 - 1922), que obteve resultados em análise e teoria dos números. Os estudos iniciais de Markov envolviam cálculos de probabilidade para consoantes e vogais, baseado na letra antecedente, em escritos de Pushkin. Não levando em consideração as letras anteriores, Markov determinou as probabilidades para as letras que viriam em sequência.Em geral, existem processos estocásticos em tempo contínuo e processos estocásticos em tempo discreto. No presente estudo, abordaremos apenas as Cadeias de Markov em tempo discreto.

As medidas estacionárias ou invariantes de uma Cadeia de Markov são uma parte interessante e fonte de muitos estudos nesta área. Intuitivamente, podemos pensar na medida estacionária como o momento em que o processo estaciona, no sentido de que com o passar do tempo, as probabilidades não devem se alterar. Em geral, veremos que determinar a medida estacionária de uma Cadeia de Markov parte da solução de um sistema linear do tipo $v = v \cdot P$, onde v é um vetor de probabilidade e P é uma matriz de transição associada a uma Cadeia de Markov.

Estudaremos algumas restrições na cadeia para a existência e unicidade de tal medida, bem como algumas classificações necessárias como, por exemplo, o conceito de cadeias aperiódicas ou periódicas e irredutíveis ou redutíveis. Para o estudo desses resultados, vamos demonstrar o Teorema de Perron-Frobenius. Veremos que podemos estudar a medida estacionária a partir de potência de matriz de transição da Cadeia de Markov, para isso, veremos um teorema de convergência por potências.

No terceiro capítulo, veremos a aplicação dos conceitos estudados durante os capítulos anteriores para o jogo de tabuleiro Snakes and Ladders. Faremos uma interpretação matricial do tabuleiro do jogo e obteremos uma matriz de transição que modela os passos do peão pelo tabuleiro. Veremos que, pelo tabuleiro utilizado, após muitas jogadas é provável que o peão encontra o fim do tabuleiro. A fim de comprovar essa hipótese, desejamos determinar a convergência através da iteração da matriz de transição. Elabora-se ainda mapas de densidade de probabilidade após alguns lançamentos de dados, e estudaremos o caso onde há mais de um peão percorrendo o tabuleiro.

Outra aplicação das Cadeias de Markov que apresentaremos no terceiro capítulo trata de um estudo de caso de crescimento populacional de algumas espécies de animais. Mais especificamente, estudaremos o caso dos esquilos cinzas e esquilos vermelhos na região da Grã-Bretanha, e vamos procurar determinar as chances de sobrevivência dessas espécies na mesma região. Extraindo os dados obtidos em questionários que expõem as aparições dos esquilos, determinamos as probabilidades de crescimento e decrescimento das espécies num intervalo de tempo. Através dos conceitos de Cadeias de Markov, desejamos compreender a evolução das espécies para um longo período de tempo e determinar as chances de sobrevivência dessas espécies.

1 Conceitos Preliminares

Neste capítulo, iremos tratar objetivamente de conceitos iniciais de Probabilidade e de Álgebra Linear que serão fundamentais para o estudo dos capítulos seguintes. Iniciaremos com resultados básicos em Probabilidade discreta, e posteriormente trataremos de aspectos elementares de Álgebra Linear que são requisitos necessários para a compreensão do trabalho como um todo.

1.1 Introdução à Probabilidade

1.1.1 Noções Básicas e Definições

Em geral, os experimentos estudados podem ter um modelo determinístico ou aleatório. Entendem-se os experimentos determinísticos como aqueles que, repetidos sobre as mesmas condições, produzirão os mesmos resultados. Por exemplo, a Lei de Kepler para o movimentação planetária pode ser entendido como um modelo determinístico, assim como as Leis da Gravitação. A área de probabilidade tem o interesse em estudar os modelos aleatórios, que são os experimentos que não produzem os mesmos resultados quando repetidos sobre as mesmas condições. Neste capítulo, estamos interessados em discutir alguns aspectos de teoria da probabilidade em experimentos discretos. Considere os seguintes exemplos de experimentos aleatórios:

Exemplo 1.1.1. *Lançamos um dado sobre uma superfície e observamos o número da face superior.*

Exemplo 1.1.2. *Efetuamos três lançamentos de uma moeda e observamos a quantidade de caras.*

Exemplo 1.1.3. *Efetuamos três lançamentos de uma moeda e observamos a sequência de caras e coroas.*

Exemplo 1.1.4. *Durante a fabricação de peças, observamos o número total de peças até que sejam produzidas 10 peças defeituosas.*

Exemplo 1.1.5. *Observamos o tempo de duração de uma lâmpada.*

Com o objetivo de criar um modelo para representar os experimentos aleatórios, podemos descrever os resultados possíveis de um experimento e, principalmente, determinar as chances de um conjunto de resultados possíveis.

Definição 1.1.1. *O conjunto S de todos os resultados possíveis de um experimento é denominado de* ***espaço amostral****.*

No Exemplo 1.1.1, que corresponde ao lançamento de um dado, o espaço amostral é o conjunto $S_1 = \{1, 2, 3, 4, 5, 6\}$. Para o Exemplo 1.1.2, teremos o espaço amostral $S_2 = \{0, 1, 2, 3\}$. Usando a notação de K para coroas e C para caras, o Exemplo 1.1.3 possui espaço amostral $S_3 = \{KKK, KKC, KCK, CKK, KCC, CKC, CCK, CCC\}$.

Note que nos Exemplos 1.1.1, 1.1.2 e 1.1.3, o espaço amostral do experimento é um conjunto com uma quantidade finita de elementos. O mesmo não ocorre com o Exemplo 1.1.4, pois não é possível garantir quando será atingida a quantidade de 10 peças defeituosas. Sendo assim, podemos entender o espaço amostral como $S_4 = \{10, 11, 12, 13, 14, \ldots\}$, um subconjunto ilimitado dos naturais e, portanto, com um infinito enumerável de elementos.

Já o Exemplo 1.1.5, onde devemos observar a vida útil de uma lâmpada, o espaço amostral é um conjunto dos números reais não-negativos, portanto com um número infinito e não-enumerável de elementos onde $S_5 = \{x \in \mathbb{R}; x \geqslant 0\}$.

Em geral, trabalharemos com o espaço amostral idealizado matematicamente e diferente do espaço amostral realizável. Por exemplo, no caso de mensurar a vida útil de uma lâmpada, nossos instrumentos para medir o tempo possuem limitações. Poderíamos ter um cronômetro com uma precisão de duas casas decimais nos segundos, e neste caso o espaço amostral passaria a ter uma quantidade infinita e enumerável de elementos. Também poderíamos supor que a lâmpada pararia de funcionar em algum momento, então o espaço amostral passaria ter uma quantidade finita de elementos. Tomaremos sempre o espaço amostral mais conveniente matematicamente.

Definição 1.1.2. *Seja S o espaço amostral associado a um experimento. Um subconjunto de S é denominado* ***evento****.*

O conceito de evento é simplesmente um conjunto de resultados possíveis do experimento. Desta maneira, o próprio espaço amostral S de um experimento é também um evento possível do mesmo experimento. De forma semelhante, o conjunto vazio $\emptyset$ também é um evento. Definiremos alguns eventos associados aos espaços amostrais dos exemplos anteriores. Com relação ao Exemplo 1.1.1, tome o evento A_1 onde ocorre um número par na face superior, ou seja, $A_1 = \{2, 4, 6\}$. Com respeito ao Exemplo 1.1.2, poderíamos ter o evento onde não ocorre cara em nenhum lançamento, sendo assim, $A_2 = \{0\}$. O subconjunto $A_3 = \{KKK, KKC, KCK, CKK\}$ define o evento onde ocorreram, pelo menos, duas coroas em três lançamentos de uma moeda como no Exemplo 1.1.3.

Tomando o Exemplo 1.1.1, suponha que o experimento seja executado duas vezes. Então poderíamos representar o conjunto dos possíveis resultados por $S \times S$. Um possível resultado $(s_1, s_2) \in S \times S$ significa que s_1 foi obtido na primeira realização do experimento e s_2 na segunda realização do experimento. Teremos o espaço amostral $S = \{(1,1);(1,2);(1,3);(1,4);...;(5,6);(6,6)\}$ associado ao experimento.

1.1.1.1 Operações entre eventos

Definição 1.1.3. *A união de dois eventos A e B de um experimento (que denotamos por $A \cup B$) é o evento que ocorre se pelo menos um deles ocorrer.*

Definição 1.1.4. *A interseção de dois eventos A e B de um experimento (que denotamos por $A \cap B$) é o evento que ocorre se ambos ocorrem.*

Definição 1.1.5. *O complementar de um evento A de um experimento (que denotamos por A^c) é evento que ocorre se, e somente se, A não ocorre.*

Definição 1.1.6. *Seja $A_1, A_2, \ldots, A_i, \ldots$, uma coleção infinita enumerável de eventos. Denotamos por $\bigcap_{i=1}^{\infty} A_i$ o evento que ocorre se, e somente se, todos os eventos A_i ocorrerem.*

Definição 1.1.7. *Seja $A_1, A_2, \ldots, A_i, \ldots$, uma coleção infinita enumerável de eventos. Denotamos por $\bigcup_{i=1}^{\infty} A_i$ o evento que ocorre se, e somente se, ao menos um dos eventos A_i ocorrer.*

Veja que as Definições 1.1.3 e 1.1.4 podem ser naturalmente estendidas para uma coleção finita de conjuntos. Em relação ao Exemplo 1.1.1, podemos associar os eventos:

1. Evento A: O número da face superior é par.
2. Evento B: O número da face superior é primo.
3. Evento C: O número da face superior é maior que 5.

Sendo assim, podemos representar os seguintes eventos:

- $A = \{2, 4, 6\}$;
- $B = \{2, 3, 5\}$;
- $A^c = \{1, 3, 5\}$;
- $A \cup B \cup C = \{2, 3, 4, 5, 6\}$;
- $B \cap C = \emptyset$;
- $A \cap B = \{2\}$;

Definição 1.1.8. *Dois eventos A e B são denominados **mutuamente excludentes**, se eles não puderem ocorrer juntos, ou seja, $A \cap B = \emptyset$.*

Observe que no Exemplo 1.1.1, os eventos B e C acima definidos são mutuamente excludentes pois não há elementos na sua interseção.

Até o momento, a respeito dos experimentos, associamos os conceitos de espaço amostral como o conjunto de todos os resultados possíveis de um experimento e o conceito de evento como um subconjunto do espaço amostral. Entretanto o interesse no estudo de um modelo probabilistico é saber as chances de determinado evento ocorrer ou não, ou seja, não temos a certeza de que um evento A ocorrerá ao realizar o experimento, diferentemente dos modelos determinísticos. Voltemos nossas atenções a fim de encontrar uma boa maneira de medir as chances de determinados eventos.

Definição 1.1.9. *Seja $n(A)$ o número de vezes em que o evento A ocorreu nas n repetições do experimento, então a razão $f_{n,A} = \dfrac{n(A)}{n}$ é denominada frequência relativa de A.*

A frequência relativa de um evento associado a um experimento surge da ideia de repetir sucessivamente o experimento, de modo que as repetições não dependam dos resultados anteriores.

Proposição 1.1.1. *Sejam A e B eventos e S o espaço amostral associado a este evento, então são válidas as seguintes propriedades:*

1. $0 \leqslant f_A \leqslant 1$.
2. $f_A = 1$ *se, e somente se, ocorrer A em repetição alguma.*
3. $f_A = 0$ *se, e somente se, não ocorrer A em todas as repetições.*
4. *Se A e B forem eventos mutuamente excludentes, então $f_{A \cup B} = f_A + f_B$.*

Tratando da dificuldade em associar um número real às chances de determinado evento ocorrer, uma possível maneira de tratar o problema seria repetir o experimento em uma grande quantidade de vezes e determinar a frequência relativa do evento. Imagine o experimento de jogar uma moeda e observar a face superior, nos 10 primeiros lançamentos poderíamos ter a ocorrência de 9 coroas, o que nos levaria a uma frequencia de coroas de 9 em 10 arremessos. Nos próximos 10 arremessos, dificilmente repetiríamos a mesma frenquência de coroas de anteriormente, então uma certa "instabilidade"no experimento seria observada. Desta maneira, não possuímos um número mínimo de lançamentos para obtermos as chances de cada evento ocorrer. O que desejamos é uma boa maneira de possuir tal número que reproduza as chances de cada evento, independente do observador e das n repetições do evento.

Definição 1.1.10. *Seja S um espaço amostral associado a um experimento. A cada evento A associamos um número real $P(A)$ denominado* ***probabilidade de*** *A, que satisfaça às seguintes propriedades:*

1. $0 \leqslant P(A) \leqslant 1$.

2. $P(S) = 1$.

3. *Se A e B forem eventos mutuamente excludentes, então* $P(A \cup B) = P(A) + P(B)$.

4. *Se A_1, A_2,...,A_n,... forem, dois a dois, eventos mutuamente excludentes, então:* $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$

A definição acima apenas caracteriza algumas propriedades gerais que a probabilidade de um evento A deverá ter. Ainda não possuímos uma maneira de calcular efetivamente $P(A)$, porém alguns resultados iniciais já podem ser estudados com a caracterização das propriedades.

Teorema 1.1.1. *Se $\emptyset$ for o conjunto vazio, então* $P(\emptyset) = 0$.

Demonstração. Seja A um evento qualquer. Da teoria de conjuntos temos que $A = A \cup \emptyset$ e os eventos A e $\emptyset$ são mutuamente excludentes. Logo, pela Propriedade 3 da Definição 1.0.2 $P(A) = P(A \cup \emptyset) = P(A) + P(\emptyset)$. Concluíndo-se que $P(\emptyset) = 0$. ■

Teorema 1.1.2. *Para todo evento A,* $P(A) = 1 - P(A^c)$.

Demonstração. Podemos escrever o espaço amostral S como a união disjunta $S = A + A^c$, como A e A^c são mutuamente excludentes, pelas Propriedades 2 e 3 da Definição 1.0.2, temos $P(S) = 1 = P(A) + P(A^c)$. Imediatamente obtemos $P(A) = 1 - P(A^c)$. ■

Teorema 1.1.3. *Sejam A e B dois eventos do espaço amostral S, tais que $A \subset B$, tem-se:*

$$P(A) \leqslant P(B)$$

Demonstração. Como $A \subset B$, então $B = A \cup (B \cap A^c)$. Por A e $B \cap A^c$ serem mutuamente excludentes, segue que $P(B) = P(A) + P(B \cap A^c)$. Pela Propriedade 1 da Definição 1.0.2, tem-se que $0 \leqslant P(B \cap A^c)$ e daí conclúi-se que $P(A) \leqslant P(B)$. ■

Teorema 1.1.4. *Se A e B forem dois eventos quaisquer, então:*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Demonstração. Sejam A e B eventos quaisquer, então podemos reescrever $A \cup B$ como $A \cup (B \cap A^c)$ onde A e $(B \cap A^c)$ são eventos mutuamente excludentes pois $B \cap A^c \subset A^c$ e $A^c \cap A = \emptyset$. B pode ser reescrito como $B = (B \cap A) \cup (B \cap A^c)$. Logo, tem-se $P(A \cup B) = P(A) + P(B \cap A^c)$ e $P(B) = P(B \cap A) + P(B \cap A^c)$. Subtraindo as equações, tem-se $P(A \cup B) - P(B) = P(A) + P(A \cap B)$ e imediatamente teremos $P(A \cup B) = P(A) + P(B) + P(A \cap B)$. ■

Pode ser feita uma generalização do Teorema 1.1.4 para uma quantidade finita de eventos $A_1, A_2, A_3, \ldots, A_n$ que iremos apenas enunciar.

Teorema 1.1.5. *Sejam $A_1, A_2, \ldots, A_n$ eventos de um espaço amostral, temos:*

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j=2}^{n} P(A_i \cap A_j) + \sum_{i<j<r=3}^{n} P(A_i \cap A_j \cap A_r) + \ldots \\ +(-1)^{n-1} P(A_1 \cap A_2 \cap \ldots \cap A_n) \tag{1.1.1}$$

Voltemos a nossa atenção para os experimentos que possuem o espaço amostral formado apenas por um número finito de elementos, ou seja, S pode ser representado por $S = \{a_1, a_2, \ldots, a_n\}$. Para encontrar uma maneira de calcular $P(A)$ para todo evento A de S, é necessária a compreensão de um eventos simples.

Definição 1.1.11. *Seja S um espaço amostral com n elementos, diz-se* ***evento simples*** *a todo evento $A = \{a_i\}$ para algum $i \in \{1, \ldots, n\}$.*

Definição 1.1.12. *Seja $\{a_i\}$ um evento simples do espaço amostral $S = \{a_1, a_2, \ldots, a_n\}$. Associaremos $\{a_i\}$ um número p_i, denominado* ***probabilidade de*** *$\{a_i\}$, que satisfaz:*

1. $p_i \geqslant 0, \; i = 1, 2, \ldots, n.$

2. $p_i + p_2 + \ldots + p_n = 1.$

Agora devemos buscar uma maneira para calcular as probabilidades dos eventos que não são simples. Sendo assim, podemos supor um evento A, constituído de k resultados, onde $1 \leqslant k \leqslant n$, ou seja, podemos escrever A como o conjunto $A = \{a_{j_1}, a_{j_2}, \ldots, a_{j_k}\}$. Onde $A \subseteq S$ e $j_m \neq j_n, \forall m, n \in \{1, 2, \ldots k\}$. Consequentemente, pela Propriedade 4 da Definição 1.1.10, podemos calcular a probabilidade $P(A)$ como a soma de eventos simples $P(A) = p_{j_1} + p_{j_2} + \ldots + p_{j_k}$.

Tratando de espaço amostral finito, é comum que uma hipótese seja utilizada para a construção de uma modelo para calcular as probabilidades de um evento. Um caso particular ocorre quando todos os n eventos simples de um espaço amostral S são verossímeis, ou seja, são igualmente prováveis como resultado do experimento.

Definição 1.1.13. *Sendo assim, a probabilidade de um evento simples $A = \{a_i\}$ é $P(A) = 1/n$ onde n é o número de resultados do espaço amostral. De forma semelhante, para um evento $B = \{b_1, b_2, \ldots, b_k\}$ com k resultados, temos que a probabilidade de B é dada por $P(B) = k/n$.*

Observação 1.1.1. *É bastante comum que o cálculo da probabilidade de um evento A seja feito da seguinte maneira:*

$$P(A) = \frac{\text{número de casos favoráveis ao evento } A}{\text{número total de resultados possíveis de } S}$$

1.1.1.2 Probabilidade Condicional

Um dos principais conceitos estudados em Teoria de Probabilidade e fundamental para o entendimento do que será proposto no estudo subsequente em Cadeias de Markov são os conceitos de Probabilidade Condicional e Independência de Eventos. Iniciaremos propondo o seguinte exemplo:

Exemplo 1.1.6. *Considere uma urna com 5 bolas pretas e 6 bolas vermelhas onde as probabilidades de retirada de cada uma das bolas sejam iguais. Suponha que sejam retiradas duas bolas, sem reposição da bola retirada. Qual é a probabilidade de retirar uma bola preta na segunda retirada, sabendo que a primeira retirada foi uma bola vermelha?*

Considere P_i as bolas pretas e V_i as bolas vermelhas. Sendo assim, temos o espaço amostral $S = \{P_1, P_2, P_3, P_4, P_5, V_1, V_2, V_3, V_4, V_5, V_6\}$. Pelo experimento proposto, temos que a probabilidade de retirar inicialmente uma bola preta é $\frac{5}{11}$, já que possuímos 5 bolas pretas em 11 bolas totais. Porém, temos a informação de que já foi retirada uma bola vermelha e que a mesma não retornou para a urna. A fim de calcular a probabilidade de retirar uma bola preta num segundo momento, as chances desse evento são diferentes das chances de retirada num primeiro momento. Isso ocorre pois a urna da primeira retirada é diferente da urna da segunda. Sendo assim, a segunda retirada possui um espaço amostral diferente do espaço amostral anterior, denotando por S_2 o espaço amostral da segunda retirada, então $S_2 = \{P_1, P_2, P_3, P_4, P_5, V_1, V_2, V_3, V_4, V_5, \}$. Ou seja, a probabilidade de retirar uma bola preta passa a ser $\frac{5}{10}$ pois permanecemos com 5 bolas pretas mas reduzimos a 10 bolas totais. Esse exemplo trata do conceito de Probabilidade Condicional já que gostaríamos de calcular a probabilidade de um evento (retirar uma bola preta na segunda retirada da urna) sabendo o resultado de um evento anterior. Isso causou uma redução do espaço amostral. Com um formalismo maior, podemos definir da seguinte maneira:

Definição 1.1.14. *Sejam A e B dois eventos de um espaço amostral e supondo que $P(A) > 0$, a probabilidade condicional de B dado A é definida por:*

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Exemplo 1.1.7. *Considere o experimento de lançar um dado duas vezes e observar o número presente na face superior em cada um dos lançamentos. Considere então os eventos A e B onde A é o evento onde "nos dois lançamentos, os números observados são* *menores do que 3"e B o evento onde "a soma dos números obtidos nos dois lançamentos é igual a 4". Queremos saber qual a probabilidade do evento B acontecer, dado que o evento A ocorreu.*

Podemos associar ao experimento o seguinte espaço amostral $S = \{(p,q) : p, q \in \mathbb{N}, 1 \leqslant p \leqslant 6, 1 \leqslant q \leqslant 6\}$, o evento $A = \{(1,1),(1,2),(2,1),(2,2)\}$, bem como o evento $B = \{(1,3),(2,2),(3,1)\}$. Segue então, que a probabilidade de B ocorrer dado que A ocorreu é $\frac{1}{4}$ pois o caso $(2,2)$ é o único caso de B que faz parte do evento A já que obrigatoriamente A deve ocorrer. Podemos perceber que o espaço amostral S reduziu-se para os casos possíveis do evento A.

Utilizando a Definição 1.1.14, podemos perceber que $A \cap B = \{(2,2)\}$, portanto $P(A \cap B) = \frac{1}{36}$, pois S possui 36 resultados possíveis. Tem-se também que $P(A) = \frac{4}{36}$ e portanto:

$$P(B|A) = \frac{\frac{1}{36}}{\frac{4}{36}} = \frac{1}{4}$$

Da Definição 1.1.14, podemos estabelecer imediatamente a fórmula:

$$P(A \cap B) = P(A) \cdot P(B|A)$$

e uma generalização dessa expressão será enunciada no lema seguinte.

Lema 1.1.2. *Sejam $A_1, A_2, \ldots, A_n$ eventos do espaço amostral S, onde está definida a probabilidade P, tem-se:*

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_2 \cap A_1) \cdot \ldots \cdot P(A_n|A_1 \cap A_2 \ldots A_{n-1})$$

1.1.1.3 Eventos independentes e fórmula das probabilidades totais

Definição 1.1.15. *Dizemos que os eventos $B_1, B_2, \ldots, B_k$ representam uma **partição do espaço amostral S** quando:*

1. $B_i \cap B_j = \emptyset$ *para todo* $i \neq j$.
2. $\cup_{i=1}^{k} B_i = S$.
3. $P(B_i) > 0$ *para todo* i.

Isso nos diz que, dado um experimento, apenas um dos eventos B_i ocorre.

Exemplo 1.1.8. *No lançamento de um dado de 6 faces, $B_1 = \{1,2\}, B_2 = \{3,4,5\}$ e $B_3 = \{6\}$ formam uma partição de $S = \{1,2,3,4,5,6\}$, enquanto $C_1 = \{1,2,3\}$ e $C_2 = \{2,4,5,6\}$ não representam pois $2 \in C_1$ e $2 \in C_2$.*

Através da utilização de partições, poderemos compreender o conceito de probabilidades totais e a fórmula que enunciaremos adiante. Uma maneira interessante de decompor um evento A como união disjunta de eventos mutuamente excludentes é na forma:

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_k)$$

Como visto anteriormente, pela Propriedade 4 da Definição 1.1.10, podemos ainda escrever a probabilidade de $P(A)$ da seguinte forma:

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \ldots + P(A \cap B_k)$$

Onde $B_1, B_2, \ldots, B_k$ é uma partição do espaço amostral. Ademais, segue que $P(A \cap B_i) = P(A|B_i) \cdot P(B_i)$. Aplicando essa propriedade, podemos reescrever $P(A)$ na **Fórmula das Probabilidades Totais**, como:

$$P(A) = P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2) + \ldots + P(A|B_k) \cdot P(B_k)$$

Definição 1.1.16. *Sejam A e B eventos, dizemos que A e B são* ***eventos independentes*** *se, e somente se, $P(A \cap B) = P(A) \cdot P(B)$.*

A definição de eventos independente nos fornece os casos onde, dado que um determinado evento aconteceu, essa informação nos acrescenta em nada para que o segundo evento ocorra. Ilustraremos melhor a definição através do seguinte exemplo proposto:

Exemplo 1.1.9. *Suponhamos que um dado seja lançado duas vezes. Definamos os eventos A e B como:*
$A = \{$o primeiro dado mostra um número par$\}$ e $B = \{$o segundo dado mostra 5 ou 6$\}$ Sendo assim, temos que: $(A \cap B) = \{(2,5);(2,6);(4,5);(4,6);(6,5);(6,6)\}$ e como temos 36 casos no total, então $P(A \cap B) = \frac{6}{36} = \frac{1}{6}$. Além disso, $P(A) = \dfrac{1}{2}$ pois em metade dos casos o primeiro dado será par e na outra metade será ímpar, e $P(B) = 1/3$ pois temos apenas 2 possibilidades em 6 para o lançamento do segundo dado. Logo, $P(A) \cdot P(B) = P(A \cap B)$. Ou seja, os eventos A e B são eventos independentes. Intuitivamente podemos perceber que o resultado no primeiro dado não interfere no resultado do segundo lançamento, e dessa maneira fica configurado a ideia de independência de eventos.

1.2 Conceitos Básicos em Álgebra Linear

Esta subseção aborda os conceitos preliminares necessários para um melhor entendimento de resultados apresentados em capítulos seguintes. Foram selecionados alguns conceitos mais importante e também partimos já de uma base que o leitor deve ter para acompanhar o trabalho. Estamos tomando como já conhecido os conceitos de determinantes e suas propriedades, espaço vetorial, sistemas lineares e outros conceitos vistos em um curso básico de Álgebra Linear. Sendo assim, apresentaremos as definições de autovalores e autovetores, independência linear e algumas proposições necessárias. Para um estudo maior, o leitor poderá encontrar mais informações em LIMA (2009), POOLE (2006) e STRANG (2006).

1.2.1 Autovalores e Autovetores

Definição 1.2.1. *Seja A uma matriz real de ordem n. Dizemos que $\lambda \in \mathbb{C}$ é* ***autovalor de A*** *se existir um vetor x não nulo tal que $Ax = \lambda x$. Ademais, nesse caso, dizemos que x é um* ***autovetor de A*** *associado ao autovalor λ.*

Observação 1.2.1. *Encontrar os autovalores de uma matriz A é determinar as soluções da equação $det(A - \lambda I) = 0$. Assim, para que $\lambda \in \mathbb{C}$ seja um autovalor da matriz A, é necessário e suficiente que $(A - \lambda I)$ seja uma matriz singular. Chama-se ainda $p_A(\lambda) = det(A - \lambda I)$ o polinômio característico de A, e dessa forma os autovalores serão as raízes do polinômio característico.*

Para uma compreensão melhor da definição anterior, vamos tomar um exemplo para um matriz de grau 2 e determinar os seus autovalores e autovetores.

Exemplo 1.2.1. *Determinar o polinômio característico, os autovalores e autovetores associados a matriz A, onde:*

$$A = \begin{pmatrix} 4 & -5 \\ 2 & -3 \end{pmatrix}.$$

O polinômio característico de A é dado por:

$$p_A(\lambda) = det(A - \lambda I) = \lambda^2 - \lambda - 2 = (\lambda + 1) \cdot (\lambda - 2).$$

Portanto, pela observação acima, temos que $\lambda_1 = -1$ e $\lambda_2 = 2$ são os autovalores da matriz A. Queremos agora determinar os autovetores associados aos autovalores encontrados. Sendo assim, determinaremos as soluções não nulas para $Ax = (-1)x$ e $Ax = (2)x$. Para $\lambda_1 = -1$, e escrevendo $x^T = (a \quad b)$ o autovetor associado, tem-se:

$$\begin{pmatrix} 5 & -5 \\ 2 & -2 \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Sendo assim, o sistema linear possui infinitas soluções, e todo vetor satisfazendo $a = b$, $b \neq 0$, será autovetor de A, ou seja, sem perda de generalidade, podemos tomar o autovetor $x^T = (1 \quad 1)$ como autovetor associado ao autovalor $\lambda_1 = -1$. Analogamente, podemos tomar o autovetor $x^T = (5 \quad 2)$ associado ao autovalor $\lambda_2 = 2$.

Em geral, determinar os autovalores de uma matriz não é algo simples, como por exemplo, para matrizes de ordem superiores a 4 não temos fórmulas fechadas para determinar as raízes do polinômio característico associado. Para isto, são interessantes algumas relações entre a matriz A e seus autovalores, como veremos a seguir: Para os resultados dos capítulos seguintes, é importante encontrar uma relação entre os autovalores de uma matriz A e os autovalores das potências de A. Sendo assim, apresentamos a seguinte proposição:

Proposição 1.2.1. *Seja A uma matriz com autovalor λ associado ao autovetor x. Então, para qualquer $n \in \mathbb{N}, n \geq 1$, λ^n será autovalor de A^n e x será um autovetor associado.*

Demonstração. Queremos demonstrar que vale a igualdade $A^n.x = \lambda^n.x$. Para isto, faremos uma demonstração por indução em n. Sendo assim, para $n = 1$, temos por hipótese que λ é autovalor de A e x seu autovetor, então $A.x = \lambda.x$. Agora, tome como hipótese de indução que seja válido para $n = k$ inteiro positivo, ou seja, $A^k.x = \lambda^k.x$.
Devemos provar que o resultado segue também para $n = k + 1$. Temos então:

$$A^{k+1}.x = A(A^k).x$$

Substituindo a hipótese de indução, $A^k.x = \lambda^k.x$, tem-se:

$$A^{k+1}.x = A\big(A^k.x\big) = A.(\lambda^k.x) = \lambda^k(A.x)$$

E como $A.x = \lambda.x$, segue finalmente que:

$$A^{k+1}.x = \lambda^k(\lambda.x) = \lambda^{k+1}.x$$

Portanto, λ^n é autovalor de A^n e x é autovetor associado a este autovalor. ■

Um outro resultado importante é a independência linear entre dois autovetores associados a autovalores distintos. Para apresentar este resultado, definiremos independência linear e demonstraremos uma proposição a seguir:

Definição 1.2.2. *Sejam $v_1, v_2, \cdots, v_n$ vetores no espaço vetorial real $\mathbb{V}$ e $a_1, a_2, \cdots, a_n$ números reais. Considere a equação vetorial $a_1v_1 + a_2v_2 + \cdots + a_nv_n = 0$. Dizemos que os vetores $v_1, v_2, \cdots, v_n$ são **linearmente independentes** se a única solução da equação for $a_1 = a_2 = \cdots = a_n = 0$. Caso contrário, dizemos que os vetores $v_1, v_2, \cdots, v_n$ são **linearmente dependentes**.*

Observação 1.2.2. *Em um conjunto de vetores linearmente dependentes, pelo menos um deles pode ser escrito como combinação linear dos demais. De fato, se* $v_1, v_2, \cdots, v_n$ *são linearmente dependentes, existem, por definição números reais não todos nulos* $a_1, a_2, \cdots, a_n$ *tais que* $a_1v_1 + a_2v_2 + \cdots + a_nv_n = 0$. *Sem perda de generalidade, suponha que* $a_1 \neq 0$. *Então* $a_1v_1 = -a_2v_2 - a_3v_3 \cdots - a_nv_n$, *e portanto* $v_1 = -\frac{a_2}{a_1}v_2 - \frac{a_3}{a_1}v_3 - \ldots - \frac{a_n}{a_1}v_n$

Uma outra característica dos autovetores de uma matriz A é que eles formam um subespaço vetorial ao adicionarmos o vetor nulo ao conjunto dos autovetores, já que por definição o vetor nulo não é autovetor. Além disso, os autovetores são de fato linearmente independentes, como podemos compreender na proposição a seguir.

Proposição 1.2.2. *Seja* $A_{p\times p}$ *uma matriz com* $\lambda_1, \lambda_2, ..., \lambda_n$ *autovalores distintos de* A *com os respectivos autovetores* $v_1, v_2, ..., v_n$. *Então* $v_1, v_2, ..., v_n$ *são linearmente independentes.*

Demonstração. Vamos fazer uma demonstração por absurdo. Suponha que $v_1, v_2, ..., v_n$ são linearmente dependentes. Sendo assim, existe um vetor v_{k+1} que pode ser escrito como combinação linear autovetores anteriores, ou seja, $v_{k+1} = c_1.v_1 + c_2.v_2 + \cdots + c_k.v_k$. Assumimos também que $v_1, v_2, \ldots, v_k$ são linearmente independentes.

Da equação $Av_{k+1} = \lambda_{k+1}v_{k+1}$, temos:

$$\begin{aligned} \lambda_{k+1}v_{k+1} &= A.(c_1.v_1 + c_2.v_2 + \cdots + c_k.v_k) \\ &= c_1.A.v_1 + c_2.A.v_2 + \cdots + c_n.A.v_k \\ &= c_1.\lambda_1.v_1 + c_2.\lambda_2.v_2 + \cdots + c_k.\lambda_k.v_k \end{aligned} \tag{1.2.1}$$

A última igualdade decorre diretamente da hipótese de λ_i ser autovetor associado a v_i para todo $i \in \{1, ..., k\}$. Multiplicando $v_{k+1} = c_1.v_1 + c_2.v_2 + \cdots c_k.v_k$ por λ_{k+1}, temos:

$$\lambda_{k+1}.v_{k+1} = c_1.\lambda_{k+1}.v_1 + c_2.\lambda_{k+1}.v_2 + \cdots c_k.\lambda_{k+1}.v_k \tag{1.2.2}$$

Subtraindo as equações (1.2.1) e (1.2.2), teremos:

$$0 = c_1.(\lambda_1 - \lambda_{k+1}) + c_2.(\lambda_2 - \lambda_{k+1}) + \cdots + c_k.(\lambda_k - \lambda_{k+1})$$

Da independência linear de $v_1, v_2, ..., v_k$, conclui-se:

$$0 = c_1.(\lambda_1 - \lambda_{k+1}).v_1 = c_2.(\lambda_2 - \lambda_{k+1}).v_1 = \cdots = c_k.(\lambda_k - \lambda_{k+1}).v_k$$

Como, por hipótese, os autovalores são distintos e os autovetores são não nulos, então para todo $i \in \{1, ..., k\}$ deveremos ter $(\lambda_i - \lambda_{k+1})v_i \neq 0$ e portanto $c_i = 0$. Logo:

$$v_{k+1} = c_1.v_1 + c_2.v_2 + \cdots c_k.v_k = 0.v_1 + 0.v_2 + \cdots + 0.v_k = 0$$

Mas isso contradiz o fato de v_{k+1} ser autovetor. Portanto, segue que $v_1, v_2, ..., v_n$ são linearmente independentes. ■

Proposição 1.2.3. *Seja $A_{n\times n}$ uma matriz. Então, A e A^T possuem os mesmos autovalores*

Demonstração. Note que $(A-\lambda I)^T = A^T - (\lambda I)^T = A^T - \lambda I$, pois λI é uma matriz diagonal. Pelo fato, $det(A) = det(A^T)$, conclui-se que $det(A-\lambda I) = det(A^T - \lambda I)$. Logo, A e A^T possuem o mesmo polinômio característico, e desta forma possuem os mesmos autovalores. ■

Proposição 1.2.4. *Seja $A_{p\times p}$ uma matriz com $\lambda_1, \lambda_2, ..., \lambda_n$ autovalores associados a $v_1, v_2, \ldots, v_n$ respectivamente. Então a matriz $A_\alpha = (A - \alpha I)$, com $\alpha \in \mathbb{R}$, possui os autovalores $(\lambda_1 - \alpha), (\lambda_2 - \alpha), ..., (\lambda_n - \alpha)$.*

Demonstração. Veja inicialmente que $A - \lambda I = A_\alpha + \alpha I - \lambda I = A_\alpha + (\alpha - \lambda)I$. Portanto, temos que $det(A - \lambda I) = det\big(A_\alpha + (\alpha - \lambda)I\big)$. Escrevendo o polinômio característico de A, temos: $p_A(x) = (\lambda - \lambda_1)(\lambda - \lambda_2) \cdot \ldots \cdot (\lambda - \lambda_n)$ pois, por hipótese, $\lambda_1, \lambda_2, ..., \lambda_n$ são os autovalores de A, e portanto as raízes do seu polinômio característico. Assim, tomando $\lambda = \lambda_i$ para todo $i \in \{1, \cdots, n\}$, segue que $det(A - \lambda_i I) = 0$. Portanto, $det(A_\alpha + (\alpha - \lambda_i)I) = 0$. Sendo assim, $-(\alpha - \lambda_i)$ são as raízes do polinômio característico de A_α, dado por $p_{A_\alpha}(\lambda) = det(A_\alpha - \lambda I)$. Segue que $(\lambda_i - \alpha)$ é autovalor de A_α para todo $i \in \{1, \cdots, n\}$. ■

Esses são alguns resultados que consideramos importantes para a discussão dos próximos capítulos. Na sequência, trataremos diretamente das Cadeias de Markov, com definição, exemplos, classificações, entre outros. Em seguida, vamos discutir a ideia de medida estacionária de uma cadeia, e como esta pode ser estimada pelo cálculo das potências da matriz de transição respectiva.

2 Cadeias de Markov

2.1 Noções básicas e definições

Um processo estocástico de parâmetro discreto pode ser visto como uma descrição das relações entre as variáveis aleatórias $X_0, X_1, X_2, \ldots$ Aqui veremos as Cadeias de Markov como um tipo especial de processos estocásticos. Estudaremos apenas as cadeias de Markov onde os eventos acontecem em pontos discretos de um espaço de estados Ω finito ou infinito enuméravel, e X_t representa o estado de um evento que ocorre no instante t, geralmente atribuída a ideia de tempo. O aspecto interessante das cadeias de Markov é o fato de esturdamos fenômenos onde o estado atual da cadeia pode ser descrito conhecendo-se apenas o estado imediatamente anterior, independentemente de todo o histórico anterior da coleção de variáveis.

Definição 2.1.1. *Seja Ω um espaço de estados. Uma* ***cadeia de Markov*** *em Ω é uma coleção $\{X_t : t \in T\}$ de variáveis aletórias com valores em Ω, tal que, se $\mathbb{P}(X_0 = x_0, X_1 = x_1, \ldots, X_t = x_t) \neq 0$ é válida a propriedade:*

$$\mathbb{P}[X_{t+1} = y | X_0 = x_0, X_1 = x_1, \ldots, X_t = x] = \mathbb{P}[X_{t+1} = y | X_t = x],$$

$\forall t \in T$ e para todos $x_0, x_1, x_2, ..., x_t \in \Omega$.

Podemos entender a propriedade acima como se a determinação de um estado dependesse unicamente do estado imediatamente anterior a esse, e todo o passado pudesse ser descartado. Como uma espécie de "perda de memória", essa propriedade é conhecida como **propriedade Markoviana**.

Observação 2.1.1. *Neste trabalho, estudaremos as cadeias de Markov em tempo discreto e fixaremos então $T = \mathbb{N}$, a menos que se diga o contrário. Também estudaremos apenas cadeias de Markov com espaço de estados finito e podemos considerar, sem perda de generalidade que $\Omega \subset \mathbb{N}$*

Durante o estudo de cadeias de Markov, dependendo do espaço de estados, teremos que recorrer ao cálculo de diversas probabilidades condicionais. Em virtude disso, se faz necessária uma boa maneira de organizar os cálculos e, para isto, introduziremos algumas notações de teoria das matrizes. Primeiramente estudaremos alguns exemplos que possibilitem uma interpretação mais simples das cadeias de Markov.

Exemplo 2.1.1. *Suponha que determinada cidade conta com três estações de aluguel temporário de bicicletas (Estação Norte, Estação Centro e Estação Sul) e desejamos estudar* *o comportamento das bicicletas ao longo dos dias conforme forem alugadas. Suponha que se determine os seguintes dados:*

- *Das bicicletas retiradas na Estação Norte, 50% retornam ao local no fim do dia, 25% finalizam o dia na Estação Centro e o restante na Estação Sul.*
- *Das bicicletas retiradas na Estação Centro, 50% terminam o dia na Estação Norte e o restante na Estação Sul.*
- *Das bicicletas retiradas na Estação Sul, 25% terminam o dia na Estação Norte, 25% na Estação Centro e o restante retorna a Estação Sul.*

Note que o processo acima pode ser descrito por uma cadeia de Markov, de estado $\Omega = \{1, 2, 3\}$, onde 1 representa a Estação Norte, 2 a Estação Centro e 3 a Estação Sul. Note que independentemente do trajeto que a bicicleta fez nos dias anteriores com outros clientes, isso não interfere no trajeto que um novo cliente irá fazer. Podemos então determinar a probabilidade da bicicleta estar num estado ao final do dia sabendo apenas o estado em que a bicicleta iniciou o dia.

Definição 2.1.2. *Sejam $i, j \in \Omega$ e $t \in \mathbb{N}$, chama-se **probabilidade de transição da cadeia** do estado i para o estado j, $p_{ij} = \mathbb{P}[X_{t+1} = j | X_t = i]$. Isto é, a probabilidade de estar no estado j sabendo que imediatamente antes estava-se no estado i da cadeia.*

Definição 2.1.3. *Sejam $p_{i,j}, \forall i, j \in \Omega$, as probabilidades de transição da cadeia de Markov, chama-se **matriz de transição da cadeia de Markov** a matriz $P = [p_{ij}]_{i,j\in\Omega}$.*

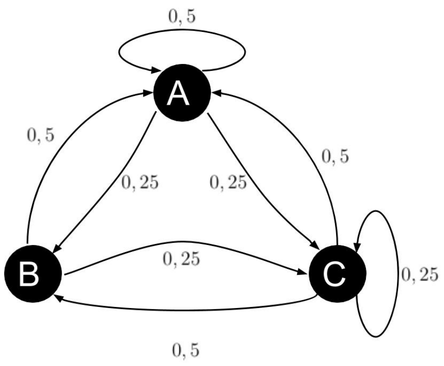
Com isso, podemos identificar as seguintes probabilidades de transição da cadeia no Exemplo 2.1.1 e ainda sistematizar tais dados de forma matricial, onde teremos:

$$\begin{cases} p_{11} = p_{2,1} = p_{2,3} = p_{3,3} = 0,5 \\ p_{1,2} = p_{1,3} = p_{3,1} = p_{3,2} = 0,25 \\ p_{2,2} = 0 \end{cases}$$

E matriz de transição P:

$$P = \begin{pmatrix} 0,5 & 0,25 & 0,25 \\ 0,5 & 0 & 0,5 \\ 0,25 & 0,25 & 0,5 \end{pmatrix}$$

Podemos ainda representar em forma de grafo, onde os estados são os vértices e existe um arco entre os vértices i e j se, e só se $p_{ij} > 0$



Notemos que cada elemento p_{ij} representa a probabilidade de estar no estado j, sabendo que imediatamente antes, estava no estado i. Sendo assim, para todo $i \in \Omega$, temos que $\sum_{j \in \Omega} p_{ij} = 1$. Em outras palavras, a soma de todos os elementos de uma mesma linha sempre vale um. Uma matriz com essa propriedade é denominada **matriz estocástica**.

Definição 2.1.4. *Seja $P = [p_{ij}]_{i,j \in \Omega}$ uma matriz quadrada. Diz-se que P é uma* ***matriz estocástica*** *se ambas as propriedades são válidas:*

1. $0 \leq p_{ij} \leq 1, \forall i, j \in \Omega$
2. $\sum_{j \in \Omega} p_{ij} = 1, \forall i \in \Omega$

É necessário expandir as transições na Cadeia de Markov para vários passos, isto é, determinar as probabilidades de transições entre estados após um período k de tempo. Ou seja, queremos determinar a probabilidade transição do estado i no tempo t para o estado j no tempo $(t + k)$ ou, melhor dizendo, a probabilidade de transição entre os estados após k passos.

Definição 2.1.5. *Sejam $i, j \in \Omega$ e $k \in \mathbb{N}$, denotaremos por $p_{ij}^{(k)}$ a* ***probabilidade de transição*** *do estado i para o estado j após k passos na cadeia de Markov, onde $p_{ij}^{(k)}$ é dada por:*

$$p_{ij}^{(k)} = \mathbb{P}[X_{t+k} = j | X_t = i]$$

Observação 2.1.2. *Para um caso particular com $k = 2$, podemos perceber que a mudança do estado i para o estado j em dois passos deve ocorrer de forma que no primeiro passo se transite do estado i para um estado n e imediatamente depois se transite do estado n para o estado j. Logo, como n é um estado qualquer de Ω, então temos o caso particular onde:*

$$p_{ij}^{(2)} = \sum_{v \in \Omega} p_{iv} \cdot p_{vj}$$

E através de uma indução podemos encontrar a forma para n passos na cadeia onde $p_{ij}^{(n)} = \sum_{v\in\Omega} p_{iv} \cdot p_{vj}^{n-1}$

Ainda no Exemplo 2.1.1, podemos verificar a probabilidade de uma bicicleta estar no estado k daqui a dois dias, sabendo que hoje ela se encontra no estado i. Tome por exemplo a pergunta: Qual a probabilidade de uma bicicleta estar na Estação Centro (2) daqui a 2 dias, sabendo que hoje ela se encontra na Estação Norte (1)? O evento da bicicleta estar na Estação Centro (2) daqui a 2 dias é a união disjunta dos eventos:

- Permanecer no Norte (1) e depois ir para o Centro (2);
- Ir para o Centro (2) e depois permanecer no Centro (2);
- Ir para o Sul (3) e depois ir para o Centro (2);

Em termos de probabilidade, permanecer no Centro (2) e depois ir para o Norte (1) é a probabilidade condicional da bicicleta ir para o Centro (2) dado que ela está no Centro (2) multiplicada pela probabilidade condicional de ir para o Norte (1) dado que ela está no Centro (2). Com raciocínio similar para os outros eventos, teremos que:

$$p_{12}^{(2)} = p_{11} \cdot p_{12} + p_{12} \cdot p_{22} + p_{13} \cdot p_{32}.$$

Note que o lado direito da equação é o produto escalar do vetor primeira linha de P pelo vetor segunda coluna de P.

Exemplo 2.1.2. *Dada uma playlist de música com músicas apenas nos estilos pop e rock, suponha que o modo de reprodução automático baseie-se apenas na música atual para determinar a próxima música que irá reproduzir. Assim, se uma música pop é reproduzida, a chance da próxima música ser pop é dada por p, e se foi reproduzida uma música de rock, a probabilidade de voltar a reproduzir uma música de rock é dada por q.*

É possível descrever o processo anterior via cadeias de Markov, com espaço de estados $\Omega = \{1, 2\}$, em que 1 indica pop e 2 indica rock, supondo que toda música só possa ser classificada em apenas um dos estilos. Podemos considerar as probabilidades de transições da seguinte maneira:

$$\begin{cases} p_{11} = p \\ p_{12} = 1 - p \\ p_{21} = 1 - q \\ p_{22} = q \end{cases}$$

Reescrevendo da sua forma matricial, teremos que a matriz de transição da cadeia é dada por:

$$P = \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix}$$

Definição 2.1.6. *Chama-se **vetor probabilidade**, qualquer vetor* $v = (v_1, v_2, ..., v_k), k \in \mathbb{N}$,*tal que* $0 \leq v_i \leq 1$*, para todo* $i \in \{1, 2, ..., k\}$ *e* $v_1 + v_2 + ... + v_k = 1$.

Observação 2.1.3. *Sendo assim, podemos tomar um vetor probabilidade para representar a **distribuição inicial** da cadeia de Markov. Ou seja, podemos caracterizar uma cadeia de Markov por um espaço de estados* Ω*, uma distribuição inicial* v_0 *e uma matriz de transição de transição* P.

Sob as condições do Exemplo 2.1.2, considere a informação que a primeira música reproduzida foi pop. Desejamos obter as probabilidades da n-ésima música reproduzida ser pop ou rock. Tomaremos o vetor $v_n = (v_n(1), v_n(2))$ que representa as probabilidades de ser pop ou rock após n trocas de músicas e o vetor $v_0 = (1 \quad 0)$, que representa a reprodução inicial ser pop.

$$\begin{cases} v_n(1) = \mathbb{P}[X_n = 1 | X_0 = 1] \\ v_n(2) = \mathbb{P}[X_n = 2 | X_0 = 1] \end{cases}$$

Então, pela Definição 2.1.5, temos:

$$\begin{aligned} v_{n+1}(1) &= \mathbb{P}[X_{t+1} = 1 | X_0 = 1] \\ &= \mathbb{P}[X_{t+1} = 1, X_t = 1 | X_0 = 1] + \mathbb{P}[X_{t+1} = 1, X_t = 2 | X_0 = 1] \\ &= \mathbb{P}[X_{t+1} = 1 | X_t = 1, X_0 = 1] \cdot \mathbb{P}[X_t = 1 | X_0 = 1] + \\ &\qquad \mathbb{P}[X_{t+1} = 1 | X_t = 2, X_0 = 1] \cdot \mathbb{P}[X_t = 2 | X_0 = 1] \\ &= p_{11} \cdot v_n(1) + p_{21} \cdot v_n(2) \\ &= p \cdot v_n(1) + (1-q) \cdot v_n(2) \end{aligned}$$

De forma análoga, podemos determinar $v_{n+1}(2) = (1-p) \cdot v_n(1) + q \cdot v_n(2)$. Perceba agora que o produto do vetor v_n pela matriz de transição P resultaria no vetor v_{n+1} encontrado, o que nos leva a próxima observação bastante útil.

Observação 2.1.4. *Seja* $\{X_i\}_{i \in \Omega}$ *uma cadeia de Markov com espaço de estados* Ω*, matriz transição* P *e o vetor* v_t *a distribuição de* X_t*:* $v_t(x) = \mathbb{P}\{X_t = x\}, \forall x \in \Omega$*. Sendo assim, podemos determinar a distribuição da cadeia no tempo* $t+1$ *como:*

$$v_{t+1} = v_t \cdot P$$

E ainda:

$$v_t = v_0 \cdot P^t$$

Isso nos mostra que a probabilidade de transição do estado i *para o estado* j *em* t *passos é a* ij*-ésima entrada da matriz de transição* P^t.

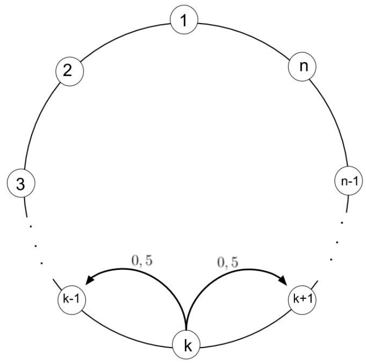
Com o auxílio do software MATLAB 2011R, calculamos algumas potências da matriz de transição P do Exemplo2.1.1:

$$A^2 = \begin{pmatrix} 0.4375 & 0.25 & 0.3125 \\ 0.375 & 0.375 & 0.25 \\ 0.4375 & 0.1875 & 0.375 \end{pmatrix} A^3 = \begin{pmatrix} 0.4218 & 0.2656 & 0.3125 \\ 0.4375 & 0.2187 & 0.3437 \\ 0.4062 & 0.2968 & 0.2968 \end{pmatrix}$$

$$A^5 = \begin{pmatrix} 0.4209 & 0.2636 & 0.3154 \\ 0.4238 & 0.2558 & 0.3203 \\ 0.4189 & 0.2685 & 0.3125 \end{pmatrix} A^{50} = \begin{pmatrix} 0.4210 & 0.2631 & 0.3157 \\ 0.4210 & 0.2631 & 0.3157 \\ 0.4210 & 0.2631 & 0.3157 \end{pmatrix}$$

Como essas matrizes A^k nos fornecem as probabilidades após k passos, analisando na ordem crescente de k podemos notar que a matriz "converge"para uma igualdade entre as linhas. Os vetores linhas na matriz A^5, por exemplo, possuem diferença nas entradas apenas na segunda casa decimal. A matriz A^{50} possui linhas iguais se tormarmos apenas 4 casas decimais. Isso nos indica que as potências de A tendem a gerar uma matriz com todas as suas linhas iguais. Já que $p_{11} = p_{21} = p_{31}$, isso indica que a probabilidade de estar no estado (1) após alguns passos independe do estado inicial da cadeia, e isso ocorre ainda para todos os outros estados da cadeia. Adiante estudaremos a distribuição estacionária de uma Cadeia de Markov e veremos sob quais condições a distribuição estacionária existe e é única.

Exemplo 2.1.3. *Considere um círculo com n pontos igualmente espaçados conforme figura a seguir:*



Denominaremos o exemplo anterior de Passeio Aleatório Simples no n-Ciclo se as probabilidade de transição entre os pontos do ciclo forem:

$$\begin{cases} p_{xy} = \dfrac{1}{2}, \text{ se } x \equiv y \pm 1 \bmod n \\ 0, \text{ caso contrário} \end{cases}$$

Dessa forma, teremos a matriz de transição $P_{n\times n}$ dada por:

$$P = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \dots & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \dots & 0 & \frac{1}{2} & 0 \end{pmatrix}$$

2.2 Classificações de uma Cadeia de Markov

O objetivo desta seção é classificar as cadeias de Markov sob algumas características necessárias para estudos posteriores. Classificaremos as cadeias quanto a sua periodicidade, bem como o fato de uma cadeia ser redutível ou irredutível.

Definição 2.2.1. *Uma Cadeia de Markov é* ***irredutível*** *se dados quaisquer $i, j \in \Omega$, existe $r \in \mathbb{N}$ tal que $p_{ij}^{(r)} > 0$, ou seja, é sempre possível transitar entre dois estados com uma quantidade r de passos. Caso contrário, dizemos que a Cadeia de Markov é* ***redutível****.*

Definição 2.2.2. *Dada uma cadeia de Markov de espaço de estado Ω finito, definimos $per(x)$ ou* ***período do estado x*** *$\in \Omega$, como:*

$$per(x) = m.d.c(T_x), \text{ onde } T_x = \{t \geq 1; p_{xx}^{(t)} > 0\}$$

Observação 2.2.1. *A definição acima nos mostra que o período de um estado é o máximo divisor comum de todos os $r's$ onde r é a quantidade de passos utilizados para retornar ao estado x.*

Definição 2.2.3. *Se para todo $x \in \Omega$, per(x)=1, então a cadeia de Markov é dita* ***aperiódica****. Caso contrário, a cadeia é dita* ***periódica****.*

Observação 2.2.2. *Se uma cadeia de Markov é irredutível, então m.d.c(T_x)=m.d.c(T_y), para todos $x, y \in \Omega$.*

Para uma melhor compreensão da Definição 2.2.3, estudaremos o Exemplo 2.1.3 do Passeio Aleatória Simples no n-Ciclo. Tal cadeia é irredutível já que qualquer estado pode ser alcançado em uma quantidade finita de passos a partir de um outro estado. Analisando apenas o período de um dos estados da cadeia, podemos verificar se tal cadeia é periódica ou aperiódica. Para isso, tome um estado k qualquer da cadeia. Uma das formas de retornar ao estado k é caminhar uma quantidade p de passos num sentido e depois retornar até o estado k no outro sentido, dando portanto $2p$ passos na cadeia. Um outra maneira de retornar ao estado k é andar apenas em um sentido da cadeia até que o ciclo seja completo (por exemplo, o caminho $1 \to 2 \to 3 \to \cdots \to n \to 1$). Neste caso, são necessários, no mínimo, n passos para o retorno. Claramente, $2 \in T_k$ e $n \in T_k$. Sendo assim:

1. Se n for par, então $m.d.c(2, n) = 2$ o que segue que o $per(k) = 2$ e a cadeia é periódica.

2. Se n for ímpar, então $m.d.c(2, n) = 1$ o que segue que $per(k) = 1$ e a cadeia é aperiódica pela Definição 2.2.3.

Definição 2.2.4. *Uma cadeia de Markov é dita* ***regular*** *se existe um $t_0 \in \mathbb{N}$, tal que para todo $t > t_0$, $p_{ij}^{(t)} > 0$, para todo $i, j \in \Omega$.*

2.3 Medida Estacionária

O estudo da medida estacionária é fundamental para a realização de previsões a longo prazo sobre o processo markoviano em questão. Nesta seção, apresentaremos a definição, alguns exemplos, e discutiremos as condições necessárias para a existência e unicidade da medida estacionária. Ademais, veremos como calcular a medida estacionária de uma Cadeia de Markov por meio das potências de sua matriz de transição.

Definição 2.3.1. *Seja $\{X_t : t \geq 0\}$ uma Cadeia de Markov com matriz de transição P. Um vetor de probabilidade π é uma medida estacionária da Cadeia de Markov se:*

$$\pi \cdot P = \pi$$

Exemplo 2.3.1. *Seja $\{X_t : t \geq 0\}$ uma Cadeia de Markov com matriz de transição P e espaço de estados $\Omega = 1, 2$ onde P é dado por:*

$$P = \begin{pmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{pmatrix}$$

A cadeia é claramente aperiódica e irredutível pois P possui todas as entradas $p_{i,j} > 0$. Queremos identificar uma medida estacionária π tal que $\pi \cdot P = \pi$. Logo, $\pi \cdot P = \pi \cdot I$ o que implica em $\pi \cdot (P - I) = 0$ e escrevendo $\pi = (\pi_1 \;\; \pi_2)$, temos:

$$(\pi_1 \;\; \pi_2) \cdot \begin{pmatrix} (0.7 - 1) & 0.3 \\ 0.2 & (0.8 - 1) \end{pmatrix} = (0 \;\; 0)$$

Temos então um sistema homogêneo que pode ser escrito como:

$$\begin{cases} \pi_1 \cdot (-0,3) + \pi_2 \cdot (0,2) = 0 \\ \pi_1 \cdot (0.3) + \pi_2 \cdot (-0,2) = 0 \end{cases}$$

Agora note que o sistema possui infinitas soluções. Porém, como o vetor π *deve ser um vetor de probabilidade, ou seja, a soma de suas entradas resulta em 1, devemos ter* $\pi_1 + \pi_2 = 1$. *Portanto, temos o sistema linear:*

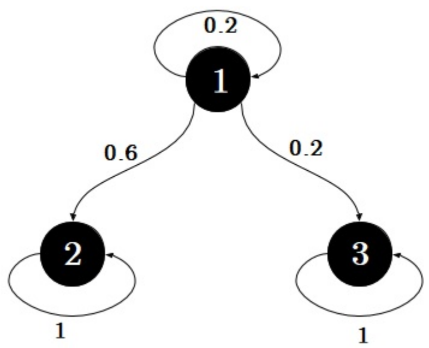
$$\begin{cases} \pi_1 \cdot (-0,3) + \pi_2 \cdot (0,2) = 0 \\ \pi_1 + \pi_2 = 1 \end{cases}$$

Como $\pi_1 = 0,4$ *e* $\pi_2 = 0,6$ *é uma solução para o sistema, então o vetor probabilidade* $\pi = (0,4 \;\; 0,6)$ *é uma medida estacionária para a cadeia.*

Observação 2.3.1. *A denominação de medida estacionária segue diretamente do fato de que quando uma Cadeia de Markov atinge uma medida estacionária, ela permanecerá nessa medida na continuidade do tempo. Ou seja, se* π *é medida estacionária e existe um* $t \in \mathbb{N}$ *tal que* $v_t = \pi$, *então* $\forall k > t$ *temos* $v_k = \pi$. *Isso decorre de* $v_{t+1} = v_t \cdot P$ *e* $v_t \cdot P = \pi$ *se* π *for medida estacionária.*

Veja também que em uma Cadeia de Markov, pode existir mais de uma medida estacionária, como no exemplo a seguir:

Exemplo 2.3.2. *Seja* $\{X_t : t \geq 0\}$ *uma Cadeia de Markov com matriz de transição* P *e estados* $\Omega = \{1, 2, 3\}$.



$$P = \begin{pmatrix} 0.2 & 0.6 & 0.2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Podemos representar a Cadeia pelo seguinte grafo:

Note que o vetores de probabilidade $z = (0 \;\; 1 \;\; 0)$ e $w = (0 \;\; 0 \;\; 1)$ são medidas estacionárias, pois:

$$(0 \;\; 1 \;\; 0) \cdot \begin{pmatrix} 0.2 & 0.6 & 0.2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = (0 \;\; 1 \;\; 0) \text{ e } (0 \;\; 0 \;\; 1) \cdot \begin{pmatrix} 0.2 & 0.6 & 0.2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = (0 \;\; 0 \;\; 1)$$

Além disso, qualquer combinação linear convexa (ou seja, combinação linear de coeficientes reais, não-negativos e soma dos coeficientes iguais a 1) entre duas medidas estacionárias é também uma medida estacionária. Note também que tal Cadeia de Markov é redutível, pois não existe um caminho entre os estados 2 e 3. Sendo assim, uma primeira característica para a unicidade da medida estacionária estará no fato da Cadeia ser irredutível.

Na próxima seção, discutiremos os teoremas principais que garantem, em certas condições, a existência e unicidade da medida estacionária em uma Cadeia de Markov.

2.3.1 Existência e Unicidade da Medida Estacionária

Teorema 2.3.1. *Seja $\{X_t : t \geq 0\}$ uma Cadeia de Markov com espaço de estados Ω finito. Então existe uma medida estacionária para esta cadeia*

Demonstração. Seja μ_0 uma distribuição de probabilidade em Ω. Temos que $\mu_t = \mu_0 \cdot P^t$. Considere a sequência de vetores $(z_n)_{n\in\mathbb{N}}$, tal que

$$z_n = \frac{\mu_0 + \mu_1 + \cdots \mu_{n-1}}{n} = \frac{1}{n}\sum_{t=0}^{n-1} \mu_t$$

Observe que para todo $n \in \mathbb{N}$, temos que $0 \leq \mu_n \leq 1$ e portanto a sequência $(z_n)_{n\in\mathbb{N}}$ é limitada em todas as suas coordenadas. Por um resultado de Análise, $(z_n)_{n\in\mathbb{N}}$ tem uma subsequência convergente, digamos $(z_{n_k})_{k\in\mathbb{N}}$ (LIMA, 2009). Defina $\pi = \lim_{k\to\infty} z_{n_k}$. Vamos mostrar que π é, de fato, uma medida de probabilidade.

Como $(z_n)_x \geq 0$ para todo $n \in \mathbb{N}$ e $x \in \Omega$, segue diretamente que $(\pi)_x \geq 0$ para todo $x \in \Omega$. Além disso, temos que:

$$\begin{aligned}
\sum_{x\in\Omega} \pi_x &= \sum_{y\in\Omega} \lim_{k\to\infty} \left(z_{n_k}\right)_y = \\
&= \sum_{y\in\Omega} \lim_{k\to\infty} \frac{1}{n_k} \sum_{t=0}^{n_k-1} (\mu_t)_y = \\
&= \lim_{k\to\infty} \sum_{y\in\Omega} \frac{1}{n_k} \sum_{t=0}^{n_k-1} (\mu_t)_y = \\
&= \lim_{k\to\infty} \frac{1}{n_k} \sum_{t=0}^{n_k-1} \sum_{y\in\Omega} (\mu_t)_y \\
&= \lim_{k\to\infty} \frac{1}{n_k} \sum_{t=0}^{n_k-1} 1 \\
&= \lim_{k\to\infty} \frac{1}{n_k} \cdot n_k \\
&= 1.
\end{aligned}$$

Vale ressaltar que a terceira igualdade ocorre por uma propriedade de limite. Ademais, na quarta igualdade trocamos a ordem dos somatórios, pois as somas são finitas. Portanto, segue que π é um vetor de probabilidade. Agora, vamos mostrar que π é uma medida estacionária da Cadeia de Markov, ou seja, $\pi = \pi \cdot P$. Fixando $x \in \Omega$, temos:

$$
\begin{aligned}
(\pi \cdot P)_x &= \sum_{y\in\Omega} \pi_y \cdot P_{y,x} = \\
&= \sum_{y\in\Omega} \lim_{k\to\infty} \frac{1}{n_k} \sum_{t=0}^{n_k-1} (\mu_0 \cdot P^t)_y \cdot P_{y,x} = \\
&= \lim_{k\to\infty} \frac{1}{n_k} \sum_{y\in\Omega} \sum_{t=0}^{n_k-1} (\mu_0 \cdot P^t)_y \cdot P_{y,x} = \\
&= \lim_{k\to\infty} \frac{1}{n_k} \sum_{t=0}^{n_k-1} \left(\sum_{y\in\Omega} (\mu_0 \cdot P^t)_y \cdot P_{y,x} \right) = \\
&= \lim_{k\to\infty} \frac{1}{n_k} \sum_{t=0}^{n_k-1} (\mu_0 \cdot P^t \cdot P)_x = \\
&= \lim_{k\to\infty} \frac{1}{n_k} \sum_{t=0}^{n_k-1} (\mu_0 \cdot P^{t+1})_x = \\
&= \lim_{k\to\infty} \frac{1}{n_k} \sum_{t=1}^{n_k-1} (\mu_0 \cdot P^t)_x = \\
&= \lim_{k\to\infty} \frac{1}{n_k} \left[\left(\sum_{t=1}^{n_k-1} (\mu_0 \cdot P^t)_x \right) + (\mu_0)_x - (\mu_{n_k})_x) - (\mu_0)_x + (\mu_{n_k})_x \right] = \\
&= \lim_{k\to\infty} \frac{1}{n_k} \left[\left(\sum_{t=0}^{n_k-1} (\mu_0 \cdot P^t)_x \right) - (\mu_0)_x + (\mu_{n_k})_x \right] = \\
&= \lim_{k\to\infty} \frac{1}{n_k} \sum_{t=0}^{n_k-1} (\mu_0 \cdot P^t)_x = \pi_x
\end{aligned}
$$

A terceira e quarta igualdade decorre de propriedades de limites e de somatórios, enquanto a sétima igualdade é uma mudança no inicio do somatório mas compensado com a potência de P e a penúltima igualdade ocorre do fato que $\frac{1}{n_k}(-(\mu_0)_x + (\mu_{n_k})) \to 0$ quando $k \to \infty$. Como x foi tomado de forma arbitrária, então o resultado vale para todas as entradas de π, portanto $\pi = \pi \cdot P$. Logo, π é medida estacionária da Cadeia de Markov correspondente. ∎

Para mostrar a unicidade da medida estacionária para cadeias irredutíveis e aperiódicas, vamos apresentar algumas definições e resultados preliminares.

Definição 2.3.2. *Seja $A \in \mathbb{R}^{n\times n}$, definimos:*

1. *A é uma matriz não-negativa se $a_{ij} \geq 0$ para todo i, j (denotamos por $A \geq 0$);*

2. *A é uma matriz positiva se $a_{ij} \geq 0$ para todo i, j e $a_{ij} > 0$ para pelo menos um par de índices i, j (denotamos por $A \gtrsim 0$);*

3. *A é uma matriz estritamente positiva se $a_{ij} > 0$ para todo i, j (denotamos por $A > 0$).*

Proposição 2.3.1. *Seja P a matriz de transição de uma cadeia de Markov irredutível e aperiódica. Então, existe $k \in \mathbb{N}$, $k \geq 1$ tal que $P^k > 0$.*

Demonstração. Considere $x \in \Omega$ e defina $T_x = \{t \geq 1, p_{x,x}^{(t)} > 0\}$, que representa o conjunto dos tempos de retorno para o estado x. Como a Cadeia de Markov é aperiódica, então $mdc(T_x) = 1$, conforme a Definição 2.2.3.

Note que se $t_1, t_2 \in T_x$, então $(t_1 + t_2) \in T_x$. De fato, temos $p_{x,x}^{t_1+t_2} = \sum_{k \in \Omega} p_{x,k}^{t_1} \cdot p_{k,x}^{t_2}$ (Equação de Chapman-Kolmogorov), e como $p_{x,x}^{t_1} > 0$ e $p_{x,x}^{t_2} > 0$, segue que $p_{x,x}^{t_1} \cdot p_{x,x}^{t_2} > 0$.

Temos também que se $mdc(T_x) = 1$, existe $t_0 \in \mathbb{N}$ tal que para todo $t > t_0$, t pode ser escrito como $t = t_1 + t_2 + \cdots + t_n$ onde $t_1, t_2, \cdots, t_n \in T_x$. Este resultado pode ser encontrado em FELLER (1993, pag. 336).

De maneira geral, podemos definir t_{0_x} para todo $x \in \Omega$, isto é, a quantidade mínima de passos para a probabilidade de retorno ser estritamente positiva. Defina ainda $k_1 = max\{t_{0_i} : i \in \Omega\}$. Do fato da cadeia ser irredutível, então temos a existência de $t_{(x,y)} \in \mathbb{N}$ tal que $p_{x,y}^{t_{(x,y)}} > 0$, para todo $x, y \in \Omega$. Tome $k_2 = max\{t_{(x,y)} : x, y \in \Omega\}$.

Seja $k = k_1 + k_2$. Vamos mostrar que $P^k > 0$. Fixe $i, j \in \Omega$, e seja $\alpha = t_{(i,j)}$ o menor tempo em que a probabilidade de transição entre os estados i e j é positiva. Tem-se que $k_2 \geq \alpha$, logo $k_2 - \alpha \geqslant 0$, então é válido que $p_{j,j}^{k_1+k_2-\alpha} > 0$ pois $k_1 \geqslant t_{0_j}$ e como $k_1 \leqslant k_1 + k_2 - \alpha$, isso implica em $k_1 + k_2 - \alpha \geqslant t_{0_j}$. Ou seja, temos a garantia da existência de um caminho entre i e j com probabilidade de transição estritamente positiva, pois $p_{i,j}^{\alpha} > 0$ e $p_{j,j}^{k_1+k_2-\alpha} \geqslant p_{j,j}^{t_{0_j}} > 0$. E como i e j foram tomados arbitrariamente, temos o resultado para todas as entradas da matriz P. Logo, $P^k > 0$. ■

Proposição 2.3.2. *Seja P uma matriz de transição $n \times n$ de uma Cadeia de Markov. Então $\lambda = 1$ é autovalor de P.*

Demonstração. Seja v um vetor coluna com todas as entradas iguais a 1. Então $P \cdot v = v$ pois como P é matriz de transição, a soma de suas linhas é igual a 1. ■

O resultado acima garante que $\lambda = 1$ é autovalor de P^T, pois uma matriz e a sua transposta tem os mesmos autovalores, conforme Proposição 1.2.3. Dessa forma, existe um vetor não nulo v tal que $P^T v = v$, e tomando a transposta, segue que $v^T P = v^T$. Com os resultados a seguir, vamos mostrar que podemos escolher v^T que satisfaça as condições de uma medida de probabilidade.

Teorema 2.3.2. ***(Perron-Frobenius).*** *Seja $A > 0$ uma matriz estritamente positiva de ordem n. Então, existem $\lambda_0 > 0$ e $x_0 > 0$ tais que:*

1. *$Ax_0 = \lambda_0 x_0$;*
2. *Se $\lambda \in \mathbb{R}$ for um autovalor de A, com $\lambda \neq \lambda_0$, então $|\lambda| < \lambda_0$;*
3. *λ_0 tem multiplicidade geométrica igual a 1.*

Demonstração. Defina $\Lambda = \{\lambda \in \mathbb{R} : Ax \geq \lambda x \text{ para algum } x \gtrsim 0\}$ e seja $\lambda_0 = sup(\Lambda)$ Note que $0 \in \Lambda$ pois se $A > 0$ e $x \gtrsim 0$, então $Ax > 0$ enquanto $\lambda x = 0$. Além disso, vamos mostrar que, de fato, Λ admite um supremo $\lambda_0 \in \Lambda$. Denotando $x = (x_1, x_2, \ldots, x_n)$ e seja $M = max\left\{\sum_{j\in\Omega} a_{ij} : i \in \{1, \ldots, n\}\right\}$ e $m = max\{x_k : k \in \{1, \ldots, n\}\}$. Sendo assim, $Ax \leq (Mm, Mm, \ldots, Mm)$, ou seja, se $\lambda > M$, então $\lambda x > (Mm, Mm \ldots, Mm)$. E como $\lambda x \geq Ax$, então $\lambda \notin \Lambda$ se $\lambda > M$. Desta forma, $0 \leq \lambda \leq M$. Ou seja, $\Lambda \subseteq [0, M]$ e portanto $\lambda_0 = sup(\Lambda) = max(\Lambda)$.

1. Devemos mostrar que $Ax_0 = \lambda_0 x_0$. Suponha que a desigualdade estrita não ocorra, ou seja, $Ax_0 \gtrsim \lambda_0 x_0$ e $Ax_0 \neq \lambda_0 x_0$. Considere então o vetor $y_0 = Ax_0$. Como $A > 0$, então $Ax > 0$ para qualquer vetor $x \gtrsim 0$. Como $y_0 \gtrsim \lambda_0 x_0$, então $(y_0 - \lambda_0 x_0) \gtrsim 0$. Sendo assim:
$$A(y_0 - \lambda_0 x_0) = Ay_0 - \lambda_0 Ax_0 = Ay_0 - \lambda_0 y_0 > 0.$$
A desigualdade acima decorre diretamente de $y_0 - \lambda_0 x_0 \gtrsim 0$. Temos então $Ay_0 > \lambda_0 y_0$. Sendo assim, podemos encontrar um $\lambda^* > \lambda_0$ tal que $\lambda^* \in \Lambda$ e $Ay_0 \geqslant \lambda^* y_0$. Só que isso contradiz a definição de λ_0 ser o supremo do conjunto Λ. Portanto, vale que $Ax_0 = \lambda_0 x_0$.

2. Seja λ o autovalor de A, com $\lambda \neq \lambda_0$, e seja y o autovetor associado a λ. Temos então que $Ay = \lambda y$. Como $A > 0$, podemos afirmar que:
$$A|y| = |A|.|y| \geqslant |Ay| = |\lambda y| = |\lambda|.|y|.$$
Sendo assim, $|\lambda| \in \Lambda$ e pela definição de λ_0, teremos $|\lambda| \leqslant \lambda_0$. Agora devemos mostrar que a desigualdade é estrita. Escolha $\alpha > 0$ de modo que a matriz $A_\alpha = A - \alpha I$ continue estritamente positiva.

   Temos que $(\lambda_0 - \alpha)$ e $(\lambda - \alpha)$ são autovalores de A_α, segundo a Proposição 1.2.4. Como $A_\alpha > 0$, seu maior autovalor deve ser $(\lambda_0 - \alpha)$. Mas se $|\lambda| = \lambda_0$, como $\lambda \neq \lambda_0$ e $\alpha > 0$, devemos ter $|\lambda - \alpha| > \lambda_0 - \alpha$, o que contradiz a conclusão do parágrafo anterior.

3. Para mostrar a multiplicidade geométrica do autovalor λ_0, suponha que exista outro autovetor $y > 0$ associado a λ_0, e linearmente independente de x_0. Portanto, é possível obter um vetor w pela combinação linear entre x_0 e y, tal que $w = x_0 + \beta y$, onde $w \gtrsim 0$, mas não vale que $w > 0$. Sendo assim, teremos $Aw = \lambda_0 w$, e portanto $w = \dfrac{1}{\lambda_0} Aw$. Como $A > 0$, segue que $w > 0$, mas isso contradiz a definição do vetor w. Portanto, não existe um outro autovetor associado a λ_0 linearmente independente de x_0. Concluímos, assim, que λ_0 tem multiplicidade geométrica 1.

■

Observação 2.3.2. *Suponha que A seja uma matriz não-negativa ($A \geqslant 0$), e que alguma potência de A seja estritamente positiva ($A^n > 0$). Então, as conclusões do Teorema 1.3.3 também são válidas para A. Isso decorre do fato dos autovalores da potência de uma matriz A serem as potências dos autovalores da matriz A, conforme Proposição 1.2.1. Ademais, também é valido que os autovetores são os mesmos para qualquer potência de A.*

Proposição 2.3.3. *Seja P uma matriz estocástica $n \times n$ com autovalor λ. Então $|\lambda| \leqslant 1$.*

Demonstração. Seja $x^T = (x_1, x_2, \cdots, x_n)$ um autovetor de P associado ao autovalor 1. Denote por x_k uma componente do vetor x que satisfaça $|x_k| \geqslant |x_i|$ para todo $i \in \{1, 2, \cdots, n\}$, e defina $|x_k| = z$. Analisando a k-ésima componente de $Px = \lambda x$, temos:

$$p_{k1}.x_1 + p_{k2}.x_2 + \cdots + p_{kn}.x_n = \lambda x_k.$$

Como $|\lambda| z = |\lambda||x_k| = |\lambda x_k|$, segue que:

$$\begin{aligned} |\lambda x_k| &= |p_{k1}x_1 + p_{k2}x_2 + \cdots + p_{kn}x_n| \\ &\leqslant |p_{k1}x_1| + |p_{k2}x_2| + \cdots + |p_{kn}x_n| \\ &= p_{k1}|x_1| + p_{k2}|x_2| + \cdots p_{kn}|x_n| \\ &\leqslant p_{k1}z + p_{k2}z + \cdots p_{kn}z \\ &= (p_{k1} + p_{k2} + \cdots p_{kn})z = z. \end{aligned}$$

Veja que a primeira desigualdade decorre da desigualdade triangular, enquanto que a segunda desigualdade segue do fato que $|x_k| \geqslant |x_i|$. Ademais, a última igualdade decorre de A ser uma matriz estocástica. Segue então que $|\lambda| z \leqslant z$, ou seja, $|\lambda| \leqslant 1$. ■

Observação 2.3.3. *Com os resultados anteriores, podemos concluir que $\lambda_0 = 1$. De fato, pela Proposição 2.3.3, temos que 1 é autovalor de P. Ademais, pela Proposição 2.3.3, $|\lambda_0| \leqslant 1$ e pelo Teorema 2.3.2, vale que $|\lambda| < \lambda_0$ quando $\lambda \neq \lambda_0$. Além disso, por uma das conclusões do Teorema de Perron-Frobenius, λ_0 tem multiplicidade geométrica igual a 1.*

Segue então um resultado fundamental para o estudo de Cadeias de Markov, que garante que a medida estacionária é única nos casos em que a cadeia é irredutível e aperiódica:

Teorema 2.3.3. ***(Unicidade da Medida Estacionária)*** *Seja $P \geqslant 0$ uma matriz de transição de uma Cadeia de Markov irredutível e aperiódica. Então existe um único vetor estacionário π tal que $\pi \cdot P = \pi$.*

Demonstração. Pela Proposição 2.3.3 e Teorema 2.3.2, temos que existe um maior e único autovalor λ_0 de P. Pela Observação 2.3.3, sabemos que este autovalor é $\lambda_0 = 1$. Além disso, temos a multiplicidade geométrica sendo 1, e portanto existe um único vetor π com entradas não-negativas tal que $\pi P = \pi$, e satisfazendo $\sum_{i \in \Omega} \pi_i = 1$. ■

Em seguida, desejamos estimar a medida estacionária através do cálculo das potências da matriz de transição de uma cadeia de Markov. Já vimos anteriormente que a distribuição de probabilidade de uma Cadeia de Markov após n passos é dada por $v_n = v_0 \cdot P^n$. Sendo assim, investigamos o resultado dessa relação quando n cresce. O resultado a seguir garante que as iterações devem convergir para uma matriz estocástica com todos vetores-linha iguais à medida estacionária da Cadeia de Markov.

Teorema 2.3.4. *Seja $P_{n \times n}$ a matriz de transição de uma Cadeia de Markov irredutível e aperiódica. Então, quando $k \to \infty$, P^k converge para uma matriz $L_{n \times n}$, cujas linhas são idênticas e iguais à medida estacionária π da Cadeia de Markov correspondente.*

Demonstração. Inicialmente, a existência do limite da sequência $\left(P^k\right)_{k \in \mathbb{N}}$ segue do fato que a sequência $\left(v_0 P^k\right)_{k \in \mathbb{N}}$ é de Cauchy para todo vetor de probabilidade v_0 (veja, por exemplo, DEMERS (2014)). Dessa forma, tomando os vetores canônicos, temos garantida a convergência em cada linha da sequência de matrizes, e portanto, existe $L = \lim_{k \to \infty} P^k$. Ademais, podemos escrever:

$$L \cdot P = (\lim_{k \to \infty} P^k) \cdot P = \lim_{k \to \infty} (P \cdot P^k) = \lim_{k \to \infty} P^{k+1} = L.$$

Denotando por z_i a i-ésima linha de L, e usando o fato que $L \cdot P = L$, temos que $z_i \cdot P = z_i$. Isto implica que toda linha de L é um autovetor de P associado a $\lambda_0 = 1$. Agora, basta mostrar que z_i é um vetor de probabilidade. De fato, seja j um vetor-coluna com todas as entradas iguais a 1. Então:

$$L \cdot j = (\lim_{k \to \infty} P^k) \cdot j = \lim_{k \to \infty} (P^k \cdot j) = \lim_{k \to \infty} j = j.$$

Finalmente, como L é matriz estocástica e $z_i \cdot P = z_i$ para todo $i \in \{1, 2, \cdots, n\}$, segue que z_i é uma medida estacionária da Cadeia de Markov por definição. Como o Teorema 2.3.5 garante a unicidade do vetor estacionário, temos que todas as linhas de L devem ser iguais à medida estacionária π da Cadeia de Markov. ■

É interessante notar que o vetor estacionário independe da distribuição inicial da Cadeia de Markov quando a Cadeia é irredutível e aperiódica. Já vimos anteriormente que as Cadeias que não são irredutíveis e aperiódicas podem conter mais de um vetor estacionário, como no Exemplo 2.3.2. Além disso, para essas Cadeias, o resultado anteriormente visto para as potências de P^k também não é valido. De fato, tomando a mesma matriz do Exemplo 2.3.2, qualquer que seja a potência k, as linhas 2 e 3 permanecerão como estão e portanto não teremos todas as linhas iguais.

Um outro detalhe interessante é que as Cadeias de Markov que não são irredutíveis e aperiódicas podem evoluir para diferentes vetores estacionários, dependendo da distribuição inicial considerada. Veja o exemplo a seguir:

Exemplo 2.3.3. *Seja* $\{X_t : t \geq 0\}$ *uma Cadeia de Markov com matriz de transição* P *e estados* $\Omega = \{1, 2, 3, 4\}$ *onde:*

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

Neste caso, qualquer vetor $v = \begin{pmatrix} 0 & 0 & \alpha & (1-\alpha) \end{pmatrix}$ *com* $0 \leq \alpha \leq 1$ *será um vetor estacionário, pois* $v \cdot P = v$. *Sendo assim, ao tomarmos a distribuição inicial* $v_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix}$, *o processo evolui para o vetor estacionário* $\pi_v = \begin{pmatrix} 0 & 0 & 0.5 & 0.5 \end{pmatrix}$. *Por outro lado, ao tomarmos a distribuição inicial* $z_0 = \begin{pmatrix} 0 & 1 & 0 & 0 \end{pmatrix}$, *o processo evolui para o vetor estacionário* $\pi_z = \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix}$.

No próximo capítulo, apresentaremos algumas aplicações de Cadeias de Markov, que ilustram a convergência para a medida estacionária mencionada nos resultados anteriores. No primeiro exemplo, referente ao jogo Snakes and Ladders, vamos analisar em quantos passos o processo se aproxima da medida em questão e, no segundo exemplo, voltado para a área de Biologia, vamos estimar a medida estacionária por meio do cálculo das potências da matriz de transição.

3 Aplicações de Cadeias de Markov

Acreditando que apenas a exposição algébrica de alguns conceitos não seja suficiente para um entendimento completo das Cadeias de Markov, fez-se necessário o estudo de algumas aplicações. Neste capítulo, através de um direcionamento diferente daquele apresentado nos livros clássicos, abordaremos duas aplicações interessantes dos resultados apresentados anteriormente. Optando por aplicações bastante distintas, podemos perceber que os conceitos estudados podem ser usados em uma gama variada de processos.

Num primeiro momento, abordaremos o jogo de tabuleiro "Snakes and Ladders" (ou Cobras e Escadas) que pertence ao grupo dos Jogos de Markov, ou seja, jogos que podem ser modelados por Cadeias de Markov. Para um aprofundamento neste sentido, outras informações podem ser encontradas em (FRANCO, 2012) *Jogos markovianos alternados sob incerteza*. Por fim, trabalharemos com uma aplicação na Biologia para o controle populacional de algumas espécies de animais, destacando a importância da medida estacionária para o processo.

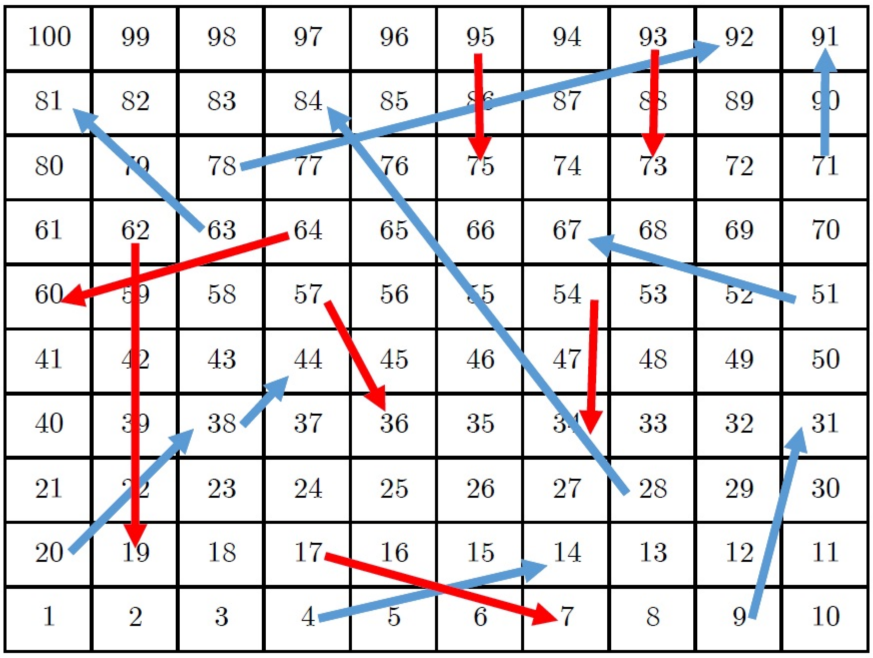
3.1 O Jogo Snakes and Ladders

O jogo Snakes and Ladders (ou Cobras e Escadas) é um jogo de tabuleiro com origem na Índia. Joga-se com o tabuleiro, um peão e um dado de 6 faces. O tabuleiro é, geralmente, quadrado e composto por "casas"numeradas de 1 a 100 (também encontra-se versões de tamanhos diferentes). O jogador começa com o peão na casa 1 e deverá leva-lo até a casa 100. Para isso, o jogador lança o dado e o número da face superior é a quantidade de casas que o peão deverá andar. Para auxiliar ou atrapalhar esse desenvolvimento do peão, temos algumas escadas e algumas cobras no tabuleiro.

As escadas servem de auxílio para o peão, sendo assim, ao parar em uma casa que possui o inicio da escada, o peão é levado imediatamente para o final da escada, pulando várias posições no tabuleiro. As cobras servem de obstáculos para o peão, ou seja, ao parar na casa que possui a cabeça de uma cobra, o peão é levado para a casa que contém o final da cobra, fazendo com que o peão volte várias casas no tabuleiro. Ao parar no final de uma cobra ou no final de uma escada, não acontece nada com o peão e ele permanece nesta posição. O jogo termina quando o peão consegue atingir a casa de número 100. Além disso, se um peão se encontra, por exemplo, na casa de número 98, a face do dado superior deve ser exatamente 2 para que o jogo termine. Se a face do dado for superior a 2, então leva-se o peão até a casa de número 100 e retorna-se com o peão na quantidade de casas que superou o 2 no lançamento do dado. Por exemplo, se sair o número 5 no dado, então o peão deverá voltar para a casa 97 pois estava a 2 casas de ganhar e teve que voltar as 3 que excederam.

Vamos considerar um tabuleiro que servirá para a nossa aplicação. Neste caso, representaremos tanto as cobras como as escadas por flechas. Quando o peão cair numa casa no início da flecha, ele irá para a outra extremidade da flecha (ponta da seta). As flechas que fazem o peão ser lançado para casas maiores são as escadas e as flechas que fazem o peão voltar são as cobras.

Tabuleiro de 100 casas



Nosso objetivo passa a ser o estudo do jogo através de um modelo markoviano. Iremos identificar cada lançamento de dado e o movimento consequente do peão como um tempo da Cadeia de Markov e cada uma das casas do tabuleiro como um estado do espaço de estados Ω da Cadeia de Markov. Como já vinhamos trabalhando em exemplos anteriores, temos um tempo discreto e um espaço de estados finito. Vamos supor também que o peão inicia a sua caminhada no estado (casa) 1 da Cadeia. Assim, estamos considerando a distribuição inicial $v_0 = \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix}$.

A respeito das probabilidades de transição entre os estados, para o primeiro lançamento do dado, temos: $p_{1,2} = p_{1,3} = p_{1,5} = p_{1,6} = p_{17} = \frac{1}{6}$, que são as probabilidades de retirar 1, 2, 4, 5 ou 6 no dado. Note que $p_{1,4} = 0$ pois no estado 4 temos uma escada que nos leva a casa 14, sendo assim, $p_{1,14} = \frac{1}{6}$. O estado 7 pode ser atingido pelo estado 1 e note que possui uma flecha, mas neste caso é uma ponta de flecha que traria o peão da casa 17 para a casa 7. Ao atingir a casa 7, o peão permanece no local até a próxima jogada, ou seja, devemos nos preocupar apenas com o inicio das flechas. Além disso, note que todas as outras casas não podem ser atingidas por um peão que se encontra na casa 1 logo após o primeiro lançamento. A seguir, apresentamos um mapa com as casas que podem ser atingidas após o primeiro lançamento:

Após 1 lançamento

100	99	98	97	96	95	94	93	92	91
81	82	83	84	85	86	87	88	89	90
80	79	78	77	76	75	74	73	72	71
61	62	63	64	65	66	67	68	69	70
60	59	58	57	56	55	54	53	52	51
41	42	43	44	45	46	47	48	49	50
40	39	38	37	36	35	34	33	32	31
21	22	23	24	25	26	27	28	29	30
20	19	18	17	16	15	14	13	12	11
1	2	3	4	5	6	7	8	9	10

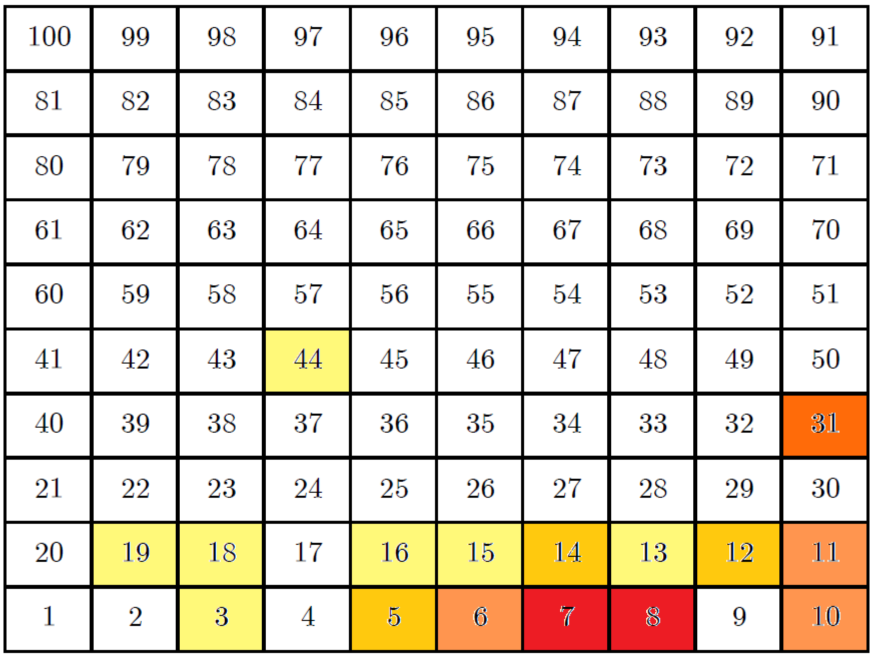
Após o segundo lançamento, as probabilidades de transição são dadas por:

- $p_{1,3}^{(2)} = p_{1,13}^{(2)} = p_{1,15}^{(2)} = p_{1,16}^{(2)} = p_{1,17}^{(2)} = p_{1,18}^{(2)} = p_{1,19}^{(2)} = p_{1,44}^{(2)} = \frac{1}{36}$
- $p_{1,12}^{(2)} = p_{1,14}^{(2)} = p_{1,15}^{(2)} = \frac{2}{36}$
- $p_{1,6}^{(2)} = p_{1,10}^{(2)} = p_{1,11}^{(2)} = \frac{3}{36}$
- $p_{1,31}^{(2)} = \frac{4}{36}$
- $p_{1,7}^{(2)} = p_{1,8}^{(2)} = \frac{5}{36}$

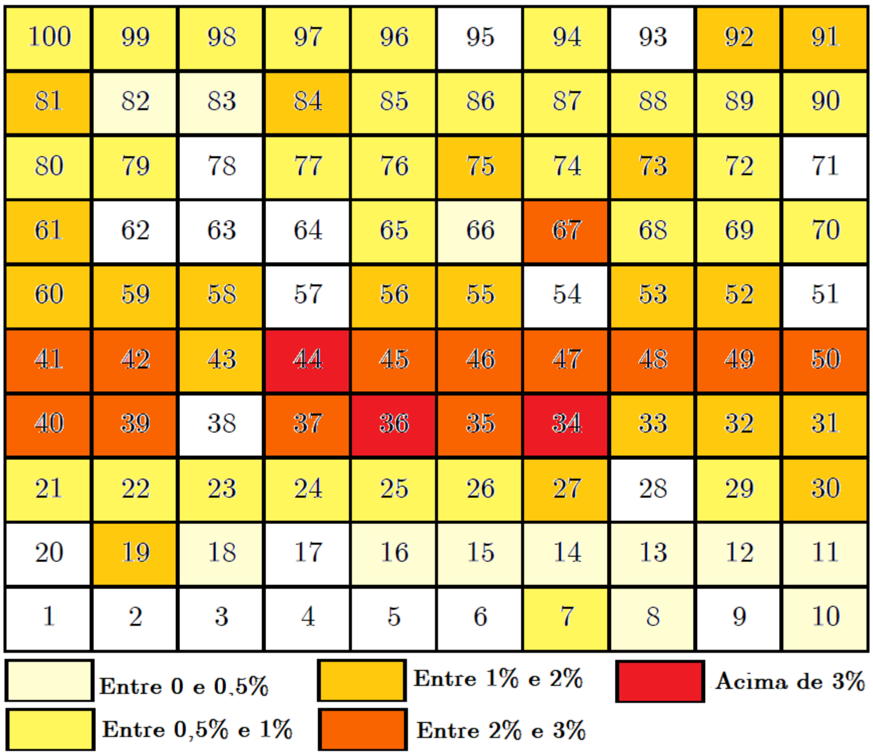
De forma análoga, podemos montar um mapa das casas que podem ser atingidas após o segundo lançamento, onde as casas menos prováveis estão em amarelo e as mais prováveis estão em vermelho. A longo prazo, veja que a cadeia deve "convergir"para a casa 100, uma vez que ao chegar na casa 100 o jogo termina e o peão não sai mais dessa casa. Como temos caminhos que nos levam a casa 100, um jogador que ficar muito tempo jogando o jogo, alcançará a casa em algum momento. Sendo assim, estamos interessados em saber as probabilidades para o peão estar após alguns lançamentos e também a probabilidade do jogo já ter terminado após esses lançamentos. Outros questionamentos ainda podem ser formulados, como por exemplo, após qual lançamento existem maior probabilidade do jogo já ter terminado do que ainda estar em andamento? Qual o número mínimo de jogadas para terminar o jogo?

Para essas questões, podemos iterar a matriz de transição e obter diversas informações. Para isso, utilizamos o software MATLAB R2017a e tomamos aproximações na quarta casa decimal. Ilustramos o mapa das probabilidades após 2 lançamentos e após 10 lançamentos.

Após 2 lançamentos



Após 10 lançamentos



Após 10 lançamentos já pudemos notar que o peão deve ficar próximo do centro do tabuleiro. As casas 34, 36 e 44 apresentaram maiores probabilidades, sendo que a casa 44 é a mais provável com $5,09\%$. Pelo tabuleiro podemos ver que existe uma escada da casa 20 para a casa 38 e outra da casa 38 para a casa 44, fazendo com que a casa 44 apresente a maior probabilidade por, digamos, terem duas escadas que levam até ela. Além disso, já é possível concluir o jogo com 10 lançamentos, a casa 100 apresenta uma probabilidade de 0.7%.

O nosso objetivo é iterar a matriz para um número de lançamentos ainda maior e perceber que a probabilidade do jogo terminar aumenta de forma significativa. Sendo assim, com 100 lançamentos já temos uma visão melhor de que o jogo deve chegar ao fim. Observe o mapa de probabilidades para 100 lançamentos:

Após 100 lançamentos

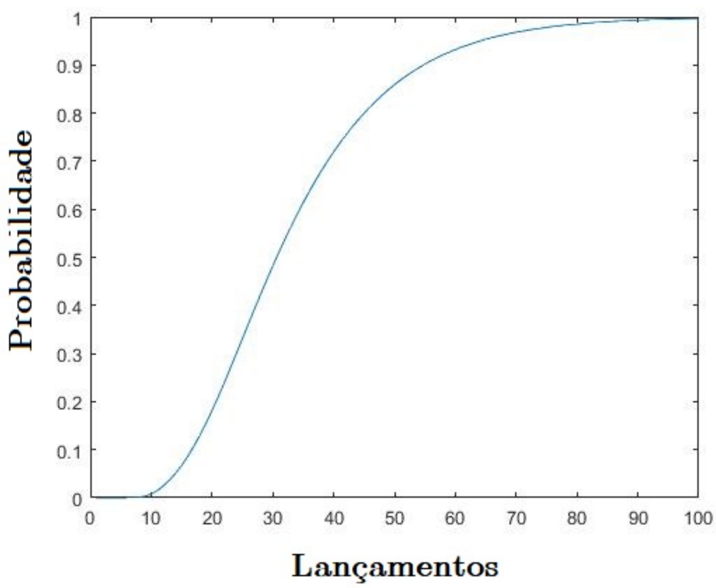
100	99	98	97	96	95	94	93	92	91
81	82	83	84	85	86	87	88	89	90
80	79	78	77	76	75	74	73	72	71
61	62	63	64	65	66	67	68	69	70
60	59	58	57	56	55	54	53	52	51
41	42	43	44	45	46	47	48	49	50
40	39	38	37	36	35	34	33	32	31
21	22	23	24	25	26	27	28	29	30
20	19	18	17	16	15	14	13	12	11
1	2	3	4	5	6	7	8	9	10

Com 100 lançamentos, obtivemos a probabilidade de $99,7\%$ do peão estar na casa de número 100, ou seja, é muito provável que o jogo termine antes de 100 rodadas. A casa 99 apresentou segunda maior probabilidade com apenas $0,04\%$.

Para ilustrar as chances do jogo finalizar após n rodadas, definimos um vetor f com 100 entradas, em que a n-ésima entrada consiste na probabilidade do peão estar na casa final após n rodadas. Algébricamente, $f_n = v_{100}^{(n)}$ onde $v^{(n)} = v_0 \cdot P^n$ e P é a matriz de transição da Cadeia de Markov. Através do software, calculamos o vetor f, que é dado por:

$$f = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0.0002 & 0.001 & 0.003 & 0.007 \\ 0.014 & 0.024 & 0.036 & 0.050 & 0.067 & 0.085 & 0.106 & 0.129 & 0.154 & 0.181 \\ 0.209 & 0.238 & 0.269 & 0.299 & 0.330 & 0.362 & 0.392 & 0.423 & 0.453 & 0.482 \\ 0.510 & 0.538 & 0.564 & 0.589 & 0.614 & 0.637 & 0.659 & 0.680 & 0.700 & 0.719 \\ 0.737 & 0.754 & 0.770 & 0.785 & 0.799 & 0.813 & 0.825 & 0.837 & 0.848 & 0.859 \\ 0.868 & 0.877 & 0.886 & 0.894 & 0.901 & 0.908 & 0.915 & 0.921 & 0.926 & 0.932 \\ 0.937 & 0.941 & 0.945 & 0.949 & 0.953 & 0.956 & 0.960 & 0.962 & 0.965 & 0.968 \\ 0.970 & 0.972 & 0.974 & 0.976 & 0.978 & 0.980 & 0.981 & 0.982 & 0.984 & 0.985 \\ 0.986 & 0.987 & 0.988 & 0.989 & 0.990 & 0.990 & 0.991 & 0.992 & 0.992 & 0.993 \\ 0.993 & 0.994 & 0.994 & 0.995 & 0.995 & 0.995 & 0.996 & 0.996 & 0.996 & 0.997 \end{pmatrix}$$

Para uma visualização melhor, construímos o gráfico da probabilidade de se atingir a casa 100 após n rodadas:

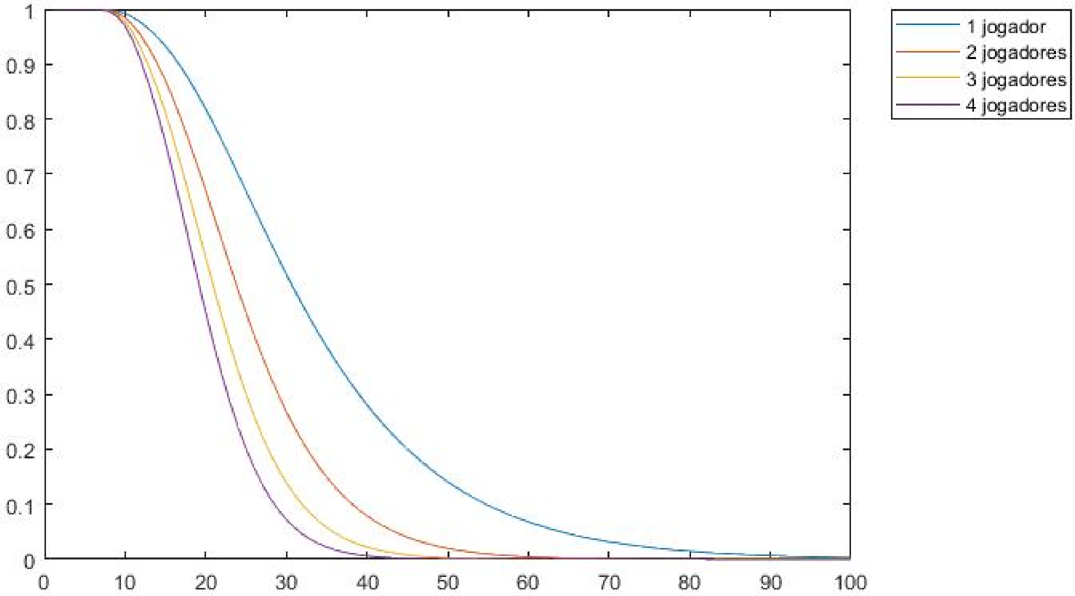


Com isso, podemos perceber que a quantidade mínima de jogadas para terminar o jogo são 7, pois é a primeira entrada positiva de f. Também é possível afirmar que, a partir de 31 rodadas, é mais provável que o jogo já tenha terminado.

Uma outra situação foi estudada. Sabemos que os jogos de tabuleiro, geralmente, possuem mais de um jogador. Existem vários peões no tabuleiro disputando para ver quem alcança o objetivo primeiro. Sendo assim, como ficaria a nossa aplicação com mais peões jogando simultaneamente? Para isso, consideramos os complementares das entradas do vetor f. Ou seja, se f representa a probabilidade de terminar o jogo após n lançamentos, então teremos um outro vetor que representa a probabilidade do jogo ainda não ter terminado após n lançamentos.

Nesse sentido, como $(1 - f_{1,n})$ representa a probabilidade de um peão não ter terminado o jogo ainda, com 2 peões em jogo basta tomar essa probabilidade ao quadrado, pois é o caso do peão 1 não ter terminado o jogo e o peão 2 também não ter terminado o jogo após n rodadas. O conceito é análogo para mais peões em jogo. Com isso, a seguir temos um gráfico que representa a probabilidade do jogo não ter terminado ainda após n rodadas com mais de um peão.

Probabilidade de não ter terminado o jogo com mais de 1 peão



É intuitivo que o jogo deve finalizar mais cedo se aumentarmos o número de jogadores. Podemos perceber pelo gráfico que a probabilidade de alguém já ter terminado o jogo é, aproximadamente, 90% perto dos 30 lançamentos quando temos 4 jogadores, enquanto apenas um jogador atinge os 90% próximo dos 50 lançamentos.

Com esta aplicação pudemos perceber que, em alguns casos, determinar apenas a medida estacionária não representa muito em informação, ou seja, já era notório que o peão iria para a casa 100, dado que existiam caminhos possíveis para isso. Sendo assim, a medida estacionária seria com 100% na casa 100. Mas, através da manipulação das potências da matriz de transição, pudemos perceber melhor o desenvolvimento do jogo e responder algumas outras questões interessantes.

3.2 Controle Populacional em espécies de animais

Uma outra aplicação interessante de Cadeias de Markov está relacionada à área da Biologia. Diversos estudos já realizados envolvem processos estocásticos para modelar o crescimento populacional de algumas espécies de animais e plantas, como por exemplo em (SILVEIRA JUNIOR, 2014), (FREITAS et al, 2005) ou ainda outra aplicação na Biologia como (FIDALGA, 2016).

Nesta sessão, estudaremos o caso das especies de esquilos cinza (*Sciurus carolinensis Gmelin*) e esquilos vermelhos (*Sciurus vulgaris L*), presentes na região da Grã-Bretanha, segundo informações disponíveis em *Some Applications of Markov Chains*, de (SWIFT, n.d.)

O esquilo vermelho é nativo dessa região e amplamente distribuído sobre o território Europeu. Por outro lado, o esquilo cinza é nativo das regiões da América, e foi levado para a Europa a partir da década de 20. Ao introduzirem o esquilo cinza nesta região, é natural o risco de um desequilíbrio biológico. Desta maneira, torna-se necessário um estudo aprofundado sobre o crescimento ou decrescimento populacional das espécies em algumas regiões, a fim de perceber um risco de extinção das espécies nativas e não nativas.

Os esquilos cinza se espalharam rapidamente por toda a Grã-Bretanha e com o passar dos anos, percebeu-se que o esquilo cinza tornou-se uma praga para as florestas. Juntamente ao crescimento dessa espécie, houve um desaparecimento dos esquilos nativos. Sendo assim, algumas medidas de controle foram necessárias. Para estudar melhor o desenvolvimento populacional, a Comissão Britânica das Florestas criou um questionário que foi entregue aos membros que frequentam essas florestas, a fim de prever as distribuições das duas espécies.

O questionário continha questões a respeito do aparecimento e desaparecimento das espécies de esquilos num período de dois anos em determinada região. Primeiramente, dividiu-se em quatro grupos:

- **C**: Registrou a presença de apenas esquilos cinza;
- **V**: Registrou a presença de apenas esquilos vermelhos;
- **A**: Registrou a presença de ambos os esquilos;
- **N**: Não registrou presença de esquilos;

Após o período de dois anos, o questionário foi novamente aplicado e colhidas as informações. Desta maneira, foi possível dividir o experimento em 16 grupos: **C** $\rightarrow$ **V** (registrou a presença de esquilos cinza e depois registrou a presença de esquilos vermelhos), **V** $\rightarrow$ **A** (registrou a presença de esquilos vermelhos e depois registrou a presença de ambos) e etc..

Os resultados obtidos no questionário podem ser vistos na tabela de frequência a seguir:

Sendo assim, podemos modelar o experimento via Cadeias de Markov. Vamos tomar como espaço de estados $\Omega = \{C(1), V(2), A(3), N(4)\}$. Os números da primeira linha representam os questionários daqueles que, na primeira resposta, registraram a presença

Tabela 1 – Questionários de distribuições de frequência

	C	V	A	N
C	2529	35	257	5
V	61	733	20	91
A	282	25	4311	335
N	3	123	310	5930

de esquilos cinza. Sendo assim, 2529 são os membros que viram esquilos cinza e continuaram a ver esquilos cinza unicamente, 35 viram esquilos cinza mas na segunda resposta registraram apenas esquilos vermelhos, 257 registraram esquilos cinza e agora registram a aparição de ambas as espécies e 5 registraram a presença dos esquilos cinza e agora não registram a aparição de nenhuma das duas espécies. A interpretação é semelhante para as demais linhas.

Podemos, desta forma, obter uma matriz de transição da Cadeia de Markov. Como desejamos que as entradas sejam não negativas e a soma de todos os elementos de uma mesma linha seja 1, tomamos a soma dos elementos de uma mesma linha da Tabela 1 e dividimos cada elemento pela soma de sua linha. Ou seja, na linha 1 teremos: $2529+35+257+5 = 2826$, e fazendo $\frac{2529}{2826} = 0.895$ obteremos a probabilidade de permanecer no estado 1, que representa a probabilidade de registrar unicamente a presença do esquilo cinza e após o período registrar novamente a presença única do esquilo cinza. Fazendo $\frac{35}{2826} = 0.013$, teremos a probabilidade de transição de $C \to V$, $\frac{257}{2826} = 0.09$ a probabilidade de transição de $C \to A$ e, finalmente, $\frac{5}{2826} = 0.002$ a probabilidade de transição de $C \to N$. Analogamente para as outras linhas, podemos formar uma matriz estocástica P dada por:

$$P = \begin{pmatrix} 0.895 & 0.013 & 0.09 & 0.002 \\ 0.067 & 0.81 & 0.022 & 0.101 \\ 0.057 & 0.001 & 0.87 & 0.072 \\ 0.0004 & 0.0193 & 0.0487 & 0.9316 \end{pmatrix}$$

Através da matriz de transição P, podemos obter algumas conclusões. Por exemplo, há uma probabilidade de 89.5% para aquelas regiões que apresentavam a presença dos esquilos vermelhos continuem com a apenas a presença desses esquilos. Enquanto isso, temos uma probabilidade de 9% para que a região seja ocupada por ambas as espécies de esquilos. Como o questionário da Comissão são reproduzidos num intervalo de 2 anos, cada período da Cadeia de Markov mostra as probabilidades de crescimento das populações de esquilos num período de 2 anos. Como já vimos anteriormente neste trabalho, para determinar as probabilidades para as populações num período de 4 anos, basta tomarmos a matriz P^2, dada por:

$$P^2 = \begin{pmatrix} 0.807 & 0.0223 & 0.1592 & 0.0114 \\ 0.1155 & 0.6589 & 0.0479 & 0.1776 \\ 0.1007 & 0.0038 & 0.7656 & 0.1299 \\ 0.0048 & 0.0337 & 0.0882 & 0.8734 \end{pmatrix}$$

Nesta aplicação, um questionamento natural é como ficam as probabilidades de transição das populações após um longo período de tempo. Como vimos anteriormente, basta tomarmos P^k quando $k \to \infty$, e teremos uma matriz com linhas iguais a medida estacionária da Cadeia de Markov. Através do MatLab R2017a, iteramos a matriz para determinar P^{200} e obtivemos:

$$P^{200} = \begin{pmatrix} 0.2091 & 0.0589 & 0.3131 & 0.4228 \\ 0.2091 & 0.0589 & 0.3131 & 0.4228 \\ 0.2091 & 0.0589 & 0.3131 & 0.4228 \\ 0.2091 & 0.0589 & 0.3131 & 0.4228 \end{pmatrix}$$

Desta forma, a matriz P^{200} nos indica um candidato para a medida estacionária deste processo, que é o vetor $\pi = \begin{pmatrix} 0.2091 & 0.0589 & 0.3131 & 0.4228 \end{pmatrix}$. E de fato, trata-se da medida estacionária da Cadeia, pois $\pi P = \pi$. Com isso, podemos afirmar, que existe uma probabilidade de $20,91\%$ que as regiões sejam ocupadas apenas por esquilos cinzas, $5,89\%$ que as regiões sejam ocupadas unicamente por esquilos vermelhos, $31,31\%$ que ambas as espécies convivam na região e $42,28\%$ que nenhuma espécie permanecerá na região.

Nesse caso, estudos de crescimento populacional podem interpretar as chances das duas espécies conviverem juntas sem colocar em risco a extinção de alguma delas, pela análise das probabilidades encontradas anteriormente. Para tanto, seria necessário um estudo mais aprofundado do ponto de vista biológico, porém o modelo markoviano mostrou-se bastante útil para determinar os valores encontrados e fica clara a importância da medida estacionária para algumas aplicações.

4 Considerações Finais

Neste trabalho, apresentamos algumas noções básicas das Cadeias de Markov, bem como a aplicação de alguns de seus aspectos em dois casos particulares. Além disso, estudamos o conceito de convergência da Cadeia para a medida estacionária via potências de matriz de transição respectiva.

Primeiramente, abordamos aspectos introdutórios de Probabilidade e de Álgebra Linear afim de proporcionar ao leitor uma compreensão de alguns resultados fundamentais para o que foi exposto nos capítulos seguintes. Destacamos, nessa discussão, a definição de probabilidade e probabilidade condicional, além de resultados envolvendo autovalores, autovetores e independência linear que deram suporte para o desenvolvimento teórico posterior.

No segundo capítulo apresentamos os principais conceitos de Cadeias de Markov. Definimos matriz estocástica, vetor de probabilidade, distribuição inicial e vimos como podemos caracterizar uma Cadeia de Markov pelo seu espaço de estado, matriz de transição e distribuição inicial. Além disso, apresentamos alguns exemplos como o Aluguel de Bicicletas e o Passeio Aleatório no n-Ciclo. Um outro aspecto importante que abordamos foi a classificação das cadeias em aperiódicas ou periódicas, e redutíveis ou irredutíveis. Vimos também que, através da noção de tempo, podemos estimar as distribuições de probabilidade calculando as potências da matriz de transição. Neste sentido, investigamos as potências de P^k em que $k \to \infty$ e demonstramos que o limite converge para uma matriz com linhas iguais a medida estacionária da Cadeia de Markov. Vimos ainda que toda Cadeia de Markov admite uma medida estacionária e, para os casos da Cadeia ser aperiódica e irredutível, podemos garantir a unicidade de tal medida. Sem essas hipóteses, uma Cadeia de Markov pode apresentar infinitas medidas estacionárias.

Finalmente, no terceiro capítulo, mencionamos duas aplicações dos conceitos estudados durante os capítulos anteriores: o jogo de tabuleiro Snakes and Ladders, bem como o estudo do crescimento populacional de espécies de esquilos não-nativos de uma determinada região. Através de uma interpretação matricial do tabuleiro do jogo, obtivemos uma matriz de transição que modela os passos do peão pelo tabuleiro. Percebemos que, pelo tabuleiro utilizado, após muitas jogadas é provável que o peão encontre o fim do tabuleiro. Afim de comprovar essa hipótese, determinamos a convergência através da iteração da matriz de transição. Através dos dados obtidos, foi possível determinar alguns mapas de densidade de probabilidade para o estado do peão após n lançamentos. Na sequência, foi possível também modelar o problema para o casos de mais de 1 jogador, quando verificamos que as chances do jogo já ter terminado após n rodadas aumenta conforme o número de jogadores cresce.

Outra aplicação das Cadeias de Markov discutida no terceiro capítulo ocorre na área da Biologia, através de estudos de caso em crescimentos populacionais de algumas espécies de animais. Nesse sentido, estudamos o caso dos esquilos cinza e esquilos vermelhos na região da Grã-Bretanha. Extraindo os dados obtidos em questionários que expõem as aparições dos esquilos, determinamos as probabilidades de crescimento e decrescimento das espécies num intervalo de tempo. Também através do cálculo das potências da matriz de transição encontrada, vimos que foi possível discutir as chances de sobrevivência das espécies estudadas após um longo período de tempo.

Referências

FELLER, W. *An introduction to probability theory and its applications*. Wiley, 1971.

FIDALGA, N. H. D. *Aplicação das Cadeias de Markov à modelação matemática de epidemias*. Institudo Politécnico de Bragança, 2016. (Acessado em 26/10/2017). Disponível em: ⟨https://bibliotecadigital.ipb.pt/handle/10198/14097⟩.

FRANCO, F. O. *Jogos markovianos alternados sob incerteza*. São Paulo: Universide de São Paulo, 2012. (Acessado em 26/10/2017). Disponível em: ⟨http://www.teses.usp.br/teses/disponiveis/45/45134/tde-19022013-093705/en.php⟩.

FREITAS, A. R. de et al. *Modelagem do crescimento populacional do rebanho bovino brasileiro*. UNICAMP, 2005. (Acessado em 26/10/2017). Disponível em: ⟨http://repositorio.unicamp.br/jspui/handle/REPOSIP/93646⟩.

JUNIOR, A. A. d. S. *Aplicação das Cadeias de Markov no estudo do controle biológico da planta aquática Eichhornia azurea*. Goiás: Universidade Federal de Goiás, 2014. (Acessado em 26/10/2017). Disponível em: ⟨https://repositorio.bc.ufg.br/tede/handle/tede/4731⟩.

LEVIN, D.; PERES, Y.; WILMER, E. *Markov Chains and Mixing Times*. American Mathematical Soc., 2008.

LIMA, E. *Álgebra linear*. Rio de Janeiro: Instituto de Matemática Pura e Aplicada, 2009.

LIMA, E. *Um curso de Análise - Vol.1*. Rio de Janeiro: Instituto de Matemática Pura e Aplicada, 2009.

MEYER, P. L. *Probabilidade: aplicações à estatística*. Livro Técnico, 1970.

MORGADO, L. B. *Tempos de Mistura em Cadeias de Markov*. Trabalho de Conclusão de Curso. UFSC, 2011.

POOLE, D. *Álgebra Linear*. Thomson Learning, 2006.

STRANG, G. *Linear Algebra and Its Applications*. Thomson, Brooks/Cole, 2006.

SWIFT, R. J. *Some Applications of Markov Chains*. California State Polytechnic University, n.d. (Acessado em 26/10/2017). Disponível em: ⟨http://www.cpp.edu/~rjswift/chain%20applications.pdf⟩.