Universidade Federal do Recôncavo da Bahia - UFRB
Centro de Ciências Exatas e Tecnológicas - CETEC
Programa de Pós Graduação em Matemática em Rede Nacional - PROFMAT
Dissertação de Mestrado

Cadeias de Markov no Ensino Médio

Otávio Augusto Rodrigues Melo

Cruz das Almas - Bahia
Julho de 2019

Cadeias de Markov no Ensino Médio

Otávio Augusto Rodrigues Melo

Dissertação de Mestrado apresentada ao curso de Mestrado Profissional em Matemática em Rede Nacional do Centro de Ciências Exatas e Tecnológicas da Universidade Federal do Recôncavo da Bahia e Sociedade Brasileira de Matemática como requisito parcial para obtenção do título de Mestre em Matemática.

Orientador: Prof. Dr. Anderson Reis Cruz.

Cruz das Almas - Bahia

Julho de 2019

FICHA CATALOGRÁFICA

M528d	Melo, Otávio Augusto Rodrigues. Cadeias de Markov no Ensino Médio / Otávio Augusto Rodrigues Melo._ Cruz das Almas, BA, 2019. 73f.; il. Orientador: Anderson Reis da Cruz. Coorientadora: Katia Silene Ferreira Lima Rocha. Dissertação (Mestrado) – Universidade Federal do Recôncavo da Bahia, Centro de Ciências Exatas e Tecnológicas. 1.Matemática – Processos de Markov. 2.Matemática – Estudo e ensino. 3.Ensino médio – Análise. I.Universidade Federal do Recôncavo da Bahia, Centro de Ciências Exatas e Tecnológicas. II.Título. CDD: 519.217

Ficha elaborada pela Biblioteca Universitária de Cruz das Almas – UFRB.
Responsável pela Elaboração – Antonio Marcos Sarmento das Chagas (Bibliotecário – CRB5 / 1615).
Os dados para catalogação foram enviados pelo usuário via formulário eletrônico.

Cadeias de Markov no Ensino Médio

Otávio Augusto Rodrigues Melo

Dissertação de Mestrado apresentada ao curso de Mestrado Profissional em Matemática em Rede Nacional do Centro de Ciências Exatas e Tecnológicas da Universidade Federal do Recôncavo da Bahia e Sociedade Brasileira de Matemática como requisito parcial para obtenção do título de Mestre em Matemática, aprovada em 25/07/2019.

Banca examinadora:

Anderson Reis da Cruz

Prof. Dr. Anderson Reis da Cruz(Orientador)
UFRB

Andressa Lima de Souza da Cruz

Profa. Dra. Andressa Lima de Souza da Cruz
UFRB

João Tiago Assunção Gomes

Prof. Dr. João Tiago Assunção Gomes
UFSB

À minha Tia Deuzinha.

Agradecimentos

Agradeço às pessoas que me ajudaram neste trabalho, em especial ao meu orientador o Professor Dr. Anderson Reis da Cruz, pela experiência, entusiasmo, dedicação e conselhos durantes as orientações até o término da dissertação. Aos demais professores pela oportunidade e por toda a contribuição neste processo de aprendizagem. Aos meus colegas, em especial ao professor Paulo e ao professor José, pelos conselhos e discussões que enriqueceram o conhecimento. À minha esposa por sua paciência e apoio durante o período de estudo. À minha familia pelo seu apoio e entusiasmo nos principais momentos de minha vida. E à Deus que me deu forças, saúde e fé para concluir mais uma etapa na minha vida.

“Não há ramo da Matemática, por mais abstrato que seja, que não possa um dia vir a ser aplicado aos fenômenos do mundo real.”

Lobachevski.

Resumo

Neste trabalho, estudaremos as Cadeias de Markov Homogêneas com o conjunto de estados discreto. Desta forma, pretende-se discutir os principais conceitos dessa teoria no ensino médio da educação básica. Visto que, é possível aprofundar os conteúdos de Probabilidade e Matrizes e suas aplicações como em uma Cadeia de Markov. Além disso, possibilita a contextualização destes tópicos e interdisciplinaridade com outras disciplinas desta etapa de ensino. No final, apresentaremos uma proposta de sequência didática para o professor fazer uso em sala de aula, com o auxílio do software matemático MAXIMA, contribuindo para uma melhor formação dos estudantes.

Palavras-chave: Probabilidade. Matrizes. Cadeias de Markov. Aplicações.

Abstract

In this work, we will study Homogeneous Markov Chains with the set of discrete states. In this way, it is intended to discuss the main concepts of this theory in the high school of basic education. Seeing that it is possible to deepen the contents of Probability and Matrices and their applications as in a Markov Chain. In addition, it allows the contextualization of these topics and interdisciplinarity with other disciplines of this stage of teaching. At the end, there is a proposal of a didactic sequence for the teacher to make use of in the classroom, with the aid of the MAXIMA mathematical software, thus contributing to a better training of students.

Keywords:Probability. Matrices. Markov Chains. Applications.

Sumário

Referências Bibliográficas **59**

Lista de Figuras

Lista de Tabelas

Introdução

Andrei Andreyevich Markov (1856−1922), foi um matemático russo que se formou e foi professor da Universidade St. Petersburg. Nesta época, realizou vários trabalhos na área da matemática, sendo os primeiros sobre limite de integrais, frações contínuas e teoria da aproximação. Depois de 1900, ele aplicou o método de frações contínuas, trabalho inicialmente realizado pelo seu professor Pafnuty Chebyshev, à teoria da probabilidade. No entanto, Markov é reconhecido por seu trabalho sobre as Cadeias de Markov possibilitou uma nova abordagem na teoria das probabilidades e também na teoria dos processos estocásticos.

As Cadeias de Markov são utilizadas para modelar matematicamente e resolver muitos problemas da atual sociedade, entre eles: processos epidemológicos, teoria dos jogos, crescimento e decrescimento de uma determinada população, identificação de genes do DNA entre outros. Além disso, ao nível da educação básica pode-se utilizar este conteúdo para discutir e analisar situações em tempo discreto.

Neste trabalho apresentaremos um primeiro capítulo com conceitos preliminares de Matrizes, Probabilidade e Variáveis Aleatórias. No segundo capítulo, tópicos básicos da teoria matemática das Cadeias de Markov, enfatizando o caso homogêneo a tempo discreto. Ou seja, tudo que é necessário para o ensino aprendizagem desta teoria na educação básica. No terceiro capítulo algumas aplicações das Cadeias de Markov como: previsão do tempo, passeio aleatório, crescimento populacionais e previsões em genética.

No último capítulo do trabalho, trazemos uma sugestão de sequência didática de Cadeias de Markov Homogênneas para o professor aplicar em sala de aula e contribuir para uma discussão matemática mais aprofundada dos conteúdos previamente mencionados.A inquietação que é comum aos alunos no por quê do estudo de algumas definições matemáticas pode ser respondida, construída e participada na discussão das situações propostas nos planos de aula. Desta forma, os objetivos propostos são: mostrar a utilização das matrizes em outras áreas do conhecimento, possibilitar que os alunos enxerguem a matriz como uma forma eficiente de representação e manipulação de dados, discutir a teoria das Cadeias de Markov, trabalhar com probabilidades dentro de matrizes e utilizar o software MAXIMA como ferramenta nos cálculos de grande iterações.

Portanto, o trabalho tem uma parte bibliográfica de aprofundamento da teoria matemática e uma parte de aplicação a educação básica com o enfoque aos alunos do segundo ano do ensino médio. Com isso, sugerimos uma sequência didática que possibilite uma maior aproximação dos estudantes a uma aplicação paupável e assim contribua para um melhor ensino e aprendizagem da matemática.

Capítulo 1

Conceitos Preliminares

Neste capítulo será abordado conceitos básicos da Teoria de Matrizes e da Teoria de Probabilidades que serão utilizados no decorrer do texto. Apresentaremos também noções de experimentos aleatórios, especificamente, discorrendo sobre variáveis aleatórias e suas propriedades. Ao leitor que desejar um estudo mais aprofundado nestes tópicos, recomendamos [7] para a Teoria de Matrizes e [12, 11, 14] para a Teoria de Probabilidades e o estudo de Variáveis Aleatórias. Nestas referências encontram-se os conceitos e propriedades discutidas neste capítulo.

1.1 Matrizes

A ideia geral de uma matriz é a disposição de elementos em linhas e colunas muito úteis para organizar dados. Por exemplo, as notas finais dos alunos de uma série no colégio podem formar uma matriz cujas linhas correspondem às matérias lecionadas naquela série e cujas colunas representam os alunos. Conforme tabela abaixo:

	Pedro	Carlos	Antônio	Juliana	Paula
Matemática	5,4	6,6	7	7,5	8
Português	3	4,3	4,5	6	9
Biologia	3	5	5	7	8
Física	2	5	5	8	8
Química	6	6	5	9	9
História	6	7,5	9	5,5	9
Geografia	7	8	9	7	9

Podemos formar uma matriz com essas informações, na qual a interseção de uma linha com uma coluna demonstra a nota que um determinado aluno tirou naquela matéria.

$$\begin{bmatrix} 5,4 & 6,6 & 7 & 7,5 & 8 \\ 3 & 4,3 & 4,5 & 6 & 9 \\ 3 & 5 & 5 & 7 & 8 \\ 2 & 5 & 5 & 8 & 8 \\ 6 & 6 & 5 & 9 & 9 \\ 6 & 7,5 & 9 & 5,5 & 9 \\ 7 & 8 & 9 & 7 & 9 \end{bmatrix}$$

As matrizes podem aparecer também como quadro de coeficientes de sistemas de equações lineares. Por exemplo,

$$\begin{cases} 4x + 5y & = 7 \\ 2x - 15y & = 9 \end{cases}$$

podemos associar a matriz dos coeficientes

$$\begin{bmatrix} 4 & 5 \\ 2 & -15 \end{bmatrix}.$$

As linhas ou colunas de uma matriz podem representar vetores em $\mathbb{R}^n$, o que as vezes é útil para solução de problemas de geometria. Observamos que, várias técnicas de resolução de problemas tomam como base o estudo de matrizes.

Na definição que adotaremos, uma matriz $m \times n$ é uma lista de números a_{ij}, onde $1 \leq i \leq m$ e $1 \leq j \leq n$, dispostas em m linhas e n colunas, no qual o elemento a_{ij} situa-se na i-ésima linha e na j-ésima coluna:

$$M = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}.$$

A lista ordenada $(a_{i1}, a_{i2}, a_{i3}, ..., a_{in})$ chama-se a i-ésima linha ou o i-ésimo vetor-linha da matriz M enquanto $(a_{1j}, a_{2j}, a_{3j}, ..., a_{mj})$ é a j-ésima coluna ou j-ésimo vetor-coluna de M. Denotamos ainda $M = [a_{ij}]_{m\times n}$ ou, quando não houver necessidade em explicitar o seu tamanho, $M = [a_{ij}]$.

A depender da quantidade de linhas ou colunas, ou ainda, pela natureza de seus elementos, destacaremos alguns tipos especiais de matrizes:

- Uma *matriz quadrada* é aquela cujo número de linhas é igual ao número de colunas $(m = n)$. Neste caso, dizemos que $M = [a_{ij}]_{m\times n}$ é uma matriz de ordem m.

- Uma *matriz nula* é aquela em que $a_{ij} = 0$, para todo i e j. Note que, não necessariamente a matriz é quadrada.

- Uma *matriz coluna* é aquela que possui uma única coluna $(n = 1)$. Analogamente, uma *matriz linha* é aquela que possui uma única linha $(m = 1)$.

- Uma *matriz diagonal* é uma matriz quadrada $(m = n)$ onde $a_{ij} = 0$, para $i \neq j$, isto é, os elementos que não estão na diagonal principal são nulos

- Um caso particular de *matriz diagonal* é a *matriz identidade,* que é uma *matriz quadrada* em que $a_{ii} = 1$ e $a_{ij} = 0$, para $i \neq j$. Denotamos tal matriz por I_n, onde n representa a sua ordem.

- Uma *matriz triangular superior* é uma *matriz quadrada* onde todos os elementos abaixo da diagonal principal são nulos, isto é, $m = n$ e $a_{ij} = 0$, para $i > j$.

- Analogamente, uma *matriz triangular inferior* é aquela em que $m = n$ e $a_{ij} = 0$, para $i < j$.

- Uma *matriz simétrica* é *uma matriz quadrada* tal que $a_{ij} = a_{ji}$, para todo $1 \leq i, j \leq n$.

- Uma *matriz antissimétrica* é uma *matriz quadrada* tal que $a_{ij} = -a_{ji}$, para todo $1 \leq i, j \leq n$. Em particular temos que $a_{ii} = -a_{ii}$, para todo $1 \leq i \leq n$, ou seja, $a_{ii} = 0$, para todo $1 \leq i \leq n$.

Definimos algumas operações entre matrizes.

Definição 1.1. A soma de duas matrizes do mesmo tipo $m \times n$ e o produto de uma matriz por um número são definidos elemento a elemento,

1. Se $A = [a_{ij}]$ e $B = [b_{ij}]$ são matrizes $m \times n$ então $A + B = [a_{ij} + b_{ij}]$.

2. $\alpha \cdot A = [\alpha a_{ij}]$ para todo α pertencente a $\mathbb{R}$.

Estas operações têm as mesmas propriedades das operações usuais de mesmo nome entre vetores de $\mathbb{R}^n$. Neste caso, o elemento neutro da soma é a matriz nula e o elemento oposto de uma matriz $A = [a_{ij}]$ é a matriz $-A = [-a_{ij}]$. De fato, o conjunto de matrizes do mesmo tipo resulta em um espaço vetorial com essas operações veja por exemplo [Boldrini].

Definimos uma operação de multiplicação entre matrizes. Intuitivamente, definiríamos tal operação também elemento a elemento. Entretanto, algumas situações práticas requerem um cálculo diferente. Vejamos o exemplo abaixo.

Exemplo 1.1. Uma empresa que possui duas padarias, chamadas de C e D, que fabrica três tipos de pão: $1, 2$ e 3, os quais são feitos de farinha, açúcar, leite, manteiga e ovos. Em cada semana, as vendas dessas duas padarias serão estimadas conforme a matriz de M de venda semanal abaixo:

Padaria	Pão tipo 1	Pão tipo 2	Pão tipo 3
C	50 unidades	30 unidades	25 unidades
D	20 unidades	20 unidades	40 unidades

Para a fabricação destes pães, o material é usado de acordo com a matriz N seguinte:

Pão	farinha	açúcar	leite	manteiga	ovos
tipo 1	600 g	250 g	400 ml	120 g	6
tipo 2	300 g	100 g	200 ml	280 g	4
tipo 3	450 g	175 g	500 ml	0 g	4

A direção da empresa, afim de atender à demanda, quer saber a quantidade de cada uma das cinco matérias primas que deve alocar as suas duas padarias. A resposta deve ser uma matriz, do tipo 2×5, onde as linhas representam as duas padarias e as colunas correspondem aos cinco materiais usados.

Padaria	farinha	açúcar	leite	manteiga	ovos
C	c_{11}	c_{12}	c_{13}	c_{14}	c_{15}
D	c_{21}	c_{22}	c_{23}	c_{24}	c_{25}

Assim, c_{ij} é quanto a i-ésima padaria deve guardar do j-ésimo material a fim de executar as vendas previstas.

Se escrevermos $M = [a_{ij}]$; para $1 \leq i \leq 2{,}1 \leq j \leq 3$ e $N = [b_{ij}]$, com $1 \leq i \leq 3$, $1 \leq j \leq 5$. Desta forma, $c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j}\ (1 \leq i \leq 2, 1 \leq j \leq 5)$.

Logo, o número de ovos necessários para a padaria C é

$$c_{15} = a_{11}b_{15} + a_{12}b_{25} + a_{13}b_{35} = 50 \times 6 + 30 \times 4 + 25 \times 4 = 520$$

.

Estendemos esta noção para caso geral a seguir:

Definição 1.2. Sejam $M = [a_{ij}]$ e $N = [b_{ij}]$ matrizes do tipo $m \times n$ e $n \times p$ respectivamente. O produto dessas matrizes é a matriz $MN = [x_{ij}]$, de tipo $m \times p$, cuja entrada ij é dado por:

$$x_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj}.$$

As operações de soma, produto por escalar e multiplicação de matrizes obedecem algumas propriedades que descreveremos nos resultados a seguir.

Proposição 1.1. *Dadas as matrizes A, B e C de mesma ordem $m \times n$, temos:*

1. $A + B = B + A$
2. $A + (B + C) = (A + B) + C$
3. $A + 0 = A$, onde 0 é a matriz nula $m \times n$.

Proposição 1.2. *Dadas as matrizes A e B de mesma ordem $m \times n$ e números reais k, k_1 e k_2, temos:*

1. $k\,(A + B) = kA + kB$
2. $(k_1 + k_2)\,A = k_1A + k_2A$
3. $0A = 0$
4. $k_1\,(k_2A) = (k_1k_2)\,A$

Observação 1.1. Como as operações de soma e produto por um número real são definidas elemento a elemento, as proposições acima são consequências imediatas das propriedades associativa, comutativa e distributiva da soma e produto de números reais.

Proposição 1.3. *Dadas A e D matrizes de ordem $m \times n$ e B, C matrizes de ordem $n \times p$, temos:*

1. $A\,(B + C) = AB + AC$
2. $(A + D)B = AB + DB$

Demonstração. Sejam $A = [a_{ij}]_{m\times n}\,, B = [b_{ij}]_{n\times p}$ e $C = [c_{ij}]_{n\times p}$. Temos que:

$$\begin{aligned} A(B+C) &= \left[\sum_{k=1}^{n} a_{ik}(b_{kj} + c_{kj})\right] = \left[\sum_{k=1}^{n} a_{ik}b_{kj} + \sum_{k=1}^{n} a_{ik}c_{kj}\right] \\ &= \left[\sum_{k=1}^{n} a_{ik}b_{kj}\right] + \left[\sum_{k=1}^{n} a_{ik}c_{kj}\right] = AB + BC \end{aligned}$$

Analogamente, provamos que $(A + D)B = AB + DB$.

□

Observação 1.2. Note que o produto não é comutativo em geral. Considere por exemplo $A = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}$ e $B = \begin{bmatrix} -1 & -1 \\ 0 & 0 \end{bmatrix}$, temos $AB \neq BA$.

1.2 Noções de Probabilidade

A Teoria das Probabilidades é um ramo da Matemática que foi iniciado através de estudos de jogos de azar. Essa teoria busca reduzir todos os acontecimentos do mesmo gênero a um certo número de casos igualmente possíveis, ou seja, tais que estejamos igualmente desacreditados de sua existência, e em determinar o número de casos favoráveis ao acontecimento cuja probabilidade é buscada. Assim, a razão deste número com o de todos os casos possíveis é a medida da probabilidade. Ou seja, uma fração onde o numerador é o número de casos favoráveis e o denominador é o número de todos os casos possíveis.

A definição de probabilidade como quociente do número de casos favoráveis sobre o número de casos possíveis foi a primeira definição formal de probabilidade, e apareceu pela primeira vez na obra Liber de Ludo Aleae de Jerônimo Cardano $(1501 - 1576)$. Para compreensão dos conceitos e propriedades iremos analisar alguns exemplos.

Exemplo 1.2. Considere o lançamento de um dado. Os resultados possíveis são $1, 2, 3, 4, 5$ e 6. Queremos saber a probabilidade de, em um lançamento obter um número ímpar. Na ideia do Ludo Aleae, tal probabilidade corresponde ao quociente entre o número de casos favoráveis (obter $1, 3$ ou 5 no lançamento) sobre o número de casos possíveis, ou seja, a probabilidade de $\frac{1}{2}$, que corresponde a 50% de chance.

Intuitivamente, entendemos um experimento como um procedimento realizado sob determinadas condições, repetido um número de vezes sob estas mesmas condições, a partir do qual observamos os resultados obtidos. Um experimento é determinístico quando sob as mesmas condições o seu resultado é completamente determinado. Por exemplo, um recipiente com água pura nas condições normais de pressão atmosférica posto a 100°C,tem como resultado a fervura da água.

Um experimento é dito ser aleatório quando o seu resultado não pode ser determinado, o que se sabe são apenas um conjunto possível de resultados. Por exemplo o lançamento de um dado pode ser considerado um experimento aleatório.

Definição 1.3. O conjunto Ω de resultados possiveis de um experimento aleatório será chamado de espaço amostral. Os elementos de Ω serão chamados de amostras.

Voltemos ao lançamento de dados. Neste caso o espaço amostral é o conjunto $\Omega = \{1, 2, 3, 4, 5, 6\}$. Podemos relacionar a obtenção de um número ímpar ao conjunto $A = \{1, 3, 5\}$,que chamaremos de evento *Laplace.*

No evento *Laplace,* os elementos de A são denominados como os casos favoráveis. Os elementos do espaço amostral Ω eram chamados os casos possíveis. Desta forma, definimos probabilidade como número de casos favoráveis sobre o número de casos possíveis.

A classe de eventos de um experimento aleatório deve ser rico suficiente de modo a ser significante.

Definição 1.4. Seja Ω um conjunto. Uma $\sigma - \acute{a}lgebra$ $\mathcal{F}$ de Ω é uma família de subconjuntos de Ω satisfazendo:

i) $\emptyset$ e Ω pertencem a $\mathcal{F}$.

ii) Se $A \in \mathcal{F}$ então $A^c \in \mathcal{F}$.

*iii)*Se $A_j \in \mathcal{F}, j = 1, 2, \ldots$ então $\underset{j}{\cup} A_j \in \mathcal{F}$.

Atribuímos o número $\frac{1}{2}$ a chance de ocorrer o evento $A = \{1, 3, 5\}$ no lançamento de dados e podemos interpretar essa chance como uma medida do conjunto A. Em geral definimos uma probabilidade no conjunto Ω como uma função que atribui a elementos de uma $\sigma - \acute{a}lgebra$ valores no intervalo $[0, 1]$.

Definição 1.5. Sejam Ω um conjunto (espaço amostral) e $\mathcal{F}$ uma $\sigma - \acute{a}lgebra$ em Ω. Uma função $P : \mathcal{F} \to [0, 1]$ é chamada de probabilidade se satisfaz as seguintes condições:

1. $P(\emptyset) = 0$ e $P(\Omega) = 1$;
2. Se $\{A_j\}_{j \in \mathbb{N}}$ é uma família de subconjuntos 2 a 2 disjuntos de $\mathcal{F}$ então $P\left(\dot{\bigcup}_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j)$.

No caso em que $\sharp\Omega < \infty$ a $\sigma - \acute{a}lgebra$ será sempre o conjunto de partes de Ω, ou seja, a família de todos os subconjuntos de Ω. Neste caso, a condição 2 reduz-se a uma união finita.

Além disso, é razoável supor que:

a) Os eventos elementares são igualmente prováveis.

b) Todo evento A é união de p eventos elementares onde $p \leq n$.

A razoabilidade da condição (a) é consequência do seguinte fato: considere um evento A resultado de um experimento aleatório. Podemos analisar a frequência deste evento em n repetições do experimento através do quociente $\frac{F_n(A)}{n}$ onde $F_n(A)$ denota o número de vezes em que o evento A aconteceu em n repetições. Observe que este é um dado empírico, obtido através da repetição do experimento um determinado número de vezes. Surge assim uma questão natural: o que acontece com $\frac{F_n(A)}{n}$ se tomarmos n suficientemente grande?

Caso tal número se aproxime de algum valor, este seria uma boa representação para a probabilidade da ocorrência do evento A. Um resultado central na Teoria de Probabilidade garante que na maioria das sequências de repetições este número se aproxima de $\frac{\sharp(A)}{\sharp(\Omega)}$, quando fazemos n suficientemente grande. Este teorema é a chamada Lei dos Grandes Números, que não apresentaremos aqui, por ser um resultado mais complexo e que foge dos nossos objetivos. Sugerimos [3] para mais informações acerca deste resultado.

Logo, podemos definir, para cada $A \subset \Omega$, a probabilidade $P(A) = \frac{\sharp(A)}{\sharp(\Omega)} = \frac{p}{n}$.
Observe que:

a) Para todo evento A, $0 \leq P(A) \leq 1$;

b) $P(\Omega) = 1$;

c) $P(\emptyset) = 0$, pois $(\sharp(\emptyset) = 0)$;

d) Se $A \cap B = \emptyset$ então $P(A \dot{\cup} B) = P(A) + P(B)$.

Portanto, P assim definida é uma probabilidade em Ω.

Exemplo 1.3. Três moedas são jogadas simultaneamente. Vamos calcular a probabilidade de obter 2 caras e a probabilidade de obter pelo menos 2 caras.

Sendo K cara e C coroa, o espaço amostral é dado por

$$\Omega = \{(KKK), (KKC), (KCK), (KCC), (CKK), (CKC), (CCK), (CCC)\}.$$

Desta forma, $\sharp(\Omega) = 8$. Se A indica o evento "obter 2 caras", então $A = \{(KKC), (KCK), (CKK)\}$. Assim, $\sharp(A) = 3$ e portanto $P(A) = \frac{3}{8}$.

Se B denota o evento "obter pelo menos duas caras", então

$$B = \{(KKC), (KCK), (CKK), (KKK)\}.$$

Assim, $\sharp(B) = 4$ e $P(B) = \frac{4}{8} = \frac{1}{2}$.

Exemplo 1.4. Dois dados são jogados simultaneamente. Vamos calcular a probabilidade de que a soma dos números mostrados nas faces de cima seja 7.

O espaço amostral Ω consiste em todos os pares (i, j) onde i e j são inteiros positivos compreendidos entre 1 e 6. Desta forma, podemos descrever o espaço amostral:

	1	2	3	4	5	6
1	$(1,1)$	$(1,2)$	$(1,3)$	$(1,4)$	$(1,5)$	$(1,6)$
2	$(2,1)$	$(2,2)$	$(2,3)$	$(2,4)$	$(2,5)$	$(2,6)$
3	$(3,1)$	$(3,2)$	$(3,3)$	$(3,4)$	$(3,5)$	$(3,6)$
4	$(4,1)$	$(4,2)$	$(4,3)$	$(4,4)$	$(4,5)$	$(4,6)$
5	$(5,1)$	$(5,2)$	$(5,3)$	$(5,4)$	$(5,5)$	$(5,6)$
6	$(6,1)$	$(6,2)$	$(6,3)$	$(6,4)$	$(6,5)$	$(6,6)$

O número de eventos elementares é igual a $\sharp(\Omega) = 36$. Seja A o conjunto dos pares (i, j) tais que $i + j = 7$.

Esses pares são justamente a anti-diagonal:

$$A = \{(6,1), (5,2), (4,3), (3,4), (2,5), (1,6)\}.$$

Desta forma, $\sharp(A) = 6$ e portanto $P(A) = \frac{6}{36} = \frac{1}{6}$.

A seguir apresentaremos algumas consequências da definição de probabilidade.

Proposição 1.4. $P\left(A^C\right) = 1 - P\left(A\right),$ *para todo* $A \in \mathcal{F}.$

Demonstração. Sabemos que $1 = P\left(\Omega\right) = P\left(A\dot{\cup}A^C\right) = P\left(A\right) + P\left(A^C\right).$ Portanto, $P\left(A^C\right) = 1 - P\left(A\right).$ □

Proposição 1.5. *Se* $A, B \in \mathcal{F}$ *com* $A \subset B$ *então* $P\left(A\right) = P\left(B\right) - P\left(B - A\right).$

Demonstração. Como $B = A\dot{\cup}\left(B - A\right),$ temos

$$P\left(B\right) = P\left(A\dot{\cup}\left(B - A\right)\right) = P\left(A\right) + P\left(B - A\right).$$

Portanto, $P(A) = P\left(B\right) - P\left(B - A\right).$ □

Corolário 1.1. *Se* $A, B \in \mathcal{F}$ *com* $A \subset B$ *então* $P\left(A\right) \leq P\left(B\right).$

Demonstração. Como $P\left(A\right) = P\left(B\right) - P\left(B - A\right)$ e $P\left(B - A\right) \geq 0,$ então $P\left(A\right) \leq P\left(B\right).$ □

Proposição 1.6. $P\left(A \cup B\right) = P\left(A\right) + P\left(B\right) - P\left(A \cap B\right).$

Demonstração. Como $P\left(A\right) = P\left(A - B\right) + P\left(A \cap B\right)$ e $P\left(B\right) = P\left(B - A\right) + P\left(A \cap B\right).$ Temos,

$$P\left(A\right) + P\left(B\right) = P\left(A - B\right) + P\left(B - A\right) + P\left(A \cap B\right) + P\left(A \cap B\right).$$

Observe que, $A \cup B = \left(A - B\right)\dot{\cup}\left(A \cap B\right)\dot{\cup}\left(B - A\right).$ Logo,

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

Portanto, $P\left(A \cup B\right) = P\left(A\right) + P\left(B\right) - P\left(A \cap B\right).$ □

Com o mesmo raciocínio usado para descrever e provar a proposição acima pode-se estabelecer uma fórmula para $P\left(A_1 \cup A_2 \cup A_n\right)$ onde $A_1, A_2, \ldots, A_n$ são n eventos, conforme enunciado abaixo:

Proposição 1.7. *Suponha* $A_j \in \mathcal{F}, j = 1, \ldots, n,$ *temos que*

$$\begin{aligned} P\left(A_1 \cup A_2 \cup \cdots \cup A_n\right) = P\left(A_1\right) + P\left(A_2\right) + \cdots \\ + P\left(A_n\right) - P\left(A_1 \cap A_2\right) - \cdots - P\left(A_{n-1} \cap A_n\right) + \\ + P\left(A_1 \cap A_2 \cap A_3\right) + \cdots + (-1)^{n-1} P\left(A_1 \cap A_2 \cap \cdots \cap A_n\right). \end{aligned}$$

As propriedades vistas acima nas proposições são válidas para qualquer probabilidade, ou seja, qualquer função de conjuntos satisfazendo as condições da definição 1.5 atende às proposições 1.6 e 1.7. Note que, sobre o mesmo espaço amostral Ω é possível definir muitas probabilidades diferentes.

Um fenômeno aleatório é representado matematicamente por um trio de objetos: o espaço amostral Ω, a família de eventos da $\sigma - \acute{a}lgebra\, \mathcal{F}$ e uma probabilidade P definida sobre os subconjuntos de Ω. O trio $(\Omega, \mathcal{F}, P)$ é chamdo de Espaço de Probabilidades.

Exemplo 1.5. Um número entre 1 e 400 é escolhido aleatoriamente. Vamos calcular a probabilidade de que ele seja divisível por 2 ou por 5.

Sejam A e B os eventos que acontecem se o número escolhido for divisível por 2 e por 5 respectivamente. Temos que calcular $P(A \cup B)$. Os números entre 1 e 400 divisíveis por 2 são 200, os divisíveis por 5 são $400/5 = 80$ e os divisíveis por 2 e por 5 ao mesmo tempo são $400/10 = 40$. Assim, $P(A) = \frac{200}{400} = \frac{1}{2}$, $P(B) = \frac{80}{400} = \frac{1}{5}$, $P(A \cap B) = \frac{40}{400} = \frac{1}{10}$. Logo, $P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{2} + \frac{1}{5} - \frac{1}{10} = \frac{6}{10}$.

1.2.1 Probabilidades Condicionais

Em certas situações é necessário compreender a relação entre dois eventos. Uma medida dessa relação é probabilidade condicional, que corresponde a probabilidade da ocorrencia de um evento A dado que ocorreu um evento B.

Definição 1.6. Dados dois eventos A e B, a probabilidade condicional de B dado A é o número $P(A \cap B)/P(A)$ onde $P(A) > 0$. Repesentaremos este número pelo símbolo $P(B \mid A)$. A equação também pode ser escrita como $P(A \cap B) = P(A) P(B \mid A)$. Se $P(B) > 0$, temos $P(A \cap B) = P(B) P(A \mid B)$.

Elencamos abaixo algumas propriedades das probabilidades condicionais. Ver [7, 12, 14].

Proposição 1.8. *Seja $A \in \mathcal{F}$ tal que $P(A) > 0$. Então*

a) $P(\emptyset \mid A) = 0$, $P(\Omega \mid A) = 1$ e $0 \leq P(B \mid A) \leq 1$, para qualquer que seja $B \in \mathcal{F}$;

b) $P((B \dot{\cup} C) \mid A) = P(B \mid A) + P(C \mid A)$.

Demonstração. a) Segue da definição 1.6 que

$$P(\emptyset \mid A) = \frac{P(\emptyset \cap A)}{P(A)} = \frac{P(\emptyset)}{P(A)} = \frac{0}{P(A)} = 0$$

$$P(\Omega \mid A) = \frac{P(\Omega \cap A)}{P(A)} = \frac{P(A)}{P(A)} = 1.$$

Como $0 \leq P(A \cap B) \leq P(A)$, temos $0 \leq \frac{P(A \cap B)}{P(A)} \leq 1$, isto é, $0 \leq P(B \mid A) \leq 1$ para qualquer $B \in \mathcal{F}$.

b)Da definição 1.6 temos

$$\begin{aligned} P((B \dot{\cup} C) \mid A) &= \frac{P((B \dot{\cup} C) \cap A)}{P(A)} = \frac{P((B \cap A) \dot{\cup} (C \cap A))}{P(A)} \\ &= \frac{P(B \cap A)}{P(A)} + \frac{P(C \cap A)}{P(A)} = P(B \mid A) + P(C \mid A). \end{aligned}$$

□

Como consequência da proposicão acima temos que, fixado $A \in \mathcal{F}$, a probabilidade condicional $P(\cdot \mid A)$ é outra probabilidade sobre o espaço amostral Ω.

Observe que para dois conjuntos A_1 e A_2 temos $P(A_1 \cap A_2) = P(A_1)\, P(A_2 \mid A_1)$. Caso tenhamos A_1, A_2, A_3, verificamos que:

$$P(A_1 \cap A_2 \cap A_3) = P(A_3 \mid (A_1 \cap A_2))\, P(A_1 \cap A_2) = P(A_3 \mid (A_1 \cap A_2))\, P(A_2 \mid A_1)\, P(A_1).$$

Usando o Princípio de Indução Completa, temos a seguinte proposição.

Proposição 1.9. *Sejam* $A_j \in \mathcal{F}, j = 1, \ldots, n$. *Se* $P(A_1 \cap A_2 \cap \cdots \cap A_n) \neq 0$, *então*

$$\begin{aligned} &P(A_1 \cap A_2 \cap \cdots \cap A_n) = \\ &\qquad = P(A_1)\, P(A_2 \mid A_1)\, P(A_3 \mid (A_1 \cap A_2)) \cdots P(A_n \mid (A_1 \cap A_2 \cap \cdots \cap A_{n-1})). \end{aligned}$$

Exemplo 1.6. Sabe-se que 80% dos habitantes de Amargosa foram vacinados contra a gripe. A probabilidade de uma vacina conseguir imunizar uma pessoa é de 40% se o habitante for da região rural e de 70% caso contrário. Ao realizar uma vacinação vamos calcular a probabilidade da vacina imunizar uma pessoa, dado que o habitante é da zona rural.

$$P(''\text{habitante de Amargosa}''\, e\, ''\text{zona rural}'') = P(B \cap C).$$

Tal probabilidade é dada por

$$P(B \cap C) = P(B) \times P(C \mid B) = 0,8 \times 0,4 = 0,32.$$

Note que, nesta situação precisamos saber a probabilidade de imunização da vacina para uma pessoa da zona rural ou da zona urbana. Assim, podemos definir o evento D como a união de dois eventos disjuntos: "O habitante é da zona rural de Amargosa e foi imunizado" e " O habitante é da zona urbana de Amargosa e foi imunizado".

Então,

$$D = (E \cap D) \cup (\bar{E} \cap D).$$

Logo,

$$P(D) = P(E \cap D) + P(\bar{E} \cap D),$$

onde

$$P(E \cap D) = P(E) \times P(D \mid E) = 0,8 \times 0,4 = 0,32$$

e

$$P(\bar{E} \cap D) = P(\bar{E}) \times P(D \mid \bar{E}) = 0,2 \times 0,7 = 0,14.$$

Logo, $P(D) = 0,32 + 0,14 = 0,46.$

Podemos escrever a probabilidade de um evento através de probabilidades condicionais como segue:

Proposição 1.10 (Teorema da Probabilidade Total)**.** *Se B é um evento contido em uma união de eventos disjuntos $A_1, A_2, \ldots, A_n$, ou seja, $\cup_{i=1}^{n} A_i = \Omega$ e $P(A_1) > 0, P(A_2) > 0, \ldots, P(A_n) > 0$ então*

$$P(B) = P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + \ldots + P(A_n)P(B \mid A_n).$$

Demonstração. Como $\cup_{i=1}^{n} A_i = \Omega$, segue que $B = (A_1 \cap B) \cup (A_2 \cap B) \cup \cdots \cup (A_n \cap B)$, onde $(A_i \cap A_j) = \emptyset, \forall i \neq j$. Logo,

$$\begin{aligned} P(B) &= P((A_1 \cap B) \cup (A_2 \cap B) \cup \cdots \cup (A_n \cap B)) \\ &= P(A_1 \cap B) + P(A_2 \cap B) + \ldots + P(A_n \cap B) \\ &= P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + \ldots + P(A_n)P(B \mid A_n). \end{aligned}$$

□

Definição 1.7. Seja A um conjunto não vazio. Uma partição de um conjunto A é qualquer coleção C de subconjuntos não vazios de A dotada da seguinte propriedade: todo elemento de A pertence a um e apenas um dos elementos de C.

Assim, uma coleção de conjuntos $C = \{A_1, A_2, \ldots, A_n\}$ é uma partição do conjunto A, se as seguintes condições forem simultaneamente satisfeitas:

1. $A_i \neq \emptyset$, para $i = 1, 2, \ldots, n$;

2. $A_i \subset A$, para $i = 1, 2, \ldots, n$;

3. $A = A_1 \dot{\cup} A_2 \dot{\cup} \cdots \dot{\cup} A_n$;

4. $A_1, A_2, \ldots, A_n$ são mutuamente disjuntos, isto é, $A_i \cap A_j = \emptyset$, para $i \neq j$, com $i, j = 1, 2, \ldots, n$.

Proposição 1.11. *Sejam $A_1, A_2, \ldots, A_n$ uma partição do espaço amostral Ω e B e C dois eventos quaisquer de Ω. Então*

$$P(B \mid C) = \sum_{i=1}^{n} P(B \mid A_i \cap C)P(A_i \mid C)$$

Demonstração. Como $A_1, A_2, \ldots, A_n$ é uma partição de Ω temos que $A_i \cap A_j = \emptyset$, $\forall i, j \in \mathbb{N}$.

$$
\begin{aligned}
P(B \mid C) = P(B \cap (\dot{\cup}_{i=1}^{n} A_i) \mid C) = \frac{P(B \cap (\dot{\cup}_{i=1}^{n} A_i) \cap C)}{P(C)} &= \frac{P((\dot{\cup}_{i=1}^{n} (B \cap A_i)) \cap C}{P(C)} \\
= \sum_{i=1}^{n} \frac{P(B \cap A_i \cap C)}{P(C)} = \sum_{i=1}^{n} \frac{P(B \cap A_i \cap C)}{P(C)} \cdot \frac{P(A_i \cap C)}{P(A_i \cap C)} \\
= \sum_{i=1}^{n} \frac{P(B \cap A_i \cap C)}{P(A_i \cap C)} \cdot \frac{P(A_i \cap C)}{P(C)} = \sum_{i=1}^{n} P(B \mid A_i \cap C) P(A_i \mid C).
\end{aligned}
$$

□

1.3 Variáveis Aleatórias

Na observação de um experimento aleatório podemos associar algumas quantidades. Por exemplo, numa pesquisa sobre o perfil de determinada população (espaço amostral) podemos indagar sobre a altura, idade ou peso de cada indivíduo (amostra). Podemos caracterizar cada uma destas quantidades por uma função

$$X : \Omega \to \mathbb{R}.$$

Devido a natureza de um experimento aleatório, seus resultados não são determinados, mas as suas probabilidades são. No estudo de quantidades associadas a esses experimentos estamos interessados em analisar probabilidades do tipo $P\left(X^{-1}\left(I\right)\right)$ onde I é um intervalo da reta. Ou seja, analisaremos a probabilidade da função X atingir seus valores no intervalo I.

Um experimento aleatório é totalmente caracterizado por seu espaço amostral Ω, a família de seus eventos ($\sigma - \acute{a}lgebra\, \mathcal{F}$) e uma probabilidade definida sobre os seus intervalos. Nos referimos ao trio $(\Omega, \mathcal{F}, P)$ como um espaço de probabilidade. Uma variável aleatória será então uma função $X : \Omega \to \mathbb{R}$, tal que possamos mensurar $X^{-1}\left(I\right)$ para todo intervalo I, ou seja, tal que $X^{-1}\left(I\right) \in \mathcal{F}$, para todo intervalo I.

Definição 1.8. Seja $(\Omega, \mathcal{F}, P)$ um espaço de probabilidade. Denominamos de variável aleatória, qualquer função $X : \Omega \to \mathbb{R}$ tal que $X^{-1}\left(I\right) = \{w \in \Omega : X\left(w\right) \in I\} \in \mathcal{F}$, para todo intervalo $I \subset \mathbb{R}$. Ou seja, X é tal que sua imagem inversa de intervalos $I \subset \mathbb{R}$ pertence a $\sigma - \acute{a}lgebra\, \mathcal{F}$.

Em geral representamos uma variável aleatória por letra maiúscula do alfabeto. Para cada elemento $w \in \Omega$, associamos um número real $X\left(w\right)$. Garantimos o cálculo de probabilidades com variáveis aleatórias ao exigir que para qualquer $I \subset \mathbb{R}$, o conjunto

$X^{-1}(I)$ seja um evento, ou seja, $X^{-1}(I)$ é um elemento da $\sigma - \acute{a}lgebra\, \mathcal{F}$. Formalmente, dizemos que a variável aleatória é qualquer função real mensurável em $\mathcal{F}$.

O espaço de probabilidade $(\Omega, \mathcal{F}, P)$ pode ser escrito simplificadamente $(\Omega, \mathcal{F})$, quando não houver dúvidas quanto a probabilidade considerada em $\mathcal{F}$.

Para verificarmos que $X : \Omega \to \mathbb{R}$ é uma variável aleatória basta verificar que $X^{-1}(-\infty, r] \in \mathcal{F}$ para todo r. Isto se deve ao fato que qualquer intervalo I pode ser gerado por intervalos da forma $(-\infty, r]$. Veja o exemplo de [9, Capítulo 2].

Exemplo 1.7. Sejam $\Omega = \{1, 2, 3, 4\}$, $\mathcal{F} = \{\emptyset, \Omega, \{1, 2\}, \{3, 4\}\}$ e considere os conjuntos $A = \{1, 2\}$ e $B = \{1, 3\}$. Vamos verificar que I_A, onde I_A é definido por

$$I_A = \begin{cases} 1, & se\, x \in A \\ 0, & se\, x \notin A \end{cases}$$

é variável aleatória em $(\Omega, \mathcal{F})$, mas I_B definida analogamente não é.

Como $A \in \mathcal{F}$ e $B \notin \mathcal{F}$, então,

$$I_A^{-1}(-\infty, x] = \{w \in \Omega : I_A(w) \in (-\infty, x]\} = \begin{cases} \emptyset, \; se\, x < 0 \\ A^c, \; se\, 0 \leq x < 1 \\ \Omega, \; se \geq 1 \end{cases}.$$

Portanto, a imagem inversa de I_A para intervalos do tipo $(-\infty, x], x \in \mathbb{R}$ e $x \in \mathcal{F}$.

Por outro lado, para $0 \leq x \leq 1$, temos $I_B(-\infty, x] = \{w \in \Omega : I_B(w) \in (-\infty, x]\} = B^C$, que não está em $\mathcal{F}$. Assim, I_B não é variável aleatória em $(\Omega, \mathcal{F})$.

Note que, podemos tornar I_B uma variável aletatória modificando a $\sigma - \acute{a}lgebra$. Assim, tomando $\sigma - \acute{a}lgebra$ suficientemente "grande" podemos trabalhar com todas as variáveis que nos interessem. O conjunto das partes é a maior $\sigma - \acute{a}lgebra$ possível. Se $\mathcal{F} = P(\Omega)$, então qualquer função $X : \Omega \to \mathbb{R}$ é uma variável aleatória.

Denotamos, $X^{-1}(I) = \{w \in \Omega : X(w) \in I\} = [X \in I]$. Por exemplo, para $I = (-\infty, 0)$, escreveremos $[X < 0]$ para denotar $X^{-1}(I)$. Podemos simplificar a notação da probabilidade destes eventos por $P(X \in I)$.

No entanto, ao caracterizarmos uma variável aleatória precisamos determinar a distribuição de suas probabilidades.

Definição 1.9 (Função de Distribuição). Seja X uma variável aleatória em $(\Omega, \mathcal{F}, P)$, sua função de distribuição é definida por $F_X(x) = P(X \in (-\infty, x]) = P(X \leq x)$, com x percorrendo todos os números reais.

A função também é chamada função de distribuição acumulada, por acumular as probabilidades dos valores inferiores ou iguais a x. Ao se referir a função de distribuição o subescrito X pode ser omitido e escreveremos F ao invés de F_X.

Proposição 1.12 (Propriedades da Função de Distribuição). *Uma função de distribuição de uma variável aleatória X em $(\Omega, \mathcal{F}, P)$ obedece às seguintes propriedades:*

$(F1)$ $lim_{x\to-\infty} F(x) = 0$ e $lim_{x\to\infty} F(x) = 1$;

$(F2)$ F é não decrescente, ou seja, $F(x) \leq F(y)$ sempre que $x \leq y, \forall x, y \in \mathbb{R}$.

No item 1.3 $\lim_{x\to-\infty} F(x)$ representa o comportamento da função F quando x toma valores cada vez menores, ou seja, se a imagem $F(x)$ se aproxima de algum valor quando tomamos x cada vez menor. Analogamente, $\lim_{x\to+\infty} F(x)$ representa o comportamento da função F quando x toma valores cada vez maiores. A propriedade 1.3 diz então que $P(X \leq x)$ se aproxima de zero conforme o x descresce e $P(X \leq x)$ se aproxima de 1 conforme o x cresce. A propriedade 1.3 é consequência imediata da definição de uma probabilidade (Definição 1.5e do Corolário 1.1), uma vez que $\{\omega \in \Omega | X(\omega) \leq x\} \subset \{\omega \in \Omega \mid X(\omega) \leq y\}$, se $x \leq y$.

Destacamos que, se desejarmos saber a variável, estamos na verdade querendo descobrir qual é a função de distribuição. Essa caracterização é única a menos de eventuais conjuntos que tenham probabilidade zero. Veremos que existem funções, relacionadas com a função de distribuição, que também caracterizam unicamente uma variável.

Exemplo 1.8. Para o lançamento de uma moeda, seja $\Omega = \{cara, coroa\}, \mathcal{F}$ o conjunto das partes de Ω e P dada por $P(cara) = P(coroa) = 1/2$. Definimos uma função X de Ω em $\mathbb{R}$ da seguinte forma:

$$X(w) = \begin{cases} 1, & se\, w = cara \\ 0, & se\, w = coroa \end{cases}.$$

Para qualquer conjunto $I \subset \mathbb{R}, X^{-1}(I) \in \mathcal{F}$ e, portanto, X é variável aleatória.

Agora, se $I = (-\infty, 0)$ temos $X^{-1}(I) = \emptyset$; Entretanto, para $I = (0, 2]$ temos $X^{-1}(I) = \{cara\}$. Nos dois casos a imagem inversa pertence a $\mathcal{F}$. Para obter a função de distribuição analisamos os casos: Para $x < 0$, $P(X \leq x) = 0$, uma vez que o menor valor assumido pela varável é 0. No intervalo $0 \leq x < 1$, temos $P(X \leq x) = P(X = 0) = \frac{1}{2}$. quando $x \geq 1$, ganhamos $P(X \leq x) = P(X = 0) + P(X = 1) = 1$. Assim, $F(x) = P(X \leq x)$ foi definida para qualquer valor real. Resumidamente, temos:

$$F(x) = \begin{cases} 0, & se\, x < 0 \\ 1/2, & se\, 0 \leq x < 1 \\ 1, & se\, x \geq 1 \end{cases}.$$

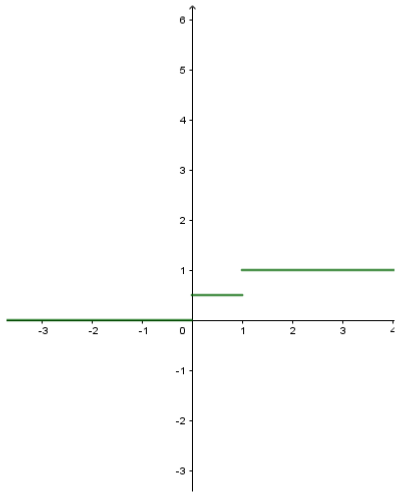

Figura 1.3.1: Gráfico de $F(x)$.

Definição 1.10 (Variável Aleatória Discreta e Função de Probabilidade). Uma variável aleatória é classificada como discreta, se assume somente um número enumerável de valores (finito ou infinito). A função de probabilidade de uma variável discreta é uma função que atribue probabilidade a cada um dos possíveis valores assumidos pela variável, ou seja, sendo X uma variável com valores $x_1, x_2, \ldots$, temos para $i = 1, 2, \ldots$

$$p(x_i) = P(X = x_i) = P\left(\{w \in \Omega : X(w) = x_i\}\right)$$

Proposição 1.13. *Se X é uma variável aleatória então a sua função de probabilidade de X em $(\Omega, \mathcal{F}, P)$ satisfaz:*

(FP1) $0 \leq p\left(x_i\right) \leq 1, \forall i = 1, 2, \ldots;$

(FP2)$\sum\limits_{i} p(x_i) = 1.$

com a soma percorrendo todos os possíveis valores (eventualmente infinitos).

Demonstração. Por serem probabilidades, os valores da função de probabilidade estão sempre entre 0 e 1. Além disso, cada valor da variável induz um subconjunto disjunto, cuja união é Ω, o que implica 1.13. □

Para as variáveis discretas, a função de distribuição tem a forma de escada sendo descontínua nos valores assumidos pela variável. Da função de probabilidade obtemos a função de distribuição e vice-versa. Se já sabemos a função de probabilidade obtemos, $\forall x \in \mathbb{R},\ F(x) = \sum_{i \in A_i} p(x_i)$,com $A_i = \{i : x_i \leq x\}$; onde a somatória se estende aos índices para os quais $x_i \leq x$. Por outro lado, dada a função de distribuição temos $p(x_i) = F(x_i) - F(x_i^-)$, em que $F(x_i^-)$ é o limite de F tendendo a x_i pela esquerda (isto é, por valores anteriores à x_i).

Exemplo 1.9. Considere dois lançamentos independentes de uma moeda equilibrada. Com o espaço de probabilidade sendo o usual, defina X como sendo o número de coroas nos dois lançamentos. A variável X será discreta e sua função de probabilidade será dada por:

X	0	1	2
$p(x_i)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Tabela 1.1: Probabilidades da v.a. X.

Desta forma, a função de distribuição será:

$$F(x) = \begin{cases} 0, & se\, x < 0; \\ 1/4, & se\, 0 \leq x < 1; \\ 3/4, & se\, 1 \leq x < 2; \\ 1, & se\, x \geq 2. \end{cases}$$

O gráfico de $F(x)$ é em escada como apresentado abaixo. Note que, os pontos de descontinuidade ocorrem nos valores assumidos pela variável, sendo o tamanho do salto a probabilidade da variável assumir aquele determinado valor.

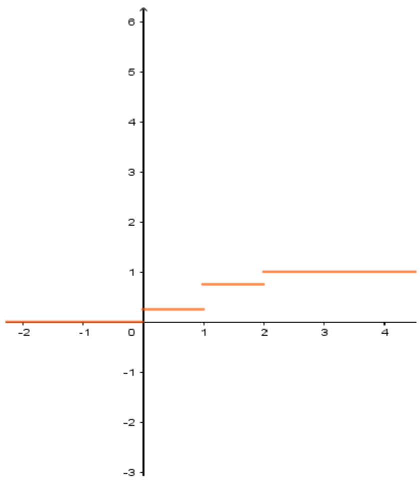

Figura 1.3.2: Função de distribuição para o número de coroas.

1.4 Alguns Modelos Discretos

Como visto a função de distribuição caracteriza completamente uma variável aleatória. Apresentaremos a seguir algumas funções de probabilidade que descrevem variável aleatória específicas. É importante ressaltar que existem vários modelos. No entato, traremos alguns modelos discretos, os quais podem a vir ser utilizados posteriormente e cujos exemplos foram discutidos na obra [5].

Definição 1.11 (**Modelo Uniforme Discreto**). Uma variável é caracterizada pelo modelo Uniforme Discreto, com valores $x_1, x_2, \ldots, x_k$, se tem função de probabilidade dada por: $p(x_i) = 1/k$, para $i = 1, 2, \ldots, k$. Usamos a notação $X \sim U_d[E]$, com E sendo o conjunto de seus valores e no caso em que $\sharp E < \infty$.

Neste modelo temos situações onde os possíveis valores da variável são equiprováveis, não existindo restrição aos valores da variável, podendo ser qualquer número real. No entanto, o número de valores diferentes precisa ser finito. Ou seja, $\sharp E < \infty$. A função de distribuição de uma variável Uniforme Discreta é uma função escada e os pontos de descontinuidade são os valores assumidos pela variável.

Exemplo 1.10. Em uma urna existem bolas numeradas de 1 até 6. Ao escolher uma determinada bola observamos o número que ocorreu.

Sendo X essa variável, é fácil verificar que $X \sim U_d[E]$, com $E = \{1, 2, 3, 4, 5, 6\}$. Sua função de probabilidade é $p(x) = 1/6$ para $x = 1, 2, \ldots, 6$.

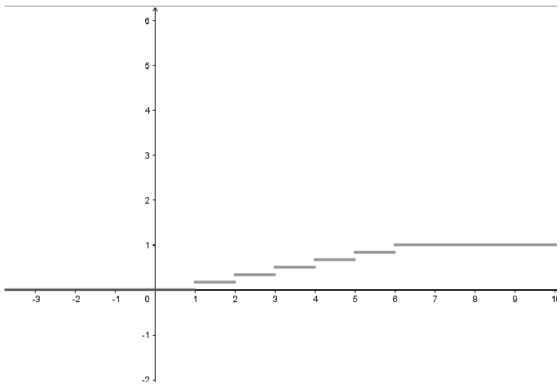

Figura 1.4.1: Função de distribuição da Uniforme Discreta em $\{1, 2, \ldots, 6\}$.

Definição 1.12 (**Modelo Bernoulli**). Uma variável aleatória segue o modelo Bernoulli, se assume apenas os valores 0 e 1. Sua função de probabilidade é dada por:

$$p(1) = P(X = 1) = p$$
$$p(0) = P(X = 0) = 1 - p$$

A notação que será utilizada será $X \sim Bernoulli(p)$.

Nesta probabilidade, p é denominada de parâmetro do modelo. É de praxe considerar sucesso a ocorrência de 1 e fracasso a ocorrência de 0. Assim, denominamos por ensaio de Bernoulli, o experiemento que tem resposta dicotômica do tipo sucesso-fracasso.

Uma boa exemplificação é o tiro ao alvo. Definindo sucesso como o acerto do alvo e fracasso como o erro do alvo, temos:

$$X = \begin{cases} 1, & se\, acertou; \\ 0, & se\, errou. \end{cases}$$

Então, $X \sim Bernoulli\,(p)$, com $p = P(acerto)$. Logo, $p = \frac{1}{2}$.

A repetição de sucessivos ensaios de Bernoulli serve de estudo para vários problemas. A partir de uma sequência de n ensaios de Bernoulli independentes, outros modelos podem ser contruídos.

Definição 1.13 (**Modelo Binomial**)**.** Seja X o modelo total de sucesso obtidos na observação de n ensaios de Bernoulli independentes. X segue o modelo Binomial com parâmetros n e p e sua função de probabilidade é dada por:

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x},\ x = 0, 1, \ldots, n.$$

A notação será $X \sim B(n, p)$.

Exemplo 1.11. A taxa de imunização de uma vacina é 70%. Se um grupo de 20 pessoas foram vacinadas, desejamos saber o comportamento probabilístico do número de pessoas imunizadas desse grupo.

Seja Xa variável de interesse. Para cada pessoa do grupo, a probabilidade de estar imunizada é $0,7$ e admitimos, ainda, independência entre os resultados das várias pessoas vacinadas. Assim, teremos $X \sim B\,(n = 20, p = 0,7)$. Por exemplo, a probabilidade de 10 estarem imunizados é dada por:

$P(X = 10) = p(10) = \binom{20}{10} 0,7^{10} 0,3^{20-10} = 0,0308.$

Definição 1.14 (**Modelo Geométrico**)**.** Considere uma sequência de ensaios de Bernoulli independentes. Sendo X o número de fracassos anteriores ao primeiro sucesso, ou simplesmente, o tempo de espera para o primeiro sucesso. A variável Xsegue o modelo Geométrico com parâmentro $p, 0 < p < 1$, e tem função de probabilidade dada por:

$$p(x) = p(1-p)^x, x = 0, 1, \ldots.$$

A notação será $X \sim Geo(p)$.

A restrição dos valores de p para aqueles estritamente entre 0 e 1 evita os casos imediatos mas, cuidadosamente, os extremos 0 e 1 poderiam ser considerados. O modelo Geométrico também pode ser definido como o número de fracassos até o primeiro sucesso, ou simplesmente, como o número de ensaios até o primeiro sucesso. Neste caso, a função de probabilidade sofre modificação, pois a variável inicia seus valores em 1 ao invés 0.

Exemplo 1.12. Uma linha de fabricação de um equipamento de precisão é interrompida na primeira ocorrência de um defeito. A partir da manutenção, o equipamento tem probabilidade de $0,01$ de apresentar defeito em um dia qualquer. Deseja-se planejar o cronograma de manutenção preventiva e, para tal, decidiu-se avaliar probabilisticamente a espera até a produção ser interrompida. Seja X a variável aleatória que conta o número de dias que antecedem a interrupção. Admitindo que o desemprenho, nos sucessivos dias, sejam independentes, temos que $X \sim Geo(p = 0,01)$. Dessa forma, $P(X = x) = 0,01 \times 0,99^x, x = 0, 1, \ldots$.

Por exemplo, para uma interrupção no sexto dia temos $P(X = 5) = 0,01 \times 0,99^5 = 0,0095$.

Qual seria o intervalo ideal para uma manutenção preventiva, se desejamos uma probabilidade de, pelo menos, $0,90$ de que o defeito não ocorrerá? A pergunta estará respondida, se determinamos quantos dias são necessários para acumular uma probabilidade de defeito próxima de $0,10$. Ou ainda, obter k tal que $P(X \leq k) = 1 - 0,99^{k+1} \simeq 0,10$. Com o auxílio de uma planilha calculadora, obtemos $P(X \leq 9) = 0,0956$ e $P(X \leq 10) = 0,1047$. Para atender o requisito desejado, a manutenção preventiva deverá ser feita após 9 dias de operação. Assim, teremos probabilidade de $0,9044$ de um defeito não ocorrer entre essas manutenções.

Capítulo 2

Cadeias de Markov

Nesta parte do trabalho, abordaremos os conceitos básicos da teoria das Cadeias de Markov e seus exemplos retirados da obra [5] os quais serão utilizados nas aplicações que virão em sequência. Além disso, com este aporte teórico será possível sugerir também uma sequência didática. Caso o leitor se interesse em um maior aprofundamento da teoria, sugerimos além de [5], as referências [2, 10, 13, 15].

2.1 Exemplos de Processos Estocásticos

Fenômenos aleatórios que evoluem ao longo do tempo são modelados por processos estocásticos. Devido a natureza randômica destes fenômenos essas mudanças não são totalmente previsíveis. O estudo da sua evolução se dará através de suas distribuições de probabilidade. Diversos fenômenos reais permitem a modelagem através de um processo estocástico. Observe alguns exemplos da obra de [5].

Exemplo 2.1. Os parafusos de uma cadeia de montagem após uma supervisão à que são submetidos, podem ser considerados defeituosos ou não. Se o n-ésimo parafuso não tiver defeito, atribuímos valor 1 a esse parafuso. Caso o n-ésimo parafuso seja defeituoso atribuímos o valor 0 a esse parafuso. Podemos definir uma variável aleatória que represente este fenômeno, digamos X_n, com espaços de estados $\{0, 1\}$, que representa a imperfeição ou não do n-ésimo parafuso. Temos assim uma família de variáveis aleatórias $\{X_n\}_{n\in\mathbb{N}}$ em que cada um delas corresponde a imperfeição do n-ésimo parafuso.

Não temos como determinar o estado do n-ésimo parafuso. Questionamos então acerca da probabilidade do parafuso ser defeituoso ou não, isto é, $P(X_n = 0)$ e $P(X_n = 1)$. Como um processo evolutivo, tais probabilidades podem depender ou não do estado dos $n-1$ parafusos anteriores. Suponha, por exemplo que um parafuso é defeituoso independente dos outros já produzidos e que a probabilidade seja p. Neste caso $\{X_n\}_{n\in\mathbb{N}}$ é uma fórmula de variáveis aleatórias independentes de *Bernoulli* com parâmetro p de sucesso, conforme 1.12.

Exemplo 2.2. Uma pequena loja de equipamentos eletrônicos vende um tipo de microsystem. No entanto, ela só pode ter em estoque no máximo sete unidades. Então, se no final do dia a loja tem no estoque somente uma unidade ou nenhuma, o gerente manda buscar o restante das unidades necessárias para ter sete unidades na loja no dia seguinte antes de começar o expediente de atendimento. Desta forma, podemos ter as seguintes quantidades no final do n-ésimo dia $\{0, 1, 2, 3, 4, 5, 6, 7\}$. Podemos definir uma variável aleatória que represente este fenômeno, digamos X_n, que é justamente à quantidade de unidades na loja no final do n-ésimo dia. Elas podem ser consideradas variáveis aleatórias, pois é fácil supor que não tem como prever a quantidade de microsystems que serão comprados cada dia.

Exemplo 2.3. Considere a história de várias gerações de uma família que ao longo do tempo tem somente um filho. Neste modelo, observe que a classe social (alta, média ou baixa) da família para cada geração permite descrever sua evolução social ao longo do tempo.

Neste caso, uma sociedade composta por famílias deste tipo, podemos escolher ao acaso uma família e para cada geração X_n uma quantidade que valerá 1 se a família é de classe alta, 2 se a família é de classe média e 3 se a família é de classe baixa. Desta forma, cada X_n será uma variável aleatória e sua evolução ao longo do tempo, permitiria concluir sobre as mudanças na estrutura da sociedade.

Exemplo 2.4. Suponhamos que uma empresa de seguros na praça receba c unidades monetárias ($u.m.$) pelo total dos prêmios que ela cobra dos afiliados dentro de um determinado tempo (mês, semestre, ano). Sabe-se também que a seguradora coleta os prêmios regularmente e que as idenizações são pagas quando os sinistros ocorrem. Além disso, as despesas administrativas serão desconsideradas, ou seja, ganhos ou perdas por investimentos, etc. Desta forma, a reserva da empresa de seguros será afetada somente pela cobrança dos prêmios ou por pagamentos dos idenizados na corrência dos sinistros. Assim, o lucro da empresa no n-ésimo período será $c - Z_n$ $u.m.$ sendo Z_n o valor total das idenizações pago pela empresa nesse período. Se chamar de L_n ao lucro da seguradora desde que ela começa a operar até o final do n-ésimo período, teremos

$$L_n = cn - \sum_{j=1}^{n} Z_j \, .$$

Definição 2.1. Um processo estocástico é uma família de variáveis aleatórias $\{X_t\}_{t \in T}$. Chamamos T o **espaço de parâmetros**.

Em geral t é utilizado como tempo, mas poderia indicar outros parâmetros como uma área ou medida de massa, conforme 2.1. Nos exemplos 2.1, 2.2, 2.3 e 2.4 temos que T $= \mathbb{N}$. No caso em que T é enumerável dizemos que o processo estocástico correspondente é a **tempo discreto**. Se T for um intervalo, o processo estocástico será chamado a **tempo contínuo**. Note que, os exemplos 2.1, 2.2, 2.3 e 2.4 são todos discretos.

Estados serão os valores que tomam as variáveis do processo e o conjunto destes estados chamamos de **espaço de estados**, que denotamos por E. Podemos classificar o **espaço de estados** como **discreto** ou **contínuo**, caso E seja enumerável ou não. No exemplo 2.1 temos que $E = \{0, 1\}$, no exemplo 2.2 $E = \{0, 1, 2, 3, 4, 5, 6, 7\}$, no exemplo 2.3 $E = \{1, 2, 3\}$ e no exemplo 2.4 temos que $E = \mathbb{R}$. Logo, nos exemplos 2.1, 2.2 e 2.3 temos **espaços de estados discretos**, enquanto que no exemplo 2.4 temos o **espaço de estados contínuos**.

	E enumerável	E não enumerável
T enumerável	Processo a tempo discreto com espaço de estados discreto	Processo a tempo discreto com espaço de estados contínuo
T não enumerável	Processo a tempo continuo com espaço de estados discreto	Processo a tempo contínuo com espaço de estados contínuo

Tabela 2.1: Classificação dos processos estocásticos.

Uma **trajetória** de um processo estocástico $\{X_t\}_{t\in T}$ é uma coleção de valores no espaço de estados de cada variável aleatória X_t, ou seja, $\{x_t : t \in T\}$. No exemplo 2.1 poderíamos ter a seguinte trajetória $\{1, 0, 0, 0, 0, 0, 0, ...\}$. Ou seja, um primeiro parafuso sem defeito e o restante com defeito.

Pode-se entender os processos estocásticos como generalizações de variáveis e vetores aleatórios. Um vetor aleatório é justamente um vetor em que suas componentes são variáveis aleatórias. Se o processo estocástico $\{X_t\}_{t\in T}$ é tal que a cardinalidade do espaço de parâmetros T é finita, então o processo será um vetor aleatório. Caso T $= \{1, 2\}$, pode representar pelo vetor aleatório (X_1, X_2). Caso T tenha apenas um elemento, então temos somente uma variável aleatória.

O comportamento probabilístico de variáveis aleatórias é descrito através da função de distribuição. No caso de vetores aleatórios usa-se a função de distribuição conjunta. A distribuição conjunta é uma função $P : A \to [0, 1]$ tal que $A = \{(x, y) \mid x, y \in E\}$, satisfazendo as seguintes condições:

1. $0 \leq P(x, y) \leq 1$;

2. $\sum_x \sum_y P(x, y) = 1$.

A partir de agora, um processo estocástico $\{X_t\}_{t\in T}$ será denotado apenas por X, para simplificar a notação.

Ao analisar os estados de um processo X pode-se encontrar relações de dependência. Utilizaremos a interpretação do espaço de parâmetros T como tempo.

A relação de dependência entre variáveis mais simples seria a ausência total dela. Chamamos de **processo de estados independentes** àquele processo estocástico cujos estados constituem uma família de variáveis aleatórias independentes, ou seja, para as variaveis aleatórias X e Y temos que $P(X = x_i, Y = y_i) = P(X = x_i) \cdot P(Y = y_i)$,com $i \in \mathbb{N}$.

Um processo estocástico X é dito um **processo de Markov, ou Markoviano,** se para todo $a, b, a_1, \ldots, a_n \in E$, vale:

$$P[a \leq X_t \leq b \mid X_{t_1} = a_1, X_{t_2} = a_2, \ldots, X_{t_n} = a_n] = P[a \leq X_t \leq b \mid X_{t_n} = a_n], \quad (2.1.1)$$

quaisquer que sejam os instantes $t_1, t_2, t_3, \ldots, t_n, t \in \mathrm{T}$, com $t_1 < t_2 < \cdots < t$.

Ou seja, o estado X_t do processo depende da sua história anterior nos instantes $t_1, t_2, \ldots, t_n$ somente através do **presente** X_{t_n}e não do **passado** $X_{t_1}, X_{t_2}, \ldots, X_{t_{n-1}}$. Os processo de estados independentes são exemplos muito simples de **processos de Markov.**

2.2 Cadeias de Markov a Tempo Discreto

Estamos interessados em processos estocásticos que admitam uma representação matricial. Veremos que se o espaço de estados é discreto podemos representar o processo estocástico, pela matriz de suas probabilidades. Ademais, se o processo é Markoviano, podemos estudar a sua evolução através do produto destas matrizes. Vejamos o seguinte exemplo.

Exemplo 2.5. Considere um jogo no qual em cada aposta você perde um real com probabilidade $0,6$ ou ganha um real com probabilidade $0,4$. Suponha que você deseja parar de jogar se a sua fortuna atingir N reais e se ela atingir 0 reais o cassino não deixa vcê jogar mais.

Seja X_n a quantidade de dinheiro que você tem depois de n apostas. Note que, esta quantidade e o resultado do próximo sorteio vão determinar a sua fortuna depois da aposta seguinte. Qualquer que tenha sido a evolução da sua fortuna no "passado"(ou seja, os valores $X_{n-1}, X_{n-2}, \ldots, X_0$), não intefere na previsão do próximo estado X_{n+1}. É suficiente conhecer a sua fortuna no "presente" (X_n) . De fato, se $X_n = i$, com $0 < i < N$, então independente dos valores $i_0, \ldots, i_{n-1}$,teremos que:

$$P(X_{n+1} = i + 1 \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_o = i_o) = 0,4.$$

Desta forma, se você ganhar a aposta $n + 1$, a sua fortuna vai ser acrescentada em um real e portanto é suficiente conhecer o valor de sua fortuna no presente.

Ao decorrer das apostas sua fortuna irá aumentar ou diminuir em um real com uma chance que não depende do número de apostas que você fez. Portanto, esta probabilidade condicional não depende de n. Fixe o valor de $N = 5$. Então os valores que pode ser a sua fortuna são $\{0, 1, 2, 3, 4, 5\}$. Suponha que depois de certa quantidade de apostas, você tem $R\$2,00$. Note que, pode acontecer de você possuir $R\$1,00$ ou $R\$3,00$, na próxima aposta, dependendo de sua sorte. Logo, é possível arranjar as apostas através de um vetor linha da forma $(0; 0,6; 0; 0,4; 0)$. E quando fazemos o somatório da probabilidade sempre teremos o resultado igual a 1, pois estamos considerando todos os valores possíveis da sua fortuna, ou seja, o vetor é justamente uma distribuição de probabilidade. Fazendo isso para cada valor possível da sua fortuna conseguimos uma matriz, onde cada linha corresponde a um vetor de probabilidades:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0,6 & 0 & 0,4 & 0 & 0 & 0 \\ 0 & 0,6 & 0 & 0,4 & 0 & 0 \\ 0 & 0 & 0,6 & 0 & 0,4 & 0 \\ 0 & 0 & 0 & 0,6 & 0 & 0,4 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Definição 2.2. Uma cadeia de Markov a tempo discreto é um processo estocástico discreto $\{X_n\}_{n \in \mathrm{T}}$, com o espaço de estados E finito ou enumerável e que possui a propriedade de Markov:

$$P(X_{n+1} = j \mid X_0 = i_0, \ldots, X_n = i_n) = P(X_{n+1} = j \mid X_n = i_n),$$

para todos os estados $i_0, \ldots, i_n, j$ e todo instante n.

Observe que a propriedade de Markov 2.2 é caso particular da propriedade apresentada em 2.1.1. Uma cadeia de Markov a tempo discreto é um proesso de Markov discreto com espaço de estados finito.

Na definição 2.2, o estado *futuro* do processo, $X_{n+1} = j$, não depende do *passado*, $X_0 = i_o, \ldots, X_{n-1} = i_{n-1}$, e só depende do *presente*, $X_n = i_n$. A probabilidade condicional $P(X_{n+1} = j \mid X_n = i_n)$ é chamada de **probabilidade de transição**(i_n para j no tempo n para $n + 1$).

No nosso trabalho nos restringimos a ***cadeias de Markov homogêneas***, cujas probabilidades de transição independem de n, ou seja,

$$P(X_{n-1} = j \mid X_n = i) = \cdots = P(X_1 = j \mid X_0 = i) = P_{i,j}, \forall i, j \in E.$$

Ademais, pediremos que $\sharp E < \infty$.

Assim, $P_{i,j}$ é a probabilidade de passar, em qualquer instante, do estado i ao estado j.

A matriz de transição é justamente o arranjo das probabilidades de transição $P_{i,j}$ numa matriz P. Se E é finito, por exemplo $E = \{0, 1, 2, \ldots, N\}$, então:

$$P = \begin{bmatrix} P_{0,0} & P_{0,1} & \cdots & P_{0,N} \\ P_{1,0} & P_{1,1} & \cdots & P_{1,N} \\ \vdots & \vdots & \ddots & \vdots \\ P_{N,0} & P_{N,2} & \cdots & P_{N,N} \end{bmatrix}.$$

No exemplo 2.5, temos que:

$$P_{i,i+1} = 0.4,\ P_{i,i-1} = 0,6, se\, 0 < i < N,\ P_{0,0} = 1 = P_{N,N}, P_{ij} = 0,\ caso\, contrário.$$

Caso E seja infinito, por exemplo $E = \mathbb{N}$, temos infinitas probabilidades de transição, que poderíamos arranjar em uma matriz infinita P descrita abaixo.

$$P = \begin{bmatrix} P_{0,0} & P_{0,1} & P_{0,2} & \cdots \\ P_{1,0} & P_{1,1} & P_{1,2} & \cdots \\ P_{2,0} & P_{2,1} & P_{2,2} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Podemos descrever estas transições através da seguinte interpretação geométrica:

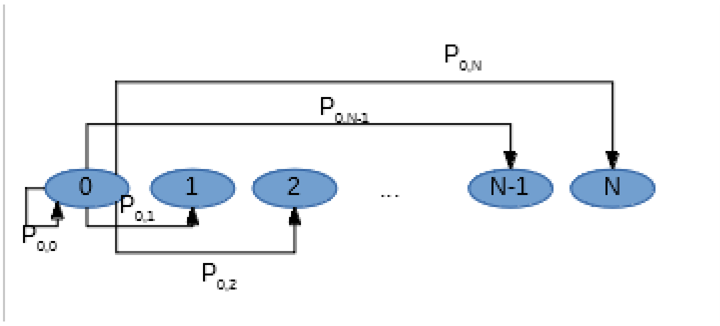

Figura 2.2.1: Grafo chamado de Topologia da Cadeia.

Contudo, ferramentas e conceitos mais complexos serão requeridos para o estudo deste tipo de processo. Por isso, nos concentramos em cadeias de Markov homogêneas com espaço de estados finito.

Exemplo 2.6 (Cadeias de Ehrenfest). Suponhamos que o total de bolas contidas em duas urnas é N. A cada instante de tempo n, pegamos uma bola da primeira urna e colocamos na segunda urna. Definimos X_n como a quantidade de bolas na primeira urna. Então X_n é uma cadeia de Markov com espaço de estados $E = \{0, 1, \ldots, N\}$. Podemos calcular as probabilidades de transição.

Note que, se em algum instante não tivermos bolas da primeira urna então necessariamente no instante seguinte teremos que passar uma bola da segunda urna para a primeira. Desta forma, $P_{0,1} = 1$. Analogamente, $P_{N,N-1} = 1$. Se $1 < i < N$, então $P_{i,i-1} = i/N$ e $P_{i,i+1} = (N - i)/N$.

Para $N = 3$, temos a seguinte matriz de transição

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

Portanto, uma cadeia de Markov homogênea discreta corresponde a uma matriz de transição. A seguir iremos abordar as principais caracterísitcas destas matrizes:

Definição 2.3. Uma matriz $P = [P_{i,j}]_{i,j \in E}$ é uma matriz estocástica se:

$i)$ $P_{i,j} \geq 0$, para todo $i, j \in E$;

$ii)$ $Para\, todo\, i \in E, \sum_{j \in E} P_{i,j} = 1.$

Note que, toda matriz de transição é uma matriz estocástica. Pois, através dos exemplos colocados anteriormente e do que foi citado acima temos que todas as entradas de uma matriz estocástica são não negativas e qualquer linha tem soma igual a 1.

Exemplo 2.7. Uma cadeia de *Markov* com espaço de estado $E = \{0, 1, \ldots, d\}$ e com probabilidade de transição

$$P_{i,j} = \begin{cases} q_i, & se\, j = i - 1 \\ r_i, & se\, j = i \\ p_i, & se\, j = i + 1 \end{cases}$$

onde $p_i + r_i + q_i = 1$ e $q_0 = 0$ e $p_d = 0\, se\, d < \infty$, é chamado de *Processo de Nascimento e Morte.*

Para exemplificar melhor temos a seguinte situação. Em um hospital, foi feito um levantamento junto aos pacientes para prever o possível aumento, diminuição ou estagnação da verba para oferecer um melhor tratamento de saúde. No caso, os pacientes foram classificados da seguinte forma: (0) para o paciente que venha a óbito, (1) para o paciente curado e (2) para o paciente doente. Desta forma, um paciente que esteja curado (1) pode permanecer curado (1), adoecer (2) ou vim a óbito (0). Já um paciente que esteja doente (2), ele pode continuar doente (2), ser curado (1) ou então vim a óbito (0). Já o paciente que venha a óbito (0), ele deverá permanecer nesta situação. Assim, podemos construir a matriz de transição que é dada abaixo.

$$P = \begin{bmatrix} P_{0,0} & P_{0,1} & P_{0,2} \\ P_{1,0} & P_{1,1} & P_{1,2} \\ P_{2,0} & P_{2,1} & P_{2,2} \end{bmatrix}.$$

2.3 Matrizes de Transição de Ordem Superior

No início deste capítulo, vimos no exemplo 2.3 que $\{X_n\}$ é um processo estocástico. Assumindo que o processo X_n é uma cadeia de Markov, consideremos a seguinte matriz de transição

$$\begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.4 & 0.4 \end{bmatrix}.$$

Desta forma, a matriz indica que a probabilidade de que os filhos de uma família de classe baixa (1) permaneça nessa classe é $0,7$.

Suponhamos que a família começa na classe média (2) na geração 0. Qual a probabilidade que a geração 1 ascenda à classe alta (3) e a geração 2 desça para a baixa (1)?

Assim, iremos calcular a probabilidade $P\left(X_1 = 3, X_2 = 1 \mid X_0 = 2\right)$. Note que, é a probabilidade de ir em um passo do estado (2) para o (1) e depois do (1) para o (3). Pela propriedade de Markov, esta probabilidade deve ser $P_{2,1}P_{1,3}$. Logo, pela Proposição 1.9, temos que

$$\begin{aligned} P(X_1 = 3, X_2 = 1 \mid X_0 = 2) &= \frac{P(X_2 = 1, X_1 = 3, X_0 = 2)}{P(X_1 = 3, X_0 = 2)} \frac{P(X_1 = 3, X_0 = 2)}{P(X_0 = 2)} \\ &= P(X_2 = 1 \mid X_1 = 3, X_0 = 2) P(X_1 = 3 \mid X_0 = 2) \\ &= P(X_2 = 1 \mid X_1 = 3) P(X_1 = 3 \mid X_0 = 2). \end{aligned}$$

Em geral, temos o seguinte resultado:

Teorema 2.1. *Sejam* $i_0, i_1, i_2, \ldots, i_n \in E$.

$$
\begin{aligned}
P(X_{n+1} = i_1, \ldots, X_{n+m} = i_m \mid X_n = i_0) &= \prod_{j=1}^{m} P\left(X_{n+j} = i_j \mid X_{n+j-1} = i_{j-1}\right) \\
&= P_{i_{m-1},i_m} P_{i_{m-2},i_{m-1}} \ldots P_{i_0,i_1} \\
&= P_{i_0,i_1} P_{i_1,i_2} \ldots P_{i_{m-1},i_m}.
\end{aligned}
$$

Exemplo 2.8. Agora considere que a família começa na classe média (2) na geração 0. Vamos calcular a probabilidade que a geração 2 ascenda para a classe baixa (1).

Considerando os três casos possíveis para a geração 1 e usando o Teorema *2.1* visto anteriormente.

$$
\begin{aligned}
P(X_2 = 1 \mid X_0 = 2) &= \sum_{k=1}^{3} P(X_1 = k, X_2 = 1 \mid X_0 = 2) \\
&= \sum_{k=1}^{3} P_{2,k} P_{k,1} \\
&= 0.3 \cdot 0.7 + 0.5 \cdot 0.3 + 0.2 \cdot 0.2 \\
&= 0.4.
\end{aligned}
$$

Podemos representar através do seguinte gráfico:

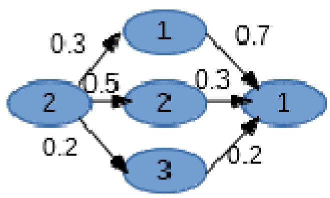

Figura 2.3.1: Gráfico de árvore.

Analogamente, podemos provar que para $i, j \in \{1, 2, 3\}$ vale

$$
P(X_2 = j \mid X_0 = i) = \sum_{k=1}^{3} P_{i,k} P_{k,j}. \tag{2.3.1}
$$

Note que, o termo da direita na igualdade anterior é o coeficiente (i, j) da matriz P^2. E se for homogênea, o termo da esquerda é justamente a probabilidade de passar do estado i para o estado j em dois passos. Desta forma, vale que a probabilidade de transição em m passos de uma cadeia de Markov X é dada por

$$
P_{i,j}^{(m)} := P(X_{m+n} = j \mid X_n = i) = P(X_m = j \mid X_0 = i),
$$

onde (i,j) é a entrada da $m-ésima$ potência da matriz de transição P. Ou seja, $P^m = \underbrace{P \cdot P \cdot \ldots \cdot P}_{m\,vezes}$ existe m termos no produto. Vejamos:

Como é homogênea qualquer transição de ordem 2 é dada por 2.3.1. Assim, para as transições de **ordem dois,** ou seja, entre os tempos n e $n+2$ vale:

$$\begin{aligned} P_{i,j}^{(2)} = P(X_{n+2} = j \mid X_n = i) &= \sum_{k\in E} P(X_{n+2} = j, X_{n+1} = k \mid X_n = i) \\ &= \sum_{k\in E} P(X_{n+2} = j \mid X_{n+1} = k) P(X_{n-1} = k \mid X_n = i) \\ &= \sum_{k\in E} P_{i,k}P_{k,j}. \end{aligned}$$

Novamente, temos que o termo da direita da última expressão é o elemento (i,j) da matriz $P^2 = P \cdot P$. Analogamente podemos encontrar as transições de ordem **três**, **quatro** e assim por diante. Ou seja, entre os tempos n e $n+3$, n e $n+4$, que é equivalente a 0 a 3 e 0 a 4.

Em geral, vale que $P_{i,j}^{(m)}$ é o elemento da (i,j) da matriz P^m. Isto decorre do seguinte resultado:

Teorema 2.2. ***Equações de Chapman-Kolmogorov***

$$P_{i,j}^{(m+n)} = \sum_{k\in E} P_{i,k}^{(m)} P_{k,j}^{(n)}.$$

Demonstração. Temos que,

$$P(X_{n+m} = j \mid X_0 = i) = \sum_{k\in E} P(X_{n+m} = j, X_m = k \mid X_0 = i).$$

Utilizando a definição da probabilidade condicional, cada um dos somandos pode ser escrito da forma,

$$\begin{aligned} P(X_{n+m} = j, X_m = k \mid X_0 = i) &= \frac{P(X_{n+m} = j, X_m = k, X_0 = i)}{P(X_0 = i)} \\ &= \frac{P(X_{n+m} = j, X_m = k, X_0 = i)}{P(X_m = k, X_0 = i)} \frac{P(X_m = k, X_0 = i)}{P(X_0 = i)} \\ &= P(X_{n+m} = j \mid X_m = k, X_0 = i) P(X_m = k \mid X_0 = i) \\ &= P(X_{n+m} = j \mid X_m = k) P(X_m = k \mid X_0 = i). \end{aligned}$$

Note que, na última linha usamos a propriedade de Markov. Substituindo na igualdade acima obtemos o resultado desejado. $\square$

Note que, ao chamarmos de $P^{(m)} = (P_{i,j}^{(m)})$ à matriz de transição de ordem m, teremos que o teorema acima afirma que $P^{(m+n)} = P^{(m)} \cdot P^{(n)}$. Como $P^{(1)} = P$, temos então que $P^{(n+1)} = P \cdot P^{(n)}$ e, usando o argumento indutivo obtemos que $P^{(n)} = P^n$. De fato, para $n = 1$ temos que $P^{(1)} = P^1 = P$ que é válida pela definição. Suponhamos que a sentença seja válida para $n = m$, com $m \in \mathbb{N}$ e $m > 1$, isto é, $P^{(m)} = P^m$, queremos mostrar que a sentença é válida quando $n = m + 1$. Usando a *Equação de Chapman-komogorov*, para $n = 1$, temos que

$$P^{(m+1)} = P_{i,j}^{(m+1)} = \sum_{k \in E} P_{i,k}^{(1)} \cdot P_{k,j}^{(m)} = \sum_{k \in E} P_{i,k}^{1} \cdot P_{k,j}^{m} = P \cdot P^m = P^{m+1}.$$

Agora, iremos ver como as distribuições conjuntas 2.1 de estados do processo estão determinadas pela matriz de transição e a distribuição de probabilidade do estado incial.

Definição 2.4. Seja $\pi_0 : E \to [0, 1]$ uma distribuição de probabilidade no conjunto E, tal que

1. $\pi_0(i) \geq 0, para\, todo\, i \in E$
2. $\sum_{i \in E} \pi_0(i) = 1.$

Dizemos que π_0 é uma distribuição inicial da cadeia de Markov $\{X_t\}_{t \in \mathbb{N}}$ se para todo $i \in E$ vale $P(X_0 = i) = \pi_0(i)$.

Logo, a distribuição inicial de uma cadeia é simplesmente a função de probabilidade do seu estado incial X_0.

Através do teorema da probabilidade total podemos obter a distribuição de qualquer um dos estados em função da matriz de transição e da distribuição inicial. Assim, para todo $n \in \mathbb{N}$,

$$\begin{aligned} P(X_n = k) &= \sum_{i \in E} P(X_n = k \mid X_0 = i) P(X_0 = i) \\ &= \sum_{i \in E} P_{i,k}^{(n)} \pi_0(i). \\ &= \pi_0 P^n. \end{aligned}$$

Note que, π_0 é justamente o *vetor linha* dos valores da distribuição inicial da cadeia. Considere $\pi_n = \begin{pmatrix} P(X_n = 1) & P(X_n = 2) & P(X_n = 3) \end{pmatrix}$ que fornece $\pi_n = \pi_0 P^n$.

Através do teorema 2.1 podemos obter o seguinte resultado,

Proposição 2.1. *Seja π_0 a distribuição inicial da cadeia $\{X_n\}_{n\geq 0}$ que tem matriz de transição $P = (P_{i,j})_{i,j\in E}$. Sejam $i_0, i_1, i_2, \ldots, i_m \in E$, então vale*

$$P(X_0 = i_0, X_1 = i_1, \ldots, X_m = i_m) = \pi(i_0)P_{i_0,i_1}\cdots P_{i_{m-1},i_m}.$$

Demonstração. Usando o Teorema 2.2, temos que

$$\begin{aligned} P(X_0 = i_0, X_1 = i_1, \ldots, X_m) = i_m &= P(X_1 = i_1, \ldots, X_m = i_m \mid X_0 = i_0)P(X_0 = i_0) \\ &= P_{i_{m-1},i_m}\cdots P_{i_0,i_1}\pi_0(i_0). \end{aligned}$$

□

2.4 Cadeias com Dois Estados

Observamos que no caso em que temos o espaço de estados finito as probabilidades de transição de ordem superior foi calculada multiplicando a matriz de transição P por ela mesma. Assim, em alguns casos depois de certa ordem as filas vão se aproximando entre si. Ou seja, os valores das entradas nas matrizes vão se aproximando. Agora nesta parte do trabalho iremos abordar os casos em que temos cadeias com dois estados. Neste caso particular, fornecemos estimativas sobre o comportamento assintótico da evolução da cadeia de Markov.

Exemplo 2.9. Considere a cadeia de Markov X_n com espaço de estados $E = \{0, 1\}$ e matriz de transição,

$$P = \begin{bmatrix} 0.5 & 0.5 \\ 0.3 & 0.7 \end{bmatrix}.$$

Então:
$P^2 = \begin{bmatrix} 0.40 & 0.60 \\ 0.36 & 0.64 \end{bmatrix}$, $P^3 = \begin{bmatrix} 0.38 & 0.62 \\ 0.372 & 0.628 \end{bmatrix}$, $P^4 = \begin{bmatrix} 0.3760 & 0.6240 \\ 0.3744 & 0.6256 \end{bmatrix}$.
No caso em que E tem apenas dois elementos a matriz de transição toma a forma:

$$P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}, \textit{ onde } 0 \leq p \leq 1, 0 \leq q \leq 1.$$

Ou seja,

- $P(X_{n+1} = 0 \mid X_n = 0) = 1 - p$,
- $P(X_{n+1} = 1 \mid X_n = 0) = p$,

- $P(X_{n+1} = 0 \mid X_n = 1) = q$,
- $P(X_{n+1} = 1 \mid X_n = 1) = 1 - q$.

Veremos que, se $p - q > 0$ a matriz P pode ser escrita como:

$$P = \frac{1}{p+q}\begin{bmatrix} q & p \\ q & p \end{bmatrix} + \frac{1-p-q}{p+q}\begin{bmatrix} p & -p \\ -q & q \end{bmatrix}.$$

Usando as seguintes relações:

$$\begin{bmatrix} q & p \\ q & p \end{bmatrix}^2 = p+q \begin{bmatrix} q & p \\ q & p \end{bmatrix},$$

$$\begin{bmatrix} q & p \\ q & p \end{bmatrix}\begin{bmatrix} p & -p \\ -q & q \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

e

$$\begin{bmatrix} p & -p \\ -q & q \end{bmatrix} = (p+q)\begin{bmatrix} p & -p \\ -q & q \end{bmatrix},$$

é possível provar por argumento indutivo que

$$P^n = \frac{1}{p+q}\begin{bmatrix} q & p \\ q & p \end{bmatrix} + \frac{(1-p-q)^n}{p+q}\begin{bmatrix} p & -p \\ -q & q \end{bmatrix}.$$

De fato, para $n = 1$,temos que:

$$P^1 = \frac{1}{p+q}\begin{bmatrix} q & p \\ q & p \end{bmatrix} + \frac{(1-p-q)^1}{p+q}\begin{bmatrix} p & -p \\ -q & q \end{bmatrix} = P.$$

Suponhamos que para $m \in \mathbb{N}$, a sentença também é válida. Queremos mostrar que para $m+1$ a sentença também é válida. Para simplificarmos a escrita escrevemos $A = \begin{bmatrix} q & p \\ q & p \end{bmatrix}$ e $B = \begin{bmatrix} p & -p \\ -q & q \end{bmatrix}$. Observe que:

$$\begin{aligned}
P^{m+1} &= P^m \cdot P \\
&= \left(\frac{1}{p+q} \cdot A + \frac{(1-p-q)^m}{p+q} \cdot B\right)\left(\frac{1}{p+q} \cdot A + \frac{1-p-q}{p+q} \cdot B\right) \\
&= \frac{1}{p+q} \cdot \frac{1}{p+q} \cdot A^2 + \frac{1}{p+q} \cdot \frac{1-p-q}{p+q} \cdot A \cdot B + \frac{(1-p-q)^m}{p+q} \cdot \frac{1}{p+q} \cdot B \cdot A \\
&\quad + \frac{(1-p-q)^m}{p+q} \cdot \frac{1-p-q}{p+q} \cdot B^2
\end{aligned}$$

Atente também que $A^2 = (p+q)\cdot A$, $A\cdot B = B\cdot A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ e $B^2 = (p+q)\cdot B$.

Então,

$$\begin{aligned} P^{m+1} &= \frac{1}{(p+q)^2}\cdot(p+q)\cdot A + \frac{(1-p-q)^{m+1}}{(p+q)^2}\cdot(p+q)\cdot B \\ &= \frac{1}{p+q}\begin{bmatrix} q & p \\ q & p \end{bmatrix} + \frac{(1-p-q)^{m+1}}{p+q}\begin{bmatrix} p & -p \\ -q & q \end{bmatrix}. \end{aligned}$$

Portatno, pelo Princípio da Indução Finita, a setença é válida para todo $n \in \mathbb{N}$.

Estudaremos o comportamento de P^n quando $n \to \infty$. Analisemos os três casos a seguir.

Se $p+q=0$. **Ou seja,** $p=0$ **e** $q=0:$

P é a matriz identidade de dimensão dois, vale que $P^n = P$ para todo $n \in \mathbb{N}$ e portanto as linhas não se aproximam entre si. Ou seja, os valores de entrada destas matrizes permanecem constantes. A cadeia vai visitar em todos os instantes o estado do qual ela começou.

Se $p+q=2$. **Ou seja,** $p=1$ **e** $q=1:$

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

Se n for par, teremos que P^n é a matriz identidade de ordem dois e para n ímpar, $P^n = P$. Como consequência disto temos que o limite de P^n quando $n \to \infty$ não existe pois a matriz oscila entre duas matrizes fixas. Ou seja, ao analisar este processo ao longo do tempo percebemos que os valores de entrada da matriz não convergem.

Se $0 < p+q < 2:$

Note que, $[0 > (-1)\cdot(p+q) > (-1)\cdot 2 \Rightarrow 0 > -p-q > -2 \Rightarrow 0+1 > 1-p-q > (-2)+1 \Rightarrow 1 > 1-p-q > -1 \Longleftrightarrow \mid 1-p-q \mid < 1]$ e portanto $(1-p-q)^n \to 0$ quando $n \to \infty$. Assim,

$$\lim_{n\to\infty} P^n = \begin{bmatrix} \frac{q}{p+q} & \frac{p}{p+q} \\ \frac{q}{p+q} & \frac{p}{p+q} \end{bmatrix}$$

No último caso, as linhas da matriz convergem para o vetor de probabilidades de uma distribuição que denotaremos como π_∞, isto é,

$$\lim_{n\to\infty} P^n_{0,0} = \lim_{n\to\infty} P^n_{1,0} = \pi_\infty(0) = \frac{q}{p+q},$$

$$\lim_{n\to\infty} P^n_{0,1} = \lim_{n\to\infty} P^n_{1,1} = \pi_\infty(1) = \frac{p}{p+q}.$$

Através do argumento indutivo também é possível obter uma estimativa para a taxa de convergência das probabilidades de transição em n passos para a distribuição π_∞.

Proposição 2.2. *Para uma cadeia de Markov $\{X_n\}_{n>1}$ com dois estados, $E = \{0, 1\}$ e tal que $0 < p + q < 2$, vale*

$$| P^n_{i,0} - \pi_\infty(0) |=| P^n_{i,0} - \frac{q}{p+q} |\leq| 1 - p - q |^n,\, com\, i = 0\, ou\, i = 1.$$

Estas probabilidades de transição se aproximam de π_∞ com velocidade exponencial. Por isso, de forma geral, quando $0 < p + q < 2$, observamos a proximidade entre as linhas de P^n mesmo para valores de n pequenos.

Exemplo 2.10. No exemplo **2.9,** temos que $p = 0,5$ e $q = 0,3$, portanto, $0 < p+q < 2$ e as linhas da matriz devem convergir para $\begin{bmatrix}\pi_\infty(0) & \pi_\infty(1)\end{bmatrix} = \begin{bmatrix}\frac{3}{8} & \frac{5}{8}\end{bmatrix} = \begin{bmatrix}0,375 & 0,625\end{bmatrix}$. A diferença entre elas vai para zero mais rápido que $(0,2)^n$. Observe que, para $n = 4$, obtivemos uma precisão de duas casas decimais.

Quando existe π_∞, temos que para instantes de tempo grandes, a matriz P^n se aproxima da matriz de transição de ordem n de uma sequência de variáveis aleatórias. Ou seja, para instantes tempo grandes, os estados irão se tornando "menos dependentes" e é natural pensar que a cadeia "ignora" o estado onde ela começou, em outras palavras, sua distribuição inicial. De fato,

Supondo a distribuição inicial da cadeia igual π_0, ou seja, temos $P(X_0 = 0) = \pi_0(0)$ e $P(X_0 = 1) = \pi_0(1)$, sabemos que $(P(X_n = 0), P(X_n = 1)) = (\pi_0(0), \pi_0(1)) \cdot P^n$. Usando a expressão que temos para P^n e o fato que π_0 é uma distribuição, obtemos:

$$P(X_n = 0) = \frac{q}{p+q} + (1 - p - q)^n(\pi_0(0) - \frac{q}{p+q}),$$

$$P(X_n = 1) = \frac{p}{p+q} + (1 - p - q)^n(\frac{q}{p-q} - \pi_0(0)).$$

Como no caso que estamos considerando vale $(1 - p - q)^n \to 0$ quando $n \to \infty$, concluimos que

$$\lim_{n\to\infty} P(X_n = 0) = \frac{q}{p+q} = \pi_\infty(0),$$

$$\lim_{n\to\infty} P(X_n = 1) = \frac{p}{p+q} = \pi_\infty(1).$$

As quantidades $\pi_\infty(0) = \frac{q}{p+q}$ e $\pi_\infty(1) = \frac{p}{p+q}$ podem ser interpretadas como as probabilidades da cadeia estar a longo prazo no estado 0 ou no estado 1, respectivamente, por isto π_∞ é chamada de **distribuição assintótica** da cadeia.

Capítulo 3

Exemplos de Cadeias de Markov para o Ensino Médio

Neste capítulo, exemplificaremos processos que podem ser abordados na educação básica. Ou seja, existe a possibilidade de montar projetos e roteiros de aula de uma maneira paupável e até mesmo interdisciplinar, pois estes foram sugeridos através de fatos do cotidiano dos estudantes.

3.1 Previsão do Tempo

Ao observar o mês de junho na cidade de Santo Antônio de Jesus na Bahia, a condição climática ao dia foi classificada como *Ensolarado, Chuvoso ou Nublado.* Assim, os dias deste mês podem ser classificados conforme a tabela abaixo:

Dia 1	Dia 2	Dia 3	Dia 4	Dia 5
Nublado	Ensolarado	Nublado	Nublado	Chuvoso
Dia 6	Dia 7	Dia 8	Dia 9	Dia 10
Ensolarado	Ensolarado	Nublado	Ensolarado	Nublado
Dia 11	Dia 12	Dia 13	Dia 14	Dia 15
Nublado	Chuvoso	Chuvoso	Nublado	Chuvoso
Dia 16	Dia 17	Dia 18	Dia 19	Dia 20
Chuvoso	Ensolarado	Ensolarado	Ensolarado	Nublado
Dia 21	Dia 22	Dia 23	Dia 24	Dia 25
Ensolarado	Ensolarado	Nublado	Nublado	Chuvoso
Dia 26	Dia 27	Dia 28	Dia 29	Dia 30
Ensolarado	Ensolarado	Nublado	Ensolarado	Nublado

Tabela 3.1: Condições climáticas de Santo Antônio de Jesus no mês de junho.

Note que, através da informações da tabela temos as seguintes probabilidades $P_{Ensolarado} = \frac{12}{30}, P_{Nublado} = \frac{12}{30}$ e $P_{Chuvoso} = \frac{6}{30}$. Desta forma, conseguimos corresponder o estado 1 ao dia *Ensolarado*, o estado 2 ao dia *Nublado* e ao estado 3 ao dia *Chuvoso*. Então, o vetor inicial é dado por $\pi^{(0)} = \begin{pmatrix} 0.4 & 0.4 & 0.2 \end{pmatrix}$.

Suponhamos que o Instituto Nacional de Pesquisas Espaciais - INPE, tenha ao longo de 10 anos de observação gerado uma matriz de transição do tempo para a cidade de Santo Antônio de Jesus com as seguintes entradas:

$$P = \begin{bmatrix} 0.3 & 0.4 & 0.3 \\ 0.1 & 0.4 & 0.5 \\ 0.5 & 0.4 & 0.1 \end{bmatrix}.$$

Através da matriz, se hoje temos um dia ensolarado há uma probabilidade de 40% de termos um dia nublado amanhã, ou ainda, se está nublado hoje temos 10% de o dia ser ensolarado amanhã. Como nossa variável é mensal, ao multiplicar o vetor de probabilidade inicial $\pi^{(0)}$ pela matriz de transição P teremos o vetor de probabilidade $\pi^{(1)}$ que representa a distribuição de probabilidades para o próximo mês e assim sucessivamente.

Pelo exemplo, podemos concluir que o tempo em uma cidade é um fenômeno que de fato independe do seu estado há um ano atrás, mas é influenciado pelo estado em que se encontra no dia anterior, ou seja, este é um fenômeno que possui a **propriedade Markoviana**. O outro ponto positivo é que a matriz de transição possui uma quantidade pequena de estados, possibilitando facilmente sua interpretação.

3.2 Previsões em Genética

De acordo com a genética clássica, algumas características das plantas e dos animais são determinadas por um par de genes, cada um dos quais podendo ser de dois tipos, denotados por A e a. Com isso, temos três genótipos possíveis: $AA, Aa\,(aA)$ e aa. Neste caso, o indivíduo se chama dominante(D) se tem o genótipo AA, heterozigoto (H) se tem o genótipo Aa e recessivo (R) se tem o genótipo aa. A previsão de genótipos em uma população é bastante útil, pois é usada na monitoração e o controle de algumas características da seleção artificial, sejam elas desejáveis ou não, como doenças genéticas por exemplo.

As Cadeias de Markov podem ser úteis nesse tipo de previsão. Considerando os estados como os genótipos D, H e R, e obtendo a probabilidade de cada genótipo para cruzamentos entre indivíduos $D \times D$, $D \times H$, $D \times R$, $H \times H$, $H \times R$ e $R \times R$. As probabilidades serão assim distribuídas:

Genótipo/Cruzamento	D(AA)	H(Aa)	R(aa)
D x D	1	0	0
D x H	0.5	0.5	0
D x R	0	1	0
H x H	0.25	0.5	0.25
H x R	0	0.5	0.5
R x R	0	0	1

Tabela 3.2: Tabela de genótipos.

Observando o cruzamento com indivíduos do genótipo H,temos

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

Ou seja, P é uma matriz de transição. Assim, é possível prever o comportamento genético para as futuras gerações. Suponhamos que, no processo de seleção artificial, a primeira geração a cruzar com o genótipo H seja formada exclusivamente por indivíduos dominantes (D), a previsão para o período de três gerações é dado por:

$$\pi^{(3)} = \pi^{(0)} \cdot P^3 = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \cdot \begin{bmatrix} 0.3125 & 0.5 & 0.1875 \\ 0.25 & 0.5 & 0.25 \\ 0.1875 & 0.5 & 0.3125 \end{bmatrix} = \begin{pmatrix} 0.3125 & 0.5 & 0.875 \end{pmatrix}.$$

3.3 Previsões Populacionais

Na região metropolitana de Salvador, a cada ano, três por cento da população de Dias Dávila migra para a cidade de Salvador, enquanto que apenas um por cento da população de Salvador migra para a cidade de Dias Dávila. Se todas as demais condições permanecerem estáveis, as condições políticas não mudaram, e estas porcentagens de migração continuam as mesmas, qual deve ser a relação entre as populações de Salvador e Dias Dávila ao longo do tempo?

Sendo três por cento da população de Dias Dávila migrando para Salvador, a probabilidade de migração de Dias Dávila para Salvador é de 0.03, já a probabilidade de não migração é de 0.97. Como um por cento da população de Salvador migra para Dias Dávila a probabilidade de migração de Salvador para Dias Dávila é de 0.01 e a de não migração é de 0.99. Denotando por S a cidade de Salvador e por D a cidade de Dias Dávila, temos a matriz das probabilidades de transição:

$$P = \begin{bmatrix} P_{S,D} & P_{S,S} \\ P_{D,S} & P_{D,D} \end{bmatrix}$$

$$P = \begin{bmatrix} 0.01 & 0.99 \\ 0.03 & 0.97 \end{bmatrix}$$

Logo, a longo prazo as probabilidades P_D, de morar em Dias Dávila, e P_S, de morar em Salvador, devem satisfazer

$$\begin{pmatrix} P_S & P_D \end{pmatrix} \cdot \begin{bmatrix} 0.01 & 0.99 \\ 0.03 & 0.97 \end{bmatrix} = \begin{pmatrix} P_S & P_D \end{pmatrix},$$

onde $99P_S = 3P_D$ e, como $P_S + P_D = 1$, temos $P_D = 0.25$ e $P_S = 0.75$. Portanto, a longo prazo, e se não houver modificações nas tendências de migração, teremos 25% da população morando em Dias Dávila e 75% da população habitando em Salvador.

3.4 Passeios Aleatórios

Quando temos o espaço de estados igual aos números inteiros, ou seja, $E = \mathbb{Z}$ então teremos um processo aleatório simples. As transições neste caso ocorrem entre estados vizinhos,

$$P_{i,i+1} = p = 1 - P_{i,i-1},\, com\, 0 \leq p < 1.$$

Se $p = 0$ as transições são somente para a esquerda e se $p = 1$ elas são somente para a direita.

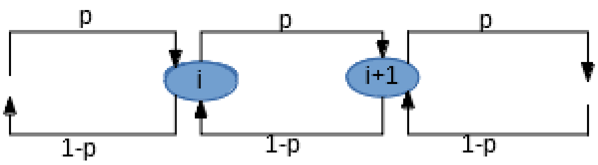

Figura 3.4.1: Topologia da cadeia para o passeio aleatório simples com infinitos estados.

Já no caso em que $p = \frac{1}{2}$ as transições são dadas por

$$P_{i,j} = \begin{cases} \frac{1}{2}, & j = i - 1\, ou\, j = i + 1, \\ 0, & caso\, contr\acute{a}rio. \end{cases}$$

A cadeia vai de um estado para o da esquerda ou para o da direita com a mesma probabilidade.

Analisemos o seguinte exemplo: Um homem está num ponto de coordenada inteira sobre o eixo x entre a origem e o ponto 4. Ele dá um passo de uma unidade para a direita com probabilidade p ou para a esquerda com probabilidade $q = 1 - p$, exceto quando ele estiver na origem, caso em que só pode dá um passo para a direita chegando ao ponto 1, ou quando estiver sobre o ponto 4, caso em que dá exclusivamente um passo para esquerda chegando ao ponto 3. Indiquemos por X_n a sua posição após n passos. Logo, o espaço de estados é dado por $E = \{0, 1, 2, 3, 4\}$, em que cada estado representa que o homem está sobre o eixo x no ponto de abscissa i, com $i = 0, 1, 2, 3, 4$. Assim, a matriz de transição é dada por

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

Cada coluna da matriz, exceto a primeira e a última, correspondem ao fato de que o homem se move de um estado presente para um estado futuro, com probabilidade p ou para o estado passado com probabilidade $q = 1 - p$. A primeira coluna corresponde ao fato de que o homem passa do estado inicial 1 sempre para o próximo estado 2 e a última coluna corresponde ao fato de que o homem sempre passa do estado 4 para o estado 3.

Suponha agora que quando o homem atinge um estado de fronteira ele permanece nesse estado, com probabilidade $\frac{1}{2}$ e se move para o outro estado de fronteira também com probabilidade $\frac{1}{2}$. Nesse caso, a matriz de transição é dada por

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

Capítulo 4

Sugestão de Sequência Didática

A sequência didática sugerida a seguir, possui 4 etapas, onde cada uma pode ter o tempo de até 4 horários de aula, resultando em um período de até 12 horários de aula. A proposta é proporcionar a investigação, a prática e o entrelaçamento dos conteúdos abordados anteriormente, sendo estes utilizados em atividades para alunos do segundo e terceiro ano do ensino médio da educação básica.

A divisão das quatro etapas será feita da seguinte forma:

- Primeira etapa: contecerá a introdução do tema e ,logo em seguida, a proposta de um exercício;

- Segunda etapa o professor apresentará as definições e as fórmulas

$$P_{i,j}^{(m)} = P(X_{m+n} = j \mid X_n = i), P(X_n = k) = \pi_0 P^n \, e \, P = \begin{bmatrix} P_{0,0} & P_{0,1} & \cdots & P_{0,N} \\ P_{1,0} & P_{1,1} & \cdots & P_{1,N} \\ \vdots & \vdots & \ddots & \vdots \\ P_{N,0} & P_{N,2} & \cdots & P_{N,N} \end{bmatrix},$$

  que possibilitarão a resolução dos exercícios propostos;

- Terceira etapa: o professor mostrará aos alunos o software matemático *Maxima,* com a finalidade de simplificar as operações entre matriz e vetor e com potência de matrizes;

- Quarta etapa: o professor discutirá com os alunos os resultados encontrados e apresentará mais problemas que possam a vim ser trabalhados em sala de aula.

A proposta deste trabalho é mostrar ao professor a utilização de conteúdos da educação básica para modelagem de problemas que refletem o cotidiano dos alunos. Ou seja, pretende-se discutir e analisar os resultados de inquietações de estudantes sobre a importância do estudo de Matrizes e Probabilidade. Para isso, sugere-se ao professor seguir cada etapa através dos planos de aula abaixo.

4.1 Primeira Etapa

PLANO DE AULA 1

Público alvo: 2^{o} e 3^{o} ano do ensino médio
Temática da aula: Cadeias de Markov
Conteúdo (s):

1. Cadeia de Markov;
2. Matriz de Transição;
3. Diagrama de Transção.

Objetivo:

1. Apresentar as Cadeias de Markov;
2. Modelar um problema prático através de uma representação matricial que é justamente a matriz de transição;
3. Analisar o problema possibilitando a formalização do que ocorre através de um diagrama de transição.

Número de aulas: 3
Material necessário:

- Quadro branco;
- Piloto;
- Folha de papel com a atividade proposta.

Desenvolvimento:

Primeiro momento: O professor deve introduzir o conceito de Cadeias de Markov, que poderá ser feito no quadro branco;

Segundo momento: Exemplificar uma situação simples de Cadeia de Markov. A sugestão é o exemplo 2.3 apresentado no capítulo 2;

Terceiro momento: Analisar o exemplo **2.3** juntamente com os alunos e construirá a matriz de transição e uma ilustração geométrica da situação que será o diagrama de transição;

Quarto momento: O professor irá entregar a cada aluno o exercício proposto abaixo e deverá esperar o tempo de no máximo 25 minutos para começar a discussão os resultados encontrados pelos estudantes.

Exercicío Proposto

Em uma família, existe um levantamento em relação a classe social de cada geração. Ou seja, cada geração pode estar na classe baixa (1), na classe média (2) ou na classe alta (3). Sobre a possibilidade de mudança de classes temos a seguinte matriz com suas probabilidades.

$$\begin{bmatrix} 0.3 & 0.3 & 0.4 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}$$

a) Faça uma interpretação geométrica da situação (diagrama de transição).

b) Se a família está na primeira geração e encontra-se na classe média, qual a probabilidade da família se tornar classe alta, quando a pesquisa for realizada na terceira geração?

Resolução:

a) Faça uma interpretação geométrica da situação (diagrama de transição).

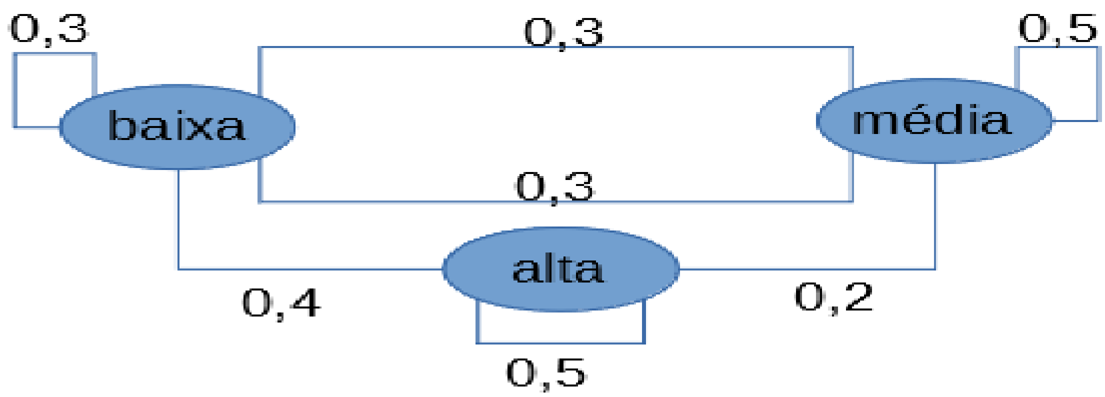

Figura 4.1.1: Grafo de transições entre classes.

b) Se a família está na primeira geração e encontra-se na classe média, qual a probabilidade da família se tornar classe alta, quando a pesquisa for feita na terceira geração?

O aluno ao responder este tópico está sendo induzido a pensar de maneira investigativa, ou seja, sabendo que a família encontra-se na classe média a probabilidade para a próxima geração é dada por:

classe 1	classe 2	classe 3
0.3	0.5	0.2

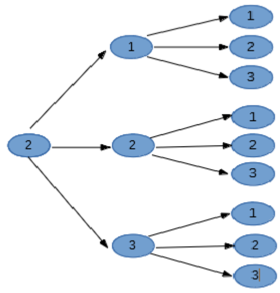

Figura 4.1.2: Transições com estado inicial 2

Desta forma, a probabilidade da família na terceira geração ser da classe 3 é dada por

$P(MÉDIA\, e\, BAIXA\, e\, ALTA) = 1 \times 0.3 \times 0.4 = 0.12$
$P(MÉDIA\, e\, MÉDIA\, e\, ALTA) = 1 \times 0.5 \times 0.2 = 0.10$
$P(MÉDIA\, e\, ALTA\, e\, ALTA) = 1 \times 0.2 \times 0.5 = 0.10$

Logo, a probabilidade da terceira geração ser da classe alta será de $0.12 + 0.10 + 0.10 = 0.32$. Ou seja, 32%.

O professor deve enfatizar que foi utilizado a probabilidade condicional para chegar ao resultado e deverá deixar claro que para uma futura geração nem sempre os cálculos serão rápidos. Por exemplo, a probabilidade agora da décima quinta geração ser da classe alta. Os cálculos a mão ficam quase impossível para a proposta de tempo da aula. Desta forma, o aluno começa a perceber a necessidade da teoria matemática para ser utilizada como uma ferramenta na resolução de muitos problemas.

4.2 Segunda Etapa

PLANO DE AULA 2

Público alvo: 2º e 3ºano do ensino médio
Temática da aula: Cadeias de Markov
Conteúdo (s):

1. Cadeias de Markov Homogêneas a tempo discreto;
2. Matriz de Transição;
3. Vetor de Probabilidade Inicial.

Objetivo:

1. Apresentar as fórmulas para facilitar os cálculos do exercício proposto na aula anterior e de grande iterações;
2. Aplicar os conceitos de Cadeias de Markov Homogênea a tempo discreto para resolver o próximo exercício proposto.

Número de aulas: 3

Material necessário:

- Quadro branco;
- Piloto;
- Folha de papel com a atividade proposta.

Desenvolvimento:

Primeiro momento: O professor irá introduzir uma noção de Cadeias de Markov Homogêneas a tempo discreto;

Segundo momento: Apresentar as fórmulas e construir o vetor inicial da situação colocada no exercício proposto da última aula e resolvê-lo no quadro junto com os estudantes;

Terceiro momento: O professor irá aplicar uma segunda atividade direcionada ao conteúdo previamente apresentado pelo professor.

Na primeira parte da aula, o professor irá falar de forma exemplificada sobre Cadeias de Markov a tempo discreto. Em seguida, deverá apresentar algumas fórmulas que ajudarão a resolver uma inquietação da aula anterior.

$$P(X_n = k) = \pi_0^t P^n \text{ e } P_{i,j}^{(m)} = P(X_{m+n} = j \mid X_n = i)$$

Exercício Proposto (última aula)

Em uma família, existe um levantamento em relação a classe social de cada geração. Ou seja, cada geração pode estar em uma classe baixa (1), classe média (2) ou classe alta (3). Sobre a possibilidade de mudança de classes temos a seguinte matriz com suas probabilidades.

$$\begin{bmatrix} 0.3 & 0.3 & 0.4 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}$$

b) Desta forma, se a família está na primeira geração e encontra-se na classe média, qual a probabilidade da família se tornar classe alta, quando a pesquisa for realizada na terceira geração?

Resolução

Sabendo que a primeira geração está na classe média então o vetor inicial é $\pi^{(0)} = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}$. Em seguida, o professor irá calcular o vetor $\pi^{(1)}$ e depois o $\pi^{(2)}$ e mostrar que é o mesmo resultado ao fazer:

$$\pi^{(2)} = \pi^{(0)} \cdot P^2$$

$$= \begin{pmatrix} 0 & 1 & 0 \end{pmatrix} \cdot \begin{bmatrix} 0.26 & 0.36 & 0.38 \\ 0.28 & 0.4 & 0.32 \\ 0.25 & 0.36 & 0.39 \end{bmatrix}$$

$$= \begin{pmatrix} 0.28 & 0.4 & 0.32 \end{pmatrix}.$$

Portanto, a probabilidade é de 0.32 da terceira geração ser de classe alta, ou seja, 32%.

Exercício Proposto

Na cidade de Amargosa a chance de fazer sol ou chuva amanhã depende somente das condições climáticas de hoje. Através de observações realizadas previamente, tem-se que a probabilidade de fazer sol no próximo dia, sendo que hoje foi ensolarado, é de 75% e 25% de chover, mas a probabilidade de fazer sol no próximo dia sendo que hoje foi um dia chuvoso é de 50%.

De acordo com o conteúdo visto em sala de aula, responda os seguintes tópicos:

a) Determine a matriz de transição relacionada a situação.

b) Se o estado incial for um dia ensolarado, qual a previsão dos próximos dois dias para a cidade de Amargosa.

c) Se o estado inicial for o vetor $\begin{pmatrix} 0.25 & 0.75 \end{pmatrix}$, verifique a previsão do tempo para os próximos três dias em Amargosa.

Resolução

a) Determine a matriz de transição relacionada a situação.

Neste caso, os alunos deverão apresentar como solução a seguinte matriz $\begin{bmatrix} 0.75 & 0.25 \\ 0.5 & 0.5 \end{bmatrix}$.

b) Se o estado incial for um dia ensolarado, qual a previsão dos próximos dois dias para a cidade de Amargosa.

O aluno irá determinar o vetor inicial e fazer uso das fórmulas apresentadas pelo professor anteriormente:

$$\pi^{(0)} = \begin{pmatrix} 1 & 0 \end{pmatrix}.$$

$$\pi^{(1)} = \pi^{(0)} \cdot P = \begin{pmatrix} 1 & 0 \end{pmatrix} \cdot \begin{bmatrix} 0.75 & 0.25 \\ 0.5 & 0.5 \end{bmatrix} = \begin{pmatrix} 0.75 & 0.25 \end{pmatrix}.$$

$$\pi^{(2)} = \pi^{(1)} \cdot P = \begin{pmatrix} 0.75 & 0.25 \end{pmatrix} \cdot \begin{bmatrix} 0.75 & 0.25 \\ 0.5 & 0.5 \end{bmatrix} = \begin{pmatrix} 0.6875 & 0.3125 \end{pmatrix}.$$

c) Se o estado inicial for o vetor $\begin{pmatrix} 0,25 & 0,75 \end{pmatrix}$, verifique a previsão do tempo para os próximos três dias em Amargosa.

Agora o aluno irá determinar um novo vetor inicial e novamente fazer uso das fórmulas apresentadas pelo professor anteriormente.

$$\pi^{(0)} = \begin{pmatrix} 0.25 & 0.75 \end{pmatrix}.$$

$$\pi^{(1)} = \pi^{(0)} \cdot P = \begin{pmatrix} 0.25 & 0.75 \end{pmatrix} \cdot \begin{bmatrix} 0.75 & 0.25 \\ 0.5 & 0.5 \end{bmatrix} = \begin{pmatrix} 0.5625 & 0.4375 \end{pmatrix}.$$

$$\pi^{(2)} = \pi^{(1)} \cdot P = \begin{pmatrix} 0.5625 & 0.4375 \end{pmatrix} \cdot \begin{bmatrix} 0.75 & 0.25 \\ 0.5 & 0.5 \end{bmatrix} = \begin{pmatrix} 0.646025 & 0.359375 \end{pmatrix}.$$

$$\pi^{(3)} = \pi^{(2)} \cdot P = \begin{pmatrix} 0.646025 & 0.359375 \end{pmatrix} \cdot \begin{bmatrix} 0.75 & 0.25 \\ 0.5 & 0.5 \end{bmatrix} = \begin{pmatrix} 0.66015625 & 0.33984375 \end{pmatrix}$$

4.3 Terceira Etapa

PLANO DE AULA 3

Público alvo: 2º e 3ºano do ensino médio

Temática da aula: MAXIMA

Conteúdo (s):

1. Cadeias de Markov Homogêneas a tempo discreto;

2. Software matemático MAXIMA.

Objetivo:

1. Apresentar o software matemático MAXIMA com suas principais funções para operações com Matrizes.

2. Utilizar o MAXIMA para os cálculos de envolvendo grandes potências de matrizes.

Número de aulas: 4

Material necessário:

- Quadro branco;
- Piloto;
- Computadores;
- Notebooks;
- Data-show.

Desenvolvimento:

Primeiro momento: O professor apresentará o software MAXIMA através de um vídeo tutorial;

Segundo momento: O professor irá junto aos estudantes interagir com o software na aprendizagem dos comandos básicos;

Terceiro momento: Os estudantes deverão fazer uso do MAXIMA para realizar operações com matrizes;

Quarto momento: O exercício proposto na primeira aula deverá ser refeito e com uma maior quantidade de iterações através dos cálculos realizados no MAXIMA.

Como visto anteriormente em aula, a partir de grandes iterações fica inviável realizar os cálculos a mão. Por isso, é necessário a utilização de ferramentas que auxiliem o aluno nesta tarefa.

Hoje, podemos fazer uso de software matemáticos gratuitos e de fácil utilização. O MAXIMA é um software neste formato, ou seja, livre e de fácil utilização para a educação básica. Desta forma, o professor deverá seguir o seguinte roteiro para sua utilização:

1. Baixar a versão correspondente a plataforma utilizada no ambiente escolar através do site http://maxima.sourceforge.net/pt/download.html;

2. Seguir os passos da instalação e ler atentamente as informações;

3. Ao terminar a instalação abrir o programa para verificar se aparece alguma mensagem de erro, caso não apareça, parabéns a instalação foi concluída com sucesso.

Agora, o professor apresentará a turma o software, vale lembrar que os estudantes devem acompanhar passo a passo as orientações no data-show e nos respectivos notebooks ou computadores.

A primeira parte da apresentação o professor irá abrir o seguinte link:

https://www.youtube.com/watch?v=VQq9yMiebvU.

Neste vídeo, será abordado funções básicas e que serão relevantes depois na operação entre matrizes. No segundo momento, cada dupla de alunos fará uso de um computador ou notebook e fará os seguintes passos:

- Abrir o programa MAXIMA;

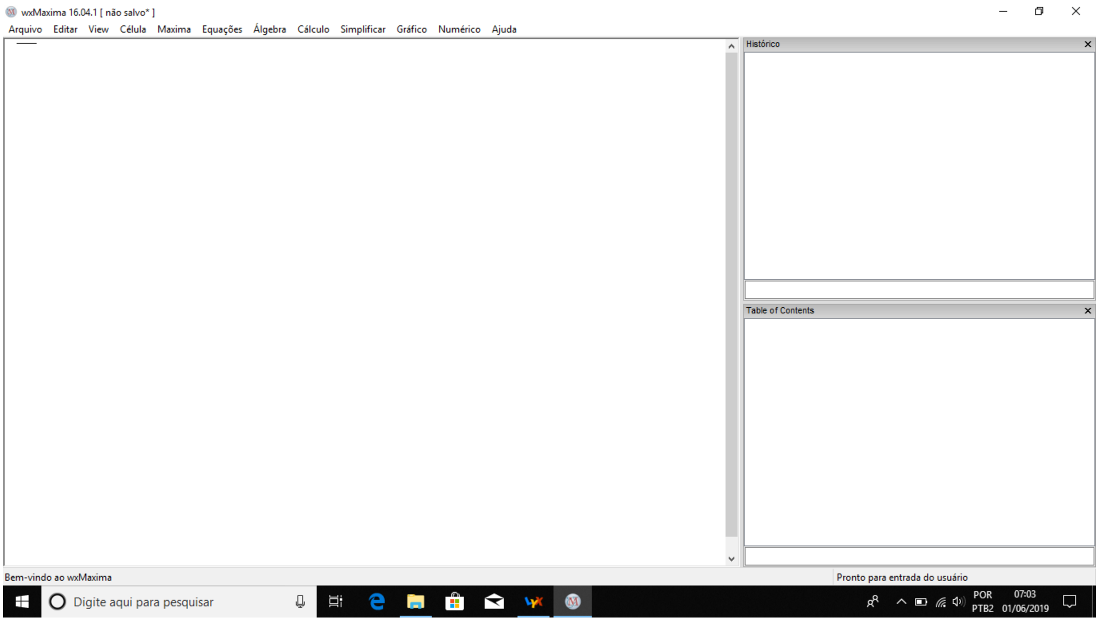

Figura 4.3.1: Página Inicial do wxMaxima.

- Na barra de comandos clique em Álgebra e depois em Introduzir matriz;

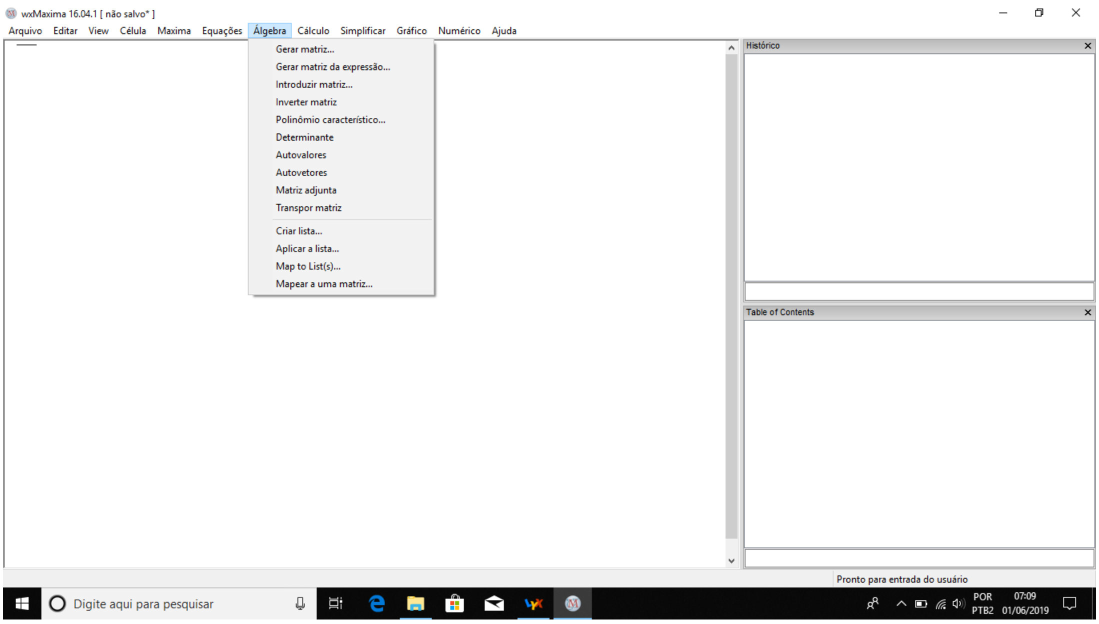

Figura 4.3.2: Comando para introduzir matriz.

- Preencher o tipo e o nome da matriz;

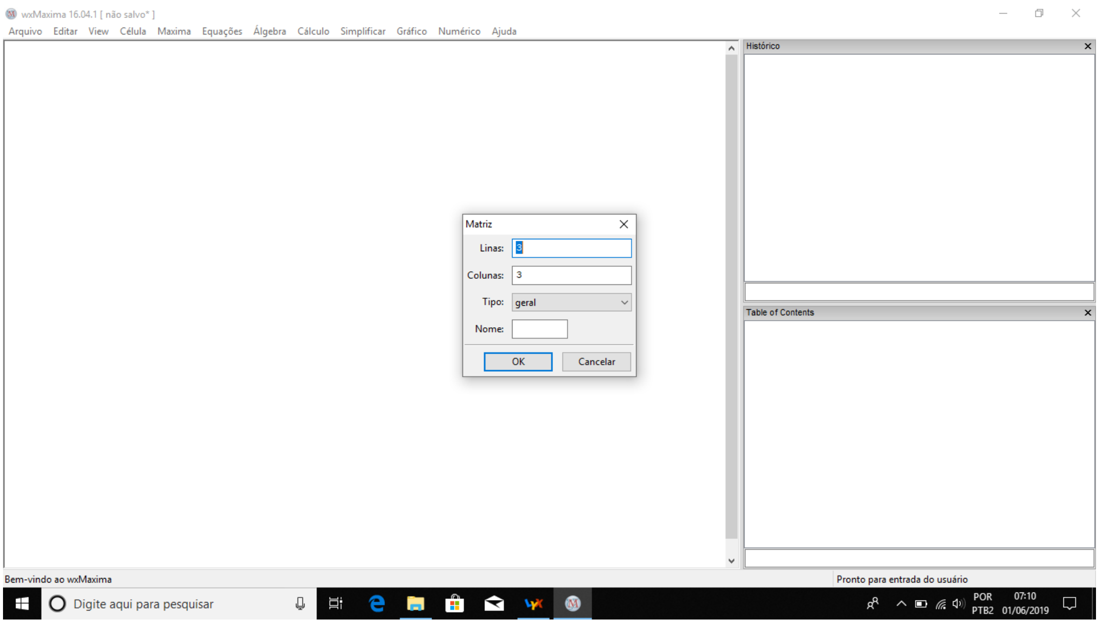

Figura 4.3.3: Caixa da matriz.

- Ao clicar em OK no passo anterior, abrirá uma nova caixa para preencher com os elementos da matriz;

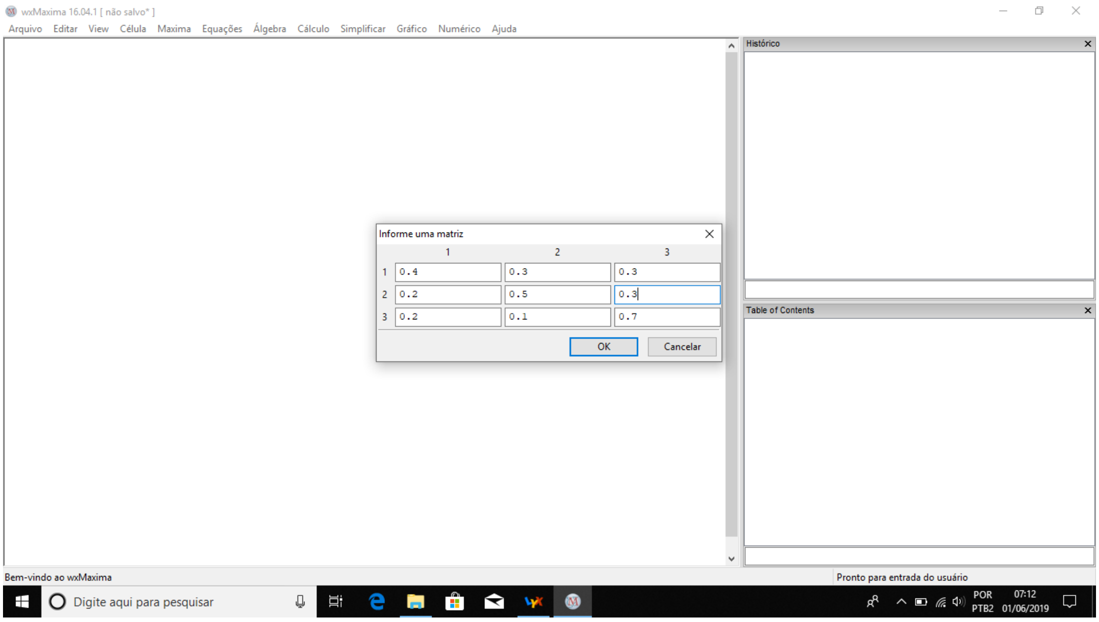

Figura 4.3.4: Caixa com os elementos da matriz.

- Depois de preencher a caixa com os elementos da matriz, clique em OK e o programa irá gerar a matriz;

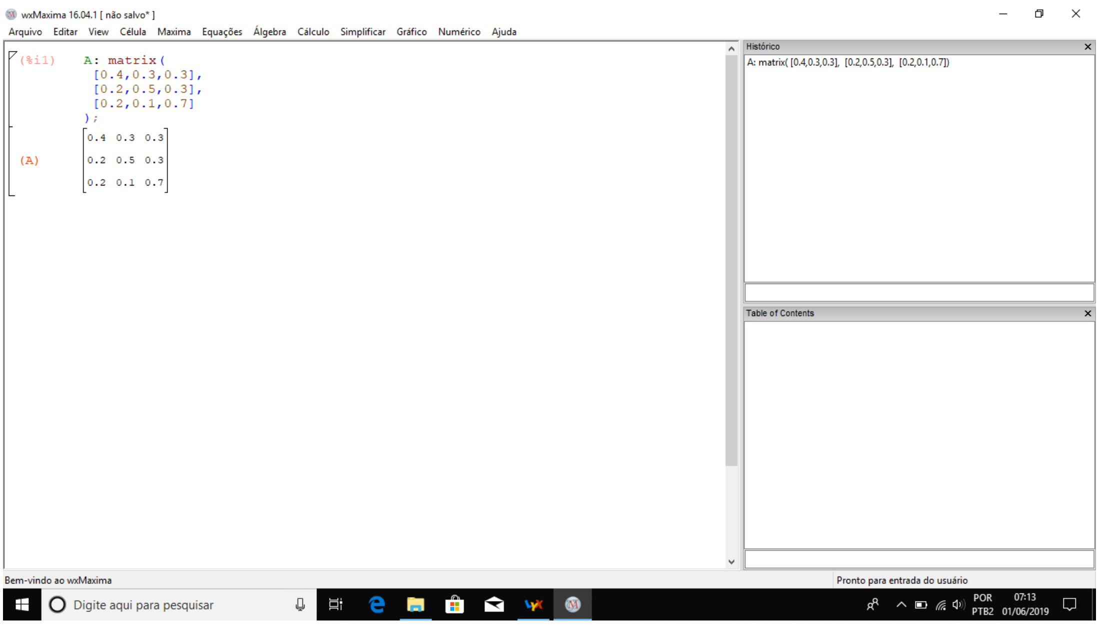

Figura 4.3.5: Matriz gerada pela caixa de elementos.

- Para realizar a operação de adição de matrizes o aluno deverá seguir o roteiro anterior e criar uma nova matriz (B) do mesmo tipo, em seguida, digitar na janela do MAXIMA os comandos: $A + B$ depois *control* mais *enter;*

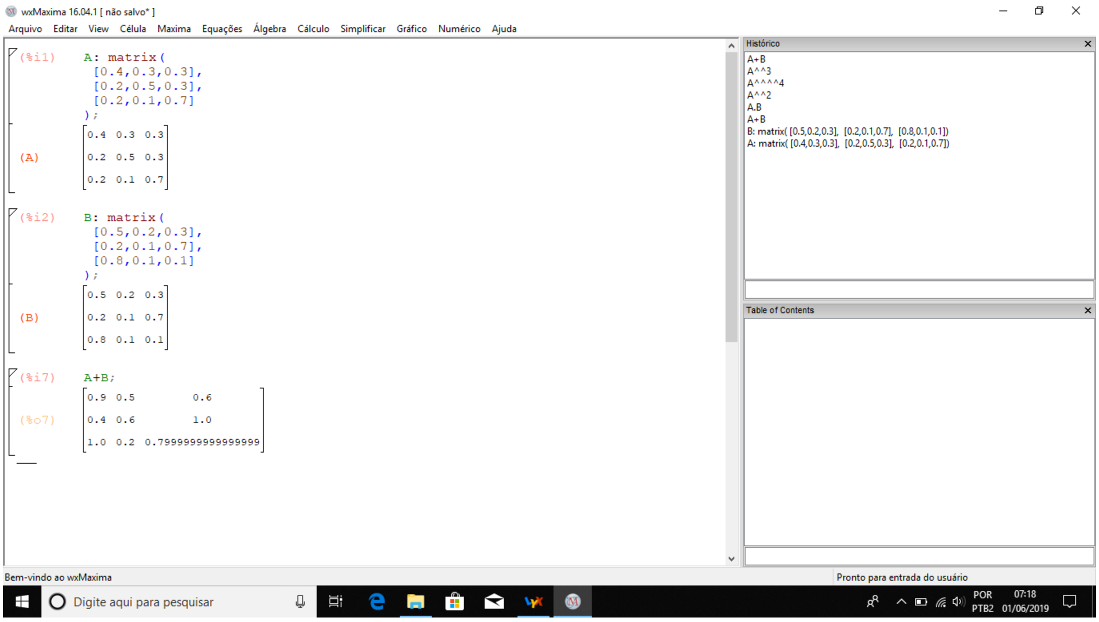

Figura 4.3.6: Adição de matrizes.

- Na operação de multiplicação entre matrizes, o aluno deverá digitar $A.B$ depois *control* mais *enter*;

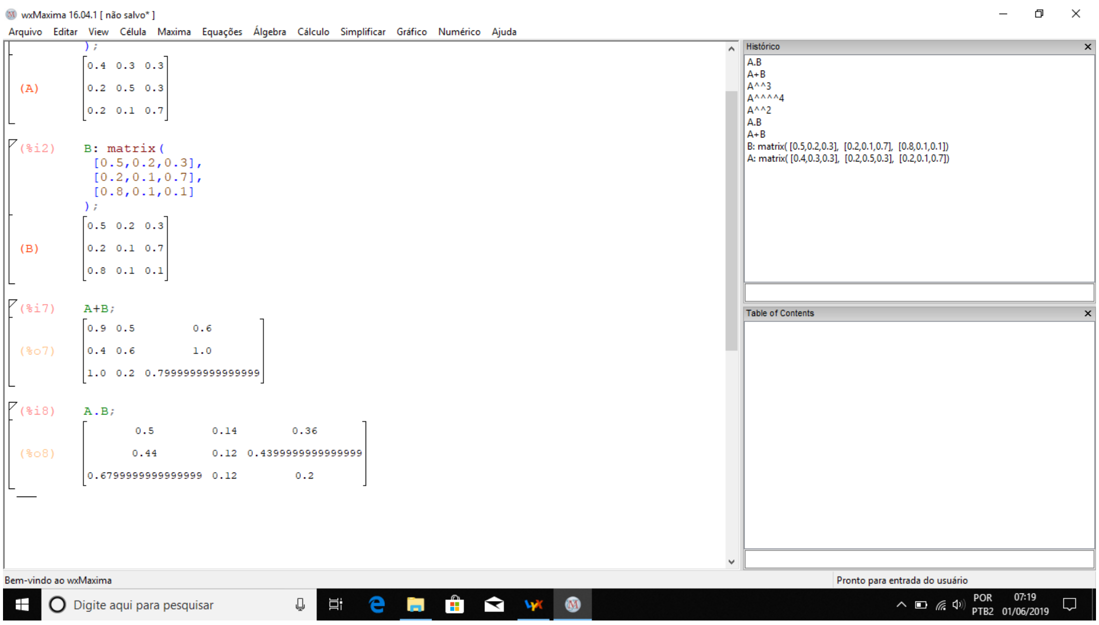

Figura 4.3.7: Multiplicação de matrizes.

- Agora o aluno irá realizar a potência de matriz, para isso ele deve utilizar os seguintes comandos: A^^2 depois *control* mais *enter*. Nesta operação a matriz A foi multiplicada por ela mesma, pois estava elevado ao quadrado;

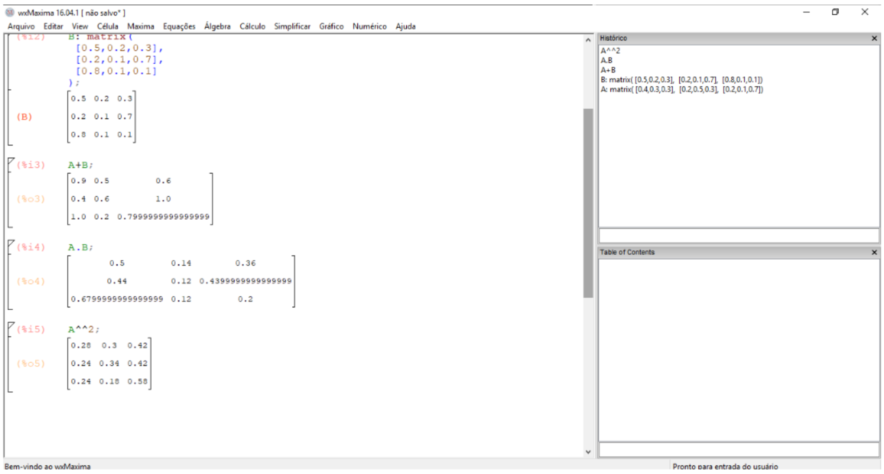

Figura 4.3.8: Potência de matriz.

Neste momento da aula, o professor irá retomar o primeiro exercício proposto.

Exercicío Proposto

Em uma família, existe um levantamento em relação a classe social de cada geração. Ou seja, cada geração pode estar em uma classe baixa (1), classe média (2) ou classe alta (3). Sobre a possibilidade de mudança de classes temos a seguinte matriz com suas probabilidades.

$$\begin{bmatrix} 0.3 & 0.3 & 0.4 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}$$

Desta forma, se a família está na primeira geração e encontra-se na classe média, qual a probabilidade da família se tornar classe alta, quando a pesquisa for realizada na décima quinta geração?

Resolução

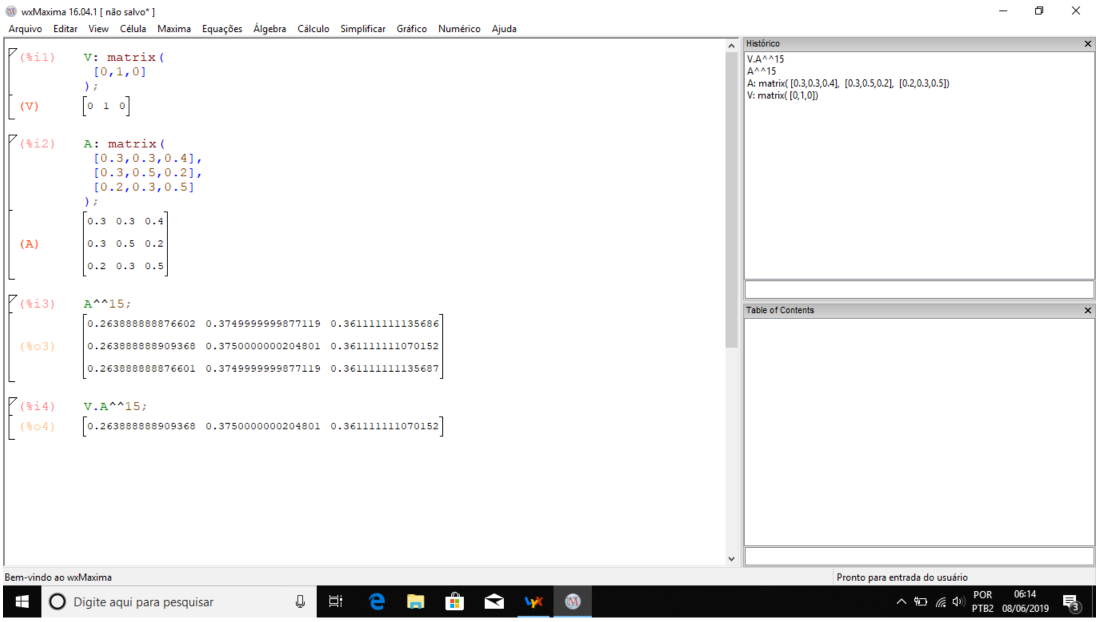

Figura 4.3.9: Solução para quinze iterações.

Logo, $36,11\%$.

4.4 Quarta Etapa

PLANO DE AULA 4

Público alvo: 2º e 3ºano do ensino médio

Temática da aula: MAXIMA

Conteúdo (s):

1. Cadeias de Markov Homogêneas a tempo discreto.

Objetivo:

1. Avaliar a participação dos estudantes com o conteúdo proposto;
2. Discutir os resultados encontrados com uma maior quantidade de iterações;
3. Sugerir aos alunos outros problemas que poderão ser modelados com o mesmo conteúdo.

Número de aulas: 2

Material necessário:

- Quadro branco;
- Piloto;

- Notebooks;
- Data-show.

Desenvolvimento:

Primeiro momento: O professor irá avaliar a participação dos alunos através de todo o processo realizado, ou seja, é possível atribuir uma pontuação ao estudante de acordo com o trabalho realizado por cada um deles no decorrer das aulas. Por exemplo: pontuação para representação matricial do problema, representação por diagrama do problema, interpretação sobre o vetor inicial, cálculos realizados manualmente e cálculos de grandes iterações realizados com ajuda do software;

Segundo momento: O professor irá deixar claro a necessidade da utilização do software MAXIMA para o cálculo de grandes iterações e observar o que poderia ter sido encontrado no caso de cada instante do tempo (geração). Já que os exercícios propostos abordam apenas Cadeias de Markov a tempo discreto;

Terceiro momento: Deixar uma atividade para que os alunos possam realizar em outro momento.

ATIVIDADE

O café tem uma colheita anual classificada como boa, média ou ruim. Na região do Vale do Jiquiriça na Bahia, observou-se que após uma colheita ruim, existe uma probabilidade de $0,7$ e $0,2$ de a colheita no ano seguinte ser classificada como média ou boa, respectivamente. Além disso, também foi observado que após uma colheita média, existe a probabilidade de $0,3$ e $0,2$ de a próxima ser classificada como boa ou ruim, respectivamente. E após uma colheita boa, há probabilidade de $0,5$ e $0,1$ de a próxima colheita ser classificada como média ou ruim, respectivamente.

Sabendo que a colheita do ano posterior é influenciada somente pela colheita do ano anterior a ela, responda:

a) Represente através do diagrama de transição.

b) Monte a matriz de transição P.

c) Em 4 anos, qual é a probabilidade de uma colheita vir a ser classificada como boa, dado que a colheita atual é ruim?

d) Em 10 anos, qual a probabilidade de uma colheita vir a ser classificada como média, dado que a colheita atual é média?

Resolução

a) Represente através do diagrama de transição.

b) Monte a matriz de transição P.

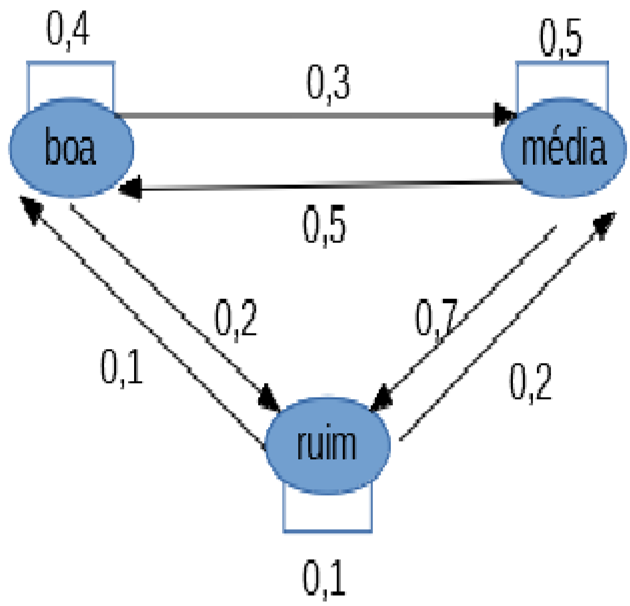

Figura 4.4.1: Diagrama de transição.

$$P = \begin{bmatrix} 0.4 & 0.5 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.7 & 0.1 \end{bmatrix}.$$

c) Em 4 anos, qual é a probabilidade de uma colheita vir a ser classificada como boa, dado que a colheita atual é ruim?

Se a colheita atual é ruim, então $\pi^{(0)} = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}$. Daí,

$\pi^{(4)} = \pi^{(0)} \cdot P^4 = \begin{pmatrix} 0.31620 & 0.53040 & 0.15340 \end{pmatrix}$.

Logo, $31,62\%$ é a probabilidade da colheita ser boa.

d) Em 10 anos, qual a probabilidade de uma colheita vir a ser classificada como média, dado que a colheita atual é média?

Se a colheita atual é média, então $\pi^{(0)} = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}$.Daí,

$\pi^{(10)} = \pi^{(0)} \cdot P^{10} = \begin{pmatrix} 0.31633 & 0.53061 & 0.15306 \end{pmatrix}$.

Logo, $53,061\%$ é a probabilidade da colheita ser média.

Referências Bibliográficas

[1] J. P. d. A. Albuquerque, J. M. P. Fortes, and W. A. Finamore. *Probabilidade, variáveis aleatórias e processos estocásticos*. Editora PUC–Rio e Editora Interciência, Rio de Janeiro, Brasil, 2008.

[2] J. L. Boldrini, S. I. Costa, V. Figueredo, and H. G. Wetzler. *Álgebra linear*. Harper & Row, 1980.

[3] K. L. Chung. *Elementary probability theory with stochastic processes*. Springer Science & Business Media, 2012.

[4] A. Golmakani, A. A. Silva, E. M. S. Freire, M. K. Barbosa, P. H. G. Carvalho, and V. L. Alves. Cadeias de markov.

[5] A. Hinojosa and A. Milanés. Uma introdução aos processos estocásticos com aplicações, departamento de estatística. *ICEx. UFMG*.

[6] E. L. Lima. *Curso de Análise, vol. 2*. Instituto de Matemática Pura e Aplicada, 1976.

[7] E. L. Lima, P. C. P. Carvalho, E. Wagner, and A. Morgado. *A matemática do ensino médio*, volume 2 of *COLEÇÃO DO PROFESSOR DE MATEMÁTICA*. Instituto de Matemática Pura e Aplicada, 2016.

[8] E. L. Lima, P. C. P. Carvalho, E. Wagner, and A. Morgado. *A matemática do ensino médio*, volume 3 of *COLEÇÃO DO PROFESSOR DE MATEMÁTICA*. Instituto de Matemática Pura e Aplicada, 2016.

[9] M. N. Magalhães. *Probabilidade e variáveis aleatórias*. Edusp, 2006.

[10] G. Modica and L. Poggiolini. *A first course in probability and Markov Chains*. Wiley Online Library, 2013.

[11] P. A. Morettin and W. O. BUSSAB. *Estatística básica*. Editora Saraiva, 2017.

[12] A. C. d. O. Morgado, J. B. P. d. Carvalho, P. C. P. Carvalho, and P. Fernandez. *Análise combinatória e probabilidade.* Instituto de Matemática Pura e Aplicada, 1991.

[13] J. R. Norris. *Markov chains.* Number 2. Cambridge university press, 1998.

[14] J. I. D. Pinheiro, S. B. d. Cunha, S. R. Carvajal, and G. C. Gomes. *Estatística básica: a arte de trabalhar com dados.* Elsevier: Campus, 2009.

[15] N. Privault. *Understanding Markov Chains: Examples and Applications.* Springer, 2013.

Universidade Federal do Recôncavo da Bahia - UFRB
Centro de Ciências Exatas e Tecnológicas / Mestrado Profissional em Matemática em Rede Nacional

Rua Rui Barbosa, s/n, Campus Universitário de Cruz das Almas - BA
CEP: 44380-000
<http://www.ufrb.edu.br/profmat>
<http://www.profmat-sbm.org.br>