Sérgio da Silva Medeiros

Cadeias de Markov Ocultas

Santo André, 2017

UNIVERSIDADE FEDERAL DO ABC

CENTRO DE MATEMÁTICA, COMPUTAÇÃO E COGNIÇÃO

SÉRGIO DA SILVA MEDEIROS
Sérgio da Silva Medeiros

CADEIAS DE MARKOV OCULTAS

Orientador: Prof. Dr. Daniel Miranda Machado

Dissertação de mestrado apresentada ao Centro de Matemática, Computação e Cognição para obtenção do título de Mestre

ESTE EXEMPLAR CORRESPONDE A VERSÃO FINAL DA DISSERTAÇÃO
DEFENDIDA PELO ALUNO SÉRGIO DA SILVA MEDEIROS,
E ORIENTADA PELO PROF. DR. DANIEL MIRANDA MACHADO.

SANTO ANDRÉ, 2017

Sistema de Bibliotecas da Universidade Federal do ABC
Elaborada pelo Sistema de Geração de Ficha Catalográfica da UFABC
com os dados fornecidos pelo(a) autor(a).

da Silva Medeiros, Sérgio
CADEIAS DE MARKOV OCULTAS / Sérgio da Silva Medeiros. — 2017.

65 fls. : il.

Orientador: Daniel Miranda Machado

Dissertação (Mestrado) — Universidade Federal do ABC, Mestrado Profissional em Matemática em Rede Nacional - PROFMAT, Santo André, 2017.

1. Probabilidade. 2. Matrizes. 3. Cadeias de Markov. I. Miranda Machado, Daniel. II. Mestrado Profissional em Matemática em Rede Nacional - PROFMAT, 2017. III. Título.

Este exemplar foi revisado e alterado em relação à versão original, de acordo com as observações levantadas pela banca no dia da defesa, sob responsabilidade única do autor e com a anuência de seu orientador.

Santo André, 04 de Outubro de 2017.

Assinatura do autor: ______________________

Assinatura do orientador: ______________________

MINISTÉRIO DA EDUCAÇÃO
Fundação Universidade Federal do ABC
Programa de Pós-Graduação em Mestrado Profissional em Matemática em Rede Nacional
Avenida dos Estados, 5001 – Bairro Santa Terezinha – Santo André – SP
CEP 09210-580 · Fone: (11) 4996-0017
profmat@ufabc.edu.br

FOLHA DE ASSINATURAS

Assinaturas dos membros da Banca Examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado do candidato Sérgio da Silva Medeiros, realizada em 16 de agosto de 2017:

Daniel M. Machado

Prof.(a) Dr.(a) **Daniel Miranda Machado** (Universidade Federal do ABC) – Presidente

---

Prof.(a) Dr.(a) **Rodney Carlos Bassanezi** (Universidade Estadual de Campinas) – Membro Titular

Coletti; Cristian Favio.

Prof.(a) Dr.(a) **Cristian Favio Coletti** (Universidade Federal do ABC) – Membro Titular

---

Prof.(a) Dr.(a) **Vinicius Cifú Lopes** (Universidade Federal do ABC) – Membro Suplente

---

Prof.(a) Dr.(a) **Pedro Jose Catuogno** (Universidade Estadual de Campinas) – Membro Suplente

Agradecimentos

A Deus.
À minha família, em especial minha esposa Silvana e minha filha Manuela.
À CAPES, pelo apoio financeiro.
Ao coordenador professor Dr. Rafael Grisi e docentes do PROFMAT da UFABC.
Aos colegas de turma.
Em especial, ao professor Dr. Daniel Miranda Machado, por sempre ter sido solícito desde suas aulas de recursos computacionais, como também por sua dedicação às orientações, tornando-as grandes momentos de aprendizagem.

Resumo

O foco principal deste trabalho é o estudo das Cadeias de Markov e das Cadeias de Markov Ocultas. As cadeias de Markov fornecem uma forma prática para o estudo de conceitos probabilísticos e matriciais.
Procuramos utilizar de forma contextualizada a aplicação do produto e potência de matrizes associados ao *software* Geogebra.
Além dos exemplos, estão contidas questões de aprendizagem, sempre com objetivo de torná-los aliados e valiosos ao aprendizado referente a este tema.

Palavras-chave: Cadeias de Markov Ocultas, Cadeias de Markov, Probabilidades

Abstract

The main focus of this work is the study of Markov Chains and the Markov Hidden Chains, which in turn brings the study of probabilistic and matrix concepts into practice.
We seek to use in a contextualized way the application of the multiplication and potency of matrices associated to the software Geogebra.
In addition to the examples, are contained learning issues, always with the goal of making them allies and valuable to the learning related to this theme.

Keywords:Markov Hidden Chains, Markov Chains, Probabilities

Conteúdo

Lista de Figuras xi

Lista de Tabelas xiii

1 Probabilidades 3
- 1.1 Introdução aos conjuntos . . . 3
  - 1.1.1 União e Interseção de Conjuntos . . . 5
  - 1.1.2 Complementar de um conjunto . . . 6
- 1.2 Análise Combinatória . . . 7
  - 1.2.1 Arranjos, Permutações e Combinações . . . 8
- 1.3 Introdução às Probabilidades . . . 9
  - 1.3.1 Experimento Aleatório . . . 9
  - 1.3.2 Espaço Amostral e Eventos . . . 10
- 1.4 Axiomas de Probabilidade . . . 12
- 1.5 Probabilidades em espaços amostrais equiprováveis . . . 13

2 Probabilidade Condicional 15
- 2.1 Probabilidade Condicional e Eventos Independentes . . . 15
- 2.2 Probabilidade Total . . . 17
- 2.3 Teorema de Bayes . . . 20

3 Cadeias de Markov 23
- 3.1 Introdução . . . 23
- 3.2 Diagrama de Transição de Estado . . . 26
- 3.3 Vetor Probabilidade . . . 27
- 3.4 Matrizes . . . 28
- 3.5 Matriz de Transição . . . 29
- 3.6 Equações de Chapman-Kolmogorov . . . 31
- 3.7 Classificação dos estados em Cadeias de Markov . . . 32
- 3.8 Probabilidades limite . . . 34
- 3.9 Problema da Ruína do Jogador . . . 36

4 Cadeias de Markov Ocultas 41
4.1 Abordando um exemplo 41
4.2 Notação 44
4.3 Os três problemas fundamentais 46
4.4 Problema 1 47
4.5 Problema 2 47
4.6 Problema 3 47
4.7 Discussão 47
4.8 As três soluções 48
4.8.1 Solução para o problema 1 48
4.8.2 Solução para o Problema 2 49
4.8.3 Solução para o Problema 3 50

5 Atividades para salas de aulas 53
5.1 Geogebra 53
5.2 Atividade 1 57
5.3 Atividade 2 58
5.4 Atividade 3 59

Índice 63

Lista de Figuras

2.1 Três urnas 18
2.2 Moeda de Bertrand 19

3.1 Apresentação do Sistema 24
3.2 Árvore característica 24
3.3 Árvore característica 25
3.4 Estímulo de Skinner 27
3.5 Diagrama de transição do modelo experimental de Skinner 27
3.6 Diagrama de estados da ruína do jogador 39

4.1 Anéis de crescimento de árvores [5] 43
4.2 Modelo Oculto de Markov 44

5.2 Janela principal do Geogebra 54
5.1 Divisões dos comandos e áreas de trabalho no Geogebra 54
5.3 Confecção da Matriz M no Geogebra 54
5.4 Confecção da Matriz N no Geogebra 55
5.5 Produto de matrizes M e N 55
5.6 Quadrado da Matriz M 56
5.7 Cubo da Matriz M 56
5.8 Diagrama de transição entre os estados do problema 1 58
5.9 Diagrama de transição entre os estados do problema 2 58
5.10 Agente de trânsito 59
5.11 Diagrama de transição de estados do agente de trânsito 60

Lista de Tabelas

4.1 Probabilidades de sequências de estados 46

4.2 Probabilidades HMM 46

5.1 Vetores de estados do agente de trânsito 61

Introdução

O propósito deste trabalho é apresentar os conceitos, bem como propriedades e também a resolução de algumas situações-problema envolvendo Cadeias de Markov, e em especial as Cadeias de Markov Ocultas, representadas por HMM. Para a realização deste propósito, além da abordagem do citado anteriormente, fizemos uso do *software* Geogebra, visando obtenção de resultados com maior precisão e rapidez.
Este trabalho aborda as Cadeias de Markov, que por sua vez faz uso de conceitos , propriedades e demonstrações de Probabilidades , Matrizes e Sistemas Lineares. Em relação ao estudo das probabilidades, a maioria dos casos são constituídos de probabilidades condicionais, nunca descartando a hipótese ou situações de outros cálculos probabilísticos.
Quando se aborda o estudo de probabilidades, a importância do mesmo vai além da aprendizagem esperada em tal conteúdo. Trabalhar, quando possível, de forma prática e contextualizada o relacionamento entre probabilidades e matrizes, proporciona um rico aprendizado composto por abstração e dinamização da aplicabilidade de matrizes à assuntos relacionados ao cotidiano. Neste trabalho, a abordagem das Cadeias de Markov começa com um breve histórico à respeito do matemático que a desenvolveu, seguindo suas teorias, propriedades e demonstrações, entre elas, matriz de transição de estados, diagrama de transição de estados, probabilidade limite, equações de Chapmman-Kolmogorov, chegando por sua vez às Cadeias de Markov Ocultas.
Dentre tantas importantes utilizações das Cadeias de Markov, destaca-se a resolução de situações-problema envolvendo Teoria das Probabilidades, e em especial os processos estocásticos que são processos que evoluem no tempo de maneira probabilística.
No capítulo 1, ocorre a abordagem do estudo da teoria dos Conjuntos e Probabilidades em espaços amostrais equiprováveis. No capítulo 2, a abordagem do teorema da probabilidade total. Nos capítulos 3 e 4, as cadeias de Markov e as cadeias de Markov Ocultas, finalizando com atividades sugeridas a serem desenvolvidas em salas de aulas, compostas por situações-problema de forma contextualizada, com o objetivo de favorecer a aprendizagem e o conhecimento a respeito do tema abordado.

1 Probabilidades

1.1 Introdução aos conjuntos

Nesta seção, há uma sucinta descrição das principais noções da teoria dos conjuntos, servindo como prévia base para o estudo das teorias das probabilidades e Cadeias de Markov.
O agrupamento de objetos com características comuns constitui a formação de um conjunto, que independente de suas apresentações, os mesmos são chamados de elementos ou membros do conjunto.
Os nomes dados aos conjuntos são representados geralmente por letras maiúsculas do nosso alfabeto, e os elementos por letras minúsculas. As duas formas em que se pode descrever um conjunto, podem ser:
• **Propriedade dos elementos**: É utilizada quando se pretende representar os elementos sem escrevê-los individualmente, para isso, escreve-se todos os elementos em uma única propriedade caracterizada, genericamente descreve-se:{ x|x tem a propriedade P}

Exemplo 1.1.

a) {x|x é divisor inteiro de 3 } é uma maneira de indicar o conjunto :{-1; 1; -3; 3 };

b) {x|x é inteiro e $0 \leqslant x \leqslant 500$} pode ser indicado por: {0; 1; 2; ... ; 500}.

• **Enumeração dos elementos**: Quando um conjunto é dado pela enumeração de seus elementos, devemos indicá-lo escrevendo seus elementos entre chaves. Essa notação também é empregada quando o conjunto é infinito, escreve-se alguns elementos que evidenciam a lei de formação com a utilização de reticências.

Exemplo 1.2.

a) Conjunto dos números ímpares positivos: {1,3,5,7,... }

b) Conjunto dos números primos positivos: $\{2,3,5,7,11,13,\ldots\}$

Conjuntos unitário e vazio

Chama-se conjunto **unitário** aquele que possui um único elemento.

Exemplo 1.3.

a) Conjunto dos divisores de 1, inteiros e positivos: $\{1\}$

b) Conjunto das soluções da equação $3x + 1 = 10$: $\{3\}$.

Chama-se conjunto vazio aquele que não possui elemento algum. O símbolo usual para o conjunto vazio é o $\varnothing$.
A obtenção de um conjunto vazio ocorre quando se descreve um conjunto por meio de uma propriedade **P** logicamente falsa.

Exemplo 1.4.

a) $\{\ x|x \neq x\ \} = \varnothing$

b) $\{\ x|x$ é ímpar e múltiplo de $2\ \} = \varnothing$

c) $\{\ x|x > 0$ e $x < 0\ \} = \varnothing$

Conjunto Universo

No desenvolvimento de um tema em Matemática, o conjunto **universo** representado por $\cup$, é aquele que engloba todos os elementos utilizados na abordagem de tal assunto. Este conjunto, que por sua vez traz características distintas em relação ao conjunto vazio, possui uma definição relativa associada à situação do assunto estudado, ou seja, ele pode mudar de acordo com o contexto.
Ao descrevermos um conjunto através da propriedade dos elementos, considerando uma determinada propriedade por **P**, é importante fixar o conjunto universo $\cup$ em que se está trabalhando, desa forma escreve-se: $\{\ x \in \cup \mid x$ possui a propriedade $\mathbf{P}\ \}$.

Igualdade de conjuntos

Dois conjuntos A e B são iguais quando todo elemento de A pertence a B e, reciprocamente, todo elemento de B pertence a A. Simbolicamente é representado por:

$$A = B \Leftrightarrow \forall x, x \in A \Leftrightarrow x \in B$$

1.1.1 União e Interseção de Conjuntos

Considerando dois conjuntos A e B, chama-se **união** de A e B o conjunto formado pelos elementos que pertencem a A ou a B.

$$A \cup B = \{\text{x} \mid \text{x} \in \text{A ou x} \in \text{B}\}$$

Para que haja existência da união de dois ou mais conjuntos, é necessário que os elementos pertençam a pelo menos um dos conjuntos.
Propriedades da união conjuntos:
Para A, B e C conjuntos quaisquer, são válidas as seguintes propriedades:

1. $A \cup A = A$ (idempotente)

2. A $\cup$ $\varnothing$ = A (elemento neutro)

3. A $\cup$ B = B $\cup$ A (comutativa)

4. (A $\cup$ B) $\cup$ C = A $\cup$ (B$\cup$ C) (associativa)

Considerando dois conjuntos A e B, chama-se **interseção** de A e B 0 conjunto formado pelos elementos que pertencem a A e a B.

$$A \cap B = \{ \text{ x} \mid \text{x} \in \text{A e x} \in \text{B}\}$$

Para que haja existência da interseção de dois ou mais conjuntos, é necessário que os elementos pertençam aos conjuntos envolvidos simultaneamente.

Propriedades da Interseção de conjuntos

Para A, B e C conjuntos quaisquer, são válidas as seguintes propriedades:

- 1ª) $A \cap A = A$ (idempotente)

- 2ª) $A \cap \cup = A$ (elemento neutro)

- 3ª) $A \cap B = B \cap A$ (comutativa)

- 4ª) $A \cap (B \cap C\,) = (\,A \cap B) \cap C$ (associativa)

Ainda, em relação à interseção de dois conjuntos, quando ocorrer o conjunto vazio, diz-se que tais conjuntos são disjuntos. Neste caso, considerando dois conjuntos quaisquer A e B, pode-se escrever:

$$A \cap B = \varnothing$$

1.1.2 Complementar de um conjunto

Considerando um conjunto A, seu **complementar** representado por $\overline{A}$ é aquele constituído por todos os elementos que não estão contidos em A, porém, estão no conjunto universo.

$$\overline{A} = \{x | x \notin A\}$$

Exemplo 1.5. Supondo $\cup = \{\ 1;2;3;4;5;6;7;8;9;10;\}$, $A = \{\ 1;2;3;4\}$ e $B = \{3;4;5;6\}$, analisando os conceitos de união, interseção e complementação de conjuntos, pode-se verificar que:

$$\overline{A} = \{5;\ 6;\ 7;\ 8;\ 9;\ 10\}$$

$$\overline{B} = \{1;\ 2;\ 7;\ 8;\ 9;\ 10\}$$

$$A \cup B = \{1;\ 2;\ 3;\ 4;\ 5;\ 6\}$$

$$A \cap B = \{3;\ 4\}$$

Proposição 1.1. *Sendo A, B e C conjuntos quaisquer, as seguintes propriedades valem para a inter-relação da união, interseção e complementar de um conjunto:*

- $A \cup (\,B \cap C) = (\,A \cup B\,) \cap (\,A \cup C\,)$

- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

- $A \cap \varnothing = \varnothing$

- $A \cup \varnothing = A$

- $\overline{(A \cap B)} = \overline{A} \cup \overline{B}$

- $\overline{(A \cup B)} = \overline{A} \cap \overline{B}$

- $\overline{\overline{A}} = A$

1.2 Análise Combinatória

Considerando um determinado conjunto não-vazio com uma quantidade n de elementos, a Análise Combinatória através de métodos, faz com que seja possível a contagem do número de agrupamentos desse considerado conjunto.
A abordagem, não detalhada, do estudo de Análise Combinatória neste trabalho tem importante papel para o cálculo do número de elementos de eventos, experimentos e espaços amostrais das Probabilidades que possam ocorrer no decorrer do mesmo.
Utilizando a notação matemática $n(A)$ para indicar o número de elementos de um conjunto A, e analogamente este conceito de notação a qualquer ou quaisquer outros conjuntos, seguem alguns exemplos:
1º) A é o conjunto de números de dois algarismos distintos formados a partir dos dígitos 1, 2 e 3.
$A = \{$ 12, 13, 21, 23, 31, 32 $\}$; $n(A) = 6$
2º) T é o conjunto das sequências de letras que se obtêm, mudando a ordem das letras da palavra RUA (anagramas da palavra RUA).
$T = \{$ RUA, RAU, URA, UAR, ARU, AUR $\}$; $n(B) = 6$

Príncipio Multiplicativo

Proposição 1.2. *Sendo os conjuntos $A = \{ a_1, a_2, \ldots, a_m \}$ e $B = \{ b_1, b_2, \ldots, b_n \}$, o número de pares ordenados (a_i, b_j), oriundos desses conjuntos, pode ser obtido pelo produto $m.n$, em que $a_i \in A$ e $b_j \in B$.*

Demonstração. Fixando o primeiro elemento do par ordenado e fazendo variar o segundo elemento, numa quantidade de m linhas tem-se:

$$
\begin{gathered}
(a_1, b_1),\ (a_1, b_2),\ \ldots,\ (a_1, b_n) \longrightarrow n \text{ pares} \\
(a_2, b_1),\ (a_2, b_2),\ \ldots,\ (a_2, b_n) \longrightarrow n \text{ pares} \\
\vdots \\
(a_m, b_1),\ (a_m, b_2),\ \ldots,\ (a_m, b_n) \longrightarrow n \text{ pares}
\end{gathered}
$$

Assim o número de pares ordenados será: $\underbrace{n + n + n + \ldots + n}_{m \quad vezes} = m.n$

□

Fatorial

Definição 1.3. Sendo n um inteiro positivo, define-se o fatorial de n, denotado por $n!$ como

$$n! = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 1$$

1.2.1 Arranjos, Permutações e Combinações

Havendo n elementos distintos num determinado conjunto, é possível calcular as permutações dos mesmos, formando-se arranjos constituídos de n elementos.
Denotando-se por P_n as permutações de n elementos, temos:

$$P_n = n!$$

Exemplo 1.6. P_6 = 6! = 6.5.4.3.2.1 = 720

Quando, dentre n elementos distintos, resolvermos tomar (considerar) uma quantidade k desses elementos, $0 \leq k \leq n$, permutando-se os k tomados, temos um Arranjo de n elementos tomados k a k.
Denotando-se por $A_{n,k}$ o número de arranjos de n elementos tomados k a k, temos:

$$A_{n,k} = \frac{n!}{(n-k)!}; n \geq k$$

Exemplo 1.7. $A_{8,5} = \frac{8!}{(8-5)!} = \frac{8!}{3!}$ = 8.7.6.5.4 = 6720

Considerando um conjunto com n elementos, chama-se de combinações de n elementos tomados k a k, aos subconjuntos constituídos de k elementos.
Denotando-se por $C_{n,k}$ ou $\binom{n}{k}$ o número de combinações de n elementos tomados k a k, temos:

$$C_{n,k} = \frac{A_{n,k}}{k!} = \frac{n!}{(n-k)!\,k!}; n \geq k$$

Exemplo 1.8. $C_{10,3} = \binom{10}{3} = \frac{10!}{3!7!} = 120$

1.3 Introdução às Probabilidades

Etimologicamente, a palavra probabilidade deriva de *probabilitas, probabilis* ou *probare*, que resumidamente, todas convergem aos significados: provar, testar, examinar.
Segundo pesquisas científicas e dados históricos, o cálculo de probabilidades surgiu na Idade Média, período no qual os jogos de azar eram habitualmente realizados, trazendo consigo ao mundo da Matemática a procura de uma modelagem, ou melhor, matematização de tais jogos. Vale ressaltar que com o decorrer das teorias das probabilidades, a mesma demonstrou ter diversas aplicações de grande importância em outras ciências, entre elas destacam-se : Biologia, Economia, Engenharias e outras.

1.3.1 Experimento Aleatório

O entendimento de experimento aleatório se torna mais evidente quando o mesmo está associado à exemplos sobre tal abordagem.
Alguns exemplos:

a) Lança-se uma moeda quatro vezes e observa-se a sequência de caras e coroas

b) Números de peças defeituosas num determinado lote

c) No lançamento de um dado, ocorrer face par, considerando a face voltada para cima

Num experimento aleatório, conhecemos as possíveis saídas, e o mesmo pode ser repetido sob as mesmas condições indefinidamente, apresentando variações de resultado e tendo como característica a indeterminação por todo período que o antecede. Nesse texto consideraremos que os possíveis resultados podem ser descritos de forma enumerável.

Estaremos interessados em experimentos aleatórios que quando repetidos uma grande quantidade de vezes, torna-se observável uma regularidade em relação à frequência.

1.3.2 Espaço Amostral e Eventos

Considerando um experimento aleatório ε, define-se **espaço amostral** como o conjunto de todos os resultados possíveis de ε, por questões simbólicas, costuma ser representado por Ω. Quando o número de elementos de um espaço amostral for um número natural, o mesmo é considerado finito, caso contrário, infinito enumerável ou infinito não-enumerável.

Exemplo 1.9.

a) Lançar uma moeda duas vezes e observar o número de caras;

$$\Omega = \{\ 0, 1, 2\} : \text{espaço amostral finito}$$

b) Uma moeda é lançada até que o resultado cara ocorra pela primeira vez. Observa-se em qual lançamento esse fato ocorre;

$$\Omega = \{\ 1, 2, 3, 4,\ \dots\ \} : \text{espaço amostral infinito numerável}$$

c) Um experimento consiste em registrar o retorno percentual diário do investimento em uma carteira de ações. O espaço amostral é formado por um intervalo contínuo de valores que, por conveniência, assumi-se estar entre - 100% e 100%.

$$\Omega = \{-100,\ 100\} : \text{espaço amostral infinito não-numerável.}$$

Eventos

Um **evento**, que costuma ser representado por uma letra maiúscula do nosso alfabeto, é um subconjunto de um espaço amostral Ω. Para que haja a ocorrência de um evento, é necessário que o experimento aleatório realizado pertença ao evento.
Adotando n(Ω) ou # (Ω) como número de elementos de um espaço amostral, se n(Ω) $= n$, $n \in \mathbb{N}$; Ω terá 2^n subconjuntos e, portanto, 2^n eventos.

Exemplo 1.10. No lançamento de uma moeda por três vezes consecutivas, observa-se a sequência de caras e coroas;
Ω = {(K,K,K) ; (K,K,C) ; (K,C,K) ; (K,C,C,) ; (C,K,K) ; (C,K,C) ; (C,C,K) ; (C,C,C) }.
Alguns eventos:

A: Ocorrência de cara (k) no 1º lançamento
A = { (K,K,K) ; (K,K,C) ; (K,C,K) ; (K,C,C) }.
B: Ocorrência de três coroas
B = { (C,C,C) }

Vale ressaltar dois eventos especiais:

- Evento impossível, representado por $\varnothing$;
- Evento certo, representado por Ω, onde o próprio evento é o espaço amostral.

Quando um experimento é realizado, seu resultado é inesperado. É necessário que haja uma atribuição com teor de "escala" para verificação do grau de ocorrência de um determinado evento.
Considerando o seguinte procedimento: Supondo que repetidas n vezes o experimento ε, e sejam A e B dois eventos associados a ε. Admitindo que sejam, respectivamente, n_A e n_B o número de vezes que o evento A e o evento B ocorram nas n repetições.

Definição 1.4. A frequência relativa do evento A nas n repetições de ε é denominada $f_A = \frac{n_A}{n}$

A frequência relativa f_A apresenta as seguintes propriedades, de fácil verificação:

- Propriedade (1) : $0 \leq f_A \leq 1$.
- Propriedade (2) : $f_A = 1$,se e somente se, A ocorrer em todas as n repetições.
- Propriedade (3) : $f_A = 0$, se e somente se, A nunca ocorrer nas n repetições.
- Propriedade (4) : Se A e B forem eventos mutuamente excludentes, e se $f_{A\cup B}$ for a frequência relativa associada ao evento $A \cup B$, então, $f_{A\cup B} = f_A + f_B$.
- Propriedade (5) : Para n repetições do experimento , quando $n \to \infty$, f_A "converge" em sentido probabilístico para $\mathbb{P}(A)$.

Exemplo 1.11. No lançamento de uma moeda, considerando a face voltada para cima como resultado ocorrido, temos duas situações, cara(K) ou coroa(C). Denotando a i-ésima face por X_i, onde $X_i = 1$ se a face for cara, e $X_i = 0$ se for coroa, observa-se uma sequência, por exemplo, 1, 1, 0, 1, 0, 0, 0, 1, 0, 0, 1. É evidente que não há a menor previsão do resultado que irá ocorrer. Calculando-se a frequência relativa em relação ao evento $k = \{$ ocorrer face cara $\}$ e baseada nas observações das ocorrências anteriores, temos: $1, 1, \frac{2}{3}, \frac{3}{4}, \frac{3}{5}, \frac{3}{6}, \frac{3}{7}, \frac{4}{8}, \ldots$. Nota-se que há um elevado grau de variação com maior evidência no início, analisando sua continuidade de forma indefinida, essas frequências relativas chegariam próximas ao valor $\frac{1}{2}$.

A essência desta propriedade é que, se um experimento for executado um grande número de vezes, a frequência relativa da ocorrência de algum evento tenderá a variar cada vez menos à medida que o número de repetições for aumentada. Esta característica é também conhecida como *regularidade estatística*.

1.4 Axiomas de Probabilidade

Definição 1.5. Seja ε um experimento. Seja Ω um espaço amostral associado a ε. A cada evento A, associaremos um número real representado por $\mathbb{P}(A)$ e denominado probabilidade de A, que satisfaça às seguintes propriedades:

- Axioma (1): $0 \leqslant \mathbb{P}(A) \leqslant 1$
- Axioma (2): $\mathbb{P}(\Omega) = 1$
- Axioma (3): Se A_1, A_2, $\ldots$, A_n,forem, dois a dois, eventos mutuamente excludentes, então: $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$

Teorema 1.6. *Se $\varnothing$ for o conjunto vazio, então $\mathbb{P}(\varnothing) = 0$.*

Demonstração. Para qualquer evento A, podemos escrever $A = A \cup \varnothing$. Uma vez que A e $\varnothing$ são mutuamente excludentes, decorre do Axioma (3), que $\mathbb{P}(A) = \mathbb{P}(A \cup \varnothing)=$ $\mathbb{P}(A) + \mathbb{P}(\varnothing)$. □

Teorema 1.7. *Se $\overline{A}$ for o evento complementar de A, então:*

$$\mathbb{P}(A) = 1 - \mathbb{P}(\overline{A})$$

Demonstração. Podemos escrever $\Omega = A \cup \overline{A}$ e, empregando os Axiomas (2) e (3), obteremos $1 = \mathbb{P}(A) + \mathbb{P}(\overline{A})$ □

Teorema 1.8. *Se A e B forem dois eventos quaisquer, então $\mathbb{P}(A \cup B) = \mathbb{P}(A)+\mathbb{P}(B)$-$\mathbb{P}(A \cap B)$.*

Demonstração. $A \cup B = A \cup (B \cap \overline{A})$ e $B = (A \cap B) \cup (B \cap \overline{A})$
Por consequência:
$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \cap \overline{A})$
$\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(B \cap \overline{A})$

$\mathbb{P}(A \cup B)$ - $\mathbb{P}(B)$ = $\mathbb{P}(A)$+$\mathbb{P}(B \cap \overline{A})$ - $\mathbb{P}(A \cap B)$ - $\mathbb{P}(B \cap \overline{A})$
$\mathbb{P}(A \cup B)$ - $\mathbb{P}(B)$ = $\mathbb{P}(A)$ - $\mathbb{P}(A \cap B)$
$\mathbb{P}(A \cup B)$ = $\mathbb{P}(A)$ + $\mathbb{P}(B)$ - $\mathbb{P}(A \cap B)$ □

Teorema 1.9. *Se A, B e C forem três eventos quaisquer, então:*
$\mathbb{P}(A \cup B \cup C)$ = $\mathbb{P}(A)$+$\mathbb{P}(B)$+$\mathbb{P}(C)$ - $\mathbb{P}(A \cap B)$- $\mathbb{P}(A \cap C)$ - $\mathbb{P}(B \cap C)$ + $\mathbb{P}(A \cap B \cap C)$.

Demonstração. Considerando A_1, A_2, ..., A_k ; quaisquer k eventos.
$\mathbb{P}(A_1 \cup A_2 \cup \ldots \cup A_k)$= $\Sigma_{i=1}^{k}$ $\mathbb{P}(A_i)$ - $\Sigma_{i<j=2}^{k}$ $\mathbb{P}(A_i \cap A_j)$ + $\Sigma_{i<j<r=3}^{k}$ $\mathbb{P}(A_i \cap A_j \cap A_r)$ + ... + $(-1)^{k-1}$ $\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_k)$. □

Teorema 1.10. *Se $A \subset B$, então $\mathbb{P}(A) \leq \mathbb{P}(B)$*

Demonstração. Decompondo B em dois eventos mutuamente excludentes, $B = A \cup (B \cap \overline{A}\,)$.
$\mathbb{P}(B)$ = $\mathbb{P}(A)$ + $\mathbb{P}(B \cap \overline{A}) \geq \mathbb{P}(A)$; pois $\mathbb{P}(B \cap \overline{A}) \geq 0$, pelo Axioma (1). □

1.5 Probabilidades em espaços amostrais equiprováveis

Considerando um espaço amostral Ω, formado por k pontos amostrais:

$$\Omega = \{a_1, a_2, a_3, \cdots a_k\}$$

Vamos associar a cada um desses pontos amostrais um número real, $\mathbb{P}a_i$, ou simplesmente, $\mathbb{P}_i$, chamado probabilidade do evento a_i,tal que:
i. $0 \leqslant \mathbb{P}_i \leqslant 1$
ii. $\sum_{i=1}^{k} P_i = 1$, isto é,$\mathbb{P}_1 + \mathbb{P}_2 + \cdots + \mathbb{P}_k = 1$

Teorema 1.11. *Num espaço amostral equiprovável, a probabilidade de ocorrer determinado evento é dada pela razão entre o número de casos favoráveis e o número de casos possíveis.*

Demonstração. Denotando por $\mathbb{P}$ a probabilidade de ocorrência de cada um dos pontos amostrais de Ω:

$$\underbrace{\mathbb{P} + \mathbb{P} + ... + \mathbb{P}}_{k \quad vezes} = 1 \Rightarrow k.\mathbb{P} = 1 \Rightarrow \mathbb{P} = \frac{1}{k}$$

A probabilidade de ocorrência de um evento E, formado por r pontos amostrais E = $\{a_1, a_2, \cdots, a_r\}$, com r $\leqslant$ k, é dada por:

$$\begin{aligned} \mathbb{P}(E) &= \mathbb{P}_1 + \mathbb{P}_2 + \cdots + \mathbb{P}_r \\ &= \underbrace{\frac{1}{k} + \frac{1}{k} + \cdots + \frac{1}{k}}_{r \quad vezes} \\ &= \frac{r}{k} \\ &= \frac{n(E)}{n(\Omega)} \end{aligned} \tag{1.1}$$

$\square$

Como $E \subset \Omega$,temos que n(E) $\leqslant$ n(Ω). Dessa forma,temos :

$$\mathbb{P}(E) = \frac{n(E)}{n(\Omega)} tal \quad que \quad 0 \leqslant \mathbb{P}(E) \leqslant 1$$

2 Probabilidade Condicional

2.1 Probabilidade Condicional e Eventos Independentes

Probabilidade Condicional

Trata-se do estudo das possibilidades de um acontecimento condicionado a outro, ou seja, é a probabilidade de ocorrer um evento com base num evento anterior. Obviamente, dentro de um espaço amostral finito, haverá tais eventos como conjuntos não-vazios.
Sendo A e B dois eventos associados a um experimento ε, denota-se por $\mathbb{P}(B\,|\,A)$ a probabilidade condicionada do evento B, quando A tiver ocorrido. Ao calcular-se $\mathbb{P}(B\,|\,A)$, calcula-se essencialmente $\mathbb{P}(B)$ em relação ao espaço amostral A, ao invés de realizá-lo em relação ao espaço amostral original, comumente representado por Ω .
É importante destacar outra probabilidade também inserida no contexto de Probabilidade Condicional, trata-se de $\mathbb{P}(A \cap B)$, " probabilidade da interseção dos eventos A e B ".Utilizando o conceito de frequência relativa para uma definição formal de $\mathbb{P}(A\,|\,B)$, temos:

Heurística 1. Para um experimento aleatório repetido n vezes, sejam n_A, n_B e $n_{A\cap B}$ o número de vezes que ocorrem A, B e $(A \cap B)$, respectivamente. A frequência relativa de A nos resultados em que B ocorre é obtida por $\frac{n_{A\cap B}}{n_B}$, ou seja, a frequência relativa de A condicionada à ocorrência de B.

$$\frac{n_{A\cap B}}{n_B} = \frac{\frac{n_{A\cap B}}{n}}{\frac{n_B}{n}} = \frac{f_{A\cap B}}{f_B}$$

Considerando n "grande", $f_{A\cap B}$ se torna bem próxima e com desconsiderada margem de erro de $\mathbb{P}(A \cap B)$ e f_B se torna próxima de $\mathbb{P}(B)$. Logo, podemos considerar:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}; \quad \mathbb{P}(B) > 0$$

Exemplo 2.1. Dois dados, d_x e d_y são lançados respectivamente, fornecendo o seguinte espaço amostral:

$$\Omega = \begin{pmatrix} (1,1) & (1,2) & \cdots & (1,6) \\ (2,1) & (2,2) & \cdots & (2,6) \\ \vdots & \vdots & \ddots & \vdots \\ (6,1) & (6,2) & \cdots & (6,6) \end{pmatrix}$$

Sejam os eventos:
A: o dado d_x apresenta resultado 2,
B: a soma dos pontos nos dois dados é 6.
Calculando-se $\mathbb{P}(A|B)$, temos:
A = {(2,1), (2,2), (2,3), (2,4), (2,5), (2,6) } $\rightarrow \mathbb{P}(A) = \frac{6}{36} = \frac{1}{6}$
B = { (1,5), (2,4), (3,3), (4,2), (5,1) } $\rightarrow \mathbb{P}(B) = \frac{5}{36}$
$(A \cap B)$ = { (2,4) } $\rightarrow \mathbb{P}(A \cap B) = \frac{1}{36}$
$\therefore \mathbb{P}(A|B) = \frac{\frac{1}{36}}{\frac{5}{36}} = \frac{1}{5}$

Probabilidade de Eventos Independentes

Considerando um espaço amostral Ω e dois eventos A e B contidos no mesmo, pode-se dizer que A independe de B, se $\mathbb{P}(A|B) = \mathbb{P}(A)$, ou seja, A independe de B se a ocorrência de B não afeta a probabilidade de A.
Da mesma forma, quando se diz que $\mathbb{P}(A|B) = \mathbb{P}(A)$, pode-se dizer também que $\mathbb{P}(B|A) = \mathbb{P}(B)$.

Heurística 2.

$$\begin{aligned} \mathbb{P}(B|A) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \\ &= \frac{\mathbb{P}(B).\mathbb{P}(A|B)}{\mathbb{P}(A)} \\ &= \frac{\mathbb{P}(B).\mathbb{P}(A)}{\mathbb{P}(A)} \\ &= \mathbb{P}(B) \end{aligned}$$

Levando em consideração que se A independe de B, então B independe de A, pode-se deduzir que o cálculo das probabilidades de dois eventos independentes pode ser definida por:

Definição 2.1.

$$\mathbb{P}(A \cap B) = \mathbb{P}(A).\underbrace{\mathbb{P}(B|A)}_{\mathbb{P}(B)} = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

Exemplo 2.2. Lança-se uma moeda por três vezes consecutivas e analisa-se os eventos e o espaço amostral:

Ω = { (K,K,K), (K,K,C), (K,C,K), (K,C,C), (C,K,K), (C,K,C), (C,C,K), (C,C,C) }
A: ocorrer pelo menos duas caras,
B: ocorrer resultados iguais nos três lançamentos.
A = { (K,K,K), (K,K,C), (K,C,K), (C,K,K) } $\rightarrow \mathbb{P}(A) = \frac{4}{8} = \frac{1}{2}$
B = { (K,K,K), (C,C,C) } $\rightarrow \mathbb{P}(B) = \frac{2}{8} = \frac{1}{4}$
$(A \cap B)$ = { (K,K,K) } $\rightarrow \mathbb{P}(A \cap B) = \frac{1}{8}$
Logo,$\underbrace{\mathbb{P}(A \cap B)}_{\frac{1}{8}} = \underbrace{\mathbb{P}(A)}_{\frac{1}{2}}.\underbrace{\mathbb{P}(B)}_{\frac{1}{4}}$
$\therefore A$ e B são independentes.

Exemplo 2.3. Uma moeda é lançada sete vezes. Qual a probabilidade de se observar cara nos sete lançamentos?
A_1: ocorrer cara no 1º lançamento
A_2: ocorrer cara no 2º lançamento
$\vdots$
A_7: ocorrer cara no 7º lançamento
Pode-se admitir que A_1, A_2, $\cdots$, A_7 são eventos independentes, pois cada lançamento não é afetado pelos outros.
$\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_7) = \mathbb{P}(A_1).\mathbb{P}(A_2). \ldots . \mathbb{P}(A_7)$
A probabilidade de ocorrer cara em qualquer lançamento é $\frac{1}{2}$.
Logo:
$\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_7) = \underbrace{\frac{1}{2}.\frac{1}{2}\ldots\frac{1}{2}}_{7vezes} = \frac{1}{128}$

2.2 Probabilidade Total

Dizemos que os eventos $E_1, E_2, \ldots, E_n$; formam uma partição do espaço amostral Ω, quando:

- $\mathbb{P}(E_k) > 0$; $\forall k$
- $E_i \cap E_j = \varnothing$ para $i \neq j$
- $\bigcup_{i=1}^{n} E_i = \Omega$

Deduz-se que, os eventos $E_1, E_2, \ldots, E_n$ são dois a dois mutuamente exclusivos e exaustivos, ou seja, sua união é Ω.
Sendo Ω um espaço amostral, A um evento qualquer de Ω e $E_1, E_2, \ldots, E_n$ uma parti-

ção de Ω, vale a seguinte relação:

$$A = (E_1 \cap A) \cup (E_2 \cap A) \cup (E_3 \cap A) \cup \ldots \cup (E_n \cap A)$$

Em casos onde o cálculo da probabilidade apresentar maior dificuldade no processo resolutivo, pode-se utilizar o Teorema da Probabilidade Total como método facilitador. Considerando a relação acima, podemos escrevê-la como

$$\mathbb{P}(A) = \mathbb{P}(E_1 \cap A) + \mathbb{P}(E_2 \cap A) + \mathbb{P}(E_3 \cap A) + \ldots + \mathbb{P}(E_n \cap A)$$

Exemplo 2.4. Existem três urnas, A, B e C. Sabe-se que a urna A contém 2 bolas vermelhas (V) e 3 bolas brancas (B); a urna B contém 3 bolas vermelhas (V) e 1 branca (B); a urna C contém 4 bolas vermelhas (V) e 2 bolas brancas (B). Seleciona-se uma urna ao acaso, e dela extraí-se uma bola. Qual a probabilidade de sortear-se aleatoriamente uma bola vermelha?

Resolução:

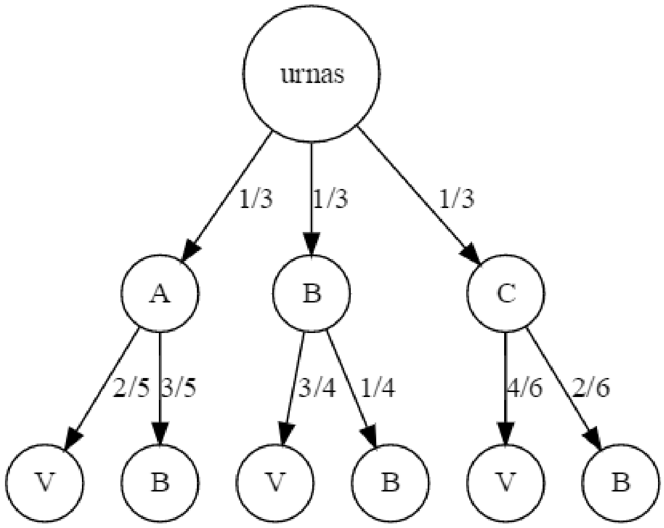


Figura 2.1: Três urnas

A partição de Ω é composta pelos eventos:
E_A : sair urna A
E_B : sair urna B
E_C : sair urna C
Sendo V o evento "sair bola vermelha", pelo teorema da probabilidade total:

$$\mathbb{P}(V) = \mathbb{P}(E_A \cap V) + \mathbb{P}(E_B \cap V) + \mathbb{P}(E_C \cap V)$$

Analisando o diagrama e utilizando o teorema da multiplicação:
$\mathbb{P}(E_A \cap V) = \frac{1}{3} \cdot \frac{2}{5} = \frac{2}{15}$
$\mathbb{P}(E_B \cap V) = \frac{1}{3} \cdot \frac{3}{4} = \frac{1}{4}$
$\mathbb{P}(E_C \cap V) = \frac{1}{3} \cdot \frac{4}{6} = \frac{2}{9}$
Logo:
$\mathbb{P}(V) = \frac{2}{15} + \frac{1}{4} + \frac{2}{9} = \frac{109}{180}$

Exemplo 2.5. *Problema da moeda de Bertrand*
Existem três caixas idênticas. A 1ª(I) contém duas moedas de ouro, a 2ª (II) contém uma moeda de ouro e outra de prata, e a 3ª (III), duas moedas de prata. Uma caixa é selecionada ao acaso e da mesma é escolhida uma moeda ao acaso. Se a moeda escolhida for de ouro, qual a probabilidade de que a outra moeda da caixa escolhida também seja de ouro?

Resolução:
Podemos analisar esse problema também de outra forma: "Escolhendo-se a moeda de ouro, qual a probabilidade de que ela tenha vindo da caixa (I)"?

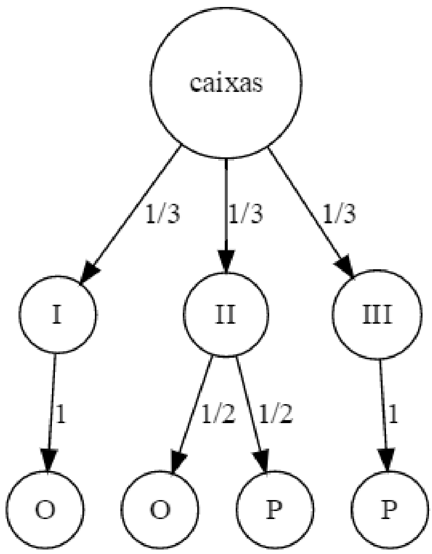


Figura 2.2: Moeda de Bertrand

A partição de Ω é composta pelos eventos:
E_I: sortear a caixa I
E_{II}: sortear a caixa II
E_{III}: sortear a caixa III
Sendo O o evento "sortear moeda de ouro", pelo teorema da probabilidade total:

$$\begin{aligned}\mathbb{P}(O) &= \mathbb{P}(\cup_I \cap O) + \mathbb{P}(\cup_{II} \cap O) + \mathbb{P}(\cup_{III} \cap O)\\ &= \frac{1}{3}\cdot 1 + \frac{1}{3}\cdot\frac{1}{2} + \frac{1}{3}\cdot 0\\ &= \frac{1}{2}\end{aligned}$$

Pela condição da moeda escolhida ser de ouro:

$$\mathbb{P}(\cup_I|O) = \frac{\mathbb{P}(\cup_I \cap O)}{\mathbb{P}(O)} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

2.3 Teorema de Bayes

O Teorema de Bayes é um encadeamento dedutivo do Teorema da Probabilidade Total, que de forma conceitual, admitindo os eventos A e B e por consequência suas probabilidades, apresenta-se por:

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

Teorema 2.2. *A probabilidade de ocorrência de um evento E_i, supondo-se a ocorrência do evento A, é dada por:*

$$\mathbb{P}(E_i|A) = \frac{\mathbb{P}(E_i)\mathbb{P}(A|E_i)}{\sum_{j=1}^{n}\mathbb{P}(C_j)P(A|C_j)}, \forall i = 1, 2, \ldots, n$$

Exemplo 2.6. A administração de um fundo de investimentos em ações pretende divulgar, após o encerramento do pregão, a probabilidade de queda de um índice da bolsa no dia seguinte, baseando-se nas informações disponíveis até aquele momento. Supondo que a previsão inicial seja de 0,10. Após encerrado o pregão, nova informação sugere uma alta do dólar frente ao real. A experiência passada indica que, quando houve queda da bolsa no dia seguinte, 20% das vezes foram precedidas por esse tipo de notícia, enquanto, nos dias em que a bolsa esteve em alta, apenas 5% das vezes houve esse tipo de notícia no dia anterior.

Resolução:
Chamando de E o evento que indica "queda da bolsa", a sua probabilidade *a priori* é $\mathbb{P}(E)$ = 0,10, enquanto a probabilidade de alta é $\mathbb{P}(\overline{E})$ = 0,90. Se B indicar alta do dólar, então as verosimilhanças são dadas por:

$$\mathbb{P}(B|E) = 0,20 \quad \mathbb{P}(B|\overline{E}) = 0,05$$

Pelo Teorema de Bayes:

$$\begin{aligned}\mathbb{P}(E|B) &= \frac{\mathbb{P}(E)\mathbb{P}(B|E)}{\mathbb{P}(E)\mathbb{P}(B|E) + \mathbb{P}(\overline{E})\mathbb{P}(B|\overline{E})} \\ &= \frac{(0,10)(0,20)}{(0,10)(0,20) + (0,90)(0,05)} \\ &= \frac{0,02}{0,065} = 0,31\end{aligned}$$

Portanto, a nova informação aumenta a probabilidade de que haja queda na bolsa de 10% para 31%.

3 Cadeias de Markov

3.1 Introdução

Em meados de 1907, Andrei Andrejevitch Markov, matemático russo, realizou estudos a respeito de um importante tipo de processo, no qual o resultado dos estados futuros dependiam apenas do estado atual e não dos estados que o antecediam. Estes processos possuíam uma propriedade conhecida como "perda de memória "(*memory loss*) ou Propriedade de Markov.

A aplicabilidade das Cadeias de Markov está relacionada de forma concisa e ampla à diversos setores e fenômenos associados ao nosso cotidiano, entre eles: previsão do tempo, Economia, jogos, confiabilidade e outras. Tendo como princípio um estado inicial, e este por sua vez, realizando transição com outro ou outros estados, por meio de certas probabilidades envolvidas num intervalo de tempo discreto com probabilidade de transição dependendo apenas do estado naquele determinado momento, dá-se o nome de Processo de Markov. Este processo é importante no estudo de confiabilidade pelo fato de permitir que um determinado estado apresentando necessidade de "reparos" seja reparado, e ao mesmo tempo sua probabilidade de transição ou permanência também serão conhecidas.

Exemplo 3.1. Considere um sistema formado por duas situações distintas, denominadas estados A e B. Representando simultaneamente, as situações descritas anteriormente, a inoperância e o funcionamento de um gerador de energia elétrica, assume-se que as probabilidades são constantes para todos os instantes de tempos discretos no futuro. Sabe-se que a probabilidade de permanecer no estado A é de 0,5, de permanência em B de 0,75; de mudança, ou seja, transição de A para B de 0,5 e de B para A de 0,25.

Qual a probabilidade de se encontrar no estado A no terceiro instante de tempo, numa situação iniciada pelo estado A ? E iniciada pelo estado B?

Solução:

Observação: Pelo fato do movimento entre os estados ocorrer em tempos discretos, temos uma Cadeia de Markov a tempo discreto.
Apresentação do sistema

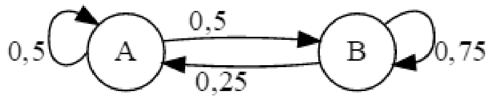


Figura 3.1: Apresentação do Sistema

Árvore característica do sistema iniciada no estado A

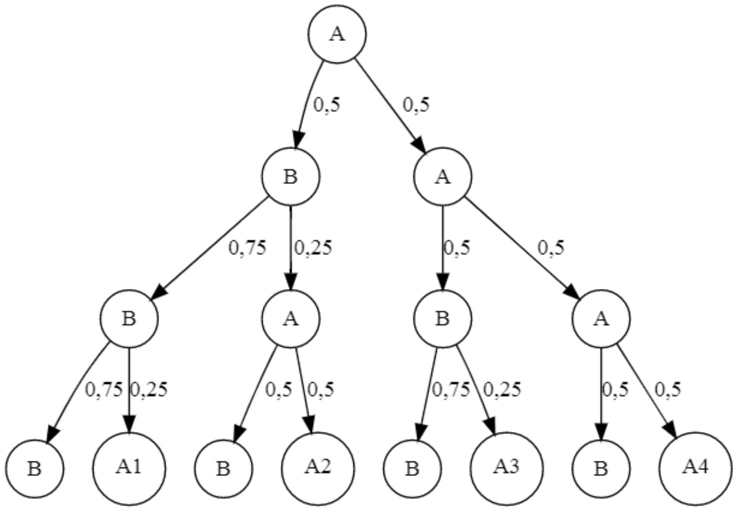


Figura 3.2: Árvore característica

Cálculo da probabilidade:
$A_1 = 0,5 \cdot 0,75 \cdot 0,25 = 0,09375$
$A_2 = 0,5 \cdot 0,25 \cdot 0,5 = 0,0625$
$A_3 = 0,5 \cdot 0,5 \cdot 0,25 = 0,0625$
$A_4 = 0,5 \cdot 0,5 \cdot 0,5 = 0,125$
Assim, a probabilidade de se encontrar no estado A no terceiro tempo $= A_1 + A_2 + A_3 + A_4 = 0,34375$.

Árvore característica do sistema iniciada no estado B

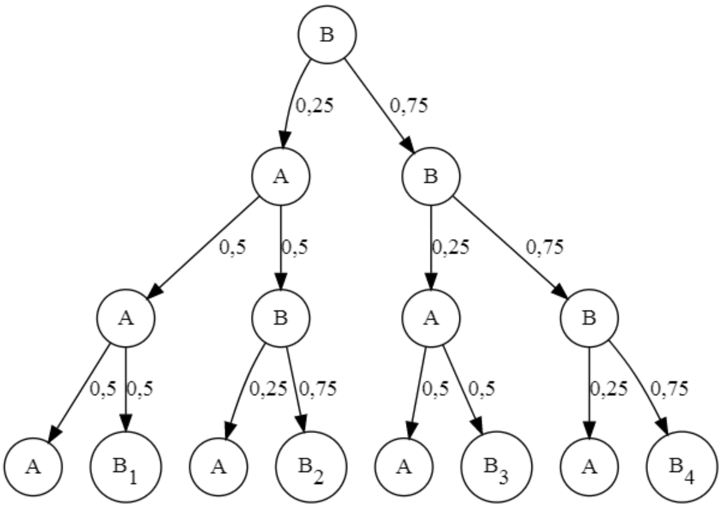


Figura 3.3: Árvore característica

Cálculo da probabilidade:
$B_1 = 0,25 \cdot 0,5 \cdot 0,5 = 0,0625$
$B_2 = 0,25 \cdot 0,5 \cdot 0,75 = 0,09375$
$B_3 = 0,75 \cdot 0,25 \cdot 0,5 = 0,09375$
$B_4 = 0,75 \cdot 0,75 \cdot 0,75 = 0,42187$
A probabilidade de se encontrar no estado B no terceiro tempo $= B_1 + B_2 + B_3 + B_4 = 0,67187$.

$\triangleleft$

Definição 3.1. Um **Processo Markoviano** é uma sequência de variáveis aleatórias $\{X_n, n = 0, 1, 2, 3, \dots\}$ satisfazendo:

$$\mathbb{P}\{X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0\} = \mathbb{P}\{X_{n+1} = j | X_n = i\} = P_{ij}$$

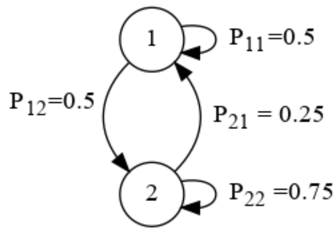
Um Processo Markoviano é dito ser uma **Cadeia de Markov** quando as variáveis randômicas estão definidas em um espaço de estado discreto.

O valor P_{ij} representa a probabilidade de que o processo, quando no estado i, faça uma transição para o estado j. Uma vez que as probabilidades são não-negativas e uma vez que o processo deva fazer uma transição para algum estado, obtemos:

$$P_{ij} \geq 0; \quad i, j \geq 0; \quad \sum_{j=0}^{\infty} P_{ij} = 1; \quad i = 0, 1, \ldots$$

3.2 Diagrama de Transição de Estado

Para se representar uma Cadeia de Markov, utiliza-se um recurso gráfico denominado **Diagrama de Transição de Estados** composto por círculos ou elipses - tendo estes, função de denominar os estados envolvidos - segmentos orientados (localizados entre os estados de transição) - funcionando como indicações das transições entre os estados, e acima ou ao lado desses segmentos orientados encontram-se na maioria dos casos as probabilidades de transições entre os estados.
Retomando a apresentação do sistema do exemplo 3.1 envolvida com o conceito citado anteriormente pelo modelo generalizado P_{ij} de representação de probabilidades e considerando o estado A como 1 e estado B como 2, temos:



Exemplo 3.2. Baseado num experimento de estímulos de Skinner, que aborda experiências sob influência de estímulos sensoriais, um cachorro foi colocado experimentalmente num local com 3 divisões representadas por 1, 2 e 3 conforme a figura abaixo. Sabe-se que por influência de estímulo sonoro, o animal muda de repartição em busca de um determinado alimento, sabe-se que continua mantendo-se igual probabilidade de escolha a qualquer uma das repartições, bem como, não haverá influência em relação às escolhas anteriores.

Referente ao exemplo acima, podemos citar um diagrama de transição de estados da seguinte forma:

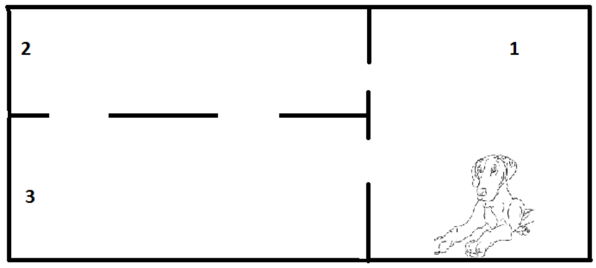


Figura 3.4: Estímulo de Skinner

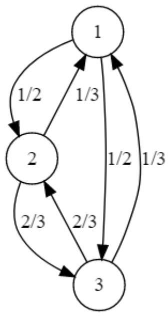


Figura 3.5: Diagrama de transição do modelo experimental de Skinner

3.3 Vetor Probabilidade

As probabilidades de transição entre os estados num intervalo de tempo discreto na forma de matriz linha contidos numa Cadeia de Markov compreendem o que se caracteriza por Vetor Probabilidade.
Genericamente, pode-se representar por:

$$V_i = [P_{i1} \quad P_{i2} \quad P_{i3} \quad \ldots \quad P_{ik}]$$

Em que:

P_{i1}: representa a probabilidade de transição do estado i para o estado 1.

$$\ldots$$

P_{ik}: representa a probabilidade de transição do estado i para o estado k.

Considerações importantes:

I) Os elementos que constituem um vetor probabilidade são positivos, e seguindo os conceitos probabilísticos, a soma dos mesmos será sempre igual a 1.

II) Quando nos referimos a uma certa observação, seja ela a enésima observação, por exemplo, escrevemos $V_i^{(n)}$, dessa forma, seguindo esta última consideração:
$V_i^{(0)}$: vetor de probabilidade numa referida observação inicial;
$V_i^{(1)}$: vetor de probabilidade numa primeira observação.
Utilizando os exemplos 3.1 e 3.2 para construção de seus respectivos Vetores Probabilidades, temos:
No caso do exemplo 3.1 :
$V_1 = [\ \frac{1}{2} \quad \frac{1}{2}\ ]$
$V_2 = [\ \frac{1}{4} \quad \frac{3}{4}\ ]$

No caso do exemplo 3.2 :
$V_1 = [\ 0 \quad \frac{1}{2} \quad \frac{1}{2}\ ]$
$V_2 = [\ \frac{1}{3} \quad 0 \quad \frac{2}{3}\ ]$
$V_3 = [\ \frac{1}{3} \quad \frac{2}{3} \quad 0\ ]$

Observa-se que as probabilidades de permanência nos estados P_{11}, P_{22} e P_{33} são todas iguais a zero, justificando que o cachorro quando submetido ao estímulo sonoro sempre mudará de um estado para outro.

3.4 Matrizes

Historicamente, os matemáticos Joseph Sylvester e Arthur Cayley há aproximadamente 150 anos , foram os precursores do estudo de matrizes, cabendo ao segundo, maior aprofundamento em relação às demonstrações e aplicabilidades à respeito da mesma.
Segundo os Parâmetros Curriculares Nacional (PCN), a abordagem do estudo de matrizes deve ocorrer no ensino médio, salvo exceções. Dentro da abordagem do estudo de matrizes, na maioria dos casos, são focadas suas propriedades, bem como exercícios para fixação, porém o vasto mundo das aplicações costumam ser deixadas para um segundo plano. Tanto no ramo da Matemática, quanto em outros, as aplicações de matrizes são várias, entre elas: computação gráfica, confecções de tabelas, previsões climáticas e outras. Na área de Economia, por exemplo, onde são utilizados comumente gráficos e tabelas, as matrizes desenvolvem importante papel no armazenamento de dados, disponibilizando-os de forma organizada e precisa.

Nesse trabalho apresentaremos uma aplicação das matrizes na descrição das Cadeias de Markov.

Definição de matrizes

Chama-se matriz de ordem *m* por *n* (*m* x *n*) elementos, dos quais numéricos, um conjunto de linhas e colunas disponibilizando os mesmos nesta ordem, possuindo suas características e propriedades matemáticas definidas. Resumidamente, podemos definir como uma tabela de informações codificadas de forma numérica.

Representação de uma Matriz

Para uma determinada matriz A, a mesma pode ser representada abreviadamente por $A = [a_{ij}]$, i variando de 1 a m $(i = 1, 2, \ldots, m)$ e j variando de 1 a n $(j = 1, 2, \ldots, n)$

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \ldots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \ldots & a_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \ldots & a_{mn} \end{bmatrix}$$

Dessa forma, $[a_{ij}]$, representa qualquer matriz A de ordem m por n.
Didaticamente, costuma-se chamar m por número de linhas e n por número de colunas.

3.5 Matriz de Transição

É a matriz formada pelos vetores de probabilidades envolvidos numa Cadeia de Markov de k estados, seguindo as propriedades já mencionadas.
Representa-se a matriz de transição por $M = [P_{ij}]$.

$$M = \begin{bmatrix} P_{11} & P_{12} & P_{13} & \ldots & P_{1k} \\ P_{21} & P_{22} & P_{23} & \ldots & P_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ P_{k1} & P_{k2} & P_{k3} & \ldots & P_{kk} \end{bmatrix}$$

Exemplo 3.3. Supondo que se hoje chove ou não, venha depender apenas das condições climáticas anteriores referentes aos últimos dois dias. Especificamente, supondo que se choveu nos últimos dois dias, então choverá amanhã com probabilidade de 0,7; se choveu hoje mas não ontem, então choverá amanhã com probabilidade de 0,5; se choveu ontem mas não hoje, então choverá amanhã com probabilidade 0,4; se choveu nos últimos dois dias, então choverá amanhã com probabilidade 0,2. Exiba, utilizando uma matriz de transição, a distribuição das probabilidades envolvidas.

Resolução: Se deixarmos o estado no tempo n depender somente se está ou não chovendo no tempo n, então não caracterizará uma Cadeia de Markov. No entanto, podemos transformar este modelo em uma Cadeia de Markov dizendo que o estado a qualquer momento é determinado pelas condições meteorológicas tanto naquele dia como no dia anterior. Dessa forma, podemos expor o processo da seguinte forma:

- Estado 0: se chove hoje e choveu ontem

- Estado 1: se chove hoje, mas não choveu ontem

- Estado 2: se choveu ontem, mas não chove hoje

- Estado 3: se não choveu ontem ou hoje

$$P = \begin{bmatrix} 0,7 & 0 & 0,3 & 0 \\ 0,5 & 0 & 0,5 & 0 \\ 0 & 0,4 & 0 & 0,6 \\ 0 & 0,2 & 0 & 0,8 \end{bmatrix}$$

Exemplo 3.4. Considere um jogador que, em cada jogo, ganha R$ 1,00 com probabilidade p ou perde R$ 1,00 com probabilidade $1-p$. Supondo que o referido jogador deixa de jogar quando ele perde ou alcança uma fortuna de R$ N, então a fortuna do jogador é uma Cadeia de Markov com probabilidade de transição dada por:

$$p_{i,i+1} = p = 1 - p_{i,i-1} \quad ; \quad i = 1, 2, \ldots, N-1$$

$$p_{00} = p_{NN} = 1$$

Os estados 0 e N são chamados de estados absorventes, pois uma vez que se encontrar nos mesmos, não é possível deixá-los.

3.6 Equações de Chapman-Kolmogorov

As Equações de Chapman-Kolmogorov possibilitam obter o resultado a partir de uma matriz de transição de passo n, n passos no tempo. Considerando P_{ij}^{n} como a probabilidade de transição do estado i para o estado j de passo n e ξ representando os estados, podemos interpretar como:

$$P_{ij}^{n} = P\{X_{n+k} = j | X_k = i\}, \quad n \geq 0; \quad i,j \geq 0$$
$$P_{ij}^{(n+m)} = \sum_{k \in \xi} P_{ik}^{n} P_{kj}^{m} \quad para \quad todo \quad i, \quad j, \quad m,n \geq 0.$$

As probabilidades de transição P_{ik}^{n} e P_{kj}^{m} representam a probabilidade que a partir de i o processo vá para o estado j em $n+m$ transições através de um caminho que conduz ao estado k na n-ésima transição. Assim, somando-se todos os estados intermediários k, resultará a probabilidade de que o processo estará no estado j após transições $n+m$. De forma generalizada, pode-se representar por:

$$\begin{aligned} P_{ij}^{n+m} &= P\{X_{n+m} = j | X_0 = i\} \\ &= \sum_{k \in \xi} P\{X_{n+m} = j, X_n = k | X_o = i\} \\ &= \sum_{k \in \xi} P\{X_{n+m} = j, X_n = k, X_o = i\} P\{X_n = k | X_o = i\} \\ &= \sum_{k \in \xi} P_{kj}^{m} P_{ik}^{n} \end{aligned}$$

Como P^n denota a matriz de probabilidades de transição de $passon$, com base na equação citada anteriormente, podemos afirmar que $P^{(n+m)} = P^n.P^m$.
Por indução, pode-se demonstrar que:

$$P^n = P^{(n-1+1)} = P^{n-1}.P = P^n$$

Dessa forma, a matriz de transição de n passos, pode ser obtida, multiplicando-se a matriz P por si mesma n vezes. Para isso, observemos um caso em particular:

$$P^2 = P^{(1+1)} = P.P = P^2$$

Exemplo 3.5. Suponha que a chance de chuva amanhã dependa das condições climáticas anteriores somente se está ou não chovedo hoje, e não das condições climáticas do passado, e também, que se chover hoje, então choverá amanhã com probabilidade α, e se não chover hoje, então choverá amanhã com probabilidade β.
Considerando que o processo está no estado 0 quando chove e estado 1 quando não

chove, então o precedente é uma Cadeia de Markov de dois estados cujas probabilidades de tansição são dadas por:

$$P = \begin{bmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{bmatrix}$$

Admitindo α = 0,7 e β = 0,4, qual a probabilidade de chuva quatro dias a partir de hoje, dado que hoje está chovendo?

Resolução:
A matriz de probabilidade de transição de passo 1 é dada por:

$$P = \begin{bmatrix} 0,7 & 0,3 \\ 0,4 & 0,6 \end{bmatrix}$$

$$P^2 = \begin{bmatrix} 0,7 & 0,3 \\ 0,4 & 0,6 \end{bmatrix} . \begin{bmatrix} 0,7 & 0,3 \\ 0,4 & 0,6 \end{bmatrix} = \begin{bmatrix} 0,61 & 0,39 \\ 0,52 & 0,48 \end{bmatrix}$$

$$P^4 = (P^2)^2 = \begin{bmatrix} 0,61 & 0,39 \\ 0,52 & 0,48 \end{bmatrix} . \begin{bmatrix} 0,61 & 0,39 \\ 0,52 & 0,48 \end{bmatrix} = \begin{bmatrix} 0,5749 & 0,4251 \\ 0,5668 & 0,4332 \end{bmatrix}$$

A probabilidade em questão é P^4_{00}, que equivale a 0,5749.

3.7 Classificação dos estados em Cadeias de Markov

a) **Alcançável**: Um estado j é dito ser alcançável ou acessível, a partir de um estado i, quando $P^n_{ij} > 0$ para algum $n > 0$. Isto implica na probabilidade do sistema entrar no estado j eventualmente quando este começa no estado i.

$$P[\text{entrar em } j| \text{ começando em } i] = P\{\cup_{n=0}^{\infty}[X_n = j]|X_o = i\}$$
$$\leq \sum_{n=0}^{\infty} PX_n = j|X_o = i$$
$$\sum_{n=0}^{\infty} P^n_{ij}$$

b) **Comunicante**: Consiste em um estado j ser alcançável a partir do estado i, assim como, o estado i ser alcançável a partir do estado j. Costuma-se representar tal relação entre os estados comunicantes por $i \leftrightarrow j$.

Quando qualquer estado se comunicar com seu próprio início, podemos definir como:

$$P_{ii}^{0} = \mathbb{P}\{X_o = i | X_o = i\} = 1$$

A relação de comunicação entre os estados satisfaz três propriedades:

- i) Estado i se comunica com estado i, para todo $i \geq 0$

- ii) Se o estado i se comunica com o estado j, então, estado j se comunica com estado i

- iii) Se o estado i se comunica com o estado j, e o estado j se comunica com estado k, então o estado i se comunica com o estado k.

As propriedades (i) e (ii) seguem claramente a definição de comunicação.
Para provar a propriedade (iii), supõe-se que i se comunica com j, e j se comunica com k. Portanto, existem inteiros n e m tais que $P_{ij}^{n} \geq 0$, $P_{jk}^{m} \geq 0$. Pelas equações de Chapman-Kolmogorov, temos:

$$P_{ik}^{n+m} = \sum_{r=0}^{\infty} P_{ir}^{n} P_{rk}^{m} \geq P_{ij}^{n} P_{jk}^{m} > 0$$

Consequentemente, o estado k é acessível a partir do estado i. Similarmente, podemos expor que o estado i é acessível a partir do estado k. Portanto, os estados i e k se comunicam.
Quando dois estados são comunicantes entre si, podemos dizer que fazem parte da mesma classe, e no caso em que todos os estados pertencem a uma única classe, é dita ser **irredutível**.

c) **Transiente**: Uma cadeia é dita transiente, quando ao entrar neste estado, o processo pode nunca retornar novamente ao mesmo. Portanto, o estado i é transiente, se e somente se, existe um estado j, $j \neq i$, que é alcançável a partir do estado i mas não vice-versa, ou seja, o estado i não é alcançável a partir do estado j. Por consequência, um estado transiente será visitado somente um número finito de vezes.

d) **Recorrente**: uma cadeia é dita recorrente, quando entrando neste estado, o processo definitivamente retornará ao mesmo. Quando um estado é recorrente, não pode ser transiente. Considerando o citado acima:

$$I_n = 1, se \quad X_n = i; \quad 0, se X_n \neq i$$

Representando o número de períodos do processo no estado i por:

$$E\left[\sum_{n=0}^{\infty} I_n | X_0 = i\right] = \sum_{n=0}^{\infty} E\left[I_n | X_o = i\right]$$

$$= \sum_{n=0}^{\infty} \mathbb{P}\{X_n = i | X_o = i\}$$

$$\sum_{n=0}^{\infty} P_{ii}^n$$

Dessa forma, pode-se dizer que, ocorre estado recorrente se $\sum_{n=1}^{\infty} P_{ii}^n = \infty$ e estado transiente, se $\sum_{n=1}^{\infty} P_{ii}^n < \infty$.

e) **Absorvente**: ocorre quando, ao entrar neste estado, o processo nunca irá deixá-lo. Portanto, um estado i é absorvente, se e somente se, $P_{ii} = 1$. Desse forma, pode-se dizer que um estado absorvente é um caso especial de um estado recorrente.

3.8 Probabilidades limite

Em diversos casos, estamos interessados nas probabilidades de estados após um número de períodos considerável, ou seja estamos interessados precisamente quando $n \to \infty$. Quando essas probabilidades convergem a valores constantes; tal fato é conhecido como Probabilidades Limite dos estados.

Exemplo 3.6. Segundo uma pesquisa realizada sobre a prescrição de uma droga ao tratamento de pacientes com certa infecção, foram avaliados três laboratórios, respectivamente A, B e C. Esta pesquisa tem como estimativa de prescrição médica a confiança em certo laboratório de 80%, possibilitando aos demais em questão, uma chance de 20% para mudança. Os tratamentos são observados, podendo haver mudança de acordo com a resposta ao tratamento num período de observação estipulado. Foram notadas que as mudanças de indicação em relação aos laboratórios são as seguintes:

- Mudança do laboratório A para os laboratórios B e C são 12% e 8%, respectivamente;

- Mudança do laboratório B para os laboratórios A e C são 15% e 5%, respectivamente;
- Mudança do laboratório C para os laboratórios A e B são 13% e 7%, respectivamente;

$$M = \begin{bmatrix} 0,8 & 0,12 & 0,08 \\ 0,15 & 0,8 & 0,05 \\ 0,13 & 0,07 & 0,8 \end{bmatrix}$$

Para realização dos cálculos, substituiremos os estados A, B e C respectivamente por 1, 2 e 3.
Foi necessária a utilização de um *software* matemático, visando um melhor entendimento e menos dispêndio de tempo para realização dos produtos de Matrizes.
Levando em consideração que uma equipe médica mantém a confiança de prescrição da droga fabricada pelo laboratório B, podemos analisar os vetores estados futuros da seguinte forma:

$$V_2^0 = V_2^1 = [0,15 \quad 0,8 \quad 0,05]$$

$$V_2^2 = V_2^1 \cdot M = [0,15 \quad 0,8 \quad 0,05] \cdot \begin{bmatrix} 0,8 & 0,12 & 0,08 \\ 0,15 & 0,8 & 0,05 \\ 0,13 & 0,07 & 0,8 \end{bmatrix} = [0,247 \quad 0,662 \quad 0,092]$$

$$V_2^3 = V_2^1 \cdot M^2 = [0,15 \quad 0,8 \quad 0,05] \cdot \begin{bmatrix} 0,8 & 0,12 & 0,08 \\ 0,15 & 0,8 & 0,05 \\ 0,13 & 0,07 & 0,8 \end{bmatrix}^2 = [0,308 \quad 0,565 \quad 0,126]$$

$$\vdots$$

$$V_2^{10} = V_2^1 \cdot M^9 = [0,15 \quad 0,8 \quad 0,05] \cdot \begin{bmatrix} 0,8 & 0,12 & 0,08 \\ 0,15 & 0,8 & 0,05 \\ 0,13 & 0,07 & 0,8 \end{bmatrix}^9 = [0,411 \quad 0,358 \quad 0,231]$$

$$V_2^{11} = V_2^1 \cdot M^{10} = [0,15 \quad 0,8 \quad 0,05] \cdot \begin{bmatrix} 0,8 & 0,12 & 0,08 \\ 0,15 & 0,8 & 0,05 \\ 0,13 & 0,07 & 0,8 \end{bmatrix}^{10} = [0,413 \quad 0,352 \quad 0,236]$$

$$V_2^{12} = V_2^1 \cdot M^{11} = [0,15 \quad 0,8 \quad 0,05] \cdot \begin{bmatrix} 0,8 & 0,12 & 0,08 \\ 0,15 & 0,8 & 0,05 \\ 0,13 & 0,07 & 0,8 \end{bmatrix}^{11} = [0,414 \quad 0,347 \quad 0,239]$$

$$\vdots$$

$$V_2^{30} = V_2^1 \cdot M^{29} = [0,15 \quad 0,8 \quad 0,05] \cdot \begin{bmatrix} 0,8 & 0,12 & 0,08 \\ 0,15 & 0,8 & 0,05 \\ 0,13 & 0,07 & 0,8 \end{bmatrix}^{29} = [0,414 \quad 0,340 \quad 0,246]$$

Observa-se que, para todos os períodos maiores que $n = 10$, os vetores-probabilidade se aproxima de um vetor constante, fato esse, chamado de Probabilidade Limite.
Pelo vetor-probabilidade em que $n = 30$, podemos deduzir que a probabilidade de mudança de prescrição médica da droga fabricada pelo laboratório B para o laboratório A é de aproximadamente 41,4%; de permanecer no laboratório B é de aproximadamente 34% e de mudança do laboratório B para C é de aproximadamente 24,6%.

3.9 Problema da Ruína do Jogador

Considerando um jogador que em cada jogo tem probabilidade *p* de ganhar uma unidade e probabilidade *q= 1-p* de perder uma unidade. Assumindo que sucessivas jogadas são independentes, qual é a probabilidade de que, começando com *i* unidades, a fortuna do jogador alcance um referido valor *N* antes de atingir *0*?
Se X_N denota a fortuna do jogador no tempo *n*, então o processo $\{X_n, n = 0, 1, 2, \cdots\}$ é uma cadeia de Markov com probabilidade de transição

$$P_{00} = P_{NN} = 1$$

$$P_{i,i+1} = p = 1 - P_{i,i-1}; \quad i = 1, 2, \cdots, N-1$$

Esta cadeia de Markov tem três classes, ou seja, $\{0\}$, $\{1, 2, \cdots, N-1\}$ e $\{N\}$; sendo a primeira e a terceira, classes recorrentes e a segunda transitória. Uma vez que cada estado transitório é visitado, apenas finitamente muitas vezes, segue-se que, após alguma consideração finita de tempo, o jogador vai atingir seu objetivo de atingir N ou ir perdendo.
Seja $P_i, \quad i = 0, 1, \cdots, N$, denota-se a probabilidade de que, a partir de *i*, a fortuna do jogador eventualmente alcance *N*. Ao condicionar o resultado do jogo inicialmente, obtemos:

$$P_i = pP_{i+1} + qP_{i-1}, \quad i = 1, 2, \cdots, N-1$$

Ou equivalente, uma vez que $p + q = 1$

$$pP_i + qP_i = pP_{i+1} + qP_{i-1}$$

Portanto, como $P_0 = 0$, obtemos da linha precedente que

$$P_2 - P_1 = \frac{q}{p}(P_1 - P_0) = \frac{q}{p}P_1$$

ou

$$P_{i+1} - P_i = \frac{q}{p}\left(P_i - P_{i-1}\right), \quad i = 1, 2, \cdots, N-1$$

Portanto, como $P_0 = 0$, obtemos da linha precedente que

$$P_2 - P_1 = \frac{q}{p}(P_1 - P_0) = \frac{q}{p}P_1$$

$$P_3 - P_2 = \frac{q}{p}(P_2 - P_1) = \left(\frac{q}{p}\right)^2 P_1$$

$$\vdots$$

$$P_i - P_{i-1} = \frac{q}{p}(P_{i-1} - P_{i-2}) = \left(\frac{q}{p}\right)^{i-1} P_1$$

$$\vdots$$

$$P_N - P_{N-1} = \frac{q}{p}(P_{N-1} - P_{N-2}) = \left(\frac{q}{p}\right)^{N-1} P_1$$

A adição da primeira consideração *i -1* dessas equações produz

$$P_i - P_1 = P_1 \left[\left(\frac{q}{p}\right) + \left(\frac{q}{p}\right)^2 + \cdots + \left(\frac{q}{p}\right)^{i-1}\right]$$

ou

$$P_i = \begin{cases} \frac{1-\left(\frac{q}{p}\right)^i}{1-\left(\frac{q}{p}\right)} P_1; & se \quad \frac{q}{p} \neq 1 \\ iP_1; & se \quad \frac{q}{p} = 1 \end{cases}$$

Agora, usando o fato de que $P_N = 1$, obtemos

$$P_1 = \begin{cases} \frac{1-\left(\frac{q}{p}\right)}{1-\left(\frac{q}{p}\right)^N}; & se \quad p \neq \quad \frac{1}{2} \\ \frac{1}{N}; & se \quad p = \frac{1}{2} \end{cases}$$

e consequentemente

$$P_i = \begin{cases} \frac{1-\left(\frac{q}{p}\right)^i}{1-\left(\frac{q}{p}\right)^N}; \quad se \quad p \neq \quad \frac{1}{2} \\ \frac{i}{N}; \quad se \quad p = \frac{1}{2} \end{cases}$$

Nota-se que, como $N \to \infty$

$$P_i = \begin{cases} 1 - \left(\frac{q}{p}\right)^i; \quad se \quad p > \quad \frac{1}{2} \\ 0 \quad ; se \quad p \leqslant \quad \frac{1}{2} \end{cases}$$

Assim, se $p > \frac{1}{2}$, há uma probabilidade positiva de que a fortuna do jogador irá aumentar indefinidamente; enquanto que se $p \leqslant \frac{1}{2}$, o jogador vai, com probabilidade 1, ir à falência contra um adversário infinitamente rico.

Exemplo 3.7. Num determinado cassino, um jogador resolve apostar no jogo da roleta. Sabe-se que o mesmo possui apenas R$ 50,00 para realizar inicialmente suas apostas.
Em cada rodada ele aposta R$ 25,00 no vermelho, se ocorrer vermelho como resultado ele ganhará R$ 25,00, caso contrário, ele perde seus R$ 25,00. De forma prática, podemos considerar que as chances de ganhar equivalem a 50%, porém deve-se acreditar que de forma real, tal chance seria bem menor.
A regra para analisar tal jogo é que ele para de jogar quando perde tudo ou adquire R$ 75,00 no total. Analisando este processo e seu comportamento a longo prazo em uma Cadeia de Markov, obtemos:

Diagrama de estados

Considerando os valores limites em relação aos acréscimos ou decréscimos de R$ 25,00, temos:
Estado 1 : R$ 0
Estado 2 : R$ 25,00
Estado 3 : R$ 50,00
Estado 4 : R$ 75,00
Baseando-se em conceitos vistos sobre Cadeias de Markov, é importante ressaltar que os estados 1 e 4 são classificados como absorventes, pois uma vez que se entra nestes estados, é impossível sair dos mesmos .

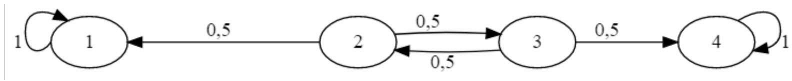


Figura 3.6: Diagrama de estados da ruína do jogador

$$P = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} & P_{34} \\ P_{41} & P_{42} & P_{43} & P_{44} \end{bmatrix} P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0,5 & 0 & 0,5 & 0 \\ 0 & 0,5 & 0 & 0,5 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Probabilidades a longo prazo

$$P^2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0,5 & 0,25 & 0 & 0,25 \\ 0,25 & 0 & 0,25 & 0,5 \\ 0 & 0 & 0 & 1 \end{bmatrix} P^{10} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0,666 & 0,001 & 0 & 0,333 \\ 0,333 & 0 & 0,001 & 0,666 \\ 0 & 0 & 0 & 1 \end{bmatrix} P^{25} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0,667 & 0 & 0 & 0,333 \\ 0,333 & 0 & 0 & 0,667 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

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

Esta matriz não é regular, portanto a melhor forma para generalizarmos seria :

$$P^{\infty} = [\text{ x 0 0 1-x}]$$

Mas, relembrando, nosso jogador iniciou com R$ 50,00 caracterizando início no estado 3. Portanto podemos utilizar começando um vetor de probabilidade da seguinte forma:

$$V = [0\ 0\ 1\ 0]$$

$$VP^{50} = [0\ 0\ 1\ 0\,] \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0,5 & 0 & 0,5 & 0 \\ 0 & 0,5 & 0 & 0,5 \\ 0 & 0 & 0 & 1 \end{bmatrix}^{50}$$

$$VP^{50} \approx \begin{bmatrix} \frac{1}{3} & 0 & 0 & \frac{2}{3} \end{bmatrix}$$

Portanto, quando iniciando com R$ 50,00, nosso referido jogador tem $\frac{1}{3}$ de chance de ir ao estado 1 e $\frac{2}{3}$ de chance de adquirir R$ 25,00, ou seja, chegar ao estado 4.
Se o jogador iniciasse com R$ 25,00, ele iniciaria no estado 2. Portanto, poderia ser representado por um vetor de probabilidade da seguinte forma:

$$V = [0\ 1\ 0\ 0]$$

$$VP^{50} = [0\ 1\ 0\ 0\,] \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0,5 & 0 & 0,5 & 0 \\ 0 & 0,5 & 0 & 0,5 \\ 0 & 0 & 0 & 1 \end{bmatrix}^{50}$$

$$VP^{50} \approx \begin{bmatrix} \frac{2}{3} & 0 & 0 & \frac{1}{3} \end{bmatrix}$$

Porém, iniciando com R$ 25,00 nosso referido jogador tem $\frac{2}{3}$ de chance de ir ao estado 1 e $\frac{1}{3}$ de adquirir R$ 25,00.

4 Cadeias de Markov Ocultas

Num Modelo Markoviano dito padrão ou regular, é possível ao observador enxergar o estado desejado. Entre os estados existem as probabilidades de transições de estados, bem como a referida distribuição de probabilidade em relação aos possíveis resultados. Tais itens são chamados de parâmetros.
Tratando-se de um Modelo de Markov Oculto, os parâmetros até então tidos como observáveis no modelo de Markov regular, se tornam ocultos e também algo a serem determinados. O desafio é determinar os parâmetros ocultos a partir dos parâmetros observáveis. Tal modelo é conhecido pelas importantes aplicações em diversos ramos da Ciência, entre eles, diversos tipos de séries temporais e Biologia computacional, também conhecida como Bioinformática.

4.1 Abordando um exemplo

Suponha que queremos determinar a temperatura média anual em um determinado local da Terra ao longo de uma série de anos. Para tornar essa busca mais interessante, consideraremos que num passado distante aconteciam possíveis verificações erradas a respeito do mesmo , numa época antes da invenção dos termômetros. Uma vez que não podemos voltar no tempo, em vez disso, procuramos indícios indiretos da temperatura.
Para simplificar o problema, consideramos apenas duas temperaturas anuais,sendo elas quente ou fria, representadas por H ou C respectivamente . Supondo que dados meteorológicos contemporâneos indiquem que a probabilidade de um ano quente ser seguido por outro ano quente seja 0,7 e a probabilidade de que um ano frio seja seguido por outro ano frio é 0,6.
Vamos supor que essas probabilidades também aconteceram no passado distante citado anteriormente, sendo estado 1 representado por quente e estado 2 por frio, as informações podem ser resumidas como:

$$\begin{bmatrix} 0,7 & 0,3 \\ 0,4 & 0,6 \end{bmatrix}.$$

A pesquisa atual apresenta uma correlação entre os estados e três níveis de escalas de graduação. Para simplificar, consideramos apenas três escalas diferentes, chamadas de anéis, Pequeno, Médio e Grande, ou S, M e L, respectivamente. Estes anéis são obtidos através de um corte transversal ao tronco de uma árvore, agindo como fonte de observações e dados associados ao estudo da temperatura.
Finalmente, supondo que com base nas evidências meteorológicas, a relação probabilística entre a temperatura anual e as escalas de graduação são representadas por

$$\begin{bmatrix} 0,1 & 0,4 & 0,5 \\ 0,7 & 0,2 & 0,1 \end{bmatrix}.$$

Onde, as linhas representam quente (H) e frio (C) sucessivamente, assim como as colunas estão ordenadas em S, M e L sucessivamente.
Para este sistema, o estado é a temperatura média anual H ou C. A transição de um estado para o outro é um processo de Markov (de ordem 1), uma vez que o próximo estado depende apenas do estado atual e das probabilidades fixadas . No entanto, os estados estão escondidos, uma vez que não podemos observar diretamente a temperatura no passado.

Embora não possamos observar o estado (temperatura) no passado, podemos observar o tamanho dos anéis envolvidos. Os anéis fornecem informações probabilísticas a respeito da temperatura. Desde que os estados estejam "escondidos", este tipo de sistema é visto como um *Hidden Markov Model* "Modelo de Markov Oculto" (HMM). O objetivo é fazer uso efetivo e eficiente das informações observáveis de modo a ganhar a introspecção em vários aspectos do processo de Markov.
Sendo a matriz de transição de estado A, dada por:

$$\begin{bmatrix} 0,7 & 0,3 \\ 0,4 & 0,6 \end{bmatrix}$$

e a matriz de observação B sendo:

$$\begin{bmatrix} 0,1 & 0,4 & 0,5 \\ 0,7 & 0,2 & 0,1 \end{bmatrix}.$$

Neste exemplo, supomos que a distribuição de estado inicial, denotada por π, é

$$\pi = \begin{bmatrix} 0,6 & 0,4 \end{bmatrix}$$

As matrizes π, A e B são *estocásticas de linha*, significando que cada elemento é uma probabilidade e os elementos de cada linha somam 1, ou seja, cada linha é uma distribuição de probabilidade.
Agora considerando um período particular de quatro anos de interesse do passado distante, para o qual observamos a série de anéis S, M, S, L.

Figura 4.1: Anéis de crescimento de árvores [5]

Aceitando 0 representar S, 1 representar M e 2 representar L, esta sequência de observação é

$$\mathcal{O} = (0, 1, 0, 2)$$

Podemos querer determinar a sequência de estado mais provável do processo de Markov dada observação. Isto é, podemos querer saber as temperaturas médias mais prováveis durante o período de quatro anos de interesse. Isto não é tão claro quanto parece, uma vez que há provavelmente diferentes interpretações possíveis de "provavelmente". Por um lado, pode-se razoavelmente definir "provavelmente" como a sequência de estado com maior probabilidade dentre todas as possíveis sequências de estado, ocorrer a de comprimento quatro. Por outro lado, podemos razoavelmente definir "provavelmente" como a sequência de estado que maximiza o número esperado de estados corretos. HMM pode ser usado para encontrar esta sequência.

Apresentamos um dos aspectos mais desafiadores dos HMM, ou seja, a notação. Em seguida, discutimos os três problemas fundamentais relacionados aos HMM e fornecemos algoritmos para suas soluções. Consideramos também algumas questões computacionais críticas que serão abordadas ao escrever qualquer programa de computador que aborde HMM. Concluímos com um exemplo, que não requer qualquer conhecimento especializado, mas ilustra bem a força da abordagem HMM.

4.2 Notação

Consideremos:

- T = Comprimento da sequência de observação
- N = Número de estados no modelo
- M = Número de símbolos de observação
- $Q = q_0, q_1, ..., q_{N-1}$ = Estados distintos do processo de Markov
- V = 0, 1, ..., M – 1 = conjunto de possíveis observações
- A = Probabilidades de transição de estado
- B = Matriz de probabilidade de observação
- π = Distribuição de estado inicial
- $O = (O_0, O_1, ..., O_{T-1})$ = Sequência de observação

As observações são sempre denotadas por $\{0, 1, ..., M - 1\}$, uma vez que isso simplifica a notação sem perda de generalidade. Isto é, $O_i \in V$ para $i = 0, 1, ..., T - 1$.

Um modelo genérico oculto de Markov é ilustrado na Figura 4.1, onde os X_i representam a sequência de estado oculto e toda outra notação é como dado acima. O processo de Markov está escondido acima da linha tracejada e, é determinado pelo estado atual e pela matriz A. Somente podemos observar os O_i, que estão relacionados com os estados (ocultos) do processo de Markov pela matriz B.

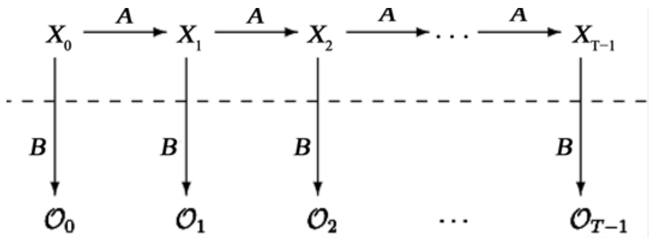


Figura 4.2: Modelo Oculto de Markov

Para o exemplo de temperatura anterior, com a sequência de observações , temos $T = 4$; $N = 2$; $M = 3$; $Q = H, C$; $V = 0, 1, 2$ (onde considerar os 0, 1, 2 representam pequeno, médio e grande anéis, respectivamente). Neste caso, as Matrizes A, B e π são dadas por (3), (4) e (5), respectivamente. A matriz $A = \{a_{ij}\}$ é N x N com

$$a_{ij} = \mathbb{P}(\text{estado } q_j \text{ em t + 1}|\text{estado } q_i \text{ em t})$$

e A é linha estocástica. Observe que as probabilidades a_{ij} são independentes de t. A matriz $B = \{bj(k)\}$ é um N x M com

$$b_j(k) = \mathbb{P}(\text{observação k em t} \mid \text{estado } q_j \text{ em t})$$

Como a matriz A, a matriz B é de linha estocástica e as probabilidades $b_j(k)$ são independentes de t. A notação incomum $b_j(k)$ é padrão no mundo HMM.
Um HMM é definido por A, B e π (e, implicitamente, pelas dimensões N e M). O HMM é denotado por $\lambda = (A, B, \pi)$.
Considere uma sequência genérica de estado de comprimento quatro

$$X = (x_0, x_1, x_2, x_3)$$

Com observações correspondentes

$$O = (O_0, O_1, O_2, O_3)$$

Então π_{x0} é a probabilidade de começar no estado x_0. Além disso, $b_{x0}(O_0)$ é a probabilidade de iniciarmos observando O_0 e $a_{x0,x1}$ é a probabilidade de transitar do estado x_0 para o estado x_1. Continuando, vemos que a probabilidade da sequência de estados X é dada por

$$\mathbb{P}(X, O) = \pi_{x_0} b_{x_0}(O_0) a_{x_0,x_1} b_{x_1}(O_1) a_{x_1,x_2} b_{x_2}(O_2) a_{x_2,x_3} b_{x_3}(O_3)$$

Considere novamente o exemplo de temperatura na Seção 4.1 com a sequência de observação $O = (0, 1, 0, 2)$, conforme dado . Usando-a, podemos calcular, digamos, $\mathbb{P}(HHCC)$ = 0,6·(0,1)·(0,7)·(0,4)·(0,3)·(0,7)·(0,6)·(0,1) = 0,000212.
Da mesma forma, podemos calcular diretamente a probabilidade de cada possível sequência em relação ao estado de comprimento quatro, assumindo a sequência de observação dada. Apresentamos esses resultados na seguinte tabela, onde as probabilidades na última coluna são normalizadas de modo que a soma seja igual a 1.
Para obter a sequência de estado ótima, deve-se escolher a sequência com maior probabilidade, ou seja, $CCCH$. Para obter o melhor no sentido HMM, escolhemos o símbolo mais provável em cada posição. Para isso somamos as probabilidades na tabela que têm um H na primeira posição. Fazendo isso, temos a probabilidade (normalizada) de H na primeira posição que é 0,18817 e, portanto, a probabilidade de C na primeira posição é 0,81183. Assim, o HMM escolhe o primeiro elemento da sequência ótima para ser C. Repetimos isto para cada elemento da sequência, obtendo as probabilidades na tabela 2.
A partir da tabela 2, temos que a sequência ótima no sentido HMM é $CHCH$.

Estado	Probabilidade	Probabilidade Normalizada
HHHH	0,000412	0,042787
HHHC	0,000035	0,003635
HHCH	0,000706	0,073320
HHCC	0,000212	0,022017
HCHH	0,000050	0,005193
HCHC	0,000004	0,000415
HCCH	0,000302	0,031364
HCCC	0,000091	0,009451
CHHH	0,001098	0,114031
CHHC	0,000094	0,009762
CHCH	0,001882	0,195451
CHCC	0,000564	0.058573
CCHH	0,000470	0,048811
CCHC	0,000040	0,004154
CCCH	0,002822	0,293073
CCCC	0,000847	0,087963

Tabela 4.1: Probabilidades de sequências de estados

4.3 Os três problemas fundamentais

Existem três problemas fundamentais que podemos resolver usando HMM. Aqui, descreveremos esses três problemas e, na próxima seção, fornecemos algoritmos para soluções.

	Elemento			
	0	1	2	3
$\mathbb{P}(H)$	0,188182	0,519576	0,228788	0,804029
$\mathbb{P}(C)$	0,811818	0,480424	0,771212	0,195971

Tabela 4.2: Probabilidades HMM

4.4 Problema 1

Dado o modelo $\lambda = (A, B, \pi)$ e uma sequência de observações O, buscando $\mathbb{P}(O|\lambda)$. Aqui queremos determinar a probabilidade da sequência observada O, dado o modelo.

4.5 Problema 2

Dado $\lambda = (A, B, \pi)$ e uma sequência de observação O, busca-se uma sequência de estado ótima para o Processo de Markov subjacente. Em outras palavras, queremos descobrir a parte oculta do Modelo de Markov Oculto.

4.6 Problema 3

Dada uma sequência de observação O e as dimensões N e M, e o modelo $\lambda = (A, B, \pi)$ que maximiza a probabilidade de O. Isso pode ser visto como um modelo de treinamento para melhorar dados observados. Alternativamente, podemos ver isso como uma subida de colina (discreta) no parâmetro espaço representado por A, B e π.

4.7 Discussão

Considere o problema do reconhecimento de fala (ocorre por ser uma das mais conhecidas aplicações de HMM). Podemos usar a solução para o Problema 3 para treinar um HMM, digamos, λ_0 para reconhecer a palavra falada *NÃO* e treinar outro HMM, digamos, λ_1 para reconhecer a palavra *SIM*. Em seguida, dada uma palavra falada desconhecida, podemos usar a solução para o Problema 1 para marcar esta palavra contra λ_0 e também contra λ_1 para determinar se é mais provável *NÃO, SIM ou NENHUM*. Neste caso, não é necessário resolver o problema 2, mas é possível que tal solução que descobre os estados ocultos possa fornecer uma visão adicional sobre o modelo de fala subjacente.

4.8 As três soluções

4.8.1 Solução para o problema 1

Seja $\lambda = (A, B, \pi)$ um modelo dado e seja $O = (O_0, O_1, \cdots, O_{T-1})$ uma série de observações. Queremos encontrar $\mathbb{P}(O|\lambda)$.
Seja $X = (x_0, x_1, \cdots, x_{T-1})$ uma sequência de estados. Então, pela definição de B, temos

$$\mathbb{P}(O|X, \lambda) = bx_o(O_0)b_{x_1}(O_1) \cdots b_{x_{T-1}}(O_{T-1})$$

E pela definição de π e A segue-se que

$$\mathbb{P}(X|\lambda) = \pi_{x_0} a_{x_0,x_1} a_{x_1,x_2} \cdots a_{x_{T-2},x_{T-1}}.$$

Desde que

$$\mathbb{P}(O, X|\lambda) = \frac{\mathbb{P}(O \cap X \cap \lambda)}{\mathbb{P}(\lambda)}$$

e

$$\mathbb{P}(O|X, \lambda)\mathbb{P}(X|\lambda) = \frac{\mathbb{P}(O \cap X \cap \lambda)}{\mathbb{P}(X \cap \lambda)} \cdot \frac{\mathbb{P}(X \cap \lambda)}{\mathbb{P}(\lambda)} = \frac{\mathbb{P}(O \cap X \cap \lambda)}{\mathbb{P}(\lambda)}$$

temos

$$\mathbb{P}(O, X|\lambda) = \mathbb{P}(O|X, \lambda)\mathbb{P}(X|\lambda).$$

Ao somar todas as possíveis sequências de estado, obtemos

$$\begin{aligned}
\mathbb{P}(O|\lambda) &= \sum_X \mathbb{P}(O, X|\lambda) \\
&= \sum_X \mathbb{P}(O|X, \lambda)\mathbb{P}(X|\lambda) \\
&= \sum_X \pi_{x_0} b_{x_0}(O_0) a_{x_0,x_1} b_{x_1}(O_1) \cdots a_{x_{T-2},x_{T-1}} b_{x_{T-1}}(O_{T-1})
\end{aligned}$$

No entanto, esta computação direta é geralmente inviável, uma vez que requer cerca de $2TN^T$ multiplicações. A força de HMM deriva em grande parte do fato de que existe um algoritmo eficiente para obter o mesmo resultado.
Para encontrar $\mathbb{P}(O|\lambda)$, utiliza-se o chamado *algoritmo progressivo*, ou *passo-α*. Para $t = 0, 1, \cdots, T-1$ e $i = 0, 1, \cdots, N-1$, define

$$\alpha_t(i) = \mathbb{P}(O_0, O_1, \cdots, O_t, x_t = q_i|\lambda)$$

Então $\alpha_t(i)$ é a probabilidade da sequência de observação parcial até o tempo t, onde o Processo de Markov está no estado q_i no tempo t. O discernimento crucial aqui é que o $\alpha_t(i)$ pode ser computado recursivamente como segue.
1. Seja $\alpha_0(i) = \pi_i b_i(O_0)$, para $i = 0, 1, \cdots, N-1$
2. Para $t = 1, 2, \cdots, T-1$ e $\quad i = 0, 1, \cdots, N-1$:

$$\alpha_t(i) = \left[\sum_{j=0}^{N-1} \alpha_{t-1}(j) a_{ji}\right] b_i(O_t).$$

Sendo:

$$\mathbb{P}(O|\lambda) = \sum_{i=0}^{N-1} \alpha_{T-1}(i).$$

O algoritmo progressivo precisa de N^2T operações.

4.8.2 Solução para o Problema 2

Dado o modelo $\lambda = (A, B, \pi)$ e uma sequência de observações O, nosso objetivo é encontrar a sequência de estado mais provável. Como mencionado acima, há diferentes interpretações possíveis de "provavelmente". Para HMM queremos maximizar o número esperado de estados corretos. Em contraste, um programa dinâmico encontra o caminho global de maior pontuação. Como vimos, essas soluções não são necessariamente as mesmas.
Primeiro, definimos o *algoritmo retrocesso*, ou *passo-*β. Isso é análogo ao *passo-* α discutido acima, distinguindo-se porque ele começa no final e funciona de volta para o início.
Para $t = 0, 1, \cdots, T-1$ e $i = 0, 1, N-1$, define

$$\beta_t(i) = \mathbb{P}(O_{t+1}, O_{t+2}, \cdots, O_{T-1}|x_t = q_i, \lambda)$$

Então o $\beta_t(i)$ pode ser computado recursivamente (e eficientemente) como segue.
1. Seja $\beta_{T-1}(i) = 1$, para $i = 0, 1, N-1$.
2. Para $t = T-2, T-3, \cdots, 0$ e $i = 0, 1, N-1$, computando

$$\beta_i = \sum_{j=0}^{N-1} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)$$

Para $t = 0, 1, \cdots, T-1$ e $i = 0, 1, \cdots, N-1$, define

$$\gamma_t(i) = \mathbb{P}(x_t = q_i|O, \lambda)$$

Uma vez que $\alpha_t(i)$ mede a probabilidade relevante até o tempo t e $\beta_t(i)$ mede a probabilidade após o tempo t,

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\mathbb{P}(O|\lambda)}$$

Vale lembrar que o denominador $\mathbb{P}(O|\lambda)$ é obtido pela soma $\alpha_{T-1}(i)$ sobre i. Pela definição de $\gamma_t(i)$ segue-se que o estado mais provável no tempo t é o estado q_i para o qual $\gamma_t(i)$ é máxima, onde o máximo é tomado sobre o índice i.

4.8.3 Solução para o Problema 3

Aqui, queremos ajustar os parâmetros do modelo para aprimorar as observações. Os tamanhos das matrizes N e M são fixadas mas os elementos de A, B e π devem ser determinados, para a condição estocástica de linha. O fato de que podemos reestimar o próprio modelo é um dos aspectos mais surpreendentes de HMM.
Para $t = 0, 1, \cdots, T-2$ e $i, j \in \{0, 1, \ldots, N-1\}$, define-se "$(di - \gamma)''$ como

$$\gamma_t(i,j) = \mathbb{P}(x_t = q_i, x_{t+1} = q_j|O, \lambda).$$

Então $\gamma_t(i,j)$ é a probabilidade de estar no estado q_i no tempo t e em trânsito para o estado q_j no tempo $t+1$. Os "di-gammas" podem ser escritos em termos de α, β, A e B como

$$\gamma_t(i,j) = \frac{\alpha_t(i)a_{i,j}b_j(O_{t+1})\beta_{t+1}(j)}{\mathbb{P}(O|\lambda)}$$

Para $t = 0, 1, \ldots, T-2$, o $\gamma_t(i)$ e $\gamma_t(i,j)$ são relacionados por

$$\gamma_t(i) = \sum_{j=0}^{N-1} \gamma_t(i,j)$$

Considerando o γ e "di-gama" verificamos abaixo que o modelo $\lambda = (A, B, \pi)$ pode ser reestimada do seguinte modo.

1. Para $i = 0, 1, \ldots, N-1$, temos

$$\pi_i = \gamma_0(i)$$

2. Para $i = 0, 1, \ldots, N-1$, e $j = 0, 1, \ldots, N-1$, computa-se

$$a_{ij} = \frac{\sum_{t=0}^{T-2} \gamma_t(i,j)}{\sum_{t=0}^{T-2} \gamma_t(i)}$$

3. Para $j = 0, 1, \ldots, N-1$ e $k = 0, 1, \ldots, M-1$, computa-se

$$b_j(k) = \frac{\sum_{t \in \{0,1,\ldots,T-1\} O_t = k} \gamma_t(j)}{\sum_{t=0}^{T-1} \gamma_t(j)}$$

O numerador do a_{ij} reestimado pode ser visto para dar o número esperado de transições do estado q_i para o estado q_j, enquanto o denominador é o número esperado de transições de q_i para qualquer estado. Então, a razão é a probabilidade de transitar do estado q_i para o estado q_j, que é o valor desejado de a_{ij}.
O numerador do $b_j(k)$ reestimado é o número esperado de vezes que o modelo é no estado q_j com observação k, enquanto o denominador é o número esperado de vezes que o modelo está no estado q_j. A razão é a probabilidade de observar o símbolo k, dado que o modelo está no estado q_j, que é o valor desejado de $b_j(k)$.

A reestimação é um processo iterativo. Primeiro, inicializamos $\lambda = (A, B, \pi)$ com um melhor adivinho ou, se nenhuma suposição razoável estiver disponível, escolhemos valores aleatórios tais que $\pi_i \approx \frac{1}{N}$ e $a_{ij} \approx \frac{1}{N}$ e $b_j(k) \approx \frac{1}{M}$. É crítico que A, B e π seja randomizado, uma vez que valores uniformes resultarão em um máximo local a partir do qual o modelo não pode subir. Como sempre, π, A e B devem ser estocásticos de linha. A solução para o problema 3 pode ser resumida como se segue.

1. Inicializar, $\lambda = (A, B, \pi)$
2. Calcular $\alpha_t(i), \beta_t(i), \gamma_t(i,j) e \gamma_t(i)$
3. Reestimar o modelo $\lambda = (A, B, \pi)$
4. Se $\mathbb{P}(O|\lambda)$ aumentar consideravelmente, redirecione-se para 2.

Claro, pode ser desejável parar se $\mathbb{P}(O|\lambda)$ não aumentar pelo menos em alguns predeterminados e, ou para definir um número máximo de iterações.

Exemplo 4.1. Numa sala de aula, um professor de Matemática propôs a seus alunos a seguinte ideia em relação a uma atividade a ser realizada envolvendo perguntas a respeito de alguns tópicos de Matemática. Os referidos tópicos eram simbolizados numericamente conforme a distribuição a seguir: Equações (1), Probabilidades (2), Funções (3), Logaritmos (4), Trigonometria (5), Geometria Analítica (6) e Análise Combinatória (7). Os alunos eram submetidos às questões de acordo com os tópicos através de sorteio, sabe-se que havia N urnas e dentro das mesmas estavam as M bolas numeradas de 1 a 7 conforme representadas anteriormente.
O processo pode ser descrito com detalhes da seguinte maneira: Inicialmente uma urna é escolhida aleatoriamente, uma bola numerada de 1 a 7 é escolhida ao acaso

observando-se seu respectivo número e por sua vez, o número se torna a observação do processo. Recoloca-se a bola na urna de origem, e então, uma nova urna é escolhida de forma aleatória. Este processo é realizado (repetido) muitas vezes, gerando uma sequência de observações constituídas pelos números das bolas.
É possível modelar esse processo por HMM, pois há duas formações de um processo estocático: o processo observável "número de bolas" e o processo não observável "escolha aleatória da urna". Os elementos que constituem um HMM em relação ao referido exemplo são:

1. N representa o número de estados do modelo, que neste caso são representados pelas urnas;

2. $S = \{S_1, S_2, \ldots, S_N\}$ representa o conjunto dos estados individuais do modelo. Neste exemplo, seria cada urna, como por exemplo S_1 seria a primeira urna e S_4 seria a quarta urna;

3. q_t representa o estado no tempo t . Neste exemplo, podemos dizer que $q_2 = S_3$ significa que a segunda urna escolhida aleatoriamente foi a terceira urna;

4. M representa o número de observações distintas por estado;

5. $V = \{v_1, v_2, \ldots, v_M\}$ representa o conjunto de observações individuais, neste exemplo seria o conjunto de bolas de uma urna;

6. $B = \{b_j(k)\}$ representa a distribuição de probabilidade da observação no estado

   Supondo que dentre N urnas, a escolhida foi a urna três na primeira escolha e dentro dela haja duas bolas (1), uma bola (2), duas bolas (4) e uma bola (6), a representação da distribuição de probabilidades das observações neste estado são:

$$b_3(1) = \mathbb{P}(v_1 \text{em t = 1} | q_1 = S_3) = \frac{2}{6} = \frac{1}{3}$$
$$b_3(2) = \mathbb{P}(v_2 \text{em t = 1} | q_1 = S_3) = \frac{1}{6}$$
$$b_3(4) = \mathbb{P}(v_3 \text{em t = 1} | q_1 = S_3) = \frac{2}{6} = \frac{1}{3}$$
$$b_3(6) = \mathbb{P}(v_4 \text{em t = 1} | q_1 = S_3) = \frac{1}{6}$$

7. $\pi = \{\pi_i\}$ representa a distribuição inicial dos estados.

No caso deste exemplo, a probabilidade de escolher aleatoriamente uma urna é considerada igual a todas. Logo:

$$\pi_i = \frac{1}{N} \quad \text{para todo i e j}$$

5 Atividades para salas de aulas

As questões de aprendizagem a seguir visam dinamizar em sala de aula os conteúdos até aqui abordados. O público alvo são alunos do ensino médio, e em especial aqueles que estão no término do processo de aprendizagem sobre probabilidades com conhecimentos prévios em operações com matrizes.
Para tais atividades, é necessário a utilização de um *software* capaz de facilitar as resoluções de potências e multiplicações de matrizes, ficando a escolha ao encargo do professor, respeitando disponibilidades ao mesmo. Utilizamos e recomendamos a utilização do Geogebra.

5.1 Geogebra

Geogebra é um aplicativo matemático composto por conceitos, propriedades geométricas e algébricas. Possui distribuição livre, o que possibilita maior acesso e possibilidades para utilização na aprendizagem de muitos conceitos matemáticos. Neste trabalho, o Geogebra desempenha importante papel como ferramenta de cálculo em relação às multiplicações e potências de matrizes.
Dentre tantas funções contidas neste aplicativo, abordaremos através de simples exemplos, as confecções , o produto de duas matrizes e duas potências de matrizes com o propósito de servir como referências às demais situações de cálculos das probabilidades de transições das Cadeias de Markov.

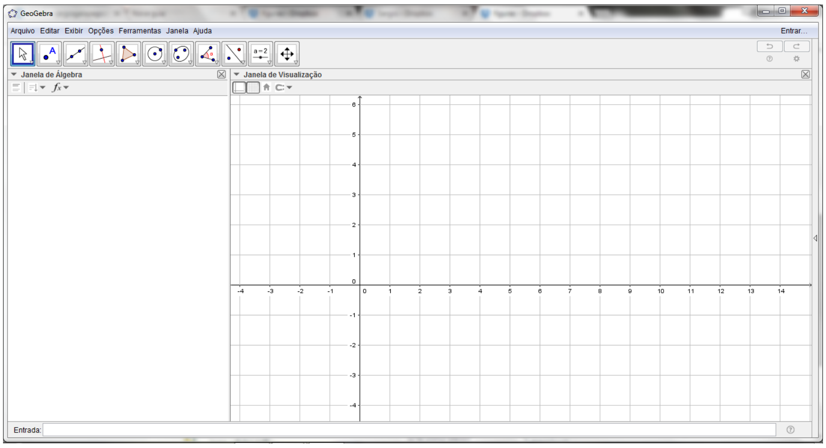


Figura 5.2: Janela principal do Geogebra

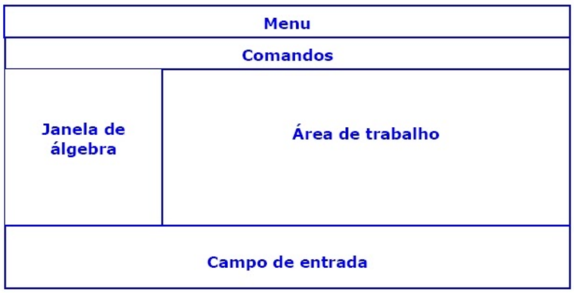


Figura 5.1: Divisões dos comandos e áreas de trabalho no Geogebra

Exemplo 5.1. Para construir uma matriz M de ordem 3, na qual os elementos das linhas serão 1,2 e 3; 4, 5 e 6; 7, 8 e 9 respectivamente, devemos realizar:

- Na área campo de entrada, digiar: $M = \{\{1,2,3\},\{4,5,6\},\{7,8,9\}\}$

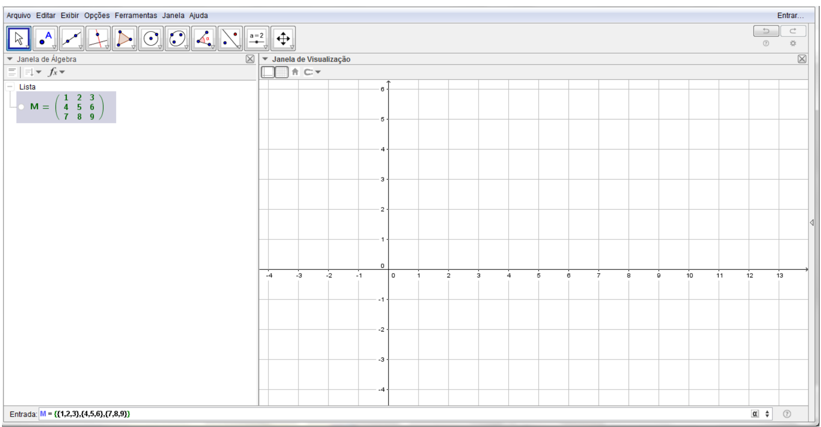


Figura 5.3: Confecção da Matriz M no Geogebra

Para construirmos uma matriz N de ordem 1 X 3 , composta pelos elementos 3, 5 e 5, devemos realizar:

- Na área campo de entrada, digiar: $N = \{\{3, 5, 5\}, \}$

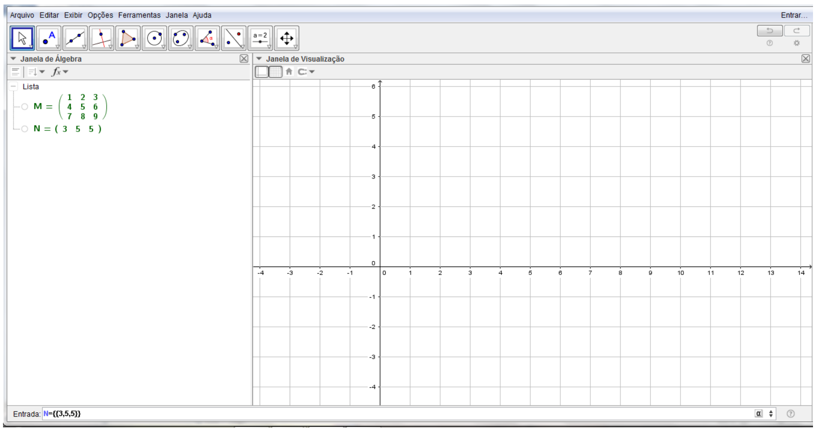


Figura 5.4: Confecção da Matriz N no Geogebra

Exemplo 5.2. Para obtermos o produto de tais matrizes, M e N, devemos utilizar uma letra como representação do produto e o comando de multiplicação "*" (asterisco), da seguinte forma:

- Na área campo de entrada, digitar: $P = N^*M$

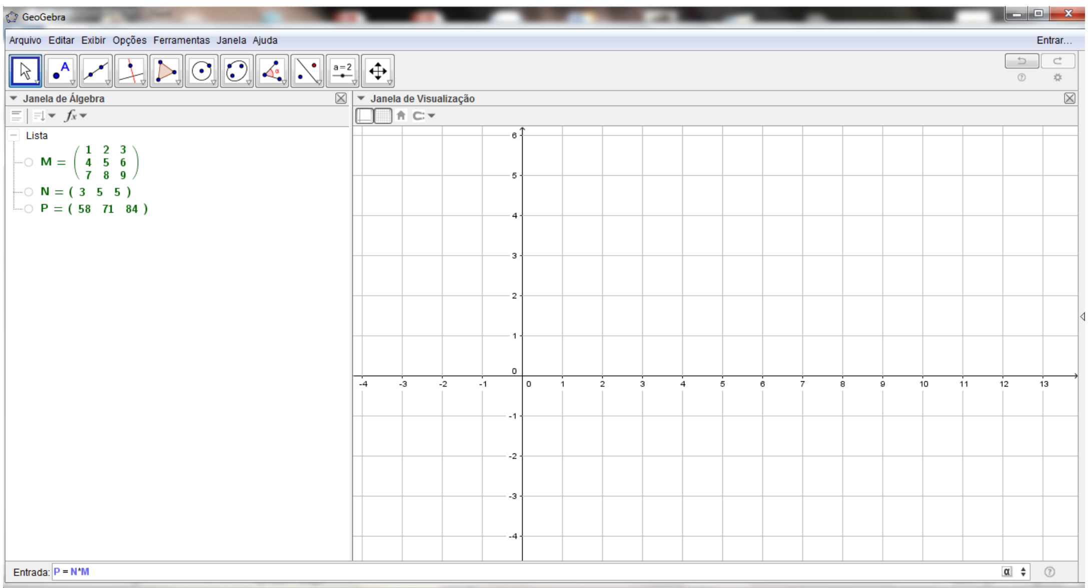


Figura 5.5: Produto de matrizes M e N

Exemplo 5.3. O comando "^" é utilizado na obtenção de potências, neste caso, a matriz M elevada ao quadrado e ao cubo respectivamente, são obtidas da seguinte forma

- Na área de entrada, digitar Q = M^2 para elevá-la ao quadrado e R = M^3 para elevá-la ao cubo.

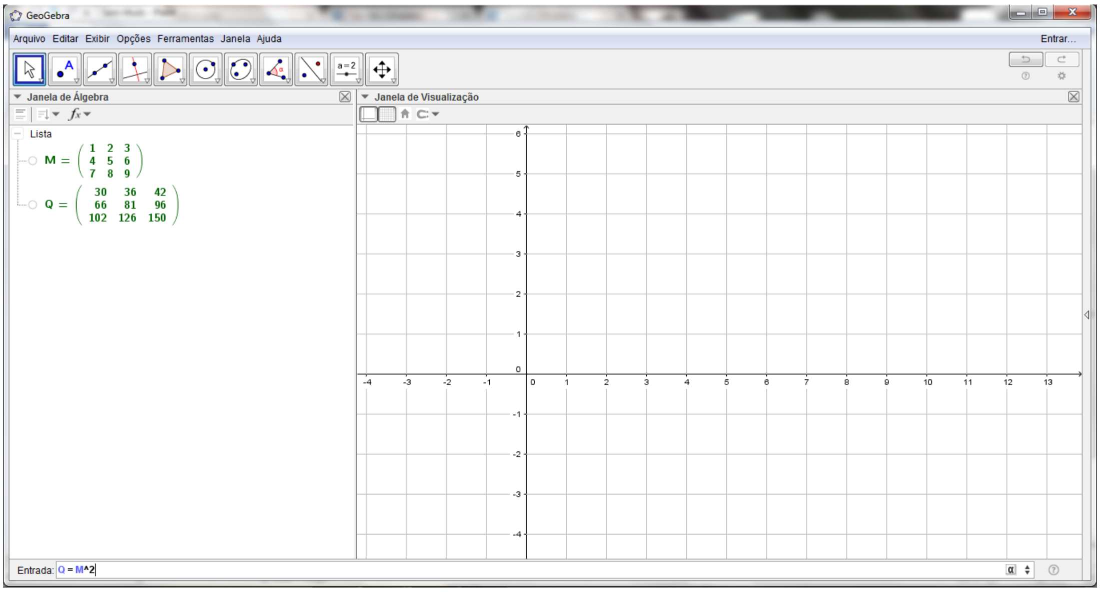


Figura 5.6: Quadrado da Matriz M

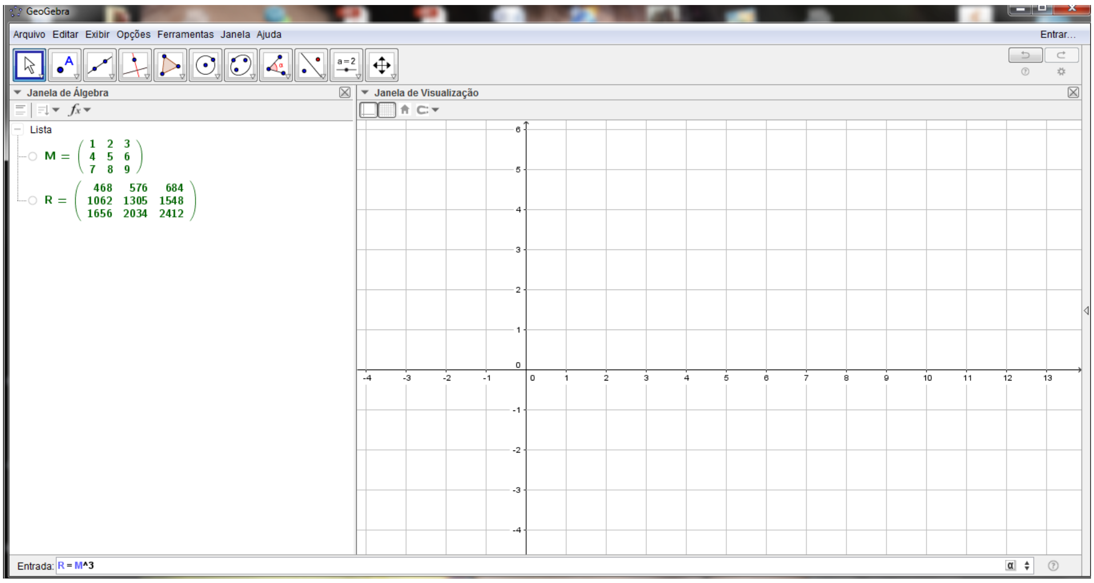


Figura 5.7: Cubo da Matriz M

Para as seguintes atividades serem aplicadas em salas de aula, consideramos três itens de grande importância:

Objetivo

Motivar as operações matriciais utilizando Cadeias de Markov. Dentre tais operações, destaca-se produto e potência de matrizes.

Público-alvo

Alunos do ensino médio, salvo exceções. Em especial, aqueles que estejam no término do processo de aprendizagem sobre probabilidades com conhecimentos prévios em operações matriciais.

Recurso necessário

Utilização de um *software*, neste caso, o Geogebra.

5.2 Atividade 1

Um biólogo analisou o comportamento de um sapo em relação à permanência do mesmo em dois locais distintos , água e terra. Os locais são representadas por 1 e 2 respectivamente. Através de uma série de observações, foi registrado que a probabilidade de permanência no local 1 é de 0,8 e no local 2 é de 0,4. Ocorre também as probabilidades de transições entre os locais , sendo que do local 1 para o local 2 é de 0,2 e do local 2 para o local 1 é de 0,6.

Conforme o enunciado , determine:

1. O vetor de estado inicial, levando em consideração que a observação foi iniciada no local 1.
2. A matriz de transição P.
3. O diagrama de transição entre os locais.
4. Qual o vetor de probabilidade no 10º passo começando no local 1? E começando no local 2?

Soluções 1.

1. E = $\{1, 0\}$
2. $P = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \Rightarrow P = \begin{bmatrix} 0,8 & 0,2 \\ 0,6 & 0,4 \end{bmatrix}$

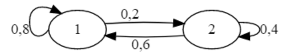


Figura 5.8: Diagrama de transição entre os estados do problema 1

3.

4. $Estado1 : V^{10} = [0{,}75\ 0{,}25]$ $\quad Estado2 : V^{10} = [0{,}75\ 0{,}25]$

5.3 Atividade 2

Admitindo que um jogador tenha inicialmente R$ 1,00 e, com probabilidade $p > 0$, ele ganhe R$ 1,00 e, com probabilidade $1 - p$, perca R$ 1,00. O mesmo participa de apostas, onde a regra do jogo é terminar quando acumular R$3,00 ou perder tudo. Baseado no enunciado , determine:

1. A representação do espaço de estados E em relação à quantia de dinheiro.
2. A matriz P de transição de estados.
3. Observando a matriz de transição acima, quais estados podemos classificá-los como absorventes? Por quê?
4. Construa um diagrama de transição de estados ao referido enunciado.

Soluções 2.

1. E = $\{0, 1, 2, 3\}$
2. $P = \begin{bmatrix} P_{00} & P_{01} & P_{02} & P_{03} \\ P_{10} & P_{11} & P_{12} & P_{13} \\ P_{20} & P_{21} & P_{22} & P_{23} \\ P_{30} & P_{31} & P_{32} & P_{33} \end{bmatrix} \Rightarrow P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1-p & 0 & p & 0 \\ 0 & 1-p & 0 & p \\ 0 & 0 & 0 & 1 \end{bmatrix}$
3. Estado (0) : P_{00} e Estado (3): P_{33}.
   Quando se entra nestes estados, nunca mais os deixam. Os mesmos representam as duas situações do término do jogo.

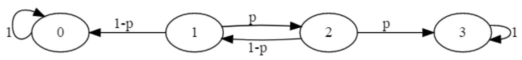


Figura 5.9: Diagrama de transição entre os estados do problema 2

4.

5.4 Atividade 3

Numa cidade, um agente de trânsito é responsável pelo controle e fluidez do tráfego em oito cruzamentos. Sabe-se que seu tempo de permanência em cada cruzamento é de 1 hora, onde o mesmo pode mudar para outro cruzamento adjacente ou permanecer no mesmo. Por questões logísticas à fiscalização, seu local de controle a cada hora é realizada de maneira aleatória, ocorrendo por exemplo, caso ele esteja de acordo com a figura 5.3 no cruzamento 4, o próximo cruzamento pode ser 1, 3, 5 e 7, e em caso de permanência, o 4.
Realizada sua jornada de trabalho, no dia seguinte, a mesma será iniciada no cruzamento em que parou no dia anterior

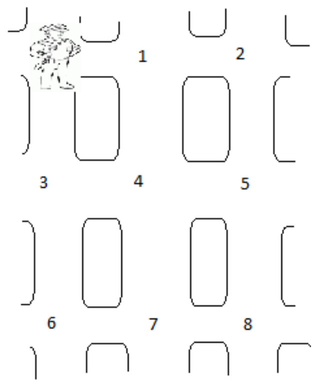


Figura 5.10: Agente de trânsito

Baseando-se no enunciado, determine:

1. O vetor estado inicial.
2. Um esboço do diagrama de transição em relação aos dados.
3. A matriz de transição P.
4. Nesta situação, podemos dizer que existe algum estado absorvente? Além desse, quais outras classificações de estados ocorrem? Justifique.
5. Utilizando recurso computacional como ferramenta de aprendizagem, obtenha os cinco primeiros vetores de estado , e depois os do 20º ao 22º vetores . Faça uma análise do comportamento dos mesmos em relação ao agente de trânsito iniciando no cruzamento 4.

Soluções 3.

1. Como o agente de trânsito inicia no cruzamento 4, o vetor deve ser:
   $V^0 = [00010000]$

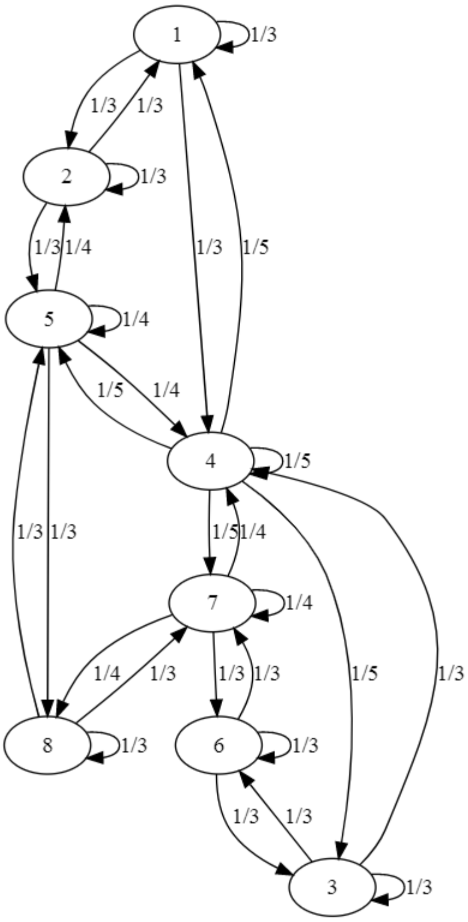


Figura 5.11: Diagrama de transição de estados do agente de trânsito

2.

3. $P = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 \\ \frac{1}{5} & 0 & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & 0 & \frac{1}{5} & 0 \\ 0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{4} \\ 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$

4. Não ocorre estado absorvente. Outras classificações de estados que ocorrem são as seguintes:

- Estado recorrente: uma vez que se entra neste estado, um eventual retorno é assegurado.
- Estado ergódico: uma vez que se entra neste estado, um retorno ao estado é assegurado dentro de um número finito de passos, porém o estado não é periódico e pode voltar antes de qualquer passo n.

n	$X_n^{[1]}$	$X_n^{[2]}$	$X_n^{[3]}$	$X_n^{[4]}$	$X_n^{[5]}$	$X_n^{[6]}$	$X_n^{[7]}$	$X_n^{[8]}$
0	0	0	0	1	0	0	0	0
1	0,2	0	0,2	0,2	0,2	0	0,2	0
2	0,106	0,116	0,106	0,273	0,09	0,116	0,09	0,1
3	0,109	0,109	0,109	0,182	0,132	0,109	0,139	0,104
4	0,108	0,108	0,108	0,179	0,141	0,107	0,141	0,107
5	0,107	0,107	0,108	0,179	0,143	0,109	0,143	0,107
20	0,107	0,107	0,107	0,179	0,143	0,107	0,143	0,107
21	0,107	0,107	0,107	0,179	0,143	0,107	0,143	0,107
22	0,107	0,107	0,107	0,179	0,143	0,107	0,143	0,107

Tabela 5.1: Vetores de estados do agente de trânsito

5.

Analisando os vetores de estados a partir de $n \geq 20$, observa-se a convergência a um vetor de probabilidade. Daí podemos dizer que a probabilidade de estar no estado 4 é 17,9%, nos estados 1, 2, 3, 6 e 8 é de 10,7% e nos estados 5 e 7 é de 14,3%.

Índice

Absorvente, 34
Alcançável, 32

Cadeia de Markov, 25
complementar, 6
Comunicante, 32
conjuntos, 3

Diagrama de Transição de Estados, 26

espaço amostral, 10
evento, 10

interseção, 5

matriz de transição, 29

Processo Markoviano, 25

Recorrente, 33

Transiente, 33

união, 5
unitário, 4
universo, 4

Bibliografia

1 HAGGSTROM, O. **Finite Markov chains and algorithmic applications**. [**sineloco**]: Cambridge University Press, 2002. volume 52.

2 MEYER, P. L. **Probabilidade: aplicaçõesà estatística**. [**sineloco**]: Livro Técnico, 1970.

3 MORETTIN, P. A.; BUSSAB, W. O. **Estatística básica**. [**sineloco**]: Editora Saraiva, 2000.

4 RAMAGE, D. **Hidden Markov models fundamentals**. [**sinelocosinenomine**], 2007.

5 RIZZATO, B. **Anéis de crescimento de árvores podem prever o futuro sobre as próximas mudanças climáticas**. 2016. URL: <`http://www.jornalciencia.com/aneis-de-crescimento-de-arvores-podem-prever-o-futuro-sobre-as-proximas-mudancas-climaticas/`>. Acedido em: 9 de ago. de 2017. Ver página 43.

6 STAMP, M. **A revealing introduction to hidden Markov models**. [**sinelocosinenomine**], 2004.

7 STEINBRUCH, A.; WINTERLE, P. **Introdução à álgebra linear**. [**sineloco**]: McGraw-Hill Medical, 1990.