Cadeias de Markov

Práticas e Aplicações

Tutorial desenvolvido por:
Fabrício Bueno

CEFETSC/Araranguá, 2008

Sumário

1 Introdução....1
2 Diagrama de transição....2
3 Vetor de probabilidade....3
4 Matriz de Transição....5
5 Determinação de probabilidades futuras....6
6 Cadeias ergóticas....8
7 Cadeias Regulares....12
8 Regime estacionário....14
9 Exercícios....16
Apêndice A – Resposta dos exercícios....20
Apêndice B – Modelagem dos exercícios no Octave....22
Apêndice C – Exemplo de aplicação de Cadeias de Markov....23
Bibliografia....33

Índice de Figuras

Figura 1: Diagrama de Transição..2
Figura 2: Diagrama de Transição parcial do exemplo de classes sócio-econômicas...........................4
Figura 3: Diagrama de Transição completo do exemplo de classes sócio-econômicas.......................5
Figura 4: Diagrama de Transição de uma cadeia ergótica...9
Figura 5: Diagrama de Transição de uma cadeia não-ergótica..9
Figura 6: Diagrama de Transição de uma cadeia não-ergótica..10
Figura 7: Diagrama de Transição de cadeia ergótica...11
Figura 8: Diagrama de Transição de cadeia ergótica...11

Índice de Tabelas

Tabela 1: Cruzamento entre genótipos..24
Tabela 2: Probabilidades de cruzamento entre genótipos...31

Cadeias de Markov

1 Introdução

Em atividades industriais, comerciais e humanas, bem como em fenômenos naturais, uma alto grau de incerteza está sempre presente. Portanto, modelos matemáticos probabilísticos, como o processo de Markov[1], que permitam uma previsão estimada do futuro, são bastante úteis na tomada de decisão(Boldrini, 1980).

Dá-se o nome de processo de Markov a um dado fenômeno que possa ser classificado em estados finitos e discretos, e cuja probabilidade de transição entre tais estados, num intervalo de tempo também discreto, dependa apenas do estado corrente e do estado seguinte[2]. À seqüência de estados seguindo este processo dá-se o nome de cadeia de Markov(Shamblin, 1979).

Estas definições podem ser exemplificadas por fenômenos sociais. Suponha a existência de três possíveis classificações sociais para um indivíduo de uma população: classes A, B e C. Têm-se aí três estados discretos, ou seja, não há meio termo entre as classes. A probabilidade de um indivíduo sair da classe C e ir para a classe B pode ser determinada por estudos estatísticos que observem a taxa de indivíduos que, ao longo de um determinado período, migram entre estas classes. Têm-se aí um Processo de Markov. Logo, a transição destes indivíduos entre as classes constitui uma Cadeia de Markov.

[1]Andrei Andrejevitch Marckov (1856-1922) – matemático russo
[2]Definição relativa a cadeia de Markov de primeira ordem. Em cadeias de Markov de segunda ordem, há dependência de dois estados correntes.

Para o entendimento de cadeias de Markov e suas aplicações, é essencial o conhecimento dos conceitos: diagrama de transição, vetor de probabilidade, matriz de transição, cadeia ergótica, cadeia regular, e regime estacionário.

2 Diagrama de transição

O diagrama de transição é uma representação gráfica de uma Cadeia de Markov. Neste diagrama são visualizados os estados (representado por círculos), as transições (representadas por arcos) e as probabilidades das transições. Generalizando, pode-se representar os estados e as probabilidades de transição, respectivamente, por E_i e p_{ij}, onde i e j são um índices que identificam os vários estados possíveis (logo p_{ij} é a probabilidade de haver uma transição do estado E_i para o estado E_j). A partir desta generalização, pode-se desenhar um diagrama, conforme a Figura 1:

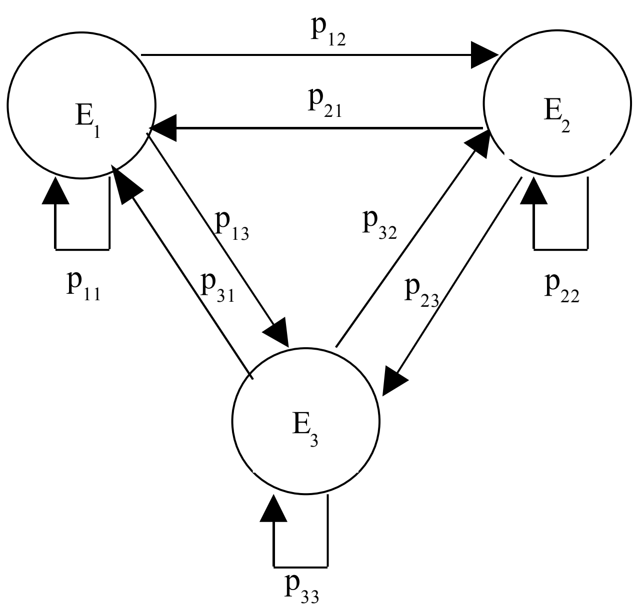


Figura 1: Diagrama de Transição

3 Vetor de probabilidade

O vetor de probabilidade contém as probabilidades de transição de um estado para outros estados em um intervalo de tempo discreto. A generalização do vetor de probabilidade é dada por:

$$V_i = [p_{ij} \quad p_{ik} \quad p_{il}]$$

Onde p_{ij} indica a probabilidade de haver transição do estado E_i para o estado E_j, p_{ik} indica a probabilidade de haver transição do estado E_i para o estado E_k, e p_{il} indica a probabilidade de haver transição do estado E_i para o estado E_l. A soma dos elementos de um vetor de probabilidade sempre será igual a 1.

Utilizando o exemplo das classes sócio-econômicas, um estudo estatístico pode ter determinado que é nula a probabilidade de um indivíduo na classe C ir diretamente para a classe A, a probabilidade de ir para a classe B ser 0,1, e a de continuar na classe C ser 0,9. Logo, o vetor de probabilidade de um indivíduo na classe C será:

$$V_C = [0 \quad 0{,}1 \quad 0{,}9]$$

A Figura 2 ilustra estas o diagrama de transição para estas probabilidades.

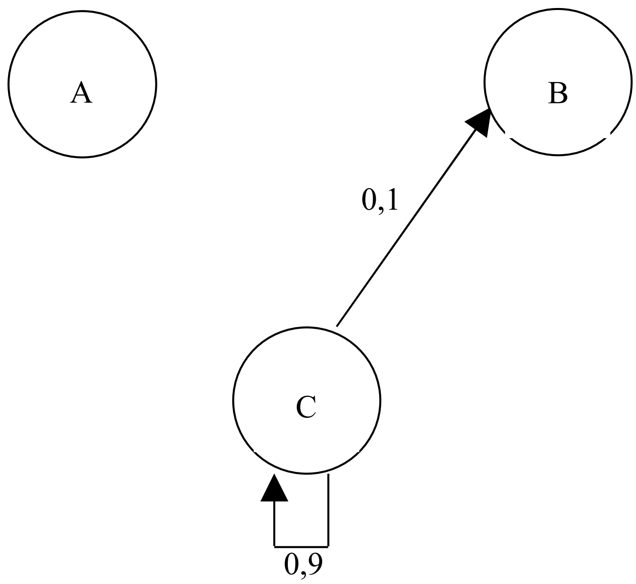


Figura 2: Diagrama de Transição parcial do exemplo de classes sócio-econômicas

Este diagrama apresenta apenas as transições cujas probabilidades são conhecidas. Note que não há um arco representando a transição de C para A, uma vez que há probabilidade nula de esta transição ocorrer.

Uma vez que todas as probabilidades sejam conhecidas, é possível montar um diagrama completo para este exemplo. Considerando os vetores de probabilidade dos estados A e B exibidos a seguir, têm-se o diagrama de transições completo, conforme a Figura 3:

$$V_A=[0{,}3 \quad 0{,}5 \quad 0{,}2]$$

$$V_B=[0{,}3 \quad 0{,}2 \quad 0{,}5]$$

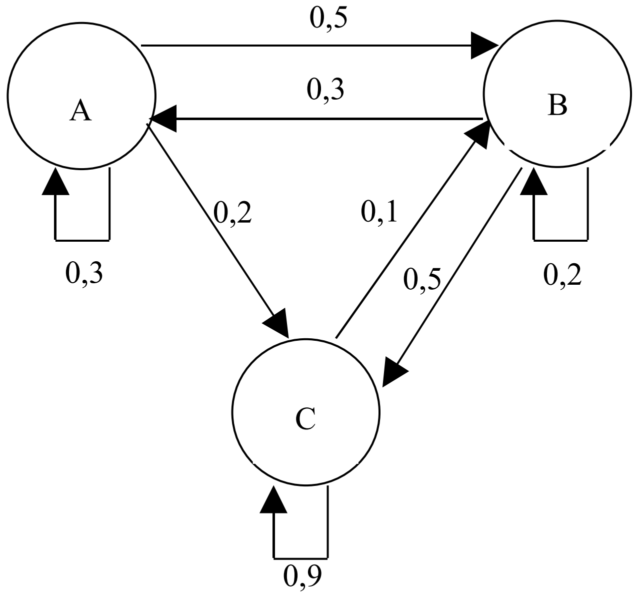


Figura 3: Diagrama de Transição completo do exemplo de classes sócio-econômicas

4 Matriz de Transição

Para cada estado deve haver um vetor de probabilidade. À união de todos os vetores de probabilidade em uma matriz dá-se o nome de matriz de transição. Esta matriz sempre será quadrada(Boldrini, 1980) ou seja, o número de linhas e colunas será sempre equivalente. A seguir é exibido um modelo genérico desta matriz:

$$M = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}$$

Onde o elemento p_{ij}, conforme já citado anteriormente, indica a probabilidade de haver transição do estado E_i para o estado E_j.

No exemplo das classes sociais, considerando os vetores de probabilidade das classes A, B e C respectivamente, [0,3 0,5 0,2], [0,2 0,3 0,5] e [0 0,1 0,9] têm-se a seguinte matriz de transição.

$$M = \begin{bmatrix} 0{,}3 & 0{,}5 & 0{,}2 \\ 0{,}2 & 0{,}3 & 0{,}5 \\ 0 & 0{,}1 & 0{,}9 \end{bmatrix}$$

Nesta matriz, o elemento p_{23}, que equivale a *0,5*, indica a probabilidade 0,5 de haver transição do estado E_B para o estado E_C. Outra interpretação possível é transformar estas probabilidades em percentagens e considerá-las como taxas de transição da população em estudo. Por exemplo, p_{23} = 0,5= 50%, indica que, em um dado momento, 50% da população no estado E_B pode passar para o estado E_C.

5 Determinação de probabilidades futuras

O vetor de probabilidade e a matriz de transição são úteis na determinação de probabilidades ao longo do tempo (Shamblin, 1979). Para tanto, temos a seguinte equação:

$$V_i^t = V_i * M^{t-1}$$

onde

- V é um vetor de probabilidade;
- *t* é o período para o qual se quer obter a probabilidade;
- *i* é índice do estado a partir do qual se quer fazer a previsão;
- M é a matriz de probabilidade.

O vetor resultante desta equação (V_i^t) conterá as probabilidades de transição de um estado E_i após um período *t.*

Por exemplo, caso se queira obter a probabilidade de um indivíduo na classe B ir para a classe A em três anos, têm-se a seguinte equação e sua resolução:

$$V_B^3 = V_B * M^2$$

$$V_B^3 = [0{,}2 \quad 0{,}3 \quad 0{,}5] * \begin{bmatrix} 0{,}3 & 0{,}5 & 0{,}2 \\ 0{,}2 & 0{,}3 & 0{,}5 \\ 0 & 0{,}1 & 0{,}9 \end{bmatrix}^2$$

$$V_B^3 = [0{,}2 \quad 0{,}3 \quad 0{,}5] * \begin{bmatrix} 0{,}19 & 0{,}32 & 0{,}49 \\ 0{,}12 & 0{,}24 & 0{,}64 \\ 0{,}02 & 0{,}12 & 0{,}86 \end{bmatrix}$$

$$V_B^3 = [0{,}084 \quad 0{,}196 \quad 0{,}72]$$

O vetor $V_B^3 = [0{,}084 \quad 0{,}196 \quad 0{,}72]$ indica que um indivíduo na classe B têm uma probabilidade de 0,084 de estar na classe A após três anos. Assim como tem a probabilidade de 0,196 de continuar na classe B, e uma probabilidade de 0,72 de estar na classe C.

Uma alternativa para esta equação é utilizar um vetor identidade (Boldrini, 1980), cujo elemento não nulo será a posição *i*, conforme exibido a seguir:

$$V_i^t = V_i^0 * M^t$$

Logo, podemos resolver o exemplo anterior com esta equação alternativa, chegando ao mesmo resultado.

$$V_B^3 = V_B^0 * M^3$$

$$V_B^3 = [0 \quad 1 \quad 0] * \begin{bmatrix} 0{,}3 & 0{,}5 & 0{,}2 \\ 0{,}2 & 0{,}3 & 0{,}5 \\ 0 & 0{,}1 & 0{,}9 \end{bmatrix}^3$$

$$V_B^3 = [0 \quad 1 \quad 0] * \begin{bmatrix} 0{,}121 & 0{,}24 & 0{,}639 \\ 0{,}084 & 0{,}196 & 0{,}72 \\ 0{,}03 & 0{,}132 & 0{,}83 \end{bmatrix}$$

$$V_B^3 = \begin{bmatrix} 0{,}084 & 0{,}196 & 0{,}72 \end{bmatrix}$$

Adotaremos esta equação ao longo do texto devido a sua maior praticidade.

6 Cadeias ergóticas

Um fenômeno em que haja probabilidade não nula de qualquer estado poder ser alcançado através de uma ou mais transições a partir de qualquer outro estado constitui uma cadeia ergótica. Esta definição pode ser melhor entendida através de diagramas de transição.

A Figura 4, assim como a Figura 3, exibe uma cadeia ergótica, pois qualquer estado pode ser alcançado a partir de outro estado qualquer. A Já na Figura 5 trata-se de uma cadeia não-ergótica, uma vez que o estado E_3 não pode ser alcançado de modo algum. A Figura 6 também traz uma cadeia não-ergótica, pois, uma vez no estado E_3, não se pode ir para nenhum outro estado. Este último caso é chamado de cadeia absorvente(Shamblin, 1979).

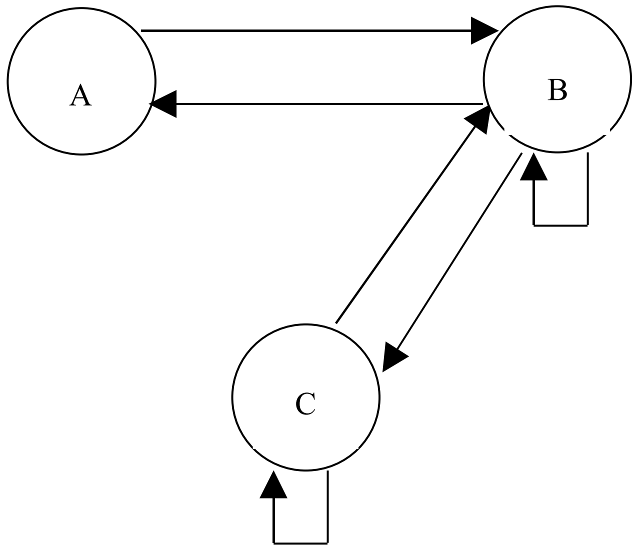


Figura 4: Diagrama de Transição de uma cadeia ergótica

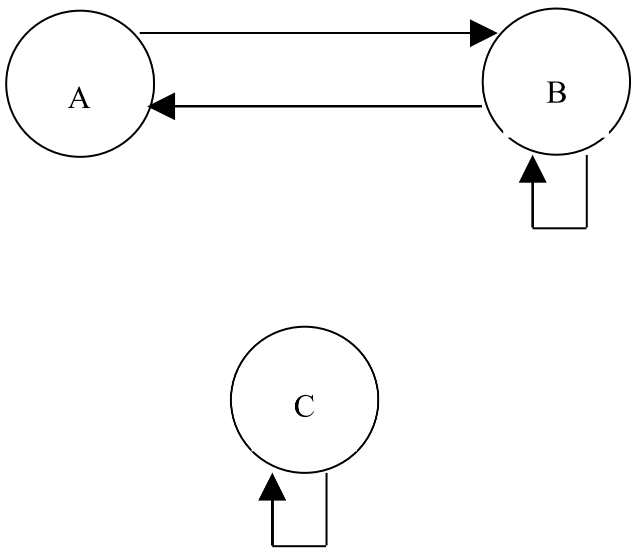


Figura 5: Diagrama de Transição de uma cadeia não-ergótica

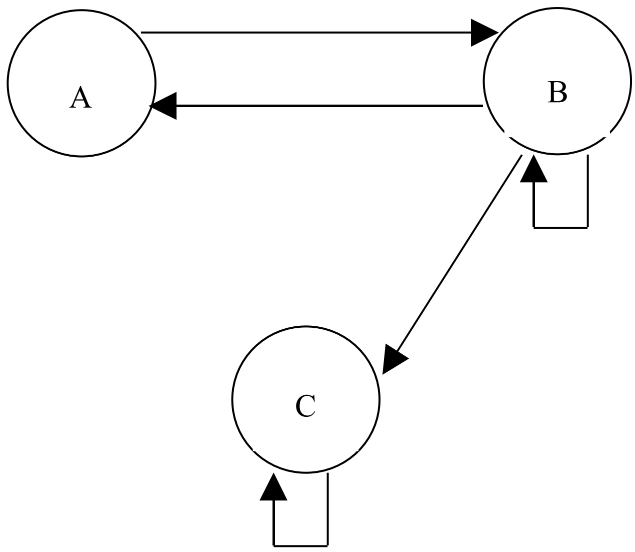


Figura 6: Diagrama de Transição de uma cadeia não-ergótica

A partir de uma matriz de transição, também pode-se determinar se uma cadeia é ergótica. Para tanto, basta verificar se há probabilidades nulas e se estas tornam algum estado inalcançável. Se não houver nenhuma probabilidade nula, certamente trata-se de uma cadeia ergótica.

Por exemplo, a seguinte matriz de transição é ergótica, pois, apesar de haver duas probabilidades nulas, é possível chegar a estado em uma ou mais transições. Isto pode ser verificado na Figura 7.

$$M = \begin{bmatrix} 0{,}1 & 0{,}5 & 0{,}4 \\ 0{,}2 & 0{,}8 & 0 \\ 0 & 0{,}1 & 0{,}9 \end{bmatrix}$$

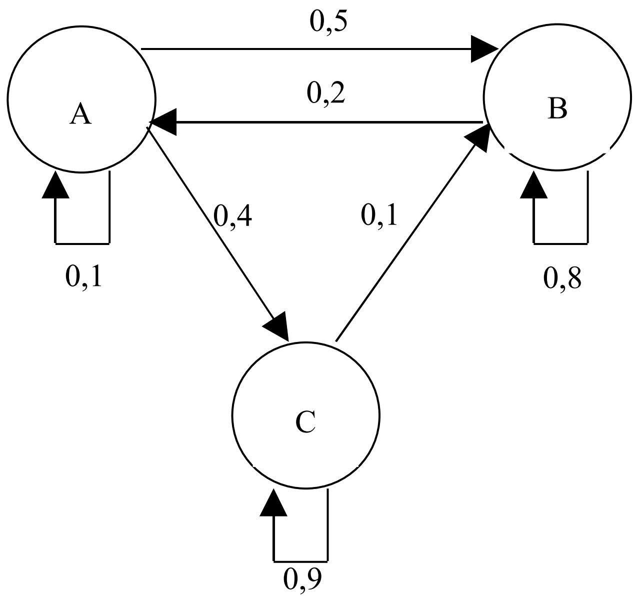


Figura 7: Diagrama de Transição de cadeia ergótica

Já a próxima matriz apresenta uma cadeia não-ergótica. Conforme pode ser verificado na Figura 8 que, uma vez no estado B, é impossível chegar a outro estado.

$$M = \begin{bmatrix} 0,1 & 0,5 & 0,4 \\ 0 & 1 & 0 \\ 0 & 0,1 & 0,9 \end{bmatrix}$$

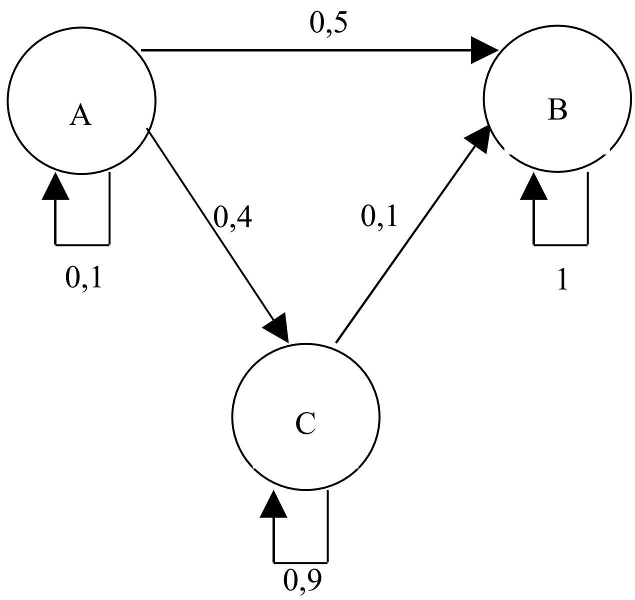


Figura 8: Diagrama de Transição de cadeia ergótica

7 Cadeias Regulares

Em fenômenos cuja matriz de transição, em alguma potência, não tenha elementos nulos, trata-se de uma cadeia regular. Toda cadeia regular é também ergótica, conforme poderá ser verificado adiante.

A matriz a seguir possui um elemento nulo na primeira linha. Porém, a potência M^2 não possui nenhum elemento nulo. Logo, trata-se de uma cadeia regular.

$$M = \begin{bmatrix} 0 & 0{,}5 & 0{,}5 \\ 0{,}2 & 0{,}3 & 0{,}5 \\ 0{,}3 & 0{,}1 & 0{,}6 \end{bmatrix}$$

$$M^2 = \begin{bmatrix} 0{,}25 & 0{,}2 & 0{,}55 \\ 0{,}21 & 0{,}24 & 0{,}55 \\ 0{,}2 & 0{,}24 & 0{,}56 \end{bmatrix}$$

Conforme foi verificado na seção “Determinação de probabilidades futuras”, uma potência da matriz de transição indica todas as probabilidades de transição em um momento futuro (por exemplo: M^3 pode indicar uma matriz de transição no terceiro ano do fenômeno estudado). Logo, se em uma potência (ou seja, em um momento futuro) a matriz de transição não possui elementos nulos, isto indica que, em algum momento, todas as transições serão possíveis, logo a cadeia é ergótica.

Já as potências das próximas matrizes formam padrões em que as probabilidades nulas não desaparecem. Logo, estas matrizes representam cadeias não regulares.

$$M = \begin{bmatrix} 0,5 & 0,5 & 0 \\ 0,1 & 0 & 0,9 \\ 0 & 0 & 1 \end{bmatrix}$$

$$M^2 = \begin{bmatrix} 0,3 & 0,25 & 0,45 \\ 0,05 & 0,05 & 0,9 \\ 0 & 0 & 1 \end{bmatrix}$$

$$M^3 = \begin{bmatrix} 0,175 & 0,15 & 0,675 \\ 0,03 & 0,025 & 0,945 \\ 0 & 0 & 1 \end{bmatrix}$$

$$M^4 = \begin{bmatrix} 0,102 & 0,0875 & 0,81 \\ 0,0175 & 0,015 & 0,9675 \\ 0 & 0 & 1 \end{bmatrix}$$

$$M = \begin{bmatrix} 0 & 1 & 0 \\ 0.1 & 0 & 0,9 \\ 0 & 1 & 0 \end{bmatrix}$$

$$M^2 = \begin{bmatrix} 0,1 & 0 & 0,9 \\ 0 & 1 & 0 \\ 0,1 & 0 & 0,9 \end{bmatrix}$$

$$M^3 = \begin{bmatrix} 0 & 1 & 0 \\ 0,1 & 0 & 0,9 \\ 0 & 1 & 0 \end{bmatrix}$$

$$M^4 = \begin{bmatrix} 0,1 & 0 & 0,9 \\ 0 & 1 & 0 \\ 0,1 & 0 & 0,9 \end{bmatrix}$$

Como estes padrões se repetem indefinidamente, tais matrizes de transição nunca deixarão de ter elementos nulos, porém, isso não indica que se trata obrigatoriamente de uma cadeia não-ergótica, uma vez que há cadeias ergóticas não-regulares.

8 Regime estacionário

Para toda cadeia ergótica regular existe um regime estacionário, onde as probabilidades de transição se tornam constantes ao longo do tempo. Em uma cadeia em regime estacionário, a probabilidade de transição é baseada apenas no próximo estado, e não mais no par estado corrente/próximo estado.

Seja π o vetor de probabilidades de uma cadeia em regime estacionário, têm-se a equação:

$$\pi = [\pi_1 \ \ \pi_2 \ \ \pi_3]$$

Onde $\sum_{i=1}^{I} \pi_i = 1$, sendo *I* o número de estados. O valor de π_1, por exemplo, é a probabilidade, em regime estacionário, de haver uma transição para o estado E_1. Este vetor pode ser obtido através da resolução do seguinte sistema.

$$\pi = \pi * M$$
$$\sum_{i=1}^{I} \pi_i = 1$$

Considerando a representação genérica de *M* e a representação do vetor π, pode ser reescrito da seguinte forma:

$$\pi_1 * p_{11} + \pi_2 * p_{21} + \pi_3 * p_{31} = \pi_1$$
$$\pi_1 * p_{12} + \pi_2 * p_{22} + \pi_3 * p_{23} = \pi_2$$
$$\pi_1 * p_{31} + \pi_2 * p_{32} + \pi_3 * p_{33} = \pi_3$$
$$\pi_1 + \pi_2 + \pi_3 = 1$$

A partir da resolução deste sistema, através de métodos lineares (Boldrini, 1980), é possível obter as probabilidades em regime estacionário. Voltando ao exemplo das classes sociais, pode-se encontrar o vetor de transição do regime estacionário a partir do sistema a seguir:

$$
\begin{aligned}
0{,}3\,\pi_1 + 0{,}2\,\pi_2 + 0\,\pi_3 &= 0 \\
0{,}5\,\pi_1 + 0{,}3\,\pi_2 + 0{,}1\,\pi_3 &= 0 \\
0{,}2\,\pi_1 + 0{,}5\,\pi_2 + 0{,}9\,\pi_3 &= 0 \\
\pi_1 + \pi_2 + \pi_3 &= 1
\end{aligned}
$$

Resolvendo este sistema, têm-se como resultado o vetor de probabilidade:

$$\pi = [0{,}04 \quad 0{,}146 \quad 0{,}812]$$

A partir destes resultados, pode-se fazer as afirmações:

1. A longo prazo, a probabilidade de um determinado indivíduo estar na classe A é de 0,04. A probabilidade de estar na classe B é 0,146 e 0,812 de estar na classe C;
2. A longo prazo, 4% da população estará na classe A, 14,6% na classe B e 81,2% na classe C.

Estas afirmações, referentes ao regime estacionário, só podem ser feitas se o fenômeno estudado se tratar de uma cadeia ergótica regular. Caso se trate de uma cadeia ergótica não-regular, a primeira afirmação não pode ser feita com segurança, uma vez que haverá elementos nulos cíclicos na matriz de transição (Shamblin, 1979).

9 Exercícios

1. Em um censo populacional de uma cidade de médio porte, foi constatado que a cada ano 7% da população rural migra para a zona urbana e que 2% da população urbana migra para a zona rural. Supondo que este fenômeno social seja estável, não havendo mudanças nestas taxas, temos os seguintes itens:

   a) Represente o diagrama de transição.

   b) Monte a matriz de transição.

   c) Em 5 anos, qual a probabilidade de um indivíduo, atualmente na zona urbana, ter migrado para a zona rural?

   d) Em 10 anos, qual a probabilidade de um indivíduo, atualmente na zona rural, ter migrado para a zona urbana?

   e) A longo prazo, qual a probabilidade de um indivíduo migre para a zona urbana?

   f) A longo prazo, qual será a taxa de migração da população para a zona urbana e para a zona rural desta cidade (desconsiderando o crescimento populacional da cidade)?

2. Uma pesquisa de mercado de um produto comercializado em três diferentes marcas constatou as seguintes probabilidades:

- Um consumidor da marca W deste produto, a cada compra, tem probabilidade 0,8 de manter-se fiel à marca, probabilidade 0,05 de escolher a marca G e probabilidade 0,15 de escolher a marca R;
- Um consumidor da marca G deste produto, a cada compra, tem probabilidade 0,9 de manter-se fiel à marca, probabilidade 0,01 de escolher a marca W e probabilidade 0,09 de escolher a marca R;
- Um consumidor da marca R deste produto, a cada compra, tem probabilidade 0,5 de manter-se fiel à marca, probabilidade 0,3 de escolher a marca G e probabilidade 0,2 de escolher a marca W;

Com base nestas informações, responda aos itens:

a) Represente o diagrama de transição.

b) Monte a matriz de transição.

c) Em 6 compras, qual a probabilidade de um consumidor da marca G optar pela marca W?

d) Em 8 compras, qual a probabilidade de um consumidor da marca R optar pela marca G?

e) A longo prazo, qual a probabilidade de um indivíduo optar pela marca G?

f) A longo prazo, qual taxa de indivíduos que terá optado pela marca R?

3. Um determinado fruto tem sua safra classificada como superior, média e pobre. Estudos revelam que, após uma safra pobre, há probabilidades 0,6 e 0,3 de a safra no ano posterior ser classificada como média ou superior, respectivamente. Após uma safra média, há probabilidades 0,4 e 0,1 de a próxima safra ser classificada como superior ou pobre, respectivamente. E após uma safra superior, há probabilidades 0,5 e 0,1 de a próxima safra ser classificada como média ou pobre, respectivamente.

   Com base nestas informações, responda aos itens:

   a) Represente o diagrama de transição.

   b) Monte a matriz de transição.

   c) Em 4 anos, qual a probabilidade de uma safra vir a ser classificada como superior, dado que a safra atual é pobre?

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

   e) A longo prazo, qual a probabilidade de se ter uma safra superior?

4. Uma máquina tem pode estar em três estados: operando, estragada e em manutenção corretiva. Em levantamentos estatísticos anteriormente feitos verificou-se que, mensalmente, uma máquina, quando colocada em funcionamento, tem probabilidade 0,9 de continuar funcionando e 0,1 de vir a apresentar algum defeito. Uma máquina em manutenção, tem probabilidade 0,4 de voltar a operar em um mês e 0,6 de continuar em manutenção. Já uma máquina estragada, tem probabilidade 0,8 de entrar em manutenção e 0,2 de continuar estragada aguardando manutenção.

Com base nestas informações, responda aos itens:

a) Represente o diagrama de transição.

b) Monte a matriz de transição.

c) Em 3 meses, qual a probabilidade de uma máquina continuar funcionando sem problemas?

d) Qual a probabilidade de uma máquina permanecer 2 meses em manutenção?

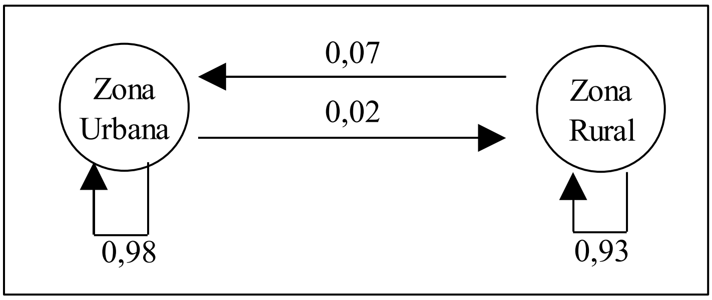
e) A longo prazo, qual a probabilidade de uma máquina apresentar defeito?

f) Para garantir uma probabilidade de pelo menos 0,73 de uma máquina continuar em operação, de quanto em quanto tempo deve-se realizar uma manutenção preditiva?

Apêndice A – Resposta dos exercícios

1.

a)



b) $M = \begin{bmatrix} 0,98 & 0,02 \\ 0,07 & 0,93 \end{bmatrix}$

c) Aproximadamente 0,083.

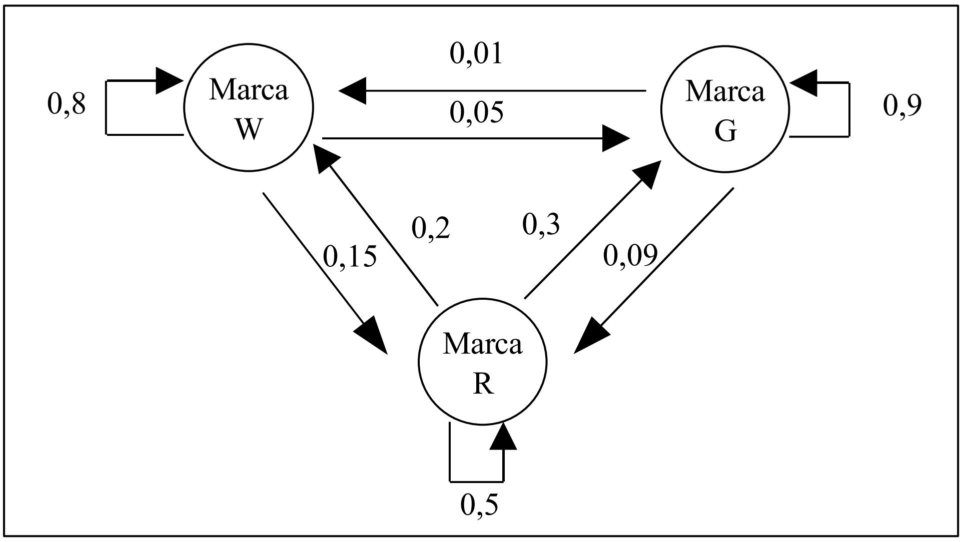
d) Aproximadamente 0,475.

e) Aproximadamente 0,77.

f) A taxa de migração da população para a zona urbana será aproximadamente 77,78% e para a zona rural será 22,22%.

2.

a)



b) $M = \begin{bmatrix} 0{,}8 & 0{,}05 & 0{,}15 \\ 0{,}01 & 0{,}9 & 0{,}09 \\ 0{,}2 & 0{,}3 & 0{,}5 \end{bmatrix}$

c) Aproximadamente 0,125.

d) Aproximadamente 0,58.

e) Aproximadamente 0,62.

f) Aproximadamente 17,33%.

Apêndice B – Modelagem dos exercícios no Octave[3]

1.

```
%criação da matriz de transição do exercicio 1
M1=[0.98 0.02; 0.07 0.93]

%vetor probabilidade da zona urbana após 5 anos
V_zona_urbana=[1 0]*M1^5

%vetor probabilidade da zona rural após 5 anos
V_zona_rural=[0 1]*M1^10

%probabilidades em regime estacionário
M1^10000
```

2.

```
%criação da matriz de transição do exercício 2
M2=[0.8 0.05 0.15; 0.01 0.9 0.09;  0.2  0.3 0.5]

%vetor probabilidade da marca G pós 6 anos
V_marca_w=[0 1 0]*M2^6

%vetor probabilidade da marca R pós 8 anos
V_marca_R=[0 0 1]*M2^8

%probabilidades em regime estacionário
M2^10000
```

[3]GNU Octave é uma linguagem de alto nível inicialmente desenvolvida para algoritmos numéricos. É compatível com o Matlab e pode ser obtida gratuitamente no site: www.octave.org.

Apêndice C – Exemplo de aplicação de Cadeias de Markov

Cadeias de Markov aplicadas à Genética

A genética permite a previsão probabilística de características em uma população. Devido a estas características poderem ser representadas em estados discretos e a probabilidade de transição destas características, em períodos também discretos, dependam do estado inicial da população (não abordaremos aqui fenômenos como migração, mutação e seleção natural), pode-se utilizar Cadeias de Markov para uma modelagem probabilística.

Tomaremos como base alguns conceitos da genética clássica, criada por Gregor Mendel (1822 – 1884). De acordo com a genética clássica, as características de plantas e animais são determinadas por um par de genes, que podem ser dominantes (A) ou recessivos (a). Há três possíveis combinações destes genes: AA, Aa (que equivale a aA) e aa. Estas combinações podem gerar, respectivamente, genótipos dominantes (D), heterozigotos (H) ou recessivos (R). Em alguns casos, estes três genótipos resultam em três características diferentes, e em outros AA e Aa resultam na mesma característica.

Nas próximas seções serão tratados alguns casos de aplicação de cadeias de Markov, e suas variações, à genética.

Previsão de genótipos com cruzamentos controlados

A previsão de genótipos em uma população pode ser bastante útil quando se faz necessária a monitoração e possível controle (por seleção artificial) de algumas características, sejam elas desejáveis ou não (como doenças genéticas, por exemplo).

As cadeias de Markov podem ser úteis neste tipo de previsão. O primeiro passo, para o exemplo a ser desenvolvido nesta seção, será definir os estados (que no caso serão os genótipos D, H e R) e obter a probabilidade de cada genótipo para cruzamentos entre indivíduos DxD, DxH, DxR DxR, HxH, HxR, e RxR. As probabilidades destes cruzamentos são exibidos na Tabela 1.

Tabela 1: Cruzamento entre genótipos

Genótipo/cruzamento	**D(AA)**	**H(Aa)**	**R(aa)**
DxD	1	0	0
DxH	0,5	0,5	0
DxR	0	1	0
HxH	0,25	0,5	0,25
HxR	0	0,5	0,5
RxR	0	0	1

Com base nas probabilidades, pode-se construir uma cadeia de Markov para, por exemplo, fazer uma previsão de genótipos para três períodos, ou seja, após três gerações. Neste exemplo, os indivíduos da população farão cruzamento com indivíduos de genótipo H[4]. A matriz de transição a seguir é formada pelos vetores de probabilidade dos cruzamentos DxH, HxH e RxH, respectivamente.

[4] Em processos de seleção artificial, pode-se, em uma população de variados genótipos de mesmo sexo, introduzir um determinado genótipo do sexo oposto, de forma que as combinações gênicas sejam controladas e, conseqüentemente, seus resultados previsíveis. Esta prática é bastante comum na agropecuária (Campos, 2002).

$$M = \begin{bmatrix} 0,5 & 0,5 & 0,0 \\ 0,25 & 0,5 & 0,25 \\ 0 & 0,5 & 0,5 \end{bmatrix}$$

Considerando que, no processo de seleção artificial, a primeira geração a cruzar com o genótipo H seja formada exclusivamente por indivíduos homozigotos dominantes (D), a previsão para o período de três gerações é dada pela equação:

$$V_D^3 = V_D^0 M^3 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 0,5 & 0,5 & 0,0 \\ 0,25 & 0,5 & 0,25 \\ 0 & 0,5 & 0,5 \end{bmatrix}^3 \cong \begin{bmatrix} 0,31 & 0,5 & 0,19 \end{bmatrix}$$

O vetor de probabilidade resultante indica que, após três gerações (sendo a geração inicial de genótipos D, e os cruzamentos sempre envolvendo um genótipo H) traz as seguintes informações:

- Há probabilidade de, aproximadamente, 0,31 de um indivíduo da terceira geração ter genótipo D, 0,5 de ter genótipo H, e 0,19 de ter genótipo R;
- Pode-se estimar que: 31% da terceira geração terá genótipo D, 50% terá genótipo H, e 19% genótipo R.

A este método de previsão utilizado, deve-se fazer algumas observações:

- Após a primeira geração, é bastante provável que surjam genótipos diferentes dos inicialmente adotados (neste exemplo, D e H);

- Embora a matriz de transição permita que todos os genótipos sejam representados, ela traz apenas as probabilidades de cruzamento com o genótipo pré-determinado;
- Para que o resultado encontrado seja válido, é preciso que haja controle dos cruzamentos, de forma que um dos genótipos, em cada cruzamento, seja do tipo pré-determinado, ou seja, genótipo H. Porém, nem sempre o controle absoluto dos cruzamentos é possível. Na próxima seção será exibido um exemplo sem esta limitação.

O cálculo de probabilidade envolvendo cruzamentos exclusivamente com genótipos D, considerando a mesma geração inicial, segue o mesmo raciocínio, porém a matriz de probabilidade irá representar os cruzamentos DxD, DxH e DxR:

$$M = \begin{bmatrix} 1 & 0 & 0 \\ 0,5 & 0,5 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

A equação a seguir exibe o cálculo do vetor de probabilidade após três gerações.

$$V_D^3 = V_D^0 M^3 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 \\ 0,5 & 0,5 & 0 \\ 0 & 1 & 0 \end{bmatrix}^3 \cong \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$$

Este resultado pode parecer bastante estranho à primeira vista, porém, deve ser levado em conta o fato de o cruzamento de dois genótipos D só poder originar genótipos D, uma vez que os genes envolvidos são DD. E como os cruzamentos serão controlados, estes genes estarão integralmente presentes, geração após geração.

Portanto, é desnecessário qualquer método de previsão quando se realiza cruzamento exclusivamente entre genótipos homogêneos. Caso a geração inicial consistisse nos genótipos H ou R, a previsão poderia trazer resultados mais relevantes.

Supondo, por exemplo, a geração inicial de genótipos H, a equação seguinte exibe a previsão após três gerações.

$$V_H^3 = V_H^0 M^3 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 \\ 0,5 & 0,5 & 0 \\ 0 & 1 & 0 \end{bmatrix}^3 \cong \begin{bmatrix} 0,87 & 0,13 & 0 \end{bmatrix}$$

Interpretando o resultado, pode-se afirmar:

- Há probabilidade de, aproximadamente, 0,87 de que um indivíduo da terceira geração seja do genótipo D, 0,13 que seja do genótipo H, e 0 que seja do genótipo R;
- Pode-se estimar que: 87% da terceira geração terá genótipo D, 13% terá genótipo H, e 0% genótipo R.

Por fim, o cálculo de previsão para três gerações, considerando que os cruzamentos envolvam o genótipo R e a geração inicial seja de genótipos H (lembrando que uma população de genótipos R dispensaria qualquer previsão), é dado pela equação.

$$V_R^3 = V_R^0 M^3 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} * \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0,5 & 0,5 \\ 0 & 0 & 1 \end{bmatrix}^3 \cong \begin{bmatrix} 0 & 0,125 & 0,875 \end{bmatrix}$$

Interpretando-se o resultado, pode-se afirmar:

- Não há probabilidade de que um indivíduo da terceira geração seja do genótipo D, há probabilidade aproximada de 0,125 de que seja do genótipo H, e 0,875 que seja do genótipo R;
- Pode-se estimar que: 0% da terceira geração terá genótipo D, 12,5% terá genótipo H, e 87,5% genótipo R.

Previsão de tendência de genótipos ao longo do tempo

Os exemplos da seção anterior servem como base para previsão de genótipos em várias gerações, sendo predefinidos o genótipo inicial da população e o genótipo utilizado nos cruzamentos. Porém esta abordagem traz algumas limitações: ela não permite a fácil identificação de uma tendência a longo prazo, e nem sempre é possível dispor de uma população inicial com genótipos idênticos. Portanto, faz-se necessário aplicar as propriedades do regime estacionário de cadeias de Markov.

A partir da matriz de transição para cruzamentos com o genótipo H pode-se obter o vetor de transição em regime estacionário:

$$\begin{aligned}
&0{,}5\pi_1 + 0{,}25\pi_2 + 0\pi_3 = \pi_1 \\
&0{,}5\pi_1 + 0{,}5\pi_2 + 0{,}5\pi_3 = \pi_2 \\
&0\pi_1 + 0{,}25\pi_2 + 0{,}5\pi_3 = \pi_3 \\
&\pi_1 + \pi_2 + \pi_3 = 1
\end{aligned}$$

$$\pi = \begin{bmatrix} 0{,}25 & 0{,}5 & 0{,}25 \end{bmatrix}$$

O resultado exibido na equação traz as seguintes informações:

- A longo prazo, há probabilidade aproximada de 0,25 de que um dado indivíduo seja do genótipo D, 0,5 do genótipo H e 0,25 que seja do genótipo R;
- A longo prazo, pode-se estimar que: 25% da população terá genótipo D, 50% terá genótipo H, e 0,25% genótipo R.

Estas informações são válidas uma vez que a cadeia expressa na matriz é ergótica regular. É interessante observar que, no regime estacionário, o estado corrente não influi no cálculo da probabilidade. Neste exemplo, isto equivale a dizer que não importa o genótipo da população inicial, basta determinar qual genótipo estará presente em todos os cruzamentos (neste caso, genótipo H) para se obter a matriz de transição, e, a partir dela, calcular o vetor de probabilidade em regime estacionário.

Para a previsão de tendências de genótipos para cruzamentos com genótipos D, deve-se obter o vetor de probabilidades em regime estacionário.

$$\begin{aligned}
&\pi_1 + 0{,}5\pi_2 + 0\pi_3 = \pi_1 \\
&0\pi_1 + 0{,}5\pi_2 + 1\pi_3 = \pi_2 \\
&0\pi_1 + 0\pi_2 + 0\pi_3 = \pi_3 \\
&\pi_1 + \pi_2 + \pi_3 = 1
\end{aligned}$$

$$\pi = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$$

Este vetor indica que, a longo prazo, a tendência da população, submetida a sucessivos cruzamentos com genótipos D, é ser composta apenas de genótipos D. Matematicamente este resultado se deve ao fato de a matriz de transição representar uma cadeia absorvente (Shamblin, 1979). Do ponto de vista da genética, pode-se comprovar o resultado obtido simulando os sucessivos cruzamentos em cada geração. Na décima geração, aproximadamente 0,1% dos indivíduos terá genótipo H (percentual que tende a se anular nas próximas gerações) e nenhum indivíduo terá genótipo R.

Previsão de tendência de genótipos utilizando matrizes aumentadas

As seções anteriores apresentaram modelagens com limitações, seja quanto à população inicial, seja quanto a cruzamentos controlados. Nesta seção será apresentada uma modelagem (baseada nos processos de Markov, mas com algumas modificações(Boldrini, 1980)) livre destas restrições (Haetinger e Dullius,2006).

Nesta modelagem, é necessário um estudo prévio da população, de forma que se possa estimar as porcentagens de cada genótipo na geração corrente. Os genótipos D, H e R, terão esta porcentagem dada por p_d^1, p_h^1 e p_r^1, respectivamente.

Logo, a probabilidade de cruzamento entre indivíduos dominantes é dada por $p_d^1 * p_d^1$. A probabilidade de cruzamento entre indivíduos recessivos é dada por $p_r^1 * p_r^1$. O mesmo raciocínio se aplica para o cruzamento entre indivíduos híbridos: $p_h^1 * p_h^1$

Já a probabilidade de cruzamento entre indivíduos dominante e um recessivo deve levar em conta as duas probabilidades: $p_d^1 * p_r^1$ e $p_r^1 * p_d^1$ (resultando na probabilidade $2(p_d^1 * p_r^1)$). Da mesma forma, o cruzamento entre híbridos e dominantes, e entre recessivos e híbridos é dada, respectivamente, por: $2(p_d^1 * p_h^1)$ e $2(p_r^1 * p_h^1)$.

Estas probabilidades são apresentadas na Tabela 2:

Tabela 2: Probabilidades de cruzamento entre genótipos

Cruzamento	Probabilidade
DxD	$p_d^1 * p_d^1$
DxH	$2(p_d^1 * p_h^1)$
DxR	$2(p_d^1 * p_r^1)$
HxH	$p_h^1 * p_h^1$
HxR	$2(p_r^1 * p_h^1)$
RxR	$p_r^1 * p_r^1$

A partir das probabilidades exibidas na Tabela 1 e na Tabela 2, pode-se calcular a porcentagem de cada um dos genótipos na segunda geração.

$$\begin{bmatrix} p_d^2 \\ p_h^2 \\ p_r^2 \end{bmatrix} = \begin{bmatrix} 1 & 0,5 & 0 & 0,25 & 0 & 0 \\ 0 & 0,5 & 1 & 0,5 & 0,5 & 0 \\ 0 & 0 & 0 & 0,25 & 0,5 & 1 \end{bmatrix} * \begin{bmatrix} p_d^1 * p_d^1 \\ 2(p_d^1 * p_h^1) \\ 2(p_d^1 * p_r^1) \\ p_h^1 * p_h^1 \\ 2(p_r^1 * p_h^1) \\ p_r^1 * p_r^1 \end{bmatrix}$$

Uma vez obtidas as porcentagens da segunda geração, pode-se obter as porcentagens da terceira geração, e, sucessivamente, das próximas gerações.

Esta modelagem pode ser bastante útil para Agricultura na previsão de resistências genéticas a certos tipos de doenças e e pesticidas. Porém, ela também apresenta algumas limitações:

- Deve-se levar em consideração que indivíduos de uma geração anterior não mais realizem cruzamentos (o que é bastante comum em populações de insetos) (Haetinger e Dullius,2006);
- Deve-se desconsiderar migração, mutações e seleção natural, e considerar que os dois sexos aparecem em quantidades iguais(Boldrini, 1980);
- Estas matrizes não representam cadeias de Markov, logo, o regime estacionário não pode ser obtido. Portanto, os cálculos de gerações futuras tendem a ser cada vez mais demorados, uma vez que não podem ser obtidos diretamente (Haetinger e Dullius,2006);

Bibliografia

BOLDRINI, José Luiz. **Álgebra Linear**. São Paulo: Harbra,1980
SHAMBLIN, James E.. **Pesquisa Operacional: uma abordagem básica**. São Paulo: Atlas,1979
CAMPOS, M. A.. **Probabilidade Intervalar e Cadeias de Markov Intervalares no MAPLE.**. SBMAC: Tendências em Matemática Aplicada, Nº 2 , 2002
HAETINGER, Claus. DULLIUS, Madalena. **Álgebra Linear e Geometria Analítica**, 2006, http://ensino.univates.br/~chaet(Acessado em 10/09/2007)