Cadeias de Markov

Francisco A. Rodrigues

Instituto de Ciências Matemáticas e de Computação

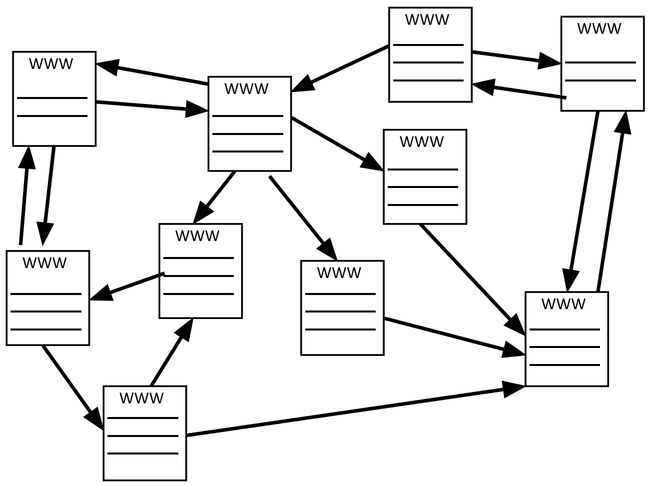
Universidade de São Paulo

1. Introdução

Se visitarmos o sítio na Wikipedia sobre cadeias de Markov:

https://pt.wikipedia.org/wiki/Cadeias_de_Markov

vamos encontrar diversos links para outras páginas, com assuntos como "matemática", "processo estocásticos" e sobre o matemático Andrei Andreyevich Markov. Se clicarmos no link "matemática", encontraremos o termo "raciocínio lógico", que quando clicado, abrirá uma outra página com o termo "lógica". Essa navegação em que clicamos em links na Wikipedia pode ser representada pelo grafo abaixo, sendo cada página um assunto e as conexões definem a relação entre as páginas.



Essas páginas definem um grafo e quando clicamos em páginas em sequência, ou seja, em uma página a cada passo de tempo, estamos realizando uma caminhada aleatória nesse grafo. Essa caminhada aleatória define uma cadeia de Markov, que é o assunto dessa aula.

Além da Wikipedia, esse conceito de caminha aleatória é usado pelo Google Pagerank, que é o algoritmo usado pela empresa Google, para ordenar as páginas da web por importância. Ou seja, quando realizamos uma busca no Google, a ordem em que as páginas aparecem é definida pelo algoritmo Pagerank. Para entender o algoritmo Pagerank precisamos definir o que é uma cadeia de Markov.

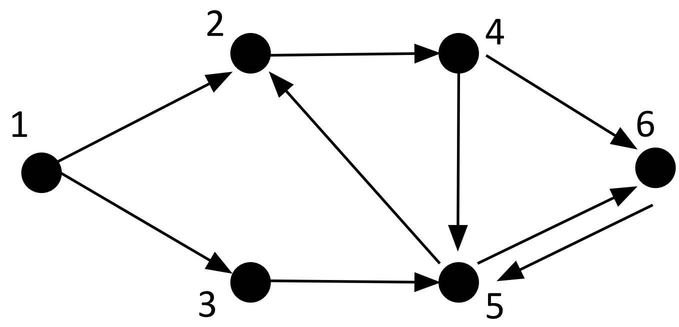
Definição 1: Considere o processo estocástico $\{X_n, n \in \mathbb{N}\}$. Nós dizemos que este processo é uma cadeia de Markov se:

$$P(X_{n+1} = j \,\big|\, X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = P(X_{n+1} = j \,\big|\, X_n = i)$$

Para todos os estados $i, j, i_0, i_1, \ldots, i_n$.

Ou seja, a probabilidade de visitar o estado j no passo n+1 depende apenas do estado em que a cadeia estava no passo n. Portanto, a transição para o próximo estado depende apenas do passo atual. Notem que o tempo é discreto, ou seja, dividido em intervalos igualmente espaçados. Podemos imaginar que a cada segundo ocorre uma transição na cadeia e entre as transições nada ocorre.

Na figura a seguir, temos um grafo, onde cada círculo, chamado vértice no grafo, são conectados por arestas. Esse grafo pode representar uma cadeia de Markov, onde os vértices são estados da cadeia e as arestas indicam as transições entre os estados. Imagine que um agente está posicionado no vértice 1. Esse agente pretende se movimentar pelo grafo e escolhe um vizinho de forma aleatória. Ou seja, o agente pode se movimentar para os vértices 2 ou 3. Vamos supor que esse agente se moveu para o vértice 3. No próximo passo de tempo, ele só pode se movimentar para o vértice 5. No vértice 5, ele pode visitar os vértices 2 ou 6. Vamos supor que o agente se move para o vértice 2. Notem que essa movimentação do vértice 5 para o 2 independe dos estados visitados anteriormente. Portanto, para chegar no vértice 5, o agente pode ter vindo dos vértices 3, 4 ou 6. Essa informação não é levada em conta quando o agente se movimenta para o estado 2 a partir do 5. Logo, a probabilidade se mover para o próximo vértice (estado 2) depende apenas do estado atual (estado 5). Essa independência com relação aos tempos anteriores define a propriedade de “falta de memória” da cadeia de Markov e ilustra conceitualmente o significado do teorema anterior.



Definição 2: Seja P uma matriz quadrada cujas entradas P_{ij} são definidas para todos os estados i e j. Então, P é chamada matriz Markoviana (ou matriz de probabilidade de transição) se:

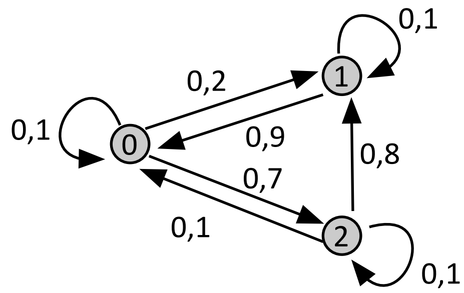
- Para todo $i \in S, P_{ij} \geq 0$, onde S é o conjunto de estados.
- $\sum_{j \in S} P_{ij} = 1, \forall i \in S$.

Um processo Markoviano é totalmente definido pela sua matriz de probabilidades de transição e a distribuição de probabilidade de X_0.

Exemplo: Uma cadeia de Markov $\{X_n, n \geq 0\}$, nos estados 0, 1, e 2, tem matriz de probabilidade de transição

$$P = \begin{array}{c} 0 \\ 1 \\ 2 \end{array} \begin{pmatrix} 0{,}1 & 0{,}2 & 0{,}7 \\ 0{,}9 & 0{,}1 & 0 \\ 0{,}1 & 0{,}8 & 0{,}1 \end{pmatrix}$$

Notem que a soma dos elementos das linhas da matriz é igual a 1. Esta matriz pode ser representada pelo grafo abaixo, sendo que cada elemento P_{ij} da matriz é igual à probabilidade de ir do estado i para o estado j.



Por exemplo, temos que

$$P(X_{n+1} = 1 \,|\, X_n = 0) = P_{10} = \ 0{,}2;$$

$$P(X_{n+1} = 2 \,|\, X_n = 0) = P_{02} = \ 0{,}7;$$

$$P(X_{n+1} = 0 \,|\, X_n = 2) = P_{20} = \ 0{,}7;$$

$$P(X_{n+1} = 1 \,|\, X_n = 2) = P_{12} = \ 0;$$

e assim sucessivamente. Notem que essas probabilidade são válidas para qualquer $n \geq 0$.

A probabilidade de visitar uma sequência de estados pode ser escrita da seguinte forma:

$$P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n\ ) =$$
$$= P(X_n = i_n\ |X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1})P(X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}),$$

Onde usamos a probabilidade condicional:

$$P(X|Y) = \frac{P(X,Y)}{P(Y)} \Longrightarrow P(X,Y) = P(X|Y)P(Y).$$

Usando a definição 1, que afirma que a probabilidade de visitar próximo estado depende apenas do estado atual, temos:

$$P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n\ ) =$$
$$= P(X_n = i_n\ |X_{n-1} = i_{n-1})P(X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}),$$

Repetindo a mesma ideia para o segundo termo do lado direito:

$$P(X_0 = i_0, X_1 = i_1, \dots, X_n\ = i_n\ ) =$$
$$= P(X_n = i_n\ |X_{n-1} = i_{n-1})P(X_{n-1} = i_{n-1}\ |X_{n-2} = i_{n-2})P(X_0 = i_0, \dots, X_{n-2} = i_{n-2}),$$

Ou seja, temos:

$$P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n\ ) =$$

$$= P(X_n = i_n \,|X_{n-1} = i_{n-1}) \times P(X_{n-1} = i_{n-1} \,|X_{n-2} = i_{n-2}\,) \ldots \times$$

$$\ldots \times P(X_1 = i_1 \,|X_0 = i_0)P(\,X_0 = i_0)$$

Portanto,

$$P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n\,) = P_{i_0 i_1}\, P_{i_1 i_2} \ldots P_{i_{n-1} i_n}$$

Esse resultado permite calcular a probabilidade de ir de um estado a outro na cadeia. Essa probabilidade na verdade é calculada considerando-se uma caminhada no grafo que começa em $X_0 = \; i$ e termina em $X_n = j$.

Exemplo: Uma cadeia de Markov $\{X_n, n \geq 0\}$, nos estados 0, 1, e 2, tem matriz de probabilidade de transição

$$P = \begin{array}{c} 0 \\ 1 \\ 2 \end{array} \begin{pmatrix} 0{,}1 & 0{,}2 & 0{,}7 \\ 0{,}9 & 0{,}1 & 0 \\ 0{,}1 & 0{,}8 & 0{,}1 \end{pmatrix}.$$

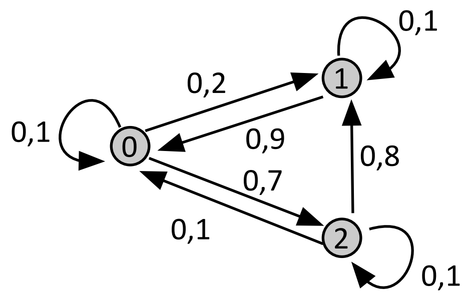
E distribuição de probabilidades inicial:

$$P(X_0 = 0) = 0{,}3, P(X_0 = 1) = 0{,}4 \text{ e } P(X_0 = 2) = 0{,}3.$$

Determine:

$a)\; P(X_0 = 0, X_1 = 1, X_2 = 0, X_3 = 2, X_4 = 0)$

Conforme vimos anteriormente, essa matriz pode ser representada pelo grafo abaixo.



Temos que $P(X_0 = 0, X_1 = 1, X_2 = 0, X_3 = 2, X_4 = 0)$ é a probabilidade de iniciar no estado 0, depois se mover para o estado 1, depois para o estado 0, para o estado 2 e finalmente, para o estado 0. Portanto, usando a notação matricial dada pela definição anterior, podemos escrever:

$$P(X_0 = 0, X_1 = 1, X_2 = 0, X_3 = 2, X_4 = 0) = P(X_0 = 0)P_{01}\, P_{10}\, P_{02}\, P_{20} =$$
$$= 0{,}3 \times 0{,}2 \times 0{,}9 \times 0{,}7 \times 0{,}1 = 0{,}003.$$

Notem que consideramos a probabilidade de iniciar no estado 0, isto é, $P(X_0 = 0) = 0{,}3$.

b) Calcule $P(X_1 = 1, X_2 = 0 | X_0 = 0)$:

Nesse caso, não precisamos levar em conta a probabilidade inicial, pois já foi informado que o processo iniciou no estado zero (notem a probabilidade condicional, informando que $X_0 = 0$). Assim, temos que o processo inicia no estado 0, move para o estado 1 e depois move-se novamente para o estado 0. Ou seja,

$$P(X_1 = 1, X_2 = 0 \,\big|\, X_0 = 0) = P_{01}\, P_{10} = 0{,}2 \times 0{,}9 = 0{,}18.$$

c) Calcule $P(X_2 = 2, X_3 = 0 \,\big|\, X_1 = 0)$:

Nesse problema, notem que o tempo inicial é igual n=1 ($X_1 = 0$), ao invés de ser zero, como vimos nos exemplos anteriores. Embora o tempo inicial não seja igual a zero, a análise é análoga à que realizamos anteriormente. Na verdade, o que importa é a diferença entre os tempos e não quando começamos a contar. Isto é, as probabilidades abaixo são equivalentes:

$$P(X_2 = 2, X_3 = 0 \,\big|\, X_1 = 0) = P(X_1 = 2, X_2 = 0 \,\big|\, X_0 = 0) = P(X_{n+1} = 2, X_{n+2} = 0 \,\big|\, X_n = 0)$$

Para qualquer $n \geq 0$.

Assim, temos:

$$P(X_2 = 2, X_3 = 0 \,\big|\, X_1 = 0) = P_{02}P_{20} = 0{,}7 \times 0{,}1 = 0{,}07.$$

Exemplo: Uma cadeia de Markov $\mathrm{X}_0, \mathrm{X}_1, \ldots$, tem uma matriz de probabilidade de transição:

$$P = \begin{pmatrix} 0{,}5 & 0{,}3 & 0{,}2 \\ 0{,}1 & 0{,}8 & 0{,}1 \\ 0{,}6 & 0{,}1 & 0{,}3 \end{pmatrix}$$

Determine $P(X_{11} = 1, X_{12} = 2, X_{13} = 2 | X_9 = 0, X_{10} = 1)$

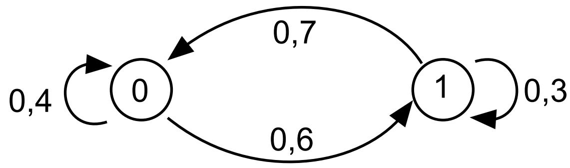
Para calcularmos a probabilidade pedida, usamos a propriedades de falta de memória da cadeia e a relação entre os índices:

$$P(X_{11} = 1, X_{12} = 2, X_{13} = 2 | X_9 = 0, X_{10} = 1) =$$
$$P(X_{11} = 1, X_{12} = 2, X_{13} = 2 | X_{10} = 1) = P(X_1 = 1, X_2 = 2, X_3 = 2 | X_0 = 1) =$$
$$= P_{11} P_{12} P_{22} = 0{,}8 \times 0{,}1 \times 0{,}3 = 0{,}024$$

Portanto

$$P(X_{11} = 1, X_{12} = 2, X_{13} = 2 | X_9 = 0, X_{10} = 1) = 0{,}024.$$

Exemplo: Considere a cadeia de Markov com dois estados:



Calcule $P(X_1 = 0)$. Assuma $P(X_0 = 0) = 0{,}75$ e $P(X_0 = 1) = 0{,}25$.

Para que a cadeia esteja no estado 0 no passo n=1, então no passo n=0 a cadeia estava no estado 0 ou no estado 1. Ou seja,

$$P(X_1 = 0) = P(X_1 = 0, X_0 = 0) + P(X_1 = 0, X_0 = 1)$$

Usando probabilidade condicional ($P(A, B) = P(A|B)P(B)$):

$$P(X_1 = 0) = P(X_1 = 0 | X_0 = 0) P(X_0 = 0) + P(X_1 = 0 | X_0 = 1) P(X_0 = 1)$$

Assim:

$$P(X_1 = 0) = P_{00} P(X_0 = 0) + P_{10} P(X_0 = 1) = 0{,}4 \times 0{,}75 + 0{,}7 \times 0{,}25 = 0{,}475.$$

Teorema: (Equações de Chapman-Kolmogorov) A probabilidade de transição no n-ésimo passo para uma cadeia de Markov satisfaz:

$$P_{ij}^{n} = P(X_{n+m} = j \,|\, X_m = i) = \sum\nolimits_{k=0}^{\infty} P_{ik}\, P_{kj}^{n-1}$$

Onde definimos:

$$P_{ij}^{0} = \begin{cases} 1 \; se \; i = j \\ 0 \; se \; i \neq j \end{cases}$$

Ou seja, esse teorema afirma que

$$P_{ij}^{n+m} = \sum_{k=0}^{\infty} P_{ik}^{n}\, P_{kj}^{m}, \ \ \forall n, m \geq 0$$

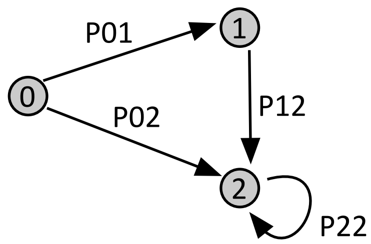
Assim, para calcular a probabilidade de sair de i e chegar em j seguindo 2, basta que multipliquemos a matriz P por ela mesma. Se desejarmos m passos, então multiplicamos a matriz por ela mesma m vezes.

Exemplo: Uma cadeia de Markov $\{X_n, n \geq 0\}$, nos estados 0, 1, e 2, tem matriz de probabilidade de transição

$$P = \begin{pmatrix} 0 & P_{01} & P_{02} \\ 0 & 0 & P_{12} \\ 0 & 0 & P_{22} \end{pmatrix}$$

Calcule $P(X_2 = 2|X_0 = 1)$.

Se considerarmos o grafo abaixo, temos que essa probabilidade é a chance de sair do vértice 1 e chegar no vértice 2 seguindo uma caminha aleatória de comprimento dois.



Se multiplicarmos a matriz por ela mesma, obtemos:

$$P^2 = \begin{pmatrix} 0 & 0 & P_{01}P_{12} + P_{02}P_{22} \\ 0 & 0 & P_{12}P_{22} \\ 0 & 0 & P_{22}^2 \end{pmatrix}$$

Podemos verificar esse resultado no grafo acima. Ou seja, para chegar no estado 2 partindo do estado 0 em dois passos, temos as possibilidades:

$a)\ X_0 = 0, X_1 = 1, X_2 = 2,$

$b) X_0 = 0, X_1 = 2, X_2 = 2,$

Considerando as probabilidades, obtemos

$$P(X_2 = 2 \,|\, X_0 = 0) = P_{01}\, P_{12} + P_{02}\, P_{22}$$

Assim, os elementos P_{ij}^n definem a probabilidade de sair do estado i e chegar no estado j seguindo n passos. Verifiquem esse resultado para os demais elementos da matriz acima.

Exemplo: Uma cadeia de Markov $\{X\_n, n \geq 0\}$, nos estados 0, 1, e 2, tem matriz de probabilidade de transição

$$P = \begin{array}{c} 0 \\ 1 \\ 2 \end{array} \begin{bmatrix} 0{,}1 & 0{,}2 & 0{,}7 \\ 0{,}2 & 0{,}2 & 0{,}6 \\ 0{,}6 & 0{,}1 & 0{,}3 \end{bmatrix}$$

Calcule:

a) $P(X_3 = 1 | X_1 = 0)$

Para calcularmos a probabilidade de ir do estado 0 ao estado 1 seguindo dois passos, multiplicamos a matriz P por ela mesma, obtendo:

$$P^2 = \begin{array}{c} 0 \\ 1 \\ 2 \end{array} \begin{bmatrix} 0{,}47 & 0{,}13 & 0{,}40 \\ 0{,}42 & 0{,}14 & 0{,}44 \\ 0{,}26 & 0{,}17 & 0{,}57 \end{bmatrix}$$

Os elementos dessa matriz definem as probabilidades:

$$P(X_{n+2} = j \,\big|\, X_n = 1) = P_{ij}^2$$

Assim, pela matriz, obtemos:

$$P(X_3 = 1 | X_1 = 0) = P_{01}^2 = 0{,}13.$$

b) $P(X_4 = 1 | X_2 = 1)$

De maneira similar ao caso anterior:

$$P(X_4 = 1 | X_2 = 1) = P_{11}^2 = 0{,}14.$$

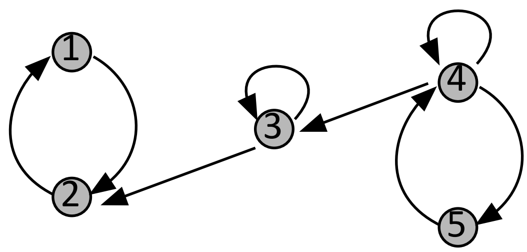
c) $P(X_2 = 0 | X_0 = 2)$

Usando matriz acima novamente:

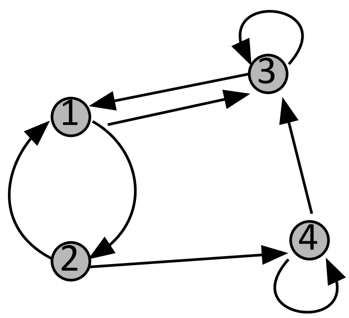
$$P(X_2 = 0 | X_0 = 2) = P_{20}^2 = 0{,}26.$$

2. Distribuição estacionária

Vamos considerar a cadeia de Markov a seguir. Se iniciarmos a caminhada aleatória no estado 1, podemos apenas nos mover para o estado 2 e voltar para o estado 1. Ou seja, a caminhada fica limitada a esses dois estados. O mesmo ocorre com os estados 4 e 5. Por outro lado, se iniciarmos no estado 3, após deixarmos esse estado, não o visitamos mais. Os estados que podemos visitar após um número infinito de passos é chamado recorrente, enquanto que estados que não serão visitados são chamados transientes. O estado 3, nesse exemplo, é transiente, enquanto os demais são recorrentes.



Por outro lado, na figura a seguir, temos uma cadeia em que todos os estados são recorrentes, ou seja, mesmo quando deixamos qualquer um dos estados, a caminhada volta a visitar esse estado.



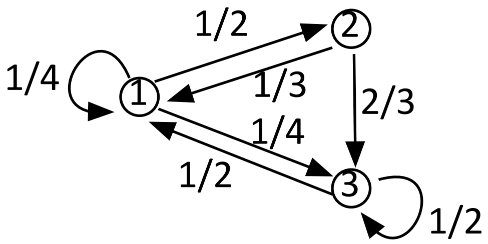
Cadeias de Markov em que todos os estados são recorrentes e para ao menos um dos estados i $P_{ii} > 0$, são chamadas ergódicas.
Para uma cadeia ergódica, podemos calcular a distribuição de probabilidade estacionária usando o seguinte teorema:

Teorema: Em uma cadeia de Markov irredutível e aperiódica com estados $j = 0, 1, 2, \dots, n$, o conjunto $(\pi)_{i=0}^{n}$ é chamado distribuição estacionária da cadeia de Markov, sendo calculada por:

$$\lim_{n\to\infty} P_{ij}^{n} = \pi_j = \sum_{i=0}^{n} \pi_i P_{ij}$$

$$\sum_{i=0}^{n} \pi_i = 1.$$

Exemplo: Encontre a distribuição estacionária para a cadeia de Markov abaixo.



Para essa cadeia, a matriz de probabilidade de transição é dada por:

$$P = \begin{array}{c} 0 \\ 1 \\ 2 \end{array} \begin{bmatrix} 1/4 & 1/2 & 1/4 \\ 0{,}2 & 0{,}2 & 0{,}6 \\ 0{,}6 & 0{,}1 & 0{,}3 \end{bmatrix}$$

Usando o teorema anterior, temos o sistema de equações:

$$\pi_j = \sum_{i=0}^{n} \pi_i P_{ij} \text{ e } \sum_{i=0}^{n} \pi_i = 1.$$

$$\begin{cases} \pi_1 = \pi_1 P_{11} + \pi_2 P_{21} + \pi_3 P_{31} \\ \pi_2 = \pi_1 P_{12} + \pi_2 P_{22} + \pi_3 P_{32} \\ \pi_3 = \pi_1 P_{13} + \pi_2 P_{23} + \pi_3 P_{33} \\ \pi_1 + \pi_2 + \pi_3 = 1 \end{cases}$$

Para resolvermos esse sistema de equações, podemos usar o site:
https://www.wolframalpha.com/calculators/system-equation-calculator

Resolvendo o sistema de equações, obtemos:

$$\pi_1 = \frac{3}{8};\ \pi_2 = \frac{3}{16};\ \pi_3 = \frac{7}{16}.$$

Notem que a última equação $\pi_1 + \pi_2 + \pi_3 = 1$ deve ser satisfeita, pois π é uma distribuição de probabilidades.

A distribuição estacionária π representa a probabilidade de chegar em cada estado após um número infinito de passos. Assim, podemos escrever:

$$\pi_i = \lim_{n \to \infty} P(X_n = i)$$

Se multiplicarmos a matriz acima por ela mesma 1000 vezes, vamos obter:

$$P^{1000} = \begin{array}{c} 0 \\ 1 \\ 2 \end{array} \begin{bmatrix} 0{,}375 & 0{,}1875 & 0{,}4375 \\ 0{,}375 & 0{,}1875 & 0{,}4375 \\ 0{,}375 & 0{,}1875 & 0{,}4375 \end{bmatrix}$$

Cada coluna da matriz representa exatamente os valores que obtivemos anteriormente:

$$\pi_1 = \frac{3}{8} = 0{,}375;\ \pi_2 = \frac{3}{16} = 0{,}1875;\ \pi_3 = \frac{7}{16} = 0{,}4375.$$

Portanto, a distribuição estacionária fornece a probabilidade de acessar cada estado depois de um número elevado de passos.