Cadeias de Markov

Estatística e Probabilidade para Ciência da Computação

Código: MAD243

Oferecido por:

Prof. Hugo Carvalho (hugo@dme.ufrj.br)
Departamento de Métodos Estatísticos - DME
Instituto de Matemática - IM/UFRJ

Cadeias de Markov – Motivação e história

O que são Cadeias de Markov?

Figura 1: Motivador para Cadeias de Markov: um jogo de tabuleiro

O que são Cadeias de Markov?

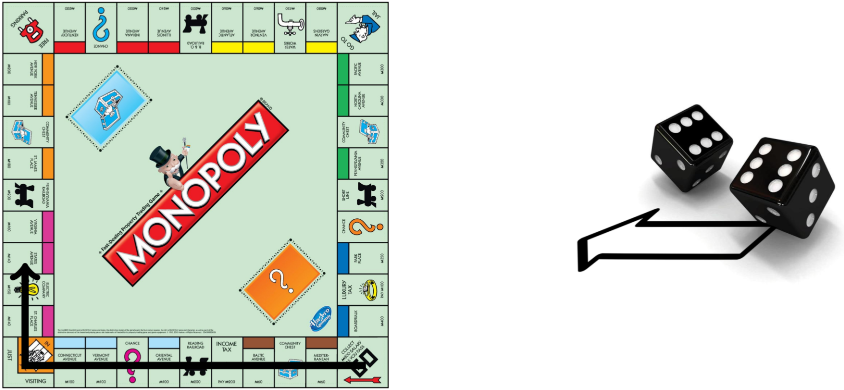


Figura 2: Movimenta-se no jogo através do resultado do lançamento independente de dois dados

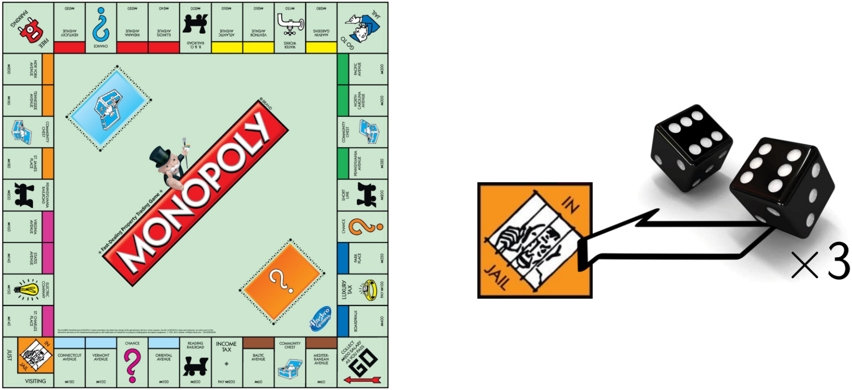


Figura 3: Obs.: Existem outras maneiras de ir para a prisão no Monopoly que não estão presentes no Banco Imobiliário.

O que são Cadeias de Markov?

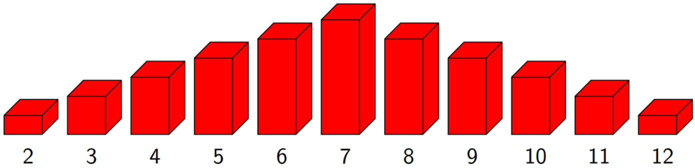


Figura 4: Função de probabilidade da variável aleatória representando a soma no lançamento de dois dados honestos.

O que são Cadeias de Markov?

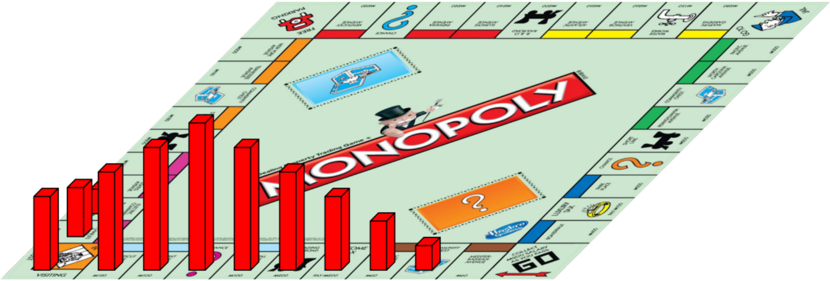


Figura 5: Probabilidade da localização no tabuleiro após uma rodada, levando em consideração somente os dados.

O que são Cadeias de Markov?

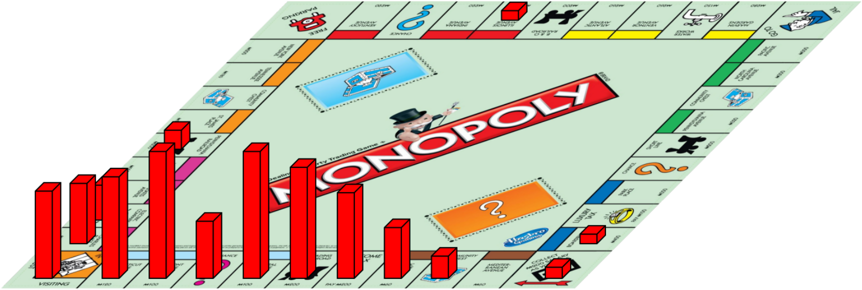


Figura 6: Probabilidade da localização no tabuleiro após uma rodada, com as regras completas do jogo.

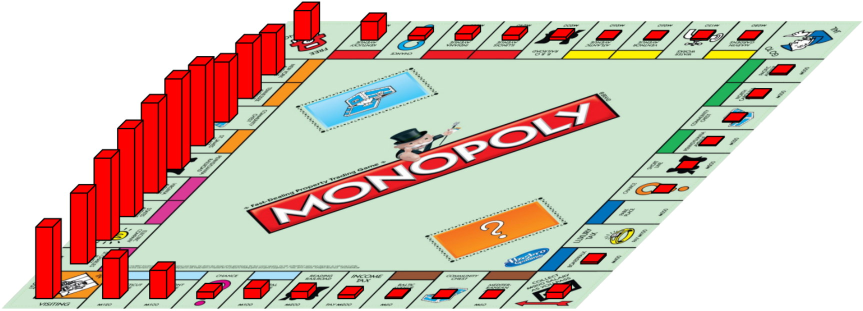


Figura 7: Probabilidade da localização no tabuleiro após duas rodadas.

O que são Cadeias de Markov?

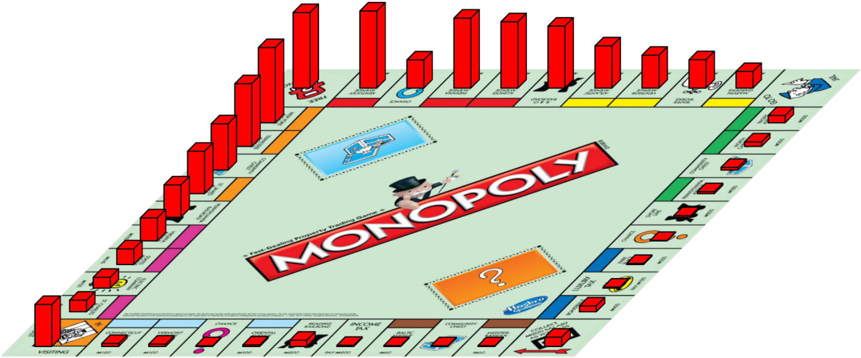


Figura 8: Probabilidade da localização no tabuleiro após três rodadas.

O que são Cadeias de Markov?

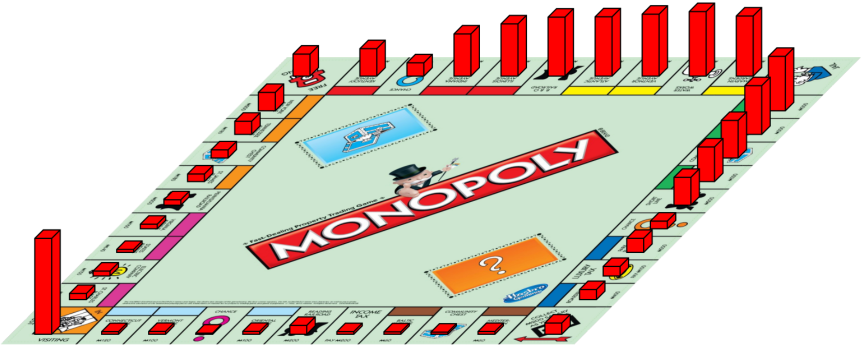


Figura 9: Probabilidade da localização no tabuleiro após quatro rodadas.

O que são Cadeias de Markov?

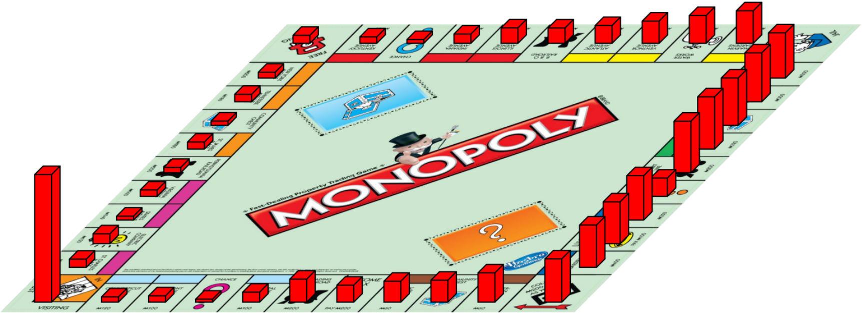


Figura 10: Probabilidade da localização no tabuleiro após cinco rodadas.

O que são Cadeias de Markov?

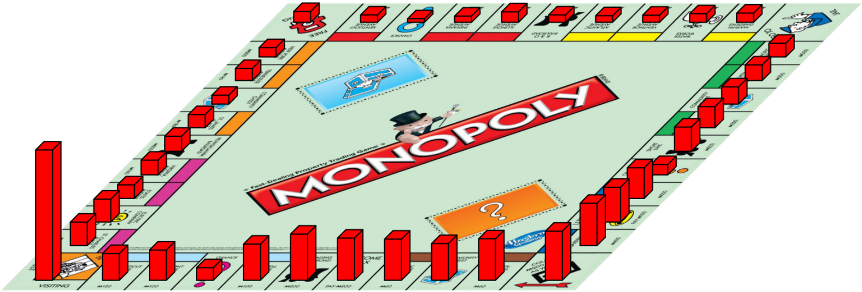


Figura 11: Probabilidade da localização no tabuleiro após seis rodadas.

O que são Cadeias de Markov?

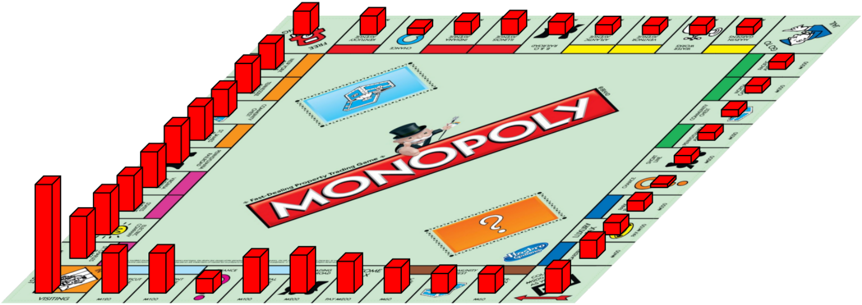


Figura 12: Probabilidade da localização no tabuleiro após sete rodadas.

O que são Cadeias de Markov?

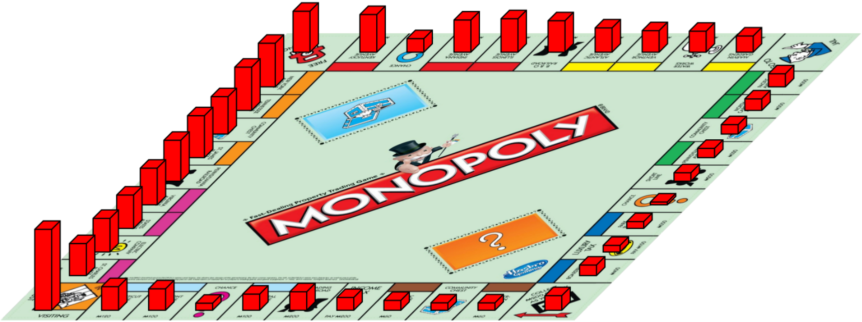


Figura 13: Probabilidade da localização no tabuleiro após oito rodadas.

O que são Cadeias de Markov?

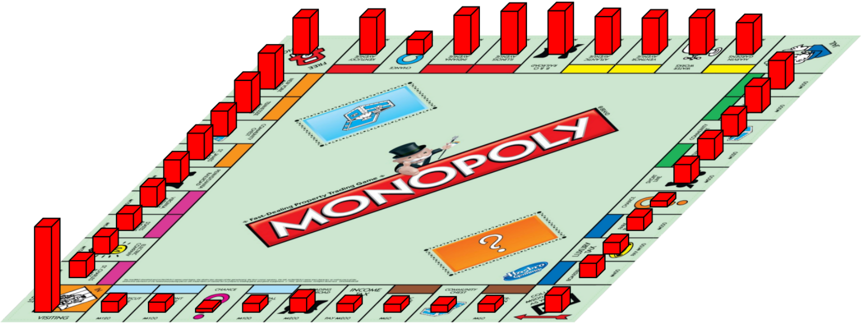


Figura 14: Probabilidade da localização no tabuleiro após nove rodadas.

O que são Cadeias de Markov?

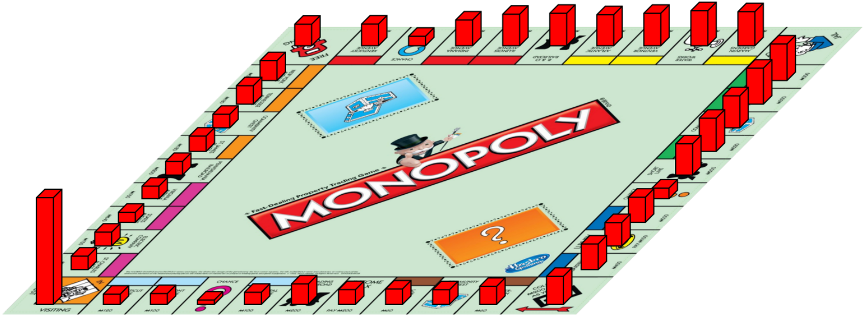


Figura 15: Probabilidade da localização no tabuleiro após dez rodadas.

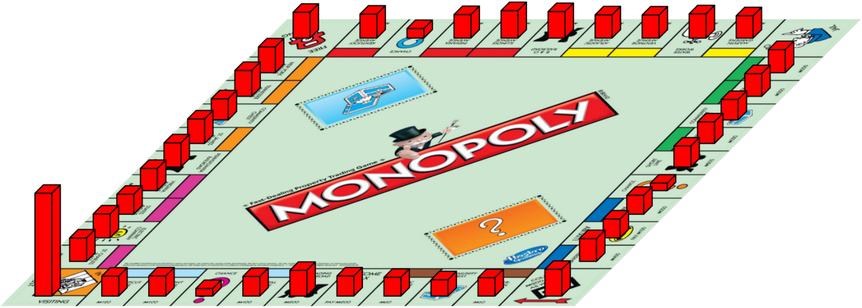


Figura 16: Probabilidade da localização no tabuleiro após infinitas rodadas.

Fonte: `http://carlabernard.ch/beni/downloads/bernard_monopoly.pdf`

Conceitos importantes

- Estados de uma cadeia
  $\Longrightarrow$ casas do tabuleiro

Conceitos importantes

- Estados de uma cadeia
  $\Longrightarrow$ casas do tabuleiro
- Probabilidades de transição
  $\Longrightarrow$ soma das faces para cima no lançamento de dois dados

Conceitos importantes

- Estados de uma cadeia
  $\Longrightarrow$ casas do tabuleiro
- Probabilidades de transição
  $\Longrightarrow$ soma das faces para cima no lançamento de dois dados
- Distribuição estacionária
  $\Longrightarrow$ localização no tabuleiro após infinitas rodadas

Cadeias de Markov

Definição: Uma *cadeia de Markov* é uma sequência infinita de variáveis aleatórias $X_0, X_1, X_2, \ldots, X_n, \ldots$

Cadeias de Markov

Definição: Uma *cadeia de Markov* é uma sequência infinita de variáveis aleatórias $X_0, X_1, X_2, \ldots, X_n, \ldots$ tais que X_{n+1} é independente de $X_{n-1}, X_{n-2}, \ldots, X_2, X_1, X_0$ caso se conheça X_n.

Definição: Uma *cadeia de Markov* é uma sequência infinita de variáveis aleatórias $X_0, X_1, X_2, \ldots, X_n, \ldots$ tais que X_{n+1} é independente de $X_{n-1}, X_{n-2}, \ldots, X_2, X_1, X_0$ caso se conheça X_n.

Definição: Os possíveis valores assumidos pelas variáveis aleatórias são os *estados* da cadeia.

Cadeias de Markov

Definição: Uma *cadeia de Markov* é uma sequência infinita de variáveis aleatórias $X_0, X_1, X_2, \ldots, X_n, \ldots$ tais que X_{n+1} é independente de $X_{n-1}, X_{n-2}, \ldots, X_2, X_1, X_0$ caso se conheça X_n.

Definição: Os possíveis valores assumidos pelas variáveis aleatórias são os *estados* da cadeia.

$\Rightarrow$ É possível fazer uma previsão acerca do futuro conhecendo-se somente o valor atual!

História

- Andrei Markov, 1906: *"Extensão da lei dos grandes números para quantidades que dependem uma das outras"*

А. А. Марков (1886).

História

- Trabalho não motivado por razões práticas, mas por uma disputa de ideias com Pavel Nekrasov

- Suas disputas iam além da Matemática: religião, filosofia, política, etc...

“História”

- Caixa com 3.000 pedras brancas e 2.000 pedras pretas
- Experimento: escolher pedra aleatoriamente, com reposição
- X_t = cor da pedra escolhida no sorteio t, para $t = 1, 2, 3, \ldots$
- $\mathcal{C} = \{\text{preta}, \text{branca}\}$

“História”

- Caixa com 3.000 pedras brancas e 2.000 pedras pretas
- Experimento: escolher pedra aleatoriamente, com reposição
- X_t = cor da pedra escolhida no sorteio t, para $t = 1, 2, 3, \ldots$
- $\mathcal{C} = \{\text{preta}, \text{branca}\}$
- $(X_t)_{t \in \mathbb{N}}$ são variáveis aleatórias independentes:

$$\mathbb{P}(X_{t_1} = c_{t_1}, \ldots, X_{t_k} = c_{t_k}) = \mathbb{P}(X_{t_1} = c_{t_1}) \ldots \mathbb{P}(X_{t_k} = c_{t_k}),$$

para todos $k \geq 2$ e $c_{t_1}, \ldots, c_{t_k} \in \mathcal{C}$, se $t_1 < t_2 < \cdots < t_k$.

“História”

- Caixa com 3.000 pedras brancas e 2.000 pedras pretas
- Experimento: escolher pedra aleatoriamente, com reposição
- $X_t =$ cor da pedra escolhida no sorteio t, para $t = 1, 2, 3, \ldots$
- $\mathcal{C} = \{\text{preta}, \text{branca}\}$
- $(X_t)_{t \in \mathbb{N}}$ são variáveis aleatórias independentes:

$$\mathbb{P}(X_{t_1} = c_{t_1}, \ldots, X_{t_k} = c_{t_k}) = \mathbb{P}(X_{t_1} = c_{t_1}) \ldots \mathbb{P}(X_{t_k} = c_{t_k}),$$

  para todos $k \geq 2$ e $c_{t_1}, \ldots, c_{t_k} \in \mathcal{C}$, se $t_1 < t_2 < \cdots < t_k$.
- $\mathbb{P}(\text{pedra preta ser escolhida}) = 2/5$

“História”

E se não soubéssemos a proporção verdadeira de pedras brancas e pretas na caixa?

"História"

E se não soubéssemos a proporção verdadeira de pedras brancas e pretas na caixa?

- Fazer uma "longa" sequência de retiradas
- Calcular $\dfrac{\text{pedras pretas retiradas}}{\text{total de pedras retiradas}}$

“História”

E se não soubéssemos a proporção verdadeira de pedras brancas e pretas na caixa?

- Fazer uma “longa” sequência de retiradas
- Calcular $\dfrac{\text{pedras pretas retiradas}}{\text{total de pedras retiradas}}$
- Jacob Bernoulli, 1713: *“Com um número suficientemente grande de observações, a proporção computada, com alta probabilidade, estará próxima do valor verdadeiro de 2/5”* – Lei Fraca dos Grandes Números

“História”

E se não soubéssemos a proporção verdadeira de pedras brancas e pretas na caixa?

- Fazer uma “longa” sequência de retiradas
- Calcular $\dfrac{\text{pedras pretas retiradas}}{\text{total de pedras retiradas}}$
- Jacob Bernoulli, 1713: *“Com um número suficientemente grande de observações, a proporção computada, com alta probabilidade, estará próxima do valor verdadeiro de $2/5$”* – Lei Fraca dos Grandes Números
- Lei Forte dos Grandes Números: *“a probabilidade de observar uma sequência de infinitas observações que leve a uma estimação incorreta do valor verdadeiro de $2/5$ é nula”*

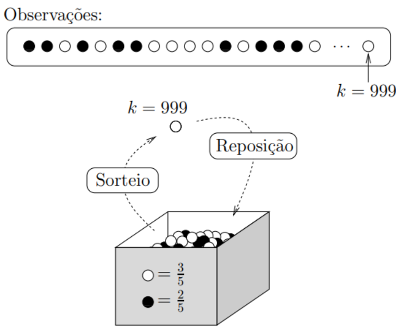
“História”

E se não soubéssemos a proporção verdadeira de pedras brancas e pretas na caixa?

- Fazer uma “longa” sequência de retiradas
- Calcular $\dfrac{\text{pedras pretas retiradas}}{\text{total de pedras retiradas}}$
- Jacob Bernoulli, 1713: *“Com um número suficientemente grande de observações, a proporção computada, com alta probabilidade, estará próxima do valor verdadeiro de 2/5”* – Lei Fraca dos Grandes Números
- Lei Forte dos Grandes Números: *“a probabilidade de observar uma sequência de infinitas observações que leve a uma estimação incorreta do valor verdadeiro de 2/5 é nula”*

É assumida a independência entre as variáveis aleatórias!

"História"



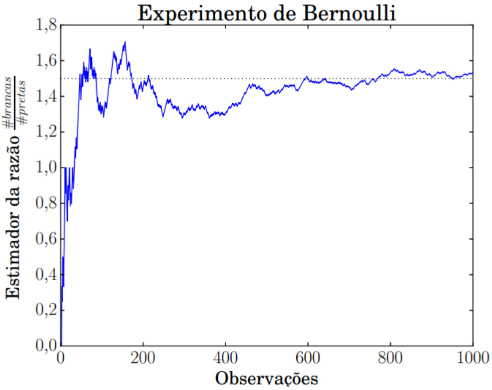


Figura 17: Ilustração do processo de estimação descrito acima.

Fonte: Lucas Simões Maia - *Formalismos da Composição Algorítmica: Um Experimento com Canções Folclóricas Brasileiras* (dissertação de Mestrado em Engenharia Elétrica)

"História"

- Agora, temos duas caixas:
  - Caixa I: 2.000 pedras brancas e 2.000 pedras pretas
  - Caixa II: 3.000 pedras brancas e 2.000 pedras pretas
- Experimento (sorteios com reposição):

“História”

- Agora, temos duas caixas:
  - Caixa I: 2.000 pedras brancas e 2.000 pedras pretas
  - Caixa II: 3.000 pedras brancas e 2.000 pedras pretas
- Experimento (sorteios com reposição):
  1) A primeira pedra c_1 é escolhida ao acaso da caixa II

“História”

- Agora, temos duas caixas:
  - Caixa I: 2.000 pedras brancas e 2.000 pedras pretas
  - Caixa II: 3.000 pedras brancas e 2.000 pedras pretas
- Experimento (sorteios com reposição):
  1) A primeira pedra c_1 é escolhida ao acaso da caixa II
  2) para $k \geq 2$:
     i) Se a pedra anterior c_{k-1} é preta, a pedra c_k é escolhida ao acaso da caixa I
     ii) Caso contrário, se a pedra anterior c_{k-1} é branca, a pedra c_k é escolhida da caixa II

“História”

- Agora, temos duas caixas:
  - Caixa I: 2.000 pedras brancas e 2.000 pedras pretas
  - Caixa II: 3.000 pedras brancas e 2.000 pedras pretas
- Experimento (sorteios com reposição):
  1) A primeira pedra c_1 é escolhida ao acaso da caixa II
  2) para $k \geq 2$:
     i) Se a pedra anterior c_{k-1} é preta, a pedra c_k é escolhida ao acaso da caixa I
     ii) Caso contrário, se a pedra anterior c_{k-1} é branca, a pedra c_k é escolhida da caixa II
- Agora $(X_t)_{t\in\mathbb{N}}$ não são independentes!

“História”

- Agora $\frac{\text{pedras retiradas da caixa I}}{\text{total de pedras retiradas}}$ converge para a probabilidade de se retirar uma pedra preta?

“História”

- Agora $\dfrac{\text{pedras retiradas da caixa I}}{\text{total de pedras retiradas}}$ converge para a probabilidade de se retirar uma pedra preta?
    - **Observação**: Note que só se retira uma pedra da caixa I se a pedra anteriormente sorteada é preta!

“História”

- Agora $\dfrac{\text{pedras retiradas da caixa I}}{\text{total de pedras retiradas}}$ converge para a probabilidade de se retirar uma pedra preta?
  - **Observação**: Note que só se retira uma pedra da caixa I se a pedra anteriormente sorteada é preta!
  - P. Nekrasov, 1902: *“Não! Independência é necessária para valer uma Lei dos Grandes Números”* – argumentos religiosos, baseados em livre arbítrio, não rigorosos

“História”

- Agora $\dfrac{\text{pedras retiradas da caixa I}}{\text{total de pedras retiradas}}$ converge para a probabilidade de se retirar uma pedra preta?
  - **Observação**: Note que só se retira uma pedra da caixa I se a pedra anteriormente sorteada é preta!
  - P. Nekrasov, 1902: *“Não! Independência é necessária para valer uma Lei dos Grandes Números”* – argumentos religiosos, baseados em livre arbítrio, não rigorosos
  - A. Markov, 1906: *“Talvez! Existem processos dependentes para os quais vale uma Lei dos Grandes Números”* – argumentos rigorosos

“História”

- Agora $\dfrac{\text{pedras retiradas da caixa I}}{\text{total de pedras retiradas}}$ converge para a probabilidade de se retirar uma pedra preta?
  - **Observação**: Note que só se retira uma pedra da caixa I se a pedra anteriormente sorteada é preta!
  - P. Nekrasov, 1902: *“Não! Independência é necessária para valer uma Lei dos Grandes Números”* – argumentos religiosos, baseados em livre arbítrio, não rigorosos
  - A. Markov, 1906: *“Talvez! Existem processos dependentes para os quais vale uma Lei dos Grandes Números”* – argumentos rigorosos
- Para mais detalhes, veja E. Seneta, 2013 – “A Tricentenary history of the Law of Large Numbers”

Aplicações

- Primeiro uso "prático" de cadeias de Markov (Markov, 1913):
  - Análise do poema *Eugene Onegin*, de Aleksandr Pushkin
  - Sequência de 20.000 caracteres (!)
  - Probabilidade (estacionária) de vogais: 0,432
  - Probabilidade de uma vogal ser seguida de uma vogal: 0,128
  - Probabilidade de uma consoante ser seguida de uma vogal: 0,663

Aplicações

- Primeiro uso "prático" de cadeias de Markov (Markov, 1913):
  - Análise do poema *Eugene Onegin*, de Aleksandr Pushkin
  - Sequência de 20.000 caracteres (!)
  - Probabilidade (estacionária) de vogais: 0,432
  - Probabilidade de uma vogal ser seguida de uma vogal: 0,128
  - Probabilidade de uma consoante ser seguida de uma vogal: 0,663
- Redescoberto em 1917, A. Erlang – Teoria de Filas

Aplicações

- Primeiro uso “prático” de cadeias de Markov (Markov, 1913):
  - Análise do poema *Eugene Onegin*, de Aleksandr Pushkin
  - Sequência de 20.000 caracteres (!)
  - Probabilidade (estacionária) de vogais: 0,432
  - Probabilidade de uma vogal ser seguida de uma vogal: 0,128
  - Probabilidade de uma consoante ser seguida de uma vogal: 0,663
- Redescoberto em 1917, A. Erlang – Teoria de Filas
- C. Shannon, 1948 – Teoria da Informação

Aplicações

- Primeiro uso "prático" de cadeias de Markov (Markov, 1913):
  - Análise do poema *Eugene Onegin*, de Aleksandr Pushkin
  - Sequência de 20.000 caracteres (!)
  - Probabilidade (estacionária) de vogais: 0,432
  - Probabilidade de uma vogal ser seguida de uma vogal: 0,128
  - Probabilidade de uma consoante ser seguida de uma vogal: 0,663
- Redescoberto em 1917, A. Erlang – Teoria de Filas
- C. Shannon, 1948 – Teoria da Informação
- Algoritmo PageRank (Google)

Aplicações

- Primeiro uso "prático" de cadeias de Markov (Markov, 1913):
    - Análise do poema *Eugene Onegin*, de Aleksandr Pushkin
    - Sequência de 20.000 caracteres (!)
    - Probabilidade (estacionária) de vogais: 0,432
    - Probabilidade de uma vogal ser seguida de uma vogal: 0,128
    - Probabilidade de uma consoante ser seguida de uma vogal: 0,663
- Redescoberto em 1917, A. Erlang – Teoria de Filas
- C. Shannon, 1948 – Teoria da Informação
- Algoritmo PageRank (Google)
- Modelar linguagens

Aplicações

- Primeiro uso "prático" de cadeias de Markov (Markov, 1913):
  - Análise do poema *Eugene Onegin*, de Aleksandr Pushkin
  - Sequência de 20.000 caracteres (!)
  - Probabilidade (estacionária) de vogais: 0,432
  - Probabilidade de uma vogal ser seguida de uma vogal: 0,128
  - Probabilidade de uma consoante ser seguida de uma vogal: 0,663
- Redescoberto em 1917, A. Erlang – Teoria de Filas
- C. Shannon, 1948 – Teoria da Informação
- Algoritmo PageRank (Google)
- Modelar linguagens

Mais aplicações na Wikipedia!

Cadeias de Markov

Definição

Definição: Uma sequência $(X_t)_{t \in \mathbb{N}}$ de variáveis aleatórias é dita uma *cadeia de Markov* se assumem valores no mesmo conjunto finito $\mathcal{C}$ e satisfazem

$$\begin{aligned}\mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_2 = x_2, X_1 = x_1) \\ = \mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n),\end{aligned}$$

para todos $n \in \mathbb{N}$ e $x_{n+1}, \ldots, x_1 \in \mathcal{C}$.

Definição

Definição: Uma sequência $(X_t)_{t\in\mathbb{N}}$ de variáveis aleatórias é dita uma *cadeia de Markov* se assumem valores no mesmo conjunto finito $\mathcal{C}$ e satisfazem

$$\begin{aligned}\mathbb{P}(X_{n+1} = x_{n+1}|X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_2 = x_2, X_1 = x_1)\\ = \mathbb{P}(X_{n+1} = x_{n+1}|X_n = x_n),\end{aligned}$$

para todos $n \in \mathbb{N}$ e $x_{n+1}, \ldots, x_1 \in \mathcal{C}$.

- Elementos de $\mathcal{C}$: *estados* da cadeia

Definição

Definição: Uma sequência $(X_t)_{t\in\mathbb{N}}$ de variáveis aleatórias é dita uma *cadeia de Markov* se assumem valores no mesmo conjunto finito $\mathcal{C}$ e satisfazem

$$\begin{aligned}\mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_2 = x_2, X_1 = x_1) \\ = \mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n),\end{aligned}$$

para todos $n \in \mathbb{N}$ e $x_{n+1}, \ldots, x_1 \in \mathcal{C}$.

- Elementos de $\mathcal{C}$: *estados* da cadeia
- $p_{ij}(n) := \mathbb{P}(X_{n+1} = x_j | X_n = x_i)$: *probabilidades de transição*
  - *"probabilidade de ir para o estado x_j no tempo $n+1$, dado que no tempo n a posição é x_i"*

Definição

Definição: Uma sequência $(X_t)_{t\in\mathbb{N}}$ de variáveis aleatórias é dita uma *cadeia de Markov* se assumem valores no mesmo conjunto finito $\mathcal{C}$ e satisfazem

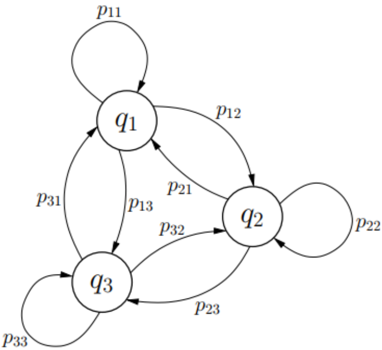
$$\begin{aligned}\mathbb{P}(X_{n+1} = x_{n+1}|X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_2 = x_2, X_1 = x_1)\\ = \mathbb{P}(X_{n+1} = x_{n+1}|X_n = x_n),\end{aligned}$$

para todos $n \in \mathbb{N}$ e $x_{n+1}, \ldots, x_1 \in \mathcal{C}$.

- Elementos de $\mathcal{C}$: *estados* da cadeia
- $p_{ij}(n) := \mathbb{P}(X_{n+1} = x_j|X_n = x_i)$: *probabilidades de transição*
    - *"probabilidade de ir para o estado x_j no tempo $n+1$, dado que no tempo n a posição é x_i"*
- $p_{ij}(n)$ não depende de $n \implies$ *cadeia homogênea*:

$$p_{ij} = \mathbb{P}(X_{n+1} = x_j|X_n = x_i), \ \forall n$$

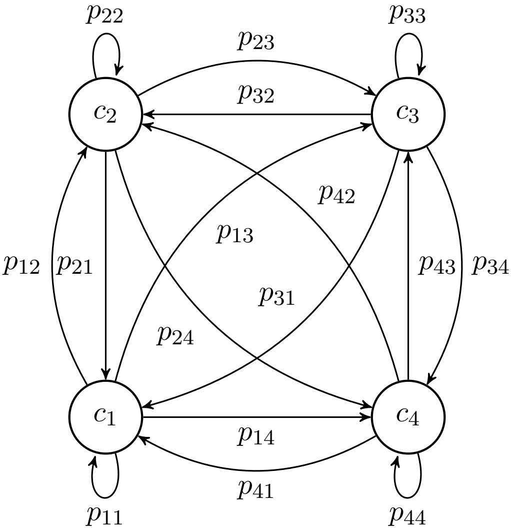
Diagrama e matriz de transição



$$
\begin{array}{ccc}
 & & X_{k+1} \\
 & & \begin{array}{ccc} q_1 & q_2 & q_3 \end{array} \\
X_k & \begin{array}{c} q_1 \\ q_2 \\ q_3 \end{array} & \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}
\end{array}
$$

Fonte: Lucas Simões Maia - *Formalismos da Composição Algorítmica: Um Experimento com Canções Folclóricas Brasileiras* (dissertação de Mestrado em Engenharia Elétrica)

Diagrama e probabilidades de transição (genérica):



Um exemplo

Matriz de transição (genérica e específica):

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix} = \begin{bmatrix} 0 & 0.5 & 0.3 & 0.2 \\ 0.1 & 0 & 0.3 & 0.6 \\ 0.2 & 0.6 & 0 & 0.2 \\ 0.5 & 0.1 & 0.4 & 0 \end{bmatrix}$$

$$p_{ij} = \mathbb{P}(X_{n+1} = x_j | X_n = x_i), \ \forall n \implies \text{cadeia homogênea}$$

Um exemplo

Matriz de transição (genérica e específica):

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix} = \begin{bmatrix} 0 & 0.5 & 0.3 & 0.2 \\ 0.1 & 0 & 0.3 & 0.6 \\ 0.2 & 0.6 & 0 & 0.2 \\ 0.5 & 0.1 & 0.4 & 0 \end{bmatrix}$$

Um exemplo

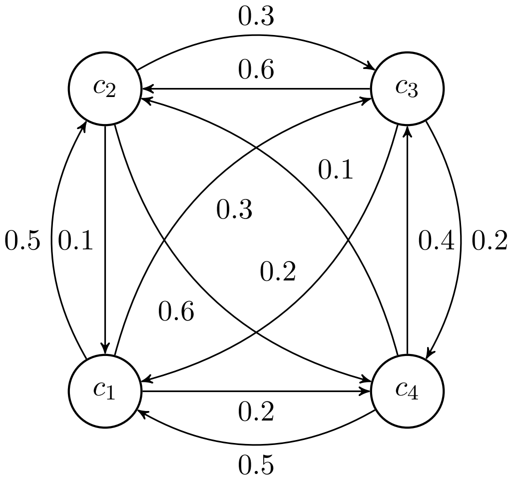
Matriz de transição (genérica e específica):

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix} = \begin{bmatrix} 0 & 0.5 & 0.3 & 0.2 \\ 0.1 & 0 & 0.3 & 0.6 \\ 0.2 & 0.6 & 0 & 0.2 \\ 0.5 & 0.1 & 0.4 & 0 \end{bmatrix}$$

$$p_{ij} = \mathbb{P}(X_{n+1} = x_j | X_n = x_i), \ \forall n \implies \text{cadeia homogênea}$$

Um exemplo

Diagrama e probabilidades de transição (específica):



Um exemplo

Algumas perguntas:

1) Dado que o estado inicial é c_1, qual é a distribuição de probabilidade do segundo estado?
2) E do terceiro estado?
3) Em média, após um longo tempo, qual proporção de tempo a cadeia passa em cada estado?

Um exemplo

1) Dado que o estado inicial é c_1...

$$\boldsymbol{\pi}_0 = [1 \quad 0 \quad 0 \quad 0] \implies \text{distribuição inicial}$$

Um exemplo

1) Dado que o estado inicial é c_1...

$$\boldsymbol{\pi}_0 = [1 \quad 0 \quad 0 \quad 0] \implies \text{distribuição inicial}$$

qual é a distribuição de probabilidade do segundo estado?

Um exemplo

1) Dado que o estado inicial é c_1...

$$\boldsymbol{\pi}_0 = [1 \quad 0 \quad 0 \quad 0] \implies \text{distribuição inicial}$$

qual é a distribuição de probabilidade do segundo estado?

$$\begin{aligned}
\boldsymbol{\pi}_1 &= \boldsymbol{\pi}_0 \mathbf{P} \\
&= [1 \quad 0 \quad 0 \quad 0] \begin{bmatrix} 0 & 0.5 & 0.3 & 0.2 \\ 0.1 & 0 & 0.3 & 0.6 \\ 0.2 & 0.6 & 0 & 0.2 \\ 0.5 & 0.1 & 0.4 & 0 \end{bmatrix} \\
&= [0 \quad 0.5 \quad 0.3 \quad 0.2]
\end{aligned}$$

Um exemplo

1) Dado que o estado inicial é c_1...

$$\boldsymbol{\pi}_0 = [1 \quad 0 \quad 0 \quad 0] \implies \text{distribuição inicial}$$

qual é a distribuição de probabilidade do segundo estado?

$$\begin{aligned}
\boldsymbol{\pi}_1 &= \boldsymbol{\pi}_0 \mathbf{P} \\
&= [1 \quad 0 \quad 0 \quad 0] \begin{bmatrix} 0 & 0.5 & 0.3 & 0.2 \\ 0.1 & 0 & 0.3 & 0.6 \\ 0.2 & 0.6 & 0 & 0.2 \\ 0.5 & 0.1 & 0.4 & 0 \end{bmatrix} \\
&= [0 \quad 0.5 \quad 0.3 \quad 0.2]
\end{aligned}$$

Um exemplo

2) Qual a distribuição de probabilidade do terceiro estado?

Um exemplo

2) Qual a distribuição de probabilidade do terceiro estado?

$$
\begin{aligned}
\boldsymbol{\pi}_2 &= \boldsymbol{\pi}_1 \mathbf{P} \\
&= \boldsymbol{\pi}_0 \mathbf{P}\mathbf{P} \\
&= (\boldsymbol{\pi}_0 \mathbf{P})\mathbf{P} \text{ ou } \boldsymbol{\pi}_0(\mathbf{P}\mathbf{P})
\end{aligned}
$$

Um exemplo

2) Qual a distribuição de probabilidade do terceiro estado?

$$\begin{aligned}\boldsymbol{\pi}_2 &= \boldsymbol{\pi}_1\mathbf{P}\\ &= \boldsymbol{\pi}_0\mathbf{P}\mathbf{P}\\ &= (\boldsymbol{\pi}_0\mathbf{P})\mathbf{P} \text{ ou } \boldsymbol{\pi}_0(\mathbf{P}\mathbf{P})\end{aligned}$$

$$\begin{aligned}(\boldsymbol{\pi}_0\mathbf{P})\mathbf{P} &= \boldsymbol{\pi}_1\mathbf{P}\\ &= \begin{bmatrix}0 & 0.5 & 0.3 & 0.2\end{bmatrix}\begin{bmatrix}0 & 0.5 & 0.3 & 0.2\\ 0.1 & 0 & 0.3 & 0.6\\ 0.2 & 0.6 & 0 & 0.2\\ 0.5 & 0.1 & 0.4 & 0\end{bmatrix}\\ &= \begin{bmatrix}0.21 & 0.20 & 0.23 & 0.36\end{bmatrix}\end{aligned}$$

Um exemplo

2) Qual a distribuição de probabilidade do terceiro estado?

$$\begin{aligned}\boldsymbol{\pi}_2 &= \boldsymbol{\pi}_1\mathbf{P} \\ &= \boldsymbol{\pi}_0\mathbf{P}\mathbf{P} \\ &= (\boldsymbol{\pi}_0\mathbf{P})\mathbf{P} \text{ ou } \boldsymbol{\pi}_0(\mathbf{P}\mathbf{P})\end{aligned}$$

$$\begin{aligned}(\boldsymbol{\pi}_0\mathbf{P})\mathbf{P} &= \boldsymbol{\pi}_1\mathbf{P} \\ &= \begin{bmatrix}0 & 0.5 & 0.3 & 0.2\end{bmatrix}\begin{bmatrix}0 & 0.5 & 0.3 & 0.2 \\ 0.1 & 0 & 0.3 & 0.6 \\ 0.2 & 0.6 & 0 & 0.2 \\ 0.5 & 0.1 & 0.4 & 0\end{bmatrix} \\ &= \begin{bmatrix}0.21 & 0.20 & 0.23 & 0.36\end{bmatrix}\end{aligned}$$

Um exemplo

$$
\begin{aligned}
\boldsymbol{\pi}_0(\mathbf{PP}) &= \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \left(\begin{bmatrix} 0 & 0.5 & 0.3 & 0.2 \\ 0.1 & 0 & 0.3 & 0.6 \\ 0.2 & 0.6 & 0 & 0.2 \\ 0.5 & 0.1 & 0.4 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0.5 & 0.3 & 0.2 \\ 0.1 & 0 & 0.3 & 0.6 \\ 0.2 & 0.6 & 0 & 0.2 \\ 0.5 & 0.1 & 0.4 & 0 \end{bmatrix} \right) \\
&= \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.21 & 0.20 & 0.23 & 0.36 \\ 0.36 & 0.29 & 0.27 & 0.08 \\ 0.16 & 0.12 & 0.32 & 0.40 \\ 0.09 & 0.49 & 0.18 & 0.24 \end{bmatrix} \\
&= \begin{bmatrix} 0.21 & 0.20 & 0.23 & 0.36 \end{bmatrix}
\end{aligned}
$$

Um exemplo

$$\boldsymbol{\pi}_0(\mathbf{PP}) = \begin{bmatrix}1 & 0 & 0 & 0\end{bmatrix} \left(\begin{bmatrix} 0 & 0.5 & 0.3 & 0.2 \\ 0.1 & 0 & 0.3 & 0.6 \\ 0.2 & 0.6 & 0 & 0.2 \\ 0.5 & 0.1 & 0.4 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0.5 & 0.3 & 0.2 \\ 0.1 & 0 & 0.3 & 0.6 \\ 0.2 & 0.6 & 0 & 0.2 \\ 0.5 & 0.1 & 0.4 & 0 \end{bmatrix} \right)$$

$$= \begin{bmatrix}0.21 & 0.20 & 0.23 & 0.36\end{bmatrix}$$

Um exemplo

$$\boldsymbol{\pi}_0(\mathbf{PP}) = [1 \quad 0 \quad 0 \quad 0] \left(\begin{bmatrix} 0 & 0.5 & 0.3 & 0.2 \\ 0.1 & 0 & 0.3 & 0.6 \\ 0.2 & 0.6 & 0 & 0.2 \\ 0.5 & 0.1 & 0.4 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0.5 & 0.3 & 0.2 \\ 0.1 & 0 & 0.3 & 0.6 \\ 0.2 & 0.6 & 0 & 0.2 \\ 0.5 & 0.1 & 0.4 & 0 \end{bmatrix} \right)$$

$$= [1 \quad 0 \quad 0 \quad 0] \begin{bmatrix} 0.21 & 0.20 & 0.23 & 0.36 \\ 0.36 & 0.29 & 0.27 & 0.08 \\ 0.16 & 0.12 & 0.32 & 0.40 \\ 0.09 & 0.49 & 0.18 & 0.24 \end{bmatrix}$$

$$= [0.21 \quad 0.20 \quad 0.23 \quad 0.36]$$

Um exemplo

$$\mathbf{P}^2 = \begin{bmatrix} 0.21 & 0.20 & 0.23 & 0.36 \\ 0.36 & 0.29 & 0.27 & 0.08 \\ 0.16 & 0.12 & 0.32 & 0.40 \\ 0.09 & 0.49 & 0.18 & 0.24 \end{bmatrix} = \begin{bmatrix} p_{11}^{(2)} & p_{12}^{(2)} & p_{13}^{(2)} & p_{14}^{(2)} \\ p_{21}^{(2)} & p_{22}^{(2)} & p_{23}^{(2)} & p_{24}^{(2)} \\ p_{31}^{(2)} & p_{32}^{(2)} & p_{33}^{(2)} & p_{34}^{(2)} \\ p_{41}^{(2)} & p_{42}^{(2)} & p_{43}^{(2)} & p_{44}^{(2)} \end{bmatrix}$$

Um exemplo

$$\mathbf{P}^2 = \begin{bmatrix} 0.21 & 0.20 & 0.23 & 0.36 \\ 0.36 & 0.29 & 0.27 & 0.08 \\ 0.16 & 0.12 & 0.32 & 0.40 \\ 0.09 & 0.49 & 0.18 & 0.24 \end{bmatrix} = \begin{bmatrix} p_{11}^{(2)} & p_{12}^{(2)} & p_{13}^{(2)} & p_{14}^{(2)} \\ p_{21}^{(2)} & p_{22}^{(2)} & p_{23}^{(2)} & p_{24}^{(2)} \\ p_{31}^{(2)} & p_{32}^{(2)} & p_{33}^{(2)} & p_{34}^{(2)} \\ p_{41}^{(2)} & p_{42}^{(2)} & p_{43}^{(2)} & p_{44}^{(2)} \end{bmatrix}$$

$$\begin{aligned} p_{ij}^{(2)} &= \mathbb{P}(X_{n+2} = x_j | X_n = x_i),\ (\forall n) \\ &= \sum_k \mathbb{P}(X_{n+2} = x_j | X_{n+1} = x_k) \mathbb{P}(X_{n+1} = x_k | X_n = x_i) \end{aligned}$$

Um exemplo

$$\mathbf{P}^2 = \begin{bmatrix} 0.21 & 0.20 & 0.23 & 0.36 \\ 0.36 & 0.29 & 0.27 & 0.08 \\ 0.16 & 0.12 & 0.32 & 0.40 \\ 0.09 & 0.49 & 0.18 & 0.24 \end{bmatrix} = \begin{bmatrix} p_{11}^{(2)} & p_{12}^{(2)} & p_{13}^{(2)} & p_{14}^{(2)} \\ p_{21}^{(2)} & p_{22}^{(2)} & p_{23}^{(2)} & p_{24}^{(2)} \\ p_{31}^{(2)} & p_{32}^{(2)} & p_{33}^{(2)} & p_{34}^{(2)} \\ p_{41}^{(2)} & p_{42}^{(2)} & p_{43}^{(2)} & p_{44}^{(2)} \end{bmatrix}$$

$$\begin{aligned} p_{ij}^{(2)} &= \mathbb{P}(X_{n+2} = x_j | X_n = x_i), \ (\forall n) \\ &= \sum_k \mathbb{P}(X_{n+2} = x_j | X_{n+1} = x_k)\mathbb{P}(X_{n+1} = x_k | X_n = x_i) \end{aligned}$$

Mais geralmente...

$$\mathbf{P}^m \text{ tem elementos } p_{ij}^{(m)} = \mathbb{P}(X_{n+m} = x_j | X_n = x_i), \ \forall n$$

Um exemplo

$$\mathbf{P}^2 = \begin{bmatrix} 0.21 & 0.20 & 0.23 & 0.36 \\ 0.36 & 0.29 & 0.27 & 0.08 \\ 0.16 & 0.12 & 0.32 & 0.40 \\ 0.09 & 0.49 & 0.18 & 0.24 \end{bmatrix} = \begin{bmatrix} p_{11}^{(2)} & p_{12}^{(2)} & p_{13}^{(2)} & p_{14}^{(2)} \\ p_{21}^{(2)} & p_{22}^{(2)} & p_{23}^{(2)} & p_{24}^{(2)} \\ p_{31}^{(2)} & p_{32}^{(2)} & p_{33}^{(2)} & p_{34}^{(2)} \\ p_{41}^{(2)} & p_{42}^{(2)} & p_{43}^{(2)} & p_{44}^{(2)} \end{bmatrix}$$

$$\begin{aligned} p_{ij}^{(2)} &= \mathbb{P}(X_{n+2} = x_j | X_n = x_i), \ (\forall n) \\ &= \sum_k \mathbb{P}(X_{n+2} = x_j | X_{n+1} = x_k) \mathbb{P}(X_{n+1} = x_k | X_n = x_i) \end{aligned}$$

Mais geralmente...

$$\mathbf{P}^m \text{ tem elementos } p_{ij}^{(m)} = \mathbb{P}(X_{n+m} = x_j | X_n = x_i), \ \forall n$$

$$\Longrightarrow \text{ Equações de Chapman-Kolmogorov}$$

Um exemplo

3) Em média, após um longo tempo, qual proporção de tempo a cadeia passa em cada estado?

Um exemplo

3) Em média, após um longo tempo, qual proporção de tempo a cadeia passa em cada estado?

$$
\begin{aligned}
p_{ij}^{(m)} &= \mathbb{P}(X_{n+m} = x_j | X_n = x_i) \\
&= \mathbb{P}(X_m = x_j | X_0 = x_i)
\end{aligned}
$$

Um exemplo

3) Em média, após um longo tempo, qual proporção de tempo a cadeia passa em cada estado?

$$\begin{aligned} p_{ij}^{(m)} &= \mathbb{P}(X_{n+m} = x_j | X_n = x_i) \\ &= \mathbb{P}(X_m = x_j | X_0 = x_i) \end{aligned}$$

$$\implies p_{ij}^{(\infty)} = \mathbb{P}(X_\infty = x_j | X_0 = x_i)$$

Um exemplo

3) Em média, após um longo tempo, qual proporção de tempo a cadeia passa em cada estado?

$$
\begin{aligned}
p_{ij}^{(m)} &= \mathbb{P}(X_{n+m} = x_j | X_n = x_i) \\
&= \mathbb{P}(X_m = x_j | X_0 = x_i)
\end{aligned}
$$

$$
\Longrightarrow \; p_{ij}^{(\infty)} = \mathbb{P}(X_\infty = x_j | X_0 = x_i)
$$

$$
\mathbf{P}^\infty \approx \begin{bmatrix} 0.2083 & 0.2807 & 0.2508 & 0.2602 \\ 0.2083 & 0.2807 & 0.2508 & 0.2602 \\ 0.2083 & 0.2807 & 0.2508 & 0.2602 \\ 0.2083 & 0.2807 & 0.2508 & 0.2602 \end{bmatrix}
$$

Um exemplo

$$\boldsymbol{\pi}_\infty = [0.2083 \quad 0.2807 \quad 0.2508 \quad 0.2602]$$

$$\Longrightarrow \text{ Distribuição estacionária}$$

Um exemplo

$$\boldsymbol{\pi}_\infty = [0.2083 \quad 0.2807 \quad 0.2508 \quad 0.2602]$$

$\Longrightarrow$ Distribuição estacionária

$$\begin{aligned}\boldsymbol{\pi}_\infty \mathbf{P} &= [0.2083 \quad 0.2807 \quad 0.2508 \quad 0.2602] \begin{bmatrix} 0 & 0.5 & 0.3 & 0.2 \\ 0.1 & 0 & 0.3 & 0.6 \\ 0.2 & 0.6 & 0 & 0.2 \\ 0.5 & 0.1 & 0.4 & 0 \end{bmatrix} \\ &= [0.2083 \quad 0.2807 \quad 0.2508 \quad 0.2602] \\ &= \boldsymbol{\pi}_\infty\end{aligned}$$

$\Longrightarrow$ A distribuição estacionária satisfaz $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$!

Um exemplo

$$\boldsymbol{\pi}_\infty = [0.2083 \;\; 0.2807 \;\; 0.2508 \;\; 0.2602]$$

$$\Longrightarrow \text{ Distribuição estacionária}$$

$$\begin{aligned}
\boldsymbol{\pi}_\infty \mathbf{P} &= [0.2083 \;\; 0.2807 \;\; 0.2508 \;\; 0.2602] \begin{bmatrix} 0 & 0.5 & 0.3 & 0.2 \\ 0.1 & 0 & 0.3 & 0.6 \\ 0.2 & 0.6 & 0 & 0.2 \\ 0.5 & 0.1 & 0.4 & 0 \end{bmatrix} \\
&= [0.2083 \;\; 0.2807 \;\; 0.2508 \;\; 0.2602] \\
&= \boldsymbol{\pi}_\infty
\end{aligned}$$

Um exemplo

$$\boldsymbol{\pi}_\infty = [0.2083 \;\; 0.2807 \;\; 0.2508 \;\; 0.2602]$$

$$\implies \text{Distribuição estacionária}$$

$$\begin{aligned}
\boldsymbol{\pi}_\infty \mathbf{P} &= [0.2083 \;\; 0.2807 \;\; 0.2508 \;\; 0.2602] \begin{bmatrix} 0 & 0.5 & 0.3 & 0.2 \\ 0.1 & 0 & 0.3 & 0.6 \\ 0.2 & 0.6 & 0 & 0.2 \\ 0.5 & 0.1 & 0.4 & 0 \end{bmatrix} \\
&= [0.2083 \;\; 0.2807 \;\; 0.2508 \;\; 0.2602] \\
&= \boldsymbol{\pi}_\infty
\end{aligned}$$

$$\implies \text{A distribuição estacionária satisfaz } \boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}!$$

Teorema ("Perron-Frobenius" e "Markov"): Seja $\mathbf{P}$ a matriz de transição de uma cadeia de Markov $(X_n)_{n\in\mathbb{N}}$. Assuma que $\mathbf{P}$ é *regular*, ou seja, $\mathbf{P}^m$ tem todas as entradas positivas para algum valor de $m \geq 1$. Então vale que:

i) A cadeia tem uma única distribuição estacionária $\boldsymbol{\pi}$;

ii) Quando $m \to \infty$, $\mathbf{P}^m$ converge para uma matriz cujas linhas são todas iguais a $\boldsymbol{\pi}$;

iii) Vale o seguinte teorema limite: $\displaystyle\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} X_n \xrightarrow{p} \mathbb{E}[\boldsymbol{\pi}]$.

Obs.: Ao dispensar a independência, mais hipóteses são necessárias para garantir uma LGN!

Distribuições estacionária vs. limite

Teorema: Se $\boldsymbol{\pi}$ satisfaz a *equação de balanço detalhado*

$$\pi_i p_{ij} = p_{ji} \pi_j,$$

então $\boldsymbol{\pi}$ é uma distribuição estacionária para $\mathbf{P}$.

Distribuições estacionária vs. limite

Teorema: Se $\boldsymbol{\pi}$ satisfaz a *equação de balanço detalhado*

$$\pi_i p_{ij} = p_{ji} \pi_j,$$

então $\boldsymbol{\pi}$ é uma distribuição estacionária para $\mathbf{P}$.

- Distribuição estacionária – inalterada por $\mathbf{P}$

Distribuições estacionária vs. limite

Teorema: Se $\boldsymbol{\pi}$ satisfaz a *equação de balanço detalhado*

$$\pi_i p_{ij} = p_{ji} \pi_j,$$

então $\boldsymbol{\pi}$ é uma distribuição estacionária para $\mathbf{P}$.

- Distribuição estacionária – inalterada por $\mathbf{P}$
- Distribuição limite – π_j = proporção de tempo gasta no estado x_j

Distribuições estacionária vs. limite

Teorema: Se $\boldsymbol{\pi}$ satisfaz a *equação de balanço detalhado*

$$\pi_i p_{ij} = p_{ji} \pi_j,$$

então $\boldsymbol{\pi}$ é uma distribuição estacionária para $\mathbf{P}$.

- Distribuição estacionária – inalterada por $\mathbf{P}$
- Distribuição limite – π_j = proporção de tempo gasta no estado x_j

Exemplo: $\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ tem $\boldsymbol{\pi} = [0.5 \;\; 0.5]$ como distribuição estacionária, mas não tem distribuição limite!

Distribuições estacionária vs. limite

Teorema: Se $\boldsymbol{\pi}$ satisfaz a *equação de balanço detalhado*

$$\pi_i p_{ij} = p_{ji}\pi_j,$$

então $\boldsymbol{\pi}$ é uma distribuição estacionária para $\mathbf{P}$.

- Distribuição estacionária – inalterada por $\mathbf{P}$
- Distribuição limite – π_j = proporção de tempo gasta no estado x_j

Exemplo: $\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ tem $\boldsymbol{\pi} = [0.5 \;\; 0.5]$ como distribuição estacionária, mas não tem distribuição limite!

Atenção: Estamos assumindo a matriz $\mathbf{P}$ conhecida!

Estimando uma matriz de transição

- $X_1, X_2, X_3, \ldots$ cadeia de Markov (homogênea)

Estimando uma matriz de transição

- $X_1, X_2, X_3, \ldots$ cadeia de Markov (homogênea)
- $\mathcal{J} = \{z_1, z_2, z_3, \ldots, z_n\}$ observações – realização da cadeia

Estimando uma matriz de transição

- $X_1, X_2, X_3, \ldots$ cadeia de Markov (homogênea)
- $\mathcal{J} = \{z_1, z_2, z_3, \ldots, z_n\}$ observações – realização da cadeia
- $\widehat{\mathbf{M}}$ matriz de contagem: m_{ij} é a quantidade de vezes que a transição $x_i \to x_j$ acontece em $\mathcal{J}$

Estimando uma matriz de transição

- $X_1, X_2, X_3, \ldots$ cadeia de Markov (homogênea)
- $\mathcal{J} = \{z_1, z_2, z_3, \ldots, z_n\}$ observações – realização da cadeia
- $\widehat{\mathbf{M}}$ matriz de contagem: m_{ij} é a quantidade de vezes que a transição $x_i \to x_j$ acontece em $\mathcal{J}$
- $\widehat{\mathbf{P}}$ matriz de transição estimada: $\widehat{p}_{ij} = \dfrac{m_{ij}}{\sum_j m_{ij}}$,proporção das transições $x_i \to x_j$ dentre todas as transições partindo de x_i

Estimando uma matriz de transição

- $X_1, X_2, X_3, \ldots$ cadeia de Markov (homogênea)
- $\mathcal{J} = \{z_1, z_2, z_3, \ldots, z_n\}$ observações – realização da cadeia
- $\widehat{\mathbf{M}}$ matriz de contagem: m_{ij} é a quantidade de vezes que a transição $x_i \to x_j$ acontece em $\mathcal{J}$
- $\widehat{\mathbf{P}}$ matriz de transição estimada: $\widehat{p}_{ij} = \dfrac{m_{ij}}{\sum_j m_{ij}}$,proporção das transições $x_i \to x_j$ dentre todas as transições partindo de x_i
- Estimativa de máxima verossimilhança: "$\widehat{\mathbf{P}}$ *é a matriz de transição para a qual as observações* $\mathcal{J}$ *são as mais prováveis*"

Estimando uma matriz de transição

- $X_1, X_2, X_3, \ldots$ cadeia de Markov (homogênea)
- $\mathcal{J} = \{z_1, z_2, z_3, \ldots, z_n\}$ observações – realização da cadeia
- $\widehat{\mathbf{M}}$ matriz de contagem: m_{ij} é a quantidade de vezes que a transição $x_i \to x_j$ acontece em $\mathcal{J}$
- $\widehat{\mathbf{P}}$ matriz de transição estimada: $\widehat{p}_{ij} = \dfrac{m_{ij}}{\sum_j m_{ij}}$,proporção das transições $x_i \to x_j$ dentre todas as transições partindo de x_i
- Estimativa de máxima verossimilhança: *"$\widehat{\mathbf{P}}$ é a matriz de transição para a qual as observações $\mathcal{J}$ são as mais prováveis"*
- Estimador consistente: *"se $n \to \infty$, $\widehat{\mathbf{P}}$ converge para a matriz 'verdadeira' $\mathbf{P}$"*