Teoria da Informação (#3)

Processos estocásticos, cadeias de Markov, distribuição estacionária, ritmo de entropia

Miguel Barão

ESCOLA DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE INFORMÁTICA

Processo determinístico: o futuro é perfeitamente determinado.

Processos determinísticos e estocásticos

Processo determinístico: o futuro é perfeitamente determinado.

Processo estocástico: o futuro é incerto, existindo várias possibilidades de evolução com probabilidades associadas.

Processo determinístico: o futuro é perfeitamente determinado.

Processo estocástico: o futuro é incerto, existindo várias possibilidades de evolução com probabilidades associadas.

Exemplos

1. Execução de um algoritmo é um *processo determinístico*.
2. Lançamento de um dado é um *processo estocástico*.
3. Cotação das acções na bolsa de valores é um *processo estocástico*.
4. Evolução da temperatura ao longo do tempo é um *processo estocástico*.

- A evolução de um processo estocástico é caracterizada pela evolução de uma variável aleatória X_t ao longo do tempo t. Por exemplo:

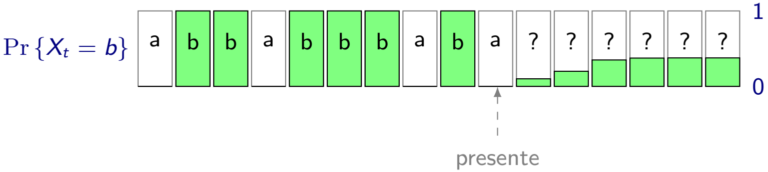
$$\underbrace{\ldots, a, b, b, a, b, b, b, a, b,}_{\text{passado}} \; \underbrace{a,}_{\text{presente}} \; \underbrace{?, ?, ?, ?, ?, ?, \ldots}_{\textit{futuro}}$$

Evolução no tempo

- A evolução de um processo estocástico é caracterizada pela evolução de uma variável aleatória X_t ao longo do tempo t. Por exemplo:

$$\underbrace{\ldots, a, b, b, a, b, b, b, a, b,}_{\text{passado}} \; \underbrace{a,}_{\text{presente}} \; \underbrace{?, ?, ?, ?, ?, ?, \ldots}_{\textit{futuro}}$$

- Em cada instante de tempo t, a variável aleatória X_t tem distribuição de probabilidade $p(x_t)$.

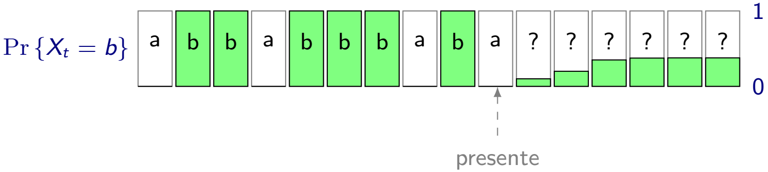


Evolução no tempo

- A evolução de um processo estocástico é caracterizada pela evolução de uma variável aleatória X_t ao longo do tempo t. Por exemplo:

$$\underbrace{\ldots, a, b, b, a, b, b, b, a, b,}_{\text{passado}} \; \underbrace{a,}_{\text{presente}} \; \underbrace{?, ?, ?, ?, ?, ?, \ldots}_{\textit{futuro}}$$

- Em cada instante de tempo t, a variável aleatória X_t tem distribuição de probabilidade $p(x_t)$.



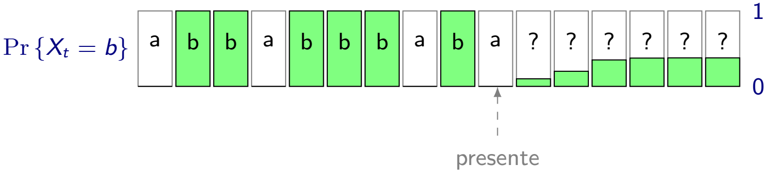
- Os valores passados de X_t não contêm incerteza uma vez que já foram observados. Assim a probabilidade $\Pr\{X_t = b\}$ ou é 0 ou 1.

Evolução no tempo

- A evolução de um processo estocástico é caracterizada pela evolução de uma variável aleatória X_t ao longo do tempo t. Por exemplo:

$$\underbrace{\ldots, a, b, b, a, b, b, b, a, b,}_{\text{passado}} \; \underbrace{a,}_{\text{presente}} \; \underbrace{?, ?, ?, ?, ?, ?, \ldots}_{\textit{futuro}}$$

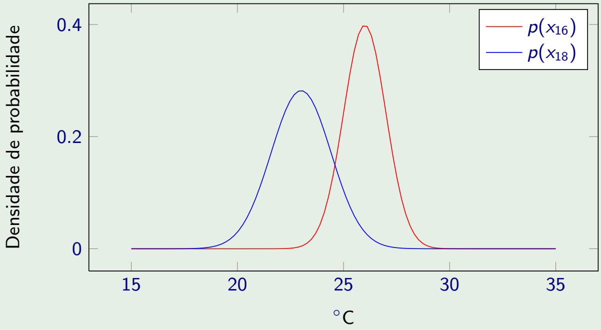
- Em cada instante de tempo t, a variável aleatória X_t tem distribuição de probabilidade $p(x_t)$.



- Os valores passados de X_t não contêm incerteza uma vez que já foram observados. Assim a probabilidade $\Pr\{X_t = b\}$ ou é 0 ou 1.
- A variável X_t em instantes de tempo futuros é uma variável aleatória com distribuição de probabilidade $p(x_t)$. Esta distribuição pode variar ao longo do tempo (ver figura).

Exemplo (Variáveis aleatórias contínuas)

- Neste momento são 14:00 e estão 28.1°C.
- Às 16:00 horas (futuro) a temperatura é uma variável aleatória X_t com uma certa densidade de probabilidade $p(x_t)$.
- Às 18:37 a temperatura tem uma densidade de probabilidade diferente.



A temperatura X_t é um processo estocástico em que a variável aleatória X_t é contínua (o alfabeto são os números reais) e evolui em tempo contínuo. A função densidade de probabilidade $p(x_t)$ varia ao longo do tempo.

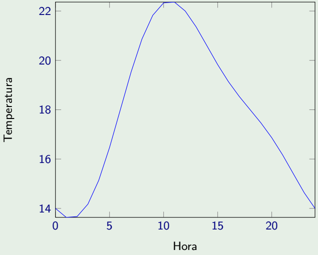
Tempo contínuo e tempo discreto

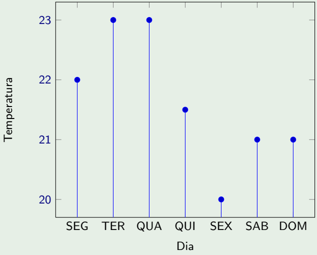
- Uma variável pode evoluir em tempo contínuo ou tempo discreto.

- Uma variável pode evoluir em tempo contínuo ou tempo discreto.

Exemplos

- Evolução da temperatura ao longo do dia → *tempo contínuo*;
- Temperaturas máximas ao longo da semana → *tempo discreto*.





Definição (Processo estocástico)

Um processo estocástico é uma sequência de variáveis aleatórias $X_1, X_2, \ldots, X_n$ e é, em geral, caracterizado pela distribuição de probabilidade conjunta $p(x_1, x_2, \ldots, x_n)$.

O exemplo anterior da evolução da temperatura máxima ao longo de uma sequência de dias é um processo estocástico.

Definição (Processo estocástico)

Um processo estocástico é uma sequência de variáveis aleatórias $X_1, X_2, \ldots, X_n$ e é, em geral, caracterizado pela distribuição de probabilidade conjunta $p(x_1, x_2, \ldots, x_n)$.

O exemplo anterior da evolução da temperatura máxima ao longo de uma sequência de dias é um processo estocástico.

Definição (Processo estocástico estacionário)

Um processo estocástico diz-se estacionário se a distribuição de probabilidade conjunta de qualquer número n de v.a. mantém-se inalterada quando se efectuam deslocamentos no tempo

$$\Pr\{X_1 = x_1, \ldots, X_n = x_n\} = \Pr\{X_{1+m} = x_1, \ldots, X_{n+m} = x_n\}$$

para qualquer deslocamento m.

O exemplo das temperaturas máximas não é um processo estocástico estacionário, pois no inverno é mais provável termos temperaturas mais baixas que no verão. As probabilidades do estado do tempo variam ao longo do ano.

Definição (Processo de Markov)

Diz-se que um processo estocástico é um processo de Markov se cada variável aleatória depende da precedente e é condicionalmente independente das restantes:

$$p(x_n|x_{n-1},\ldots,x_{n-m}) = p(x_n|x_{n-1}).$$

Definição (Processo de Markov)

Diz-se que um processo estocástico é um processo de Markov se cada variável aleatória depende da precedente e é condicionalmente independente das restantes:

$$p(x_n|x_{n-1},\dots,x_{n-m}) = p(x_n|x_{n-1}).$$

A sequência de variáveis aleatórias $X_1, X_2, X_3, X_4, \dots$, com alfabeto $X_i \in \mathcal{X}$, é um processo de Markov quando se verificam as dependências:

$$X_1 \xrightarrow[p(x_2|x_1)]{} X_2 \xrightarrow[p(x_3|x_2)]{} X_3 \xrightarrow[p(x_4|x_3)]{} X_4 \longrightarrow \cdots$$

Definição (Processo de Markov)

Diz-se que um processo estocástico é um processo de Markov se cada variável aleatória depende da precedente e é condicionalmente independente das restantes:

$$p(x_n|x_{n-1},\ldots,x_{n-m}) = p(x_n|x_{n-1}).$$

A sequência de variáveis aleatórias $X_1, X_2, X_3, X_4, \ldots$, com alfabeto $X_i \in \mathcal{X}$, é um processo de Markov quando se verificam as dependências:

$$X_1 \xrightarrow[p(x_2|x_1)]{} X_2 \xrightarrow[p(x_3|x_2)]{} X_3 \xrightarrow[p(x_4|x_3)]{} X_4 \longrightarrow \cdots$$

Num processo de Markov, as variáveis não têm necessariamente de ter o mesmo alfabeto. Por exemplo, as variáveis X, Y, Z podem ter alfabetos $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ diferentes e formar um processo de Markov

$$X \xrightarrow[p(y|x)]{} Y \xrightarrow[p(z|y)]{} Z$$

Exemplos

1. Seja X_t a variável aleatória que indica se chove ou não num determinado dia. A sequência $X_1, X_2, \ldots, X_n$, que indica o estado do tempo em dias consecutivos, forma um processo estocástico não estacionário, pois a probabilidade de chuva varia ao longo do ano.

Exemplos

1. Seja X_t a variável aleatória que indica se chove ou não num determinado dia. A sequência $X_1, X_2, \ldots, X_n$, que indica o estado do tempo em dias consecutivos, forma um processo estocástico não estacionário, pois a probabilidade de chuva varia ao longo do ano.
2. Uma sequência de lançamentos de um dado é um processo estocástico estacionário pois as probabilidades das faces são constantes ao longo dos vários lançamentos.

Exemplos

1. Seja X_t a variável aleatória que indica se chove ou não num determinado dia. A sequência $X_1, X_2, \ldots, X_n$, que indica o estado do tempo em dias consecutivos, forma um processo estocástico não estacionário, pois a probabilidade de chuva varia ao longo do ano.
2. Uma sequência de lançamentos de um dado é um processo estocástico estacionário pois as probabilidades das faces são constantes ao longo dos vários lançamentos.
3. No jogo da glória, a sequência $X_1, X_2, \ldots, X_n$ representa a posição de uma peça ao longo do tempo. Esta sequência é um processo de Markov: a probabilidade da peça se encontrar numa dada casa depende apenas da posição anterior e não de toda a história de jogo passada (independência condicional).

Definição (Cadeia de Markov)

Um processo de Markov em que a variável X_n toma valores num alfabeto finito (ou infinito enumerável) chama-se cadeia de Markov.

Definição (Cadeia de Markov)

Um processo de Markov em que a variável X_n toma valores num alfabeto finito (ou infinito enumerável) chama-se cadeia de Markov.

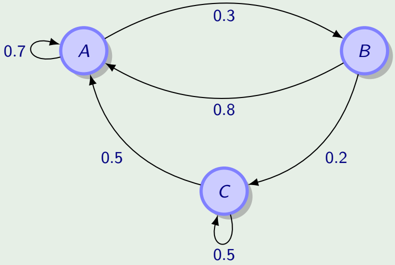
- Funciona como uma máquina de estados onde as transições são estocásticas.
- As cadeias de Markov podem ser representadas por grafos.

Definição (Cadeia de Markov)

Um processo de Markov em que a variável X_n toma valores num alfabeto finito (ou infinito enumerável) chama-se cadeia de Markov.

- Funciona como uma máquina de estados onde as transições são estocásticas.
- As cadeias de Markov podem ser representadas por grafos.
- nós representam os estados.
- arcos representam as probabilidades de transição $p(x_t|x_{t-1})$ entre os estados

Exemplo (Cadeia de Markov)



Estados: $\mathcal{X} = \{A, B, C\}$

Definição (Cadeia de Markov homogénea no tempo)

Diz-se que a cadeia de Markov é homogénea no tempo se as probabilidades de transição são constantes ao longo do tempo.

Definição (Cadeia de Markov homogénea no tempo)

Diz-se que a cadeia de Markov é homogénea no tempo se as probabilidades de transição são constantes ao longo do tempo.

Atenção

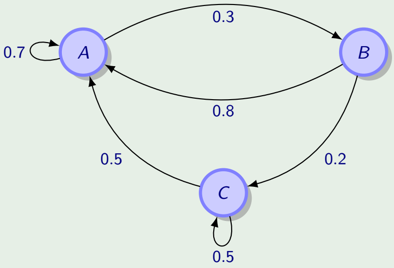
Não confundir "homogéneo no tempo" com "estacionário".

Homogéneo no tempo diz respeito ao mecanismo de funcionamento (probabilidades de transição)

Estacionário diz respeito ao estado (probabilidade da cadeia se encontrar em certos estados)

Uma cadeia de Markov homogénea no tempo pode ou não encontrar-se em regime estacionário.

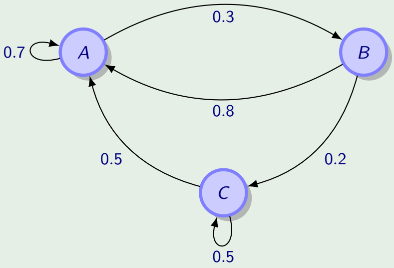
Exemplo (Probabilidade de uma sequência de estados)



Questões:

1. Esta cadeia de Markov é homogénea no tempo?

Exemplo (Probabilidade de uma sequência de estados)



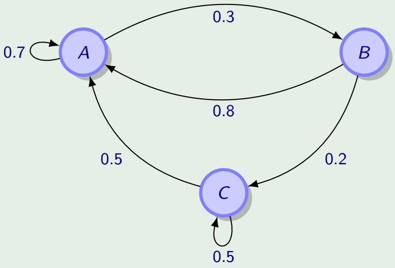
Questões:

1. Esta cadeia de Markov é homogénea no tempo?

Respostas:

1. É homogénea no tempo porque as probabilidades de transição são constantes.

Exemplo (Probabilidade de uma sequência de estados)



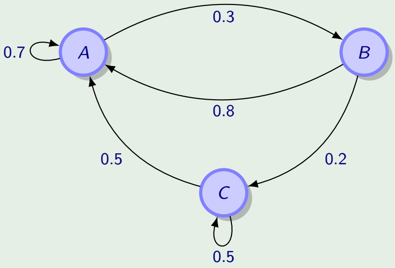
Questões:

1. Esta cadeia de Markov é homogénea no tempo?
2. Sabendo que o estado inicial é $X_0 = A$, qual a probabilidade de observar a sequência $X_0X_1X_2X_3 = ABCA$?

Respostas:

1. É homogénea no tempo porque as probabilidades de transição são constantes.

Exemplo (Probabilidade de uma sequência de estados)



Questões:

1. Esta cadeia de Markov é homogénea no tempo?
2. Sabendo que o estado inicial é $X_0 = A$, qual a probabilidade de observar a sequência $X_0X_1X_2X_3 = ABCA$?

Respostas:

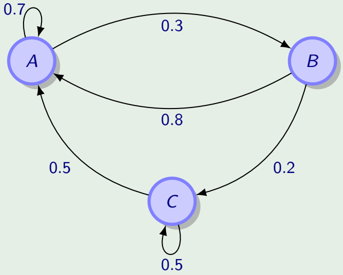
1. É homogénea no tempo porque as probabilidades de transição são constantes.
2. Usando as probabilidades condicionadas temos que

$$p(x_0, x_1, x_2, x_3) = p(x_0)p(x_1|x_0)p(x_2|x_1)p(x_3|x_2),$$

então

$$p(ABCA) = 1 \times 0.3 \times 0.2 \times 0.5 = 0.03.$$

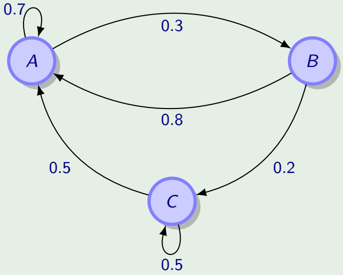
Exemplo (Probabilidade de atingir um certo estado)



Questão:

1. Sabendo que o estado inicial é $X_0 = A$, qual a probabilidade de obtermos $X_3 = C$?
   (*i.e.*, a probabilidade de atingirmos o estado C ao fim de exactamente três instantes de tempo.)

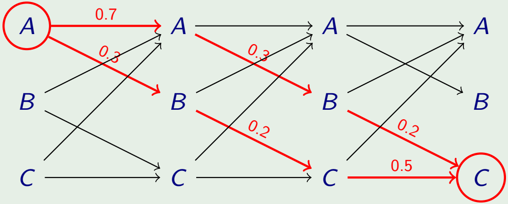
Exemplo (Probabilidade de atingir um certo estado)



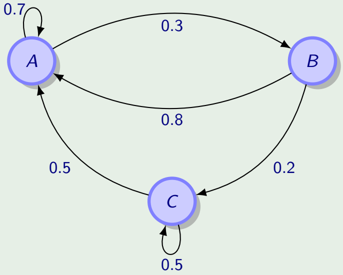
Questão:

1. Sabendo que o estado inicial é $X_0 = A$, qual a probabilidade de obtermos $X_3 = C$?
(*i.e.*, a probabilidade de atingirmos o estado C ao fim de exactamente três instantes de tempo.)

A probabilidade de atingir $X_3 = C$ a partir de $X_0 = A$ é calculada percorrendo todos os caminhos que atingem C em três passos:



Exemplo (Probabilidade de atingir um certo estado)



Questão:

1. Sabendo que o estado inicial é $X_0 = A$, qual a probabilidade de obtermos $X_3 = C$? (*i.e.*, a probabilidade de atingirmos o estado C ao fim de exactamente três instantes de tempo.)

A probabilidade de atingir $X_3 = C$ a partir de $X_0 = A$ é calculada percorrendo todos os caminhos que atingem C em três passos:

$$\begin{aligned}
\Pr\{X_3 = C\} &= \Pr\{(X_0X_1X_2X_3 = AABC) \cup (X_0X_1X_2X_3 = ABCC)\} \\
&= \Pr\{X_0X_1X_2X_3 = AABC\} + \Pr\{X_0X_1X_2X_3 = ABCC\} \\
&= 0.7 \times 0.3 \times 0.2 + 0.3 \times 0.2 \times 0.5 \\
&= 0.072
\end{aligned}$$

Embora correcta, esta não é a maneira mais fácil pois o número de percursos tende a crescer exponencialmente com o o número de passos!

A cadeia de Markov tem a propriedade de que o estado X_n depende apenas de X_{n-1}. A distribuição condicionada $p(x_n|x_{n-1})$ descreve essa dependência. Podemos fazer uso deste facto e calcular a distribuição incrementalmente ao longo do tempo a partir da calculada anteriormente.

Evolução da cadeia de Markov

A cadeia de Markov tem a propriedade de que o estado X_n depende apenas de X_{n-1}. A distribuição condicionada $p(x_n|x_{n-1})$ descreve essa dependência. Podemos fazer uso deste facto e calcular a distribuição incrementalmente ao longo do tempo a partir da calculada anteriormente.

Dada a distribuição $p(x_{n-1})$, a distribuição no instante seguinte é

$$p(x_n) = \sum_{x_{n-1} \in \mathcal{X}} p(x_n, x_{n-1}) = \sum_{x_{n-1} \in \mathcal{X}} p(x_n|x_{n-1})p(x_{n-1}).$$

A cadeia de Markov tem a propriedade de que o estado X_n depende apenas de X_{n-1}. A distribuição condicionada $p(x_n|x_{n-1})$ descreve essa dependência. Podemos fazer uso deste facto e calcular a distribuição incrementalmente ao longo do tempo a partir da calculada anteriormente.

Dada a distribuição $p(x_{n-1})$, a distribuição no instante seguinte é

$$p(x_n) = \sum_{x_{n-1} \in \mathcal{X}} p(x_n, x_{n-1}) = \sum_{x_{n-1} \in \mathcal{X}} p(x_n|x_{n-1}) p(x_{n-1}).$$

Escrevendo as distribuições de probabilidade como vectores coluna, a fórmula anterior é equivalente a

$$\underbrace{\begin{bmatrix} \mu_n^A \\ \mu_n^B \\ \mu_n^C \end{bmatrix}}_{p(x_n)} = \underbrace{\begin{bmatrix} 0.7 & 0.8 & 0.5 \\ 0.3 & 0 & 0 \\ 0 & 0.2 & 0.5 \end{bmatrix}}_{p(x_n|x_{n-1})} \underbrace{\begin{bmatrix} \mu_{n-1}^A \\ \mu_{n-1}^B \\ \mu_{n-1}^C \end{bmatrix}}_{p(x_{n-1})}$$

ou de modo mais compacto

$$\mu_n = P \mu_{n-1}.$$

Evolução da cadeia de Markov

Partindo de uma distribuição inicial μ_0 podemos obter todas as seguintes fazendo

$$\begin{aligned}
\mu_1 &= P\mu_0 \\
\mu_2 &= P\mu_1 = P(P\mu_0) = P^2\mu_0 \\
\mu_3 &= P\mu_2 = P(P(P\mu_0)) = P^3\mu_0 \\
&\vdots \\
\mu_n &= P^n\mu_0.
\end{aligned}$$

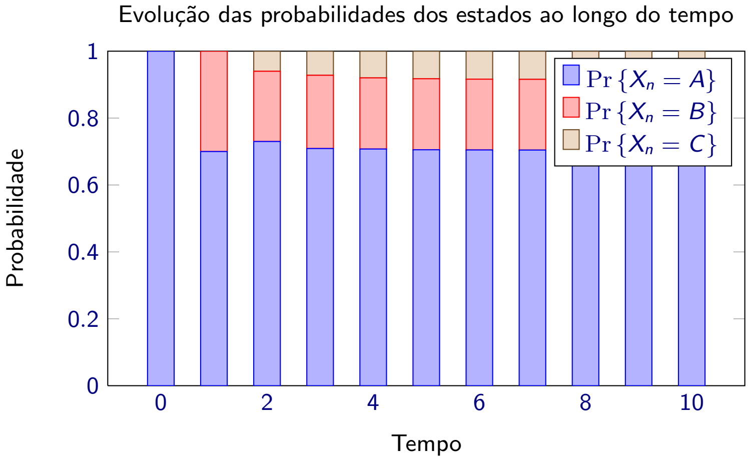
No caso do exemplo, com estado inicial $X_0 = A$ temos a distribuição

$$\mu_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

Multiplicando três vezes por P obtém-se

$$\mu_3 = P^3\mu_0 = \begin{bmatrix} 0.709 \\ 0.219 \\ 0.072 \end{bmatrix}.$$

O slide seguinte mostra a evolução das probabilidades ao longo do tempo.



Em muitas situações (mas não em todas) as probabilitades dos estados tendem para valores estacionários. *I.e.*, ao fim de algum tempo, a probabilidade da cadeia de Markov se encontrar num dado estado não depende do estado inicial.

Problema

Um jogador tem duas moedas: a moeda A, perfeitamente equilibrada; e a moeda B, desequilibrada, em que as probabilidades de "cara" e "coroa" são respectivamente 0.7 e 0.3. O jogador efectua uma sequência de lançamentos de uma das moedas, trocado de moeda sempre que sai "caras", e anunciando em voz alta qual a moeda usada.

1. *Represente a cadeia de Markov na forma de grafo, indicando os estados e as probabilidades de transição.*
2. *A cadeia de Markov é homogénea no tempo?*
3. *Supondo que a primeira moeda usada é a A, qual a probabilidade de a moeda usada no 3º lançamento ser usada a moeda B.*

Definição (Cadeia de Markov irredutível)

Uma cadeia de Markov é irredutível se é possível transitar entre qualquer par de estados num número finito de passos com probabilidade não nula.

Definição (Cadeia de Markov irredutível)

Uma cadeia de Markov é irredutível se é possível transitar entre qualquer par de estados num número finito de passos com probabilidade não nula.

Exemplo

Não é irredutível pois não é possível transitar dos estados C e D para os estados A e B. Diz-se que os estados $\{A, B\}$ não são acessíveis de $\{C, D\}$.

Definição (Estados períodicos e cadeia de Markov aperiódica)

Um estado é periódico com período k se só pode ser visitado em instantes de tempo múltiplos de k. Formalmente, o período é

$$k = \gcd\{n : \Pr\{X_n = i | X_0 = i\} > 0\}$$

Uma cadeia de Markov é aperiódica se todos os estados são aperiódicos (*i.e.*, não há estados periódicos).

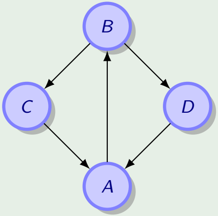
Definição (Estados períodicos e cadeia de Markov aperiódica)

Um estado é periódico com período k se só pode ser visitado em instantes de tempo múltiplos de k. Formalmente, o período é

$$k = \gcd\{n : \Pr\{X_n = i | X_0 = i\} > 0\}$$

Uma cadeia de Markov é aperiódica se todos os estados são aperiódicos (*i.e.*, não há estados periódicos).

Exemplo



A cadeia de Markov não é aperiódica. É periódica com período 3.

Definição

Chama-se distribuição estacionária da cadeia de Markov à distribuição de probabilidade dos estados tal que a distribuição no instante n é igual à distribuição no instante anterior $n-1$, *i.e.*

$$p(x_n) = p(x_{n-1})$$

Definição

Chama-se distribuição estacionária da cadeia de Markov à distribuição de probabilidade dos estados tal que a distribuição no instante n é igual à distribuição no instante anterior $n-1$, *i.e.*

$$p(x_n) = p(x_{n-1})$$

Teorema

Se uma cadeia de Markov é simultaneamente irredutível e aperiódica, então tem uma única distribuição estacionária. Nesse caso, partindo de qualquer distribuição inicial, a variável X_n tende para a distribuição estacionária μ, i.e.

$$p(x_n) \to \mu \text{ quando } n \to \infty$$

Como se viu, a distribuição estacionária é tal que $p(x_n) = p(x_{n-1})$.
Já sabemos que, dada a distribuição $p(x_{n-1})$, a distribuição no instante seguinte é

$$p(x_n) = \sum_{x_{n-1} \in \mathcal{X}} p(x_n | x_{n-1}) p(x_{n-1}),$$

[1] A matriz de transição também é conhecida como matriz estocástica ou matriz de Markov.

Como se viu, a distribuição estacionária é tal que $p(x_n) = p(x_{n-1})$.
Já sabemos que, dada a distribuição $p(x_{n-1})$, a distribuição no instante seguinte é

$$p(x_n) = \sum_{x_{n-1} \in \mathcal{X}} p(x_n | x_{n-1}) p(x_{n-1}),$$

ou de modo matricial

$$\mu_n = P\, \mu_{n-1}$$

onde P é a matriz de transição[1] e μ_n, μ_{n-1} são vectores de probabilidades.

A distribuição é estacionária se os vectores de probabilidades forem iguais em instantes de tempo consecutivos, ou seja, a distribuição estacionária satisfaz a equação matricial

$$\mu = P\mu.$$

[1] A matriz de transição também é conhecida como matriz estocástica ou matriz de Markov.

Como se viu, a distribuição estacionária é tal que $p(x_n) = p(x_{n-1})$.
Já sabemos que, dada a distribuição $p(x_{n-1})$, a distribuição no instante seguinte é

$$p(x_n) = \sum_{x_{n-1} \in \mathcal{X}} p(x_n | x_{n-1}) p(x_{n-1}),$$

ou de modo matricial

$$\mu_n = P \ \mu_{n-1}$$

onde P é a matriz de transição[1] e μ_n, μ_{n-1} são vectores de probabilidades.

A distribuição é estacionária se os vectores de probabilidades forem iguais em instantes de tempo consecutivos, ou seja, a distribuição estacionária satisfaz a equação matricial

$$\mu = P\mu.$$

Será que conseguimos encontrar uma solução desta equação?

[1] A matriz de transição também é conhecida como matriz estocástica ou matriz de Markov.

Queremos encontrar a distribuição estacionária que é solução μ da equação

$$P\,\mu = \mu$$

- Esta equação é semelhante a um problema de álgebra linear que consiste em calcular os valores próprios λ e vectores próprios v de uma matriz A. Nesse contexto resolve-se a equação $Av = \lambda v$ em ordem aos escalares λ e depois aos vectores v correspondentes.

Queremos encontrar a distribuição estacionária que é solução μ da equação

$$P\,\mu = \mu$$

- Esta equação é semelhante a um problema de álgebra linear que consiste em calcular os valores próprios λ e vectores próprios v de uma matriz A. Nesse contexto resolve-se a equação $Av = \lambda v$ em ordem aos escalares λ e depois aos vectores v correspondentes.
- No caso aqui em estudo, $P\mu = \mu$, o valor próprio é 1, e a distribuição estacionária μ é um vector próprio associado ao valor próprio 1.

Queremos encontrar a distribuição estacionária que é solução μ da equação

$$P\,\mu = \mu$$

- Esta equação é semelhante a um problema de álgebra linear que consiste em calcular os valores próprios λ e vectores próprios v de uma matriz A. Nesse contexto resolve-se a equação $Av = \lambda v$ em ordem aos escalares λ e depois aos vectores v correspondentes.
- No caso aqui em estudo, $P\mu = \mu$, o valor próprio é 1, e a distribuição estacionária μ é um vector próprio associado ao valor próprio 1.
- Acontece que um vector próprio não é único: qualquer vector com a mesma direcção é igualmente um vector próprio associado ao mesmo valor próprio (os vectores próprios determinam uma direcção e portanto o seu comprimento, a norma, pode ser arbitrária).

Distribuição estacionária de uma cadeia de Markov

Queremos encontrar a distribuição estacionária que é solução μ da equação

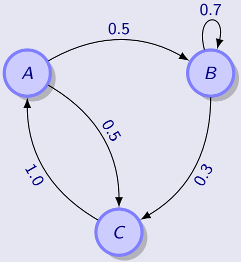
$$P\,\mu = \mu$$

- Esta equação é semelhante a um problema de álgebra linear que consiste em calcular os valores próprios λ e vectores próprios v de uma matriz A. Nesse contexto resolve-se a equação $Av = \lambda v$ em ordem aos escalares λ e depois aos vectores v correspondentes.
- No caso aqui em estudo, $P\mu = \mu$, o valor próprio é 1, e a distribuição estacionária μ é um vector próprio associado ao valor próprio 1.
- Acontece que um vector próprio não é único: qualquer vector com a mesma direcção é igualmente um vector próprio associado ao mesmo valor próprio (os vectores próprios determinam uma direcção e portanto o seu comprimento, a norma, pode ser arbitrária).
- Para que o vector próprio μ corresponda a uma distribuição de probabilidade é necessário que as suas componentes, as probabilidades, somem 1. Assim, a solução é obtida normalizando μ.

Problema

Considere uma cadeia de Markov com três estados $\{A, B, C\}$ *e probabilidades de transição dadas na tabela seguinte:*

$p(x_n \mid x_{n-1})$	$X_{n-1} = A$	$X_{n-1} = B$	$X_{n-1} = C$
$X_n = A$	0	0	1
$X_n = B$	0.5	0.7	0
$X_n = C$	0.5	0.3	0



1. *Esta cadeia de Markov é irredutível e aperiódica?*
2. *Qual a distribuição de probabilidade estacionária?*

Resolução usando valores/vectores próprios em `octave`

```
>> P = [0, 0, 1; 0.5, 0.7, 0; 0.5, 0.3, 0];
>> [vec, val] = eig(P)
vec =
   0.80855  -0.45750   0.59231
  -0.30588  -0.76249  -0.78285
  -0.50268  -0.47750   0.19055

val =
  -0.62170         0         0
         0   1.00000         0
         0         0   0.32170
```

Como o valor próprio 1 é o segundo da diagonal, então o vector próprio correspondente é a segunda coluna de `vec`. Normalizando obtém-se a distribuição estacionária:

```
>> mu = vec(:,2) / sum(vec(:,2))
mu =
   0.27273
   0.45455
   0.27273
```

Resolução usando valores/vectores próprios em `python/numpy/scipy`

```
>> P = array([[0, 0, 1], [0.5, 0.7, 0], [0.5, 0.3, 0]])
>> (val, vec) = eig(P)
>> val
array([-0.62169906,  1.        ,  0.32169906])
```

Como o vector próprio 1 é o segundo, então o vector próprio correspondente é a segunda coluna de `vec`. Normalizando obtém-se a distribuição estacionária:

```
>> vec[:,1] / sum(vec[:,1])
array([0.27272727,  0.45454545,  0.27272727])
```

Resolução usando sistemas de equações lineares em `octave`

```
>> P = [0, 0, 1; 0.5, 0.7, 0; 0.5, 0.3, 0];
>> [P-eye(3); ones(1,3)] \ [zeros(3,1); 1]
ans =
   0.27273
   0.45455
   0.27273
```

A função `eye(3)` gera uma matriz identidade 3×3. As funções `zeros` e `ones` geram matrizes de zeros e uns, respectivamente, com as dimensões indicadas.

Este método consiste essencialmente em resolver o seguinte problema:

$$P\,\mu = \mu, \qquad \sum_{i=1}^{N} \mu_i = 1.$$

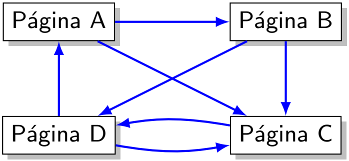
A primeira equação é equivalente a $(P - I)\mu = 0$. Então podemos escrever tudo numa única equação linear com N incógnitas e $N + 1$ equações:

$$\begin{bmatrix} & P - I & \\ 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_N \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

Exemplo: PageRank (google)

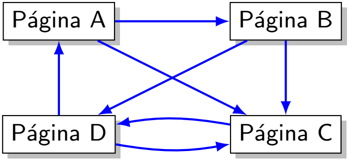
O motor de pesquisa do google apresenta os resultados ordenados de acordo com o *ranking* de cada página – o PageRank.

Consideram-se as páginas na internet como nós num grafo direccionado e os hyperlinks como arestas. Por exemplo:



Exemplo: PageRank (google)

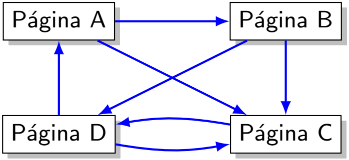
O motor de pesquisa do google apresenta os resultados ordenados de acordo com o *ranking* de cada página – o PageRank.
Consideram-se as páginas na internet como nós num grafo direccionado e os hyperlinks como arestas. Por exemplo:



Assume-se que um internauta:

- Clica aleatoriamente nos links com igual probabilidade.
- Não se cansa e continua eternamente a ler as páginas e a clicar nos links.

O motor de pesquisa do google apresenta os resultados ordenados de acordo com o *ranking* de cada página – o PageRank.
Consideram-se as páginas na internet como nós num grafo direccionado e os hyperlinks como arestas. Por exemplo:



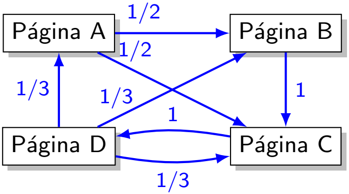
Assume-se que um internauta:

- Clica aleatoriamente nos links com igual probabilidade.
- Não se cansa e continua eternamente a ler as páginas e a clicar nos links.

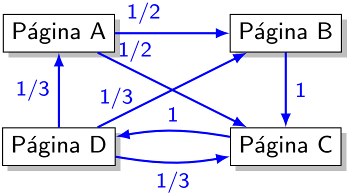
A importância de uma página (*PageRank*) é a frequência com que ela é visitada.

Como calcular o PageRank?

Se a probabilidade de clicar nos links é uniforme, então temos uma cadeia de Markov com probabilidades de transição:



Se a probabilidade de clicar nos links é uniforme, então temos uma cadeia de Markov com probabilidades de transição:



Ao longo de um período grande de tempo, a frequência com que as páginas são visitadas é dada pela distribuição estacionária da cadeia de Markov.
Assim, o PageRank é a solução de

$$\begin{bmatrix} \mu_A \\ \mu_B \\ \mu_C \\ \mu_D \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1/3 \\ 1/2 & 0 & 0 & 1/3 \\ 1/2 & 1 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_A \\ \mu_B \\ \mu_C \\ \mu_D \end{bmatrix}$$

A solução é aproximadamente $(0.12, 0.18, 0.35, 0.35)$. As páginas C e D têm o PageRank mais elevado e aparecem primeiro. A página A tem o PageRank mais baixo e aparece no final da lista.

Ritmo de entropia de uma cadeia de Markov

O ritmo de entropia de uma sequência de variáveis aleatórias descreve quanto a entropia conjunta cresce ao longo do tempo a cada nova ocorrência.

Definição (Ritmo de entropia)

O ritmo de entropia é definido por

$$H'(\mathcal{X}) = \lim_{n\to\infty} \frac{H(X_1, X_2, \dots, X_n)}{n}.$$

- Para uma cadeia de Markov, o ritmo de entropia é igual à entropia condicional

  $$H(\mathcal{X}) = H(X_n | X_{n-1}).$$

  onde a entropia condicional é calculada usando a distribuição estacionária.
- Para uma sequência de variáveis aleatórias independentes, o ritmo de entropia é igual à entropia

  $$H(\mathcal{X}) = H(X).$$