MPEI 2018-2019

20 – Cadeias de Markov

?!

- O que esteve na origem do grande sucesso inicial da Google ?

- O que tem em comum esse sucesso com a capacidade de interagir por voz com computadores, robôs e smartphones ?
  - No reconhecimento de fala ?
  - Na síntese de fala ?

Exemplo 1

- Suponhamos que em cada dia que têm aulas de MPEI acordam e decidem se vêm ou não à aula.
- Se vieram à aula anterior, a probabilidade de virem é 70%; se faltaram à anterior, essa probabilidade é 80%.

- Algumas questões:
  – Se vieram à aula esta segunda, qual a probabilidade de virem na aula de SEGUNDA da próxima semana ?
  – Assumindo que o semestre tem duração infinita (que horror!), qual a percentagem aproximada de aulas a que estariam presentes ?

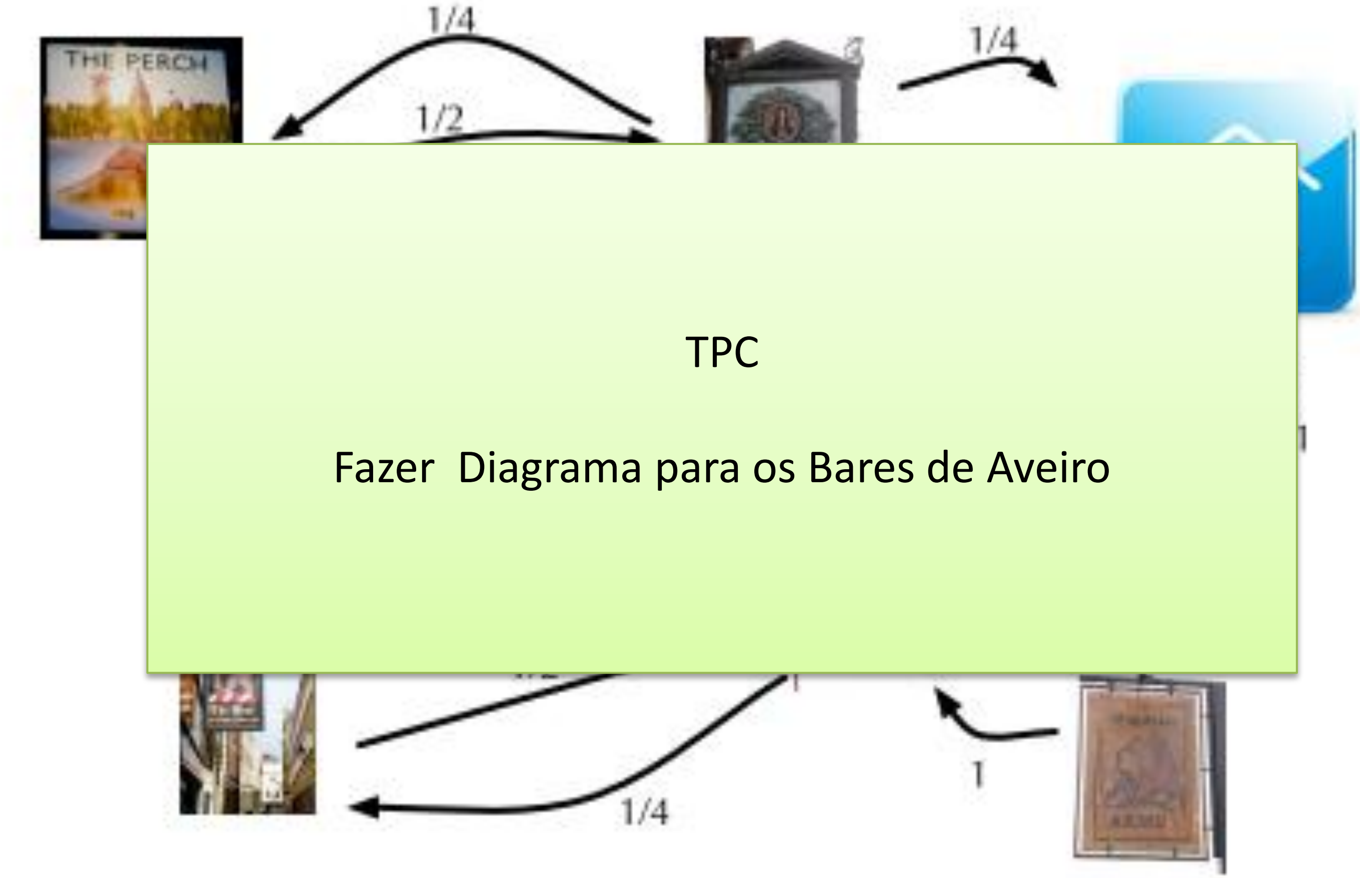
Exemplo 2

- Dividir a turma em 3 grupos A, B e C no início do semestre
- No final de cada aula:
- 1/3 do grupo A vai para o B e outro 1/3 do grupo A vai para o grupo C
- ¼ do grupo B vai para A e ¼ de B vai para C
- ½ do grupo C vai para o grupo B

- Como ficarão os grupos ao fim de n aulas ?

Exemplo 3 – “Pub Crawl”

- Bares junto a uma conhecida Universidade:



TPC

Fazer Diagrama para os Bares de Aveiro

Outro exemplo

- Passeio aleatório (random walk)

Lançar moeda

Cara Coroa Coroa … Cara …

Muitas áreas de aplicação

- Muitas vezes estamos interessados na transição de algo entre certos estados.
- Exemplos:
  - Movimento de pessoas entre regiões
  - Estado do tempo
  - Movimento entre as posições num jogo de Monopólio
  - Pontuação ao longo de um jogo
  - Estado de Filas de atendimento

Princípios básicos

Processos estocásticos

- Lidam com a dinâmica da teoria de probabilidades
- O conceito de processo estocástico estende o conceito de variável aleatória
- Uma v.a. X mapeia um acontecimento $s \in \Omega$ num número $X(s)$
- O processo mapeia o evento para números diferentes em tempos diferentes
  - O que implica que em lugar de termos um número $X(s)$ temos $X(t, s)$
  - Sendo t$\in T$ geralmente um conjunto de tempos

Processos estocásticos

- Se fixarmos $s, X(t)$ é uma função real do tempo
- $X(t, s)$ pode então ser vista como uma colecção de funções no tempo

- Se fixarmos t temos uma função $X(s)$ que depende apenas de s, ou seja uma variável aleatória

- Um nome alternativo é processos aleatórios

Classificação de processos estocásticos

- Podem ser classificados segundo o parâmetro t e os valores que $X(t, s)$ pode assumir (estados do processo)
- Quanto a t:
  - Tempo contínuo: Se T é um intervalo contínuo
  - Tempo discreto: Se T é um conjunto contável
    - Também chamada sequência aleatória e representada por $X[n]$
- Quanto ao conjunto de estados (E):
  - Contínuo
  - Discreto

Definição

- Um processo de Markov é um processo estocástico em que a probabilidade de o sistema estar num estado específico num determinado período de observação depende apenas do seu estado no período de observação imediatamente precedente
  - O futuro apenas depende do presente e não do passado

Tipos de processos de Markov

- Discretas/contínuas

		Espaço de estados	
		Discreto	Contínuo
Tempo	Discreto	**Cadeia de Markov tempo discreto**	Processo de Markov em tempo discreto
	Contínuo	Cadeia de Markov tempo contínuo	Processo de Markov em tempo contínuo

- Focaremos a nossa atenção em cadeias de Markov de tempo discreto

Cadeias de Markov discretas

- X_n: estado após n transições
  - Pertence a um conjunto finito,
    - Em geral $\{1, 2, \dots, m\}$

  - X_0 é dado ou aleatório

Questões comuns relativas a cadeias de Markov

- Qual a probabilidade de transição entre dois estados em *n* observações ?

- Existe algum equilíbrio ?

- Existe uma estabilidade a longo prazo ?

Propriedade/Suposição de Markov

- Probabilidade de transição do estado i para o estado j:

- $\boldsymbol{p_{ji}} = P(X_{n+1} = j | X_n = i)$

 $= P(X_{n+1} = j | X_n = i, X_{n-1}, \ldots, X_0)$

- Quando estas probabilidades $\boldsymbol{p_{ji}}$ não dependem de n a cadeia diz-se **homogénea**
  - Focaremos a nossa atenção neste tipo de cadeias de Markov

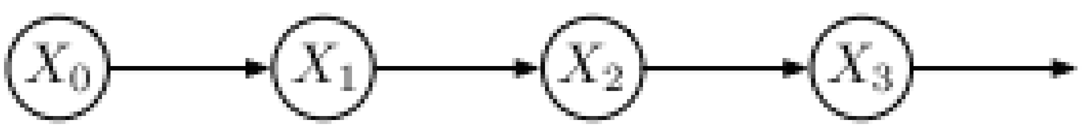
Propriedade/Suposição de Markov

- $P(X_0 = x_0, X_1 = x_1, X_2 = x_2 \dots) = ?$

$$= P(X_0 = x_0)\, P(X_1 = x_1 | X_0 = x_0)\, P(X_2 = x_2 | X_0 = x_0, X_1 = x_1) \dots$$

$$= P(X_0 = x_0)\, P(X_1 = x_1 | X_0 = x_0)\, P(\boldsymbol{X_2 = x_2 | X_1 = x_1}) \dots$$

- O futuro é independente do passado, dado o presente
- O processo “não tem memória”



Especificação de uma cadeia

- Identificar os estados possíveis
- Identificar as transições possíveis
- Identificar as probabilidades de transição

Aplicando ao exemplo 1 – faltar ou não faltar à aula

- Estados ?

- Transições ?

- Probabilidades de transição ?

Aplicando ao exemplo 1 – faltar ou não faltar à aula

- Estados ?
  - 2: {faltar, não faltar}

- Transições ?
  - Faltar-> não faltar
  - Não faltar -> faltar
  - Faltar -> faltar
  - Não faltar -> não faltar

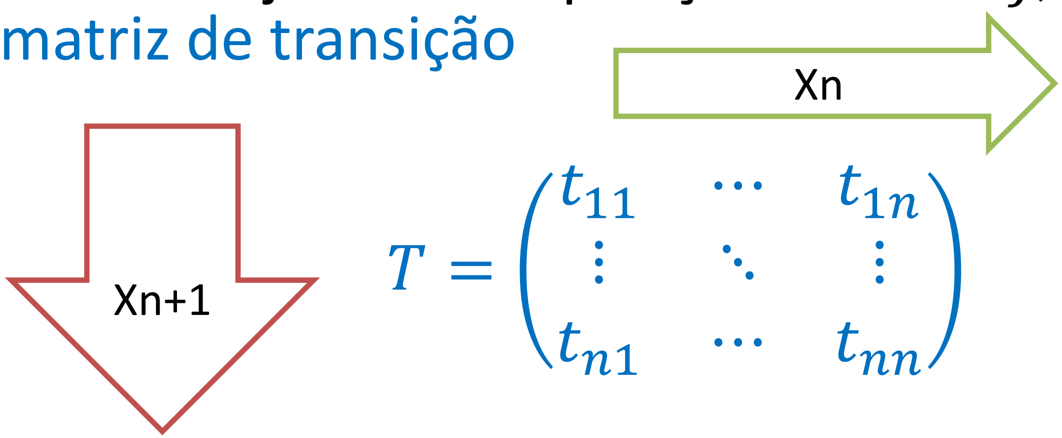
- Probabilidades de transição ?

  - Faltar-> não faltar : 0,8
  - Não faltar -> faltar : 0,3

  - Faltar -> faltar : 0,2
  - Não faltar -> não faltar: 0,7

Matriz de transição

- É usual representar as probabilidades de transição através de uma matriz, chamada de matriz de transição
- Tendo o sistema n estados possíveis, para cada par i, j fazemos t_{ji} igual à probabilidade de mudar **do estado *i* para o estado *j***.
- A matriz T cujo valor na posição $\mathrm{linha} = j, coluna = i$ é t_{ji} é a matriz de transição



$$T = \begin{pmatrix} t_{11} & \cdots & t_{1n} \\ \vdots & \ddots & \vdots \\ t_{n1} & \cdots & t_{nn} \end{pmatrix}$$

- Nota: Alguns autores adoptam t_{ij} como a probabilidade de mudar do estado i para o estado j

Matriz T do Exemplo 1

- $T = \begin{pmatrix} ? & ? \\ ? & ? \end{pmatrix}$

- Considerando estado 1 “não faltar”, temos

- $T = \begin{array}{l} não\ faltar \rightarrow \\ faltar \rightarrow \end{array} \begin{pmatrix} 0,7 & 0,8 \\ 0,3 & 0,2 \end{pmatrix}$

Matriz T do Exemplo 2

$$T = \begin{array}{c|ccc} & A & B & C \\ \hline A & ? & ? & ? \\ B & ? & ? & ? \\ C & ? & ? & ? \end{array}$$

Matriz T do Exemplo 2

$$T = \begin{array}{c|ccc} & A & B & C \\ \hline A & 1/3 & & \\ B & 1/3 & & \\ C & 1/3 & & \end{array}$$

Futuro Estado

Matriz T do Exemplo 2

$$T = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{ccc} A & B & C \\ \left|\begin{array}{ccc} 1/3 & 1/4 & 0 \\ 1/3 & 1/2 & 1/2 \\ 1/3 & 1/4 & 1/2 \end{array}\right| \end{array}$$

Futuro Estado

Matriz T é estocástica

- A matriz de transição reflecte propriedades importantes das probabilidades:

  - Todas as entradas são não-negativas
  - Os valores em cada COLUNA somados dão sempre resultado 1

- Devido a estas propriedades a matriz é denominada de **matriz estocástica**

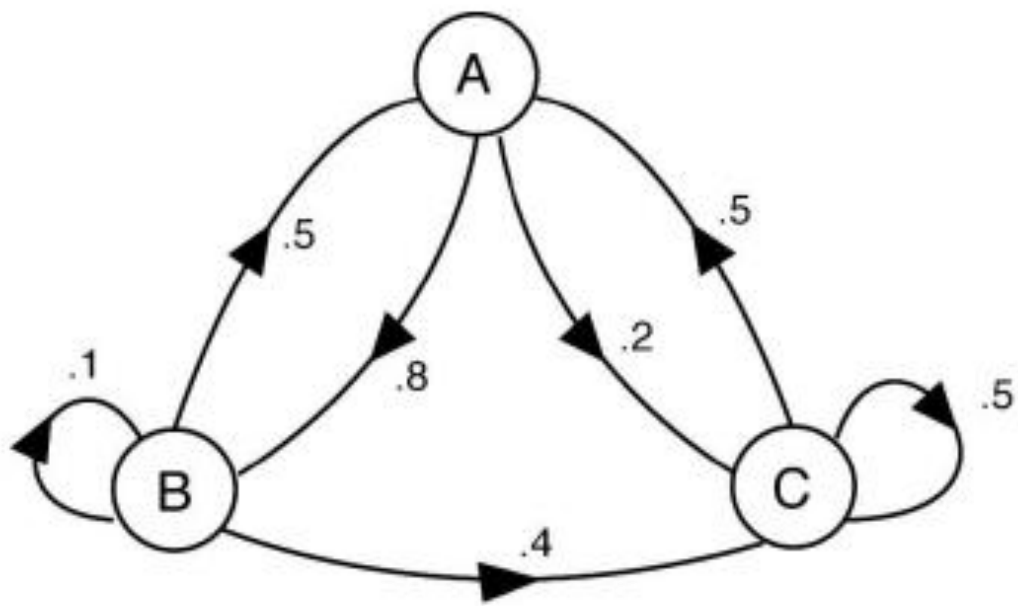
Representação gráfica da cadeia

- Apropriada e possível para número de estados pequeno

- Nós: representam todos os estados
- Setas: para todas as transições permitidas (one-step)
  - Ou seja, seta entre i e j apenas de $p_{ji} > 0$

Representação gráfica da cadeia

- Exemplo:



Simulação / Visualização dinâmica

- Estão disponíveis online formas de visualizar as transições entre estados ao longo do tempo ...
- Um desses exemplos é **Markov Chains - A visual explanation by Victor Powell**
  - http://setosa.io/blog/2014/07/26/markov-chains/index.html

  Que inclui:
  - http://setosa.io/markov/index.html#%7B%22tm%22%3A%5B%5B0.5%2C0.5%5D%2C%5B0.5%2C0.5%5D%5D%7D

- Para usar precisamos apenas de introduzir a matriz T
  - Que define o número de estados, quais as transições possíveis e as probabilidades associadas a essas transições

Simulando os nossos exemplos

- Exemplo 1:
  - Matriz:
    [[0.7, 0.3],
      [0.8, 0.2]]
  - http://setosa.io/markov/index.html#%7B%22tm%22%3A%5B%5B0.7%2C0.3%5D%2C%5B0.8%2C0.2%5D%5D%7D

- Exemplo 2:
  - Matriz:
  [[0.33,0.33,0.34],
    [0.25,0.5,0.25],
    [0,0.5,0.5]]

- Outro exemplo
  [[0,1,0,0],
    [0,0,1,0],
    [0,0,0,1],
    [0.2,0.3,0.3,0.2]]

  - O que vamos ver ?

  - Acesso directo:
  http://setosa.io/markov/index.html#%7B%22tm%22%3A%5B%5B0%2C1%2C0%2C0%5D%2C%5B0%2C0%2C1%2C0%5D%2C%5B0%2C0%2C0%2C1%5D%2C%5B0.2%2C0.3%2C0.3%2C0.2%5D%5D%7D

Estado da cadeia num determinado instante

- O **estado** de uma cadeia de Markov com n estados no tempo (time step) k é dado pelo vector estado

$$\mathbf{x}^{(k)} = \begin{pmatrix} \boldsymbol{p_1}^{(\boldsymbol{k})} \\ \boldsymbol{p_2}^{(\boldsymbol{k})} \\ \\ \boldsymbol{p_n}^{(\boldsymbol{k})} \end{pmatrix}$$

- Onde ${p_j}^{(k)}$ é a probabilidade de o sistema estar no estado j no tempo k

Vector estado/probabilidade

- Considerando o exemplo 1:
- Suponhamos que após 10 aulas a probabilidade de faltar e não faltar são iguais
- Então o vector representativo do estado (state vector) seria:

$$\mathbf{x}^{(10)} = \begin{pmatrix} 0{,}5 \\ 0{,}5 \end{pmatrix}$$

- Este vector também se designa por **vector de probabilidade**
  - Todos elementos não-negativos
  - Soma dos elementos igual a um

Exemplo 2

- Supondo que começávamos com 20 estudantes no grupo A e 10 estudantes nos outros dois grupos, o vector relativo ao estado inicial seria

- $\mathbf{x}^{(0)} = \begin{pmatrix} 0{,}5 \\ 0{,}25 \\ 0{,}25 \end{pmatrix}$

Vector estado após uma transição

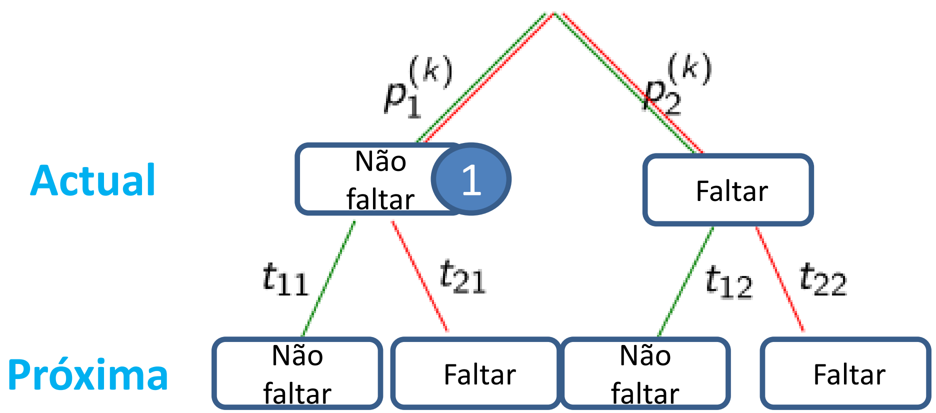
- Como obter $\mathbf{x}^{(k+1)}$?

- O vector de estado $\mathbf{x}^{(k+1)}$ no período de observação $k+1$ pode ser determinado a partir do vector $\mathbf{x}^{(k)}$ através de:
- $\mathbf{x}^{(k+1)} = T\mathbf{x}^{(k)}$

- Que resulta da probabilidade condicional:
- $P(\,estado\ j\ em\ t = k+1)$
- $= \sum_{\mathrm{i}=1}^{n} P(transição\ do\ estado\ i\ para\ o\ j)P(estado\ \mathrm{i}\ em\ t = k)$

Exemplo de aplicação – Exemplo 1

- De que forma depende a probabilidade de ir à aula seguinte da probabilidade de estar na aula actual ?



$$p_1^{(k+1)} = t_{11} p_1^{(k)} + t_{12} p_2^{(k)}$$
$$p_2^{(k+1)} = t_{21} p_1^{(k)} + t_{22} p_2^{(k)}$$

Estado após múltiplas transições

- Ataquemos agora problemas do “tipo”:
  - Qual a probabilidade de transição entre dois estados em n observações/transições ?

- Exemplo 1:
  - Qual a probabilidade dos que estiveram na aula de uma segunda virem à aula na segunda seguinte
    - Assumindo as probabilidades do nosso exemplo !
    - Tendo em conta que temos aulas segunda e quinta (TP2) ou segunda e terça (TP1).

Equações de Chapman-Kolmogorov

- Definindo a transição em n passos ${p_{ji}}^{n}$ como a probabilidade de **um processo no estado $\boldsymbol{i}$ se encontrar no estado *j* após *n* transições** adicionais. Ou seja:
- ${p_{ji}}^{n} = P(X_{n+k} = \boldsymbol{j} | X_k = \boldsymbol{i}),\ n \geq 0, i, j \geq 0$
- Obviamente ${p_{ji}}^{1} = p_{ji}$

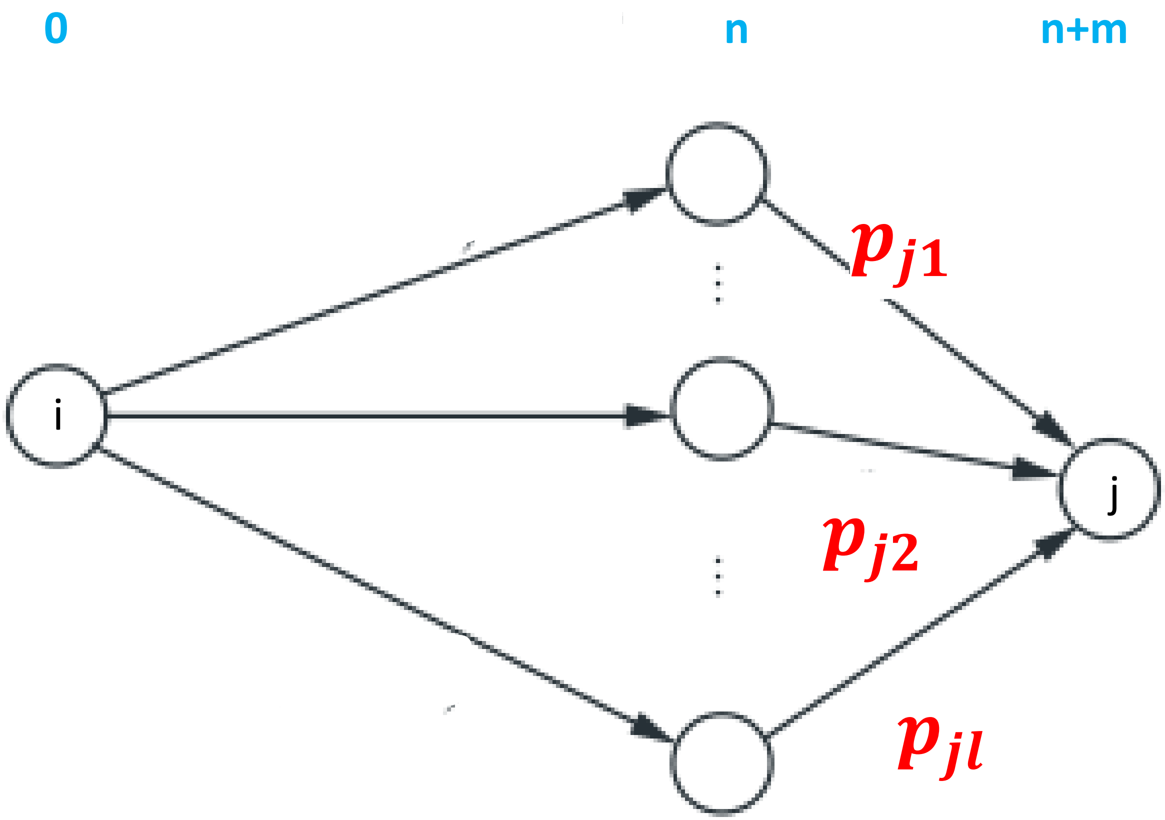
- As **equações de Chapman-Kolmogorov** permitem calcular estas probabilidades

$${p_{ji}}^{n+m} = \sum_k {p_{ki}}^{n}\ {p_{jk}}^{m} \quad \forall n, m \geq 0, \forall i, j$$

Interpretação

- É fácil de compreender se tivermos em conta que ${p_{ki}}^n {p_{jk}}^m$ representa a probabilidade de:
  - Começando em i o processo ir para o estado j em $n + m$ transições..
  - Através de um caminho que o leva ao estado k na transição n
- Logo, somando para todos os estados intermédios k obtém-se a probabilidade de estar no estado j ao fim de $n + m$ transições

Interpretação



"Demonstração" Eqs. Chapman-Kolmogorov

- ${p_{ji}}^{n+m} = P(X_{n+m} = j|X_0 = i)$

- $= \sum_k P(X_{n+m} = j, X_n = k|X_0 = i)$

- $= \sum_k P(X_{n+m} = j|X_n = k)\; P(X_n = k|X_0 = i)$

- $\sum_k \quad {p_{jk}}^m {p_{ki}}^n$

Em termos de matrizes

- Se usarmos $\mathbf{T}^{(\mathbf{n})}$ para representar a matriz com as probabilidades de n transições, a equação anterior transforma-se em:

$$\mathbf{T}^{(n+m)} = \mathbf{T}^{(n)} \,.\, \mathbf{T}^{(m)}$$

Em que o "." significa multiplicação de matrizes

- Desta equação obtém-se facilmente:

$$\mathbf{T}^{(2)} = \mathbf{T}^{(1+1)} = \mathbf{T}\,.\,\mathbf{T} = \mathbf{T}^{2}$$

- E por indução $\mathbf{T}^{(n)} = \mathbf{T}^{(n-1+1)} = \mathbf{T}^{n-1}\,.\,\mathbf{T} = \mathbf{T}^{n}$
  - Ou seja, a matriz de transição relativa a n transições pode ser obtida multiplicando $\mathbf{T}$ por si própria n vezes

Aplicação ao Exemplo 1

- Voltando a uma questão colocada no início da aula …

- *Se vieram à aula esta segunda, qual a probabilidade de virem na aula de SEGUNDA da próxima semana ?*

- Solução:
- Temos $\mathbf{x}^{(0)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, significando “não faltar”
- Pretendemos $\mathbf{x}^{(2)}$, 0= hoje

...

- $\mathbf{x}^{(2)} = \mathrm{T}\mathbf{x}^{(1)} = T\left(T\mathbf{x}^{(0)}\right) = T^2\mathbf{x}^{(0)}$

$$= \begin{pmatrix} 0.7 & 0.8 \\ 0.3 & 0.2 \end{pmatrix}^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.7 & 0.8 \\ 0.3 & 0.2 \end{pmatrix} \begin{pmatrix} 0.7 \\ 0.3 \end{pmatrix} = \begin{pmatrix} 0.73 \\ 0.27 \end{pmatrix}$$

- Ou seja 73% de probabilidade de virem na próxima Segunda

Terminologia

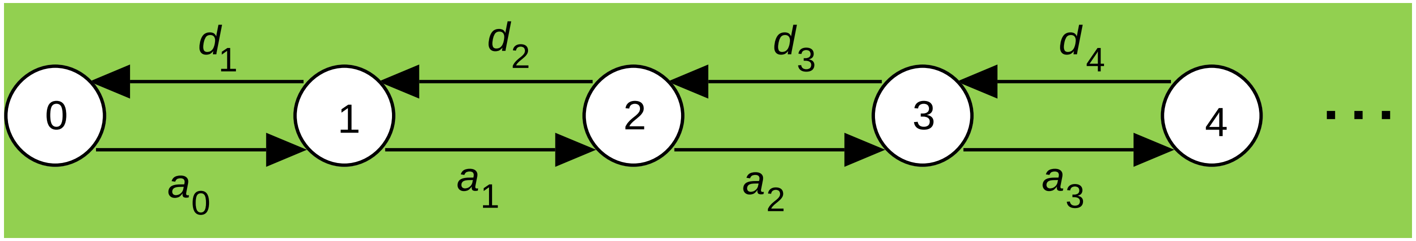
Tipos de estados

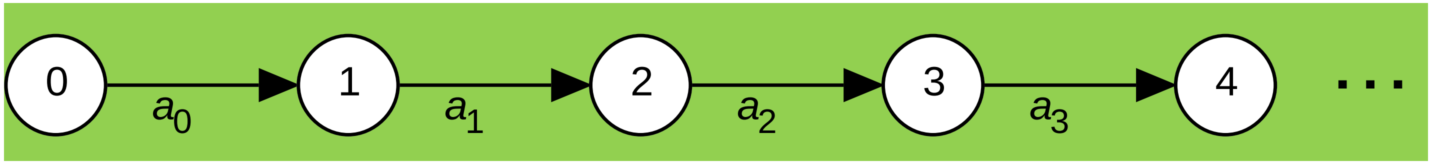
Tipos de matrizes de transição

...

Acessibilidade de um estado

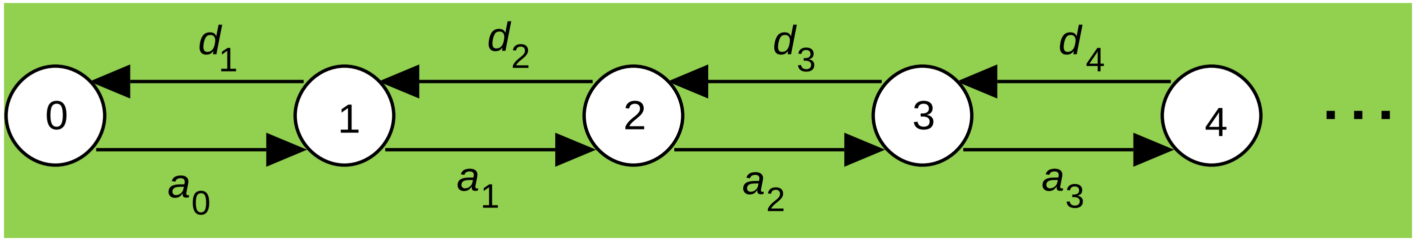
- Possibilidade de ir do estado i para o estado j (existe caminho na cadeia de i para j).





Estados comunicantes

- Dois estados comunicam se ambos são acessíveis a partir do outro.



- Um sistema é não redutível (irreducible) se todos os estados comunicam

- **Classe**: conjunto de estados que comunicam entre si

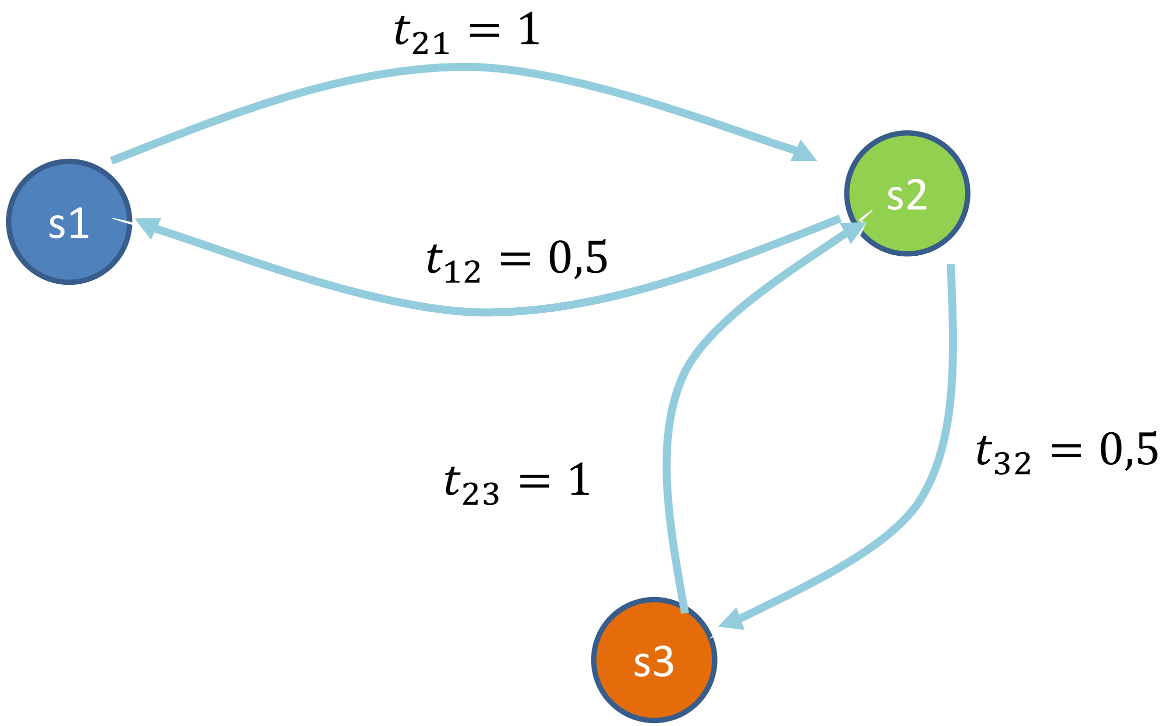
Estado recorrente

- Um estado s_i é um **estado recorrente** se o sistema poder sempre voltar a ele (depois de sair dele).

- De uma forma mais formal: s_i é um **estado recorrente** se, para todos os estados s_j, a existência de um inteiro r_j tal que $p_{ji}{}^{(r_j)} > 0$ implica que existe um inteiro r_i tal que $p_{ij}{}^{(r_i)} > 0$

- Um estado não recorrente é transiente

Estados recorrentes ?



- Os 3 estados são recorrentes

Estado transiente

- Um estado é transiente se existe um outro estado qualquer para o qual o processo de Markov pode transitar, mas do qual o processo não pode retornar

- Ou seja, se existe um estado s_j e um inteiro l tal que ${p_{ji}}^{(l)} \neq 0$ e ${p_{ij}}^{(r)} = 0$ para $r = 0,1,2, \ldots$

- A probabilidade destes estados tende para zero quando *n* tende para infinito
  - Pois apenas são visitados um número finito de vezes

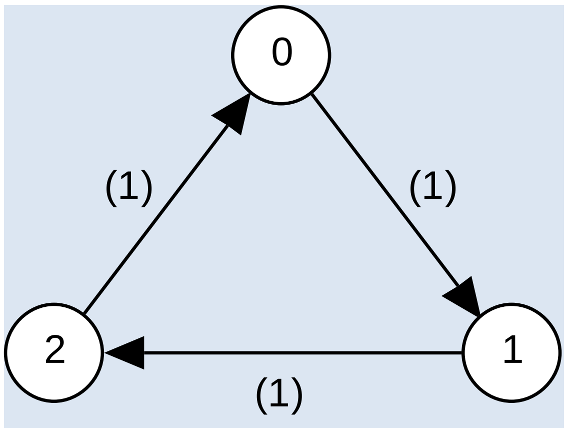
Estado periódico

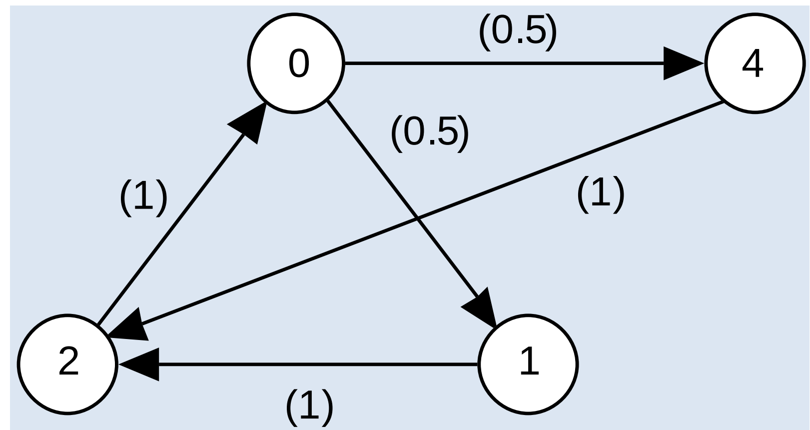
- Um estado é **periódico** se apenas se pode regressar a ele após um número fixo de transições superior a 1 (ou múltiplos desse número).

- Formalizando:
  – Um estado recorrente s_i diz-se **periódico** se existe um inteiro $c > 0$ tal que ${p_{ii}}^{(r)}$ é igual a zero para todos os valores de r excepto $r = c, 2c, 3c, \ldots$

Estado periódico

•





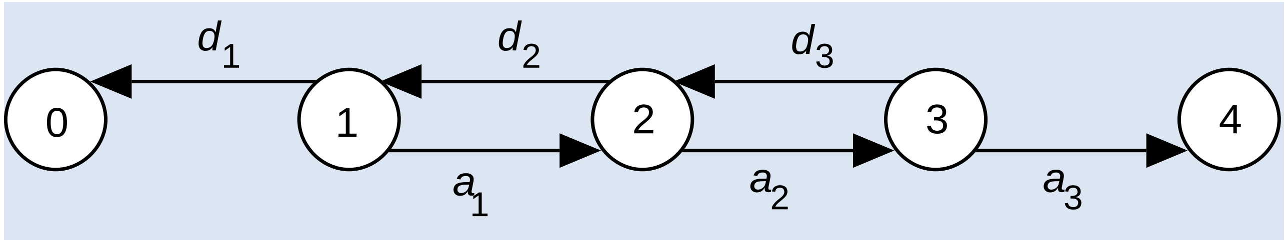
Todos estados visitados em múltiplos de 3 iterações

Todos estados visitados em múltiplos de 3 iterações

- Um estado não periódico é **aperiódico**
  - Como era de esperar!

Estado absorvente

- Um estado **absorvente** é um estado do qual não é possível sair (ou seja transitar para outro estado)
- Uma cadeia é absorvente se tiver pelo menos um estado absorvente

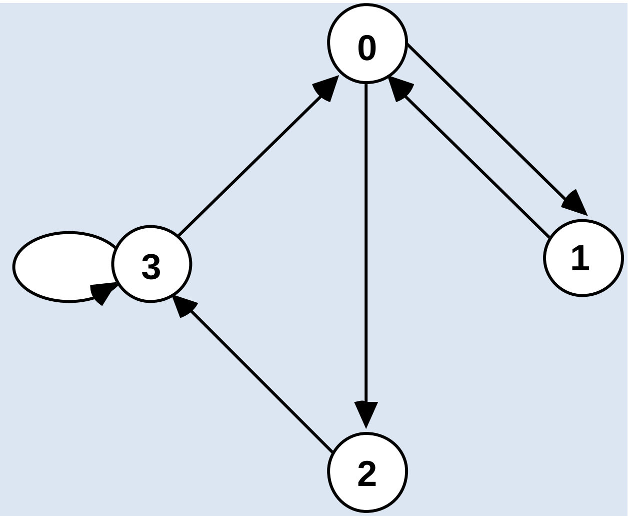


- Os estados 0 e 4 são absorventes

Aplicação dos conceitos

- Exemplo:

$$\begin{array}{c c} \text{State} & \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0 & X & X & 0 \\ X & 0 & 0 & 0 \\ 0 & 0 & 0 & X \\ X & 0 & 0 & X \end{bmatrix} \end{array}$$



- Todos os pares de estados comunicam, formando um única classe recorrente
  - Os estados são aperiódicos
- Em consequência o processo é aperiódico e irredutível