Processo de Markov: a probabilidade do sistema estar num estado específico num determinado período de observação só depende do seu estado no período de observação imediatamente precedente

Probabilidade de transição do estado i para o estado j (p_{ji})

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

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

→ Se não depender de n → cadeia homogénea

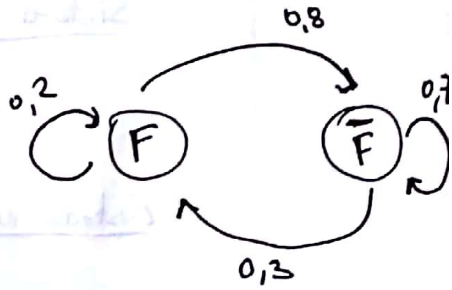
Exemplo 1:

- Se um aluno foi à aula anterior, a probabilidade de vir à aula de hoje é 70%.
- Se faltaram à anterior, a probabilidade de vir hoje é de 80%.

① Identificar estados:

2 estados → Faltar, Vir à aula

② Identificar transições:



faltar → faltar (20%) ; não faltar (80%)

não faltar → faltar (30%) ; não faltar (70%)

③ Probabilidades de transição

Matriz transição:

$$T = \underbrace{t_{ji}}_{\text{passar de } i \text{ para } j} = \begin{array}{c|cc} & \text{inicial } \bar{F} & F \\ \text{finais } \bar{F} & 0{,}7 & 0{,}8 \\ F & 0{,}3 & 0{,}2 \end{array}$$

→ • Entradas não negativas
• soma da coluna = 1

⇓

Matriz estocástica

tor Estado

$$\boxed{x^{(k)} = \begin{pmatrix} p_1^{(k)} \\ \vdots \\ p_M^{(k)} \end{pmatrix}}$$

p_j^{k} → probabilidade do sistema estar no estado j no tempo K

Ex:

Se, após 10 aulas, a probabilidade de faltar ou não faltar fosse igual, teríamos:

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

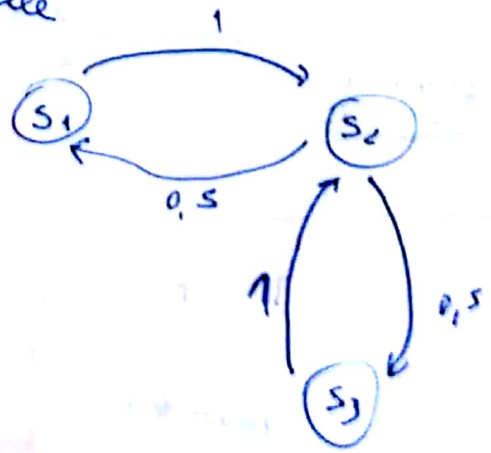
Após as transições:

$$\boxed{x^{(k+1)} = T \cdot x^{(k)}} \Rightarrow x^{(2)} = T^2 \cdot \underbrace{x^{(0)}}_{\text{estado inicial}}$$

Terminologias:

Sistema não redutível → todos os estados são comunicantes (comunicam entre si)

Estado recorrente → depois de sair dele, o sistema volta a entrar nele



Estado transiente: existe um outro estado qlq para o qual se pode transitar, mas do qual não se pode retornar.

Estado periódico: Se apenas se pode regressar a ele após n.º de transições fixo (superior a 1)

Estado absorvente: estado que não tem nenhuma saída.

Equilíbrio.

Ao fim de muitas transições as cadeias de Markov atingem um equilíbrio, sendo que a probabilidade de qualquer estado se torna independente do passo e das condições iniciais

Matriz regular: • Se alguma das suas potência tiver todos os valores não nulos

• ∃ uma probabilidade de mudar de qlq estado para qlq estado

é //

Cadeia Ergódica:

Podem-se efetuar transições de qlq estado para qlq outro estado

$$\lim_{n \to +\infty} T^n = \begin{bmatrix} u_1 & u_1 & u_1 \\ u_2 & u_2 & u_2 \\ u_3 & u_3 & u_3 \end{bmatrix}$$

vetor estado estacionário (u)

Quando $n \to +\infty \Rightarrow T^n \cdot x = u$ e $Tu = u$

Em matlab

$T = [0{,}7 \quad 0{,}8 \;;\; 0{,}3 \quad 0{,}2]$

$M = [T - eye(2) \;;\; ones(1,2)]$

$x = [0 \;;\; 0 \;;\; 1]$

$u = M/x$

Cadeia Absorvente:

1. Tem, pelo menos, 1 estado absorvente
2. Podemos ir de um estado não absorvente para um absorvente num número finito de passos.

Forma canónica:

Numa matriz de transição agrupamos todos os estados absorventes

→ No início retemos os não absorventes

→ No fim os absorventes

$$
\begin{array}{rc}
 & \begin{array}{cc} Tr & Abs \end{array} \\
\text{Transientes} \leftarrow Tr & \multirow{2}{*}{$\left(\begin{array}{c|c} Q & 0 \\ \hline R & I \end{array}\right)$} \\
\text{absorventes} \leftarrow Abs &
\end{array}
$$

$I \rightarrow$ Matriz identidade

Matriz Fundamental :

$$\boxed{F = (I - Q)^{-1}}$$ ⟹ representa o número médio de visitas a cada estado antes da absorção!

↳ Matriz

Tempo médio até à absorção:

t = soma do nº de visitas a todos os estados transientes até à absorção ⟶ soma de cada coluna de F → Valor esperado do tempo necessário até à absorção

$$\boxed{t = F' \, 1}$$

↳ vetor coluna com uns

$$t = \text{sum}(F)$$

Probabilidades de absorção

$$B = R \cdot F$$

MatLAB :

estados = [1 2 3 4]

$$T = \left[\begin{array}{ccc|c} 0,8 & 0 & 0,3 & 0 \\ 0,2 & 0,9 & 0,2 & 0 \\ 0 & 0,1 & 0,4 & 0 \\ \hline 0 & 0 & 0,1 & 1 \end{array}\right]$$

Q (bloco superior esquerdo), 0 (bloco superior direito), R (bloco inferior esquerdo), I (bloco inferior direito)

$Q = T(1{:}3, 1{:}3)$

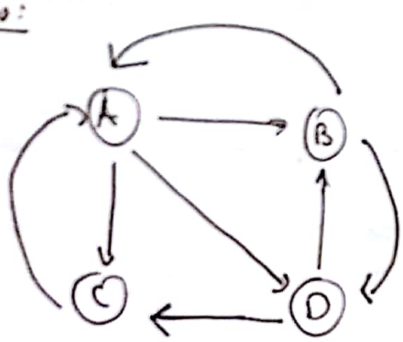
$F = \text{inv}(\text{eye}(\text{size}(Q)) - Q)$ ⟶ $F = (I - Q)^{-1}$

$t = F' \cdot \text{ones}(3,1)$ ⟶ $t = F' \cdot \underline{1}$

Page Rank:

- Consideramos a Web como um grafo orientado, em que as páginas são nodos → se existir um arco entre A e B, então existe 1 ou + links de A para B

Exemplo:



- A tem links para os outros 3
- B tem links para A e D
- C tem 1 link para A, apenas
- D tem links para B e C

Se o surfista começar em A, então, tem igual probabilidade ($\frac{1}{3}$) de ir para B, C ou D.

Se começar em B, tem $P(A) = \frac{1}{2}$, $P(D) = \frac{1}{2}$, $P(C) = 0$

Ex: Se o google apontar para o meu Blog, então vai ter um bom pagerank

Forma matricial (matriz dos Hyperlinks)

$$H_{ji} = \begin{cases} \frac{1}{d_i}, & \text{se } \exists \text{ link de } i \text{ para } j \\ 0, & \text{caso contrário} \end{cases}$$

→ contém as probabilidades de transição das páginas

$$r^{(k+1)} = H \underbrace{r^{(k)}}_{\text{vetor com pageranks na iteração } k}$$

$$H_{\text{exemplo anterior}} = \begin{bmatrix} 0 & 1/2 & 1 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix}$$

Estado estacionário:

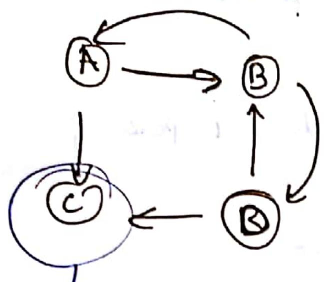
$$r = Hr$$

Aplicando $r^{(k+1)} = H r^{(k)}$ recursivamente e iniciando com $\frac{1}{n}$

$$\begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}, \begin{bmatrix} 9/24 \\ 5/24 \\ 5/24 \\ 5/24 \end{bmatrix} \cdots \begin{bmatrix} 3/9 \\ 2/9 \\ 2/9 \\ 2/9 \end{bmatrix}$$

Problemas nos passeios aleatórios:

Dead ends:



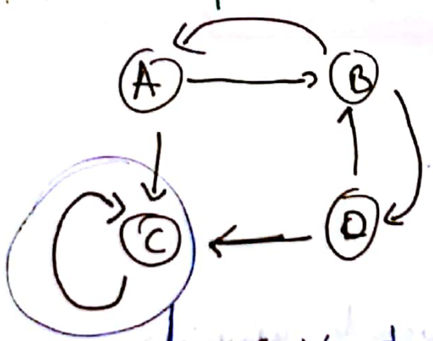
Dead end $\longrightarrow$ estado estacionário $= \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$

$\Longrightarrow$ matriz não é estocástica

Solução

- Teleport! $\longrightarrow$ ajustar a matriz de forma a seguir um link com probabilidade 1

Spider traps



Spider trap $\longrightarrow$ vetor estacionário $= \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$

Solução

- Seguir um link aleatório

Teleport

- Nunca se fica preso numa spider trap
- Torna a matriz estocástica → Não há dead ends

→ Há sempre para onde ir!

Solução do google:

- Teleports random

Page rank no matlab:

Ex: considerar um subconjunto de páginas (N = 20) recebidas a partir de www.ua.pt

Passos:

① Obter informação das páginas
② obter H
③ criar a matriz A, aplicando o método do google
④ Aplicar power method
⑤ Apresentar resultados

① [U, L] = surfer("www.ua.pt", 20)

surfer → download desta função

U → URLS

L → ligações

20 → N=20

2 e 3 Obter H e A

```
H = full(L);
c = sum (full(L));          -> numero de ligações
H = H./repmat(c, N, 1)

P = 0,85
A = P.H + (1-p).ones(N)/N       -> matriz google
A(isnan(A)) = 1/N               -> resolver dead ends
```

4 Aplicar power method

```
x0 = ones(N,1)/N        Matriz google = β.H+(1-β).ones(N)/N
iter = 1;
x = x0;
epsilon = 1e-3;

while 1
    fprintf(1, 'iteração %d\n', iter);
    xold = x;
    x = A.x;
    if (max(abs(x-xold)) < epsilon
        break
    end
    iter = iter+1;
end
```

⑤ Apresentar resultado

```
[xs idx] = sort(x, 'descend');
for p = 1:N
    fprintf(1, 'Page rank = %.3f: %s\n', x(idx(p)),
            U{idx(p)});
```

Podemos confirmar estes resultados com r = pagerank(U, L)

⟹ x - r tem que dar 0!

Problemas:

$$\underbrace{r^{(k+1)}}_{\downarrow} = A \cdot r^{k}$$

web é enorme ⟶ falta de memória!

↓

codificar a matriz apenas nas entradas não nulas

Soma de variáveis aleatórias

Média da soma de n variáveis = soma das médias $\begin{cases} Z = X + Y \\ E[Z] = E[X] + E[Y] \end{cases}$

Variância da soma de n variáveis = soma de todas as variâncias e covariâncias

→ se as variáveis são independentes ⟹ $Cov(X_j, Y_k) = 0$

$$Var(S_n) = \sum_{i=1}^{n} Var(X_i)$$

→ se forem independentes e identicamente distribuídas:

- $\underbrace{E[S_n]}_{\text{média}} = n \mu$, $\mu \to E[X_i]$
- $Var(S_n) = n \sigma^2$, $\sigma^2 \to Var(X_i)$

Funçoẽ distribuição da soma de 2 variáveis aleatórias independentes

$$Z = X + Y \implies P_Z(z) = P(X + Y = z)$$
$$= \underbrace{P_X(x) \cdot P_Y(z)}_{\text{Convolução discreta de } P_X \text{ e } P_Y}$$

Combinações lineares de variáveis aleatórias:

$$Y_n = c_1 X_1 + c_2 X_2 + \dots + c_n X_n$$

$c_1, c_2, \dots, c_n$ → constantes

$$E[Y_n] = c_1 E[X_1] + c_2 E[X_2] + \dots$$

$$Var(Y_n) = \sum_{i=1}^{n} (c_i)^2 Var(X_i) + \sum_{i=1} \sum_{j} c_i \cdot c_j \, Cov(X_i, X_j)$$

Se independentes:

$$Var(Y_n) = \sum_{i=1}^{n} c_i^2 Var(X_i)$$

Média de variáveis aleatórias:

Se criarmos a variável aleatória relativa à média de n variáveis independentes e identicamente distribuídas (IID):

- $M_n = \frac{S_n}{n}$

- $E[M_n] = E\left[\frac{S_n}{n}\right] = \frac{\sum_{i=1}^{n} E[x_i]}{n} = E[x_i] = \mu$

- $\underbrace{Var(M_n) = Var\left(\frac{S_n}{n}\right) = \frac{\sigma^2}{n}}$

Quanto mais experiências rasa é a variância da estimativa da média

Desigualdade de Markov e Chebyshev:

- Permitem estabelecer majorantes

Desigualdade de Markov:

$$P(X \geq a) \leq \frac{E[X]}{a}$$

} Dá-nos um limite superior para a probabilidade da função X ser maior ou igual a um determinado valor

Exemplo:

Média de alturas = 1,66 m

P (altura > 2 metros) num grupo

$$P(x \geq 2) \leq \frac{E[X]}{2} = 0,83$$

Média da turma = 15,2

P (nota ≥ 17) numa

$$P(x \geq 17) \leq \frac{15,2}{17} = 0,8941$$

Desigualdade de Chebyshev:

Probabilidade da diferença entre a variável e o seu valor esperado ser superior, ou inferior a um determinado valor:

- $P(|X - E[X]| \geq a) \leq \frac{Var(X)}{a^2}$
- $P(|X - E[X]| < a) \geq 1 - \frac{Var(X)}{a^2}$

} Desigualdade de Chebyshev

⇒ Se expressarmos a em função do desvio padrão, $a = h\sigma$ temos:

$$P(|X - E[X]| \geq h\sigma) \leq \frac{\sigma^2}{(h\sigma)^2} = \frac{1}{h^2}$$

A probabilidade de obter um valor que dista da média de h desvios padrão, ou mais, é menor ou igual a $\frac{1}{h^2}$

Com variáveis aleatórias

- $P(|M_n - E[M_n]| \geq \varepsilon) \leq \frac{Var(M_n)}{\varepsilon^2} = \frac{\frac{\sigma^2}{n}}{\varepsilon^2}$
- $P(|M_n - E[M_n]| \leq \varepsilon) \geq 1 - \frac{\sigma^2}{n\varepsilon^2}$

$\lim_{n \to +\infty}$ ⟹ média estabiliza!

⟹ variância vai para 0!

Lei fraca dos grandes números:

$$\lim_{n \to +\infty} P\left(|M_n - E[M_n]| < \varepsilon \right) = 1$$

$\Rightarrow$ Afirma que, para um valor de n suficientemente elevado, a média das amostras estará muito próxima do valor esperado

Lei forte dos grandes números:

$$\boxed{P\left(\lim_{n \to \infty} M_n = \mu \right) = 1} \longrightarrow$$ garante que é certo que o limite para o qual tende a média das amostras é o valor esperado

Seja $I_j \Rightarrow$ variável aleatória indicadora da ocorrência de A de ordem j

Número de ocorrências de A nas n experiências $\Rightarrow N_n = I_1 + I_2 + \dots + I_n$

$$\rightarrow f_A(n) = \frac{I_1 + I_2 + \dots + I_n}{n} \Rightarrow f_A$$ = média das amostras das variáveis aleatórias I_j

- $\lim_{n \to \infty} P\left(| f_A(n) - p(A) | < \varepsilon \right) = 1$
- $p\left(\lim_{n \to \infty} f_A(n) = p(A) \right) = 1$

A freq. relativa é uma boa estimativa da probabilidade

Informação e probabilidade:

Ao comunicar a ocorrência de um acontecimento transfere-se uma quantidade de informação que depende da probabilidade do acontecimento

$\rightarrow$ informação é uma função da probabilidade $\Longrightarrow I(p)$

- Acontecimento certo $\Longrightarrow$ Não há comunicação $\Longrightarrow I(p)=0$
- Quanto mais improvável é o acontecimento maior é a informação associada

$$p\downarrow \;\Longrightarrow\; I(p)\uparrow$$

p e q $\Longrightarrow$ probabilidades de 2 acontecimentos independentes

$\rightarrow$ $I(pq) = I(p) + I(q)$

$$= \log\left(\frac{1}{pq}\right)$$

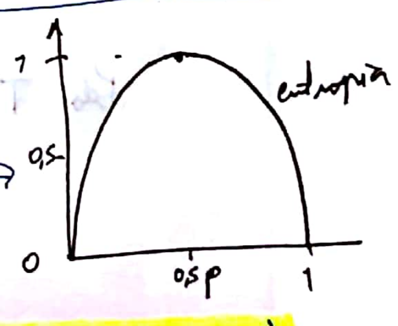
Informação de comunicar o resultado de uma experiência que tem 2 resultados equiprováveis é, em bits:

$$I\left(\tfrac{1}{2}\right) = \log \frac{1}{\frac{1}{2}} = \log(2) = 1 \text{ bit}$$

NOTA $\Longrightarrow$ Para armazenar n bits de dados nem sempre são necessários n bits de memória

Entropia:

- mede a imprevisibilidade
- dada pela média da informação



$$H(x) = E[I(x)] = \sum_{k=1}^{M} p(x_k)\cdot\log\left(\frac{1}{p(x_k)}\right)$$

- Se a variável aleatória tomar 2 valores equiprováveis $\Rightarrow H(X) = \frac{1}{2}\log(2) + \frac{1}{2}\log(2) = 1$ bit

- Se as probabilidades forem p e 1-p $\Rightarrow H(X) = -p \cdot \log(p) - (1-p)\log(1-p)$

- Se a variável aleatória X tomar 1 de 2^n valores equiprováveis $\Rightarrow H(X) = \sum_{n=1}^{2^n} 2^{-n} \cdot \log(2^n) = 2^n \cdot 2^{-n} \cdot n$

$= n$ bits

Teorema do Limite central:

Def • A soma de variáveis independentes e identicamente distribuídas (iid) tende para uma distribuição normal quando o número de variáveis é grande.

X_1 e $X_2 \Rightarrow$ variáveis aleatórias IID

• $E[X_i] = \mu$ / $Var[X_i] = \sigma^2$

• $S_n \Rightarrow$ soma das n primeiras variáveis

• $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} \Rightarrow$ variável aleatória de média nula e variância unitária

Pelo TLC $\Rightarrow \lim_{n \to \infty} P(Z_n \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{x^2}{2}} dx$

$$M_n = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$$

$\hookrightarrow$ Pelo TLC $\Rightarrow$ $M_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$, quando $n \to \infty$

TLC $\Rightarrow$ resultado extremamente importante e abre caminho a muitas aplicações.

Exemplo:

despesas de um restaurante são v.a. i.i.d. com $\mu = 6,5€$ e $\sigma =$

P(100 i-ésimos clientes gastarem um total > 600 €) = ?

- $S_{100} = X_1 + X_2 + \dots + X_{100}$
- $E[S_{100}] = 100\mu = 650€$
- $n\sigma^2 = 100 \cdot 2.5^2 = 625€$

$$\left\{ Z_n = \frac{S_{100} - 650}{25} \right.$$

Como, pelo TLC Z_{100} segue uma lei $N(0,1)$:

$$P(S_{100} > 600) = P\left(Z_{100} > \frac{600-650}{25}\right)$$

$= P(Z_{100} > -2)$

$= 1 - \phi(-2) = 1 - (1 - \phi(2)) =$

$= \phi(2) = 0,97725$

Em Matlab:

Como obter $\phi(2)$:

$z = 2$

$m = 0$

$sigma = 1$

$p = \underline{cdf}('Normal', z, m, sigma)$

$= 0{,}9772$

No Matlab aplica-se diretamente a S_{100}:

$s = 600 \longrightarrow P(S_{100} > 600)$

$m = 650 \longrightarrow$ média de S_{100}

$sigma = 25 \longrightarrow \sigma$ de S_{100}

$p = 1 - \underline{cdf}("Normal", s, m, sigma)$

Exemplo: Quantas pessoas devemos inquirir para ter uma probabilidade de 95% de que não cometemos um erro superior a 5%.

; dua
ão que
; futebol

$$\boxed{P(|M_n - p| \leq 0{,}05) \geq 0{,}95}$$

probabilidade de $|M_n - p| \leq 0{,}05 = ?$

$$P\left(\left|\frac{S_n - np}{n}\right| \leq 0{,}05\right)$$

$Z_n = \frac{S_n - np}{\sigma\sqrt{n}}$ $\xrightarrow{\text{manipulação}}$ $P\left(\left|\frac{S_n - np}{\sqrt{n}\,\sigma}\right| \leq \frac{0{,}05\sqrt{n}}{\sigma}\right)$

$\hookrightarrow$ Tende para $N(0,1)$, logo:

$$P(|M_n - p| \leq 0{,}05) \simeq P\left(|Z| \leq 0{,}05\frac{\sqrt{n}}{\sigma}\right)$$

Usar majorante para obter Variância:

$$p(1-p) \leq \frac{1}{4} \qquad (6 = 0,5)$$

$$P(|\Pi_n - p| \leq 0,05) \leq P(|Z| \leq 0,1\sqrt{n})$$

$$= P(-0,1\sqrt{n} \leq Z \leq 0,1\sqrt{n})$$
$$= F_{N(0,1)}(0,1\sqrt{n}) - F_{N(0,1)}(-0,1\sqrt{n})$$

Colocar em função de $Q(z)$:

$$1 - F_{N(0,1)}(z) \quad e \quad F_{N(0,1)}(-z) = Q(z)$$

$$1 - Q(0,1\sqrt{n}) - Q(0,1\sqrt{n}) =$$
$$= 1 - 2Q(0,1\sqrt{n})$$

Assim:

$$\boxed{1 - 2Q(0,1\sqrt{n}) \geq 0,95}$$

$$\Rightarrow Q(0,1\sqrt{n}) \geq 0,025$$

tabela

$$\Rightarrow 0,1\sqrt{n} \geq 1,96$$
$$\Rightarrow \sqrt{n} = 19,6$$
$$\Rightarrow n = 384,16 \Longrightarrow 385 \text{ pessoas}$$