

UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
CURSO DE MATEMÁTICA

*PageRank: Uma aplicação prática para as Cadeias
de Markov*

Renato José Policani Borseti

Rio de Janeiro
2013

Renato José Policani Borseti

*PageRank: Uma aplicação prática para as Cadeias de
Markov*

Monografia apresentada ao Curso de Matemática da UNIRIO, como requisito para a obtenção parcial do grau de LICENCIADO em Matemática.

Orientadora: Luzia da Costa Tonon Martarelli
Mestre em Matemática - UNIRIO

Rio de Janeiro
2013

Borseti, Renato

PageRank: Uma aplicação prática para as Cadeias de Markov /
Renato Borseti - 2013

xx.p

1.Probabilidade 2. Processos Estocásticos. I.Título.

CDU 536.21

Renato José Policani Borseti

PageRank: Uma aplicação prática para as Cadeias de Markov

Monografia apresentada ao Curso de Matemática da UNIRIO, como requisito para a obtenção parcial do grau de LICENCIADO em Matemática.

Aprovado em XX de dezembro de 2013.

BANCA EXAMINADORA

Luzia da Costa Tonon Martarelli

Mestre em Matemática - UNIRIO

Beatriz Malajovich

Mestre em Matemática Aplicada - UNIRIO

Bruno Francisco Teixeira Simões

Doutor em Engenharia de Produção - UNIRIO

*A Deus, pois sem Sua vontade nada é possível.
A minha querida esposa e meu amado filho,
pela paciência e apoio durante todos esses anos.
Aos amigos e professores, pelo apoio e compa-
nheirismo.*

Resumo

A internet possui bilhões de sites e, a cada hora, é criado um novo site com novas informações. Como saber quais sites possuem as informações desejadas? Ou qual é o site que possui a maior probabilidade de ter a informação solicitada pelo usuário? Para responder a essas questões dois ex-alunos da Universidade de Stanford - Larry Page e Sergey Brin os fundadores do Google - criaram um processo chamado PageRank. PageRank é uma família de algoritmos de análise de rede que atribui pesos numéricos a cada elemento de uma coleção de documentos hiperligados, como as páginas da Internet, com o propósito de medir a sua importância nesse grupo por meio de um motor de busca. O algoritmo pode ser aplicado a qualquer coleção de objetos com ligações recíprocas e referências. O Google tem os direitos de licença exclusivos sobre a patente de PageRank. Neste trabalho será apresentado a matemática que apoia o PageRank, como as cadeias de Markov e suas matrizes de transição, e explorado o Método das Potências como método para obtenção da matriz estacionária. O objetivo é obtermos o vetor de distribuição estacionária o qual contém a importância de cada site baseado em seus links de referência.

Palavras-chaves: Probabilidade, Cadeias de Markov, Processos Estocásticos, PageRank.

Abstract

The internet has billions of websites and every moment a new site with new information is created. How to know which sites have the information you want ? Or what is the website has the highest probability of having the information requested by the user ? To answer these questions, two former Stanford University students, Larry Page and Sergey Brin Google's founders, created a process called PageRank. PageRank is a family of algorithms for network analysis that assigns numeric values to each element of a collection of hyperlinked documents such as web pages, in order to measure its importance in this group via a search engine. The algorithm can be applied to any collection of objects with reciprocal links and referrals. Google has exclusive license rights on the patent to PageRank. In this work the math that supports the PageRank will be presented as Markov chains and their transition matrices will be explored "Method of Powers" as a method for obtaining the stationary matrix and, thus, give us the stationary distribution vector which contains the importance of each site based on your referral links.

Keywords: Probability, Markov chains, Stochastic process, PageRank algorithm.

Agradecimentos

A todos os meus parentes, pelo encorajamento e apoio. A professora Luzia Tonon pela orientação e amizade, sem a qual este trabalho não se realizaria. Agradeço especialmente os professores Ronaldo Busse, Leonardo Silvaes, Luiz Amâncio e Michel Cambrinha pelo apoio e auxílio durante todo o curso, mais do que professores, amigos para uma vida toda.

A todos os professores do Departamento de Matemática pelos seus ensinamentos e orientações e aos funcionários do curso, que durante esses anos, contribuíram de algum modo para o meu aprendizado pessoal e profissional.

“At a given moment there is only a fine layer between the ”trivial” and the impossible. Mathematical discoveries are made in this layer”.

Andrey Nikolaevich Kolmogorov

Sumário

1	Introdução	8
2	Conceitos Básicos	9
2.1	Probabilidade	9
2.1.1	Propriedades da Probabilidade	13
2.1.2	Probabilidade Condicional	14
2.1.3	Teorema de Bayes	15
2.1.4	Variáveis Aleatórias	16
2.1.5	Valor Esperado	20
2.2	Matrizes	22
2.2.1	Matriz	22
2.2.2	Transposta de uma matriz	23
2.2.3	Matrizes Quadradas	23
2.2.4	Potências de matrizes	23
2.2.5	Matriz regular	24
2.2.6	Polinômio Característico	24
2.2.7	Polinômio Mínimo	25
2.2.8	Transformações Lineares	25
2.2.9	Autovalores e Autovetores	26
2.2.10	Operadores Nilpotentes	28
2.2.11	Forma Canônica de Jordan	28
2.2.12	Teorema de Perron-Frobenius	29
2.3	Convergência em \mathbb{R}^n	30

2.3.1	Sequências no espaço euclidiano	30
3	Cadeias de Markov	32
3.1	Processos Estocásticos	32
3.2	Cadeias de Markov Discretas	34
3.3	Probabilidades de transição estacionárias	34
3.3.1	Matriz de Transição	37
3.4	Distribuição Estacionária	41
3.5	Método das Potências	42
4	Page Rank	44
4.1	Importância de um site	44
4.2	Matriz Google	48
5	Conclusão	51
	Referências Bibliográficas	52

1 Introdução

Com o advento da Internet vimos a explosão na criação de sites, blogs e redes sociais, mas nos deparamos com uma simples pergunta: como procurar uma informação dentro da Internet? Para isso foram criados os motores de busca como Google, Yahoo! e outros, cada um com um sistema de busca diferenciado. O Google se tornou o motor de busca mais famoso e utilizado da Internet por um simples motivo, o seu algoritmo de busca chamado PageRank. Este algoritmo permite colocar em ordem de importância os sites onde há uma probabilidade de encontrar a informação desejada pelo usuário, enquanto que os outros motores de busca não eram tão eficientes ao fazer o mesmo ranqueamento. Esta monografia trata da etapa em que é avaliada a importância dos sites, de forma que, quando o usuário faz uma busca e o subconjunto de dados com a informação desejada é encontrado, as páginas mais importantes sejam apresentadas nas primeiras posições da lista de resultados.

No capítulo 2, Conceitos Básicos, é apresentado uma revisão de probabilidade, matrizes e convergência no \mathbb{R}^n contemplando definições e teoremas que serão úteis ao longo do trabalho. O capítulo 3 apresenta as Cadeias de Markov, bem como os processos estocásticos e o Método das Potências para a obtenção dos autovalores e autovetores da matriz estacionária, que será utilizado no PageRank. No capítulo 4, temos o algoritmo do PageRank e a famosa matriz Google. O PageRank é uma família de algoritmos que atribuem uma *importância* a cada site em função da quantidade de links que recebem. Com isso, é possível determinar qual site é o mais importante em relação a pesquisa do usuário mas, para obtermos o vetor que determina a importância de cada site dependemos de certas condições. Quando essa matriz de transição não satisfaz a essas condições, podemos construir outra matriz a partir da modificação da matriz original utilizando um parâmetro $\alpha \in [0, 1]$. A matriz final será regular e irredutível, logo irá satisfazer as condições impostas e temos a matriz desejada. Essa matriz obtida a partir da alteração da matriz original é conhecida como **matriz Google**.

2 Conceitos Básicos

2.1 Probabilidade

As cadeias de Markov tem suas raízes na Teoria da Probabilidade, por isso aqui será feita uma revisão dos conceitos de probabilidade que são de interesse na análise das cadeias de Markov. Chamamos de *fenômenos* ou *experimentos aleatórios* as incertezas encontradas na natureza. A probabilidade pode ser definida como se segue em Viali [4]:

A probabilidade é o ramo da matemática que pretende modelar fenômenos não determinísticos, isto é, aqueles fenômenos em que o acaso representa um papel preponderante.

Os teoremas, lemas, proposições e definições aqui apresentados se encontram em [1], [2], [3], [5], [6], [9], [10] e [11].

Na probabilidade o espaço amostral, normalmente simbolizado por Ω , é o conjunto de todos os resultados possíveis de um experimento aleatório. Ele pode ser enumerável, finito ou infinito, ou não enumerável. Cada resultado obtido no experimento é chamado de *ponto* ou *elemento* de Ω e denotado por ω . Qualquer subconjunto de Ω é chamado de *evento*, que se encontram em uma σ -álgebra \mathbf{F} , e somente a esses eventos atribuímos probabilidade.

Por exemplo, quando lançamos uma moeda honesta temos o seguinte espaço amostral:

$$\Omega = \{K, C\},$$

Onde K representa o lado cara e C o lado coroa. Definamos o subconjunto $A = K$ como o evento no qual a moeda mostra o lado cara e o subconjunto $B = C$ como o evento no qual a moeda mostra o lado coroa. Ao repetimos a mesma experiência para duas moedas honestas, o novo espaço amostral será

$$\Omega = \{(K, K), (K, C), (C, K), (C, C)\}.$$

Neste espaço vamos definir o evento $A = \{(K, K)\}$ como o evento onde ambas as moedas mostram o lado cara e $B = \{(K, C), (C, K), (C, C)\}$ como o evento onde ao menos uma

moeda apresenta o lado coroa. Definimos a **união** de dois eventos A e B de um espaço amostral Ω denotado como $A \cup B$ o conjunto de todos os elementos que pertencem a A ou B ou ambos. A **interseção** de dois eventos A e B , denotado como $A \cap B$, é definido como o conjunto de todos os elementos que pertencem a A e B , e o **complementar** do evento A , denotado por A^c , é o conjunto de todos os elementos do espaço amostral que não pertencem a A .

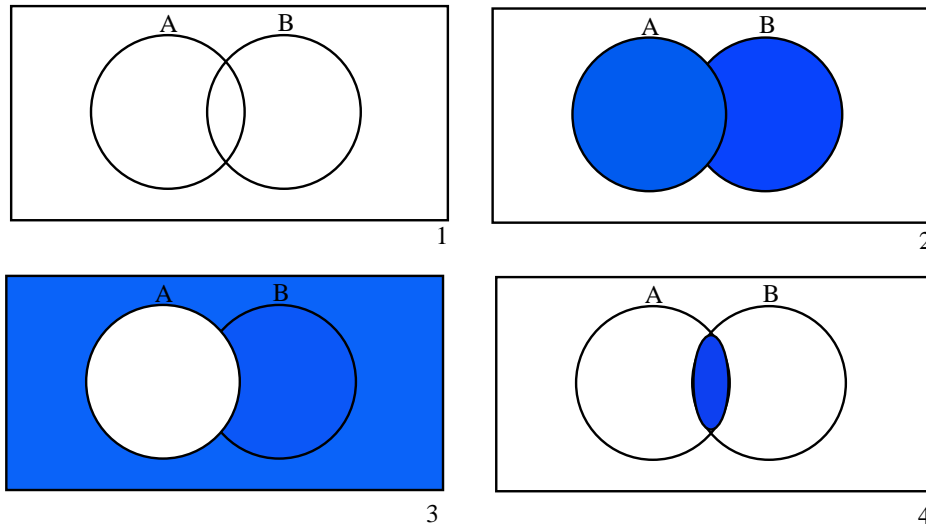


Figura 2.1: 1 - conjuntos A e B . 2 - União de A com B . 3 - Complementar de A . 4 - Interseção de A com B .

Retornando ao exemplo das duas moedas, se fizermos $A \cap B = \{\emptyset\}$, agora se fizermos $A \cup B = \{(K, K), (K, C), (C, K), (C, C)\}$ e $A^c = B$.

Definição 2.1.1. σ -álgebra

Uma classe de subconjuntos de Ω , representada por \mathbf{F} , é denominada uma σ -álgebra se satisfaz as seguintes propriedades:

- $\Omega \in F$;
- Se $A \in F$, então $A^c \in F$;
- Se $A_i \in F$, $i \geq 1$, então

$$\bigcup_{i=1}^{\infty} A_i \in F$$

Se apenas a união finita está em \mathbf{F} temos uma classe menos restrita, denominada *álgebra*.

Tomemos como exemplo o lançamento de um dado de 6 lados não viciado, isto é, a probabilidade de aparição das faces é igual para todos. Como Ω é o conjunto de todos os resultados possíveis, temos:

$$\Omega = 1, 2, 3, 4, 5, 6 \Rightarrow \#(\Omega) = 6.$$

Seja \mathbf{A} o evento onde a face de cima é ímpar. Logo:

$$A = 1, 3, 5 \Rightarrow \#(A) = 3.$$

Nos é intuitivo que, se repetirmos o experimento várias vezes obteremos um número ímpar em aproximadamente metade dos casos, ou seja, o evento \mathbf{A} vai ocorrer mais ou menos a metade das vezes. Esta intuição ocorre pelos seguintes motivos:

- os eventos são todos equiprováveis, isto é, tem a mesma probabilidade de ocorrência;
- o número de elementos de A é justamente a metade do número de elementos de Ω .

Baseada nesses itens, a probabilidade de ocorrer um evento como A é:

$$P(A) = \frac{\#(A)}{\#(\Omega)} = \frac{1}{2}.$$

Laplace denominava os elementos de A como *casos favoráveis* e os elementos do espaço amostral como *casos possíveis*. Portanto:

$$P(A) = \frac{\text{Número de elementos de } A}{\text{Número total de elementos em } \Omega}.$$

A partir dessa definição podemos resolver inúmeros problemas utilizando as técnicas de contagem da análise combinatória, mas se o número de elementos de Ω for infinito, então precisaremos tratar a definição com o uso de limites.

Se Ω for não enumerável, o conceito se aplicará ao comprimento de intervalos, áreas ou similares, dando origem a *probabilidade geométrica*. Uma outra definição, denominada *frequentista* ou *estatística*, considera o limite de frequências relativas como o

valor da probabilidade. Para tal, seja n_A o número de ocorrências de A em n repetições independentes do experimento em questão. Logo,

$$P(A) = \lim_{n \rightarrow \infty} \frac{n_A}{n}.$$

Essa definição vem da **Lei dos Grandes Números** e não será abordada neste trabalho, para o leitor mais interessado recomendamos a leitura de [1] página 323. As definições acima apresentadas, no entanto, não são suficientes para uma formulação matemática mais rigorosa da probabilidade. Por volta de 1930 Kolmogorov (1903 - 1987) apresentou um conjunto de axiomas matemáticos para definir probabilidade, permitindo incluir as definições anteriores como casos particulares. Vamos utilizar a definição que consta em [1].

Definição 2.1.2. Probabilidade

Uma função P , definida na σ -álgebra F de subconjuntos de Ω e com valores em $[0,1]$, é uma *probabilidade* se satisfaz os **Axiomas de Kolmogorov**:

- $P(\Omega) = 1$;
- Para todo subconjunto $A \in F$, $P(A) \geq 0$;
- Para toda sequência $A_1, A_2, \dots \in F$, mutuamente exclusivos, temos

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Definimos a trinca (Ω, F, P) como *espaço de probabilidade*.

2.1.1 Propriedades da Probabilidade

Proposição 2.1.1. *Dado (Ω, F, P) , suponha que os conjuntos mencionados abaixo são eventos nesse espaço de probabilidade. Temos:*

- $P(A) = 1 - P(A^c)$.
- *Sendo A e B dois eventos quaisquer, vale $P(B) = P(B \cap A) + P(B \cap A^c)$.*
- *Se $A \subset B$ então $P(A) \leq P(B)$.*
- *Regra da Adição de Probabilidades:*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Demonstração. A demonstração dessas propriedades está disponível em [1].

□

2.1.2 Probabilidade Condicional

Definição 2.1.3. Probabilidade Condicional

A ideia da probabilidade condicional é descobrir a probabilidade de ocorrer um evento B dado que um evento A já tenha ocorrido. Se A e B pertencem ao mesmo espaço de probabilidade (Ω, F, P) , denotamos a **probabilidade condicional** de B dado A como

$$P(B|A) = \frac{P(A \cap B)}{P(A)}. \quad (2.1)$$

Note que $P(B|A)$ só está definido quando $P(A) > 0$.

Observe ainda que podemos reescrever a equação (1.1) como

$$P(A \cap B) = P(B|A)P(A), \quad (2.2)$$

o que nos leva a seguinte proposição:

Proposição 2.1.2. Regra do Produto

Para os eventos A_1, A_2, \dots, A_n em (Ω, F, P) , com $P(\bigcap_{i=1}^n A_i) > 0$, a regra do produto é dada por:

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1) \dots P(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1}).$$

Demonstração. A demonstração dessa proposição pode ser encontrada em [1], [2] ou [3].

□

Definição 2.1.4. Conjunto das Partes

Dado um conjunto A , indicaremos como $\beta(A)$ o *conjunto das partes de A* . Os elementos de $\beta(A)$ são os subconjuntos de A .

$$\beta(A) = \{X; X \subset A\}.$$

Definição 2.1.5. Partição de Conjunto

Seja A um conjunto não vazio e $\beta(A)$ o conjunto das partes de A . Dizemos que o conjunto não vazio $\mathbb{P} \subset \beta(A)$ é uma **partição do conjunto A** se:

- $\forall B_1, B_2 \in \mathbb{P}, B_1 \neq B_2 \Rightarrow B_1 \cap B_2 = \emptyset$.
- $\bigcup_{B \in \mathbb{P}} B = A$.

Teorema 2.1.1. Lei da Probabilidade Total

Suponha que os eventos C_1, C_2, \dots, C_n em (Ω, F, P) formam uma partição de Ω e todos têm probabilidade positiva. Então, para qualquer evento A , vale

$$P(A) = \sum_{i=1}^n P(C_i)P(A|C_i).$$

Demonstração. A demonstração se encontra em [1]. □

Munidos dos teoremas e proposições anteriores, podemos enunciar o teorema de Bayes ¹.

2.1.3 Teorema de Bayes

Teorema 2.1.2. *Suponha que os eventos C_1, C_2, \dots, C_n em (Ω, F, P) formam uma partição de Ω e todos têm probabilidade positiva. Seja A um evento qualquer com $P(A) > 0$. Então, para todo $j = 1, 2, \dots, n$, temos*

$$P(C_j|A) = \frac{P(A|C_j)P(C_j)}{\sum_{i=1}^n P(A|C_i)P(C_i)}.$$

Demonstração. Na expressão do lado direito, o numerador é $P(A \cap C_j)$ pela proposição 2.1.2. O denominador é $P(A)$ pelo teorema 2.1.1. Portanto, pela definição de probabilidade condicional, o teorema está demonstrado. □

Para melhor compreendermos o teorema de Bayes, vamos a um exemplo. Suponhamos que uma determinada empresa de anti-vírus resolva testar um novo método para detecção de um tipo específico de vírus. O laboratório garante que o teste tem precisão de 90% mas com 5% de chances de apresentar um falso positivo. Se 15% dos arquivos em um diretório de um disco rígido tem o vírus, qual é a probabilidade de que, escolhido um arquivo ao acaso, teremos o resultado de seu teste positivo? E qual é a probabilidade de que, dado um arquivo que teve resultado positivo, realmente não possua o vírus?

Chamemos de \mathbf{D} o evento onde o arquivo seja realmente um portador do vírus, $\overline{\mathbf{D}}$ o evento onde o arquivo não possua o vírus e \mathbf{A} o evento o qual o teste tem

¹Thomas Bayes (1702 - 1761).

resultado positivo. Da hipótese do problema sabemos que $P(D) = 0,15$ e $P(\bar{D}) = 0,85$. A probabilidade do resultado do teste ser positivo para um arquivo que tenha o vírus é de $P(A|D) = 0,9$, enquanto que a probabilidade para um resultado positivo para um arquivo não portador é de $P(A|\bar{D}) = 0,05$. Portanto, a probabilidade do teste dar um resultado positivo é

$$P(A) = P(D)P(A|D) + P(\bar{D})P(A|\bar{D}) = 0,1775.$$

Daí, a probabilidade do arquivo não ter um vírus dado que o teste resulta positivo é

$$P(\bar{D}|A) = \frac{P(\bar{D})P(A|\bar{D})}{P(D)P(A|D) + P(\bar{D})P(A|\bar{D})} = 0,2394.$$

A fórmula de Bayes é muito útil para encontrar as probabilidades de um dado resultado, mas as vezes estamos mais interessados no processo gerador do resultado do que em um resultado específico. Por exemplo, tenhamos um jogo de dados onde ganha quem conseguir lançar dois dados, não viciados, e a soma dos valores de suas faces seja 7. Note que neste cenário estamos interessados no resultado e não nos pares (2,5) ou (1,6) ou (3,4) etc. Essas quantidades de interesse, ou mais formalmente, essas funções de valores reais definidas no espaço de probabilidade, são conhecidas como **variáveis aleatórias**.

2.1.4 Variáveis Aleatórias

Definição 2.1.6. Variável Aleatória

Seja (Ω, F, P) um espaço de probabilidade. Denominamos por *variável aleatória* qualquer função $X : \Omega \rightarrow \mathbb{R}$ tal que

$$X^{-1}(I) = \{\omega \in \Omega : X(\omega) \in I\} \in F,$$

para todo intervalo $I \subset \mathbb{R}$. Em palavras, X é variável aleatória se sua imagem inversa para intervalos $I \subset \mathbb{R}$ pertencem a σ -álgebra F .

Pela definição podemos ver que, por exemplo, no lançamento de uma moeda, observar o resultado obtido, cara ou coroa, não é uma variável aleatória, pois os resultados não são números. Mas se associarmos o número 0 à ocorrência de cara e o número 1 à ocorrência de coroa, teremos uma variável aleatória. Se os dados obtidos puderem assumir qualquer valor em um determinado intervalo, devemos introduzir uma outra definição:

Definição 2.1.7. Uma variável aleatória é dita discreta se sua imagem é um conjunto finito e enumerável. Se a imagem for um conjunto não enumerável, então dizemos que a variável aleatória é contínua.

Um exemplo de variável aleatória contínua é a renda mensal de uma família em uma pesquisa domiciliar, pois essa variável pode assumir qualquer valor real.

Função de Distribuição

Sendo X uma variável aleatória em (Ω, F, P) , sua *função de distribuição* é definida por

$$F_X(x) = P(X \in (-\infty, x]) = P(X \leq x), \quad x \in \mathbb{R}.$$

Segundo Magalhães em [1], conhecer a função de distribuição nos permite obter qualquer informação sobre a variável aleatória, mesmo que a variável aleatória só assuma valores em um subconjunto dos reais (a função de distribuição é definida em toda a reta). Em algumas literaturas essa função é conhecida como *função de distribuição acumulada* por acumular as probabilidades dos valores inferiores ou iguais a x . Usaremos a notação F ao invés de F_X , caso contrário será colocado uma notificação.

Definição 2.1.8. Propriedades da Função de Distribuição

Uma função de distribuição de uma variável X em (Ω, F, P) obedece às seguintes propriedades:

- Os limites de $F(x)$ no infinito são

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \text{e} \quad \lim_{x \rightarrow \infty} F(x) = 1;$$

- F é contínua à direita;
- F é não decrescente, isto é, $F(x) \leq F(y)$ sempre que $x \leq y, \forall x, y \in \mathbb{R}$.

Demonstração. A demonstração se encontra em [1]. □

Função de Probabilidade de uma variável aleatória discreta

A *função de probabilidade* de uma variável aleatória discreta é uma função que atribui probabilidade a cada um dos possíveis valores assumidos pela variável. Ou seja, se X é uma variável com valores x_1, x_2, \dots , temos, para $i = 1, 2, \dots$,

$$p(x_i) = P(X = x_i) = P(\{\omega \in \Omega : X(\omega) = x_i\}).$$

Definição 2.1.9. Propriedades da Função de Probabilidade

A função de probabilidade de X em (Ω, F, P) satisfaz:

- $0 \leq p(x_i) \leq 1, \forall i = 1, 2, \dots$
- A soma de $p(x_i)$ é $\sum_i p(x_i) = 1$,

com a soma percorrendo todos os possíveis valores (eventualmente infinitos).

Demonstração. A demonstração se encontra em [1]. □

Vamos a um exemplo encontrado em [1]. Considere dois lançamentos independentes de uma moeda equilibrada. Com o espaço de probabilidade sendo o usual, defina X como sendo o número de caras nos dois lançamentos. A variável X será discreta e sua função de probabilidade será dada por:

X	0	1	2
$p(x_i)$	1/4	1/2	1/4

A função de distribuição correspondente será:

$$F(x) = \begin{cases} 0, & \text{se } x < 0, \\ \frac{1}{4}, & \text{se } 0 \leq x < 1, \\ \frac{3}{4}, & \text{se } 1 \leq x < 2, \\ 1, & \text{se } x \geq 2. \end{cases}$$

Função de Densidade de uma variável aleatória contínua

Uma variável aleatória X em (Ω, F, P) , com função de distribuição F é dita *contínua* se existir uma função não negativa f tal que:

$$F(x) = \int_{-\infty}^x f(\omega) d\omega, \quad \forall x \in \mathbb{R}.$$

A função f é denominada *função densidade*.

Definição 2.1.10. Propriedades da Função Densidade

A função densidade de X em (Ω, F, P) , $f(x)$, satisfaz:

- $f(x) \geq 0, \quad \forall x \in \mathbb{R}.$

- $\int_{-\infty}^{\infty} f(\omega) d\omega = 1.$

Demonstração. A demonstração se encontra em [1]. □

Para obtermos a probabilidade da variável estar em um certo intervalo $(a, b]$, fazemos a integral da função densidade nesse intervalo.

$$P(a < X \leq b) = \int_a^b f(x)dx = F(b) - F(a).$$

Notemos que a integral não se altera com a inclusão ou não dos extremos a e b , ou seja, o valor seria o mesmo para (a, b) , $[a, b]$, e $[a, b)$.

Segue um exemplo encontrado em [1]. A duração, em anos, de uma certa lâmpada especial é uma variável aleatória contínua com densidade dada por:

$$f(x) = \begin{cases} 2e^{-2x} & , \quad x \geq 0, \\ 0 & , \quad \text{caso contrário.} \end{cases}$$

Para obtermos a função de distribuição $F(x) = \int_{-\infty}^x f(\omega)d\omega$, distinguimos dois casos. Para $x < 0$, $F(x) = 0$, pois a função densidade é nula nesse intervalo. Para $x \geq 0$, temos

$$F(x) = \int_{-\infty}^x f(\omega)d\omega = \int_0^x 2e^{-2\omega} d\omega = 1 - e^{-2x}.$$

Se desejarmos a probabilidade da lâmpada durar até 2 anos, calculamos $F(x)$ no ponto 2. Assim, $F(2) = 1 - e^{-4} = 0.98$.

2.1.5 Valor Esperado

Segundo Magalhães em [1], o conceito de valor esperado ou média parece ter sido historicamente desenvolvido para avaliar ganhos em jogos com apostas a dinheiro. O retorno financeiro, obtido em uma jogada de dados ou rodada de um certo jogo de cartas, seria imprevisível. A questão de interesse era avaliar esse retorno em um horizonte de várias jogadas. Seria uma espécie de balanço final de muitas jogadas, após contabilizar perdas e ganhos. Com o auxílio do formalismo matemático, essas ideias foram estabelecidas em definições rigorosas, incluindo os casos discreto e contínuo.

Valor Esperado para Variáveis Discretas

Vamos usar a definição dada em [1]. Seja X uma variável aleatória discreta com função de probabilidade p_x e valor x_i para i em um certo conjunto de índices I . O *valor esperado* ou *esperança matemática* ou *média* de X é definido por

$$E(X) = \mu_X = \sum_{i \in I} x_i p_X(x_i),$$

desde que a soma esteja determinada. A notação será simplificada para μ , sempre que não houver outras variáveis envolvidas. Segundo Magalhães em [1] “*o valor esperado pondera os valores assumidos pelas respectivas probabilidades e não precisa ser um dos valores possíveis da variável*”. O uso da denominação *média* para o valor esperado da variável tem origem histórica, mas também pode ser visto como uma referência a um resultado importante conhecido por *Lei dos Grandes Números*. Como ilustração, seja X a face obtida no lançamento de um dado honesto. O valor esperado de X é dado por

$$E(X) = \sum_{i=1}^6 x_i p_i = \frac{7}{2}.$$

Se coletarmos, de forma independente, um certo número de valores da variável X ou, em outras palavras, se coletarmos uma amostra aleatória de X , poderemos ter repetições de valores dentre aqueles que são possíveis para X . Esperaríamos que a proporção com que cada valor aparece na amostra fosse próxima da respectiva probabilidade. Dessa forma, teríamos a expectativa de que a média dos valores amostrados não ficasse distante do valor esperado da variável. Parece intuitivo admitir que, quanto maior for o tamanho da amostra, menor diferença haverá entre a proporção amostral e a respectiva probabilidade. Assim, para uma amostra suficientemente grande, podemos admitir que a média das observações se aproxima do valor esperado da variável X .

Valor Esperado para Variáveis Contínuas

Novamente, vamos definir conforme [1]. Seja X uma variável aleatória contínua com função densidade f . Definimos *valor esperado* ou *esperança matemática* ou *média* de X por

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx,$$

desde que a integral esteja bem definida. A interpretação de $E(X)$ para o caso contínuo é similar ao mencionado para variáveis discretas. O valor esperado é o centro de gravidade da distribuição da variável aleatória.

2.2 Matrizes

Nesta seção apresentaremos um resumo sobre as matrizes. É de suma importância o entendimento das operações com matrizes pois teremos matrizes criadas a partir do Webgrafo gerado pelos links que interligam os sites. Com essa matriz, podemos obter a importância de cada site e, assim, calcularmos a importância geral de um determinado site em face a pesquisa realizada pelo usuário.

2.2.1 Matriz

Uma *matriz A sobre um corpo K* ou simplesmente uma matriz A (quando K estiver subentendido) é uma tabela retangular de escalares, costumeiramente apresentada no formato

$$A = \begin{bmatrix} a_{1,1} & \dots & a_{1,n} \\ a_{2,1} & \dots & a_{2,n} \\ \dots & \dots & \dots \\ a_{m,1} & \dots & a_{m,n} \end{bmatrix}.$$

As linhas da matriz A são as m listas horizontais de escalares dadas por

$$(a_{11}, a_{12}, \dots, a_{1n}), (a_{21}, a_{22}, \dots, a_{2n}), \dots, (a_{m1}, a_{m2}, \dots, a_{mn})$$

e as colunas de A são as n listas verticais de escalares dadas por

$$\begin{bmatrix} a_{11} \\ a_{21} \\ \dots \\ a_{m1} \end{bmatrix}, \begin{bmatrix} a_{12} \\ a_{22} \\ \dots \\ a_{m2} \end{bmatrix}, \dots, \begin{bmatrix} a_{1n} \\ a_{2n} \\ \dots \\ a_{mn} \end{bmatrix}$$

Note que a_{ij} representa o elemento que consta na linha i e coluna j (também chamado de *ij-ésima entrada*). Em geral, denotamos essa matriz simplesmente escrevendo $A = [a_{ij}]$. Dizemos que uma matriz com m linhas e n colunas é uma matriz *m por n*, que escreveremos $m \times n$, onde o par m e n é dito a *ordem* da matriz A.

2.2.2 Transposta de uma matriz

A *transposta* de uma matriz A , denotada por A^T , é a matriz obtida escrevendo as colunas de A , na mesma ordem, como linhas. Por exemplo,

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \quad \text{e} \quad \begin{bmatrix} 1 & -3 & -5 \end{bmatrix}^T = \begin{bmatrix} 1 \\ -3 \\ -5 \end{bmatrix}.$$

Em outras palavras, se $A = [a_{ij}]$ é uma matriz $m \times n$, então $A^T = [b_{ij}]$ é a matriz $n \times m$ dada por $b_{ij} = a_{ji}$.

2.2.3 Matrizes Quadradas

Uma matriz é dita *quadrada* se tiver o mesmo número de linhas e colunas. Dizemos que uma matriz quadrada $n \times n$ é de *ordem* n . As operações de adição, multiplicação, multiplicação por escalar e transposição podem sempre ser efetuadas com matrizes quadradas, e o resultado da operação é uma matriz quadrada.

2.2.4 Potências de matrizes

Seja A uma matriz quadrada de ordem n com entradas de algum corpo K (no nosso trabalho, estamos sempre no corpo dos reais salvo quando explicitado). As *potências* de A são definidas como segue:

$$A^2 = AA, \quad A^3 = A^2A, \quad \dots, \quad A^{n+1} = A^nA, \quad \dots, \quad \text{e} \quad A^0 = I,$$

onde I é a matriz identidade.

2.2.5 Matriz regular

Seja A uma matriz quadrada de ordem n positiva, isto é, seus elementos são todos positivos. Dizemos que A é uma *matriz regular* se, para algum $k \geq 1$, temos $A^k > 0$.

Qualquer matriz positiva é regular. Abaixo temos alguns exemplos de matrizes regulares e não regulares:

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{e} \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

não são regulares, enquanto

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 1 \\ 4 & 3 & 2 \end{bmatrix}$$

é regular.

2.2.6 Polinômio Característico

Seja $A = [A_{ij}]$ uma matriz quadrada de ordem n . A matriz $M = A - tI_n$, em que I_n é a matriz identidade de ordem n e t é uma incógnita, pode ser obtida subtraindo t de cada elemento diagonal de A . A matriz simétrica na soma de M é a matriz $tI_n - A$ e seu determinante

$$\Delta(t) = \det(tI_n - A) = (-1)^n \det(A - tI_n).$$

que é um polinômio de grau n em t , é denominado *polinômio característico* de A .

Teorema 2.2.1. Teorema de Cayley-Hamilton

Cada matriz A é uma raiz de seu polinômio característico.

Demonstração. A demonstração se encontra em [9].

□

2.2.7 Polinômio Mínimo

Seja A uma matriz quadrada qualquer. Seja $J(A)$ o conjunto de todos os polinômios $f(t)$ para os quais A é uma raiz, isto é, para os quais $f(A) = 0$. O conjunto $J(A)$ não é vazio em virtude do teorema 2.2.1 de Cayley-Hamilton, que nos diz que o polinômio característico $\Delta_A(t)$ de A pertence a $J(A)$. Seja $m(t)$ o polinômio mônico de menor grau em $J(A)$, dizemos que $m(t)$ é o *polinômio mínimo* da matriz A . O polinômio mínimo $m(t)$ de uma matriz A divide qualquer polinômio que tenha A como raiz. Em particular, $m(t)$ divide o polinômio característico $\Delta(t)$ de A .

2.2.8 Transformações Lineares

Sejam E, F espaços vetoriais. Uma *transformação linear* $A : E \rightarrow F$ é uma correspondência que associa a cada vetor $v \in E$ um vetor $A(v) = A \cdot v = Av \in F$ de modo que valham, para quaisquer $u, v \in E$ e $\alpha \in \mathbb{R}$, as seguintes relações:

$$A(u + v) = Au + Av, \text{ e}$$

$$A(\alpha \cdot v) = \alpha \cdot Av.$$

O vetor $A \cdot v$ chama-se a *imagem* de v pela transformação A . Se $A : E \rightarrow F$ é uma transformação linear, então $A \cdot 0 = 0$. Além disso, dados $u, v \in E$ e $\alpha, \beta \in \mathbb{R}$ tem-se $A(\alpha u + \beta v) = A(\alpha u) + A(\beta v) = \alpha A(u) + \beta A(v)$. Seja A uma matriz $m \times n$ sobre um corpo K . Então A determina uma aplicação $F_A : K^n \rightarrow K^m$ por meio de

$$F_A(u) = Au,$$

em que os vetores K^n e K^m são escritos como colunas. Quando o domínio e o contradomínio de uma transformação linear coincidem, chamamos essa transformação de **operador linear**. Dizemos que dois espaços E e F são *isomorfos* sobre \mathbb{R} (escrevemos $E \cong F$) se existir uma transformação linear bijetora $T : E \rightarrow F$. Nesse caso, dizemos que T é um *isomorfismo* entre E e F .

2.2.9 Autovalores e Autovetores

Um vetor $v \neq 0$ em E chama-se um *autovetor* do operador $A : E \rightarrow F$ quando existe $\lambda \in \mathbb{R}$ tal que

$$Av = \lambda v.$$

O número $\lambda \in \mathbb{R}$, por sua vez, chama-se um *autovalor* do operador A quando existe um vetor não-nulo $v \in E$ tal que $Av = \lambda v$. Diz-se, então, que o autovalor λ corresponde, ou pertence, ao autovetor v e vice-versa. Para todo $w = \alpha v$ tem-se $Aw = \lambda w$. De fato,

$$Aw = A(\alpha v) = \alpha(Av) = \alpha(\lambda v) = \lambda(\alpha v) = \lambda w.$$

Determinando os autovalores e autovetores de um operador

Se λ é autovalor de A associado a v , então:

$$Av = v\lambda \Leftrightarrow Av - \lambda v = 0 \Leftrightarrow (A - \lambda I)v = 0 \Leftrightarrow v \in \ker(A - \lambda I).$$

Logo, $A - \lambda I$ não é injetiva, então também não é invertível, o que implica em $\det(A - \lambda I) = 0$. Vamos a um exemplo:

Queremos determinar os autovetores e autovalores do operador A descrito abaixo:

$$\begin{aligned} A &: \mathbb{R}^2 \rightarrow \mathbb{R}^2 \\ (x, y) &\mapsto (4x + 3y, x + 2y). \end{aligned}$$

Solução:

Temos

$$\begin{pmatrix} 4 & 3 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} - \lambda I \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix},$$

de modo que:

$$\det(A - \lambda I) = 0 \Rightarrow \det \begin{pmatrix} 4 - \lambda & 3 \\ 1 & 2 - \lambda \end{pmatrix} = 0 \Rightarrow (4 - \lambda)(2 - \lambda) - 3 = 0.$$

O polinômio característico de A e seus autovalores são, respectivamente,

$$\lambda^2 - 6\lambda + 8 = 0 \Rightarrow \lambda = \{1, 5\}.$$

Este operador se enquadra nas hipóteses da definição, logo \mathbb{R}^2 possui uma base formada por autovetores de $A = \{v_1, v_2\}$.

$$Av_1 = 1v_1$$

$$Av_2 = 5v_2$$

A matriz de A nessa base é:

$$\begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix}.$$

Para determinarmos os autvetores v_1 e v_2 , fazemos

$$v_1 = (a, b) \Rightarrow Av_1 = (4a + 3b, a + 2b),$$

de modo que

$$Av_1 = v_1 \Rightarrow \begin{cases} 4a + 3b = a \\ a + 2b = b \end{cases} \Rightarrow \begin{cases} 3a + 3b = 0 \\ a + b = 0, \end{cases}$$

cujas soluções são

$$S = \{(-b, b) \mid b \in \mathbb{R}\}.$$

Fazendo $b = -1$, temos $v_1 = (1, -1)$. Por outro lado,

$$v_2 = (a, b) \Rightarrow Av_2 = (4a + 3b, a + 2b).$$

Logo,

$$Av_2 = 5v_2 \Rightarrow \begin{cases} 4a + 3b = 5a \\ a + 2b = 5b \end{cases} \Rightarrow \begin{cases} 3b - a = 0 \\ a - 3b = 0, \end{cases}$$

cujas soluções são

$$S = \{(3b, b) \mid b \in \mathbb{R}\}.$$

Fazendo $b = 1$ temos $v_2 = (3, 1)$.

Assim, $\{(1, -1), (3, 1)\}$ é base de \mathbb{R}^2 formada por autovetores.

2.2.10 Operadores Nilpotentes

Dizemos que um operador linear $T : V \rightarrow V$ é *nilpotente* se $T^n = 0$, para algum inteiro positivo n . Dizemos que o *índice de nilpotência* de T é k se $T^k = 0$, mas $T^{k-1} \neq 0$. Analogamente, dizemos o mesmo para uma matriz quadrada A se existir um k inteiro positivo que satisfaça os requisitos anteriores. Claramente, o polinômio mínimo de um operador (matriz) nilpotente de índice k é $m(t) = t^k$ e, por isso, 0 é seu único autovalor.

2.2.11 Forma Canônica de Jordan

Um operador linear T pode ser colocado na forma canônica de Jordan se seus polinômios mínimo e característico admitem uma fatoração em polinômios lineares. Isso sempre vale se estivermos operando no corpo dos complexos \mathbb{C} . Em geral, sempre podemos estender o corpo base K a um corpo no qual os polinômios mínimo e característico podem ser fatorados em polinômios lineares; assim, neste sentido amplo, qualquer operador tem uma forma canônica de Jordan. Analogamente, toda matriz é semelhante a uma matriz em forma canônica de Jordan. Um *bloco de Jordan* de tamanho k é

$$B_k(\lambda) = \begin{pmatrix} \lambda & \dots & \dots & 0 \\ 1 & \lambda & \dots & \vdots \\ \vdots & 1 & \lambda & \vdots \\ 0 & \dots & 1 & \lambda \end{pmatrix}_{k \times k} = \underbrace{B_k(0)}_{\text{nilpotente}} + \text{diag}(\lambda) = J_k(0) + \text{diag}(\lambda).$$

Dizemos que uma matriz está na **forma canônica de Jordan** se ela é formada por blocos de Jordan ao longo de sua diagonal, a em cada entrada *superdiagonal* (que consiste nas entradas acima da diagonal) e zero nas demais, como no exemplo abaixo:

$$\begin{pmatrix} \lambda_1 & a & 0 & \dots & \dots & 0 \\ 1 & \lambda_1 & a & 0 & \dots & 0 \\ 0 & 1 & \lambda_1 & a & 0 & 0 \\ 0 & 0 & 0 & \lambda_2 & a & 0 \\ 0 & 0 & 0 & 1 & \lambda_2 & a \\ 0 & 0 & 0 & 0 & 0 & \lambda_3 \end{pmatrix}$$

2.2.12 Teorema de Perron-Frobenius

Teorema 2.2.2. Teorema de Perron-Frobenius² para matrizes regulares

A uma matriz quadrada regular e positiva, então

- *Existe um autovalor λ_{pf} de A real e positivo, com autovetor positivo;*
- *para qualquer outro autovalor temos $|\lambda| < \lambda_{pf}$;*
- *o autovalor λ_{pf} é simples, isto é, tem multiplicidade 1, e corresponde a um bloco de Jordan 1×1 .*

O autovalor λ_{pf} é chamado de autovalor Perron da matriz A .

Demonstração. A demonstração deste teorema foge do escopo deste trabalho, mas pode ser encontrada em detalhes em [11]. □

Teorema 2.2.3. Teorema de Perron-Frobenius para matrizes não negativas

A uma matriz quadrada e $A \geq 0$, então

- *Existe um autovalor λ_{pf} de A real e não negativo, com autovetor não negativo;*
- *para qualquer outro autovalor temos $|\lambda| \leq \lambda_{pf}$.*

Demonstração. É a mesma do teorema anterior, com a modificação de que a matriz pode conter elementos nulos. □

²Oskar Perron (1880-1975) e Ferdinand Georg Frobenius (1849-1917) - ambos matemáticos alemães

2.3 Convergência em \mathbb{R}^n

2.3.1 Sequências no espaço euclidiano

Uma sequência em \mathbb{R}^n é uma aplicação $x : \mathbb{N} \rightarrow \mathbb{R}^n$, definida no conjunto \mathbb{N} dos números naturais. O valor que essa aplicação assume no número n é indicado com x_n e chama-se o *n-ésimo termo* da sequência. Usaremos as notações $(x_n)_{n \in \mathbb{N}}$ ou $\{X_n\}_{n \in \mathbb{N}}$ para indicar uma sequência cujo *n-ésimo termo* é $x_n \in \mathbb{R}^n$.

Uma sequência $\{X_n\}_{n \in \mathbb{N}}$ em \mathbb{R}^n equivale a k sequências de números reais. Com efeito, para cada $n \in \mathbb{N}$ temos $x_n = (x_{n1}, x_{n2}, \dots, x_{nk})$ onde $x_{ni} = \pi_i(x_n) = i$ -ésima coordenada de x_n , $i = 1, 2, \dots, k$. As k sequências $(x_{ni})_{n \in \mathbb{N}}$ ($i = 1, \dots, k$) são chamadas as *sequências das coordenadas* de (x_n) . Assim, por exemplo, no plano \mathbb{R}^2 , uma sequência de pontos $z_n = (x_n, y_n)$ é o mesmo que um par de sequências $(x_n), (y_n)$ de números reais.

Diz-se que o ponto $a \in \mathbb{R}^n$ é o *limite* da sequência de pontos $x_n \in \mathbb{R}^n$ quando, para todo $\epsilon > 0$ dado, é possível obter $n_0 \in \mathbb{N}$ tal que $n > n_0 \Rightarrow |x_n - a| < \epsilon$. Nesse caso, diz-se que (x_n) *converge* para a ou *tende* para a e escreve-se $\lim x_n = a$, $\lim_{n \rightarrow \infty} x_n = a$, $\lim_{n \in \mathbb{N}} x_n = a$, ou simplesmente $x_n \rightarrow a$.

Quando existe o limite $a = \lim x_n$, diz-se que a sequência (x_n) é *convergente*. Caso contrário, diz-se que (x_n) é *divergente*. Por exemplo, uma sequência constante (a, a, \dots, a, \dots) é obviamente convergente e seu limite é a . Por outro lado, se $a \neq b$ então (a, b, a, b, \dots) é uma sequência divergente. Tem-se $\lim x_n = a \Leftrightarrow \lim |x_n - a| = 0$. Isto reduz a convergência em \mathbb{R}^n à convergência de números reais ≥ 0 .

Em termos de bolas, tem-se $\lim x_n = a$ se, e somente se, qualquer bola aberta de centro a contém todos os termos x_n salvo, possivelmente para um número finito de índices n . Temos que toda sequência convergente é limitada, mas a recíproca é falsa: Basta tomar uma sequência (a, b, a, b, \dots) , onde $a \neq b$; a sequência é limitada mas divergente.

Um fato elementar, mas essencial, é que o limite de uma sequência convergente é único. Ou seja, se $\lim x_n = a$ e $\lim x_n = b$, então $a = b$. Com efeito, para todo $n \in \mathbb{N}$, temos:

$$0 \leq |a - b| \leq |x_n - a| + |x_n - b|$$

Logo, $\lim |x_n - a| = \lim |x_n - b| = 0 \Rightarrow a = b$.

Agora podemos introduzir os conceitos de Cadeias de Markov, processos estocásticos, matrizes de transição e convergência das variáveis aleatórias no **Método das Potências**, assuntos de suma importância para este trabalho.

3 Cadeias de Markov

As cadeias de Markov são de fundamental importância para este trabalho, pois são a partir delas que temos as matrizes de transição. Daí, obtemos a matriz de importância dos sites e, assim, saberemos qual site é mais importante para o usuário em sua busca.

Segundo [3], uma **cadeia de Markov** é um tipo especial de **processo estocástico**. Antes de definirmos formalmente uma cadeia de Markov, vamos definir o que é um processo estocástico.

3.1 Processos Estocásticos

Um **processo estocástico** $\{X(t), t \in T\}$ é uma coleção de variáveis aleatórias onde, por exemplo, t é o tempo e $X(t)$ é o estado do processo no tempo t . O conjunto T é chamado de **conjunto index** do processo. Se T for enumerável, então dizemos que o processo estocástico é um processo **discreto no tempo**, caso T seja um intervalo de tempo real, então dizemos que o processo estocástico é um processo de **tempo contínuo**.

Vamos a um exemplo encontrado em [3] para facilitar o entendimento. Suponha que um escritório possua 5 linhas telefônicas e que qualquer uma dessas linhas tem a possibilidade de se encontrar em uso a qualquer momento. Durante um determinado período de tempo as linhas são observadas em intervalos regulares de 2 minutos e o número de linhas ocupadas é anotado.

Tomemos X_1 como o número de linhas ocupadas ao serem observadas no 1º intervalo, X_2 como o número de linhas que estão ocupadas quando observadas no 2º intervalo, generalizando, para $n = 1, 2, \dots$, tomemos X_n como o número de linhas que estão ocupadas quando observadas no n -ésimo intervalo de tempo.

A sequência de observações X_1, X_2, \dots , é chamada um processo estocástico, ou

processo aleatório, porque os valores de suas observações não podem ser previstos antecipadamente com precisão. No entanto, probabilidades podem ser especificadas para cada um dos diferentes valores possíveis a qualquer tempo particular. Um processo estocástico como o que foi descrito é chamado um processo com um parâmetro de tempo discreto porque as linhas são observadas somente a pontos discretos ou separados no tempo, ao contrário do continuamente no tempo.

Em um processo estocástico a primeira observação X_1 é chamada de estado inicial do processo, e para $n = 1, 2, \dots$, a observação X_n é chamado de estado do processo no tempo n . No exemplo, o estado do processo a qualquer momento é o número de linhas ocupadas naquele momento. Entretanto, cada estado deve ser um inteiro entre 0 e 5.

Em um processo estocástico com parâmetro de tempo discreto, o estado do processo varia de maneira aleatória de tempos em tempos. Para descrever um modelo probabilístico completo para um processo particular, é necessário especificar uma probabilidade para cada um dos valores possíveis do estado inicial X_1 , assim como para cada estado subsequente X_{n+1} ($n = 1, 2, \dots$) a probabilidade condicional da seguinte forma:

$$P(X_{n+1} = x_{n+1} | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n).$$

Em outras palavras, para cada tempo n , o modelo de probabilidade deve especificar a probabilidade condicional que o processo estará no estado x_{n+1} no tempo $n + 1$, dado que nos instantes $1, \dots, n$ o processo estava nos estados x_1, \dots, x_n .

3.2 Cadeias de Markov Discretas

Definição 3.2.1. Cadeias de Markov

Seja (Ω, F, P) um espaço de probabilidade. Considere a sequência de variáveis aleatórias $\{X_n\}_{n \geq 0}$ assumindo valores em um espaço de estados finito V . Dizemos que $\{X_n\}_{n \geq 0}$ é uma *cadeia de Markov discreta* se, para qualquer sequência de estados $x_1, x_2, \dots, x_{n+1} \in V$,

$$P(X_{n+1} = x_{n+1} | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = P(X_{n+1} = x_{n+1} | X_n = x_n).$$

Da proposição 2.1.2 (regra do produto), as probabilidades na cadeia de Markov devem satisfazer a relação

$$\begin{aligned} & P(X_1 = x_1, X_2 = x_2, \dots, X_{n+1} = x_{n+1}) = \\ &= P(X_1 = x_1)P(X_2 = x_2 | X_1 = x_1)P(X_3 = x_3 | X_2 = x_2) \cdots \\ &= P(X_{n+1} = x_{n+1} | X_n = x_n). \end{aligned}$$

3.3 Probabilidades de transição estacionárias

Dizemos que uma cadeia de Markov é *finita* quando há um número finito k de estados possíveis s_1, s_2, \dots, s_k e, a qualquer momento, a cadeia deve estar em um desses k estados.

A probabilidade condicional $P(X_{n+1} = s_j | X_n = s_i)$ de que a cadeia de Markov esteja no estado s_j no instante de tempo $n + 1$ se no instante n ela estiver no estado s_i é chamada de *probabilidade de transição*. Se para uma certa cadeia de Markov esta probabilidade de transição tem o mesmo valor para todo o tempo n ($n = 1, 2, \dots$), então é dito que a cadeia de Markov tem probabilidades de transição estacionárias, isto é, se para qualquer estados s_i e s_j , há uma probabilidade de transição $p_{i,j}$ tal que

$$P(X_{n+1} = s_j | X_n = s_i) = p_{i,j}, \quad \forall n = 1, 2, \dots$$

Conjunto Fechado

Definiremos **conjunto fechado** como em [12]. Considere um subconjunto C do espaço de estados V . Este conjunto é dito fechado se $p_{i,j} = 0 \quad \forall i \in C, j \notin C$. Se um conjunto fechado consiste de um único estado, este será chamado **estado absorvente**, ou seja, o estado j é absorvente se e somente se $p_{j,j} = 1$.

Uma cadeia de Markov é dita **irredutível** se não existirem conjuntos fechados não vazios exceto o próprio V . Se V tem um subconjunto fechado próprio, a cadeia é dita **reduzível**.

Estados Intercomunicantes

Dois estados, i e j , são ditos **intercomunicantes** se, para algum $n \geq 0$, $p_{i,j}^{(n)} > 0$ e para algum $m \geq 0$, $p_{j,i}^{(m)} > 0$, ou seja, nesta cadeia é possível ir de i para j em n passos e de j para i em m passos. Os inteiros m e n não precisam ser iguais.

Teorema 3.3.1. Teorema das cadeias irredutíveis

*Uma cadeia de Markov será **irredutível** se, e somente se, todos os pares de estados são intercomunicantes.*

Demonstração. A prova deste teorema será a mesma que se encontra em [12].

(\Rightarrow) Vamos supor que a cadeia seja irredutível e definamos $C_j = \{i : p_{i,j}^{(n)} = 0 \quad \forall n \geq 0\}$, isto é, C_j é o conjunto de todos os estados a partir dos quais o estado j não pode ser alcançado. O conjunto C_j é um subconjunto fechado de V e para provar isto é necessário mostrar que, se $i \in C_j$ e $k \notin C_j$, então $p_{i,k} = 0$. Contudo, se $k \notin C_j$ então para algum $m \geq 0$ segue que $p_{k,j}^{(m)} > 0$. Se $p_{i,k}$ for positivo, então $p_{i,j}^{(m+1)} = \sum_{l \in V} p_{i,l} \times p_{l,j}^{(m)} \geq p_{i,k} \times p_{k,j}^{(m)} > 0$, o que implica em $i \notin C_j$. Esta contradição nos leva a concluir que $p_{i,k} = 0$ para todo $i \in C_j$ e $k \notin C_j$. Logo C_j é fechado.

O único subconjunto fechado não-vazio de uma cadeia irredutível é o próprio V . Deste modo, $C_j = V$ ou $C_j = \emptyset$. Entretanto, como $j \notin C_j$ pois $p_{j,j} = 1$, temos que $C_j = \emptyset$, o que significa que j pode ser alcançado a partir de todos os estados. Como j foi escolhido de forma arbitrária, conclui-se que todos os estados são intercomunicantes.

(\Leftarrow) Assumamos que todos os estados sejam intercomunicantes e que χ seja um conjunto fechado não-vazio de V . Se $j \in \chi$, então para um estado arbitrário $i \in V$ existe um n_i tal que $p_{j,i}^{(n_i)} > 0$. Como o estado i pode ser alcançado a partir do estado $j \in \chi$, segue que $i \in \chi$. Mas i foi escolhido de forma arbitrária em V , assim $\chi = V$. Logo a cadeia é irredutível. \square

Estado Recorrente

Seja $\{X_n\}_{n \geq 0}$ uma cadeia de Markov discreta com espaço de estados V e matriz de transição \mathbf{P} . Dado que o estado inicial da cadeia é i , considere T_i o tempo do primeiro retorno e N_i o número total de retornos. O estado i é denominado **estado recorrente** se

$$P(T_i < \infty | X_0 = i) = 1.$$

Em caso contrário, i é chamado de **estado transiente**.

Período de um estado ou de uma cadeia

Um estado recorrente j é chamado **periódico** de período α se $\alpha \geq 2$ é o maior inteiro positivo tal qual

$$P(T_j = n\alpha \text{ para algum } n \geq 1 | X_0 = j) = 1.$$

Caso contrário, j é dito **aperiódico**.

3.3.1 Matriz de Transição

Definição 3.3.1. Matriz de Transição

Considere (Ω, F, P) um espaço de probabilidade. V um espaço de estados finito e $\{X_n\}_{n \geq 0}$ uma cadeia de Markov. A **matriz de transição** de uma cadeia de Markov é definida como uma matriz $k \times k$ com elementos $p_{i,j}$:

$$P = \begin{bmatrix} p_{1,1} & \cdots & p_{1,k} \\ p_{2,1} & \cdots & p_{2,k} \\ \cdots & \cdots & \cdots \\ p_{k,1} & \cdots & p_{k,k} \end{bmatrix}. \quad (3.1)$$

Como cada número $p_{i,j}$ é uma probabilidade, então $p_{i,j} \geq 0$ e, se somarmos os coeficientes de cada coluna, teremos $\sum_{i=1}^k p_{i,j} = 1$, para $i = 1, 2, \dots, k$.

A matriz que satisfaz as condições anteriores é chamada de **estocástica**.

Matriz de Transição para diversos passos

Anteriormente definimos a matriz de transição para um único passo. Agora, vamos definir a matriz para n -passos conforme [3].

Vamos considerar uma cadeia de Markov arbitrária com k estados possíveis s_1, s_2, \dots, s_k e a matriz \mathbf{P} dada pelo exemplo (3), e assumamos que a cadeia está no estado s_i em um dado tempo n . Desejamos determinar a probabilidade de que a cadeia esteja no estado s_j no tempo $n + 2$, isto é, a probabilidade de sair do estado s_i para o estado s_j em dois passos. Usaremos a notação $p_{i,j}^{(2)}$.

Para $n = 1, 2, \dots$, tomemos X_n o estado da cadeia no tempo n . Então, se s_r é o estado para o qual a cadeia foi migrada no instante $n + 1$, temos:

$$\begin{aligned} p_{i,j}^{(2)} &= P(X_{n+2} = s_j | X_n = s_i) \\ &= \sum_{r=1}^k P(X_{n+1} = s_r \text{ e } X_{n+2} = s_j | X_n = s_i) \\ &= \sum_{r=1}^k P(X_{n+1} = s_r | X_n = s_i) P(X_{n+2} = s_j | X_{n+1} = s_r) \\ &= \sum_{r=1}^k p_{i,r} p_{r,j}. \end{aligned}$$

O valor de $p_{i,j}^{(2)}$ pode ser determinado da seguinte maneira. Se a matriz de transição \mathbf{P} é quadrada, construímos a matriz $P^2 = PP$. O elemento da i -ésima linha e da j -ésima coluna da matriz P^2 será $\sum_{r=1}^k p_{i,r}p_{r,j}$. Entretanto, $p_{i,j}^{(2)}$ será o elemento na i -ésima linha e da j -ésima coluna de P^2 .

Por argumento similar, a probabilidade de que a cadeia vá do estado s_i para o estado s_j em três passos, ou $p_{i,j}^{(3)} = P(X_{n+3} = s_j | X_n = s_i)$, pode ser encontrado construindo a matriz $P^3 = P^2P$. Então a probabilidade $p_{i,j}^{(3)}$ será o elemento na i -ésima linha e na j -ésima coluna da matriz P^3 .

Geralmente, para qualquer valor de m ($m = 2, 3, \dots$), a m -ésima potência P^m da matriz \mathbf{P} definirá a probabilidade $p_{i,j}^{(m)}$ de que a cadeia irá de qualquer estado s_i para qualquer estado s_j em m passos. Por essa razão, a matriz P^m é chamada de *matriz de transição de m -passos* da cadeia de Markov. Simbolicamente podemos escrever

$$p_{i,j}^{(k)} = P(X_{n+k} = s_j | X_n = s_i) \quad \forall n \geq 0; s_i, s_j \geq 0.$$

Temos que $p_{i,j}^1$ é equivalente a $p_{i,j}$ e, usando essa equivalência, as equações de **Chapman-Kolmogorov** nos fornecem um método para encontrar as probabilidades de transição de m -passos. Essas equações são

$$p_{i,j}^{(n+m)} = \sum_{k=0}^{\infty} p_{i,j}^{(n)} p_{i,j}^{(m)}, \quad \forall n, m \geq 0 \forall s_i, s_j \geq 0. \quad (3.2)$$

Tomemos, como exemplo, o seguinte problema: seja A um site de Internet que está disputando um prêmio de melhor blog onde as disputas são feitas por chaves, onde cada blog disputa com outro. Quem ganha vai para a chave da próxima etapa e quem perde fica em outra chave para disputar as outras colocações. Suponhamos que o site A tem probabilidade de 0.4 de ganhar a próxima chave dado que ganhou a atual, e tem probabilidade de 0.2 de ganhar a próxima chave caso perca a atual. Sua matriz de transição para a premiação é mostrada abaixo:

$$A = \begin{bmatrix} 0.4 & 0.6 \\ 0.2 & 0.8 \end{bmatrix}.$$

Agora desejamos saber qual é a probabilidade do site A ganhar uma disputa na próxima chave dado que ele ganhou a disputa nesta chave, isto é, queremos saber $p_{0,0}^{(2)}$ em A^2 . Então

$$A^2 = A \cdot A = \begin{bmatrix} 0.4 & 0.6 \\ 0.2 & 0.8 \end{bmatrix} \cdot \begin{bmatrix} 0.4 & 0.6 \\ 0.2 & 0.8 \end{bmatrix} = \begin{bmatrix} 0.28 & 0.72 \\ 0.24 & 0.76 \end{bmatrix}.$$

Matriz de Transição Regular

Dizemos que uma matriz de transição de Markov A é regular se

- O autovalor 1 é uma raiz simples do polinômio característico $\det(\lambda I - A) = 0$;
- todos os outros autovalores λ de A tem módulo menor do que 1.

Diagramas de Transição

Uma matriz de transição de uma cadeia de Markov pode ser representada graficamente como um **diagrama de probabilidade de transição** onde cada nó representa um estado do sistema, arcos conectam diretamente o estado s_i para o estado s_j se uma transição de um passo entre s_i e s_j é permitida. A probabilidade de transição $p_{i,j}$ é escrita próxima do arco. Note que a transição de um estado para ele mesmo é representado por um *loop*.

Exemplo 1

Suponha que um usuário só possa navegar entre os sites A e B e as condições de navegação nos permitem a seguinte interpretação:

- Probabilidade de no próximo momento o usuário estar no site A dado que agora ele está no site A é de 0.7.
- Probabilidade de no próximo momento o usuário estar no site B dado que agora ele está no site A é de 0.3.

- Probabilidade de no próximo momento o usuário estar no site B dado que agora ele está no site B é de 0.4.
- Probabilidade de no próximo momento o usuário estar no site A dado que agora ele está no site B é de 0.6.

Montando uma tabela com os dados, colocando na coluna o estado “agora” e na linha o estado “no próximo momento” temos:

Condições	Site A	Site B
Site A	0.7	0.6
Site B	0.3	0.4

Tabela 3.1: Tabela referente ao exemplo 1

Da Tabela 3.1 temos o seguinte diagrama de transição:

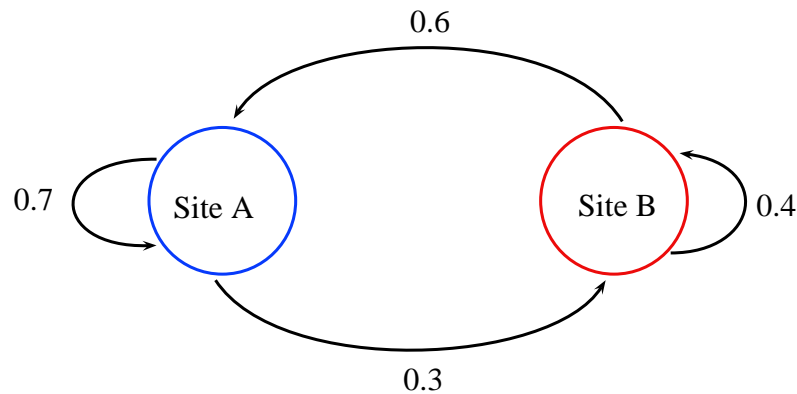


Figura 3.1: Exemplo de diagrama de probabilidade de transição para o exemplo 1.

O diagrama da Figura 3.1 nos fornece a seguinte matriz:

$$P = \begin{bmatrix} 0.7 & 0.6 \\ 0.3 & 0.4 \end{bmatrix}. \quad (3.3)$$

É fácil verificar que essa matriz é uma matriz de transição de Markov.

3.4 Distribuição Estacionária

Dado a construção da matriz de transição, nem sempre é possível que a sequência $\{X_n\}_{n \geq 0}$ convirja para um determinado estado. Porém, sob certas condições, é possível que a sequência estabilize. A existência de uma distribuição limite para X_n , quando $n \rightarrow \infty$, está estritamente ligada a existência da **distribuição estacionária**.

Chamemos o vetor π de **distribuição estacionária** de uma cadeia de Markov se possuir entradas $(\pi(j); j \in V)$ tais que

- $\pi(i) \geq 0$, $\forall i \in V$ e $\sum_{i \in V} \pi(i) = 1$;
- $\pi' = \pi' P$, onde dizemos que $\pi(j) = \sum_i \pi(i) p_{i,j} \forall j \in V$.

Temos que $\pi' P^n = \pi'$, para todo $n \geq 0$, pois,

$$\pi' = \pi' P = \pi' P^1 = (\pi' P) P^1 = \dots = \pi' P^{(n-2)} = (\pi' P) P^{(n-2)} = \pi' P^{(n-1)} = (\pi' P) P^{(n-1)} = \pi' P^n$$

A distribuição estacionária é um **estado de equilíbrio** da cadeia de Markov. Se o estado inicial for escolhido segundo π , então todos os instantes seguintes também terão distribuição π , mostrando que a cadeia é estacionária com o passar do tempo. Neste caso, π é também a distribuição limite de X_n quando $n \rightarrow \infty$, pois

$$P_\pi(X_1 = i) = \sum_{u \in V} P_\pi(X_1 = i | X_0 = u) = \sum_{u \in V} P(X_1 = i | X_0 = u) \pi(u) = \sum_{u \in V} p_{u,i} \pi(u) = \pi(i)$$

Agora, vamos assumir que P é regular, o que significa que existe um $k \geq 1$ tal que $P^k > 0$. Como $(p^k)_{ij} = P(X_{t+k} = i | X_t = j)$, as probabilidades de transição são positivas de um estado para qualquer outro em k passos. Desde que P é regular, pelo teorema 2.2.2, existe uma única distribuição estacionária π que satisfaz $\pi > 0$. O autovalor 1 é simples e dominante, então temos $\lim_{n \rightarrow \infty} p_{i,j}^{(n)} = \pi(j)$, logo temos que a distribuição de uma cadeia de Markov regular sempre irá convergir para uma distribuição estacionária única.

Agora cabe o estudo do Método das Potências para a obtenção da distribuição estacionária. O vetor π também é conhecido como *autovetor de Perron* e a equação $\pi = \pi P$ é equivalente ao sistema de equações

$$\begin{aligned}\pi_1 &= \pi_1 p_{11} + \pi_2 p_{21} + \dots + \pi_n p_{n1} \\ \pi_2 &= \pi_1 p_{12} + \pi_2 p_{22} + \dots + \pi_n p_{n2} \\ &\vdots \\ \pi_n &= \pi_1 p_{1n} + \pi_2 p_{2n} + \dots + \pi_n p_{nn} \\ 1 &= \pi_1 + \pi_2 + \dots + \pi_n.\end{aligned}\tag{3.4}$$

3.5 Método das Potências

O **Método das Potências** consiste de um método iterativo utilizado para obtenção do autovalor dominante¹ e seu correspondente autovetor.

Atribuímos uma aproximação inicial arbitrária para o autovetor correspondente ao autovalor dominante que é sucessivamente melhorada até que a precisão requerida seja encontrada. A convergência para o autovalor dominante é simultaneamente obtida. Os métodos iterativos são mais proveitosos no tratamento de matrizes de grandes dimensões e esparsas, obtendo-se boas estimativas para os seus autovetores.

Método das Potências

Considere A uma matriz de dimensão $n \times n$ com n autovalores $\lambda_i, i = 1, 2, \dots, n$ que satisfazem $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \dots \geq |\lambda_n|$, e n correspondentes autovetores v_j , para $j = 1, \dots, n$, linearmente independentes e normalizados.

Um vetor x_0 pode ser expresso pela combinação linear:

$$x_0 = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n = \sum_{i=1}^n \alpha_i v_i,$$

onde α_i , para $i = 1, \dots, n$, são escalares nem todos iguais a 0. O método iterativo é dado por:

$$x_k = Ax_{k-1}, \quad \text{para } k = 1, 2, 3, \dots\tag{3.5}$$

¹Autovalor dominante é o autovalor com a maior magnitude.

Deste modo,

$$x_k = Ax_{k-1} = A^2x_{k-2} = \dots = A^kx_0 = \sum_{i=1}^n \alpha_i \lambda_i^k v_i. \quad (3.6)$$

Como $\alpha_1, \alpha_2, \dots, \alpha_r$ são diferentes de zero, o termo do lado direito da equação (3.5) é governado pelos termos $\sum_{i=1}^r \alpha_i \lambda_i^k v_i$. Em particular, se $r = 1$, assumimos que $\alpha_1 \neq 0$ e temos

$$x_k = \lambda_1^k \left\{ \alpha_1 v_1 + \sum_{i=2}^n \alpha_i \left(\frac{\lambda_i}{\lambda_1} \right)^k v_i \right\} = \lambda_1^k \{ \alpha_1 v_1 + \epsilon_k \}$$

Para k suficientemente grande, ϵ_k é um vetor com elementos muito pequenos. O vetor x_k é uma aproximação para o autovetor não normalizado v_1 , e é exato se $\|\epsilon_k\|$ for suficientemente pequeno. Esta observação é a base para a simplicidade do cálculo do autovalor dominante no método das potências. Como

$$x_{k+1} = \lambda_1^{k+1} \{ \alpha_1 v_1 + \epsilon_{k+1} \},$$

segue, para qualquer i , que:

$$\frac{(x_{k+1})_i}{(x_k)_i} = \lambda_1 \frac{\alpha_1 (x_1)_i + (\epsilon_{k+1})_i}{\alpha_1 (x_1)_i + (\epsilon_k)_i} \rightarrow \lambda_1,$$

quando $k \rightarrow \infty$, onde $(x_k)_i$ denota a i -ésima componente de x_k .

Velocidade de convergência

A velocidade de convergência do método das potências vai depender das razões $|\frac{\lambda_2}{\lambda_1}|, \dots, |\frac{\lambda_n}{\lambda_1}|$ e, quanto menor essas razões, mais rápida será a convergência. Vale citar que, caso $|\frac{\lambda_2}{\lambda_1}|$ se aproxime de 1, a convergência será muito lenta.

As cadeias de Markov possuem inúmeras aplicações. A próxima seção será sobre uma das mais famosas atualmente, o PageRank.

4 Page Rank

4.1 Importância de um site

Quando criamos um site geralmente colocamos *links* para os sites que serviram de fontes de informações para o nosso, com isso estamos atribuindo valor a esses sites. O algoritmo do PageRank associa a cada página P uma *importância* $I(P)$, onde $I(P)$ é um valor real e positivo. Esta importância é dividida pela quantidade de *links* no *site*, de modo que cada *link* receberá uma parte dessa importância. Suponhamos que um *site* P_j tem l_j *links*, se um desses *links* for para a *site* P_i então P_j passará $1/l_j$ de sua importância para o *site* P_i . O rank de importância de P_i é a soma de todas as contribuições de cada *site* que possui P_i em suas referências (*links*). Se chamarmos de B_i o conjunto de *sites* que referenciam P_i , então podemos calcular a importância de P_i como

$$I(P_i) = \sum_{P_j \in B_i} \frac{I(P_j)}{l_j}. \quad (4.1)$$

Vale salientar que dois *links* com o mesmo *site* de origem e destino serão considerados como um só. Primeiramente, vamos criar uma matriz, chamada de matriz de *hiperlink* $H = [H_{ij}]$, na qual a entrada da i -ésima linha e j -ésima coluna é dada por:

$$H_{ij} = \begin{cases} \frac{1}{l_j} & , \quad \text{se } j \in B_i, \\ 0 & , \quad \text{caso contrário.} \end{cases} \quad (4.2)$$

Notemos que H tem todas as entradas não nulas e a soma de todos os elementos de uma coluna é igual a 1, a menos que a página correspondente não tenha links.

Agora vamos considerar o exemplo citado em [7]. Sejam **A**, **B**, **C**, **D** e **E** cinco *sites* na *web* tais que, se estivermos na página **A**, temos apenas um único link para a página **B** enquanto que, se estivermos na página **C**, temos três links que nos dão a escolha de ir para as páginas **A**, **B** ou **E**. Notemos que há, ao menos, um *link* para cada página como pode ser visto na Figura (4.1).

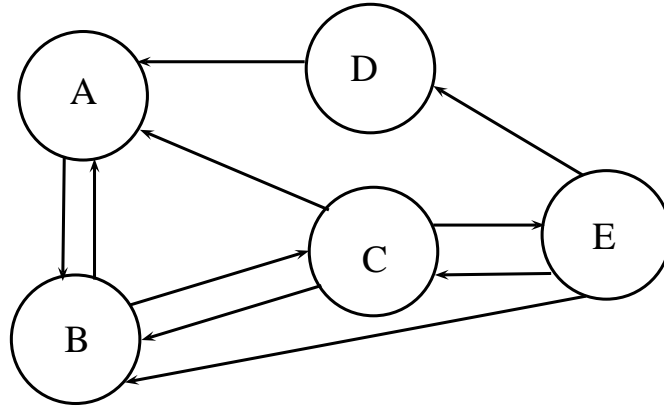


Figura 4.1: Webgrafo do exemplo

Agora vamos fazer um caminho aleatório neste grafo orientado. Iniciando em uma página, a cada passo nós escolhemos aleatoriamente um link desta página que nos leve a outra página e assim por diante. Por exemplo, se iniciarmos na página **B**, podemos ir para **A** ou para **C** com probabilidade $1/2$ para cada caso; se iniciarmos em **D** então, necessariamente, vamos para **A** com probabilidade 1. Portanto não podemos deixar de perguntar “*Onde estaremos daqui a n passos ?*”.

Para automatizar o processo, vamos criar uma matriz **P** onde cada coluna representa a página inicial (onde estamos no momento) e cada linha a página para onde vamos

$$P = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{3} & 1 & 0 \\ 1 & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & \frac{1}{3} & 0 & 0 \end{bmatrix}. \quad (4.3)$$

Consideremos o vetor $I = (I(1), I(2), \dots, I(n))$ tal que as entradas são as importâncias de todas as páginas. Por (4.1), podemos escrever o vetor I como

$$I = P \cdot I.$$

Deste modo, I é o autovetor da matriz **P** correspondente ao autovalor 1. Mas esse autovetor irá se alterar para cada passo que dermos na matriz, logo precisamos estabilizar a matriz **P** e, com isso, achar o autovetor dominante π que é a distribuição estacionária da matriz. Para acharmos o vetor π , usaremos o método das potências apresentado na Seção 3.5. Aplicando o método na matriz **P**, descobrimos que a matriz se estabiliza na potência 32, isto é, quando $n > 32$ em P^n os valores das linhas e colunas não sofrem

alterações para uma precisão de 3 casas decimais, o que nos dá a seguinte matriz

$$P^{32} = \begin{bmatrix} 0.293 & 0.293 & 0.293 & 0.293 & 0.293 \\ 0.390 & 0.390 & 0.390 & 0.390 & 0.390 \\ 0.220 & 0.220 & 0.220 & 0.220 & 0.220 \\ 0.024 & 0.024 & 0.024 & 0.024 & 0.024 \\ 0.073 & 0.073 & 0.073 & 0.073 & 0.073 \end{bmatrix}. \quad (4.4)$$

Da Seção 3.4, sabemos que

$$\pi = P^n \pi.$$

Logo,

$$P^{32} \begin{bmatrix} \pi_0 \\ \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \end{bmatrix} = \begin{bmatrix} \pi_0 \\ \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \end{bmatrix}.$$

Então π_0 está associado a página **A**, π_1 a página **B** e assim por diante. Resolvendo o sistema (3.4) em relação a essa matriz, temos

$$\pi^t = (0.293, 0.390, 0.220, 0.024, 0.073), \quad (4.5)$$

o que nos dá a informação de que a página **B** é a quem tem a maior importância dentre as 5 páginas. Ao utilizarmos o método das potências, no entanto, podemos nos deparar com algumas dificuldades. Para isso, tomemos o exemplo abaixo obtido em [8].

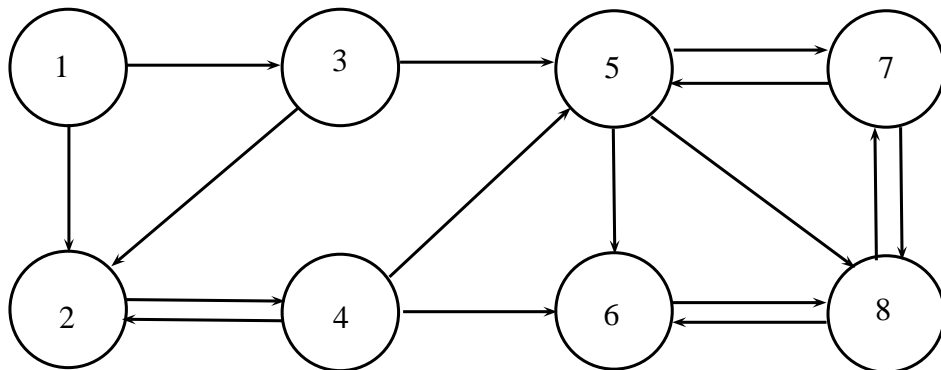


Figura 4.2: Exemplo de falha do método das potências

Para este exemplo temos a seguinte matriz \mathbf{P} :

$$P = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 1/3 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/3 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 1 & 1/2 & 0 \end{bmatrix}.$$

Utilizando o método das potências, obtemos o seguinte vetor estacionário

$$\pi^T = \begin{bmatrix} 0 & 0 & 0 & 0 & 0.12 & 0.24 & 0.24 & 0.4 \end{bmatrix}.$$

Note que as primeiras quatro páginas estão com probabilidade zero no vetor estacionário. Entretanto, isto não é verdade, pois cada uma dessas páginas tem um link oriundo de outra.

O problema com esse exemplo é que existe uma pequena rede *web* dentro dele, que está destacada no desenho abaixo:

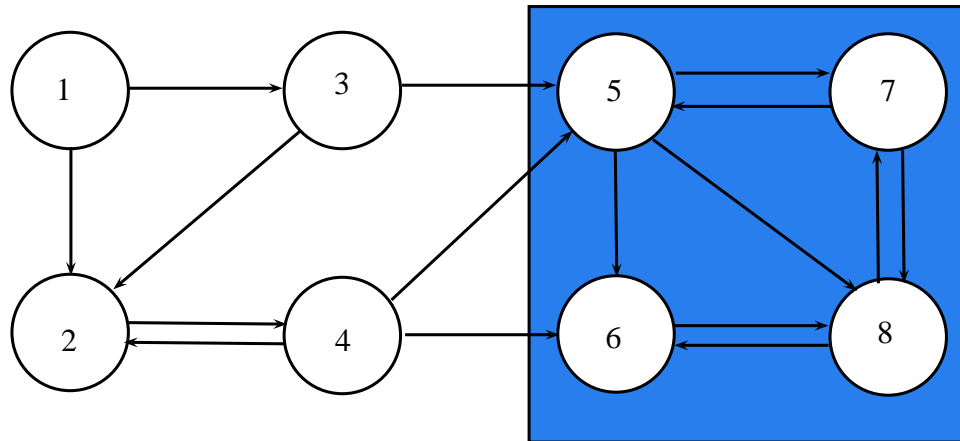


Figura 4.3: Uma pequena rede web dentro de uma rede de links

Os links “entram na área azul” mas “nenhum sai”. Dizemos que essas páginas formam um “ralo de importância” que drena a importância das outras quatro páginas. Isto acontece quando a matriz \mathbf{P} é *reduzível*, isto é, quando pode ser escrita como

$$\mathbf{P} = \begin{bmatrix} * & 0 \\ * & * \end{bmatrix}.$$

De fato, se a matriz \mathbf{P} é irreduzível, então podemos garantir que há um vetor estacionário com todas as entradas positivas. Na próxima seção será apresentado uma modificação para resolvermos esse problema dos links.

4.2 Matriz Google

Para encontrarmos uma matriz que seja regular (ou primitiva) e irreduzível, iremos modificar a forma como o nosso navegador aleatório se movimenta através da web, conforme em [8]. Neste caso, o movimento do navegador aleatório é determinado por \mathbf{S} : ou ele segue um dos links da página atual ou, se a página não tiver links, ele escolhe aleatoriamente qualquer outra página para se mover. Para fazer a nossa modificação, devemos primeiramente escolher um parâmetro $\alpha \in [0, 1]$. Com probabilidade α , ele é guiado por \mathbf{S} , e com probabilidade $1 - \alpha$, ele escolhe a próxima página uniformemente. Se denotarmos por \mathbf{K} a matriz $n \times n$ cujas entradas são todas iguais a 1, obtemos a matriz *Google*:

$$G = \alpha S + (1 - \alpha) \frac{1}{n} K. \quad (4.6)$$

Observemos que \mathbf{G} é estocástica, pois é uma combinação de matrizes estocásticas. Além disso, todas as entradas de \mathbf{G} são positivas, o que implica que \mathbf{G} é regular e irredutível, portanto possui um vetor estacionário único que pode ser obtido pelo método das potências. O parâmetro α é muito importante: se $\alpha = 1$, então $\mathbf{G} = \mathbf{S}$ e trabalharemos com a estrutura de *hyperlinks* original; se $\alpha = 0$, então $G = 1/n$ e, neste caso, consideramos que há um link entre quaisquer duas páginas, perdendo a estrutura original da web.

Obviamente desejamos manter o parâmetro α muito próximo de 1, pois assim teremos a estrutura mais próxima da web original e essa estrutura seria fortemente ponderada nos cálculos computacionais. Entretanto, temos outra consideração. Lembremos que a taxa de convergência do Método das Potências é governada pela magnitude do segundo maior autovalor $|\lambda_2|$. Para a matriz Google, foi provada que a magnitude do segundo autovalor é $|\lambda_2| = \alpha$, sua demonstração se encontra em [12]. Portanto, quando a razão $\frac{|\lambda_2|}{|\lambda_1|}$ está próximo a 1, a convergência do Método das Potências será muito lenta. Com o objetivo de satisfazer esses dois interesses concorrentes, os criadores do PageRank escolheram $\alpha = 0.85$, como informado em [8].

Uma particularidade da matriz G é sua dimensão, com cerca de 25 bilhões de linhas e colunas [8]. Cada entrada do vetor π pode ser interpretada como a importância relativa de cada página, permitindo assim, comparar a importância de quaisquer duas páginas. O vetor π deve sempre ser normalizado para que a soma das popularidades não ultrapasse o valor de 1. Devido ao tamanho da dimensão de G , o Método das Potências pode levar muito tempo para calcular o vetor π . Uma matriz estocástica S pode ser reescrita como

$$S = H + A,$$

onde H é matriz original de hiperlinks e A é uma matriz cujas entradas são zero exceto para as colunas correspondentes aos nós pendentes (páginas que não possuem links para outras páginas), em que cada entrada é $1/n$. Aplicando a equação (4.6), temos:

$$G = \alpha H + \alpha A + \frac{1 - \alpha}{n} K.$$

Portanto, para o vetor de distribuição estacionária π , temos:

$$G\pi = \alpha H\pi + \alpha A\pi + \frac{1 - \alpha}{n} K\pi. \quad (4.7)$$

Lembremos que a maioria das entradas de H são zero. Se considerarmos que cada página possua, no máximo, 10 *links*, então H terá apenas 10 entradas por colunas não-nulas. Portanto, calcular $H\pi$ irá requerer apenas 10 termos não-nulos para cada entrada no vetor resultante. Além disso, as linhas de A são todas idênticas, assim como as linhas de K . Portanto, calculamos os valores de $A\pi$ e $K\pi$ para adicionar aos atuais rankings de importância dos nós pendentes ou de todas as páginas web. Sendo isso feito apenas uma única vez.

Com o valor de α igual a 0.85, o Google diz que necessita apenas de 50 a 100 interações para obter uma boa aproximação do vetor π . Este cálculo leva poucos dias para ser concluído [8].

5 Conclusão

As cadeias de Markov possuem aplicações em diversas áreas do conhecimento humano. Na agricultura, por exemplo, as cadeias de Markov são usadas para analisar os períodos de seca e chuva em determinadas regiões do país [13]. Na área da gestão da informação, apresentamos o famoso algoritmo PageRank, que é capaz de listar em ordem de maior probabilidade os sites que possuem a informação desejada pelo usuário em sua busca. Algumas partes mais sofisticadas do algoritmo continuam a ser desenvolvidas, como a criação de matrizes que reflitam o gosto de navegação do usuário bem como propagandas pagas. O PageRank é uma ideia simples e inteligente que levou os motores de busca na web a um avanço significativo sendo que a sua implementação requer ideias de matemática “elementar”, como álgebra linear e teoria da probabilidade.

Referências Bibliográficas

- [1] Magalhães, Marcos N., *Probabilidade e Variáveis Aleatórias*, Editora da Universidade de São Paulo, 3ª edição, São Paulo, 2011.
- [2] Morgado, Augusto César et al., *Análise Combinatória e Probabilidade*, IMPA, 9ª edição, Rio de Janeiro, 1991.
- [3] DeGroot, Morris H., *Probability and Statistics*, Editora Addison Wesley, 2ª edição, USA, 1986.
- [4] Viali, Lorí., *Algumas considerações sobre a origem da teoria da Probabilidade*, Revista Brasileira de História da Matemática, Volume 8, nº 16, páginas 143-153, Publicação Oficial da Sociedade Brasileira de História da Matemática, ISSN 1519-955X.
- [5] D.L. Isaacson, R.W. Madsen, *Markov Chains Theory and Applications*, Editora John Wiley and Sons, 2ª edição, Florida, 1985.
- [6] Lima, Elon Lages, *Curso de análise volume 2*, IMPA, 11ª edição, Rio de Janeiro, 2012.
- [7] Rousseau, Christiane, *How Google works: Markov chains and eigenvalues*, Klein Project Blog, 2012.
- [8] Austin, David, *How Google find your needle in Web's Haystack*, American Mathematical Society Feature Column, 2006.
- [9] Lima, Elon Lages, *Álgebra Linear*, IMPA, 8ª edição, Rio de Janeiro, 2009.
- [10] Lipschutz, S., Lipson, Marc L., *Álgebra Linear*, Editora Bookman, 4ª edição, Porto Alegre, 2011.
- [11] Gantmacher, Feliks R., *The Theory of Matrices vol. 2*, AMS Chelsea Publishing Company, New York - USA, 1959.
- [12] Melo, Mariana Pereira, *Ordenação das Páginas do Google - "PageRank"*, Instituto de Matemática e Estatística da USP, 1ª edição, São Paulo, 2009.

-
- [13] Antonio R. Santos, et al., *Aplicação da Probabilidade Condicional e do Processo de Cadeia de Markov na Análise da Ocorrência de Períodos Secos e Chuvosos para o Município de Garanhuns, PE, Brasil*, *Ambiente e Água - An Interdisciplinary Journal of Applied Science*, vol. 4, núm. 1, Tabaté, 2009, pp. 169-182.
- [14] Gonçalves, Adilson, *Introdução à Álgebra*, IMPA, 5ª edição, Rio de Janeiro, 2009.