TADEU ALEXANDRE RODRIGUES DOS SANTOS

A MATEMÁTICA POR TRÁS DO GOOGLE

Santo André, 2014

UNIVERSIDADE FEDERAL DO ABC

CENTRO DE MATEMÁTICA, COMPUTAÇÃO E COGNIÇÃO

TADEU ALEXANDRE RODRIGUES DOS SANTOS

A MATEMÁTICA POR TRÁS DO GOOGLE

Orientador: Prof. Dr. Rafael de Mattos Grisi

Dissertação de mestrado apresentada ao Centro de Matemática, Computação e Cognição para obtenção do título de Mestre

ESTE EXEMPLAR CORRESPONDE A VERSÃO FINAL DA DISSERTAÇÃO
DEFENDIDA PELO ALUNO TADEU ALEXANDRE RODRIGUES DOS SANTOS,
E ORIENTADA PELO PROF. DR. RAFAEL DE MATTOS GRISI.

SANTO ANDRÉ, 2014

RESUMO

Neste trabalho apresentamos o algoritmo PageRank, usado pela Google para ordenar páginas no resultado de buscas. No primeiro capítulo descrevemos de maneira detalhada as estruturas matemáticas por trás do algoritmo, apresentando uma interpretação probabilística para suas estruturas e resultados. Para melhor entender a matemática do Google, nos capítulos 2 e 3 trabalhamos conceitos básicos de Cadeias de Markov, em especial a noção de medidas invariantes.

Palavras-chave: PageRank, Google, cadeias de Markov

ABSTRACT

In the present work we present the PageRank algorithm, used by Google to sort the search results on the web. At the first chapter we describe in details the mathematical structures behind the algorithm, providing a probabilistic interpretation for it's structures and results. For a better understanding of Google's math, in chapters 2 and 3 we work on some basic concepts of Markov Chains, specially the notion of invariant measures.

Keywords: PageRank, Google, Markov chains

CONTEÚDO

INTRODUÇÃO

Estamos vivendo e acompanhando uma revolução nos modos de pensar e agir das pessoas em todo o mundo e a precursora deste fato é incontestavelmente a internet. Ela tem o poder de alcançar todos os cantos do mundo, levando todo o tipo de informação e formação, aproximando pessoas e permitindo o contato com novas e diferentes culturas, quer dizer, a internet abre um leque infinito de possibilidades e, tudo isso, sem a necessidade de se levantar do sofá de casa, bastando, para isso, ter um computador conectado a web.

Com a internet qualquer pessoa do mundo pode criar uma página e expressar suas opiniões, ideias, devaneios e, além disso, divulgar, oferecer e vender uma infinidade de produtos e serviços com um enorme alcance. Agora, a grande força da internet vem do fato que qualquer pessoa conectada em qualquer lugar do mundo pode ter acesso a essa página e compartilhar do conteúdo dela, chegando em alguns casos, até a dar suas contribuições para a página acessada. Enfim, a mágica da internet está na inimaginável gama de possibilidades de interações entre as pessoas em todas as partes do mundo e a internet é o veículo para que isso ocorra. Porém, as facilidades que a grande rede oferece às pessoas gera um enorme imbróglio devido ao imenso número de páginas - da ordem de bilhões - que edificam a rede e que são criadas a todo instante e em vários lugares do mundo.

Diante disto, o presente trabalho de conclusão de curso visa investigar a matemática por trás do Google, o maior e melhor site de buscas da internet. O Google utiliza um método desenvolvido por Larry Page e Sergey Brin na Universidade de Stanford e conhecido como PageRank, que é um sistema para dar notas a páginas na web. Mesmo hoje, com todos os avanços na área de tecnologia, o algoritmo do PageRank ainda continua a ser o âmago das buscas no Google.

O PageRank usa uma enorme cadeia de hyperlinks como um indicador do valor de uma página, ou seja, o Google interpreta um link da página a para a página b como uma espécie de voto de a para b, contudo o Google vai além da quantidade de votos, ou links, que uma página recebe, ele analisa também a página que dá o voto. Os votos dados por páginas com maiores PageRanks pesam mais do que os dados por outras

com menores PageRanks. Pois bem, quando um internauta digita uma palavra ou frase no Google, a ordem dos resultados para essa busca obedecem ao PageRank que o Google atribui às páginas na internet.

No primeiro capítulo deste trabalho falaremos do ordenamento das páginas e dos problemas que surgem na determinação do PageRank de algumas destas, provendo um olhar probabilístico ao algoritmo do Google. Já no capítulo dois utilizaremos a Cadeia de Markov para investigar, um pouco mais de perto, as características do vetor estacionário de uma matriz de transição, com o intuito de entender um pouco melhor as características e interpretações do PageRank. Finalizamos com o capitulo três que trata do acoplamento de Cadeias de Markov, uma técnica simples mas poderosa, que nos permite entender como ocorre a convergência do método iterativo utilizado no algoritmo do PageRank.

1

O ALGORITMO DE ORDENAMENTO DO GOOGLE

De acordo com o site worldwidewebsize.com, a web é formada por dezenas de bilhões de páginas que se multiplicam a cada momento, devido a facilidade e praticidade de suas criações. Dados indicam que, aproximadamente, 95% dos textos em páginas da web são compostos por cerca de 10 mil palavras, ou seja, quando o usuário digita o objeto de sua pesquisa, haverá um grande número de páginas que contém, em seus conteúdos, as palavras que estão sendo buscadas. Portanto, é necessário uma forma de classificar a relevância das páginas da web que se enquadram nos critérios de pesquisas, pois, assim, nos resultados das buscas, as páginas classificadas com o maior grau de relevância aparecerão no topo da lista.

Cada página da web trata de um ou vários assuntos e, geralmente, quando um internauta acessa um site é devido a algum tipo de interesse que tem por este. Quando as páginas são construídas é natural que sejam incluídos caminhos (links) para outras páginas que, de modo geral, apresentam conteúdos adjacentes de interesse.

Na inclusão de links, o criador de um site, de certa forma, está assumindo que esse link é importante e nele existem informações valiosas e confiáveis que possivelmente podem ser úteis ao internauta . Por exemplo, um site de automóveis, possivelmente, disponibilizará links para outros sites que giram em torno do assunto automóveis.

O cerne da ideia dos criadores do PageRank, apresentado em [4] por Sergey Brin e Lawrence Page, é atribuir maior importância às páginas mais visitadas. Assim, se dentre duas páginas s_1 e s_2 que tratam do mesmo assunto, a página s_1 recebe mais acessos, então esta deveria aparecer na frente em resultados de busca.

Mas como fazer esta medida? Mesmo que toda página tivesse uma espécie de contador de visitas, o google poderia não ter acesso a estas informações. Além disso, para ter uma medida confiável seria necessário controlar a quanto tempo cada contador está no ar, dentre outros problemas de difícil controle.

Para contornar estes problemas, Brin e Page decidiram olhar para os links que apontavam para cada página. A ideia é que uma página com maior quantidade de links apontados para ela, seria mais importante. A medida feita desta forma claramente não é boa, pois uma página com muitos links de páginas pouco importantes poderiam ter menor importância do que uma página com menos links provenientes de páginas de maior importância. A seguir descrevemos a solução implementada por Brin e Page.

Seguiremos muito de perto a descrição dada em [1], mas diversos artigos e livros foram escritos a respeito. O leitor interessado pode procurar mais informações em [5, 7, 8], ou no próprio Google!

1.1 O PAGERANK DE UMA PÁGINA

Para corrigir o problema a ideia implementada foi a de “transferência de importância". Cada página “transfere” sua importância de maneira equânime a todas as páginas para as quais possui um link. Do mesmo modo, ele “receberá” peso das páginas que possuem links apontadas para ela.

Para determinar então a medida de importância de uma página web (ou PageRank) $r(s)$, suponha que a página s_j tenha l_j links, sendo um desses links a página s_i, dessa forma a razão que s_j vai passar de sua importância para s_i será de $1/l_j$. Com isso, o PageRank de s_i será a soma de todas as contribuições feitas por páginas da web que têm links de acesso para s_i . Nesse sentido, podemos chamar de B_i o conjunto de páginas que possuem links para s_i, então para cada $i \in \{1, \ldots, N\}$

$$r(s_i) = \sum_{s_j \in B_i} \frac{r(s_j)}{l_j}, \tag{1.1}$$

onde N é o total de páginas da web.

Analisando a equação (1.1), o leitor menos experiente pode a princípio pensar que estamos em um paradoxo, pois para determinarmos o PageRank de s_i, temos de saber qual o PageRank de todas as páginas que se ligam a ela. Porém, o que temos é apenas um sistema de equações (uma equação para cada página da web), e tudo o que (1.1) nos diz é que as importâncias das diversas páginas dever satisfazer as equações deste sistema.

Para simplificar a notação para cada $i \in \{1, \ldots, N\}$ chamaremos $r(s_i) := r_i$.

Para melhor representar o sistema acima, considere a matriz $H = [H_{ij}]$, que chamaremos de matriz hyperlink dada por

$$H_{ij} = \begin{cases} 1/l_j, & \text{se } s_j \in B_i \\ 0, & \text{se } s_j \text{ não tem links.} \end{cases} \tag{1.2}$$

Definindo agora o vetor coluna $\mathbf{r} = [r_i]$, no qual as componentes serão os PageRanks das páginas web, vale que

$$\mathbf{r} = H\mathbf{r}.$$

O vetor $\mathbf{r}$ assim definido é conhecido como *vetor estacionário* da matriz H.

No exemplo que segue, vamos encontrar a matriz de hyperlinks H de uma rede formada por 5 páginas da web, com os links representados por flechas, representada na figura a seguir.

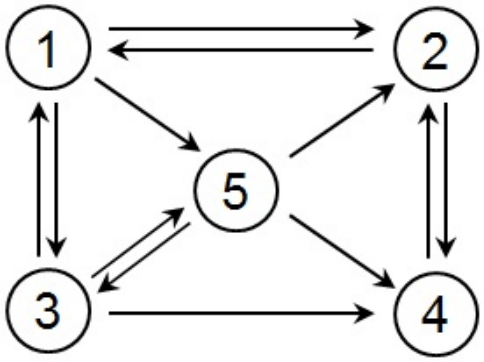


Figura 1: Rede considerada no exemplo 1.1

.

Exemplo 1.1. Considere a rede de páginas na Figura 1. Neste caso a matriz de hyperlink da rede será

$$H = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{3} & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 1 & \frac{1}{3} \\ \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 \end{bmatrix}.$$

O vetor estacionário deve satisfazer portanto

$$\begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{3} & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 1 & \frac{1}{3} \\ \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \end{bmatrix}.$$

Igualando os dois lados da equação acima linha por linha, obtemos

$$\begin{cases} \frac{1}{2}r_2 + \frac{1}{3}r_3 & = r_1 \\ \frac{1}{3}r_1 + r_4 + \frac{1}{3}r_5 & = r_2 \\ \frac{1}{3}r_1 + \frac{1}{3}r_5 & = r_3 \\ \frac{1}{2}r_2 + \frac{1}{3}r_3 + \frac{1}{3}r_5 & = r_4 \\ \frac{1}{3}r_1 + \frac{1}{3}r_3 & = r_5 \end{cases}$$

ou ainda

$$\begin{cases} -r_1 + \frac{1}{2}r_2 + \frac{1}{3}r_3 & = 0 \\ \frac{1}{3}r_1 - r_2 + r_4 + \frac{1}{3}r_5 & = 0 \\ \frac{1}{3}r_1 - r_3 + \frac{1}{3}r_5 & = 0 \\ \frac{1}{2}r_2 + \frac{1}{3}r_3 - r_4 + \frac{1}{3}r_5 & = 0 \\ \frac{1}{3}r_1 + \frac{1}{3}r_3 - r_5 & = 0 \end{cases}.$$

Se tentarmos resolver o sistema acima, perceberemos que existem infinitas soluções possíveis. Por razões que ficarão claras mais a frente, tomaremos uma solução com todos os termos positivos, e cuja soma total dos pesos é 1. Ou seja, queremos $r_i \geq 0$ para todo i e $r_1 + \cdots + r_N = 1$. Com isso, neste exemplo temos

$$\mathbf{r} = \begin{bmatrix} \frac{6}{29} \\ \frac{10}{29} \\ \frac{3}{29} \\ \frac{7}{29} \\ \frac{3}{29} \end{bmatrix} = \begin{bmatrix} 0,2069 \\ 0,3448 \\ 0,1034 \\ 0,2414 \\ 0,1034 \end{bmatrix}. \tag{1.3}$$

Assim, de acordo com o vetor $\mathbf{r}$, a página 2 é a mais importante, e aparecerá na frente das demais no resultado de uma busca. Mais do que isso, se uma busca retornar as páginas 2, 3 e 4 sua ordenação será r_2, r_4, r_3.

1.2 CALCULANDO O PAGERANK

Vimos no exemplo 1.1 que para encontramos as entradas do vetor **r**, basta resolver um sistema linear em que o somatório das componentes de **r** seja igual a 1. Porém, pensando na web, isso se torna um grande desafio, pois a matriz H é uma matriz quadrada onde cada coluna corresponde a uma página da web indexada pelo Google, assim sendo, H teria cerca de $N = 25$ bilhões de colunas e linhas. É fato que a maioria das entradas de H são iguais a 0, isso por que o total de links em uma página é normalmente muito menor que o total de páginas na web. Apesar disso simplificar a resolução do sistema, como veremos mais a frente causa problemas sérios na ordenação das páginas, o que nos forçará a alterar a matriz H, fazendo com que esta fique menos esparsa. Precisamos ter em mãos um método diferente de solução.

Retomando ao exemplo da nossa rede com 5 páginas, vamos utilizar um método iterativo para encontrarmos o vetor **r**. Primeiro escolhemos um vetor qualquer $\mathbf{r}^0$ como um candidato para **r** e, em seguida, produzimos uma sequência de vetores $\mathbf{r}^k$, definido para cada $k \geq 0$

$$por \mathbf{r}^{k+1} = H\mathbf{r}^k.$$

A ideia agora é que se $\mathbf{r}^k \sim \mathbf{v}$, para k suficientemente grande, então teremos $\mathbf{r}^{k+1} \sim \mathbf{v}$ e

$$\mathbf{v} = H\mathbf{v},$$

e portanto $\mathbf{v} = \mathbf{r}$.

Para ilustrar o método acima vamos estudar o caso em que

$$\mathbf{r}^0 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Vale ressaltar que poderíamos escolher outros valores para $\mathbf{r}^0$, seguindo apenas a regra estabelecida anteriormente de que as entradas devem ser não-negativas e somarem 1. Assim, teríamos os seguintes resultados:

Comparando com a solução encontrada em (1.3), vemos que estas coincidem com pelo menos 4 casas decimais de precisão.

$\mathbf{r}^0$	$\mathbf{r}^1$	$\mathbf{r}^2$	$\mathbf{r}^3$	...	$\mathbf{r}^{60}$
1	0	0,2778	0,0926	...	0,2069
0	0,3333	0,1111	0,5185	...	0,3448
0	0,3333	0,1111	0,1296	...	0,1034
0	0	0,3889	0,1296	...	0,2414
0	0,333	0,1111	0,1296	...	0,1034

Tabela 1: Sequência de vetores $\mathbf{r}^k$ com $\mathbf{r}^0 = (1, 0, 0, 0, 0)$.

Observe que as entradas dos vetores $\mathbf{r}^k$ são sempre não-negativas, e somam 1. Mais a frente veremos por que isso acontece.

A tabela 1 mostra uma sequência em que escolhido $\mathbf{r}^0$ e feita algumas iterações, convergimos naturalmente para o vetor estacionário $\mathbf{r}$ da matriz H. Neste exemplo, esta convergência aconteceria mesmo se escolhêssemos um outro valor para $\mathbf{r}^0$, por exemplo, se tomarmos

$$\mathbf{r}^0 = \begin{bmatrix} 0,25 \\ 0 \\ 0,25 \\ 0,25 \\ 0,25 \end{bmatrix},$$

então

$\mathbf{r}^0$	$\mathbf{r}^1$	$\mathbf{r}^2$	$\mathbf{r}^3$	...	$\mathbf{r}^{60}$
0,25	0,0833	0,2639	0,1528	...	0,2069
0	0,4167	0,2500	0,4352	...	0,3448
0,25	0,1667	0,0833	0,1157	...	0,1034
0,25	0,1667	0,3194	0,1806	...	0,2414
0,25	0,1667	0,0833	0,1157	...	0,1034

Tabela 2: Sequência de vetores $\mathbf{r}^k$ com $\mathbf{r}^0 = (1/4, 0, 1/4, 1/4, 1/4)$.

1.3 CASOS PROBLEMÁTICOS PARA A MATRIZ DE HYPERLINKS

Observando os exemplos acima, somos levados a fazer três questionamentos:

- O vetor $\mathbf{r}^k$ sempre converge?
- A convergência para o vetor $\mathbf{r}$ depende da escolha de $\mathbf{r}^0$?
- Os valores obtidos desta forma para os PageRanks sempre fornecem as informações que queremos?

Infelizmente a resposta para as três perguntas é não! A boa notícia é que se fizermos algumas modificações na matriz H teremos resposta afirmativa para as três perguntas.

1.3.1 *Páginas sem links*

É bastante comum enquanto navegamos pela web, que os alvos de nossas buscas sejam arquivos PDF, imagens ou mesmo páginas web sem links para nenhuma outra página. Estas páginas ou arquivos existem em grande quantidade, e fazem parte da rede. Infelizmente elas são também uma fonte de dor de cabeça para o PageRank.

Considere a seguinte rede com três páginas:

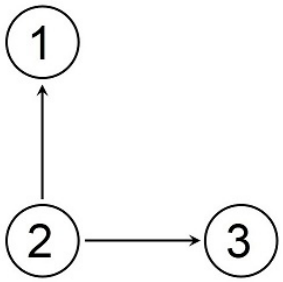


Figura 2: Rede com páginas sem link.

A matriz hyperlink correspondente a figura 2 é

$$H = \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{bmatrix}.$$

Com isso, aplicando o método iterativo, obtemos

$\mathbf{r}^0$	$\mathbf{r}^1$	$\mathbf{r}^2$	$\mathbf{r}^3 = \mathbf{r}$
0	0,5	0	0
1	0	0	0
0	0,5	0	0

Tabela 3: Sequência de vetores $\mathbf{r}^k$ para um rede com páginas sem link.

Neste exemplo, as componentes do vetor estacionário de H são iguais a zero e, portanto, não é possível estabelecer um PageRank para as três páginas. Isso ocorre devido as páginas 1 e 3 não possuírem ligações para outras páginas, ou seja, as páginas 1 e 3 tiram o grau de importância da página 2 mas não repassam para outras, e nem permanecem com ele. Páginas que não possuem links são chamadas de *dangling nodes* e na web existem milhares delas. Lidaremos com este problema modificando a matriz H, mas deixaremos isso para mais tarde, depois de entendermos os demais problemas relacionados a matriz de hyperlinks.

1.3.2 *Ciclos de Páginas*

O próximo problema é causado por páginas cujas sequências de links foram um ciclo. Podemos imaginar um site de formulários, que após passar por todas as páginas do formulário, volta à página inicial para novo preenchimento. Neste tipo de estrutura toda a importância de um site é passada para o site seguinte, só retornando para a primeira página depois de percorrer o ciclo, instante no qual recomeça o processo.

Para entender melhor este fenômeno, considere uma rede com as seguintes ligações de páginas

A matriz hyperlink correspondente a figura 3 é

$$H = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

de onde segue

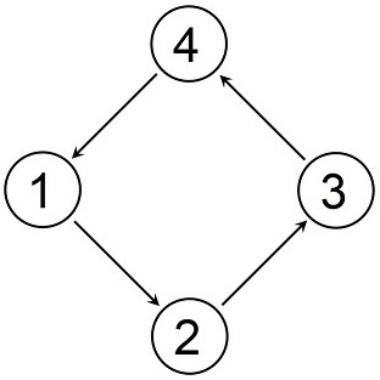


Figura 3: Páginas em ciclo.

$\mathbf{r}^0$	$\mathbf{r}^1$	$\mathbf{r}^2$	$\mathbf{r}^3$	$\mathbf{r}^4$	$\mathbf{r}^5$	$\mathbf{r}^6$	$\mathbf{r}^7$	$\mathbf{r}^8$
1	0	0	0	1	0	0	0	1
0	1	0	0	0	1	0	0	0
0	0	1	0	0	0	1	0	0
0	0	0	1	0	0	0	1	0

Tabela 4: Sequência de vetores $\mathbf{r}^k$ para uma rede circular.

O exemplo acima ilustra claramente como ocorre a transferência de importância, que faz com que vetores $\mathbf{r}^k$ se comportem de forma periódica. Isso impede que o método iterativo convirja para algum vetor estacionário $\mathbf{r}$. É importante observar que este método não necessariamente atrapalha na existência de um vetor estacionário. A princípio tal vetor poderia ainda existir, mas a sequência $\mathbf{r}^k$ pode não convergir para ele.

1.3.3 *Conjuntos de Páginas Auto-referenciadas*

Suponha que um grupo de cervejeiros caseiros resolve entrar no mercado de cervejas artesanais, e para isso abre uma escola onde ensinarão como se fazer a bebida em casa. Para divulgar a escola eles criam um site simples, com a informações dos cursos e contato. Como este é apenas o primeiro site, eles não colocam nenhum link para sites externos à escola, de modo que todas as páginas do site possuem link apenas para outras páginas do mesmo site. Para divulgar a escola eles conseguem colocar propagandas em alguns sites especializados, com um link que leva para o site da escola.

Está criada assim uma situação bastante comum na internet: conjunto de páginas auto-referenciadas, com ligações vindas de páginas externas.

A rede a seguir exemplifica tal situação.

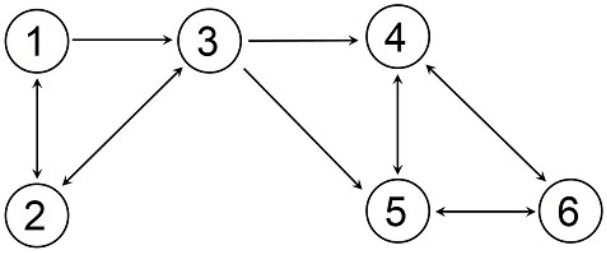


Figura 4: Rede com grupo de páginas auto-referenciadas.

Analisando a rede da figura 4, encontramos a matriz hyperlink e vetor estacionário dados respectivamente por

$$H = \begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \quad \text{e} \quad \mathbf{r} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}.$$

Escolhendo um vetor $\mathbf{r}^0$ como nos exemplos anteriores, e fazendo algumas iterações encontramos

$\mathbf{r}^0$	$\mathbf{r}^1$	$\mathbf{r}^2$	$\mathbf{r}^3$	...	$\mathbf{r}^{40}$
1	0	0,25	0,0833	...	0
0	0,5	0,1667	0,2083	...	0
0	0,5	0,25	0,2083	...	0
0	0	0,1667	0,1667	...	0,3333
0	0	0,1667	0,1667	...	0,3333
0	0	0	0,1667	...	0,3333

Tabela 5: Sequência de vetores $\mathbf{r}^k$ para uma rede com absorção.

Olhando a sequência vemos que a importância dos sites 1, 2 e 3 são lentamente absorvidas pelas demais páginas. Isso por que existem links de 3 para 4 e 5, não existem links das páginas 4, 5 e 6 para as demais, como salientado na figura 5.

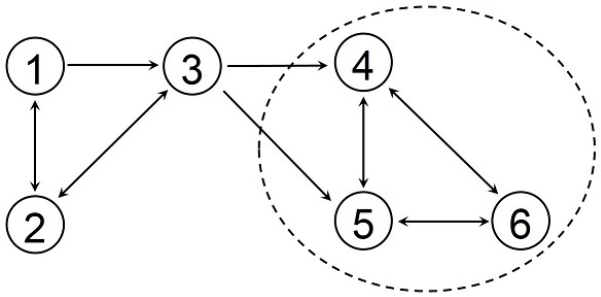


Figura 5: Rede com grupo de páginas auto-referenciadas.

Assim, quando o navegador chegar à rede em destaque, ele não encontrará links que liguem o segundo bloco ao primeiro, permanecendo para sempre alí dentro. Ligações desse tipo são dissipadoras de importância, pois drenam o grau de importância de outras três páginas. Isso não impede a convergência do método, e tampouco a existência do vetor estacionário. O principal problema está em atribuir PageRank 0 a diversas páginas, que poderiam ser inclusive mais relevantes que as páginas do grupo dissipador. Mas mesmo que não fosse, não há interesse em colocar nível de importância 0 a nenhuma página. Toda página tem seu nível de importância, mesmo que dentro de parâmetros de busca muito específicos. Isso significa que, mesmo que encontremos um vetor estacionário **r**, este não carrega consigo as informações que buscamos.

Vistos os principais problemas existentes na web, e que atrapalham o funcionamento perfeito do algoritmo de PageRank, está no hora de entendermos como resolvê-los. Mas para isso, vamos primeiro introduzir uma forma diferente de olharmos para o processo.

1.4 INTERPRETAÇÃO PROBABILÍSTICA DO ALGORITMO DE PAGERANK

Como vimos na seção anterior, a matrix de hyperlinks, se usada sozinha, possui problemas que podem inviabilizar o cálculo do vetor estacionário, falhando assim em determinar o PageRank de cada página. Por isso será necessário que façamos algumas mudanças no modelo, mas sem perder de vista nosso objetivo principal: atribuir a cada página da web um índice de importância. Para melhor justificar as mudanças que

faremos, vamos antes propor uma nova interpretação do modelo. Uma nova forma de olhar o processo, que nos permitirá justificar as mudanças a serem feitas.

Se acompanharmos de perto a navegação de um internauta fixo, muito provavelmente identificaremos padrões para sua navegação, e dificilmente poderemos dizer que suas escolhas de páginas são fruto do acaso. Mas os padrões de navegação dos diversos internautas que hoje se utilizam da web são muito variados, e assim ao modelar o comportamento de um internauta aleatoriamente escolhido, é plausível trabalharmos com a hipótese de que sua navegação é de fato aleatória. Em outras palavras, podemos supor que a cada passo da sua navegação o internauta sai de uma página, escolhendo aleatoriamente dentre os links disponíveis nesta. Outra forma de colocarmos tal hipótese é supor que, dado a diversidade de padrões de navegação na internet, é pouco razoável atribuirmos importância maior a um link com prejuízo aos demais links na mesma página, de modo que a proporção de cliques em um link deve ser igual para todos os links da página, isto é, o inverso do total de links na página.

Suponhamos então que um internauta esteja navegando pela web de forma aleatória, ou seja, uma vez que ele chega ao site s_j, que possui uma quantidade l_j de links para outras páginas, ele escolhe um dos links com probabilidade $1/l_j$, e segue para onde este link o envia. Assim, a probabilidade deste internauta ir de s_j para s_i é de $/l_j$, se existir um link de s_j para s_i, e 0 caso contrário.

Desta forma, a entrada H_{ij} da matriz H poderia ser interpretada como a probabilidade do internauta ir da página s_j para a página s_i. Como veremos mais a frente, esta interpretação da matriz H ainda possui um pequeno problema, mas que pode ser resolvido de forma bastante simples.

Antes de passarmos para a análise dos problemas listados na seção anterior, vamos antes interpretar o PageRank sob este novo prisma. Para isso, considere a rede descrita no exemplo 1.1, representada pela figura 6.

Chame de t_1, t_2, t_3, t_4 e t_5 a proporção média de tempo que o internauta passa nos sites s_1, s_2, s_3, s_4 e s_5 respectivamente. Vamos nos concentrar por um momento no site s_1. Pelo modelo descrito, toda visita feita ao site s_1 só pode ter sido originada do site s_2 ou s_3. Deste modo, o total de visitas ao site s_1 será dado pelo total de visitas a s_2 que seguiram para s_1 somada ao total de visitas a s_3 que seguiram para s_1, e a mesma argumentação vale para a proporção de tempo passada em cada site. Como a probabilidade do internauta ir de s_2 para s_1 é $1/2$, então aproximadamente metade

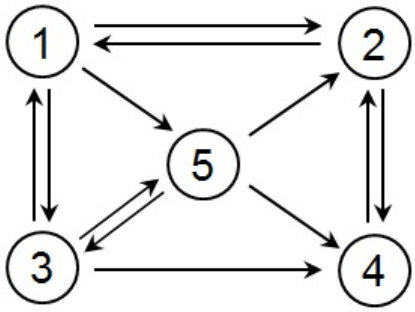


Figura 6: Exemplo de rede

das visitas a s_2 seguiram para s_1, e analogamente $1/3$ das visitas a s_3 seguiram para s_1 (note que s_3 tem 3 links). Isso tudo nos mostra que

$$t_1 = \frac{1}{2}t_2 + \frac{1}{3}t_3.$$

Fazendo o mesmo raciocínio para as demais páginas da rede, concluímos que

$$\begin{cases} \frac{1}{2}t_2 + \frac{1}{3}t_3 & = t_1 \\ \frac{1}{3}t_1 + t_4 + \frac{1}{3}t_5 & = t_2 \\ \frac{1}{3}t_1 + \frac{1}{3}t_5 & = t_3 \\ \frac{1}{2}t_2 + \frac{1}{3}t_3 + \frac{1}{3}t_5 & = t_4 \\ \frac{1}{3}t_1 + \frac{1}{3}t_3 & = t_5. \end{cases}$$

Observe agora que os valores de t_1, t_2, t_3, t_4 e t_5 obedecem ao mesmo sistema de equações que r_1, r_2, r_3, r_4 e r_5. Além disso, é interessante notar que a soma dos tempos t_1, t_2, t_3, t_4 e t_5 deve ser 1, uma vez que representa a proporção total de tempo que o internauta passou navegando.

Deste modo, nesta visão probabilística do modelo, o PageRank de cada página pode ser visto como a proporção de tempo que um internauta aleatório passa naquela página. Esta é uma visão absolutamente compatível com a noção de importância de uma página. Ou seja, uma página é mais importante que outras se um internauta aleatório passa mais tempo nela do que nas demais.

Este tipo de modelo é conhecido como *Cadeia de Markov*, que será melhor estudada nos próximos capítulos. No capítulo 2 veremos uma interpretação para os vetores $\mathbf{r}^k$, usados no cálculo de $\mathbf{r}$, o que nos levará a uma interpretação distinta (mas equivalente) do vetor estacionário $\mathbf{r}$.

1.4.1 *Alterando o modelo*

Como comentamos anteriormente, o uso apenas da matriz H pode causar problemas na determinação do PageRank, dependendo de características das conexões da rede. A seguir vamos abordar estes problemas, e ver como solucioná-los usando a interpretação probabilíssima que acabamos de estudar.

O primeiro problema que descrevemos é causado por páginas que não possuem link, ou seja, as chamadas *dangling nodes*. De fato, veja que a coluna correspondente a um dangling node é de certa forma incompatível com o passeio aleatório do internauta. Isso por que toda entrada desta coluna é nula, o que não nos permite interpretá-la da mesma forma que as demais. Note que uma vez que o internauta, em seu caminho aleatório, atinge um destes sites ele não sabe o que fazer. A probabilidade de ir para outro site, até este momento, é nula. Mas isso significaria que a probabilidade de permanecer neste site é 1. Isto por que não queremos que o passeio acabe.

Poderíamos então modificar a matriz H colocando o valor 1 na entrada H_{ii} sempre que s_i for um dangling node. Mas isso causa um outro problema. Note que agora temos uma *subrede* formada por apenas um site, que possui link apenas para si mesmo. Entramos assim no mesmo problema das redes auto-referenciadas discutidas anteriormente. Como vimos, este tipo de rede *rouba* toda a importância das demais páginas.

Para resolver esta questão imaginemos que uma vez que o internauta acabe de navegar uma *dangling node* ele volta a navegar, escolhendo uma página qualquer da rede de forma aleatória. Isso é equivalente a dizer que cada dangling node possui links para todas as outras páginas (incluindo ela mesma).

Com isso modificamos a matriz Hyperlink H, substituindo uma coluna de zeros correspondentes a uma *dangling nodes* por uma coluna na qual cada entrada é $/N$, onde N é o total de páginas na web. Desta forma, se modificarmos a matriz H, que representa o exemplo da subseção 1.3.1, e chamarmos essa nova matriz de S, ficaríamos com

$$S = \begin{bmatrix} \frac{1}{3} & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{3} \end{bmatrix}.$$

Esta modificação altera também o vetor estacionário, que agora deve responder à relação

$$\mathbf{r} = S\mathbf{r}.$$

Segue assim que

$$\mathbf{r} = \begin{bmatrix} \frac{3}{8} \\ \frac{1}{4} \\ \frac{3}{8} \end{bmatrix}.$$

Podemos pensar também na matriz S como sendo uma soma da matriz H com uma matriz que vamos chamar de A que representaria as páginas sem links, isto é, a matriz A usa o artifício de considerar cada *dangling nodes* com tendo uma saída para todas as outras páginas da rede. Assim, o exemplo da subseção 1.3.1 fica

$$H = \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{bmatrix} \quad \text{e} \quad A = \begin{bmatrix} \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} \end{bmatrix}.$$

Então

$$S = H + A = \begin{bmatrix} \frac{1}{3} & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{3} \end{bmatrix}.$$

Note que a matriz S possui agora entradas não-negativas, e a soma de todas as entradas de qualquer coluna é 1. Uma matriz com estas característica é chamada de *matriz de transição de probabilidades*, e será estudada em mais detalhes nos próximos capítulos.

Infelizmente esta modificação não resolve os problemas causados pelas páginas em ciclo e pelas redes auto-referenciadas. Para resolver estes problemas, analisemos o modelo um pouco mais de perto.

Do jeito que descrevemos o problema, a única forma do internauta chegar a uma certa página é através de um link em alguma outra página. Mas isto nem sempre é verdade. Frequentemente, ao navegarmos pela web, somos provocados por notícias, imagens, artigos ou mesmo por pura curiosidade, a buscar informações em locais na web para os quais a página que nos encontramos não possui links.

É razoável então pensarmos na seguinte modificação. Cada vez que nosso internauta estiver um uma dada página, com probabilidade α ele escolhe permanecer no padrão de navegação descrito por S, e com probabilidade $(1 - \alpha)$ ele pula para uma página escolhida aleatoriamente, com probabilidade $1/N$ para cada página.

Assim, chamando de G a matriz utilizada pelo Google, temos

$$G = \alpha S + (1-\alpha)\frac{1}{N}U, \tag{1.4}$$

onde S é a matriz de hiperlinks modificada, N é o número de páginas na web e U é a matriz $N \times N$ com todas as entradas iguais a 1.

Com isso o PageRank de cada página é dado agora pelo vetor estacionário $\mathbf{r}$, com entradas positivas somando 1, que atende ao sistema

$$\mathbf{r} = G\mathbf{r}. \tag{1.5}$$

É interessante notar que agora a matriz G possui todas as entradas positivas. Além disso G é também uma matriz de transição.

Para exemplificar, vamos calcular a matriz G para a rede descrita no problema de redes auto-referenciadas, cuja figura copiamos abaixo. Para isso consideraremos $\alpha = 3/4$.

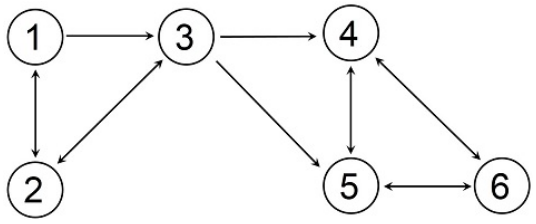


Figura 7: Rede com problemas.

Temos

$$G = \frac{3}{4}\begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} + \frac{1}{4}\cdot\frac{1}{6}\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix},$$

e portanto

$$G = \begin{bmatrix} \frac{1}{24} & \frac{5}{12} & \frac{1}{24} & \frac{1}{24} & \frac{1}{24} & \frac{1}{24} \\ \frac{5}{12} & \frac{1}{24} & \frac{7}{24} & \frac{1}{24} & \frac{1}{24} & \frac{1}{24} \\ \frac{5}{12} & \frac{5}{12} & \frac{1}{24} & \frac{1}{24} & \frac{1}{24} & \frac{1}{24} \\ \frac{1}{24} & \frac{1}{24} & \frac{7}{24} & \frac{1}{24} & \frac{5}{12} & \frac{5}{12} \\ \frac{1}{24} & \frac{1}{24} & \frac{7}{24} & \frac{5}{12} & \frac{1}{24} & \frac{5}{12} \\ \frac{1}{24} & \frac{1}{24} & \frac{1}{24} & \frac{5}{12} & \frac{5}{12} & \frac{1}{24} \end{bmatrix}.$$

Calculando $\mathbf{r}$ encontramos

$$\mathbf{r} = \begin{bmatrix} \frac{4}{51} \\ \frac{5}{51} \\ \frac{11}{102} \\ \frac{25}{102} \\ \frac{25}{102} \\ \frac{23}{102} \end{bmatrix}.$$

No artigo citeBrin98, Brin e Page comentam que em geral o valor de α é estabelecido como $0,85$.

Como mostraremos nos próximos capítulos, estas modificações resolvem todos os problemas descritos anteriormente. Em particular, podemos agora responder positivamente às três questões colocadas no início da seção 1.3. Para entender as respostas para estes questionamentos precisamos entender um pouco melhor as chamas Cadeias de Markov. Infelizmente, não conseguiremos mostrar que o vetor estacionário sempre existe, pois os conceitos necessários para tal fogem do escopo deste trabalho. No entanto apresentaremos uma demonstração acessível para a convergência no método iterativo.

Neste sentido note que, como todas as entradas da matriz são positivas, o cálculo do PageRank via resolução do sistema linear é muito complexo, uma vez que todas as variáveis estarão presentes em todas as equações. Isso torna o método iterativo ainda mais importante, uma vez que ele possui um custo operacional mais baixo (ver [11]).

2

CADEIAS DE MARKOV

”De onde eu venho não importa pois já passou O que importa é saber pra onde vou"(Senhorita - Zé Geraldo)

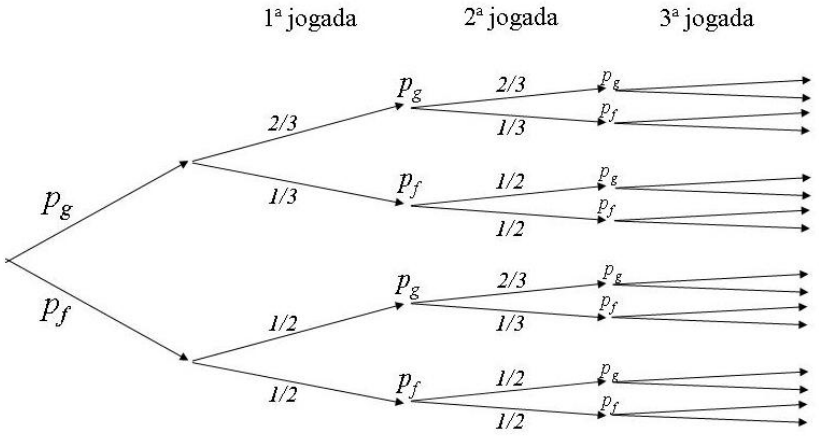
Ao longo da história homens brilhantes buscaram e buscam prever o futuro. Certamente os sentimentos de satisfação e prazer chegam ao ápice quando conseguimos enxergar ou prever o que os outros não conseguem, principalmente quando essas previsões partem da beleza de construções matemáticas.

Neste capítulo vamos estudar algumas propriedades básicas de uma estrutura matemática aleatória conhecida como Cadeia de Markov. O que estudaremos aqui é apenas o básico, suficiente para entendermos a matemática encontrada no PageRank. O leitor interessado em maiores detalhes sobre a teoria de probabilidades e cadeias de Markov, pode encontrar em [3, 6, 9, 10].

Para isso começaremos tentando entender o comportamento de um modelo simples. Suponhamos então que em um jogo de dados tenhamos a seguinte comanda: um jogador lança um dado, não viciado, com numeração de 1 a 6 em cada uma de suas faces e submetido as seguintes condições, se ao lançar o dado, este cair com uma das faces com numeração par voltada para cima, o jogador ganha e segue no jogo, porém, para ganhar na próxima jogada, ele deverá obter um número menor do que 5 na face superior do dado, dando ao jogador uma leve vantagem. Deste modo, se este jogador perde uma jogada ele teria probabilidade $1/2$ para ganhar no lançamento seguinte, pois para sair número par em um dado temos três possibilidades. Logo, a probabilidade deste jogador perder é também $1/2$. Agora, se o jogador ganhou uma dada jogada, a probabilidade de ganhar a próxima jogada é de $4/6 = 2/3$, e de perder é $2/6 = 1/3$.

O que descrevemos até agora é como o jogo se desenrola, como acontecem as transições de uma jogada para a próxima. Para o modelo ficar bem definido precisamos

definir ainda como o jogo começa! Sem uma referência anterior, não sabemos quais as probabilidades de vitória e derrota do jogador na primeira rodada. Assim, acompanhando a sequência de vitórias e derrotas do jogador a cada jogada n, precisamos de certa forma escolher se no instante 0 o jogador terá vitória ou derrota. Esta escolha pode ser feita também de forma aleatória. Defina então $p_g^{(k)}$ e $p_f^{(k)}$, $k = 0, 1, 2, \ldots$, as probabilidades de vitória e derrota, respectivamente, na $k-$ésima rodada. Vamos a seguir tentar calcular os valores de $p_g^{(k)}$ e $p_f^{(k)}$ para todo k, e estudar como se comportam estas probabilidade depois de várias jogadas.



Para iniciar, vamos tabelar as probabilidades dadas no problema. Temos

	g	f
g	$\frac{2}{3}$	$\frac{1}{2}$
f	$\frac{1}{3}$	$\frac{1}{2}$

E transformando em matriz obtemos

$$T = \begin{bmatrix} \frac{2}{3} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} \end{bmatrix}.$$

Analisando a árvore de probabilidades vemos que

$$p_g^{(1)} = \frac{2}{3}p_g^{(0)} + \frac{1}{2}p_f^{(0)}$$
$$p_f^{(1)} = \frac{1}{3}p_g^{(0)} + \frac{1}{2}p_f^{(0)}.$$

Repare também que

$$T \cdot \begin{bmatrix} p_g^{(0)} \\ p_f^{(0)} \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} \end{bmatrix} \cdot \begin{bmatrix} p_g^{(0)} \\ p_f^{(0)} \end{bmatrix} = \begin{bmatrix} \frac{2}{3}p_g^{(0)} + \frac{1}{2}p_f^{(0)} \\ \frac{1}{3}p_g^{(0)} + \frac{1}{2}p_f^{(0)} \end{bmatrix}$$

e portanto

$$\begin{bmatrix} p_g^{(1)} \\ p_f^{(1)} \end{bmatrix} = T \cdot \begin{bmatrix} p_g^{(0)} \\ p_f^{(0)} \end{bmatrix}.$$

E escrevendo

$$p^{(n)} = \begin{bmatrix} p_g^{(n)} \\ p_f^{(n)} \end{bmatrix},$$

para $n = 0, 1, 2, \cdots$, temos

$$p^{(1)} = T \cdot p^{(0)}.$$

Com o mesmo raciocínio encontramos que

$$\begin{aligned} p^{(2)} &= T \cdot p^{(1)} \\ p^{(3)} &= T \cdot p^{(2)} \\ &\vdots \\ p^{(n)} &= T \cdot p^{(n-1)}. \end{aligned} \tag{2.1}$$

É interessante observar também que para todo $n = 0, 1, 2, \ldots$

$$p^{(n)} = T \cdot p^{(n-1)} = T \cdot T \cdot p^{(n-2)} = T^2 \cdot p^{(n-2)} = \cdots = T^n \cdot p^{(0)}.$$

Ou seja

$$p^{(n)} = T^n \cdot p^{(0)}.$$

Com isso temos uma forma de calcular os valores de $p^{(n)}$ para diferentes valores de n. Mas como mostra a expressão acima, precisamos primeiro escolher valores para $p^{(0)}$. Tome então

$$p^{(0)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Calculando então algumas iterações, encontramos

$p^{(0)}$	$p^{(1)}$	$p^{(2)}$	$p^{(3)}$	$p^{(4)}$	$p^{(5)}$	$p^{(6)}$	$\ldots$	$p^{(10)}$
1	0,6667	0,6111	0,6019	0,6003	0,6001	0,6	$\ldots$	0,6
0	0,3333	0,3889	0,3981	0,3997	0,3999	0,4	$\ldots$	0,4

Tabela 6: Sequência de vetores $p^{(k)}$ para vários lançamentos.

Os valores da tabela acima são aproximados, e foram calculados com 4 casas decimais de precisão. Mas ainda assim ela parece sugerir que a medida que o jogo passa as probabilidades $p^{(n)}$ convergem para algum vetor

$$\begin{bmatrix} \frac{3}{5} \\ \frac{2}{5} \end{bmatrix}.$$

Mesmo que aceitemos que a sequência converge, o vetor limite pode ainda depender do vetor inicial $p^{(0)}$. Para verificar tal dependência, assim como tentar entender se a convergência ocorre, tome

$$p^{(0)} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Neste caso, calculando algumas iterações, encontramos

$p^{(0)}$	$p^{(1)}$	$p^{(2)}$	$p^{(3)}$	$p^{(4)}$	$p^{(5)}$	$p^{(6)}$	...	$p^{(10)}$
0	0,5	0,5833	0,5972	0,5995	0,5999	0,6	...	0,6
1	0,5	0,4167	0,4028	0,4005	0,4001	0,4	...	0,4

Tabela 7: Sequência de vetores $p^{(k)}$ para vários lançamentos.

Da mesma forma que na tabela 6, a tabela acima parece sugerir que $p^{(n)}$ converge para o mesmo vetor

$$\mu = \begin{bmatrix} \frac{3}{5} \\ \frac{2}{5} \end{bmatrix}.$$

Mas se a convergência ocorre para estas duas escolhas de vetor inicial então, para qualquer

$$p^{(0)} = \begin{bmatrix} p_g \\ p_f \end{bmatrix},$$

com $p_g + p_f = 1$, temos

$$p^{(0)} = \begin{bmatrix} p_g \\ p_f \end{bmatrix} = p_g \begin{bmatrix} 1 \\ 0 \end{bmatrix} + p_f \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Segue que

$$p^{(n)} = T^n \cdot p^{(0)} = T^n \left(p_g \begin{bmatrix} 1 \\ 0 \end{bmatrix} + p_f \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) = p_g T^n \begin{bmatrix} 1 \\ 0 \end{bmatrix} + p_f T^n \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

e assim $p^{(n)}$ converge para

$$p_g \begin{bmatrix} \frac{3}{5} \\ \frac{2}{5} \end{bmatrix} + p_f \begin{bmatrix} \frac{3}{5} \\ \frac{2}{5} \end{bmatrix} = \begin{bmatrix} \frac{3}{5} \\ \frac{2}{5} \end{bmatrix} (p_g + p_f) = \begin{bmatrix} \frac{3}{5} \\ \frac{2}{5} \end{bmatrix}.$$

Segue que $p^{(n)}$ deve convergir para

$$\mu = \begin{bmatrix} \frac{3}{5} \\ \frac{2}{5} \end{bmatrix}.$$

para qualquer escolha de $p^{(0)}$.

Uma boa forma de entender qual deve ser o vetor μ, supondo que a convergência ocorra, vem da equação (2.1). Como $p^{(n)} = T \cdot p^{(n-1)}$, se $p^{(n)}$ converge então o limite μ deve satisfazer

$$\mu = T\mu.$$

Ou seja, se

$$\mu = \begin{bmatrix} x \\ y \end{bmatrix},$$

com $x + y = 1$, então

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix},$$

de onde segue que

$$\begin{cases} \frac{2}{3}x + \frac{1}{2}y &= x \\ \frac{1}{3}x + \frac{1}{2}y &= y \\ x + y &= 1 \end{cases}.$$

Basta agora resolver o sistema acima para encontrar

$$\mu = \begin{bmatrix} \frac{3}{5} \\ \frac{2}{5} \end{bmatrix}.$$

Uma forma de interpretar tal convergência é pensar que, de certo modo, a medida que o jogo corre, o modelo "esquece" como o jogo começou e começa entrar em uma espécie de equilíbrio. Este como consequência da vantagem recebida pelo jogador a cada vitória, este equilíbrio é tal que o jogador, em média, ganha 3 a cada 5 jogadas e, consequentemente, perde 2 a cada 5 jogadas. Ou seja, o jogador ganha em média $3/5$ das jogadas, e perde $2/5$ destas.

Apesar de aqui termos modelado um simples jogo de dados, o fenômeno descrito é o mesmo estudado no algoritmo de PageRank. Se olharmos o vetor $\mathbf{r}^0$ como o vetor que

determina a probabilidade do internauta começar a navegar em cada página da web, a convergência do vetor $\mathbf{r}^k$ mostra que após algum tempo navegando, o internauta "esquece" em que página começou, fazendo o processo entrar em equilíbrio. E assim, como já comentamos anteriormente, as entradas do vetor limite $\mathbf{r}$ indicam então a proporção média de tempo que o internauta passa em cada página.

A seguir vamos explicar brevemente a estrutura matemática onde podemos encaixar estes dois problemas.

2.1 CADEIAS DE MARKOV

Processos como o descrito no exemplo anterior são conhecidos na matemática como *Cadeias de Markov*, e são frequentemente usados para modelar processos com incerteza.

Uma Cadeia de Markov nada mais é do que uma sequência aleatória de elementos, com uma regra de formação definida. Assim, para entender o que é uma Cadeia de Markov, precisamos antes entender como se dá a transição de um elemento a outro desta sequência. Para isso, considere a definição abaixo.

Definição 2.1. Uma matriz $T = [T_{i,j}]_{m\times m}$ dada por

$$\mathbf{T} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}$$

é chamada de *matriz de probabilidades de transição*, *matriz estocástica* ou simplesmente *matriz de transição* se

1. $p_{ij} \geq 0$, para quaisquer $i, j \in \{1, \ldots, m\}$;
2. Para qualquer $j \in \{1, \ldots, m\}$ vale que

$$p_{1j} + p_{2j} + \cdots + p_{mj} = 1.$$

Ou seja, a soma dos elementos de cada coluna é sempre 1.

Dada uma matriz de transição T, para construir agora o que chamaremos de cadeia de Markov, proceda da seguinte forma:

- Tome um conjunto $S = \{a_1, \ldots, a_m\}$ que chamaremos de *espaço de estados*, e em seguida escolha um elemento de S para iniciar o processo, e chame este elemento de X_0. A escolha de X_0 pode ser feita de forma aleatória ou determinística.
- A seguir, se $X_0 = a_j$ escolha o estado X_1 usando para isso a j-ésima coluna da matriz de transição T. Ou seja, a probabilidade de $X_1 = a_i$ se $X_0 = a_j$ será p_{ij}.
- Siga da mesma forma para os estados seguintes. Ou seja, sempre que a cadeia estiver em a_k, ela saltará para a_l com probabilidade p_{lk}.

O terceiro passo acima será denotado por

$$\mathbb{P}(X_n = a_l | X_{n-1} = a_k) = p_{lk},$$

e leremos "a probabilidade de X_n ser igual a a_l dado que X_{n-1} é igual a a_k é p_{lk}". É por esta razão que precisamos que a soma os elementos de cada coluna seja 1.

Temos

Definição 2.2. Uma *cadeia de Markov* ou *processo de Markov* com matriz de transição T e espaço de estados $S = \{a_1, a_2, ..., a_m\}$ é uma sequência aleatória $X_0, X_1, \ldots$, tal que

(i) $\mathbb{P}(X_n = a_i | X_{n-1} = a_j) = p_{ij}$ para quaisquer $i, j \in \{1, \ldots, m\}$;

(ii) A trajetória feita para chegar até X_{n-1} não muda as probabilidades de escolha de X_n, dependendo apenas de X_{n-1}.

Assim, o jogo descrito no início do capítulo é um exemplo de Cadeia de Markov. Da mesma forma, se considerarmos a interpretação probabilística do algoritmo de PageRank descrito no capítulo 1, temos outro exemplo de Cadeia de Markov.

Agora, seguindo os passos do exemplo anterior, vamos considerar que a cadeia começa em um estado aleatório, escolhido de acordo com um vetor de probabilidades

$$p^{(0)} = \begin{bmatrix} p_1^{(0)} \\ \vdots \\ p_m^{(0)} \end{bmatrix}.$$

Ou seja, a probabilidade de iniciarmos o processo em um estado a_k é dada por $p_k^{(0)}$. Perguntamos agora qual a probabilidade $p_i^{(1)}$ de estarmos em um estado a_i no instante

1. Seguindo o mesmo raciocínio do exemplo, percebemos que, para estar em a_i no instante 1, temos que primeiro saber onde a cadeia estava no instante 0. Assim, se a cadeia estava em a_1, o que ocorre com probabilidade $p_1^{(0)}$, ela salta para a_i com probabilidade p_{i1}. Do mesmo modo, se estava em a_2 (e isso ocorre com probabilidade $p_2^{(0)}$, a cadeia salta para a_i com probabilidade p_{i2}. Seguindo o mesmo raciocínio para cada estado de S encontramos que

$$p_i^{(1)} = p_{i1} \cdot p_1^{(0)} + p_{i2} \cdot p_2^{(0)} + \cdots + p_{im} \cdot p_m^{(0)}.$$

E assim, escrevendo as probabilidade $p_i^{(1)}$ como um vetor coluna. Ou seja, se

$$p^{(1)} = \begin{bmatrix} p_1^{(1)} \\ \vdots \\ p_m^{(1)} \end{bmatrix},$$

encontramos que

$$p^{(1)} = T \cdot p^{(0)}. \tag{2.2}$$

Vamos aqui fazer um pequeno parêntese, para notar que se $v = [v_i]_{m\times 1}$ é um *vetor de probabilidades*, isto é,

- $v_k \geq 0$ para todo $k = 1, \ldots, m$;
- $v_1 + v_2 + \cdots + v_m = 1$,

então para $u = Tv$, temos

$$u_i = T_{i1}v_1 + T_{i2}v_2 + T_{i3}v_3 + \ldots + T_{im}v_m \geq 0.$$

Além disso

$$u_1 + u_2 + \ldots + u_m = (T_{11}v_1 + \ldots + T_{1m}v_m) + (T_{21}v_1 + \ldots + T_{2m}v_m) + \ldots + (T_{m1}v_1 + \ldots + T_{mm}v_m)$$

e portanto, como T é uma matriz de transição, cada uma de suas colunas é um vetor de probabilidades, e

$$\begin{aligned} u_1 + u_2 + \ldots + u_m &= (T_{11} + T_{21} + \ldots + T_{m1})v_1 + \ldots + (T_{1m} + T_{2m} + \ldots + T_{mm})v_m \\ &= v_1 + v_2 + \ldots + v_m \\ &= 1. \end{aligned}$$

Daí se T é uma matriz de transição e v é um vetor de probabilidades, então $u = Tv$ é também um vetor de probabilidades. Concluímos então que o vetor $p^{(1)}$ definido em (2.2) é um vetor de probabilidades.

Voltando ao nosso problema, do mesmo modo que em (2.2), podemos definir *vetor de probabilidades* $p^{(n)}$ como o vetor cuja a i-ésima linha indica a probabilidade de observarmos o estado a_i após n passos. E assim, teremos

$$p^{(n)} = \begin{bmatrix} p_1^{(n)} \\ \vdots \\ p_m^{(n)} \end{bmatrix}$$

e

$$p^{(n)} = T \cdot p^{(n-1)}. \tag{2.3}$$

Assim, abrindo a relação acima, encontramos que

$$p^{(n)} = T^n \cdot p^{(0)}. \tag{2.4}$$

Observação 2.1.1. *A equação* (2.4) *nos permite interpretar a entrada* (i, j) *da matriz* T^n *como a probabilidade da cadeia ir de* a_j *para* a_i *em n passos. Para ver isso, denote as entradas de* T^n *por* b_{ij} *e observe que*

$$p_i^{(n)} = b_{i1} \cdot p_1^{(0)} + b_{i2} \cdot p_2^{(0)} + \cdots + b_{im} \cdot p_m^{(0)}.$$

2.1.1 *Medidas Invariantes*

No exemplo descrito no início do capítulo, mostramos que os vetores

$$p^{(n)} = \begin{bmatrix} p_g^{(n)} \\ p_f^{(n)} \end{bmatrix}$$

convergiam para um certo vetor μ, e que esta convergência não dependia do vetor $p^{(0)}$ escolhido. Mostramos também que o vetor é tal que $\mu = T\mu$. Surge então a questão: Isto é válido sempre?

Mais exatamente, queremos saber se dado um vetor de probabilidades $p^{(0)}$, a sequência $p^{(n)}$ definida pela equação (2.3) converge para algum vetor μ. E de que forma μ depende de $p^{(0)}$. Esta pergunta já surgiu na descrição do algoritmo PageRank, e lá mostramos que a convergência nem sempre ocorre. Surgem então algumas questões:

1. Sempre existe um vetor μ tal que $\mu = T\mu$?
2. Caso exista, é único?
3. Em que situações $p^{(n)}$ converge para μ?

4. A convergência depende de $p^{(0)}$?

Infelizmente, responder parte destas peguntas, como a existência e unicidade de tal vetor, fogem do escopo deste trabalho. De todo modo vamos tentar dar alguns passos no sentido das respostas. Começaremos estudando o vetor limite, e que características ele deve ter. No próximo capítulo apresentaremos uma construção que permite mostrar a convergência para tal vetor, uma vez que sabemos de sua existência e unicidade.

Suponhamos então que a sequência $p^{(n)}$ converge para um vetor de probabilidades

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{bmatrix}.$$

Segue da equação (2.3), fazendo n crescer, que μ deve satisfazer

$$\mu = T\mu. \tag{2.5}$$

Vetores de probabilidades satisfazendo (2.5) são conhecidos como *medidas invariantes* da cadeia de Markov com matriz de transição T, e razão para isso é que, se escolhermos o estado inicial da cadeia de acordo com o vetor μ, então a probabilidade da cadeia estar em a_i no instante n é μ_i para qualquer instante n. De fato, da equação (2.5), segue

$$T^2\mu = T(T\mu) = T\mu = \mu,$$

e seguindo deste modo

$$T^n\mu = \mu.$$

Assim, se $p^{(0)} = \mu$ então

$$p^{(n)} = T^n \cdot p^{(0)} = T^n\mu = \mu.$$

É interessante observar também que, de (2.4), uma das formas de estudar a convergência de $p^{(n)}$ é estudar a convergência da matriz T^n. Em particular, a existência e unicidade de μ é equivalente a convergência de T^n para uma matriz M com todas as colunas iguais. Isso segue do fato que, se

$$v_k = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix},$$

é o vetor com a k-ésima entrada 1 e as demais 0, então Mv_k retorna a k-ésima coluna da matriz M.

Assim se T^n converge para M com todas as colunas iguais a um vetor μ então,

$$Mv_k = \mu,$$

para todo $k = 1, \ldots, m$. E como $p^{(0)} = p_1^{(0)} v_1 + \cdots + p_m^{(0)} v_m$ então

$$p^{(n)} = T^n \cdot p^{(0)} \longrightarrow M \cdot p^{(0)}, \text{quando } n \to \infty.$$

$$\begin{aligned} Mp^{(0)} &= M\left(p_1^{(0)} v_1 + \cdots + p_m^{(0)} v_m\right) \\ &= p_1^{(0)} Mv_1 + \cdots + p_m^{(0)} Mv_m \\ &= \left(p_1^{(0)} + \cdots + p_m^{(0)}\right) \mu \\ &= \mu, \end{aligned}$$

de onde segue que para qualquer escolha de $p^{(0)}$,

$$p^{(n)} \longrightarrow \mu, \quad \text{quando } n \to \infty.$$

A invariância de μ sai diretamente de (2.3), fazendo $n \to \infty$.

Reciprocamente, se $p^{(n)} \longrightarrow \mu$, quando $n \to \infty$, para qualquer escolha de $p^{(0)}$, fazendo $p^{(0)} = v_k$ mostramos que T^n converge para uma matriz M com todas as colunas iguais a μ.

Os cálculos acima mostram que μ pode ser visto como uma espécie de equilíbrio da cadeia. Ou seja, se deixamos a cadeia rodar por tempo suficiente, o vetor de probabilidades $p^{(n)}$ fica próximo de μ, e cadeia começa a se comportar aproximadamente como se tivéssemos escolhido o estado inicial de acordo com μ. Em outras palavras, se para n suficientemente grande temos

$$p^{(n)} \sim \mu$$

, então

$$p^{(n+1)} = T \cdot p^{(n)} \sim T\mu = \mu,$$

e para todo $k > n$ teremos

$$p^{(k)} \sim \mu.$$

Assim, se quisermos calcular, por exemplo, o tempo médio que a cadeia passa em cada estado, podemos observar o comportamento da cadeia apenas quando esta entra

em equilíbrio (ou quando está próximo disso), ou então considerar que a cadeia já foi iniciada de acordo com μ.

A relação entre tempo médio passado em um estado e a medida invariante μ, é a mesma explicada no exemplo do jogo de dados, e no algoritmo de PageRank. Lembrando, no caso do jogo de dados, após um número suficientemente grande de jogadas, o jogo esquece como foi iniciado e começa a observar uma média de 3 vitórias a cada 5 jogadas, proporção esta dada pelo vetor μ.

O mesmo ocorre para uma medida invariante qualquer. Considere uma cadeia de Markov cujo estado inicial foi escolhido de acordo com uma medida invariante μ. Sabemos que a probabilidade da cadeia estar em um estado a_i em qualquer instante n é dada por μ_i, e portanto depois de um tempo n suficientemente grande, esperamos observar a_i aproximadamente $n\mu_i$ vezes. Assim, o tempo médio que a cadeia passa em a_i é exatamente μ_i.

O algoritmo PageRank pode ser visto então como uma cadeia de Markov, com matriz de transição G dada por (1.4). O vetor de PageRank $\mathbf{r}$ é portanto exatamente a medida invariante da cadeia, como vemos em (1.5). A idéia do algoritmo é se valer da convergência da sequência $\mathbf{r}^n$ e encontrar uma aproximação para $\mathbf{r}$.

2.2 MATRIZ DE TRANSIÇÃO REGULAR

Nesta seção vamos enunciar condições nas quais a medida invariante existe, é única e a convergência de $p^{(n)}$ ocorre independente de como escolhemos o estado inicial. Como já comentamos, a demonstração de existência e unicidade da medida invariante é técnica, e foge do escopo deste trabalho. A convergência, na situação que apresentaremos abaixo, é também técnica e sem interesse para nós. Mas no capítulo 3 estudaremos uma forma de mostrar tal convergência para algumas matrizes de transição, incluindo a matriz usada no algoritmo de PageRank.

Antes de mais nada vamos lembrar os principais problemas que precisamos resolver no algoritmo de PageRank.

Para começar, lembre dos problemas das páginas dispostas em ciclos. Neste caso, a convergência não ocorria pois, se iniciássemos a cadeia em um certo estado no ciclo, aconteceria um alternância entre estados, de modo que certos estados só poderiam

visados em certos instantes de tempo. Isso se reflete na sequência $\mathbf{r}^k$, de modo a torná-la uma sequência periódica e, portanto, não convergente.

Os outros dois problemas, a saber as páginas sem links e os conjuntos de páginas auto-referenciadas, levavam ao mesmo problema. Ao entrarmos nestes sites, nunca mais sairíamos deles, fazendo com que o medida invariante fosse nula fora destes pontos.

É fato que isso pode não afetar diretamente a existência da medida invariante, ou na convergência para ela, mas a existência de mais de um destes conjuntos pode sim afetar sua unicidade. Para isso basta observar que, caso tenhamos dois conjuntos auto-referenciados distintos, e a cadeia inicie seu passeio dentro do primeiro, ela não sairá mais de lá. Assim, a convergência, caso ocorra, será para uma medida invariante com entrada nula para sites do segundo conjunto. Repetindo a mesma análise para uma cadeia iniciada dentro do segundo grupo de páginas, encontramos uma possível medida invariante nula nas páginas do primeiro grupo. Encontramos assim duas medidas distintas!

Uma forma de corrigir este problema é tomar o que chamaremos de matriz de transição regular, que definimos abaixo.

Definição 2.3. Uma matriz de probabilidades de transição é *regular* se alguma de suas potências tem todos os elementos não nulos.

Em particular, a matriz G usado no algoritmo do PageRank é uma matriz de transição regular, uma vez que todas as suas entradas já são estritamente positivas.

Para terminar este capítulo, enunciaremos sem demonstrar um teorema que garante o funcionamento do algoritmo de PageRank.

Teorema 2.4. *Se $T = [T_{i,j}]_{m\times m}$ é uma matriz de transição regular, então*

(i) Existe um único vetor de probabilidades $\mu = [\mu_i]_{m\times 1}$ tal que $\mu = T\mu$;

(ii) Para qualquer vetor de probabilidades $p^{(0)}$ a sequência definida por

$$p^{(n)} = T \cdot p^{(n-1)},$$

$n \in \{1, 2, 3, \ldots\}$, converge para o vetor μ.

3

ACOPLANDO CADEIAS

Neste capítulo vamos mostrar uma maneira de construir cadeias de Markov, e a partir de uma destas construções mostraremos a convergência para medida invariante tratada do capítulo anterior. A maneira descrita abaixo é uma adaptação discreta do método apresentado em [6], onde consideraremos apenas matrizes com entradas racionais. Os cálculos para matrizes com entradas quaisquer são similares, mas pressupõe do leitor um conhecimento um pouco mais profundo de teoria de probabilidades, e por isso decidimos por esta adaptação.

3.1 SIMULANDO UMA DISTRIBUIÇÃO

Considere a seguinte distribuição de probabilidades no conjunto $\Omega = \{x_1, x_2, x_3, x_4, x_5\}$ dada por

$$P_1 = \left\{ \frac{1}{10}; \frac{3}{10}; \frac{2}{10}; \frac{1}{10}; \frac{3}{10} \right\}.$$

Ou seja, a probabilidade de sortear x_1 é $1/10$, de sortear x_2 é $3/10$ e assim por diante.

Queremos agora *simular* a distribuição P_1. Em outras palavras, queremos criar um experimento aleatório no qual a probabilidade de observar um dado valor do conjunto $\{x_1, x_2, x_3, x_4, x_5\}$ seja dada pelo valor correspondente em P_1.

A primeira ideia que vem a mente é escolher uma série de bolas coloridas de cor distinta, e marcá-las com elementos de Ω de modo que as proporções desejadas sejam respeitadas. Precisaremos então de 1 bola marcada x_1, 3 marcadas x_2, 2 com x_3, 1 com x_4 e outras 3 com x_5, em um total de 10 bolas.

Uma alternativa equivalente (e que mais a frente vai se mostrar útil) é separar 10 bolinhas de mesmo tamanho e ordená-las. Na figura 8 associamos a cada bola um

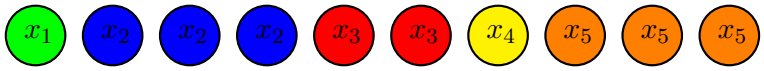


Figura 8: Simulando uma distribuição discreta.

elemento do conjunto $\{x_1, x_2, x_3, x_4, x_5\}$ de modo que a proporção de bolas de um certo tipo seja justamente a probabilidade de sortearmos este tipo. A figura 9 mostra uma maneira de fazer isso.

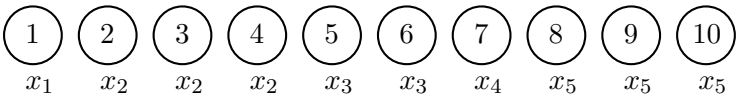


Figura 9: Simulando uma distribuição discreta.

Deste modo se sortearmos, por exemplo, a bola de número 5 estaremos sorteando o elemento x_3, e da mesma forma sortear a bola 8 equivale a sortear o elemento x_5.

É interessante observar que esta relação entre bola numerada e elemento do conjunto pode ser feito de diversas formas diferentes, e as probabilidades serão as mesmas. A figura 10 mostra outras duas formas de relacionar os elementos de Ω com bolas numeradas que simulam o mesmo sorteio.

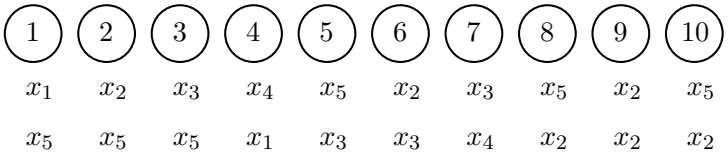


Figura 10: Duas outras formas de relacionar elementos e bolas.

Apenas para fixar ideia, vamos fazer mais um exemplo. Considere a distribuição de $\Omega = \{x_1, x_2, x_3, x_4\}$ dada por

$$P = \left\{ \frac{1}{4}, \frac{1}{3}, \frac{1}{6}, \frac{1}{4} \right\}.$$

Para determinar o número de bolas que necessitaremos, o primeiro passo é reescrever as probabilidades de P com um denominador comum. Obtemos

$$P = \left\{ \frac{3}{12}, \frac{4}{12}, \frac{2}{12}, \frac{3}{12} \right\}.$$

Tomamos então 12 bolas e relacionamos com os elementos de Ω na proporção indicada por P, como na figura 11.

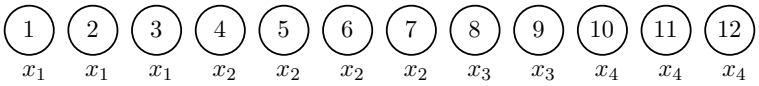


Figura 11: Duas outras formas de relacionar elementos e bolas.

Feito isso basta sortear uma das 12 bolas, e verificar qual o elemento de Ω está associado a ela. Assim, se sortearmos a bola 6, por exemplo, estaremos sorteando de fato o elemento x_2.

3.1.1 *Acoplando duas distribuições*

Considere agora duas distribuições distintas no conjunto $\{x_1, x_2, x_3, x_4, x_5\}$ dadas por

$$P_1 = \left\{\frac{1}{10}; \frac{1}{5}; \frac{3}{10}; \frac{3}{10}; \frac{1}{10}\right\}$$

e

$$P_2 = \left\{\frac{1}{5}; \frac{1}{5}; \frac{1}{5}; \frac{1}{5}; \frac{1}{5}\right\}.$$

Seguindo a mesma ideia apresentada anteriormente queremos agora fazer um único experimento aleatório que simule as duas distribuições simultaneamente.

Nós poderíamos usar dois conjuntos de bolas diferentes, e proceder da mesma forma que fizemos no caso de apenas uma distribuição, sorteando duas bolas (uma de cada conjunto). Mas queremos um pouco mais do que isso, queremos sortear apenas uma bola, e com isso simular as duas distribuições. A resposta na verdade não é muito difícil: basta fazer duas atribuições diferentes ao mesmo conjunto de bolas.

Primeiro observe que o total de bolas necessárias para simular cada uma das distribuições é distinta: para P_1 precisamos de 10 bolas, enquanto para P_2 são necessárias 5 bolas. Assim, para corrigir o problema precisamos primeiro escrever todas as probabilidades (em P_1 e P_2) usando um mesmo denominador. Temos

$$P_1 = \left\{\frac{1}{10}; \frac{2}{10}; \frac{3}{10}; \frac{3}{10}; \frac{1}{10}\right\}$$

e

$$P_2 = \left\{\frac{2}{10}; \frac{2}{10}; \frac{2}{10}; \frac{2}{10}; \frac{2}{10}\right\}.$$

Agora, assim como fizemos no caso de uma única distribuição, para cada distribuição temos que associar uma bola a um estado de Ω. A maneira mais simples de fazer isso é a sequencial, que mostramos na figura 12. Seguindo este arranjo de bolas e estados, ao sortearmos a bola de número 2, estaremos sorteando x_2 na distribuição P_1 e x_1 na distribuição P_2.

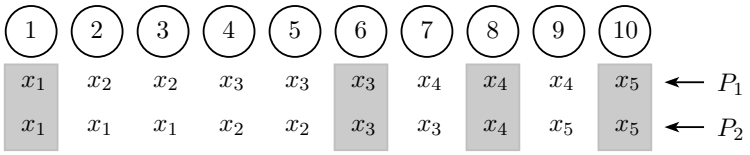


Figura 12: Simulando P_1 e P_2 simultaneamente. Em destaque estão os sorteios onde os valores de P_1 e P_2 coincidem.

É interessante observar que ao sortearmos as bolas 1, 6, 8 ou 10, estaremos sorteando o mesmo elemento de Ω nas duas distribuições. Isso significa que, nesta construção, a probabilidade dos estados coincidirem nas duas distribuições é de $4/10$. O evento

$$C = \{\text{O estado sorteado nas duas distribuições é o mesmo}\}$$

é conhecido como *evento acoplante*, e sua probabilidade $\mathbb{P}(C)$ é chamada de *probabilidade de acoplamento*.

Não é difícil ver que a probabilidade de acoplamento depende da construção (associação de bolas e estados) escolhida. Analisando a construção dada na figura 12, por exemplo, notamos que se em P_2 tivéssemos associado as bolas 2 e 3 a x_2 e as bolas 4 e 5 a x_1, faríamos aumentar as coincidências com a distribuição P_1, aumentando assim a probabilidade de acoplamento.

Surge então a próxima questão: como fazer para maximizar a probabilidade de acoplamento?

Para isso temos que aumentar o total de coincidências entre associações das bolas nas duas distribuições. Analisemos as distribuições estado por estado.

- O estado x_1 deverá ser associado a apenas uma bola em P_1 e 3 bolas em P_2, assim só devemos separar uma bola que será associada a x_1 em ambas as distribuições.
- x_2 deve se associar a 2 estados em cada uma das distribuições, e portando separamos 2 bolas para isso;

- x_3 deve ser associado a 3 estados em P_1 e 2 em P_2. Separamos então 2 bolas para isso;
- x_4 precisará de apenas 1 bola de coincidência;
- x_5 também precisará de apenas 1.

Assim podemos associar as 7 primeiras bolas aos estados de Ω de acordo com as quantidades previstas acima. As demais, podemos distribuir da maneira que quisermos, bastando que completemos as proporções necessárias para cada distribuição.

O procedimento acima pode ser descrito da seguinte forma:

- Defina $m_i = \min\{P_1(x_i); P_2(x_i)\}$;
- Defina n_i como o numerador de m_i;
- Associe as n_1 primeiras bolas à x_1, as n_2 bolas seguintes a x_2 e assim até x_5;
- Complete as proporções corretas de cada distribuição do jeito que preferir.

Veja a figura 13

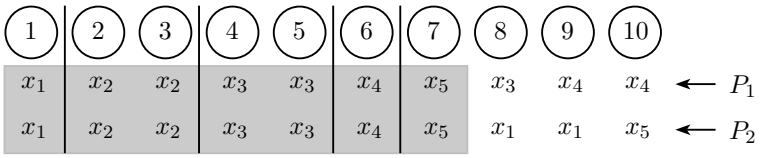


Figura 13: Acoplando P_1 e P_2.

Assim, ao sortearmos uma das bolas numeradas, o valor correspondente nas duas distribuições será o mesmo se, e somente se, a bola sorteada tiver valor menor ou igual a 7. Segue então que a probabilidade de acoplamento é igual a $7/10$.

De modo geral, para acoplar duas distribuições em $\{x_1, \ldots, x_k\}$ com entradas racionais dadas por

$$P_1 = \{P_1(x_1), \ldots, P_1(x_k)\}$$

e

$$P_2 = \{P_2(x_1), \ldots, P_2(x_k)\},$$

basta seguir os seguintes passos:

- Escreva todas as entradas de P_1 e P_2 usando um único denominador comum. Este denominador será exatamente o total de bolas que usaremos para o sorteio;

- Defina $m_i = \min\{P_1(x_i); P_2(x_i)\}$, para cada $i = 1, \ldots, k$;
- Defina n_i como o numerador de m_i;
- Associe as n_1 primeiras bolas à x_1, as n_2 bolas seguintes a x_2 e assim até x_k;
- Complete as proporções corretas de cada distribuição do jeito que preferir.

Neste caso, o acoplamento acontecerá se a bola sorteada tiver numeração menor ou igual a $n_1 + n_2 + \cdots + n_k$. E assim a probabilidade de acoplamento é igual a $m_1 + \cdots + m_k$.

3.1.2 *Acoplando três ou mais distribuições*

Agora que vimos como simular duas distribuições simultaneamente, preparar um experimento que simule conjuntamente três ou mais distribuições em um mesmo conjunto Ω não é exatamente um desafio. Seguindo os passos vistos anteriormente, começamos determinando o total de bolas necessárias, e em seguida basta associar cada bola aos estados correspondentes em cada distribuição.

Para maximizar a probabilidade de acoplamento, também podemos seguir os mesmos passos descritos na seção anterior. Para isso considere então l distribuições em $\Omega = \{x_1, \ldots, x_k\}$, todas com entradas racionais, dadas por

$$P_i = \{P_i(x_1), \ldots, P_i(x_k)\},$$

com $i = 1, \ldots, l$.

Para preparar o experimento faça o seguinte:

- Escreva todas as entradas de $P_1, P_2, \ldots, P_l$ usando um único denominador comum. Este denominador será exatamente o total de bolas que usaremos para o sorteio;
- Defina $m_i = \min\{P_1(x_i); P_2(x_i)\}$, para cada $i = 1, \ldots, k$;
- Defina n_i como o numerador de m_i;
- Associe as n_1 primeiras bolas à x_1, as n_2 bolas seguintes a x_2 e assim até x_k;
- Complete as proporções corretas de cada distribuição do jeito que preferir.

No caso de três os mais distribuições diremos que aconteceu o acoplamento se o elemento sorteado em todas as distribuições for igual. Não basta, portanto, que haja

coincidência em uma parte das distribuições simuladas. Ou seja, neste caso o evento acoplante pode ser definido como

$$C = \{\text{O estado sorteado nas } l \text{ distribuições é o mesmo}\}.$$

Para exemplificar melhor considere as seguintes distribuições em $\Omega = \{x_1, x_2, x_3, x_4, x_5\}$.

$$P_1 = \left\{\frac{1}{12}; \frac{1}{4}; \frac{1}{4}; \frac{1}{3}; \frac{1}{12}\right\},$$

$$P_2 = \left\{\frac{1}{4}; \frac{1}{3}; \frac{1}{6}; 0; \frac{1}{4}\right\},$$

$$P_3 = \left\{\frac{1}{12}; \frac{1}{4}; \frac{1}{4}; \frac{1}{3}; \frac{1}{12}\right\},$$

$$P_4 = \left\{\frac{1}{12}; \frac{5}{12}; \frac{1}{6}; \frac{1}{12}; \frac{1}{4}\right\}.$$

Escrevendo todas as entradas com um mesmo denominador temos

$$P_1 = \left\{\frac{1}{12}; \frac{3}{12}; \frac{3}{12}; \frac{4}{12}; \frac{1}{12}\right\},$$

$$P_2 = \left\{\frac{3}{12}; \frac{4}{12}; \frac{2}{12}; 0; \frac{3}{12}\right\},$$

$$P_3 = \left\{\frac{1}{12}; \frac{3}{12}; \frac{3}{12}; \frac{4}{12}; \frac{1}{12}\right\},$$

$$P_4 = \left\{\frac{1}{12}; \frac{5}{12}; \frac{2}{12}; \frac{1}{12}; \frac{3}{12}\right\}.$$

Isso nos diz que precisaremos de 12 bolas para realizar o experimento.

Para determinar as associações entre bolas e elementos de Ω, primeiro calculamos os valores de m_1, m_2, m_3, m_4, m_5. Assim

$$m_1 = \min\left\{\frac{1}{12}; \frac{3}{12}; \frac{1}{12}; \frac{1}{12}\right\} = \frac{1}{12},$$

$$m_2 = \min\left\{\frac{3}{12}; \frac{4}{12}; \frac{3}{12}; \frac{5}{12}\right\} = \frac{3}{12},$$

$$m_3 = \min\left\{\frac{3}{12}; \frac{2}{12}; \frac{3}{12}; \frac{2}{12}\right\} = \frac{2}{12},$$

$$m_4 = \min\left\{\frac{4}{12}; 0; \frac{4}{12}; \frac{1}{12}\right\} = 0,$$

$$m_5 = \min\left\{\frac{1}{12}; \frac{3}{12}; \frac{1}{12}; \frac{3}{12}\right\} = \frac{1}{12}.$$

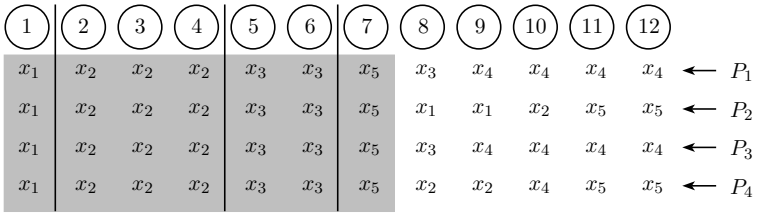


Figura 14: Simulando 4 distribuições.

De onde segue que $n_1 = 1$, $n_2 = 3$, $n_3 = 2$, $n_4 = 0$ e $n_5 = 1$, como ilustrado na figura 14.

Fica claro então que, para esta construção, a probabilidade de acoplamento é de $7/12$.

3.2 ACOPLANDO CADEIAS DE MARKOV

Nesta seção vamos ver como fazer para simular as trajetórias de uma cadeia de Markov, a partir de um estado inicial dado. Feito isso vamos estudar uma técnica para simular simultaneamente várias trajetórias de uma mesma cadeia, cada uma delas iniciando em um estado diferente, de modo que sempre que duas trajetórias se encontrem, sigam iguais para sempre.

Para deixar as contas mais claras, vamos começar trabalhando com um exemplo específico, para depois generalizar para outras cadeias. Considere então a cadeia de Markov com espaço de estados $\Omega = \{x_1, x_2, x_3, x_4\}$ e matriz de transição dada por

$$T = \begin{bmatrix} \frac{1}{6} & \frac{1}{4} & \frac{1}{12} & \frac{1}{3} \\ \frac{5}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{4} \\ \frac{1}{6} & \frac{1}{4} & \frac{1}{4} & \frac{1}{3} \\ \frac{1}{4} & \frac{5}{12} & \frac{7}{12} & \frac{1}{12} \end{bmatrix}.$$

Lembrando, cada coluna representa uma distribuição de probabilidade, associada a um estado da cadeia. Assim, sempre que a cadeia estiver no estado x_k usaremos a distribuição P_k descrita na linha k para escolher o próximo estado que iremos.

Para simular a cadeia acima precisamos primeiro simular cada uma das distribuições acima. Assim, escrevendo todas as entradas com mesmo denominador, temos

$$P_1 = \left\{ \frac{2}{12}; \frac{5}{12}; \frac{2}{12}; \frac{3}{12} \right\},$$

$$P_2 = \left\{ \frac{3}{12}; \frac{1}{12}; \frac{3}{12}; \frac{5}{12} \right\},$$
$$P_3 = \left\{ \frac{1}{12}; \frac{1}{12}; \frac{3}{12}; \frac{7}{12} \right\},$$
$$P_4 = \left\{ \frac{4}{12}; \frac{3}{12}; \frac{4}{12}; \frac{1}{12} \right\}.$$

Na figura 15 ilustramos um acoplamento simples destas distribuições, onde associamos cada bola aos estados de forma sequencial.

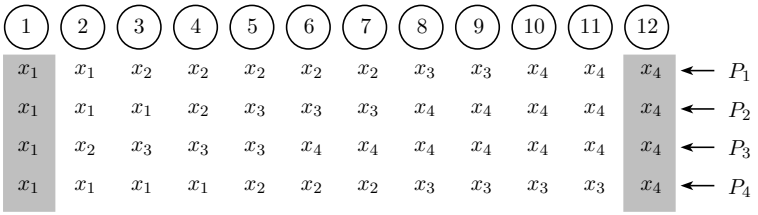


Figura 15: Acoplamento simples das distribuições relacionadas aos estados de uma cadeia.

Para simular a cadeia de Markov seguimos os seguintes passos:

- Escolhemos um estado inicial u;
- Simulamos a distribuição associada a u, e o seguimos para o estado indicado pela simulação;
- Repetimos o passo anterior usando o novo estado no lugar de u.

Para ficar mais claro, considere que escolhemos iniciar nossa cadeia no estado x_2. Assim, para dar o primeiro passo vamos simular a distribuição P_2, e para isso considere o acoplamento mostrado na figura 15. Suponha que sorteamos a bola de número 10. Isso nos leva ao estado x_3, e para o próximo passo temos que usar a distribuição P_3. Se sortearmos agora o número 9, saltamos para x_4, e usaremos P_4 para escolher o próximo passo.

Na tabela abaixo ilustramos algumas simulações possíveis, considerando diferentes estados iniciais para a cadeia. Denotaremos por X_n^i o estado na cadeia no passo n tendo iniciado no estado x_i, ou seja, $X_0^i = x_i$. Chamaremos de $U_1, U_2, U_3, \ldots$ os valores das bolas sorteadas em cada passo da cadeia.

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
U_n	-	7	4	6	9	11	2	6	4	2	4	3	8	5	12
X_n^1	x_1	x_2	x_2	x_3	x_4	x_3	x_2	x_3	x_3	x_2	x_2	x_1	x_3	x_3	x_4
X_n^2	x_2	x_3	x_3	x_4	x_3	x_4	x_1	x_2	x_2	x_1	x_2	x_1	x_3	x_3	x_4
X_n^3	x_3	x_4	x_1	x_2	x_4	x_3	x_2	x_3	x_3	x_2	x_2	x_1	x_3	x_3	x_4
X_n^4	x_4	x_2	x_2	x_3	x_4	x_3	x_2	x_3	x_3	x_2	x_2	x_1	x_3	x_3	x_4

Na figura 16 colocamos um gráfico das trajetórias governadas pelos mesmos sorteios, mas com estados iniciais distintos.

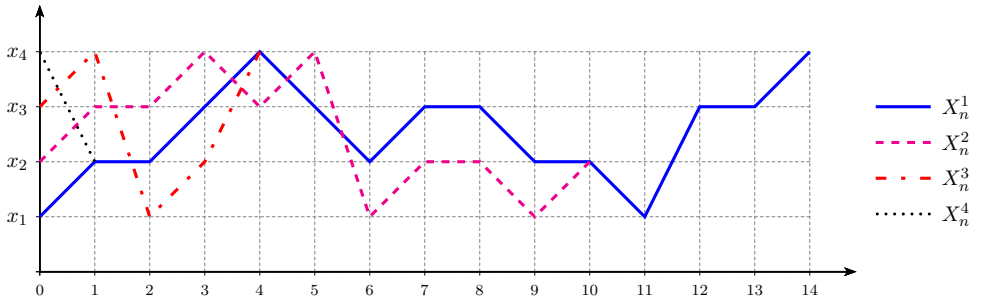


Figura 16: Simulação das trajetórias de uma Cadeia de Markov usando acoplamento simples das transições.

Observe que as trajetórias se acoplam todas no passo 10, e a partir daí seguem juntas. O instante do acoplamento é algo difícil de se determinar, mas podemos fazer algumas estimativas se considerarmos a maneira como acoplamos as distribuições de transição. Para isso, observe na figura 15 que os números 1 e 12 estão cada um deles relacionados à um único estado em todas as distribuições. Ou seja, se sortearmos 1 ou 12 em algum passo da simulação todas as trajetórias saltarão para o mesmo estado, independente de onde estejam no passo anterior. Isso não quer dizer que ao sortearmos um outro valor, não possa ocorrer o acoplamento. O exemplo acima mostra que isso é possível. A diferença é que neste caso precisaremos saber exatamente onde estão as trajetórias para poder garantir o acoplamento, e sorteando 1 ou 12 esta informação se torna desnecessária.

Podemos então estimar probabilidades relacionadas ao tempo que uma cadeia demora para se acoplar. Seja τ o tempo que todas as trajetórias simuladas como no exemplo acima demorem para acoplar. Formalmente falando temos

$$\tau = \min\{n > 0 : X_n^1 = X_n^2 = X_n^3 = X_n^4\}.$$

Note que se $\tau > n$ então $U_1 \notin \{1, 12\}, U_2 \notin \{1, 12\}, \ldots, U_n \notin \{1, 12\}$, e portanto

$$\mathbb{P}(\tau > n) \leq \left(1 - \frac{2}{12}\right)^n = \left(\frac{5}{6}\right)^n.$$

Fica claro assim que para "diminuir" o tempo que demora para as trajetórias acoplarem precisamos aumentar a probabilidade de acoplamento das distribuições de transição. Encontrar o acoplamento que maximiza esta probabilidade não é tarefa fácil, mas podemos dar um primeiro passo.

Para isso vamos considerar o acoplamento de P_1, P_2, P_3, P_4 como descrito na seção anterior. Primeiro devemos calcular os mínimos das probabilidades de cada estado. Assim

$$m_1 = \min\left\{\frac{2}{12}; \frac{3}{12}; \frac{1}{12}; \frac{4}{12}\right\} = \frac{1}{12},$$
$$m_2 = \min\left\{\frac{5}{12}; \frac{1}{12}; \frac{1}{12}; \frac{3}{12}\right\} = \frac{1}{12},$$
$$m_3 = \min\left\{\frac{2}{12}; \frac{3}{12}; \frac{3}{12}; \frac{4}{12}\right\} = \frac{2}{12},$$
$$m_4 = \min\left\{\frac{3}{12}; \frac{5}{12}; \frac{7}{12}; \frac{1}{12}\right\} = \frac{1}{12}.$$

Procedendo como na seção anterior teremos o acoplamento ilustrado na figura 17.

1	2	3	4	5	6	7	8	9	10	11	12	
x_1	x_2	x_3	x_3	x_4	x_1	x_2	x_2	x_2	x_2	x_4	x_4	$\leftarrow P_1$
x_1	x_2	x_3	x_3	x_4	x_1	x_1	x_3	x_4	x_4	x_4	x_4	$\leftarrow P_2$
x_1	x_2	x_3	x_3	x_4	x_3	x_4	x_4	x_4	x_4	x_4	x_4	$\leftarrow P_3$
x_1	x_2	x_3	x_3	x_4	x_1	x_1	x_1	x_2	x_2	x_3	x_3	$\leftarrow P_4$

Figura 17: Acoplamento das distribuições de transição de uma Cadeira de Markov.

Fazendo agora a simulação das trajetórias usando os mesmos valores sorteados na tabela anterior encontramos os valores da tabela a seguir, que são melhor visualizados na figura 18.

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
U_n	-	7	4	6	9	11	2	6	4	2	4	3	8	5	12
X_n^1	x_1	x_2	x_3	x_3	x_4	x_3	x_2	x_1	x_3	x_2	x_3	x_3	x_4	x_4	x_3
X_n^2	x_2	x_1	x_3	x_3	x_4	x_3	x_2	x_1	x_3	x_2	x_3	x_3	x_4	x_4	x_3
X_n^3	x_3	x_4	x_3	x_3	x_4	x_3	x_2	x_1	x_3	x_2	x_3	x_3	x_4	x_4	x_3
X_n^4	x_4	x_1	x_3	x_3	x_4	x_3	x_2	x_1	x_3	x_2	x_3	x_3	x_4	x_4	x_3

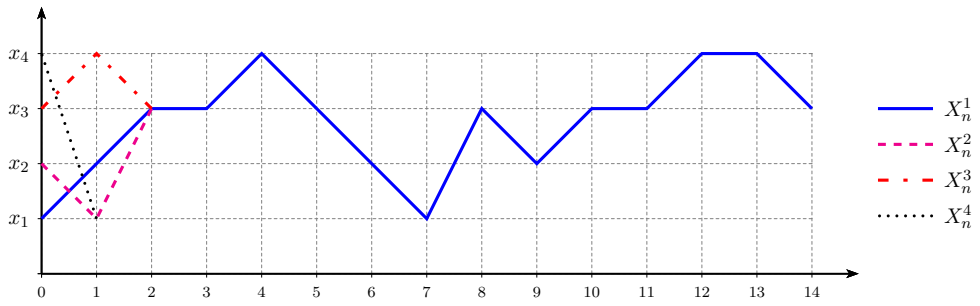


Figura 18: Simulação das trajetórias de uma Cadeia de Markov usando acoplamento "maximal" das transições.

Neste caso as trajetórias se acoplaram já no passo 2 da simulação, que foi exatamente o primeiro passo no qual sorteamos um valor abaixo de 5 (ver figura 17). Teremos assim que para este acoplamento vale que

$$\mathbb{P}(\tau > n) \leq \left(1 - \frac{5}{12}\right)^n = \left(\frac{7}{12}\right)^n,$$

melhorando a estimativa conseguida com o acoplamento simples.

O procedimento exemplificado acima pode ser usado para qualquer cadeia de Markov com transições regulares, nos levando ao seguinte teorema.

Teorema 3.1. *Seja $T = [T_{i,j}]_{m\times m}$ uma matriz de transição regular para uma cadeia de Markov com m estados. Denote por P_k, $k = 1, \ldots, m$, a distribuição de probabilidade representada na k-ésima coluna de T. Nestas condições, se*

$$\beta = \sum_{k=1}^{m} \min\{P_1(x_k); \ldots; P_l(x_k)\} > 0$$

e

$$\tau = \min\{n \geq 0; X_n^1 = \cdots = X_n^m\},$$

então existe um acoplamento $\{X_n^1; \ldots; X_n^m\}, n \geq 0$, *com* $X_0^i = x_i$ *tal que*

$$\mathbb{P}(\tau > n) \leq (1 - \beta)^n.$$

3.3 CONVERGÊNCIA PARA A MEDIDA INVARIANTE

Seja T a matriz de transição de uma cadeia de Markov. Ou seja, cada entrada p_{ij} representa a probabilidade da cadeia ir do estado j para o estado i em um passo. Já vimos que se $p_{ij}^{(n)}$ é o elemento ij da matriz T^n, n-ésima potência da matriz T, então $p_{ij}^{(n)}$ representa a probabilidade da cadeia ir de j para i em n passos.

Já estudamos também o conceito de medida invariante de uma cadeia de Markov. Intuitivamente μ é uma medida invariante se, ao escolhermos o estado inicial de acordo com μ, a probabilidade de estarmos em cada estado em um instante n qualquer é ainda dada por μ. Algebricamente, buscamos o vetor μ tal que

$$\mu = T\mu.$$

Segue que $\mu = T^n\mu$ para todo $n \geq 0$.

A seguir vamos mostrar que, sob certas condições, $p_{ij}^{(n)}$ converge para $\mu(i)$, quando $n \to \infty$, independente do estado j que iniciemos a cadeia.

Vamos antes fixar algumas notações.

Seja X_n^i uma cadeia de Markov para o qual o estado inicial é i. Ou seja,

$$\mathbb{P}(X_n^i = a) = p_{ai}^{(n)},$$

para quaisquer estados a, i e todo $n \geq 1$.

Vale também que, como $\mu = T^n\mu$, então

$$\mu(a) = \sum_{j=1}^{m} p_{aj}^{(n)} \mu(j),$$

para todo estado i e todo $n \geq 1$.

Como μ é um vetor de probabilidades, temos

$$p_{ai}^{(n)} = \sum_{j=1}^{m} \mu(j) p_{ai}^{(n)},$$

para quaisquer estados a, i.

Vamos precisar também da conhecida *desigualdade triangular*. Recordando, ela diz que para quaisquer valores $a_1, \ldots, a_m$, vale

$$|a_1 + \ldots + a_m| \leq |a_1| + \ldots + |a_m|,$$

ou ainda

$$\left| \sum_{k=1}^{m} a_k \right| \leq \sum_{k=1}^{m} |a_k|.$$

Tome agora duas cadeias X_n^i e X_n^j começando em estados distintos i e j, respectivamente. Observe que se em um instante qualquer n tivermos que $X_n^i = a$ e $X_n^j \neq a$, então teremos que $X_n^i \neq X_n^j$. Concluímos assim que

$$\mathbb{P}(X_n^i = a; X_n^j \neq a) \leq \mathbb{P}(X_n^i \neq X_n^j).$$

Se considerarmos agora que as cadeias $X_n^1, X_n^2, \ldots, X_n^m$ estão acopladas de acordo com o acoplamento visto anteriormente, encontramos que

$$\mathbb{P}(X_n^i = a; X_n^j \neq a) \leq \mathbb{P}(X_n^i \neq X_n^j) \leq \mathbb{P}(\tau > n) \leq (1-\beta)^n,$$

para quaisquer estados i, j. Ou seja

$$\mathbb{P}(X_n^i = a; X_n^j \neq a) \leq (1-\beta)^n.$$

Outro ponto importante é perceber que

$$\begin{aligned}
\mathbb{P}(X_n^i = a) &= \mathbb{P}(X_n^i = a; X_n^j \neq a) + \mathbb{P}(X_n^i = a; X_n^j = a) \\
\mathbb{P}(X_n^j = a) &= \mathbb{P}(X_n^j = a, X_n^i \neq a) + \mathbb{P}(X_n^j = a, X_n^i = a).
\end{aligned}$$

E portanto, para quaisquer estados i, j, segue pela desigualdade triangular que

$$\begin{aligned}
\left| \mathbb{P}(X_n^i = a) - \mathbb{P}(X_n^j = a) \right| &= \left| \mathbb{P}(X_n^i = a; X_n^j \neq a) - \mathbb{P}(X_n^j = a, X_n^i \neq a) \right| \\
&\leq \mathbb{P}(X_n^i = a; X_n^j \neq a) + \mathbb{P}(X_n^j = a, X_n^i \neq a) \\
&\leq 2(1-\beta)^n.
\end{aligned}$$

Com isso temos

$$\begin{aligned}
\left|p_{ai}^{(n)} - \mu(a)\right| &= \left|p_{ai}^{(n)} - \sum_{j=1}^{m} \mu(j) p_{aj}^{(n)}\right| \\
&= \left|\sum_{j=1}^{m} \mu(j) p_{ai}^{(n)} - \sum_{j=1}^{m} \mu(j) p_{aj}^{(n)}\right| \\
&= \left|\sum_{j=1}^{m} \mu(j)(p_{ai}^{(n)} - p_{aj}^{(n)})\right| \\
&\leq \sum_{j=1}^{m} \mu(j) \left|p_{ai}^{(n)} - p_{aj}^{(n)}\right| \\
&\leq \sum_{j=1}^{m} \mu(j) \left|P(X_n^i = a) - \mathbb{P}(X_n^j = a)\right| \\
&\leq 2 \sum_{j=1}^{m} \mu(j)(1-\beta)^n \\
&\leq 2(1-\beta)^n, \qquad (3.1)
\end{aligned}$$

e portanto, se $\beta > 0$

$$\left|p_{ai}^{(n)} - \mu(a)\right| \longrightarrow 0, \text{ quando } n \to \infty,$$

para todo i e todo a.

Mostramos assim o seguinte resultado.

Teorema 3.2. *Seja $T = [T_{i,j}]_{m\times m}$ uma matriz de transição regular para uma cadeia de Markov com m estados. Denote por P_k, $k = 1, \ldots, m$, a distribuição de probabilidade representada na k-ésima coluna de T. Nestas condições, se*

$$\beta = \sum_{k=1}^{m} \min\{P_1(x_k); \ldots; P_m(x_k)\} > 0,$$

então existe uma única medida invariante μ para T e, se $p^{(n)}$ é a sequência definida por

$$p^{(n)} = T \cdot p^{(n-1)},$$

vale que

$$p^{(n)} \to \mu, quando$$

$n \to \infty$,

$$paraqualquerescolhadevetordeprobabilidades$$

$p \hat{} (0)$.

Para terminar vamos verificar que a matriz G definida em (1.4), que é usada pelo Google para determinar a importância das páginas da web, satisfaz as condições do teorema 3.2.

Como já comentamos anteriormente, G é claramente regular.

Para mostrar que $\beta > 0$ bastaria notar que β é sempre maior que a menor entrada da matriz, que no caso de G são todas positivas. Mas vamos conseguir uma estimativa um pouco melhor. Note primeiro que, da equação (1.4), temos

$$G_{ij} = \alpha S_{ij} + \frac{1-\alpha}{N},$$

onde G_{ij} e S_{ij} são as entradas (i, j) das matrizes G e S respectivamente, N é o total de páginas na internet e α é a probabilidade de um internauta continuar navegando de acordo com a matriz S. Logo

$$G_{ij} \geq \frac{1-\alpha}{N}.$$

Segue que

$$\min\{G_{k1}; \ldots; G_{kN}\} > \frac{1-\alpha}{N},$$

para todo k, e portanto

$$\beta = \sum_{k=1}^{N} \min\{G_{k1}; \ldots; G_{kN}\} > \sum_{k=1}^{N} \frac{1-\alpha}{N} = 1-\alpha.$$

Fica como desafio para o leitor perceber que, de fato,

$$\min\{G_{k1}; \ldots; G_{kN}\} = \frac{1-\alpha}{N},$$

e portanto

$$\beta = 1-\alpha.$$

Assim, segue dos cálculos em (3.1) que para todo site s_i,

$$|r_i^n - r_i| < \alpha^n.$$

E se $\alpha = 0,85$ como comentado em [1], então

$$|r_i^n - r_i| < 0,85^n,$$

e com menos de 60 iterações a precisão estará além da quarta casa decimal.

BIBLIOGRAFIA

[1] David Austin, *How Google Finds Your Needle in the Web's Haystack*, `http://www.ams.org/samplings/feature-column/fcarc-pagerank`.

[2] Michael W. Berry and Murray Browne, *Understanding search engines : mathematical modeling and text retrieval*, Software, environments, tools, Society for industrial and applied mathematics, Philadelphia (Pa.), 1999.

[3] J.L. Boldrini, *Algebra linear*, HARBRA, 1986.

[4] Sergey Brin and Lawrence Page, *The anatomy of a large-scale hypertextual Web search engine*, Computer Networks and ISDN Systems **30** (1998), no. 1–7, 107–117.

[5] Kurt Bryan and Tanya Leise, *The $25,000,000,000 eigenvector: the linear algebra behind google*, SIAM Review **48** (2006), 569–581.

[6] J.A. Ferrari, P.A. e Galves, *Acoplamento e processos estocásticos*, IMPA, 1997.

[7] Taher Haveliwala and Sepandar Kamvar, *The second eigenvalue of the google matrix*, Technical Report 2003-20, Stanford InfoLab, 2003.

[8] Amy N. Langville and Carl D. Meyer, *Google's pagerank and beyond: the science of search engine rankings*, Princeton University Press, Princeton, NJ, 2006.

[9] Sheldon M. Ross, *Introduction to probability models*, Probability and Statistics, Academic Press, 2007.

[10] _____, *Probabilidade: um curso moderno com aplicações*, 8 ed., Bookman, Porto Alegre, 2010.

[11] V.L. Ruggiero, M.A.G. e da Rocha Lopes, *Cálculo numérico: aspectos teóricos e computacionais*, Makron Books do Brasil, 1996.