UM MÉTODO ESTOCÁSTICO DE ALOCAÇÃO DE MEMÓRIA

JAYME LERNER

Tese submetida ao corpo docente da coordenação dos programa de Pós-Graduação de Engenharia da Universidade Federal do Rio de Janeiro como parte dos requisitos necessários para obtenção do grau de mestre em Ciência (M.Sc.)

APROVADA POR

Nelson Maculan Filho
PRESIDENTE

RIO DE JANEIRO
ESTADO DA GUANABARA - BRASIL

NOVEMBRO DE 1973

À

Tereza e Alfredo Degens.

Meus sinceros agradecimentos à

COPPE-UFRJ, e aos Professores:

Nelson Maculan Filho

Célio Guimarães

Jack Schechtman

Flávio R. Velasco

Demétrio Alonso Ribeiro

Maurício Matos Peixoto

João Lizardo R.H. de Araújo

R E S U M O

Apresentamos o desenvolvimento da teoria de processos estocásticos aplicada a sistemas de paginação, e ao final do estudo formulamos um algoritmo de paginação.

Primeiro tomamos algumas propriedades de álgebra, e com elas desenvolvemos a teoria necessária à interpretação matemática das cadeias de páginas que ligam duas páginas do programa no computador.

A seguir notando a aleatoriedade como fator preponderante na estrutura desenvolvida, apresentamos alguns resultados que explicam a lei de formação dessas cadeias. Então introduzimos os conceitos de tempo médio de absorção, tempo médio de primeira passagem, desenvolvendo resultados que julgamos serem necessários à evolução natural do estudo.

Aplicamos esses resultados a um processo de markov particular que é o *random walk* no círculo, com o propósito de obter informações sobre as leis de formação de cadeias para esse caso particular, objetivando ter condições que nos permitirão formular um algoritmo de paginação.

A B S T R A C T

We introduce the development of the theory of stochastic process as applied to paging systems, and at the end of the study we formulate a paging algorithm.

First we borrow some properties of algebra, and with these resuts we develop the necessary theory to the mathematical interpretation of the chains of pages that link two given pages of the program in the computer.

Then noticing randomnes as a main factor in the developed structure, we introduce some results that explain the law of formation of these chains, Thereafter, we introduce the concepts of mean first passage time, mean absorption, time, and results we deem necessary to the studies natural evolution.

We apply these results to a particular markov chain namely the random walk in the circle, aiming to get informations about the chain formation laws for this particular case, results whereon we will be able to formulate a paging algorithm.

INDICE

1. INTRODUÇÃO

Desde os primórdios da computação eletrônica tem sido reconhecido que, devido ao fato de que memórias de acesso rápido serem muito caras, computadores de grande memória, precisam tê-las organizadas hierarquicamente, com pelo menos dois níveis, os quais não chamados *memória principal* e *memória auxiliar*.

Uma informação do programa (i.é. código de instrução e dados) somente podem ser referenciados quando ela reside na memória principal, assim a informação que tem a maior probabilidade de ser referenciada deve residir na memória principal, e as outras informações devem residir na memória auxiliar.

O *problema de alocação de memória* é aquele de determinar, em cada instante do tempo, como a informação deveria ser distribuída pelos níveis de hierarquia da memória.

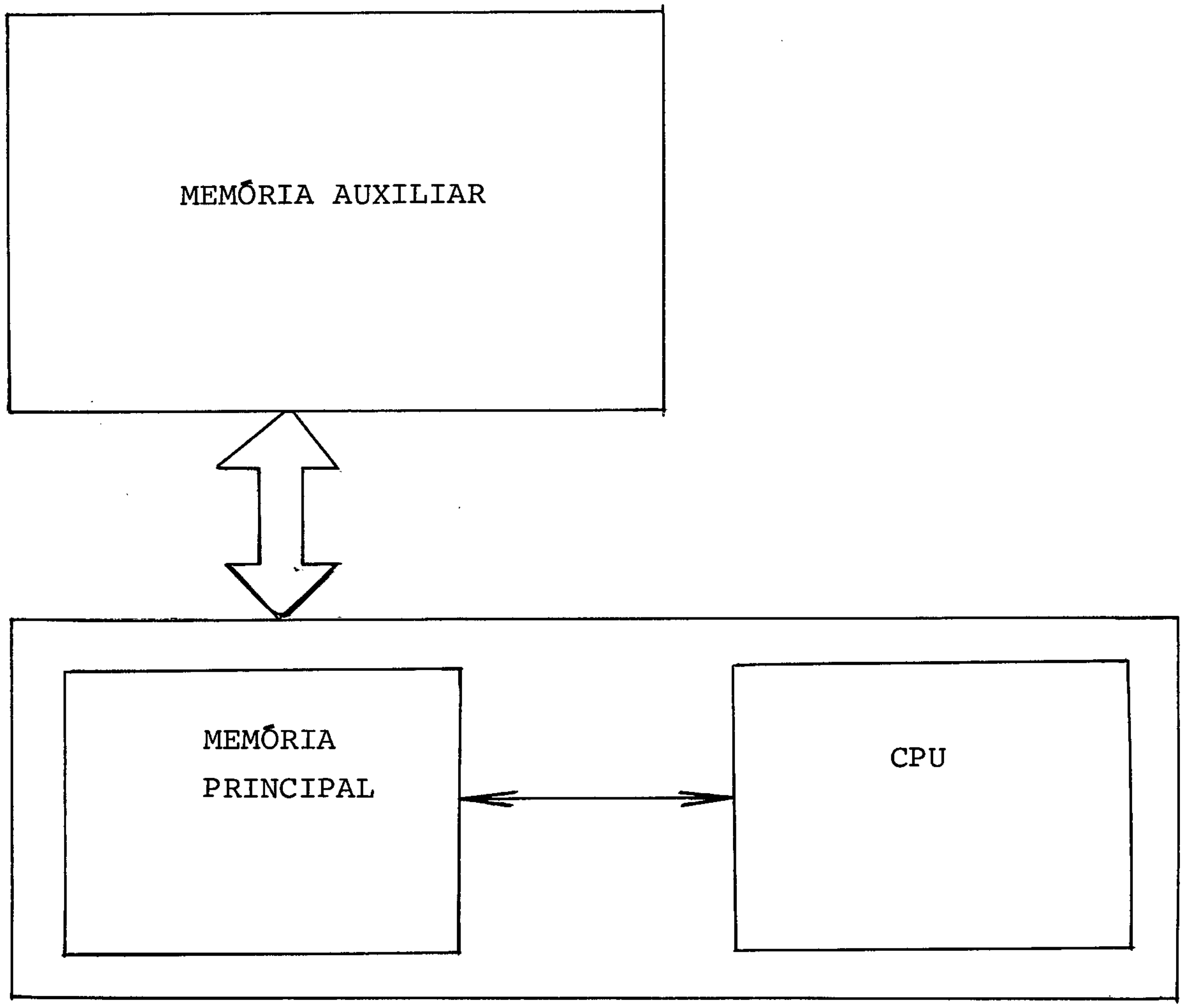
Durante os primeiros anos da computação eletrônica cada programador tinha que incorporar procedimentos de alocação no seu programa, sempre que a totalidade de informação excedia o tamanho da memória principal. Esses procedimentos eram relativamente diretos, e se resumiam em dividir o programa em uma sequência de segmentos os quais iriam alternar residência na memória principal na medida que fosse necessário.

Isso começou a mudar depois da introdução das linguagens de programação de alto nível nos anos cinquenta. Programadores foram encorajados serem mais orientados para a resolução de problemas do que com detalhes da máquina.

À medida que os programas cresciam em tamanho, crescia em tamanho os problemas de alocação de memória. Realmente em princípio dos anos sessenta, ficou claro que a eficiência do computador poderia sofrer muito, como consequência de políticas de alocação de memória que fossem inadequadas, ficou então claro que o problema de alocação de memória tinha se tornado de importância capital.

Vide Figura 1.

Figura 1



Nós nos fixamos então na seguinte imagem: temos um programa P, cujo tamanho excede em muito o tamanho da memória principal de modo que teríamos que dividir o programa funcionalmente em partes, as quais assumimos serem todas de mesmo tamanho, em princípio é claro que colocaríamos o maior número dessas partes, que chamaremos de *páginas* na memória principal, e o restante na memória auxiliar. Um mecanismo intrínseco ao sistema teria no momento que fosse necessária uma página i que estivesse na memória auxiliar, localizá-la, determinar uma página j na memória principal, e efetuar a troca de residência de i e j, para que as instruções da página i pudessem então ser executadas.

Das diversas perguntas que poderiam ser feitas nos preocupamos tão somente com uma, qual seja:

Qual seria a política ótima a adotar na troca de páginas?

Uma resposta intuitivamente probabilista seria de se esperar,pois iria depender das propriedades probabilistas do software do sistema.

Para o problema que vamos estudar, tomaríamos por hipóteses:

1) O nosso programa teria sido dividido em m páginas,numeradas de 1 até m. m primo .

2) De qualquer página i que estivesse sendo referenciada na memória principal, o mecanismo intrínseco ao sistema iria passar a referenciar a página i+1 com probabilidade p, ou a página i-1 com probabilidade q. No caso particular de i=1 então iria referenciar i+1 com probabilidade p ou m com probabilidade q, ou se i=m então iria referenciar 1 com probabilidade p ou i-1 com probabilidade q.

Qual seria a política ótima a adotar para troca de páginas, e sob que condições ela é válida?

Ótima no sentido que durante a execução do programa, o mínimo número de trocas ocorra.

Para isso somos intuitivamente levados a querer que nossa política ótima de troca de páginas, ao ser pedida uma página da memória auxiliar, implique a substituição da página do buffer que tiver a menor probabilidade de ser pedida no futuro.

Podemos então começar a desenvolver a estrutura matemática baseada na qual formularemos nossa política de troca de páginas.

Notemos que nosso objetivo é desenvolver uma estrutura a partir da qual se possam basear no futuro, pesquisas que resolvam problemas cujas hipóteses sobre o mecanismo intrínseco ao sistema sejam outras que não as formuladas, mas essencialmente probabilistas.

2. ALGUNS RESULTADOS ELEMENTARES

Tomaremos em nosso estudo a noção de *conjunto* como intuitiva. Uma coleção de objetos aos quais chamaremos de *elementos* do conjunto.

Sejam A e B dois conjuntos. Produto cartesiano de dois conjuntos A e B é a coleção de pares ordenados (a,b) tal que $a \in A$ e $b \in B$. Esse novo conjunto será denotado por A x B.

Qualquer subconjunto R de AxB será por nós chamado de *relação binária entre os conjuntos A e B*, e diremos que se $(a,b) \in R$, a tem relação R com b, escrevendo aRb.

Quando A=B, R será chamada simplesmente de *relação R em A*.

A relação R em A será *reflexiva* se e só se xRx, para todo x pertencente a A.

A relação R em A será *simétrica* se e só se sempre que xRy for verdadeira então yRx também for, para todo x,y pertencentes a A.

A relação R em A será *transitiva* se e só se sempre que xRy e yRz forem verdadeiras, então xRz também o for, para todos x,y,z pertencentes a A.

Uma relação R em A que é transitiva, simétrica e reflexiva será chamada de *relação de equivalência R em A*.

TEOREMA 2.1.

Uma relação de equivalência R em A particiona A em subconjuntos dois a dois dijuntos, e cuja união é A.

PROVA:

Se $a \in A$ então seja $[a] = \{ b | bRa \}$
temos que provar que $A = \cup [a]$, Va pertencente a A.

Como R é reflexiva $a \in [a]$ logo $[a] \neq \phi$

Suponhamos que $[a]$ interseção $[b] \neq \phi$, e seja então c pertencente a essa interseção, logo cRa e cRb logo aRb e bRa pois R é transitiva e simétrica.

Se $d \in [a]$ então dRa logo dRb pois aRb e R é transitiva então d pertence a $[b]$ e assim $[a]$ está contido em $[b]$;
Raciocinando de modo inverso $[b]$ está contido em $[a]$ e então $[a]=[b]$.

c.q.d.

Cada um desses subconjuntos determinados por R será chamado de *classe de equivalência*, e será denotado por $[a]$ onde a pertence à classe de equivalência $\{ b | bRa\}$.

Uma relação R que é reflexiva e transitiva é chamada de *relação de ordem fraca*.

Ilustremos o porque dessa definição:

Dados a,b pertencentes a A, e a relação de ordem fraca R.

Se aRb e bRa a e b coincidem em algum sentido.

Se aRb e $b\not R a$ a está na frente de b em algum sentido.

Se $a\not R b$ e bRa b está na frente de a em algum sentido.

Se $a\not R b$ e $b\not R a$ impossível comparar a e b em algum sentido.

TEOREMA 2.2.

Se R for uma relação de ordem fraca então a relação R' tal que (aR'b se e só se simultaneamente aRb e bRa) é uma relação de equivalência determinada por R.

PROVA: trivial!

Se R for uma relação de ordem fraca, e a relação de equivalência determinada por R é a identidade então R será chamada *relação de ordenação parcial*.

Intuitivamente dois elementos não são coincidentes em algum sentido.

Se R é uma relação de ordem fraca definida em A, então seja R* uma relação definida: da seguinte maneira no conjunto das classes de equivalência de A .

(a) R* (b) se e só se $\forall a' \varepsilon (a)$, a'Rb' - $\forall b' \varepsilon (b)$.

Definida dessa maneira R* é uma relação de ordem parcial chamada *relação de ordem parcial induzida por R em A.*

TEOREMA 2.3.

Um conjunto de inteiros positivos que é fechado em relação a adição contém a menos de um número finito, todos os múltiplos de seu máximo divisor comum.

3. UMA ESTRUTURA PARA O CONJUNTO DE PÁGINAS

Suponhamos que temos um conjunto de m páginas que compõem um programa P em um computador, e que este possui um mecanismo intrínseco que "a pedido" de uma página i, procura e acha uma outra página j desse programa.

Convencionamos então que $i \to j$ significava que a página i pede a página j no seguinte sentido:

(1) $i \to j$ *diretamente*

(2) $i \to j$ *indiretamente* ié $i \to i_1 \to i_2 \to \quad \to i_n \to j$,

com $i, i_1, i_2, i_3, \ldots, i_n, j \in P$.

Chamaremos uma sequência $(i, i_1, i_2, \ldots, i_n, j)$ de *caminho orientado*, ou simplesmente *caminho de* i *a* j.

Notemos que:

(1) $i \rightarrow i$

(2) Se $i \rightarrow j$ e $j \rightarrow k$ então $i \rightarrow k$ $i,k,j \varepsilon$ P.

(3) Se $i \rightarrow j$ então não necessariamente $j \rightarrow i$.

Mas (1) e (2) nos garantem que $\rightarrow$ é uma relação de ordem fraca.

Ainda notamos que a relação de equivalência determinada por $\rightarrow$ é

$$(i \rightarrow j \quad e \quad j \rightarrow i).$$

pois:

(1) $i \rightarrow i$ e $i \rightarrow i$

(2) Se ($i \rightarrow j$ e $j \rightarrow i$) então ($j \rightarrow i$ e $i \rightarrow j$).

(3) Se ($i \rightarrow j$ e $j \rightarrow i$) e ($j \rightarrow k$ e $k \rightarrow j$) então ($i \rightarrow k$ e $k \rightarrow i$)

Por meio dessa relação de equivalência, particionamos o conjunto P em classes de equivalência, cujo conjunto denotamos por P.

Duas páginas do programa P estarão na mesma classe de equivalência se e só se uma puder pedir a outra, i.é. ($i \rightarrow j$ e $j \rightarrow i$).

A ordenação parcial $\overset{*}{\rightarrow}$ induzida por $\rightarrow$ no conjunto das classes de equivalência terá o seguinte significado:

Seja $\lceil i \rceil$ e $\lceil j \rceil \varepsilon P$ então $\lceil i \rceil \overset{*}{\rightarrow} \lceil j \rceil$ se e só se $\forall\, i_1 \varepsilon \lceil i \rceil$,tivermos

$$i_1 \rightarrow j_1 \ \forall\, j_1 \varepsilon \lceil j \rceil$$

mas não conversamente, a menos que $\lceil i \rceil = \lceil j \rceil$

Agora um elemento $\lceil i \rceil$ do conjunto P das classes de equivalência determinadas por $\rightarrow$ será chamado *maximal* se:

para todo $\lceil j \rceil \varepsilon P$ tal que $\lceil j \rceil \overset{*}{\rightarrow} \lceil i \rceil$ então $\lceil i \rceil \overset{*}{\rightarrow} \lceil j \rceil$ ou seja $\lceil j \rceil = \lceil i \rceil$, ou equivalentemente se e só se não existir $\lceil k \rceil \varepsilon P$ tal que $\{k\} \overset{*}{\nrightarrow} \{i\}$ e $\{k\} \neq \{i\}$, iē se as pāginas de $\{i\}$ não puderem ser pedidas por nenhuma pāgina de outra classe de equivalência qualquer. Concluīmos então que elementos de conjuntos maximais sō podem pedir pāginas de outros conjuntos.

Um elemento $\{i\}$ do conjunto P das classes de equivalência determinadas por $\rightarrow$ serā chamado minimal se por definição $\forall \{j\} \in P$ tal que $\{i\} \overset{*}{\rightarrow} \{j\}$ então $\{j\} \overset{*}{\rightarrow} \{i\}$ ou seja $\{j\} = \{i\}$ ou equivalentemente se e sō se não existir $\{k\} \in P$ tal que $\{i\} \overset{*}{\rightarrow} \{k\}$ e $\{k\} \neq \{i\}$ iē se as pāginas de $\{i\}$ não puderem pedir pāginas de outras classes de equivalência.

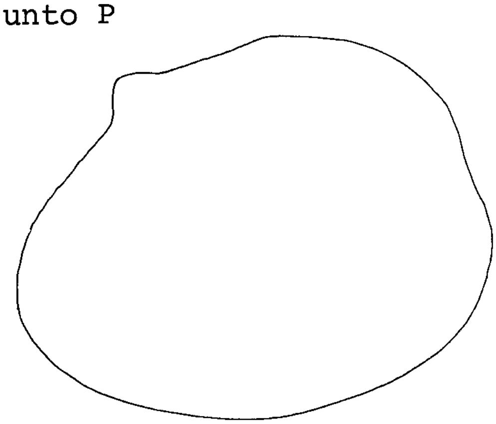
Realmente se existisse $\{j\} \in P$ tal que exista $j' \in \{j\}$ tal que $i' \rightarrow j'$ onde $i' \in \{i\}$, então $i' \rightarrow j''$ $\forall j'' \in \{j\}$ pois $j' \rightarrow j''$ $\forall j'' \in \{j\}$, mas então como $\forall i'' \in \{i\}$ $i'' \rightarrow i'$, vem que $\forall i'' \in \{i\}$, $i'' \rightarrow j''$ $\forall j'' \in \{j\}$ ou seja $\{i\} \overset{*}{\rightarrow} \{j\}$ logo $\{i\} = \{j\}$.

Concluīmos então que elementos de conjuntos minimais sō podem ser pedidos por outros elementos de outros conjuntos.

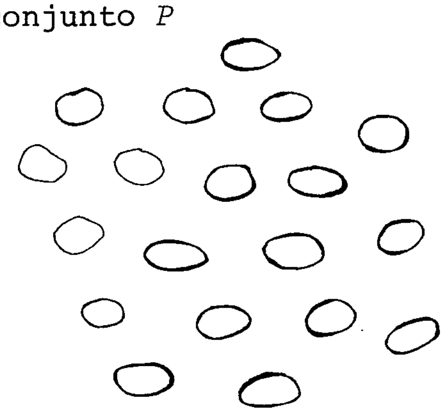
Vide Figura 2.

Figura 2

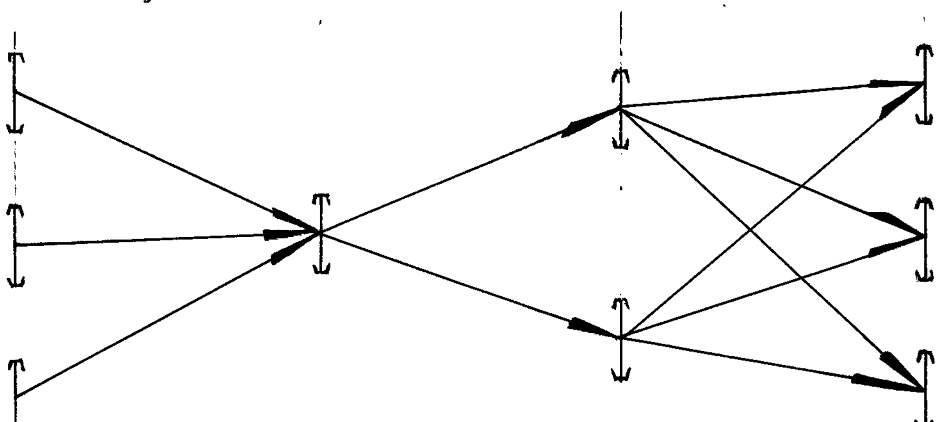
Tínhamos um conjunto P



Obtivemos um conjunto *P*



com a relação de ordem



maximais $\xrightarrow{*}$ minimais

Passemos agora ao estudo de uma classe de equivalência $\lceil i \rceil$ de P, lembrando que $\forall\ j$ e $k\ \varepsilon \lceil i \rceil$, $j \to k$ e $k \to j$.

O mecanismo intrínseco ao sistema ao receber de uma página o pedido de outra página, estabelecerá uma cadeia direta ou indiretamente entre a página que pede e a página pedida.

Como unidade de tempo tomamos o intervalo de tempo que leva para o sistema procurar e encontrar uma página pedida, diretamente.

Esse intervalo de tempo será por nós chamado de *transição*.

Seja então $M(i,j)$ o conjunto de m's tais que depois de m transições a página pedida por i é j .

Em particular $M(i,i)$ é o conjunto de intervalos de tempo, medidos na nossa unidade, que é a transição, que a página i pode pedir a si mesma.

Notemos que $M(i,i)$ é fechado em relação à adição pois se $l, n\ \varepsilon\ M(i,i)$ então $i \to i$ em l transições e $i \to i$ em n transições logo $i \to i$ em $l + n$ transições, logo $M(i,i)$ contém todos os múltiplos do seu máximo divisor comum $c(i)$ a menos de um número finito, ou seja $M(i,i)$ tem todos os seus elementos múltiplos de $c(i)$.

Notemos que $M(i,i) \neq \phi$ pois $i \to j$ $\forall i,j\ \varepsilon\ \lceil i \rceil$

TEOREMA 3.1.

(1) Se $\lceil i \rceil = \lceil j \rceil$ então $c(i) = c(j) = c$

(2) Se $l,n\ \varepsilon\ M(i,j)$ então $l \equiv n \pmod{C}$

PROVA:

Notemos que se $l \varepsilon M(i,j)$ e $n \varepsilon M(i,j)$ - e $q \varepsilon M(j,i)$ então um possível caminho para $i \to i$ seria $i \to j$ e $j \to i$ em l e q transições respectivamente, e portanto $l + q \varepsilon M(i,i)$.

Outro caminho possível seria $i \to j$ e $j \to j$ e $j \to i$ em l transições, $K.c(j)$ transições com K suficientemente grande, e q transições respectivamente, e portanto $l + K.c(j) + q \varepsilon M(i,i)$ logo $c(j)$ deve ser múltiplo de $c(i)$ pois $l + q = K_1.(c_i)$ mas

$l + K.c(j) + q = K_2.c(i)$.

Dessas duas expressões concluímos que $K\ c(j) = (K_2-K_1)\ c(i)$, então $c(j)$ é múltiplo de $c(i)$.

Mas da mesma maneira mostraríamos que $c(i)$ é múltiplo de $c(j)$, e a conclusão que segue é que $c(i) = c(j) = c$, provando (1).

Agora $i \to i$ com um possível caminho estabelecido da seguinte maneira $i \to j$ em n transações, $j \to i$ em q transições, logo $n + q \varepsilon M(i,i)$ então c divide $n+q$ e c divide $l+q$ consequentemente c divide $l-n$ ou $l \equiv n$ (mód.c) $\forall n,l \varepsilon M(i,j)$.

c.q.d.

Seja $M_o = \min \{ n | n \varepsilon M(i,j)\}$ e $r(i,j)$ definido como

$r(i)j\ ) = \text{resto}\ (\frac{M_o}{c})$, então $0 \leqslant r(i,j) < c$.

Todos os elementos de $M(i,j)$ são congruentes a $r(i,j)$ pois se $l \varepsilon M(i,j)$ então $l-M_o = K_1.\ c$, consequência de (2) do teorema anterior.

Mas $M_o = K_r\ .\ c + r(i,j)$, então $l - r(i,j) = K.c$ ou seja

$l \equiv r(i,j)$ (mód. c).

Em particular notemos que $r(i,i) = 0$, pois todos os elementos de $M(i,i)$ são múltiplos de c conforme notamos anteriormente.

TEOREMA 3.2.

Na classe de equivalência $\lceil i \rceil$ de P temos definida uma relação de equivalência R que é a seguinte:

$(i,j) \in R$ se e só se $r(i,j) = 0$.

PROVA:

(1) Já notamos que $r(i,i) = 0$, iRi .

(2) Notemos que $r(i,j) + r(j,i) \equiv 0 \pmod{c}$

pois se $l \in M(i,j)$ então $l \equiv r(i,j) \pmod{c}$

$n \in M(j,i)$ então $n \equiv r(j,i) \pmod{c}$

Então $l + n \in M(i,i)$ logo $l+n \equiv 0 \pmod{c}$.

Se iRj então $r(i,j) = 0$ logo $r(j,i) \equiv 0 \pmod{c}$ e como $0 \leqslant r(j,i) < c$ temos que $r(j,i) = 0$ e assim jRi .

(3) Notemos que $r(i,k) \equiv r(i,j) + r(j,k) \pmod{c}$

pois se $l \in M(i,j)$ então $l \equiv r(i,j) \pmod{c}$

se $n \in M(j,k)$ então $n \equiv r(j,k) \pmod{c}$

então $l+n \in M(i,k)$ logo

$l+n \equiv r(i,k) \pmod{c}$

é consequência dessas últimas três equações

$r(i,j) + r(j)k) \equiv r(i,k) \pmod{c}$

Então se $r(i,j) = 0$ e $r(j,k) = 0$ essa última relação nos diz que $r(i,k) = 0$ ou que iRj e jRk implica iRk .

Chamaremos uma classe de equivalência determinada por R em $\lceil i \rceil$ de classe cíclica, (e a denotaremos ⌈ (i) ⌉).

As perguntas que fazemos agora são as seguintes:

Quantas classes cíclicas existem em (i) ?

Como são os caminhos contidos em (i) ?

Suponhamos que c é grande para não perdermos a generalidade. Notemos então: Seja i uma página de uma classe cíclica qualquer.

Em uma transição não é possível nenhuma página de ⌈ (i) ⌉ diretamente pois a propriedade que caracteriza todos os elementos i,j de ⌈ (i) ⌉ é que r(i,j) = 0 ié de i só podemos alcançar j em intervalos de tempo múltiplos de c o qual supomos ser bastante grande, então na primeira transição estaremos em outra classe cíclica de (i) , repetimos então o raciocínio, em uma transição não podemos referenciar nenhuma página da classe cíclica de i e tão pouco nenhuma página dessa classe cíclica, vamos então para uma terceira classe cíclica, repetindo o mesmo raciocínio chegamos à conclusão que em uma transição iremos para uma quarta classe cíclica, até que chegamos a uma nésima classe cíclica em n transições, cada transição para uma classe cíclica diferente das anteriores, onde $n < c$.

Repetimos então o raciocínio, com mais uma transição não podemos voltar a nenhuma késima classe cíclica das n enumeradas até agora, pois $(n-k) + 1 < c$ e $(n-k) + 1$ seria o número de transições para ir do elemento i_k referenciado na k ésima transição ao elemento referenciado na (n+1) ésima transição que se pertencesse a ⌈ (i_k) ⌉ precisaria de no mínimo c transições para ser referenciado a partir de i_k.

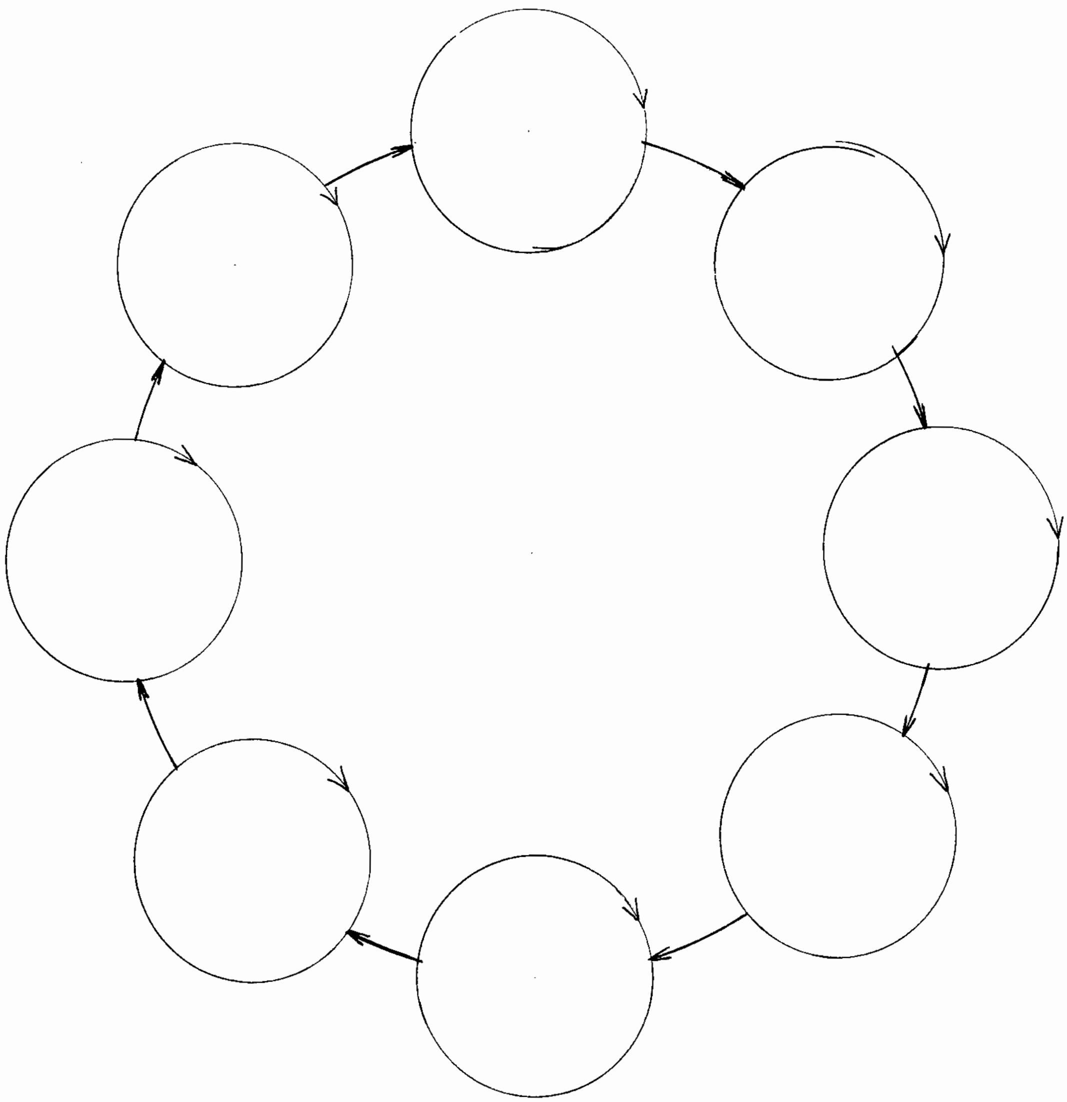
Então podemos afirmar que uma cadeia que se origina em um elemen to i da classe cíclica $((i))$, até um elemento $j \in ((j))$ e tal que ela tem comprimento n, só pode estar na cíclica $((j))$ depois de exatamente n transições.

Agora n é arbitrário e c é finito de modo que **voltamos** então a $((i))$, de maneira que percorreremos exatamente c classes cíclicas.

Respondemos assim as perguntas, e ilustramos graficamente.

Vide Figura 3.

Figura 3



A cadeia é cíclica, no sentido que ela vai de uma classe cíclica $((i))$ percorre todas as outras cíclicas e na césima transição ela retorna a classe cíclica de partida $((i))$, de modo que o ciclo tem comprimento C.

Notemos o seguinte fato não trivial, quando $c = 1$, então temos somente uma classe cíclica que é igual ao conjunto $\{i\}$, e nesse caso depois de um intervalo de tempo suficientemente grande qualquer página $i \in \{i\}$ pode pedir qualquer página $j \in \{i\}$, em um número de transições que independe das páginas do conjunto $\{i\}$.

4. ADIÇÃO DE ESTOCASTICIDADE

Dado o fato que uma página i pede outra página j, i.é. existe uma cadeia de i até j . Perguntaríamos qual seria a página pedida em uma mésima transição nessa cadeia?

Essa pergunta é válida pois lembramos que entre uma página i que pede e uma página j pedida existem vários caminhos possíveis, de modo que poderíamos ter várias respostas, e a conclusão é que esse problema não é determinístico. Reformularíamos a pergunta para:

Qual seria a probabilidade de ter uma determinada página pedida por i na mésima transição?

Para respondê-la precisaríamos de uma estrutura probabilística.

Faremos para conseguí-la, duas hipóteses:

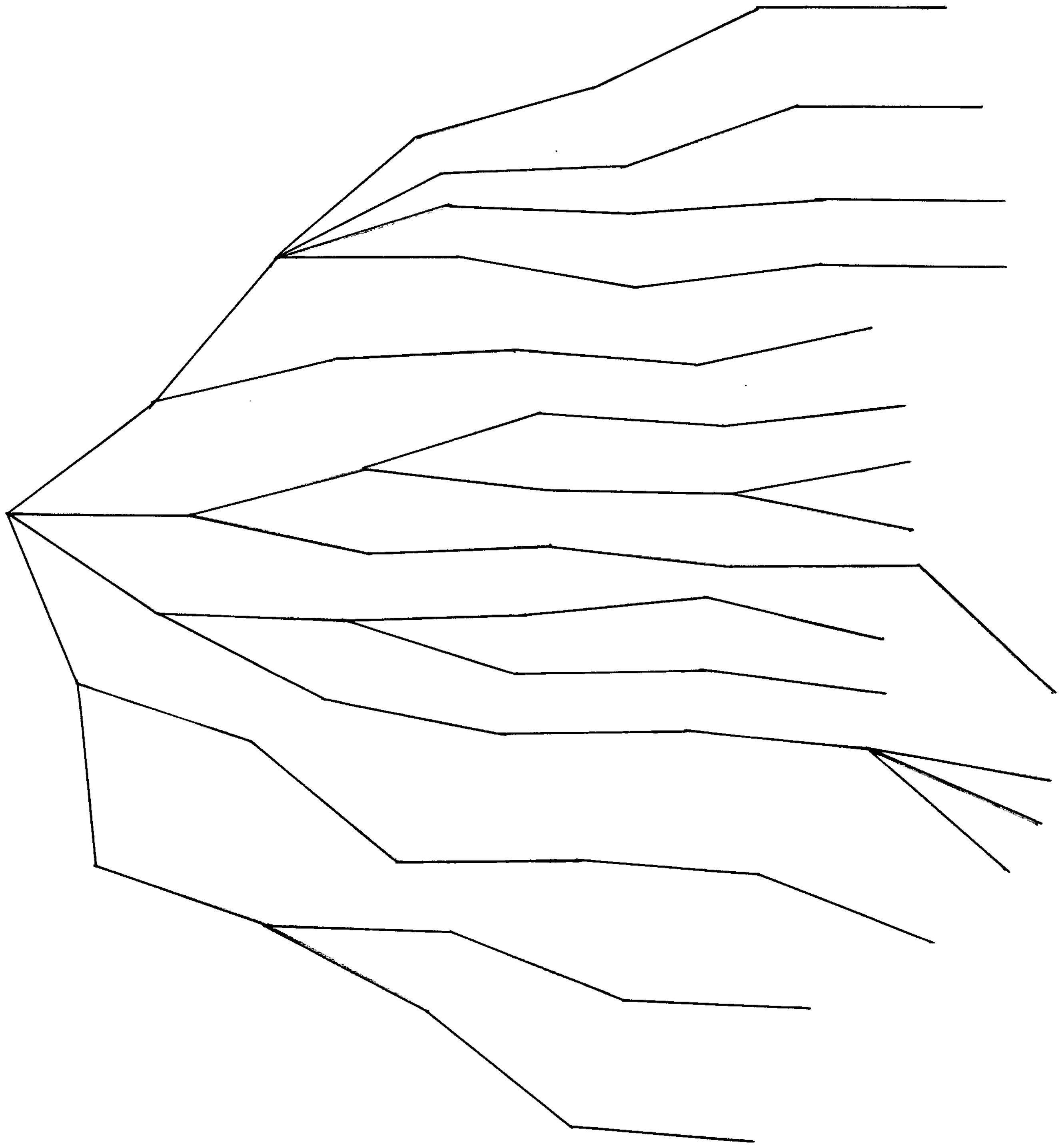
(1) A página a ser referenciada na mésima transição poderia depender das páginas que foram referenciadas anteriormente no caminho de i até ela.

(2) A probabilidade que certa página seja chamada na mésima transição pode ser calculada quando as páginas referenciadas até a (m - 1) ésima transição são conhecidas.

É intuitivo que podemos representar todos os possíveis caminhos em uma árvore, a qual chamaremos árvore de m transições.

Vide figura 4.

Figura 4

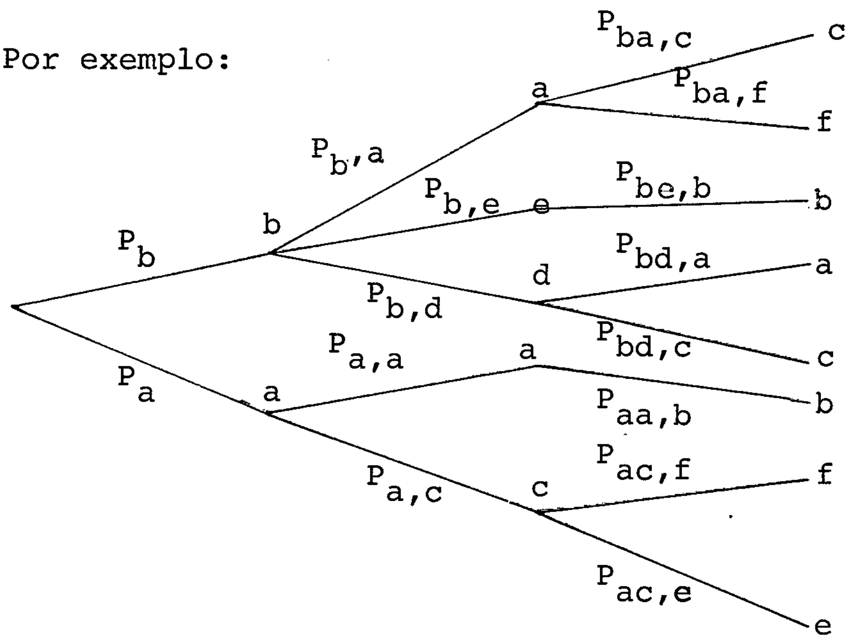


A cada ramo associaremos um número que é a probabilidade da página determinada na origem do ramo pedir *diretamente* a página no final do ramo.

Agora a cada caminho, associamos um número que é o produto das probabilidades associadas a cada ramo do caminho. Notamos então que a soma desses números para todos os caminhos de m transições é 1 .

Vide Figura 5

Por exemplo:



Assumimos que esses números P's são dados de acordo com as hipóteses formuladas. $P_{aa,b}$ = probabilidade de ocorrer b na 3ª transição dado que ocorreram a na 1ª transição e a na 2ª transição. O fato de que sejam probabilidades para as transições nos diz que os P's são positivos e que:

$P_a + P_b = 1$, $P_{b,a} + P_{b,e} + P_{b,d} = 1$, $P_{bd,a} + P_{bd,c} = 1$ etc...

como dissemos o número associado a um caminho é uma probabilidade associada àquele caminho. Computando a soma dos números associados a todos os caminhos vem que: $P_b \cdot P_{ba}\, P_{ba,c} + P_b \cdot P_{ba} \cdot P_{ba,f} + P_b \cdot P_{be} \cdot P_{be,b} + P_b \cdot P_{bd} \cdot P_{bd,a} + P_b \cdot P_{bd} \cdot P_{bd,c} + P_a \cdot P_{a,a} \cdot P_{aa,b} + P_a \cdot P_{ac} \cdot P_{ac,f} + P_a \cdot P_{a,c} \cdot P_{ac,e} = P_b \cdot P_{ba} + P_b \cdot P_{be} + P_b \cdot P_{bd} + P_a \cdot P_{aa} + P_a \cdot P_{ac} = P_b (P_{ba} + P_{be} + P_{bd}) + P_a (P_{aa} + P_{ac}) = P_a + P_b = 1$

e assim mostramos que realmente temos uma probabilidade associada com os caminhos

Então podemos falar em probabilidade de um caminho orientado ou um caminho de m transições.

Seja então uma função X_n cujo domínio é o conjunto de caminhos de uma árvore de m transições, e cujo valor em um caminho de m transições é a página pedida na mésima transição.

Chamaremos essas funções X_n de *variáveis aleatórias*, e ao conjunto $\{X_1, X_2, \ldots, X_m\}$ chamamos de *Processo Estocástico*.

A relação entre as probabilidades associadas aos ramos e às variáveis aleatórias do processo estocástico seria:

probabilidade na 1ª transição:

$$P_r \{X_1 = i_q\}$$

probabilidade na 2ª transição:

$$P_r \{X_2 = i_q \mid X_1 = i_r\}$$

probabilidade na 3ª transição:

$$P_r \{X_3 = i_q \mid X_1 = i_r , X_2 = i_s\}$$

e assim por diante.

Como os contradomínios de todas as variáveis aleatórias são subconjuntos do conjunto de m páginas do programa P, esse processo estocástico será chamado de *processo estocástico finito*.

Notemos então que nós temos uma árvore, uma probabilidade associada a cada caminho orientado, e uma sequência de funções

$\{X_1, X_2, X_3, \ldots \}$, cujo domínio de cada X_m é a subárvore até a mésima transição inclusive, e cujo contradomínio é o conjunto de páginas que podem ser pedidas na mésima transição.

Estamos particularmente interessados num processo estocástico finito com a seguinte propriedade:

$$P_r \{X_m = iq \mid X_{m-1} = i_n, X_{m-2} = is, \ldots, X_1 = i_t\} =$$

$$= P_r \{X_m = iq \mid X_{m-1} = i_r\}$$

Aos processos estocásticos com essa propriedade chamaremos de *processos de markov* .

e denotaremos

$$P_r \{X_m = i_q \mid X_{m-1} = i_r\} \quad \text{por} \quad P^{(m)}_{i_r iq}$$

Outra propriedade que assumimos que o nosso processo estocástico tenha é a seguinte:

$P^{(m)}_{i_r i_q}$ não dependa de m ié

V m,n

$$P_r \{X_n = i_q \mid X_{n-1} = i_r\} = P_r \{X_m = i_q \mid X_{m-1} = i_r\}$$

Quando então denotaremos

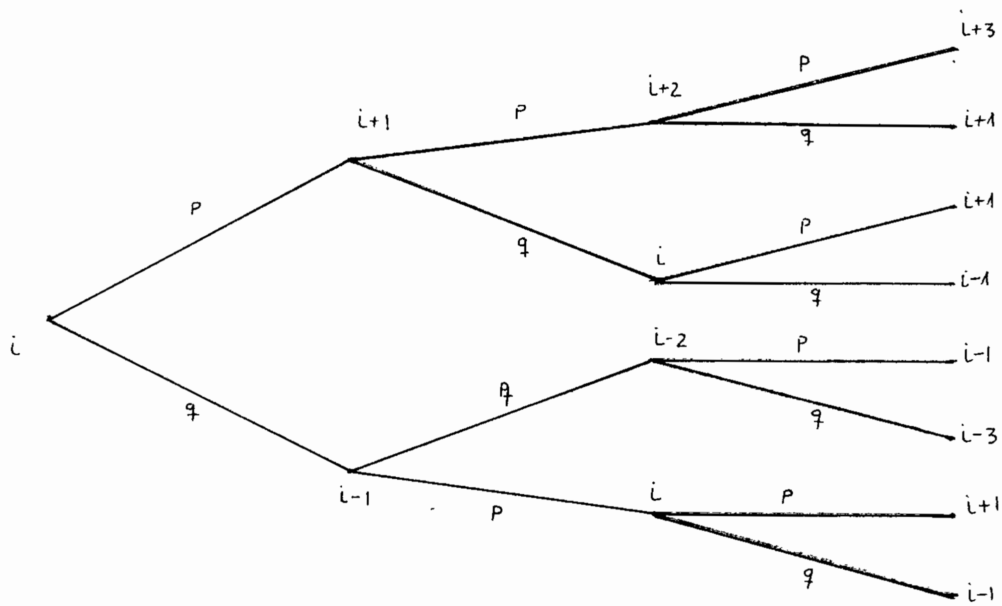
$Pi^{(m)}_{r}i_q$ por $Pi_r i_q$

e chamamos o processo de markov finito, de *cadeia de markov*.

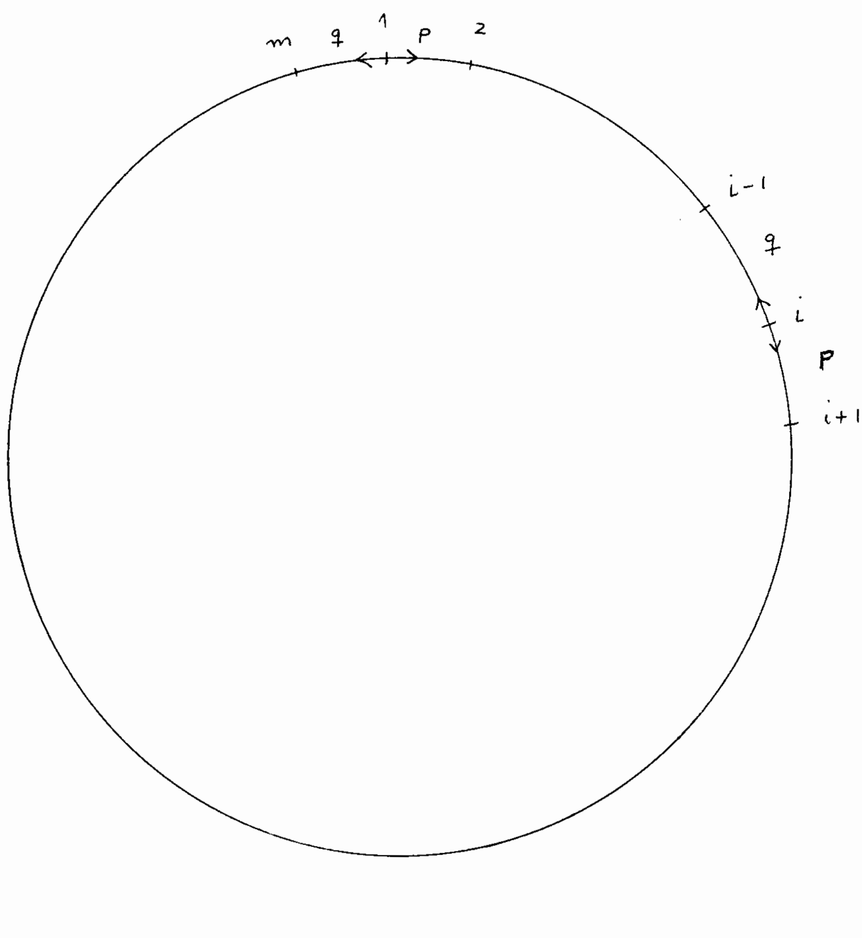
Vide Figura 6 .

Figura 6

Um exemplo de cadeia de markov:



p+q=1



Notemos agora que à cadeia de markov finita podemos associar uma matriz T a qual chamaremos matriz de transição da cadeia de markov.

$$T = \{P_{i_r i_q}\}$$

Para nossa cadeia de markov, T tem a forma:

$$T = \begin{matrix} 0 & P & 0 & 0 & \cdots & 0 & 0 \\ q & 0 & P & 0 & \cdots & 0 & 0 \\ 0 & q & 0 & P & \cdots & 0 & 0 \\ \\ P & 0 & 0 & 0 & \cdots & q & 0 \end{matrix}$$

$$T = T\ (m \times m)$$

Voltemos agora às classes de equivalência do conjunto P, cujo conjunto chamamos de P. Chamamos a atenção para o fato que o conjunto P é finito pois P é finito, fato este que tem como consequência trivial que a ordenação parcial induzida em P tem elementos minimais.

Os elementos minimais pela ordenação parcial das classes de equivalência serão chamados de conjuntos *ergódigos*.

Os outros elementos (aqueles que não são minimais) chamaremos de *transientes*.

Os elementos de um conjunto ergódigo serão por sua vez chamados de páginas ergódigas.

Afirmamos acima que existiam elementos minimais, notamos então que não é necessário que existam conjuntos transientes, o que acontece quando a cadeia consiste de um único conjunto ergódigo ou somente de vários conjuntos ergódigos que não se comunicam uns com os outros.

Recordamos que a consequência da relação de ordem parcial no conjunto de classes de equivalência P, é que quando o processo de markov deixa o conjunto transiente ele não retorna a esse conjunto, e quando ele entra num conjunto ergódigo ele não sai mais desse conjunto.

Então quando o conjunto ergódigo contém só uma página ela uma vez pedida não pede nenhuma outra página e nesse caso tal página i será chamada de *absorvente* e $P_{ii} = 1$.

Devido à relação de ordem parcial, consideramos as páginas de nosso programa P reenumeradas da seguinte maneira:

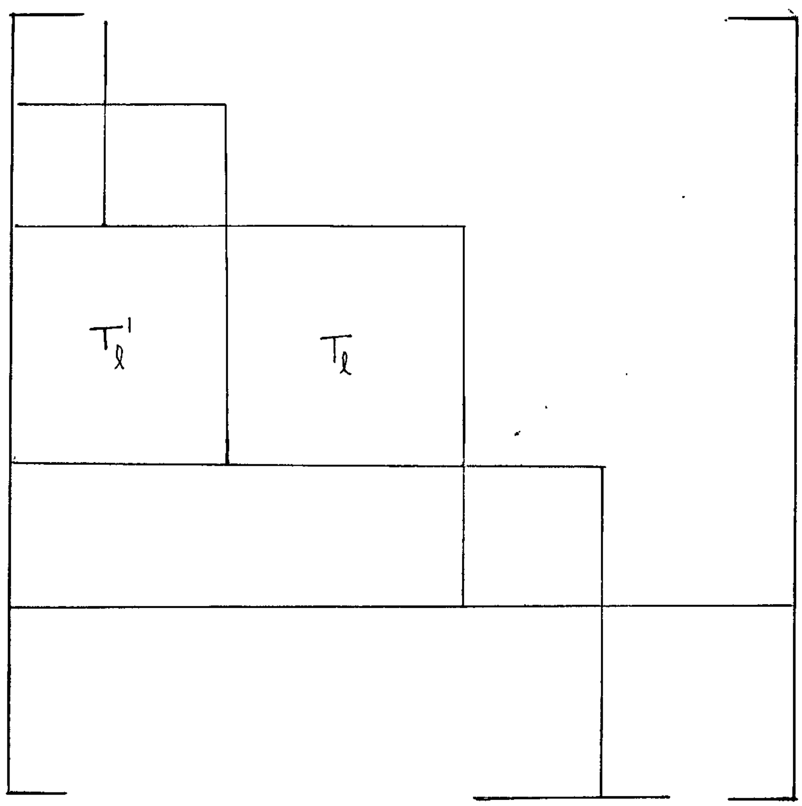
Primeiro enumeramos as páginas dos conjuntos minimais consecutivamente, depois as páginas das classes de equivalência que estiverem imediatamente acima das minimais e assim por diante.

Desse modo vemos que uma página de uma classe de equivalência pode pedir outra da mesma classe, ou de classe anterior, mas nunca de uma classe posterior.

A matriz de transição ficará com a forma ilustrada na Figura 7.

Vide Figura 7.

Figura 7



Onde aí T_1 são matrizes de transição para uma classe de equivalência, e T_1' será 0 se T_1 for ergódiga e diferente de 0 caso contrário.

Dizemos que uma cadeia de markov é *regular* se e só se ela consistir de somente uma classe de equivalência, e esta classe de equivalência de somente uma classe cíclica, em outras palavras, se todas as páginas puderem pedir todas as páginas, depois de um intervalo de tempo suficientemente grande.

Dizemos que uma cadeia de markov é absorvente, se e só se todos os conjuntos ergódigos são absorventes ié cada um é uma página i com $P_{ii} = 1$.

Vejamos agora alguns resultados para cadeias de markov cujas páginas não transientes são absorventes.

TEOREMA 4.1.:

Numa cadeia de markov finita, não importando onde o processo comece, a probabilidade depois de m transições que o processo esteja uma página absorvente tende a 1 quando $m \to$ infinito.

PROVA:

Se o processo alcança uma página absorvente, ele não pode deixá-la mais.

Suponhamos que ele comece uma página transiente, sua classe de equivalência não é minimal, portanto existe um elemento minimal que pode ter sua página pedida.

Isto significa que deve ser possível alcançar uma página absorvente em não mais de m transições (como existe somente um número finito de páginas, m é simplesmente o maior dos números, cada um per si número de transições requeridas para cada página). Portanto existe um número positivo p tal que a probabilidade de entrar uma página absorvente em no máximo m transições é pelo menos p a pedido de qualquer estado transiente. Portanto, a probabilidade de não alcançar uma página absorvente em m transições é no máximo (1 - P) que é menor que 1.

A probabilidade de não alcançar uma página absorvente em k.m transições é menor ou igual a $(1 - p)^k$ logo se $k \to$ infinito vem que $(1 - p)^k \to 0$, e o processo tende a entrar uma página absorvente com probabilidade 1.

c.q.d.

Consideremos então a *forma canônica* da matriz de transição *T*, mostrada na Figura 8.

Vide Figura 8.

Figura 8

$$\left[\begin{array}{c|c} T_a' & \\ \hline T_{at} & T_t \end{array}\right]$$

A matriz T_t (t x t) se conserne com o processo enquanto ele fica nas páginas transientes.

A matriz T_{at} (t x v) se conserne com a transição de páginas transientes para páginas absorventes.

Pelo teorema que demonstramos $T_t^m \to 0$, quando m → infinito.

A matriz T_a' (v x v) é igual a I , e só se conserne com as páginas absorventes.

TEOREMA 4.2.:

Para qualquer cadeia de markov absorvente $(I - T_t)$ tem uma inversa e

$$(I - T_t)^{-1} = I + T_t + T_t^2 + \qquad + = \sum_{l=0}^{\infty} T_t^l$$

PROVA:

$T_t^m \to 0$ como nós sabemos

notemos então que

$$(I - T_t) (I + T_t + T_t^2 + \qquad + T_t^{m-1}) = I - T_t^m$$

logo o lado direito tende para I se m → infinito.

Essa matriz tem determinante 1 logo

V m grande $(I - T_t^m)$ tem determinante diferente de zero.Como o produto dos determinantes de duas matrizes é o determinante do produto delas $(I - T_t)$ deve ter determinante não nulo, logo deve ser inversível, então $(I+T_t+ \quad +T_t^{m-1}) = (I-T_t)^{-1}(I-T_t^m)$ e toman do o limite quando $m \to$ infinito

$$(\mathcal{I} - \mathcal{T}_t)^{-1} = \sum_{l=0}^{\infty} \mathcal{T}_t^l$$

c.q.d.

Seja agora uj uma variável aleatória que diz o número de vezes que o processo está na página j (onde j é transiente).

Seja u(k,j) definida como sendo 1 se o processo está na página j depois de k transições, e 0 caso contrário.

TEOREMA 4.3.:

$$\{ E_i (uj) \} = (\mathcal{I} - \mathcal{T}_t)^{-1} \qquad \text{onde } i,j \text{ são transientes.}$$

E_i (u,j) é o valor esperado de uj dado i .

PROVA:

$$uj = \sum_{k=0}^{\infty} u(k,j)$$

$$E_i (uj) = \sum_{k=0}^{\infty} E_i(u(k,j))$$

$$= \sum_{k=0}^{\infty} \{(1 - P_{ij}^{(k)}) . 0 + P_{ij}^{(k)} . 1\}$$

$$= \sum^{\infty} P_{ij}^{(k)}$$

onde $P\{u(k,j) \mid i\} = 1 - P_{ij}^{(k)}$

$P\{u(k,j) \mid i\} = P_{ij}^{(k)}$

logo

$$\{E_i(uj)\} = \sum_{k=0}^{\infty} T_t^k$$

$$= (I - T_t)^{-1}$$

c.q.d.

Este teorema nos diz que o valor médio do número de vezes que o processo está em uma página transiente determinada j , é sempre finito e que seu valor esperado é dado pelos entries de $(I - T_t)^{-1}$.

Logo como a cadeia é finita o número de páginas transientes é finito, assim o *tempo esperado de absorção* é finito.

TEOREMA 4.4.:

Seja T uma matriz de transição (r x r) tal que nenhum de seus entries seja nulo. Seja $E>0$ o menor entry de T .

Seja X um vetor coluna de r componentes tendo componente máximo M_o e componente mínimo m_o , e seja M_1 e m_1 os componentes máximo e mínimo respectivamente do vetor $T.X$. Então $M_1 \leqslant M_o$, e $m_1 \geqslant m_o$, e $M_1 - m_1 \leqslant (1 - 2E)(M_o - m_o)$

PROVA:

Seja X' o vetor obtido de X, se trocando todas as componentes de X exceto m_o por M_o. $X \leqslant X'$. Cada componente de $T.X'$ é da forma

$$e \cdot m_o + (1 - e)M_o = M_o - e(M_o - m_o) \;, \quad \text{onde} \quad e \geqslant E \;.$$

Logo tal componente é $\leqslant M_o - E(M_o - m_o)$.

Mas como $X \leqslant X'$, nós temos $M_1 \leqslant M_o - E(M_o - m_o)$

pois $T.X \leqslant T.X'$

Agora consideremos o vetor $-X$, para esse vetor $-m_o$ é agora componente máximo, e $-M_o$ é componente mínimo.

Repetindo o raciocínio acima, para esse vetor, obtemos

$$X' = \begin{bmatrix} -m_o \\ -M_o \\ -m_o \end{bmatrix} \quad \text{e} \quad -X \leqslant X'$$

Cada componente de $T.X'$ é da forma

$$-a.M_o - (1-a)m_o = -m_o - a\,(-m_o+M_o)$$

$$\text{onde} \quad a \geqslant E \;.$$

Mas como $-X \leqslant X'$ vem então que

$$-m_1 \leqslant -m_o - a(-m_o+M_o)$$

e como $-a \leqslant -E$

$$-m_1 \leqslant -m_o + E\,(-m_o+M_o) \qquad (2)$$

Somando (1) e (2) vem que

$M_1 - m_1 \leqslant M_o - m_o - 2E(M_o-m_o)$

$$M_1 - m_1 \leqslant (1 - 2E)\ (M_o-m_o)$$

de (1) $\quad M_1 \leqslant M_o-E(M_o-m_o) \leqslant M_o$

de (2) $\quad m_o \leqslant m_1$

c.q.d.

TEOREMA 4.5.:

Se T é uma matriz de transição de uma cadeia de markov regular, então:

(1) As potências de T tendem a uma matriz probabilística E .

(2) Cada linha de E é o mesmo vetor probabilidade L.

(3) As componentes de L são positivas.

PROVA:

Suponhamos que T não tenha zeros, ou seja entries nulos.

Seja E o menor entry.

Seja cj um vetor coluna tal que a sua jésima componente seja 1, e as outras componentes sejam iguais a zero.

Seja M_n e m_n as componentes máxima e mínima respectivamente do vetor $T^n.Cj$.

Como $T^n.Cj = T.T^{n-1}.Cj$ nós temos do teorema anterior que

$$M_1 \geqslant M_2 \geqslant M_3 \geqslant \ldots \text{ e } m_1 \leqslant m_2 \leqslant m_3 \leqslant \ldots$$

e que

$$M_m - m_n \leqslant (1 - 2E)\ (M_{n-1} - m_{n-1})$$

$$\forall n \geqslant 1 .$$

fazendo $d_n = M_n - m_n$

$$d_n \leqslant (1 - 2E)^n$$

logo se $n \to$ infinito, implica que $d_n \to 0$.

o que significa que M_n e m_n tendem a um limite comum, e portanto

$T^n.Cj$ tende a um vetor cujas componentes são todas as mesmas ou seja são iguais.

Seja Cj esse valor comum. É claro que $\forall n \quad 0 < m_n \leqslant Cj \leqslant M_n \leqslant 1$.

Agora $T^n.Cj$ é a jésima coluna de T^n, logo T^n tende a matriz E com todas as linhas iguais ao vetor $L = \{L_1, L_2, \ldots, L_r\}$ pois a jésima coluna de T^n tende a um vetor com todas as componentes iguais ao valor Lj .

Como a soma das componentes de cada linha de T^n é sempre 1, o mesmo deve ser verdade do limite.

Agora se T é a matriz de transição de uma cadeia de markov regular, sabemos que existe um inteiro N tal que T^N não tem entries nulos.

Então quando $n > N$, $n \to$ infinito

$T^n \to E$ com as propriedades enunciadas.

TEOREMA 4.6.:

Se T é uma matriz de transição de cadeia de markov regular e

C é um vetor coluna tal que $T.C = C$ onde $C^t = \{C_1, C_2, \ldots, C_r\}$

então

$$C = C \begin{bmatrix} 1 \\ 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}$$

como $T \cdot C = C$ então $T^n \cdot C = C$ mas

então $E \cdot C = C$ assim $C_i = L \cdot C$, $\forall i$

logo $C_1 = C_2 = \quad = C_r = C$, C constante

c.q.d.

Para uma cadeia de markov regular, é interessante definir a variável aleatória tj, tempo para ir de uma página i a uma página j pela primeira vez, ou seja o tempo de primeira passagem tj é uma variável aleatória cujo valor é *o número de transiçõesnecessárias antes de entrar em j pela primeira vez depois de sair da página i dada*.

TEOREMA 4.7.:

Para qualquer página i , $E_i(tj)$ é finito.

PROVA:

Se assumirmos que $i \neq j$. Formamos uma nova cadeia de markov tornando j absorvente, então essa nova cadeia de markov é absorvente com um único estado absorvente j .

Portanto, o tempo esperado para ir de i a j na cadeia é o tempo esperado de absorção, o qual é finito por um teorema anterior.

Se $i = j$ então notemos que:

$$E_i(t_j) = \sum_k P_{ik} \cdot E(t_j | X_2 = k)$$

$$= \sum_{k \neq j} E(tj | X_2 = k) \cdot P_{ik} + P_{ij} \cdot E(tj | X_2 = j)$$

$$= \sum_{k \neq j} P_{ik} \cdot \{1 + E(tj | X_1 = k)\} + P_{ij}$$

$$= \sum_{k \neq j} P_{ik} \{1 + E_k(tj)\} + P_{ij}$$

para $i=j$, então definindo $E_i(t_i)$ como sendo tempo médio de recorrência da página i .

$E_i\ (t_i)$ é finito pois

$$E_i(t_i) = P_{ii} + \sum_{k \neq i} \{ 1 + P_{ik} \cdot E_k(t_i) \}$$

que é finito pela primeira parte.

c.q.d.

Notemos agora que $c_{ij} = E_i(t_j)$ pode ser colocado em forma de matriz, a qual denotaremos por E e chamaremos de *matriz do valor* *esperado do tempo de primeira passagem.*

Formulamos em seguida um importante....

TEOREMA 4.8.:

A matriz E satisfaz a equação

$$E = T\{E - E_{dg}\} + 1$$

onde E_{dg} é a matriz obtida de E fazendo todos os entries fora da diagonal igual a zero. 1 é a matriz com todos os entries iguais a 1 .

PROVA:

$$E_i(t_j) = \sum_{k \neq j} P_{ik}\{E_k(tj) + 1\} + P_{ij}$$

$$= \sum_{k \neq j} P_{ik} \cdot E_k(tj) + 1 - P_{ij} + P_{ij}$$

$$= \sum_{k} P_{ik} \cdot E_k(tj) - P_{ij} \cdot E_j(tj) + 1$$

o que em forma matricial é

$$E = T\{E - E_{dg}\} + 1$$

$$E = \{c_{ij}\}$$

c.q.d.

TEOREMA 4.9.:

Seja $L = \{L_1, L_2, \ldots, L_r\}$ o vetor probabilidade limite de T.

Então $c_{ij} = \frac{1}{L_i}$

PROVA:

Como $E = T\{E - E_{dg}\} + 1$

prê multiplicando -a por L nós obtemos

$$L.E = L.T\,(E - E_{dg}) + L \,.\, 1$$

$$L.E = L\,\{E - E_{dg}\} + L.1$$

ou $$L.E_{dg} = L.1 = \begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ 1 \end{bmatrix}$$

ou $L_i \,.\, C_{ii} = 1$

assim $c_{ii} = \frac{1}{L_i}$

TEOREMA 4.10.:

A equação $E = T\,\{E - E_{dg}\} + 1$

tem uma ũnica solução.

PROVA:

Sejam E e E' duas soluções dessa equação, então nós sabemos do teorema anterior que

$$L \cdot E_{dg} = L \cdot E'_{dg}$$

ou $\quad L_i \cdot c_{ii} = L_i \cdot c'_{ii}$

ou $\quad c_{ii} = c'_{ii}$

ou $\quad E_{dg} = E'_{dg}$

logo $\qquad E - E' = T\,\{E - E'\}$

mas então se olharmos as colunas de $E - E'$ uma de cada vez notamos que elas são vetores com a propriedade $T \cdot C = C$.

Mas por teorema anterior cada coluna é então um vetor constante.

Como $E - E'$ tem sua diagonal toda em zeros logo $E - E' = 0$ então $E = E'$.

c.q.d.

Agora observemos que não importa o meio de cálculo, uma vez achada uma matriz E que satisfaz a equação.

$$E = T\,\{E - E_{dg}\} + 1$$

achamos a matriz do valor esperado do tempo de primeira passagem, para uma cadeia de markov cuja matriz de transição é T.

A matriz da cadeia de markov que é alvo de nosso estudo tem a forma:

$$T = \begin{matrix} O & P & O & O & \ldots & O & q \\ q & O & P & O & \ldots & O & O \\ O & q & O & P & \ldots & O & O \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot \\ q & O & O & O & \ldots & P & O \end{matrix}$$

Nosso interesse é obter c_{ij}, qual seja o valor esperado do tempo de primeira passagem pela página j tendo como página de partida a página i.

Pela simétria é de se esperar que esse intervalo de tempo seja função do número de páginas do programa, da distância definida segundo algum critério entre a página inicial e a página final que se tem em mente e dos valores das probabilidades p e q respectivamente de sair de uma página k e ir diretamente para uma página $k+1$ e $k-1$ respectivamente.

Afirmamos que as seguintes matrizes E definidas por

$$c_{ij} = \frac{1}{p-q} \left\{ m \cdot \frac{r^m - r^{m-d}}{r^m - 1} - d \right\} \qquad \begin{matrix} i \neq j \\ p \neq q \end{matrix}$$

$$c_{ii} = m$$

$$c_{ij} = d(m-d) \qquad i \neq j$$

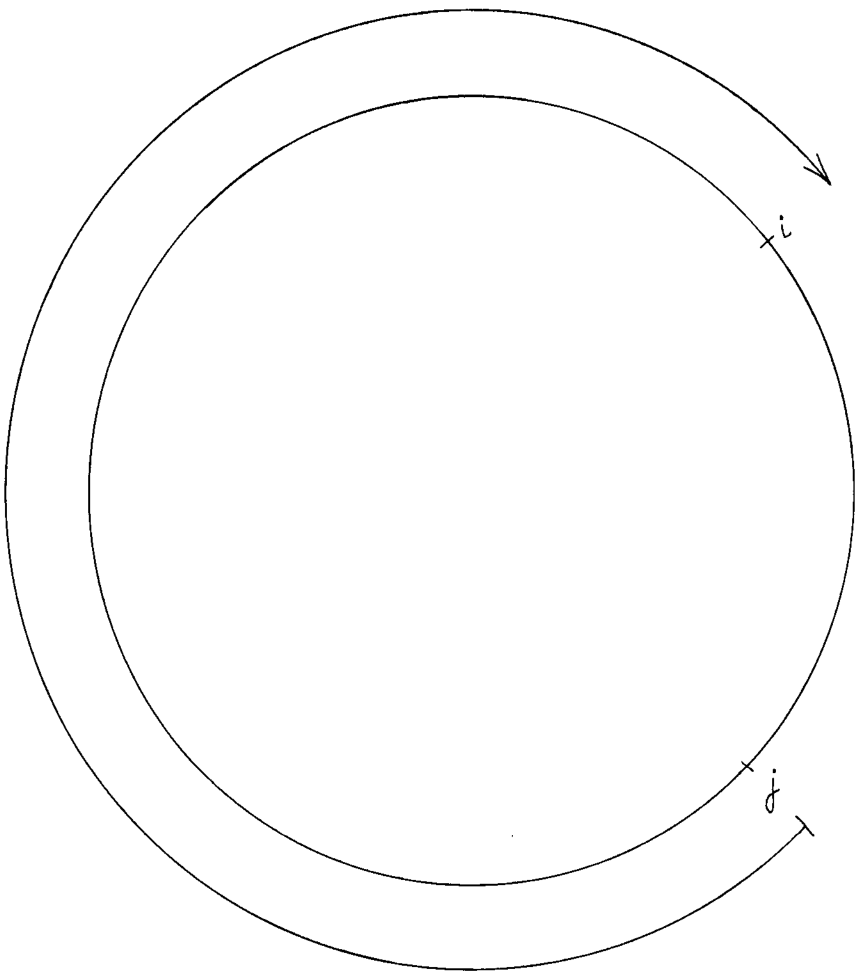
$$p = q = \frac{1}{2}$$

$$c_{ii} = m$$

onde $r = p/q$, d é a distância entre a página i e j definida da seguinte maneira, mostrada na figura , como sendo o número de transições necessárias para ir de j a i no sentido dos ponteiros de relógio sem voltar atrás.

Vide Figura 9.

Figura 9



São as matrizes E respectivamente para $p \neq q$ e $p=q=\frac{1}{2}$ para a cadeia de markov cuja matriz de transição T é a dada.

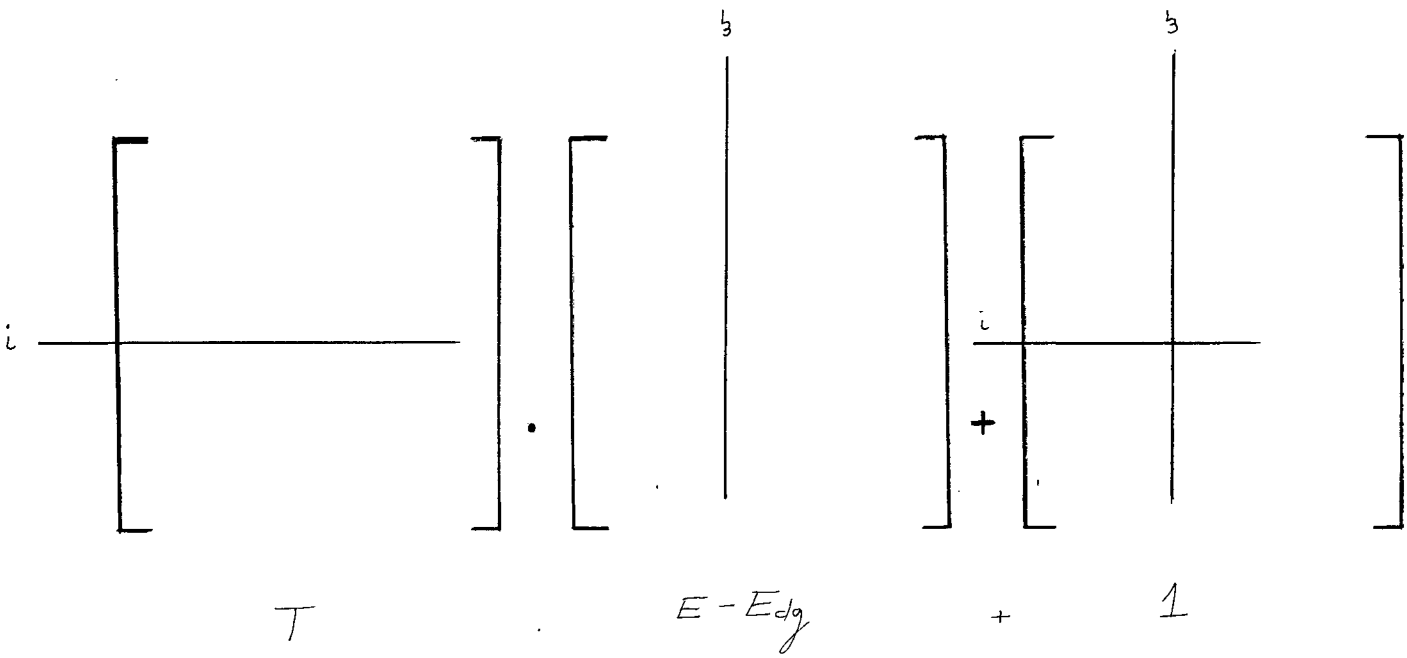
Para que não paire dúvidas basta que provemos que {cij} satisfaz a identidade:

$$E = T \{ E - E_{dg} \} + 1$$

$$T = \begin{matrix} 0 & p & 0 & 0 & \dots & 0 & q \\ q & 0 & p & 0 & \dots & 0 & 0 \\ 0 & q & 0 & p & \dots & 0 & 0 \\ \\ p & 0 & 0 & 0 & \dots & q & 0 \end{matrix}$$

Vide Figura 10.

Figura 10



substituindo T em $T\{E - E_{dg}\} + 1$

obtemos o $\{i,k\}$ésimo elemento que é

$$p \cdot c_{2k} + q \cdot c_{mk} + 1 \qquad i = 1$$

$$1 \cdot c_{(i+1)k} + q \cdot c_{(i-1)k} + 1 \qquad 1 < i < m-1$$

$$p \cdot c_{1k} + q \cdot c_{(m-1)k} + 1 \qquad j = m$$

provemos o primeiro caso $p \neq q$.

$$\frac{r}{r-1} \{ m \cdot \frac{r^m - r^{m-d-1}}{r^m - 1} - d - 1 \} + \frac{1}{r-1} \{ m \cdot \frac{r^m - r^{m-d+1}}{r^m - 1} - d + 1 \} + 1$$

$$= \frac{1}{r-1} \{ m \cdot \frac{r^{m+1} - r^{m+d} + r^m - r^{m-d+1}}{r^m - 1} - dr - r - d + 1 + r - 1 \}$$

$$= \frac{1}{r-1} \{ m \, (r+1) \frac{r^m - r^{m-d}}{r^m - 1} - d \}$$

$$= \frac{1}{p-q} \{ m \cdot \frac{r^m - r^{m-d}}{r^m - 1} - d \}$$

$$= c_{ik}$$

c.q.d.

agora provemos para o caso $p=q=\frac{1}{2}$

$$= \frac{1}{2}\ \{(d+1)\ (m-d-1) + (d-1)\ (m-d+1)\} + 1$$

$$= \frac{1}{2}\ \{\ 2md - 2d^2 - 2\ \} + 1$$

$$= (m - d)d$$

$$= cik$$

c.q.d.

logo achamos os valores esperados dos tempos de primeira passagem para a cadeia de markov que é base de nosso estudo.

Nossa pergunta agora seria: sob quais condições seria válida as seguintes ordenações?

$$E_i(t_{i+1}) < E_i(t_{i+2}) < E_i(t_{i+3}) < \qquad < E_i(t_{i+m})$$

ou

$$E_i(t_{i+1}) > E_i(t_{i+2}) > E_i(t_{i+3}) > \qquad > E_i(t_{i+m})$$

basta substituir os respectivos valores e por simplificação algébrica, chegamos às seguintes condições:

A primeira ordenação se verifica tão somente quando $r>1$ e $\frac{r^m-1}{r-1} < mr^k$, para k entre 1 e $m-1$.

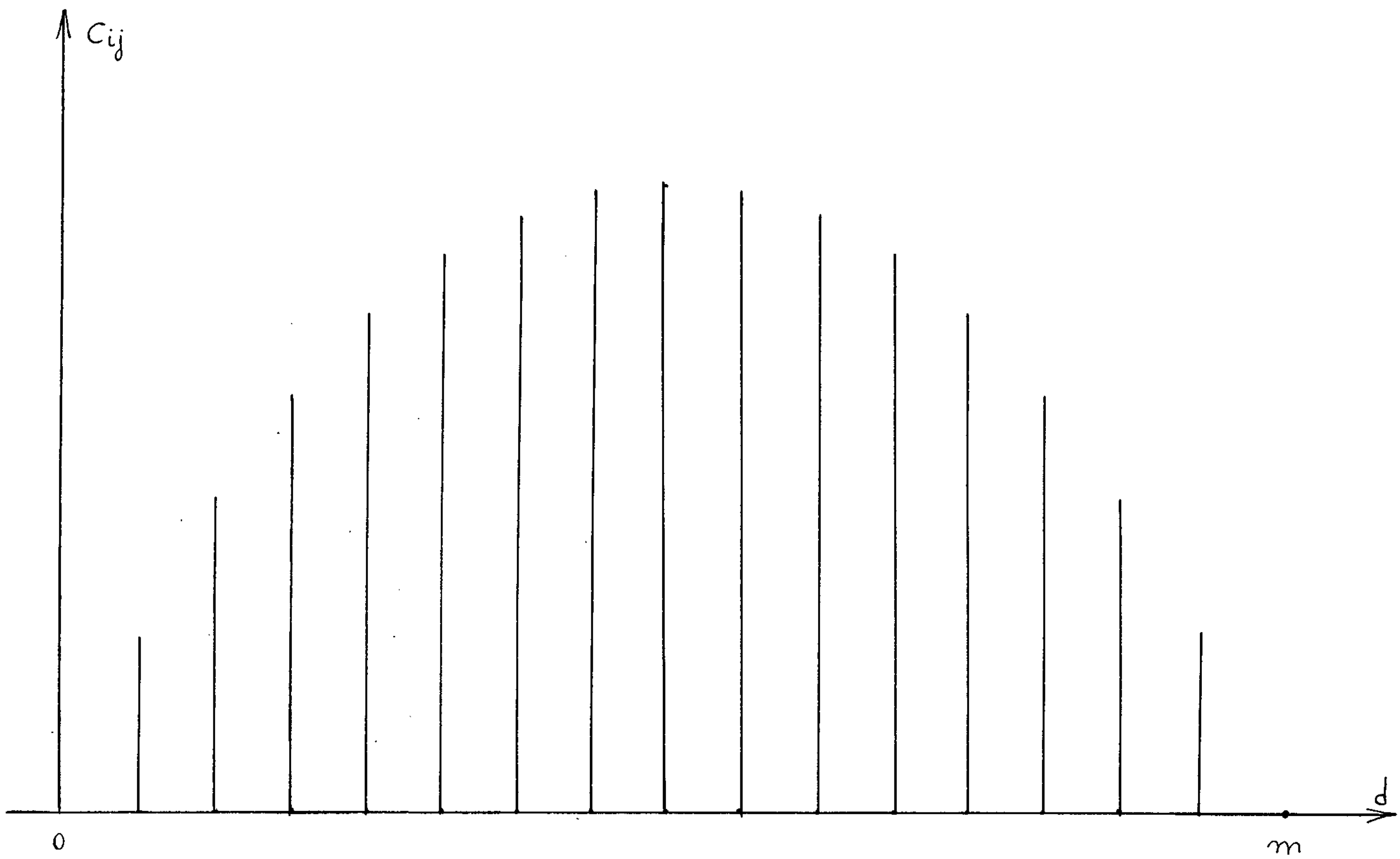
A segunda ordenação se verifica tão somente quando $r<1$ e

$\frac{r^m-1}{r-1} > mr^k$, para todo k entre 1 e $m-1$.

Agora quando $p=q=1$, examinamos a equação $cij = d(m-d)$, $i \neq j$. a qual no R^2 tem um gráfico no primeiro quadrante, repousando numa parábola conforme mostra a Figura 11.

Vide Figura 11 .

Figura 11



notamos, então, que c_{ij} é máximo quando $d = \frac{m}{2}$.

Estudemos, agora, a função f definida abaixo em A x B.

$$A \times B = \{(r, d) / r \varepsilon [0, \infty), \quad d \varepsilon [0, m]\}$$

$$f : A \times B \longrightarrow R$$

$$(r, d) \longmapsto f(r, d) = \begin{cases} \dfrac{r+1}{r-1} \cdot \left\{ m \cdot \dfrac{r^m - r^{m-d}}{r^m - 1} - d \right\}, & r \neq 1 \\ (m-d)\, d, & r = 1 \end{cases}$$

valem as seguintes propriedades:

$$f(r, d) = f(r^{-1}, m - d)$$

Se $r = 1$ é trivial

Se $r \neq 1$

$$f(r^{-1}, m - d) = \frac{r^{-1} + 1}{r^{-1} - 1} \left\{ m \cdot \frac{(r^{-1})^m - (r^{-1})^{m-[m-d]}}{(r^{-1})^m - 1} - [m - d] \right\}$$

$$= - \frac{r+1}{r-1} \left\{ m \cdot \frac{1 - r^{m-d}}{1 - r^m} - m + d \right\}$$

$$= \frac{r+1}{r-1} \left\{ m \cdot \frac{r^m - r^{m-d}}{r^m - 1} - d \right\}$$

$$= f(r, d)$$

Em palavras, as curvas representativas da equação são simétricas em relação a um eixo vertical situado em $d = \frac{m}{2}$.

(2) $f(r, d) \geqslant 0$ em $A \times B$, $f(r, d)$ é continua e até mesmo derivável em relação a d.

Agora, calculando $\frac{\partial f(r, d)}{\partial d}$, $r \neq 1$

obtemos:

$$\frac{\partial f(r, d)}{\partial d} = \left(\frac{r+1}{r-1}\right) \left(\frac{-1}{r^m-1}\right) \frac{\partial\{r^{m-d}+d\}}{\partial d}$$

$$= \left(\frac{r+1}{r-1}\right) \left(\frac{-1}{r^m-1}\right) \{r^{m-d}. (-1). Lr + r^m -1\}$$

$$= \frac{r+1}{r-1} \left\{ -1 + \frac{r^{m-d}}{r^m-1} . Lr \right\}$$

$$= \frac{r+1}{r-1} \left\{ \left(\frac{r^m}{r^m-1} . Lr\right) . r^{-d} - 1\right\}$$

$$= B \{Ar^{-d} - 1\}$$

onde, $A = \frac{r^m}{r^m-1} . Lr$, $B = \frac{r+1}{r-1}$

Em virtude da propriedade 1 basta estudar o comportamento de $\frac{\partial d(r, d)}{\partial d}$ em $(1, \infty) \times [0,m]$

notando que $A > 1$, $B > 1$ para todo $r \in (1,\infty)$ ilustramos graficamente.

Vide Figura 12.

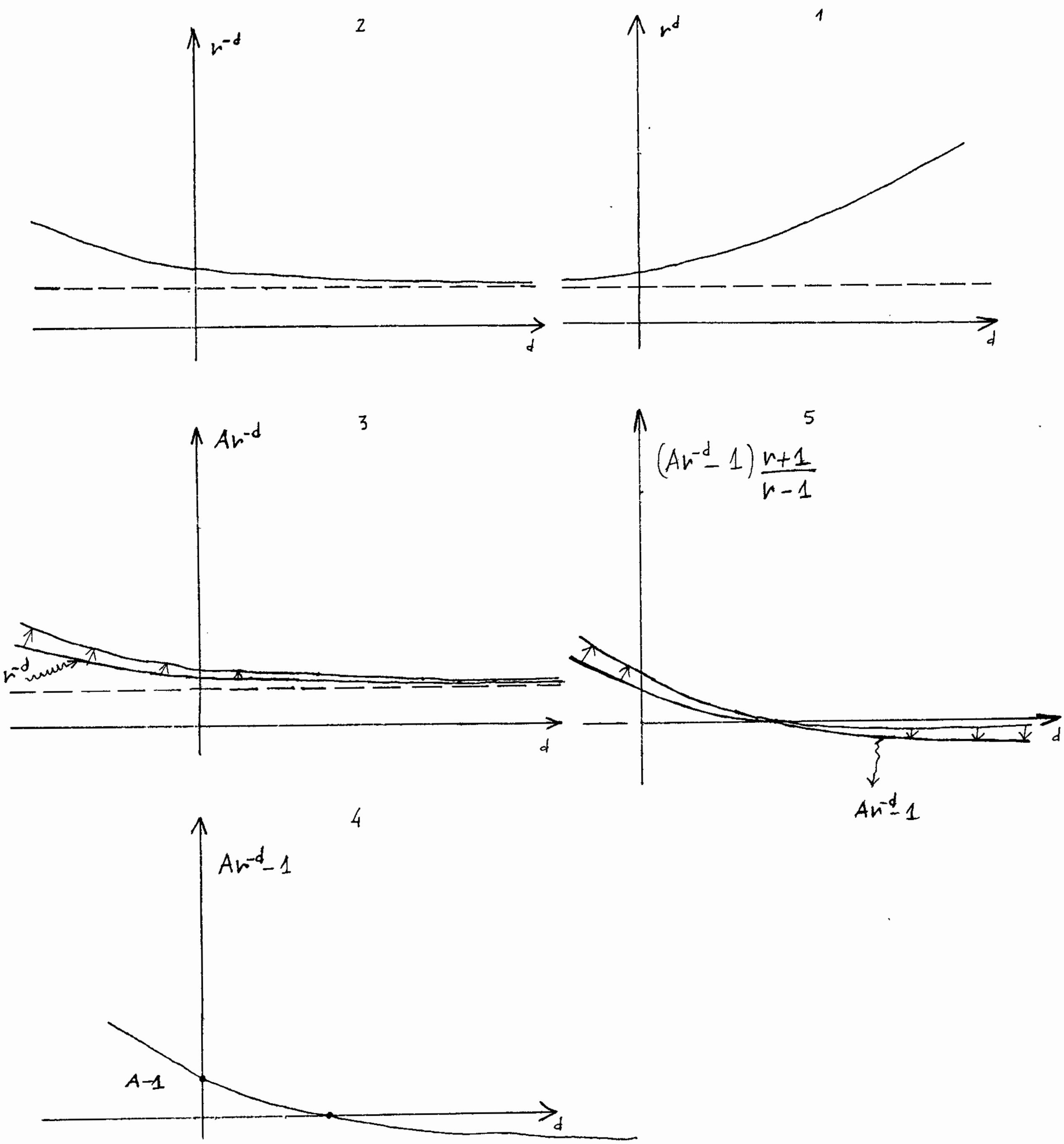
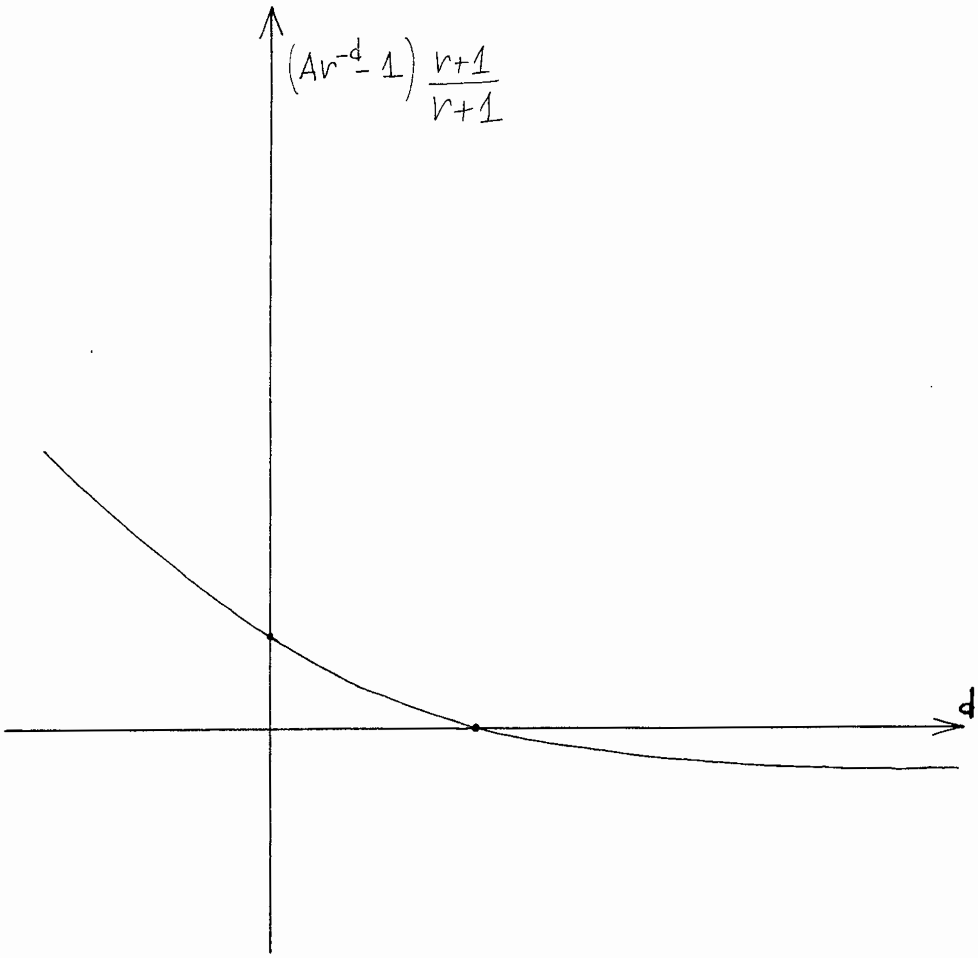


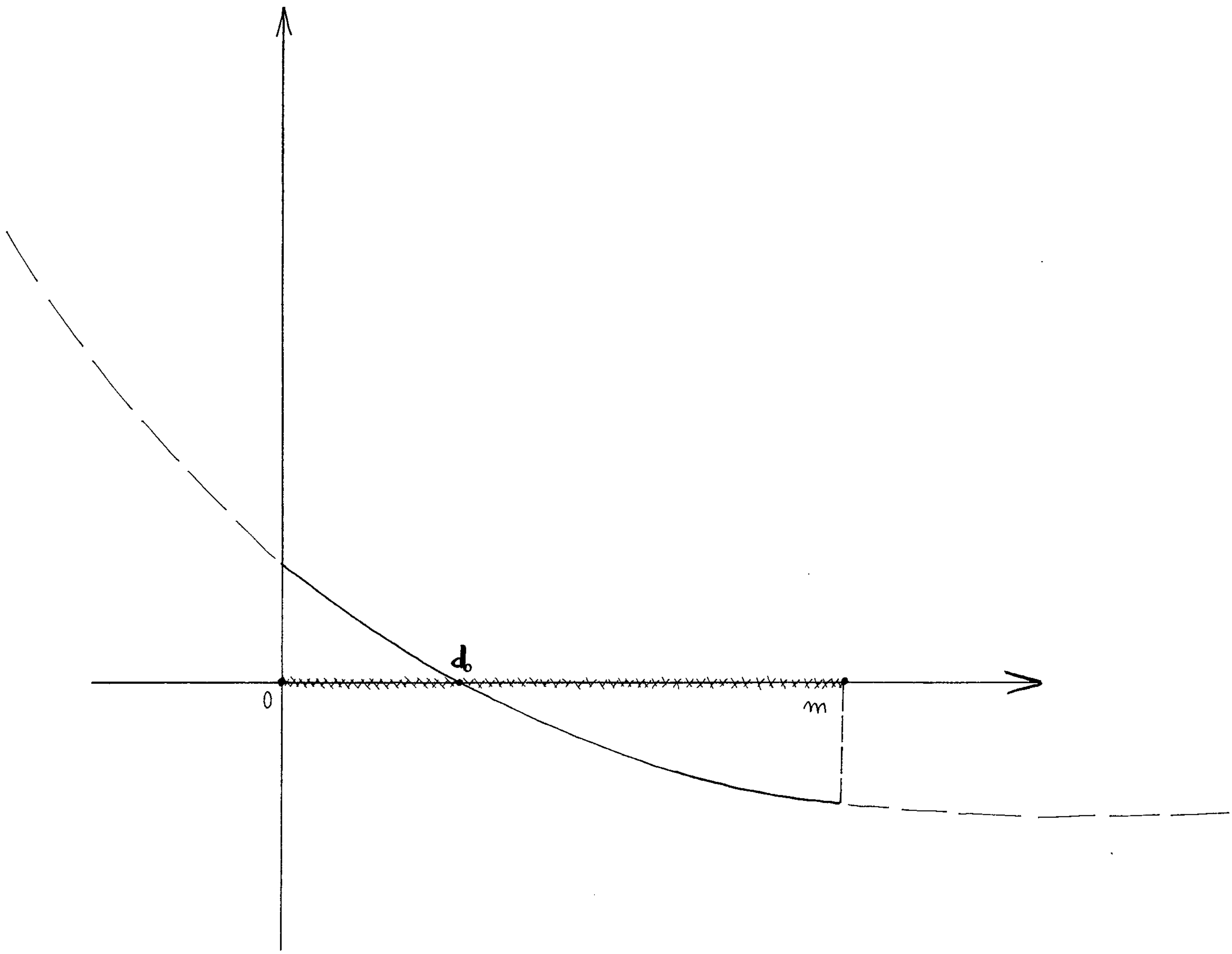
FIGURA 12



A curva representativa de $\frac{\partial f(r, d)}{\partial d}$ pode ser descrita como uma curva derivável, convexa, decrescente, e que corta o eixo horizontal em um ponto de (0,m), daí concluímos que f(r, d) para $r \neq 1$ fixo, tem a seguinte forma.

Vide figura 13.

FIGURA 13



isto é,

(3) É uma curva concava, com um máximo em (0,m) mais precisamente em d_o.

$$d_o = \begin{cases} \dfrac{LA}{Lr} \quad \text{onde} \quad A = Lr \cdot r^m \cdot \dfrac{1}{r^m - 1} & , r \neq 1 \\ \dfrac{m^2}{2} & , r = 1 \end{cases}$$

Agora nos interessa o inteiro d_o' em (0,m) tal que para esse conjunto $f(r, d_o')$ seja máximo. d_o' é obtido calculando-se o valor de $f(r, .)$ nos dois inteiros imediatamente superior e inferior a d_o respectivamente se $d_o \in [1, m-1]$.

Agora se $r \square 1$ e $d_o \in (0,1)$ então $d_o' = 1$.

se $r = 1$ e $d_o \in (m-1, m)$ então $d_o' = m-1$.

Podemos então, enunciar um Algoritmo Ótimo de Troca de Páginas:

Se a página i estiver na memória auxiliar e for pedida para ser executada em seguida, ela tem que vir para o buffer. A página do buffer que deverá ser retirada para dar lugar a página i será a página correspondente àquela página situada no círculo a qual tem as duas propriedades seguintes:

(1) Ela corresponde à uma página no buffer.

(2) das páginas correspondentes àquelas que estão no buffer ela é a que tem o maior tempo esperado de chamada.

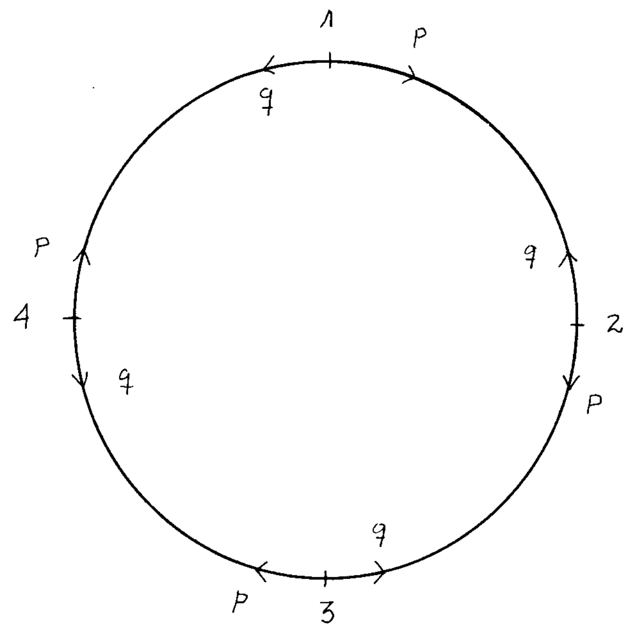
A P Ê N D I C E

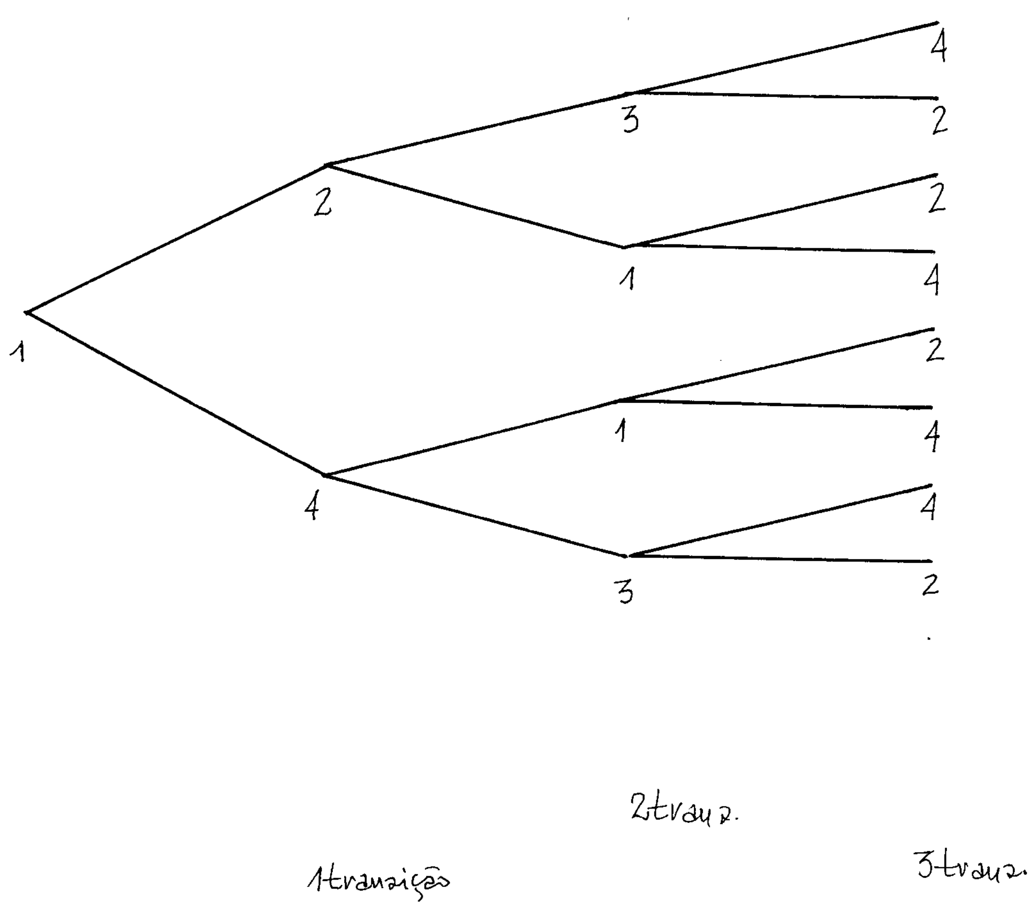
Discutimos primeiro o problema da paridade do número de páginas que o programa necessariamente deve ter.

Caso Par:

Exemplificamos uma cadeia de Markov que não é regular, e corresponde ao caso do programa ter quatro páginas.

Vide Figura 1 .


1
p
q
q
2
p
q
3
p
4
p
q


1
2
4
3
1
1
3
4
2
2
4
2
4
4
2
1transição
2trans.
3trans.

Caso Ímpar:

Esse caso generaliza o caso trivial do número de páginas do programa ser primo.

Notemos então que quando o número de páginas do programa é ímpar também vale o algorítmo. Implicitamente queremos dizer que de um estado qualquer existe probabilidade positiva de qualquer outro estado seja alcançado em um número finito de transições.

A prova é geométrica, pois raciocinamos em termos de árvores.

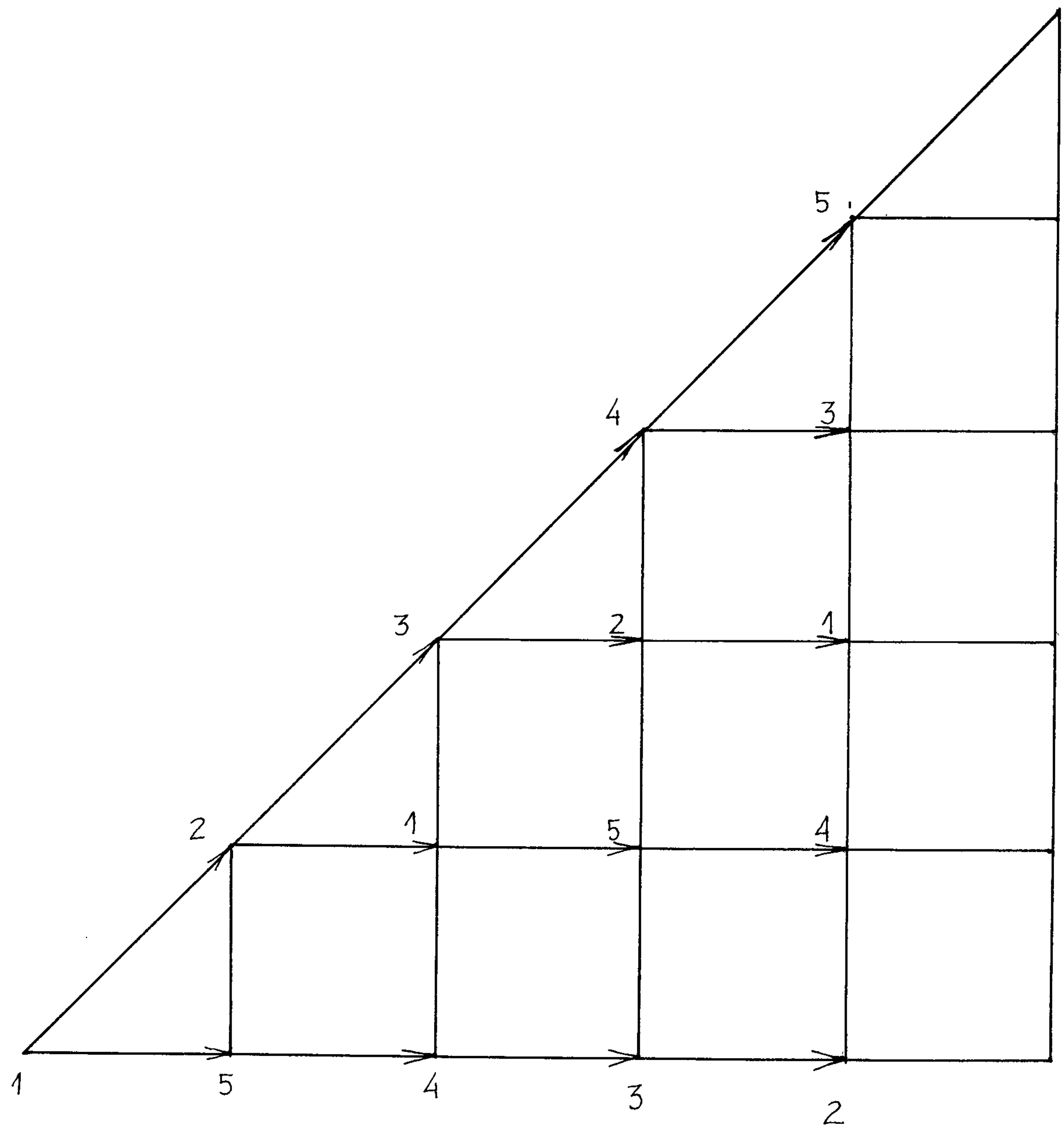
Chamamos atenção para o fato que é suficiente exibirmos uma subárvore tal que em um número finito de transições todas as páginas são alcançadas.

Antes de mostrarmos em plena generalidade ilustramos 3 casos particulares correspondentes a $m = 5$, 7 e 11 respectivamente.

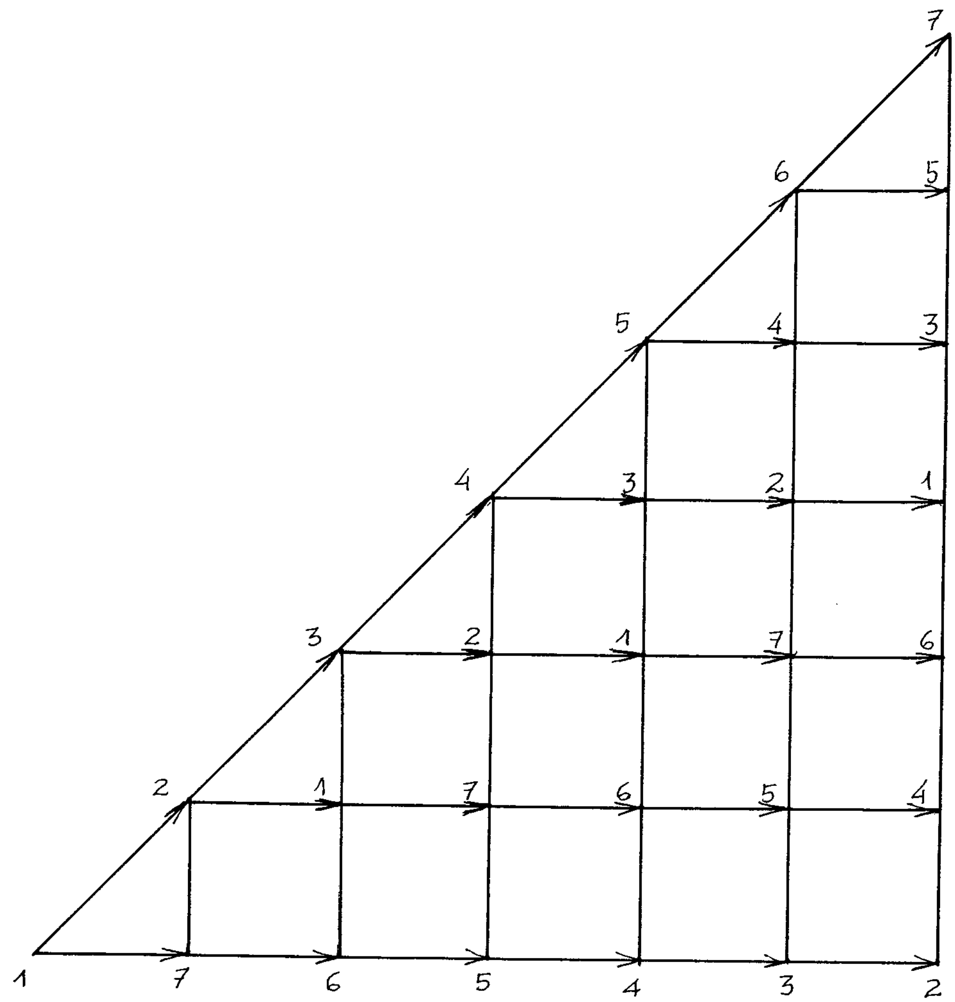
Vide Figura 2

FIGURA 2

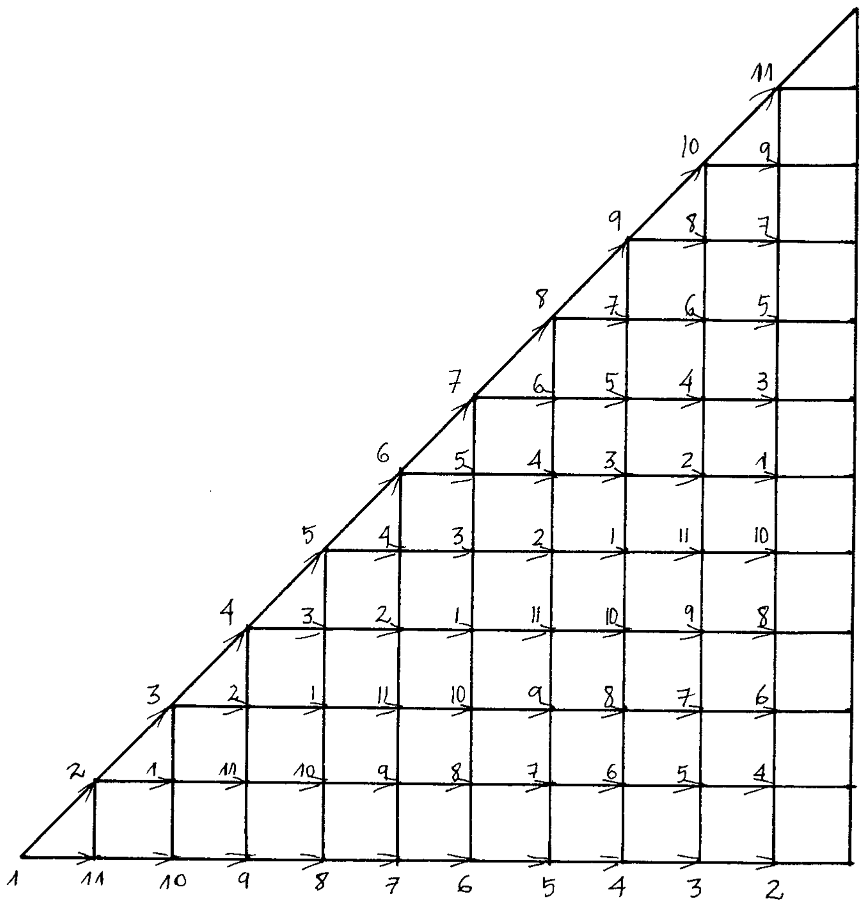
m = 5



m = 7



m = 11



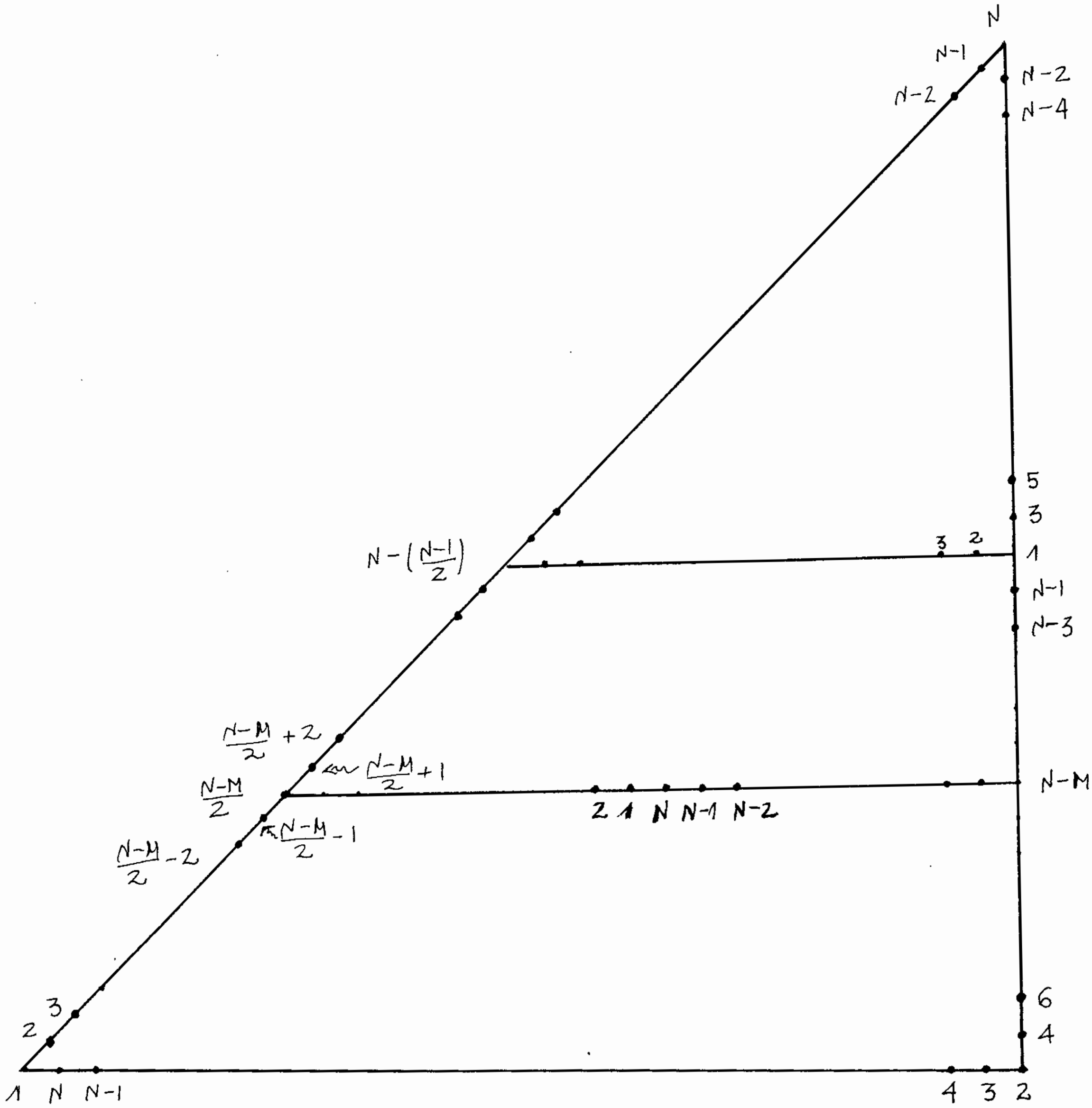
Agora mostramos o caso geral:

O propósito de ilustrarmos os casos particulares anteriormente foi para que o leitor notasse que existe uma certa ordem na árvore, de baixo para cima alcançamos em exatamente $N - 1$ transições os estados 2, 4, 6 ...,

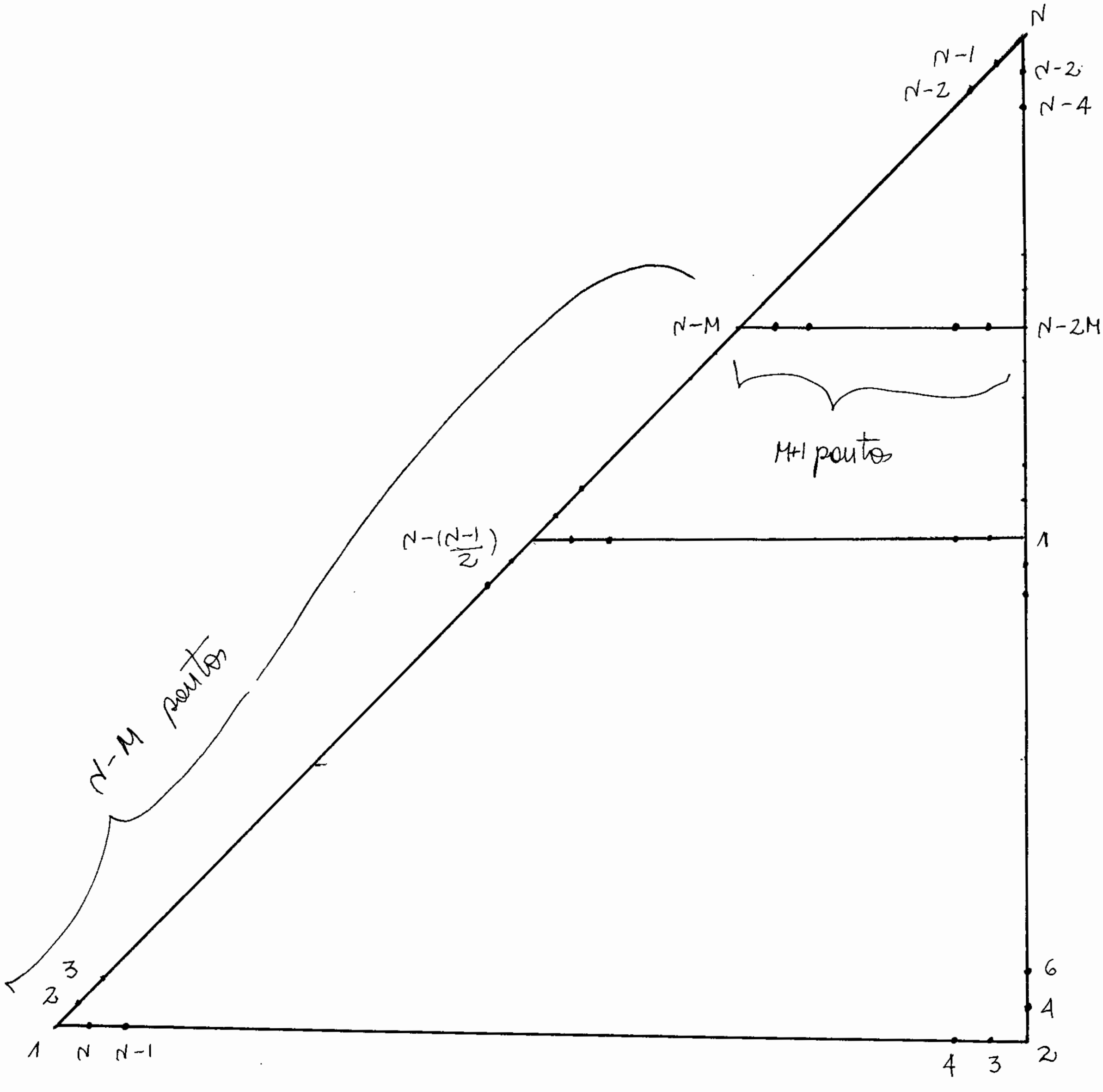
$N-3, N-1, 1, 3, 5, \ldots, N-4, N-2$ e N .

Vide Figura 3.

FIGURA 3



$$\left(\frac{N-M}{2} + \frac{N-M}{2} + M \right) - 1 = (N-1) \text{ transições}$$



{ (N-M) + M } = (N-1) transições

Agora queremos mostrar ao leitor que na demonstração da página 51, os casos $i=1$ e $j=m$ são triviais e apresentamos a seguir a demonstração para o caso que $k = i+1$ e $k= i-1$

$k = i+1$

$$p \cdot 0 + q \cdot c_{i-1\ i+1} + 1 = c_{i\,i+1} \ , \ d(i-1 \ , \ i+1) = m - 2 \ .$$

$$q \cdot \frac{1}{p-q} \{ m \cdot \frac{r^m - r^2}{r^m - 1} - (m-2) \} + 1$$

$$\frac{1}{p-q} \{ m \cdot \frac{r^m - r^2}{r^m - 1} \cdot q - mq + 2q \} + 1$$

$$m \cdot \frac{r^m - r^2}{r^m - 1} \cdot \frac{q}{p-q} - \frac{mq}{p-q} + \frac{2q}{p-q} + 1$$

$$m \cdot \frac{r^m - r^2}{r^m - 1} \cdot \frac{1}{r-1} - \frac{m}{r-1} + \frac{2}{r-1} + 1$$

$$m \cdot r \cdot \frac{(r^{m-1} - r)}{r^m - 1} \cdot \frac{1}{r-1} - \frac{m}{r-1} + \frac{2}{r-1} + \frac{r-1}{r-1}$$

$$\frac{1}{p-q} \{ m \cdot \frac{r \cdot (r^{m-1} - r)}{r^m - 1} \cdot q \cdot \frac{r-1}{r-1} - \frac{m.(r-1)}{r - 1} + \frac{2.q.(r-1)}{r-1} + (r-1).q \}$$

$$\frac{1}{p-q} \{ m \cdot r. \frac{r^{m-1} - r}{r^m - 1} \cdot q - m.q+1 \}$$

$$\frac{1}{p-q} \{ m \cdot (\frac{r^m - r^2}{r^m - 1} - 1) q + 1 \}$$

$$\frac{1}{p-q} \{ m (\frac{r^m - r^2 - r^m + 1}{r^m - 1}) q + 1 \}$$

$$\frac{1}{p-q} \{ - m . \frac{(r + 1) . (r - 1) . q}{r^m - 1} + 1 \}$$

$$\frac{1}{p-q} \{ + m \frac{1-r}{r^m-1} + 1 \}$$

$$\frac{1}{p-q} \{ m . \frac{r^m - r}{r^m - 1} + m \frac{(1 - r^m)}{r^m - 1} + 1 \}$$

$$\frac{1}{p-q} \{ m . \frac{r^m - r}{r^m - 1} - m + 1 \}$$

$$\frac{1}{p-q} \{ m . \frac{r^m - r}{r^m - 1} - (m - 1) \}$$

$$= c_{i\ i+1}$$

$$1 + p . c_{i+1\ ,\ i-1} = c_{i\ i-1}$$

$$p . \frac{1}{p-q} \{ m . \frac{r^m - r^{m-2}}{r^m - 1} - 2 \} + 1$$

$$\frac{1}{p-q}\left\{\frac{m.q.(r-1)(r^m+r^{m-1})}{r^m-1} - \frac{2q(r-1)r}{r-1} + q\ (r-1)\right\}$$

$$\frac{1}{p-q}\left\{\frac{m\ .\ q\ .\ r^{m-1}\ .\ (1+r)\ (r-1)}{r^m-1} - 1\right\}$$

$$\frac{1}{p-q}\left\{\frac{m\ .\ q\ .\ r^{m-1}\ (r^2-1)}{r^m-1} - 1\right\}$$

$$\frac{1}{p-q}\left\{\frac{m\ .\ r^{m-1}\ (r-1)}{r^m-1} - 1\right\}$$

$$\frac{1}{p-q}\left\{m\ .\ \frac{r^m-r^{m-1}}{r^m-1} - 1\right\}$$

$$= c_{i\ i-1}$$

A R T I G O S

(1) ONE LEVEL STORAGE SYSTEM
Ire Transactions. Ec-11, 2,223 - 235
T.Kilburn, D.B.G.Edwards, M.J.Lanigan, F.H.Sumner.

(2) STATISTICAL ANALYSIS OF PAGED AND SEGMENTED COMPUTER SYSTEMS
Ieee Transactions, Ec-15, 6, 855-863.

(3) PERFORMANCE EVALUATION OF COMPUTING SYSTEMS WITH MEMORY - HIERARCHIES.
Ieee Transactions on Elect. Comp. Ec-16,6,764-773.
W.Anacker, C.P.Wang.

(4) CONCEPTS FOR BUFFER STORAGE
Ieee Computer Group News, 2,8, 9-13
C.J.Conti

(5) EVALUATION TECHNIQUES FOR STORAGE HIERARQUIES
IBM Systems Journal
J.Gecsey, D.R.Slutz, I.L.Traiger.

(6) VIRTUAL MEMORY
Computing Surveys, 2,3,153 - 189
P.Denning

(7) PRINCIPLES OF OPTIMAL PAGE REPLACEMENT
Journal of the Ass.for Comp.Mach., 18,1,80-93
A.V.Aho, P.J.Denning, J.D.Ullman.

(8) PROPERTIES OF THE WORKING SET MODEL
Com. of the Ass. for Comp. Mach, 11,5,323-333
P.J.Denning, S.C.Schwartz

B I B L I O G R A F I A

(1) KARLIN, S., A First Course in Stochastic - PROCESSES.

(2) HOEL, P.G., Introduction to Stochastic Processes.

(3) PARZEN, E., Stochastic Processes.

(4) BARUCHA-REID, A.T., Elements of the Theory of Markov Processes and their Aplications.

(5) KEMENY, e outros, Denumerable Markov Chains.

(6) KEMENY, e outros, Introduction to Finite Mathematics.

(7) KEMENY, e outros, Finite Mathematical Structures.

(8) KEMENY, e outros, Finite Markov Chains.

(9) GNEDENKO, The Theory of Probability.

(10) FELLER, W., An Introduction to Probability Theory and its Aplications.

(11) BARTLE, R.G., The Elements of Real Analysis.

(12) HARDY, G.H., A Course of Pure Mathematics.

(13) FISZ, M., Probability Theory and Mathematical Statistics.

(14) CHUNG, K.L., Markov Chains With Stationary Transition Probabilities.

(15) KNOPP, K., Infinite Sequences and Series.