A abordagem de martingais para o estudo de ocorrência de palavras em ensaios independentes

Vanessa Masitéli

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA UFSCar-USP

Vanessa Masitéli

A abordagem de martingais para o estudo de ocorrência de palavras em ensaios independentes

Dissertação apresentada ao Departamento de Estatística - Des/UFSCar e ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Estatística - Programa Interinstitucional de Pós-Graduação em Estatística UFSCar-USP

Orientador: ***Prof. Dr. Renato Jacob Gava***

São Carlos
Maio de 2017

The martingale approach in the study of words occurrence in independent experiments

Vanessa Masitéli

Dissertation to be submitted to the Departament of Statistics - Des/UFSCar and the Institute of Mathematics and Computer Sciences - ICMC-USP, as part of the requirements for obtaining master's degree in Statistics at Inter Program Graduate in Statistics UFSCar-USP.

Supervisor: ***Prof. Dr. Renato Jacob Gava***

São Carlos

May 2017

UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa Interinstitucional de Pós-Graduação em Estatística

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado da candidata Vanessa Masitéli, realizada em 07/04/2017:

Renato Gava

Prof. Dr. Renato Jacob Gava
UFSCar

Élcio Lebensztayn.

Prof. Dr. Élcio Lebensztayn
UNICAMP

Geraldine Góes Bosco

Profa. Dra. Geraldine Goes Bosco
USP

Aos meus pais,
José Sanches Masitéli e Joana Aparecida Rosa Masitéli,
em forma de agradecimento.

AGRADECIMENTOS

Agradeço, primeiramente, à Deus. Pela sabedoria e persistência que me foi concedida. Obrigada por permitir a realização de mais um sonho.

Aos meus pais agradeço por sempre me mostrarem o melhor caminho, por abdicarem de alguns sonhos para viver os meus, por me ensinarem, através de exemplos, que, com honestidade, amor e fé, se consegue conquistar o mundo. Agradeço também por acreditarem em meu potencial e me incentivarem em todos os momentos. A minha irmã Viviane Sanches Masitéli agradeço por sempre estar presente, por ser minha certeza de que nunca estarei sozinha. Eles são meus pilares de sustentação.

Ao meu orientador Renato Jacob agradeço por ter aceitado a orientação e por dividir um pouco do seu conhecimento comigo, pelos conselhos acadêmicos e pessoais, pelos fins de semana dedicados às orientações que foram motivadoras na conclusão deste trabalho. Seu incentivo e sua experiência foram de extrema importância no desenvolver desta dissertação.

À docente Vera Tomazella, docente Vera Tomazella agradeço pelo apoio e exemplo durante o curso de pós graduação. Aos docentes Luiz Ernesto Salasar e Geraldine Bosco agradeço pelas sugestões apresentadas durante o exame de qualificação. E à docente Doralice Pinheiro Pereira agradeço pelas correções e apoio durante o desenvolver do trabalho.

Aos amigos da pós graduação Ricardo Ferreira, Eveliny Barroso, Vanessa Rufino, George Pezzott, Danila Silva e Arhur Necchi agradeço pelas horas de estudos que tivemos juntos, pelas discussões acadêmicas, pelos conselhos e os momentos de filosofia que foram essenciais para o meu desenvolvimento pessoal e profissional, por terem aguentado meu mal humor nas vésperas de provas e todas minhas reclamações, por terem sido meu apoio nos momentos difíceis, pelos brindes que fizemos a cada conquista (porque não é só de Lattes que vive o homem). Foram minha família de São Calos e, independente para onde formos, estarão sempre comigo. Agradeço por terem caminhado ao meu lado e deixarem meu caminhar mais leve.

Aos amigos matemáticos Letícia Sanches, Jhony Sá, Juliana Honda, Aline Monteiro, Aneliza Longhi, Daniella Porto, Wanderson Tenório, agradeço por sempre estarem ao meu lado, desde o início da graduação até o final desse ciclo tão importante da minha vida, por aguentarem minhas reclamações, minhas crises de desespero, por cada risada que tivemos juntos, por dividirem suas experiências comigo e terem permitido que eu dividisse as minhas com eles. Em especial, queria agradecer a Lê, por estar ao meu lado quando mais precisei, pelas palavras de apoio, pelas panelas de brigadeiro, por todas as vezes que atendeu aos meus "vem estudar comigo porque não tô conseguindo me concentrar", ela (mais uma vez) foi muito importante nessa fase da minha vida. Agradeço por sempre me apoiarem e fazerem parte de cada uma das minhas conquistas.

À minha amiga de toda vida Isadora Pinheiro Pereira agradeço por ter me acolhido em sua casa no começo dessa jornada e por estar presente nela até o final. Agradeço pelos momentos de descontrações que temos juntas, os quais deixam a minha vida mais leve; pelos conselhos, pelas conversas sobre a vida que sempre me acrescentam conhecimentos, por me escutar em qualquer momento e, independente de lugar, por estar sempre "aqui pro que der e vier."

Aos amigos e familiares Maria Aparecida Rossini, Fernando dos Anjos, Fernanda Rossini, Carolina Fernandes, Carlos Alberto, Marilene Fernandes, Julio César, André Sanches, Emanuele Seicenti, Daiane Seicenti, Camila Trevisani, Fernando Cabreira, Natália Guirão, Fernanda Cavallari, Leandro Christopher, Arthur Pinheiro Pereira, Grace Hung e Arthur Nagae agradeço pelo apoio profissional e psicológico, pelos momentos em que se prontificaram em me escutar sem às vezes nem entender do que eu estava falando, por ajudarem a construir meu alicerce extraindo de mim o que tenho de melhor. Meus etenos agradecimentos por estarem ao meu lado em mais uma conquista! Sou o resultado da confiança e da força de cada um de vocês.

E por fim, agradeço o ambiente acolhedor e a infraestrutura disponibilizada para o desenvolvimento desse trabalho pelos departamentos de Estatística da Universidade Federal de São Carlos e do Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo.

"O que vale na vida não é o ponto de partida e sim a caminhada.
Caminhando e semeando, no fim terás o que colher."
(Cora Coralina)

RESUMO

MASITÉLI, V.. **A abordagem de martingais para o estudo de ocorrência de palavras em ensaios independentes**. 2017. 43 f. Dissertação (Mestrado em em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação (ICMC/USP), São Carlos – SP.

Seja $\{X_n\}$ uma sequência de variáveis aleatórias i.i.d. assumindo valores num alfabeto enumerável. Dada uma coleção de palavras finita, observamos esta sequência até o momento τ em que uma dessas palavras apareça em $X_1, X_2, \ldots$. Neste trabalho utilizamos a abordagem de martingais, introduzida por Li (1980) e Gerber e Li (1981), para estudar o tempo de espera até que uma das palavras ocorra pela primeira vez, o tempo médio de τ e a probabilidade de uma palavra ser a primeira a aparecer.

Palavras-chave: martingal; cadeias de Markov; probabilidade condicional; esperança condicional; tempo de espera..

ABSTRACT

MASITÉLI, V.. **A abordagem de martingais para o estudo de ocorrência de palavras em ensaios independentes**. 2017. 43 f. Dissertação (Mestrado em em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação (ICMC/USP), São Carlos – SP.

Let $\{X_n\}$ be a sequence of i.i.d. random variables taking values in an enumerable alphabet. Given a finite collection of words, we observe this sequence till the moment τ at which one of these words appears as a run. In this work we apply the martingale approach introduced by Li (1980) and Gerber e Li (1981) in order to study the waiting time until one of the words occurs for the first time, the mean of τ and the probability of a word to be the first one to appear.

Key-words: martingale; Markov chain; conditional probability; conditional expectation; waiting time..

SUMÁRIO

CAPÍTULO

1

INTRODUÇÃO

Suponha que letras de um alfabeto $\mathscr{A}$ sejam aleatoriamente alinhadas, uma após a outra, indefinidamente. Considere uma palavra qualquer formada por letras de $\mathscr{A}$. Quantas letras temos de alinhar sucessivamente até que esta palavra apareça na sequência de letras? Qual o valor esperado e a distribuição do número de letras até que esta palavra ocorra? Por outro lado, considere uma coleção de palavras finita. Qual delas será a primeira a ocorrer na sequência de letras? Quais são a esperança e a distribuição do número de letras até que uma dessas palavras seja a primeira a ocorrer? Qual é a probabilidade de que uma determinada palavra seja a primeira a aparecer?

Matematicamente, seja $\{X_n\}_{n\geq 1}$ uma sequência de variáveis aleatórias independentes assumindo valores no alfabeto $\mathscr{A}$, que será finito ou infinito enumerável. Definimos como palavra qualquer sequência finita de elementos de $\mathscr{A}$. Inicialmente, consideremos uma coleção finita $\mathscr{C} = \{A_1, A_2, ..., A_k\}$ de palavras, podendo ser de tamanhos diferentes. Considere $A_1 = l_1...l_p$, isto é, A_1 é uma palavra de tamanho $p \in \mathbb{N}$. Seja τ_{A_1} o tempo de espera até que A_1 apareça, isto é,

$$\tau_{A_1} = \min\{n \geq 1 : X_n = l_p, ..., X_{n-p+1} = l_1\}.$$

Defina o tempo de espera

$$\tau_{Aq} = \min\{\tau_{A_1}, ..., \tau_{A_k}\}, \text{ com } q = 1, ..., k. \tag{1.1}$$

Assumimos que nenhuma palavra de $\mathscr{C}$ contenha outra palavra como subpalavra.

No Capítulo 2 introduzimos brevemente as ideias de probabilidade e esperança condicional. No Capítulo 3, enunciamos as definições de cadeias de Markov e martingais, além de alguns resultados a serem usados nos próximos capítulos. No Capítulo 4, estudamos os resultados principais de Pozdnyakov e Kulldorff (2006),Li (1980) e Gerber e Li (1981). Nas Seções 4.1 e 4.3 apresentamos a abordagem de martingais de Pozdnyakov e Kulldorff (2006) para o problema do tempo de espera para palavras. Inicialmente para uma palavra apenas, depois para um conjunto de palavras finito. Na Seção 4.4 apresentamos o teorema principal do artigo Li (1980). Por fim, na Seção 4.2 analisamos a imersão da cadeia de Markov de Gerber e Li (1981) para mostrar que o sistema linear do Teorema 7 possui uma única solução.

CAPÍTULO

2

PROBABILIDADE E ESPERANÇA CONDICIONAL

Neste capítulo revisamos as definições e propriedades de probabilidade e esperança condicional, que são fundamentais para as próximos capítulos. Algumas demonstrações nessa seção serão omitidas, porém podem ser encontradas nos livros de Ross (2002), Grimmett e Stirzaker (2001) e Gut (2009).

2.1 Probabilidade Condicional

Sejam $(\Omega, \mathscr{F}, \mathbb{P})$ um espaço de probabilidade e A, B eventos nesse espaço. Suponha que o evento A ocorra. Queremos encontrar a probabilidade do evento B ocorrer dado que o evento A ocorreu. Tal probabilidade é chamada de probabilidade condicional e denotada por $\mathbb{P}(B|A)$, isto é,

$$P(B|A) = \frac{\mathbb{P}(B \bigcap A)}{\mathbb{P}(A)}, \tag{2.1}$$

Definição 1. Sejam X e Y variáveis aleatórias que assumem valores em um conjunto discreto e que são distribuídas conjuntamente com função massa de probabilidade $p(x,y)$. A função de probabilidade condicional de $X|Y = y$ é dada por

$$p_{x|y}(x|y) = \mathbb{P}(X = x|Y = y) = \frac{p(x,y)}{p_Y(y)}, x \in X \tag{2.2}$$

para todo $y \in Y$ tal que $p_Y(y) = \mathbb{P}(Y = y) > 0$. E a função de distribuição condicional de $X|Y = y$ é dada por

$$F_{X|Y=y}(x) = \sum_{z \leq y} p_{X|Y=y}(z). \tag{2.3}$$

em que $P_y(y) = \sum_{x \in X} \mathbb{P}_{X,Y}(x,y)$, $\mathbb{P}_X(x) = \sum_{y \in Y} \mathbb{P}_{X,Y}(x,y)$ são as funções de massa marginais.

Exemplo 1. Uma moeda é jogada duas vezes. Suponha que os quatros pontos no espaço amostral $S = \{(h,h),(h,t),(t,h),(t,t)\}$ sejam igualmente prováveis, onde h representa cara e t coroa. Qual é a probabilidade condicional de que dê cara em ambas as jogadas dado que:
(i) Dê cara na primeira jogada?
(ii) Dê cara em pelo menos uma das jogadas?
Seja $X = \{(h,h)\}$ o evento em que ambas as jogadas dão cara, $Y = \{(h,h),(h,t)\}$ o evento em que aparece cara na primeira moeda e $A = \{(h,h),(h,t),(t,h)\}$ o evento em que aparece cara em pelo menos uma jogada.
(i) $\mathbb{P}(X|Y) = \dfrac{p(x,y)}{p_Y(y)} = \dfrac{p(\{(h,h)\})}{p(\{(h,h),(h,t)\})} = \dfrac{1/4}{2/4} = \dfrac{1}{2}$
(ii) $\mathbb{P}(X|A) = \dfrac{p(x,y)}{p_A(a)} = \dfrac{p(\{(h,h)\})}{p(\{(h,h),(h,t),(t,h)\})} = \dfrac{1/4}{3/4} = \dfrac{1}{3}$

2.2 Esperança Condicional

Definição 2. Sejam X e Y variáveis aleatórias com distribuição conjunta. A esperança condicional de Y dado X, $Y|X = x$, é

$$\mathbb{E}[Y|X=x] = \sum_y y p_{Y|X=x}(y), \quad \text{para o caso de ambas serem (absolutamente) discretas}$$

Para este trabalho, usaremos a Esperança Condicional para os casos absolutamente discretos. Temos os casos de Esperança Condicional para funções contínuas, estes casos deixaremos a cargo do leitor. Facilmente encontrados no livro de Ross (2002).

Exemplo 2. Sejam X e Y variáveis aleatórias binomiais independentes com parâmetros idênticos n e p. Calcule a esperança condicional de $X|X+Y=m$.
Note que $X+Y \sim Binomial(2n,p)$. Assim,

$$\begin{aligned}
\mathbb{P}(X=k|X+Y=m) &= \frac{\mathbb{P}(X=k, X+Y=m)}{\mathbb{P}(X+Y=m)} \\
&= \frac{\mathbb{P}(X=k)\mathbb{P}(Y=m-k)}{\mathbb{P}(X+Y=m)} \\
&= \frac{\binom{n}{k} p^k (1-p)^{n-k} \binom{n}{m-k} p^{m-k} (1-p)^{n-m+k}}{\binom{2n}{m} p^m (1-p)^{2n-m}} \\
&= \frac{\binom{n}{k}\binom{n}{m-k}}{\binom{2n}{m}}
\end{aligned}$$

E concluímos que $X|X+Y=m \sim$ Hipergeométrica.

Portanto,

$$\mathbb{E}[X|X+Y=m]=\frac{m}{2}$$

Proposição 1. Sejam X, Y, Y_1, Y_2 variáveis aleatórias, g uma função e c uma constante. Então:
(1) $\mathbb{E}[c|X=x]=c$
(2) $\mathbb{E}[Y_1+Y_2|X=x]=\mathbb{E}[Y_1|X=x]+\mathbb{E}[Y_2|X=x]$
(3) $\mathbb{E}[cY|X=x]=\mathbb{E}[Y|X=x]c$
(4) $\mathbb{E}[g(X,Y)|X=x]=\mathbb{E}[g(x,Y)|X=x]$
(5) $\mathbb{E}[g(x)Y|X]=g(X)\mathbb{E}[Y|X]$
(6) $\mathbb{E}[Y|X=x]=\mathbb{E}[Y]$, se X e Y são independentes.

Teorema 1. Suponha que $\mathbb{E}[|Y|]<\infty$. Então,

$$\mathbb{E}[\mathbb{E}(Y|X)]=\mathbb{E}[Y] \tag{2.4}$$

A demonstração dos itens da proposição acima pode ser encontrada nos livros Gut (2009) e Ross (2002), nos capítulos 2 e 4, respectivamente.

CAPÍTULO

3

PROCESSOS ESTOCÁSTICOS

Antes de começarmos o estudo dos problemas citados na introdução deste trabalho precisamos entender os conceitos básicos de processos estocásticos. Usaremos como base o Livro Norris (1998) para cadeias de Markov e os livros de Grimmett e Stirzaker (2001) e Ross *et al.* (1996) para definições e propriedades de martingais.

3.1 Cadeias de Markov

Processos estocásticos são caracterizados pela relação de dependência entre suas variáveis.

Existem processos estocásticos a tempo discreto e a tempo contínuo. Para o desenvolvimento desse trabalho usamos somente os conceitos de processos estocásticos a tempo discreto e começaremos essa seção definindo tais processos. Em seguida definimos e damos algumas propriedades de cadeia de Markov retiradas do livro de Norris (1998). As demonstrações serão omitidas em muitas dessas propriedades.

Definição 3. Um processo estocástico a tempo discreto é uma sequência de variáveis aleatórias $\{X_n\}_{n\geq 0}$, $n \in \mathbb{N}$, definidas no mesmo espaço de probabilidade e com valores em algum conjunto enumerável S, chamado de espaço de estados.

Denomina-se "tempo discreto"devido ao subíndice n ser interpretado como unidades de tempo e $\{X_n\}_{n\geq 0}$ representar a evolução de um certo fenômeno ao longo desse tempo. Exemplificando, se lançamos sucessivamente uma moeda e X_n representa o resultado do n-ésimo lançamento (com opção de ser cara ou coroa), a sequência $\{X_n\}_{n\geq 1}$ é um processo estocástico a tempo discreto.

Definição 4. Dizemos que $\{X_n\}_{n\geq 0}$ é uma *Cadeia de Markov* com distribuição inicial λ e matriz de transição $\mathbb{P}$ se, para $n \geq 0$ e $i_0, \ldots, i_{n+1} \in I,$ com $I \in \mathbb{N}$

(i) $\mathbb{P}(X_0 = i_0) = \lambda_{i_0}$.

(ii) $\mathbb{P}(X_{n+1} = i_{n+1} | X_0 = i_0, ..., X_n = i_n) = p_{i_n, i_{n+1}}$

Notação: Markov$(\lambda, \mathbb{P})$.

A característica de uma Cadeia de Markov, tanto a tempo discreto quanto a tempo contínuo, é a falta de memória. Assim, o comportamento do processo no futuro depende somente do estado presente, isto é, o futuro dado o presente e o passado, só depende do presente.

Exemplo 3 (Passeio aleatório em $\mathbb{Z}$)**.** Suponha que queiramos modelar o seguinte fenômeno. Em um instante de tempo $n = 0$ temos uma partícula no vértice $0 \in \mathbb{Z}$. A cada instante discreto de tempo a partícula pula um vértice à direita com probabilidade p ou um vértice à esquerda com probabilidade $1 - p$.

Se X_n representa a posição da partícula no n - ésimo instante de tempo. Então, X_n toma valores em $\mathbb{Z}$ e a sequência $(X_n)_{n \geq 0}$ é um processo estocástico a tempo discreto. Esse processo é chamado de passeio aleatório em $\mathbb{Z}$

Note que para saber a posição da partícula no instante de tempo n é suficiente saber sua posição no tempo $n - 1$, independente das possíveis posições da partícula em instantes anteriores.

Teorema 2. Um processo estocástico a tempo discreto (X_n) é Markov $(\lambda, \mathbb{P})$ se, e somente se, para todo $i_o, ..., i_N \in I$, com $I \in \mathbb{N}$

$$\mathbb{P}(X_0 = i_0, ..., X_N = i_N) = \lambda_{i_0} p_{i_0 i_1} p_{i_1 i_2} ... p_{i_{N-1}, i_N}. \tag{3.1}$$

Demonstração. $(\Rightarrow)$Suponha que $(X_n)_{n \geq 0}$ é Markov $(\lambda, \mathbb{P})$.

Defina $A_k = \{X_k = i_k\}$.

Assim,

$$\{X_0 = i_0, ..., X_N = i_N\} = \bigcap_{k=0}^{N} A_k$$

Logo,

$$\begin{aligned}
\mathbb{P}(X_0 = i_0, ..., X_N = i_N) = {} & \mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1}, ..., X_0 = i_0)\mathbb{P}(X_{n-1} = i_{n-1} | X_{n-2}, ..., X_0 = i_0) \\
& ...\mathbb{P}(X_1 = i_1 | X_0 = i_0)\mathbb{P}(X_0 = i_0) \\
& = \lambda_{i_0} p_{i_0 i_1} p_{i_1 i_2} ... p_{i_{n-1} i_n}
\end{aligned}$$

$(\Leftarrow)$ Se (3.1) é válido para N, então somando em i_N temos que

$\mathbb{P}(X_0 = i_0, ..., X_{N-1} = i_{N-1}) = \lambda_{i_0} p_{i_0 i_1} p_{i_1 i_2} ... p_{i_{n-2} i_{n-1}}$

Por indução,

$\mathbb{P}(X_0 = i_0, ..., X_N = i_N) = \lambda_{i_0} p_{i_0 i_1} p_{i_1 i_2} ... p_{i_{n-1} i_n}$

Em particular, $\mathbb{P}(X_0 = i_0) = \lambda_{i_0}$

E portanto,

$$\begin{aligned}
\mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n, ..., X_0 = i_0) &= \frac{\mathbb{P}(X_0 = i_0, ..., X_{n+1} = i_{n+1})}{\mathbb{P}(X_0 = i_0, ..., X_n = i_n)} \\
&= \frac{\lambda_0 p_{i_0 i_1} ....p_{i_{n-1} i_{in}} p_{i_n i_{n+1}}}{\lambda_{i_0} p_{i_0 i_1} p_{i_1 i_2} ... p_{i_{n-1} i_n}} \\
&= p_{i_n i_{n+1}}.
\end{aligned}$$

□

Teorema 3. Seja $(X_n)_{n \geq 0}$ Markov $(\lambda, \mathbb{P})$, com $\mathbb{P} = (\mathbb{P}_{ij})_{I \times I}$ e $\lambda = (\lambda_1, \lambda_2, ...)$.

Defina

$$(\lambda \mathbb{P})_j = \sum_i \lambda_i \mathbb{P}_{ij} = \mathbb{P}(X_1 = j)$$

e

$$(\mathbb{P}^2)_{ik} = \sum_{j \in I}^{i} \mathbb{P}_{ij} \mathbb{P}_{jk} = \mathbb{P}(X_2 = k | X_0 = i).$$

Então, $\forall n, m \geq 0$,

(i) $\mathbb{P}(X_n = j) = (\lambda \mathbb{P}^n)_j$

(ii) $\mathbb{P}_i(X_n = j) = p_{ij}^{(n)} = \mathbb{P}(X_{n+m} = j | X_m = i) = (\mathbb{P}^n)_{ij}$.

Demonstração. Similarmente, defina $\mathbb{P}^n = \mathbb{P}\mathbb{P}^{n-1}$, onde $\mathbb{P}^0 =$ Identidade e escrevemos $p_{ij}^{(n)} = P(X_n = j | X_0 = i)$.

$(\Rightarrow)$

$$\begin{aligned}
\mathbb{P}(X_n = j) &= \sum_{i_0 \in I} \sum_{i_1 \in I} ... \sum_{i_{n-1} \in I} \mathbb{P}(X_0 = i_0, X_1 = i_1, ..., X_{n-1} = i_{n-1}, X_n = j) \\
&= \sum_{i_0 \in I} \sum_{i_1 \in I} ... \sum_{i_{n-1} \in I} \lambda_{i_0} p_{i_0 i_1} p_{i_1 i_2} ... p_{i_{n-1} i_j} \\
&= (\lambda \mathbb{P}^n)_j
\end{aligned}$$

$(\Leftarrow)$

Tome $Y_n = X_{n+m}, \forall n \geq 0$.

Logo, segue que Y_n é Markov $(\lambda_i, \mathbb{P})$, onde $\delta_i = \mathbb{P}(Y_0 = i) = \mathbb{P}(X_m = i) = 1$ e de o resultado segue de (i). □

Exemplo 4. A cadeia mais comum de dois estados tem matriz de transição

$$P = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix},$$

$$\begin{aligned} P^2 &= \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix} \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix} \\ &= \begin{bmatrix} (1-\alpha)^2+\alpha\beta & \alpha(1-\alpha)+\alpha(1+\beta) \\ \beta(1-\alpha)+\beta(1-\beta) & (1-\beta)^2+\alpha\beta \end{bmatrix} \\ &= \frac{1}{\alpha+\beta}\begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix} + \frac{(1-\alpha-\beta)^2}{\alpha+\beta}\begin{bmatrix} \alpha & -\alpha \\ -\beta & \beta \end{bmatrix} \end{aligned}$$

Tome

$$A = \begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix}$$

e

$$B = \begin{bmatrix} \alpha & -\alpha \\ -\beta & \beta \end{bmatrix}$$

Assim,

$$P^2 = \frac{1}{\alpha+\beta}A + \frac{(1-\alpha-\beta)^2}{\alpha+\beta}B$$

Provemos por indução que

$$P^n = (\alpha+\beta)^{-1}[A+(1-\alpha-\beta)^n B] \tag{3.2}$$

Suponha que 3.2 vale para n.

$$\begin{aligned} p^{n+1} &= p^n p \\ &= \{(\alpha+\beta)^{-1}[A+(1-\alpha-\beta)^n B]\}P \\ &= (\alpha+\beta)^{-1}[AP+(1-\alpha-\beta)^n BP] \\ &= (\alpha+\beta)^{-1}[A+(1-\alpha-\beta)^{n+1}B], \end{aligned}$$

em que

$$AP = \begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix} \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix} = \begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix} = A$$

e

$$BP = \begin{bmatrix} \alpha & -\alpha \\ -\beta & \beta \end{bmatrix} \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix} = (1-\alpha-\beta)B$$

Note que $|1-\alpha-\beta| < 1$. Logo, quando $n \to \infty$,

$$(1-\alpha-\beta)^n B \to \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \text{isto é,}$$

$$p^n \to \begin{bmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \\ \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \end{bmatrix}.$$

3.1.1 Estrutura de Classe

Podemos quebrar o espaço de estados de uma cadeia de Markov em pedaços que facilitam entender a cadeia. Para isso identificamos as classes comunicantes da cadeia.

Dizemos que "i vai para j"e escrevemos $i \to j$ se

$$\mathbb{P}_i(X_n = j \text{ para algum } n \geq 0).$$

Dizemos que i e j são comunicantes se $i \to j$ e $j \to i$.

Notação: $i \leftrightarrow j$.

Teorema 4. Para i e j, estados distintos, são equivalentes:

(i) $i \to j$

(ii) $p_{i_0 i_1} p_{i_1 i_2} ... p_{i_{n-1} i_n} > 0$, para alguns estados $i_0, i_1, ..., i_n \in I$, com $i_0 = i$ e $i_n = j$

(iii) $p_{ij}^{(n)} > 0$, para algum $n \geq 0$.

A demonstração pode ser encontrada na página 10 de Norris (1998).

Note que " $\leftrightarrow$ " define uma classe de equivalência.

(a) $i \leftrightarrow i$

(b) $i \to j$ e $j \to k \Rightarrow i \to k$ ($p_{ik}^{(n+m)} \geq p_{ij}^{(n)} p_{(m)} > 0$)

(c) $i \leftrightarrow j$, então $j \leftrightarrow i$

Definição 5. Dizemos que uma classe $C \subset I$ é fechada se $i \in C$ e $i \to j$, então $j \in C$.

Quando I é uma única classe, dizemos que a cadeia é irredutível e quando uma classe fechada contiver apenas um elemento ($C = \{i\}$), dizemos que i é um estado absorvente.

3.1.2 Tempo de parada para uma cadeia de Markov

Definição 6. Uma variável aleatória $T : \Omega \to \{0,1,2,...\} \bigcup \{\infty\}$ é chamada de tempo de parada para $\{X_n\}_{n\geq 0}$ se o evento $\{T = n\}$ depende somente de $\{X_0,...,X_n\}, n = 0,1,2,....$ Isto é, se conhecermos a realização de $X_0,...,X_n$, então sabemos se T ocorre ou não.

Exemplo 5. Primeiro tempo de passagem:

$T = \inf\{n \geq 1; X_n = j\}$ é tempo de parada.

De fato,

$\{T = n\} = \{X_1 \neq j, X_2 \neq j, ..., X_{n-1} \neq j, X_n = j\}$.

Exemplo 6. Último tempo de passagem:

$S = \sup\{n \geq 0; X_n = j\}$.

Note que o evento $\{S = n\}$ não é tempo de parada pois depende de $X_{n+1}, X_{n+2}, ...$

Mais especificamente,

$\{X_{n+1} \neq j, X_{n+2} \neq j, ...\} \subset \{S \leq n\}$.

Exemplo 7. (Passeio aleatório simples simétrico em $\mathbb{Z}$)

Seja $\{X_n\}_{n\geq 1}$ uma sequência de variáveis aleatórias i.i.d. tal que $\mathbb{P}(X_n = 1) = \frac{1}{2} = \mathbb{P}(X_n = -1)$. Considere $S_0 = 0$ e, para $n \geq 1$, defina

$$S_n = \sum_{i=0}^{n} X_n.$$

Já vimos que S_n é um passeio aleatório simples; quando $p = 1/2$, dizemos que o passeio é simétrico. Além disso, S_n é martingal.

Agora considere $\max\{\varnothing\} = 0$ e defina

$$\gamma = \max\{n \leq 15 : S_n = 1\}$$

A variável γ nos diz a última visita ao vértice 1 para $n \in \{0,1,...,15\}$. Observe que o evento $\{\gamma \leq 5\}$ não depende apenas dos 5 primeiros ensaios, mas sim dos 15 primeiros ensaios. Assim, $\{\gamma \leq i\}$ não depende apenas dos i primeiros ensaios. Portanto, γ não é tempo de parada.

Propriedade 1 (Propriedade forte de Markov)**.** Sejam $\{X_n\}_{n\geq 0}$ Markov(λ, P) e T um tempo de parada de $\{X_n\}$. Então, dado o evento $\{T < \infty, X_T = i\}$, $\{X_{T+n}\}_{n\geq 0}$ é Markov(λ_i, P) e é independente de $X_0, X_1, ..., X_T$.

A demonstração desta propriedade pode ser encontrada na página 20 de Norris (1998).

Propriedade 2 (Propridade Fraca de Markov)**.** Seja $\{X_n\}_{n\geq 0}$ Markov(λ, P). Dado o evento $\{X_n = 1\}$, o processo $\{X_{m+n}\}_{n\geq 0}$ é Markov(δ_i, P) é independente de $X_0, \ldots, X_m$, em que

$$\delta_i = (\delta_{ij})_{j\in I} \text{ e } \delta_{ij} = \begin{cases} 1, & \text{se } i = j \\ 0, & \text{se } i \neq j \end{cases}$$

3.2 Martingais

Definição 7. Seja $\{M_n\}_{n\geq 1}$ um processo aleatório tal que $E|M_n| \leq \infty$. Chamamos $\{M_n\}_{n\geq 1}$ de martingal se, para todo $n \geq 1$,

$$E[M_{n+1}|M_1, \ldots, M_n] = M_n. \tag{3.3}$$

Em palavras, a definição de martingal nos diz que, dada a informação sobre os valores do processo $\{M_n\}_{n\geq 1}$ até o instante n, o instante $n+1$ em média terá o mesmo valor que o instante anterior.

Pensemos em M_n como uma fortuna do jogador após o n-ésimo jogo, então (3.3) afirma que essa fortuna esperada na $(n+1)$-ésima jogada é igual a sua fortuna na n-jogada independente do que aconteceu nas anteriores.

Usando as propriedades de esperança condicional e a definição de martingal temos que,

$$E[M_{n+1}] = E[E(M_{n+1}|M_1, \ldots, M_n)] = E[M_n]$$

Portanto, um martingal tem média constante para todo $n \geq 1$.

Exemplo 8. Passeio aleatório simples.
Uma partícula dá um passo para direita ou um passo para a esquerda com probabilidade p e $q = 1 - p$, respectivamente. Assumindo independência e que $S_n = X_1 + X_2 + \ldots + X_n$ é a posição da partícula após n passos, com $\mathbb{E}|S_n| < n$. Temos,

$$\mathbb{E}[S_{n+1}|X_1, X_2, \ldots, X_n] = S_n + (p - q)$$

De fato,
Temos que

$$S_{n+1} = \begin{cases} \sum_{i=1}^{n} X_i + 1, & \text{com probabilidade} \quad p \\ \sum_{i=1}^{n} X_i - 1, & \text{com probabilidade} \quad 1-p \end{cases},$$

Assim,

$$\begin{aligned}
E[S_{n+1}|X_1,...,X_n] &= p\left(\sum_{i=1}^{n} X_i + 1\right) + (1-p)\left(\sum_{i=1}^{n} X_i - 1\right) \\
&= p\sum_{i=1}^{n} X_i + p + \sum_{i=1}^{n} X_i - 1 - p\sum_{i=1}^{n} X_i + p \\
&= S_n + (p-q)
\end{aligned}$$

em que $Y_n = S_n - (p-q)$ é martingal com respeito a X.

Teorema 5 (Teorema da parada ótima de Doob)**.** Sejam τ um tempo de parada e $\{M_n\}_{n\geq 1}$ um martingale para o processo $\{X_n\}_{n\geq 1}$. Se $\mathbb{E}(\tau) < \infty$ e existe um número real $k > 0$ tal que

$$|M_{n+1} - M_n| \leq k \text{ para todo } n \,,$$

então M_τ é integrável e

$$\mathbb{E}(M_\tau) = \mathbb{E}(M_1)$$

A demonstração pode ser encontrada no livro de Williams (1991) página 100.

O teorema de parada de Martingales (Teorema de Doob) diz que em um jogo justo, se um jogador usa um tempo de parada para decidir quando parar, sua fortuna esperada será igual a inicial.

Corolário 1 (Equação de Wald)**.** Se X_i, $i \geq 1$, são i.i.d. com $\mathbb{E}|X| < \infty$ e se τ é um tempo de parada para $X_1, X_2, ...$, com $\mathbb{E}[\tau] < \infty$, então

$$\mathbb{E}\left[\sum_{i=1}^{\tau} X_i\right] = \mathbb{E}[\tau]\mathbb{E}[X]$$

Demonstração. Demonstração encontrada na página 300 livro de Ross *et al.* (1996). □

CAPÍTULO

4

ABORDAGEM DE MARTINGAIS

Neste capítulo vamos introduzir a abordagem de Li (1980) e simplificada por Pozdnyakov e Kulldorff (2006) para o cálculo do tempo de espera médio. Para este trabalho iniciaremos o estudo para uma palavra, em seguida vamos trabalhar com um grupo de k palavras usando o teorema principal desse trabalho apresentado por Li (1980).

4.1 O tempo de espera médio para uma palavra

Primeiro consideremos o caso simples, no qual temos uma palavra $A_1 = l_1 l_2 ... l_m$ e calcularemos $\mathbb{E}(\tau_{A_1})$, em que $\mathbb{P}(X_n = l_i) > 0$, para $i = 1, ..., m$.

Por simplicidade, pensemos em times de apostas para construir a ideia dos lançamentos das letras que construirá as palavras.

Suponha que estamos em um cassino e que X_n denote o resultado da n-ésima rodada, para $n = 1, 2, ...$ O jogo acaba no momento aleatório τ_{A_1} em que a palavra A_1 aparece pela primeira vez. Atribuímos um time de apostas (formado por jogadores) para a palavra A_1 da seguinte maneira:

Suponha que um novo jogador entre no jogo a cada instante de tempo $n = 1, 2, ...$ e que aposta 1 real no evento $X_n = l_1$. Se ele perde, deixa o jogo com 0 reais. Se ele ganha, ele recebe

$$\frac{1}{\mathbb{P}(X_n = l_1)} \text{ reais .}$$

Em seguida, aposta o montante total que ele recebeu no evento $X_{n+1} = l_2$ Analogamente, se ele perde deixa o jogo com 0 reais. E se ele ganhar pela segunda vez consecutiva recebe

$$\frac{1}{\mathbb{P}(X_n = l_1)} \times \frac{1}{\mathbb{P}(X_{n+1} = l_2)} \text{ reais .}$$

E assim por diante, até completar as m letras da palavra A.

Neste caso, ele sairá do jogo com

$$\frac{1}{\mathbb{P}(X_n = l_1)} \times \frac{1}{\mathbb{P}(X_{n+1} = l_2)} \times .... \times \frac{1}{\mathbb{P}(X_{n+(m-1)} = l_m)} \text{ reais .}$$

Note que se trata de um jogo justo e para todo $n \geq 1$ existe no máximo m jogadores apostando. Isto é, um martingale.

De fato, suponha que no instante n o jogador tenha u reais e no instante $n+1$ aposta na letra l, então a fortuna no instante $n+1$ será

$$F_{n+1} = \frac{u}{\mathbb{P}(X_{n+1} = l)} \mathbb{I}_{X_{n+1}=l}$$

Assim,

$$\mathbb{E}(F_{n+1}) = \frac{u}{\mathbb{P}(X_{n+1} = l)} \mathbb{P}(X_{n+1} = l) = u$$

Definição 8. Sejam $\{X_n\}_{n\geq 0}$ variáveis aleatórias discretas fixadas e Σ um conjunto de valores possíveis para $\{X_n\}_{n\geq 0}$. Sejam agora, $A = (a_1, ..., a_m)$ e $B = (b_1, ..., b_n)$ duas sequências em Σ. Para cada par (i, j) de inteiros, definimos

$$\delta_{ij} = \left[\begin{array}{cc} [\mathbb{P}(X_n = b_j)]^{-1} & \text{se } 1 \leq i \leq m \text{ , } 1 \leq j \leq n \text{ e } a_i = b_j \\ 0 & \text{caso contrário} \end{array} \right. ,$$

Assim, definimos

$$A * B = \delta_{11}\delta_{22}...\delta_{mm} + \delta_{21}\delta_{32}...\delta_{m,m-1} + ... + \delta_{m1} \tag{4.1}$$

Pensando no time de apostas, A seria os lançamentos do cassino e B os lançamentos dos jogadores. $A * B$ é o tempo que temos que esperar para sobreposição de B em A.

Em palavras, sejam $A = THHT$ e $B = THT$.

Pela definição, todos os jogadores começam apostando na primeira letra de A (T) e cada jogador lança a letra de acordo com sua posição no jogo (1º jogador lança a primeira letra de B, segundo jogador incia com a 2º letra de B, ...)

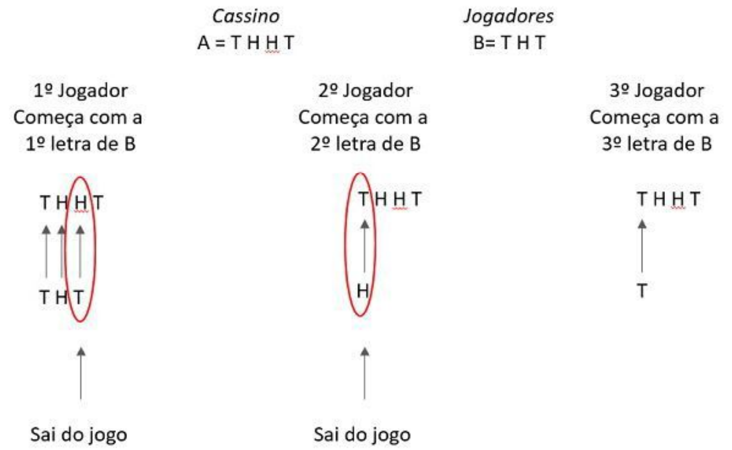
Exemplo 9. Seja $\Sigma = \{H, T\}$, com $\mathbb{P}(X_n = T) = \frac{1}{2} = \mathbb{P}(X_n = H)$.

Sejam $A = THHT$ e $B = THT$. Assim,

$$A * B = \frac{1}{1/2} 0 + 0 + \frac{1}{1/2} = 2$$

e

$$A * A = \frac{1}{1/2} \times \frac{1}{1/2} \times \frac{1}{1/2} \times \frac{1}{1/2} + 0 + \frac{1}{1/2} \times \frac{1}{1/2} + 0 = 16 + 4 = 20$$



Temos que τ_A é a quantidade total de dinheiro gasto pelos apostadores até o tempo τ_A e $A * A$ é a quantidade de dinheiro paga pelo cassino ao time de apostadores da palavra A. Logo, o lucro líquido do cassino no momento τ_A, denotado por S_{τ_A}, é dado por

$$S_{\tau_A} = \tau_A - A * A \tag{4.2}$$

Abaixo mostraremos que o tempo de espera para qualquer palavra tem esperança finita e portanto, tem probabilidade nula de ser infinito.

Sejam $\mathscr{A}$ o nosso alfabeto finito e $A = l_1 \ldots l_m$. Como anteriormente, τ_A é o tempo de espera até que ocorra A pela primeira vez na sequência $\{X_n\}_{n \geq 1}$. Defina,

$$T = \min\{q \geq 1 : X_{m(q+1)} = l_m, X_{n-1} = l_{m-1}, \ldots, X_{mq+1} = l_1\}$$

Isto é, a variável T observa a primeira ocorrência de A a cada bloco de m variáveis.

$$X_1 X_2 \ldots X_m | X_{m+1} X_{m+2} \ldots X_{2m} | X_{2m+1} X_{2m+2} \ldots X_{3m} | X_{3m} \ldots$$

Observe que $\tau_A \leq mT$ e que T tem distribuição geométrica. Logo,

$$\mathbb{E}[\tau_A] \leq \mathbb{E}[mT] = m\mathbb{E}[T] < \infty$$

Pois $\mathbb{E}[T] < \infty$ e com isso, $\mathbb{P}(\tau_A = \infty) = 0$.

E além disso o ganho total de cada jogador está limitado por

$$\prod_{i=1}^{m} \frac{1}{\mathbb{P}(X_n = l_i)}$$

pois, existem no máximo m jogadores no tempo n.

O que garante as condições do Teorema 5. Desse modo,

$$0 = \mathbb{E}[S_0] = \mathbb{E}[S_{\tau_A}] = \mathbb{E}[\tau_A] - A * A \Rightarrow \mathbb{E}[\tau_A] = A * A \tag{4.3}$$

Logo, com as condições dadas acima temos:

Proposição 2. Seja $\mathbb{E}(\tau_A)$ o tempo esperado para ocorrer a sequência A. O tempo médio de espera para atingir A sem uma sequência inicial é dado por,

$$\mathbb{E}(\tau_A) = A * A$$

Exemplo 10. O DNA é uma molécula composta por Ácido Desoxirribonucleico e é responsável por armazenar toda a informação genética de um organismo. Essa carga genética encontra-se nos cromossomos de todas as células de um organismo e pode ser transmitida para descendentes. Além disso, é responsável por controlar todo o funcionamento da maquinaria do organismo. O DNA é formado por uma fita dupla composta pelo sequenciamento de 4 moléculas orgânicas, chamadas de Bases Nitrogenadas: A (Adenina), T (Timina), C (Citosina) e G (Guanina). O código genético é a relação entre a sequência de bases nitrogenadas e os aminoácidos que compõem a proteínas formada. Suponha que um código genético seja formado por $A_1 = ACGTAC$ e $A_2 = CAACG$. Seja $\{X_n\}_{n\geq 1}$ uma sequência de variáveis aleatórias i.i.d. tal que $\mathbb{P}(X_n = A) = 1/3, \mathbb{P}(X_n = C) = 1/4, \mathbb{P}(X_n = G) = 1/4$ e $\mathbb{P}(X_n = T) = 1/6$. Então, $\mathbb{E}(\tau_{A_1})$ e $\mathbb{E}(\tau_{A_2})$ serão,

$$\begin{aligned}
\mathbb{E}[\tau_{A_1}] &= A_1 * A_1 = 3\times 4\times 4\times 6\times 3\times 4 + 0 + 0 + 0 + 3\times 4 + 0 = 3468 \text{ e} \\
\mathbb{E}[\tau_{A_2}] &= A_2 * A_2 = 4\times 3\times 3\times 4\times 4 + 0 + 0 + 0 + 0 = 576
\end{aligned}$$

Na próxima seção vamos apresentar os casos de tempo de espera para várias palavras, para isso usaremos o lema apresentado abaixo.

4.2 Tempo de primeira passagem em uma Cadeia de Markov

Neste seção, estudaremos as ideias de Gerber e Li (1981) sobre o tempo esperado em uma cadeia de Markov. Seja $X_0, X_1, X_2, \ldots$ uma cadeia de Markov com matriz de transição estacionária, isto é, $\mathbb{P}(X_{n+1} = x_{n+1}|X_n = x_n) = \mathbb{P}(X_1 = x_1|X_0 = x_0)$. Denotemos os estados por $a, b, c, \ldots$ e assumimos que o espaço de estados é finito.

Denote

$$\tau_{ab} = \min\{t : X_t = b|X_0 = a\}$$

para o tempo de espera para o estado b quando a cadeia de Markov começa no estado a e

$$e_{ab} = \mathbb{E}[\tau_{ab}],$$

para o tempo médio de espera. Assumiremos que os estados são comunicantes e que como o espaço de estados é finito temos que e_{ab} também será, para qualquer a, b.

Sejam $b_1, b_2, ..., b_n$ estados diferentes. Estamos interessados em saber qual estado será o primeiro a ser tocado se começarmos em X_0 e quando isso vai acontecer. Por simplicidade, $j = b_j$ e $\tau_j = \tau_{0j}$.

Sejam $\tau = \min\{\tau_1, ..., \tau_n\}$ e $p_{0j} = \mathbb{P}(\tau = \tau_j)$, em que p_{0j} denota a probabilidade do estado j ser o primeiro a ser tocado. Note que,

$$e_{0i} = \mathbb{E}[\tau] + \sum_{j=1}^{n} p_{0j} e_{ji} \tag{4.4}$$

De fato, usando as propriedades de esperança condicional temos

$$\begin{aligned}
e_{0i} &= \mathbb{E}[\tau_{0i}] \\
&= \mathbb{E}[\tau + N_{0i} - \tau] \\
&= \mathbb{E}[\tau] + \mathbb{E}[\tau_{0i} - \tau] \\
&= \mathbb{E}[\tau] + \mathbb{E}[\mathbb{E}(\tau_{0i} - \tau | \tau = \tau_j)] \\
&= \mathbb{E}[\tau] + \sum_{j=1}^{n} \mathbb{E}[\tau_i - \tau | \tau = \tau_j] \mathbb{P}(\tau = \tau_j) \\
&= \mathbb{E}[\tau] + \sum_{j=1}^{n} \mathbb{E}[\tau_i - \tau_j | \tau = \tau_j] \mathbb{P}(\tau = \tau_j) \\
&= \mathbb{E}[\tau] + \sum_{j=1}^{n} e_{ji} p_{0j}
\end{aligned} \tag{4.5}$$

em que $e_{ii} = 0$, $\sum_{j=1}^{n} p_{0j} = 1$ e $p_{01} + ... + p_{0n} = 1$.

Colocando em forma matricial, temos um sistema linear de $n+1$ incógnitas, tal que

$$\left[\begin{array}{c|ccc} 0 & 1 & \ldots & 1 \\ \hline 1 & & & \\ \vdots & & e_{ji} & \\ 1 & & & \end{array}\right] \begin{bmatrix} \mathbb{E}[\tau] \\ p_{01} \\ \vdots \\ p_{0n} \end{bmatrix} = \begin{bmatrix} 1 \\ e_{01} \\ \vdots \\ e_{0n} \end{bmatrix} \tag{4.6}$$

Como sabemos, matematicamente para que o sistema linear tenha solução, a matriz de coeficientes precisa ser não singular. O teorema abaixo garante que a o sistema tem uma solução no caso geral.

Teorema 6. A matriz de coeficientes de (4.6) é não singular.

Observação 1. O Teorema 6 garante-nos que o sistema linear do Teorema 7, quando olhamos a cadeia imersa de ocorrência de palavras, é possível e determinado, ou seja, possui uma única solução.

Demonstração. Seja M a matriz de coeficiente e

$$A = \left[\begin{array}{c|ccc} 0 & 1 & \dots & 1 \\ \hline 1 & & & \\ \vdots & & e_{ji} & \\ 1 & & & \end{array}\right],$$

Aplicando Cramer em (4.6) para isolar a incógnita $\mathbb{E}[\tau]$, temos

$$\mathbb{E}[\tau] = \frac{detA}{detM} \Rightarrow \mathbb{E}[\tau]detM = detA \tag{4.7}$$

Usaremos (4.7) e o fato que $\mathbb{E}[\tau] > 0$ para mostrar por indução que

$$(-1)^n detM > 0 \tag{4.8}$$

Para $n = 1$,

$$\begin{bmatrix} 0 & 1 \\ 1 & e_{11} \end{bmatrix} \Rightarrow detM = -1$$

Agora, assumimos que (4.8) mantém para todos os conjuntos de n estados diferentes e denotamos por M_{n+1} a matriz correspondente do estado $n+1$. Dividindo $detM_{n+1}$ em determinantes menores (Regra de Cramer), temos

$$detM_{n+1} = \sum_{k=1}^{n+1} (-1)^k detB_k \tag{4.9}$$

em que a matriz B_k é obtida excluindo a coluna 0 e a linha k em M_{n+1}.

Seja A_k a matriz de resultados quando a $k-$ ésima coluna da matriz B_k é induzida para frente.

Logo,

$$\begin{aligned} detM_{n+1} &= \sum_{k=1}^{n+1} (-1)^k detB_k \\ &= \sum_{k=1}^{n+1} (-1)^k (-1)^{k-1} detA_k \\ &= \sum_{k=1}^{n+1} (-1)^{2k} (-1)^{-1} detA_k \\ &= -\sum_{k=1}^{n+1} detA_k \end{aligned} \tag{4.10}$$

Porém, cada uma das matrizes A_k é do mesmo tipo que a matriz A. Assim, de (4.7) e da indução, temos que sinal $(detA_k)$ = sinal $(detM)$. Mas de (4.10) temos que sinal $(detM_{n+1})$ = - sinal $(detM)$. □

Com isso garantimos que podemos calcular a probabilidade para cada estado $b_1, b_2, ..., b_n$ ser o primeiro a ser tocado e a esperança do tempo de espera até a primeira batida.

4.3 O tempo de espera médio para várias palavras e as probabilidades de parada

Nesta seção, faremos uma generalização da **Proposição 4.1.3** para o caso de k palavras aleatórias concorrendo pela primeira ocorrência em $\{X_n\}_{n\geq 1}$.

Considere uma coleção finita $\mathscr{C} = \{A_1, ..., A_k\}$ de palavras em nosso alfabeto. Analogamente aos capítulos anteriores, supomos que nenhuma palavra seja sub-palavra da outra.

Sabemos que o tempo de parada τ é definido como

$$\tau = \min\{\tau_{A_1}, ..., \tau_{A_k}\} \tag{4.11}$$

e que $\mathbb{E}(\tau) < \infty$.

Agora supomos que temos um time de apostadores para cada k palavra, em que cada jogador do time associado a palavra A_i, $i = 1, ..., k$, aposta y_i reais no momento que entra na aposta.

Manteremos as definições da seção anterior:

S_n será a quantidade de dinheiro acumulada pelo cassino de todos os jogadores até o tempo n (incluindo n) e de todos todos os times de apostas.

Assim, no tempo de parada τ, temos que a quantidade de dinheiro acumulada do cassino será a quantia que o cassino arrecada junto aos apostadores até o tempo τ menos a quantia que ele paga a cada time de apostadores no momento τ. Isto é,

$$S_\tau = (y_1 + ... + y_k)\tau - \sum_{i=1}^{k} (y_1(A_i * A_1) + ... + y_k(A_i * A_k))\mathbb{I}_{\{\tau = \tau_{Al_i}\}}$$

Defina $\mu_i = \mathbb{P}(\tau = \tau_{A_i})$, $i \in \{1, ..., k\}$, com $\mu = (\mu_1, ..., \mu_k)^T$ e $Y = (y_1, ..., y_k)$. Defina a matriz de lucros de cada um dos times como

$$P = \begin{bmatrix} A_1 * A_1 & ... & A_1 * A_k \\ A_2 * A_1 & ... & A_2 * A_k \\ ... & ... & ... \\ A_k * A_1 & ... & A_k * A_k \end{bmatrix}, \tag{4.12}$$

Nas mesmas hipóteses da seção anterior, temos

$$0 = \mathbb{E}[S_\tau] = (y_1 + ... + y_k)\mathbb{E}[\tau] - \mu^T PY. \tag{4.13}$$

Proposição 3. Se existe uma solução $Y = (y_1, ..., y_k)^T$ para o sistema linear $PY = \mathbf{1}$, em que $\mathbf{1} = (1, ..., 1)^T$. Então,

$$\mathbb{E}(\tau) = \frac{1}{y_1 + ... + y_k} \tag{4.14}$$

Sabendo $\mathbb{E}(\tau)$, podemos escolher as apostas de modo a obter as probabilidades de parada $\mu = (\mu_1, \ldots, \mu_k)^\top$, onde $\mu_i = \mathbb{P}(\tau = \tau_{A_i})$, $i = 1, \ldots, k$.

Proposição 4. O vetor de probabilidades $\mu = (\mu_1, \ldots, \mu_k)^\top$ satisfaz a equação $P^\top \mu = \mathbb{E}(\tau)\mathbf{1}$, onde $\mathbf{1} = (1, \ldots, 1)^\top$.

Demonstração. Para $j = 1, \ldots, k$, considere o seguinte vetor de apostas

$$Y^j = (0, \ldots, \underbrace{1}_{j-\text{ésima posição}}, 0, \ldots, 0)^\top.$$

Daí, o produto PY^j nos dá a j-ésima coluna de P, isto é,

$$0 = \mathbb{E}(S_\tau) = \mathbb{E}(\tau) - (A_1 * A_j, \ldots, A_k * A_j)^\top \mu, \tag{4.15}$$

para todo $j \in \{1, \ldots, k\}$. Logo, o resultado segue. □

Exemplo 11. Considere o alfabeto $\mathscr{A} = \{H, T\}$ e as probabilidades $\mathbb{P}(X_n = H) = \mathbb{P}(X_n = T) = 1/2$ para $n \geq 1$. Defina as palavras $A_1 = HHHHH$ e $A_2 = HTHTH$. Calculemos $\mathbb{E}(\tau_{A_1})$, $\mathbb{E}(\tau_{A_2})$, $\mathbb{E}(\tau)$, onde $\tau = \min\{\tau_{A_1}, \tau_{A_2}\}$, e as probabilidades de parada μ_1, μ_2.

Para A_1, primeiramente note que $\delta_{ij} = \frac{1}{1/2} = 2$ para todo $i, j = 1, \ldots, 5$. Usando (4.1) e a Proposição 2, temos

$$\mathbb{E}(\tau_A) = 2^5 + 2^4 + 2^3 + 2^2 + 2^1 = 62.$$

Para A_2, temos que

$$\delta_{ij} = \begin{cases} 2, & \text{se } (i,j) \in \{1,3,5\}^2 \cup \{2,4\}^2 \\ 0, & \text{se } (i,j) \notin \{1,3,5\}^2 \cup \{2,4\}^2 \end{cases}$$

Daí,

$$\mathbb{E}(\tau_{A_2}) = 2^5 + 0 + 2^3 + 0 + 2^1 = 42.$$

É interessante notar que, embora A_1 e A_2 tenham mesmo tamanho e mesma probabilidade de ocorrência em 5 lançamentos consecutivos, essas palavras apresentam tempos de espera médio bem distintos. Basicamente, isso se deve ao fato de que a sobreposição de A_1 dificulta sua ocorrência.

Encontremos agora matriz de lucros P. Neste caso, temos que

$$P = \begin{bmatrix} 62 & A_1 * A_2 \\ A_2 * A_1 & 42 \end{bmatrix}.$$

Calculemos $A_1 * A_2$.

Inicialmente, usaremos a abordagem de Pozdnyakov e Kulldorff (2006). Como $\delta_{i2} = 0$ para $i \in \{1, \ldots, 5\}$ e $\delta_{51} = 2$, então $A_1 * A_2 = 0+0+0+0+2 = 2$. Por outro lado, temos que $A_2 * A_1 = 2$, já que $\delta_{2i} = \delta_{4i} = 0$ para $i \in \{1, \ldots, 5\}$ e $\delta_{51} = 2$.

Para encontrar $\mathbb{E}(\tau)$, temos de resolver o sistema linear

$$\begin{bmatrix} 62 & 2 \\ 2 & 42 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

cuja solução única é dada por $y_1^* = 10/650$ e $y_2^* = 15/650$. Usando a Proposição 3, temos que $\mathbb{E}(\tau) = 26$.

Para achar $\mu = (\mu_1, \mu_2)^\top$, de acordo com a Proposição 4, temos de resolver o sistema

$$\begin{bmatrix} 62 & 2 \\ 2 & 42 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} 26 \\ 26 \end{bmatrix},$$

cuja solução única é dada por $\mu_1 = 26/65 = 0,4$ e $\mu_2 = 39/65 = 0,6$.

4.4 O teorema de Li (1980) para várias palavras com alfabeto enumerável

Nesta seção, estudamos o resultado principal de Li (1980), o Teorema 7 logo abaixo enunciado, que versa sobre o tempo de espera até a ocorrência de uma palavra entre várias palavras. A abordagem de Li (1980) considera o alfabeto enumerável e lança mão da ideia de imersão da cadeia de Markov.

Teorema 7. Sejam $X, X_1, X_2, \ldots$ variáveis aleatórias iid e $A_1, \ldots, A_k$ sequências finitas de possíveis valores de X, não contendo uma a outra. Seja A outra sequência tal que não seja subsequência de qualquer outra A_i, $i = 1, \ldots, k$. Dada a sequência inicial A, seja μ_i a probabilidade que A_i preceda as demais sequências em $X_1, X_2, \ldots$. Então, para cada i, temos

$$\sum_{j=1}^{k} p_j A_j * A_i = \mathbb{E}[\tau] + A * A_i, \tag{4.16}$$

em que τ é o tempo de parada dado por (4.11) e $A_j * A_i$ é definido por (4.1). Em particular, quando $A = \varnothing$, temos

$$\sum_{j=1}^{k} p_j A_j * A_l = \mathbb{E}[\tau] \tag{4.17}$$

Antes de demonstrarmos o Teorema 7, necessitamos de um resultado auxiliar, cuja prova pode ser encontrada em Li (1980).

Lema 1. Dada uma sequência inicial A defina o tempo de espera até que B ocorra por τ_{AB}. Então

$$\mathbb{E}[\tau_{AB}] = B * B - A * B, \tag{4.18}$$

em que B não é uma subsequência de A.

A demonstração do Lema 1 pode ser encontrada no artigo de Li (1980).

Prova do Teorema 7. Sejam $N_l = N_{AA_l}$ N o mínimo entre $N_1, \ldots, N_n$ e

Temos

$$\mathbb{E}[N_l] = \mathbb{E}[N_l] + \mathbb{E}[N_l - N] = \mathbb{E}[N] + \sum_{j=1}^{n} p_j \mathbb{E}[N_j - N | N = N_j] \tag{4.19}$$

Pelo Lema 1 sabemos também que,

$$\mathbb{E}[N_l] = A_l * A_l - A * A_l \tag{4.20}$$

e

$$\mathbb{E}[N_j - N | N = N_j] = A_l * A_l - A_j * A_l \tag{4.21}$$

Assim,

$$\begin{aligned} A_l * A_l - A * A_l &= \mathbb{E}[N] + \sum_{j=1}^{n} p_j \mathbb{E}[N_j - N | N = N_j] \\ &= \mathbb{E}[N] + \sum_{j=1}^{n} p_j (A_l * A_l - A_j * A_l) \end{aligned}$$

Portanto,

$$\sum_{j=1}^{n} p_j A_j * A_j = \mathbb{E}[N] + A * A_j \tag{4.22}$$

□

Exemplo 12. Voltemos ao exemplo *5.2.3*. Temos um alfabeto $\mathscr{A} = \{H, T\}$ e probabilidades $\mathbb{P}(X_n = H) = \mathbb{P}(X_n = T) = 1/2$ para $n \geq 1$. Definimos as palavras $A_1 = HHHHH$ e $A_2 = HTHTH$ e as probabilidades de parada μ_1, μ_2., em que

$$\mathbb{E}(\tau_{A_1}) = 62 \text{ e } \mathbb{E}(\tau_{A_2}) = 42.$$

Usaremos a abordagem de (LI, 1980) para calcular as probabilidades de parada. Sabemos que $A_1 * A_2 = A_2 * A_1 = 2$ e $\mu_1 + \mu_2 = 1$. Assim,

$$\begin{cases} \mu_1(A_1 * A_1) + \mu_2(A_2 * A_1) = \mathbb{E}(N) \\ \mu_1(A_1 * A_2) + \mu_2(A_2 * A_2) = \mathbb{E}(N), \end{cases}$$

ou seja,

$$\begin{cases} 62\mu_1 + 2(1-\mu_1) = \mathbb{E}(N) \\ 2\mu_1 + 42(1-\mu_1) = \mathbb{E}(N). \end{cases}$$

Logo, $\mu_1 = 0,4$ e $\mu_2 = 0,6$.

CAPÍTULO

5

CONSIDERAÇÕES FINAIS

Neste trabalho, mostramos a abordagem de martingais para a ocorrência de palavras em ensaios $\{X_n\}_{n\geq 0}$ i.i.d. Começamos com uma revisão de probabilidade e processos estocásticos nos Capítulos 2 e 3 que são fundamentais para o desenvolvimento deste projeto. No Capítulo 4, introduzimos a abordagem de Li (1980) e de Pozdnyakov e Kulldorff (2006), uma simplificação da primeira, para o cálculo do tempo de parada e o tempo de espera médio da ocorrência da palavra. Nas Seções 4.1 e 4.3, apresentamos a abordagem de Pozdnyakov e Kulldorff (2006) para o problema de ocorrência de palavras, inicialmente para uma palavra apenas e depois para um conjunto de palavras. Calculamos o tempo médio de espera para uma ou mais palavras e a probabilidade de uma determinada palavra ser a primeira a aparecer num alafabeto finito. Na Seção 4.4, mostramos o tratamento de Li (1980), que nos permite calcular o tempo de espera até que ocorra uma palavra dentre várias palavras em um alfabeto infinito enumerável; tal abordagem lança mão do uso de imersão da cadeia de Markov. Na Seção 4.2, estudamos o tempo de primeira passagem em uma cadeia de Markov abordado em Gerber e Li (1981), em que usamos a definição de esperança para encontrar o tempo de espera médio de um estado ser alcançado. Nesta seção, temos um resultado fundamental para o trabalho: provamos que a matriz de coeficientes do sistema (4.6) é não singular, o que nos garante que o sistema tem uma única solução.

REFERÊNCIAS

GERBER, H. U.; LI, S.-Y. R. The occurrence of sequence patterns in repeated experiments and hitting times in a markov chain. **Stochastic Processes and their Applications**, Elsevier, v. 11, n. 1, p. 101–108, 1981. Citado 6 vezes nas páginas 9, 11, 15, 16, 32 e 41.

GRIMMETT, G.; STIRZAKER, D. **Probability and random processes**. [S.l.]: Oxford university press, 2001. Citado 2 vezes nas páginas 17 e 21.

GUT, A. **An Intermediate Course in Probability**. Springer New York, 2009. (Springer Texts in Statistics). ISBN 9781441901620. Disponível em: <https://books.google.com.au/books?id=ufxMwdtrmOAC>. Citado 2 vezes nas páginas 17 e 19.

LI, S.-Y. R. A martingale approach to the study of occurrence of sequence patterns in repeated experiments. **the Annals of Probability**, JSTOR, p. 1171–1176, 1980. Citado 10 vezes nas páginas 9, 11, 13, 15, 16, 29, 37, 38, 39 e 41.

NORRIS, J. R. **Markov chains**. [S.l.]: Cambridge university press, 1998. Citado 3 vezes nas páginas 21, 25 e 27.

POZDNYAKOV, V.; KULLDORFF, M. Waiting times for patterns and a method of gambling teams. **The American Mathematical Monthly**, JSTOR, p. 134–143, 2006. Citado 4 vezes nas páginas 15, 29, 37 e 41.

ROSS, S. **A First Course in Probability**. Prentice Hall, 2002. ISBN 9780130338518. Disponível em: <https://books.google.com.br/books?id=5ptVAAAAMAAJ>. Citado 3 vezes nas páginas 17, 18 e 19.

ROSS, S. M. *et al.* **Stochastic processes**. [S.l.]: John Wiley & Sons New York, 1996. v. 2. Citado 2 vezes nas páginas 21 e 28.

WILLIAMS, D. **Probability with Martingales**. Cambridge University Press, 1991. (Cambridge mathematical textbooks). ISBN 9780521406055. Disponível em: <https://books.google.co.jp/books?id=e9saZ0YSi-AC>. Citado na página 28.