UNIVERSIDADE ESTADUAL DA PARAÍBA
CENTRO DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA

Benedito Vicente dos Santos

Tópicos Avançados de Processos Markovianos e Aplicações

Campina Grande - PB

Dezembro de 2017

Benedito Vicente dos Santos

Tópicos Avançados de Processos Markovianos e Aplicações

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Orientador: Divanilda Maia Esteves

Campina Grande - PB

Dezembro de 2017

S237t Santos, Benedito Vicente dos.
Tópicos avançados de processos Markovianos e aplicações [manuscrito] : / Benedito Vicente dos Santos. - 2017.
54 p.

Digitado.
Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2017.
"Orientação : Profa. Dra. Divanilda Maia Esteves, Coordenação do Curso de Estatística - CCT."

1. Processos estocásticos. 2. Processos Markovianos. 3. Inferência estatística.

21. ed. CDD 519.54

Benedito Vicente dos Santos

Tópicos Avançados de Processos Markovianos e Aplicações

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Trabalho aprovado em 01 de dezembro de 2017.

BANCA EXAMINADORA

DMEsteves

Dr^{a}. Divanilda Maia Esteves
Universidade Estadual da Paraíba

Ricardo Alves de Olinda

Dr. Ricardo Alves de Olinda
Universidade Estadual da Paraíba

Tiago Almeida de Oliveira

Dr. Tiago Almeida de Oliveira
Universidade Estadual da Paraíba

Dedico este trabalho a todos que tornaram possível a realização deste trabalho e que fizeram parte direta ou indiretamente deste trabalho.

Agradecimentos

Aos meus pais que sempre me apoiaram em todos os momentos.

Aos meus irmãos, Iris, Celson, Maísa, Cristina, Rafaela e Karl Marx pelo apoio e compreensão.

A professora Divanilda Maia Esteves, por me orientar na construção deste trabalho, pela imensa dedicação, compreensão, empenho e apoio em todos os momentos da realização deste trabalho.

Aos amigos que fiz neste curso.

*"Tudo o que não é perfeito pode ser melhorado, mas é preciso fazer alguma coisa.
Tudo o que está ruim, pode piorar ainda mais, basta não fazer nada."*
(S.R.Marks)

Resumo

O objetivo deste trabalho foi estudar alguns aspectos dos Processos Markovianos de forma mais aprofundada do que aquela que é vista em geral em um curso de graduação. Um processo estocástico $\{X(t) : t \in T\}$ é uma coleção de variáveis aleatórias, isto é, para cada $t \in T$, $X(t)$ é uma variável aleatória. O conjunto T é chamado conjunto de índices e o conjunto de todos os valores que as variáveis $X(t)$ pode assumir é chamado espaço de estados do processo estocástico. Frequentemente, o índice t é interpretado como tempo t e por isso nos referimos a $X(t)$ como o estado do processo no tempo t. Daí, uma maneira alternativa de se definir um processo estocástico é: um processo estocástico é uma família de variáveis aleatórias que descreve a evolução de algum processo físico através do tempo. Um aspecto importante nas situações práticas é o conhecimento sobre a estrutura de dependência que há entre as variáveis. O caso mais considerado é aquele em que as variáveis são independentes. Isso facilita bastante operações envolvendo a verossimilhança, no entanto, muitas vezes tal estrutura não é adequada. Aqui, o enfoque será naqueles casos em que se observa uma estrutura de dependência chamada de dependência de Markov. Um processo estocástico é dito ser markoviano se, uma vez que se conhece o estado atual do processo, os estados passados não influenciam o futuro. Essa é a definição mais comum de se encontrar na literatura e trata apenas daqueles casos em que a informação mais recente que se tem sobre o processo concentra "toda" informação que o passado tem para se conhecer o futuro. Mas essa definição pode ser mais geral, considerando que não apenas a informação mais recente, mas as k informações mais recentes influenciam no futuro do processo. Esses modelos tem notável importância e amplo uso teórico e prático. No final aplicou-se tal teoria em um conjunto de dados relacionado às altas e baixas da cotação do dólar no Brasil no começo deste ano.

Palavras-chaves: Processos Estocásticos; Processos Markovianos; Inferência Estatística.

Abstract

The objective of this work was to study some aspects of the Markovian more in-depth than that which is generally seen in an undergraduate course. A stochastic process $\{X(t) : t \in T\}$ is a collection of random variables, it implies, for each $t \in T$, $X(t)$ is a random variable. The T set is called an indexes set and the set of all values which the variables can assume is called states space from the stochastic process. Often, the set t is interpreted as t time for this reason we refer to $X(t)$ as one state in the t time process. From that, an alternative manner to define a stochastic process is: One stochastic process is a family of random variables which describes the evolvement of a physic process through the time. An important aspect in the practical situations is the knowledge about the dependence of structure which exists between the variables. The most considered case is the one in which the variables are independent. This greatly facilitates the operations involving the verisimilitude, although many times this structure is not adequate. Here, the focus will be on the cases in which it is possible to observe a dependence structure called Markov dependence. A stochastic process it is recognized as Markovian once if it is knew the actual state of the process, the previous states do not influence the future. This is the most common definition to find in literature and treats only the cases in which the most recent information obtained about the process. It groups "all" new information which the past has, to know about the future. But this definition can be more general, considering not only the most recent information, but the most recent information influence in the future of the process. These models have notorious importance, a wide theoretical and practical usage. At the ending this theory was applied in a data set making relation the highs and lows of the Dollar cotation at the beginning of this year in Brazil.

Key-words: Stochastic Processes; Markovian Processes; Statistic Inference.

Sumário

1 INTRODUÇÃO **10**

2 FUNDAMENTAÇÃO TEÓRICA **11**

2.1 Processos estocásticos **11**

2.2 Cadeias de Markov **12**

2.2.1 Classificação dos estados 15

2.2.2 Representação canônica da matriz de probabilidade de transição 17

2.2.3 Cadeias de Markov finitas com estados transitórios 19

2.2.4 A matriz fundamental 20

2.3 Cadeias de Markov irredutíveis com estados ergódicos **25**

2.3.1 Comportamento transiente 26

2.3.2 Cadeias de Markov de dois estados 26

2.3.3 O método de autovalor para cadeias de Markov finitas irredutíveis 29

2.3.4 Medida invariante 31

2.3.5 Tempo de permanência em um estado 32

2.4 Processos de ramificação e outros tópicos especiais **33**

2.4.1 Processos de ramificação 33

2.4.2 Cadeias de Markov de ordem maior que 1 37

2.4.3 Cadeias de Markov inversa 38

2.5 Inferência estatística para cadeias de Markov **39**

2.5.1 Estimativa dos elementos de uma matriz de probabilidade de transição . . . 39

2.6 Teste de hipótese para cadeias de Markov **42**

2.6.1 Teste de hipótese para uma matriz de probabilidade de transição 42

2.6.2 Teste para estacionariedade da matriz de transição de probabilidade 42

2.6.3 Teste para independência versus a dependência de primeira ordem 44

2.6.4 Teste para a ordem da cadeia de Markov 44

2.6.5 Inferência estatística para processos de ramificação 45

2.7 Processos de Markov simples **46**

2.7.1 Processos de Markov: Propriedades gerais 46

3 APLICAÇÃO **50**

4 CONCLUSÃO **53**

REFERÊNCIAS **54**

1 Introdução

De modo geral, pode-se dizer que um processo estocástico é qualquer processo que evolui de maneira aleatória. Assim, é fácil encontrar aplicações dos processos estocásticos em diversas áreas de conhecimento, como por exemplo: a evolução do índice diário da bolsa de valores, o número de ligações recebidas por dia em uma central de teleatendimento ou a quantidade de chuva registrada por mês em determinada localidade. Além disso, esses processos são fundamentais para o estudo de Séries Temporais, que é uma área de notável importância e de amplo uso teórico e prático.

Mais formalmente, um processo estocástico $\{X(t) : t \in T\}$ é uma coleção de variáveis aleatórias, isto é, para cada $t \in T$, $X(t)$ é uma variável aleatória. Para que se possa modelar conjuntamente essas variáveis, é necessário ter uma ideia sobre a estrutura de dependência entre as variáveis. Aqui, o enfoque será naqueles casos em que se observa uma estrutura de dependência chamada de **dependência de Markov**. Um processo estocástico $\{X_n, n \geq 0\}$ é dito ser markoviano se para $t_0 < t_1 < \cdots < t_{n+1}$

$$P(X_{t_{n+1}} = i_{t_{n+1}} | X_{t_0} = i_{t_0}, X_{t_1} = i_{t_1}, ..., X_{t_n} = i_{tn}) = P(X_{t_{n+1}} = i_{t_{n+1}} | X_{t_n} = i_{t_n}),$$

para qualquer escolha $i_{t_0}, i_{t_1}, ..., i_{t_{n+1}}$ em S e qualquer n. Isto quer dizer que, uma vez que se conhece o estado atual do processo, os estados passados não influenciam o futuro. Essa é a definição mais comum de se encontrar na literatura e trata apenas daqueles casos em que a informação mais recente que se tem sobre o processo concentra "toda" informação que o passado tem para se conhecer o futuro. Mas essa definição pode ser mais geral, considerando que não apenas a informação mais recente, mas as k informações mais recentes influenciam no futuro do processo. Generalizando, um processo estocástico $\{X_n, n \geq 0\}$ é dito ser markoviano de ordem $k \in \mathbb{N}$ se para $t_0 < t_1 < \cdots < t_{n+1}$

$$P(X_{t_n} = i_{t_n} | X_{t_0} = i_{t_0}, X_{t_1} = i_{t_1}, ..., X_{t_{n-1}} = i_{t_{n-1}}) = P(X_{t_n} = i_{t_n} | X_{t_{n-k}} = i_{t_{n-k}}, ..., X_{t_{n-1}} = i_{t_{n-1}}),$$

para qualquer escolha $i_{t_0}, i_{t_1}, ..., i_{t_{n+1}}$ em S e qualquer n. Isto quer dizer que o estado atual do processo é influenciado pelas k observações mais recentes do processo.

Assim, a proposta deste trabalho foi estudar os processos markovianos, bem como o processo de estimação dos parâmetros e testes associados a avaliação da adequabilidade do modelo para um determinado conjunto de dados. Por fim, foi possível propor uma aplicação a dados reais para reforçar os conceitos vistos.

Neste trabalho, foi feito um estudo da teoria de processos Markovianos, seus principais aspectos estão na fundamentação teórica. Depois aplicou-se tal teoria em um conjunto de dados relacionado ás altas e baixas da cotação do dólar no Brasil no começo deste ano. Com o objetivo de estimar os parâmetros do modelo e avaliar se o modelo de cadeias de Markov é um modelo adequado para este conjunto de dados.

2 Fundamentação Teórica

Este trabalho tem caráter predominantemente teórico. Grande parte da fundamentação teórica foi baseada no livro Bhat e Miller (2002). Foram usados como referências complementares Hoel, Port e Stone (1972), Karlin e Taylor (1975), Ross (1996), Brémaud (1999), Ross (2007), Stewart (2009). A seguir, serão apresentados alguns dos principais tópicos estudado neste trabalho.

2.1 Processos estocásticos

Um processo estocástico $\{X(t) : t \in T\}$ é uma coleção de variáveis aleatórias, isto é, para cada $t \in T, X(t)$ é uma variável aleatória. O conjunto T é chamado de conjunto de índices. O conjunto de todos os valores que as variáveis $X(t)$ pode assumir será chamado espaço de estados do processo estocástico e denotaremos por S. Frequentemente, o índice t é interpretado como tempo t e por isso nos referimos a $X(t)$ como o estado do processo no tempo t. Daí, uma maneira alternativa de se definir um processo estocástico é: um processo estocástico é uma família de variáveis aleatórias que descreve a evolução de algum processo físico através do tempo.

Seja $\{X(t) : t \in T\}$ um processo estocástico com espaço de estados S e conjunto de índices T. Se S for enumerável, o processo é dito discreto ou a valores inteiros (do inglês *integer-valued*). Se S é um intervalo da reta (ou o próprio $\mathbb{R}$) então tem-se um processo a valores reais (do inglês *real-valued*). Se o conjunto de índices T for enumerável, então diz-se que o processo é a tempo discreto e, em geral, considera-se $T = \{0, 1, 2, ...\}$ e usaremos $\{X(n), n \geq 0\}$ no lugar de $\{X(t) : t \in T\}$. Se $T = [0, \infty)$, $X(t)$ é dito um processo a tempo contínuo.

Em situações práticas, há interesse em se calcular $P(X_0 = x_0, X_1 = x_1, ..., X_n = x_n)$, que é a função de verossimilhança da amostra $x_0, x_1, ..., x_n$, e que é usada, por exemplo, para estimarmos os parâmetros do modelo. Para calcular tal quantidade, é importante que saibamos qual é a estrutura de dependência que há entre as variáveis. O caso mais considerado na teoria estatística é aquele em que as variáveis são independentes. No entanto, muitas vezes tal estrutura não é adequada. Aqui, o enfoque será naqueles casos em que se observa uma estrutura de dependência chamada de dependência de Markov.

Definição 2.1 *Um processo estocástico $\{X_n, n \geq 0\}$ é dito ser Markoviano (de ordem 1) se para $t_0 < t_1 < ... < t_{n+1}$*

$$P(X_{t_{n+1}} = i_{t_{n+1}} | X_{t_0} = i_{t_0}, X_{t_1} = i_{t_1}, ..., X_{t_n} = i_{tn}) = P(X_{t_{n+1}} = i_{t_{n+1}} | X_{t_n} = i_{t_n}),$$

para qualquer escolha $i_{t_0}, i_{t_1}, ..., i_{t_{n+1}}$ em S e qualquer n. Isto quer dizer que, uma vez que conhecemos o estado atual do processo, os estados passados não influenciam o futuro.

Há diversas generalizações dessa dependência. Por exemplo, um processo estocástico $\{X_n, n \geq 0\}$ é dito ser markoviano de ordem k se para $t_0 < t_1 ... < t_{n+1}$

$$P(X_{t_n} = i_{t_n} | X_{t_0} = i_{t_0}, X_{t_1} = i_{t_1}, ..., X_{t_{n-1}} = i_{t_{n-1}}) = P(X_{t_n} = i_{t_n} | X_{t_{n-k}} = i_{t_{n-k}}, ..., X_{t_{n-1}} = i_{t_{n-1}}),$$

para qualquer escolha $i_{t_0}, i_{t_1}, ..., i_{t_{n+1}}$ em S e qualquer n. Isto quer dizer que o estado atual da cadeia é influenciado pelas k observações mais recentes do processo. Se um processo estocástico é markoviano e seu espaço de estados (S) é discreto, então o processo é dito ser uma cadeia de Markov.

2.2 Cadeias de Markov

Uma cadeia de Markov é um processo estocástico com espaço de estados discreto. Se o conjunto de índices for um conjunto enumerável, então temos uma cadeia de Markov a tempo discreto, caso contrário, temos uma cadeia de Markov a tempo contínuo. Frequentemente, nos referimos às cadeias de Markov a tempo discreto de ordem 1 simplesmente por cadeia de Markov.

Definição 2.2 *Um processo $\{X_n : n \geq 0\}$ assumindo valores em um conjunto S é uma cadeia de Markov (de ordem 1) se dado estado presente do processo, o futuro não é influenciado pelo passado, ou seja,*

$$P(X_{n+1} = i_{n+1} | X_0 = i_0, X_1 = i_1, ..., X_n = i_n) = P(X_{n+1} = i_{n+1} | X_n = i_n).$$

Geralmente consideremos apenas cadeias de Markov que são homogêneas no tempo. Dizemos ainda que uma cadeia de Markov é homogênea no tempo se

$$P(X_{n+1} = j | X_n = i) = P(X_1 = j | X_0 = i), \tag{2.1}$$

para qualquer $n \geq 0$.

A probabilidade dada pela Equação (2.1) pode ser chamada de probabilidade de transição a um passo da cadeia e para facilitar a notação usaremos

$$P(X_{n+1} = j | X_n = i) = P(X_1 = j | X_0 = i) = P(i, j),$$

ao que diremos a probabilidade de ir do estado i ao estado j em um passo. Nota-se que o termo passo refere-se ao intervalo de tempo entre as observações. Em outras palavras, se a cadeia de Markov está no estado i no tempo n, ela tem probabilidade $P(i, j)$ de estar em j no próximo passo, e isso justifica o fato de as quantidades $P(i, j)$ serem chamadas probabilidades de transição a um passo da cadeia.

Quando temos uma cadeia que é homogênea no tempo, podemos representar suas probabilidades de transição matricialmente, na forma

$$\mathbf{P} = ||\mathbf{P}_{ij}|| = \begin{bmatrix} P_{00} & P_{01} & P_{02} & \dots \\ P_{10} & P_{11} & P_{12} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Neste caso, $\mathbf{P} = ||\mathbf{P}_{ij}||$ é a matriz de transição a um passo da cadeia, ou simplesmente matriz de transição da cadeia. Observa-se que todas as entradas da matriz são não negativas e que a soma dos elementos de cada linha deve ser igual a 1. A notação $||\mathbf{P}_{ij}||$ significa que estamos usando todos os elementos da matriz $\mathbf{P}_{ij}$.

Para estudar a evolução da cadeia é necessário conhecer a sua função de transição e também sabermos as chances associadas ao início da cadeia, ou seja, precisamos modelar o início, o X_0 da cadeia.

Definição 2.3 *A função $\pi_0(i), i \in S$, definida por*

$$\pi_0(i) = P(X_0 = i),$$

é chamada distribuição inicial da cadeia e é tal que

$$\pi_0(i) \geq 0 \qquad i \in S,$$

$$\sum_{i \in S} \pi_0(i) = 1.$$

A distribuição inicial de uma cadeia é simplesmente a função de probabilidade do seu estado inicial X_0. Frequentemente, apresentamos a distribuição inicial da cadeia na forma de um vetor:

$$\pi_0 = [\pi_0(0), \pi_0(1), \pi_0(2), ...].$$

Teorema 2.1 *A distribuição conjunta de $X_0, X_1, ..., X_n$ pode ser expressa em termos da função de transição e da distribuição inicial da seguinte maneira:*

$$P(X_0 = x_0, X_1 = x_1, ..., X_n = x_n) = \pi_0(x_0) \times P(x_0, x_1) \times P(x_1, x_2) \times ... \times P(x_{n-1}, x_n).$$

A ideia da demonstração é usar o fato de que

$$\begin{aligned} P(X_0 = x_0, X_1 = x_1, ..., X_n = x_n) &= P(X_n = x_n | X_0 = x_0, X_1 = x_1, ..., X_{n-1} = x_{n-1}) \\ &\times P(X_{n-1} = x_{n-1} | X_0 = x_0, X_1 = x_1, ..., X_{n-2} = x_{n-2}) \\ &\times ... \times P(X_1 = x_1 | X_0 = x_0) P(X_0 = x_0). \end{aligned}$$

Definição 2.4 *A função de transição a m passos da cadeia é definida como*

$$P^m(i,j) = P(X_m = j|X_0 = i), \qquad i,j \in S,$$

que é a probabilidade da cadeia ir de um estado i a um estado j em m passos.

Assim como fizemos com a função de transição a um passo, a função de transição a m passos será representada na forma de uma matriz como

$$\mathbf{P}^{(m)} = ||\mathbf{P}_{ij}^{(m)}|| = \begin{bmatrix} P_{00}^{(m)} & P_{01}^{(m)} & P_{02}^{(m)} & \dots \\ P_{10}^{(m)} & P_{11}^{(m)} & P_{12}^{(m)} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Uma forma alternativa de se calcular as probabilidades de transição a m passos é usar as equações de Chapman-Kolmogorov que são

$$P^{(m+n)}(i,j) = \sum_{z \in S} P^{(m)}(i,z)P^{(n)}(z,j), \qquad \forall\, m,n \geq 0 \quad e \quad \forall\, i,j \in S. \tag{2.2}$$

Probabilisticamente a equação de Chapman-Kolmogorov diz que, para a cadeia fazer uma transição em $m+n$ passos do estado i para o j, ela precisa fazer uma transição em m etapas para algum estado z seguida de uma outra transição em n etapas do estado z para o estado j.

As probabilidades $P_{ij}^{(m+n)}$ podem ser obtidas por meio do produto de matrizes das probabilidades de transição em m e n etapas respectivamente. Isto é $\mathbf{P}^{(m+n)} = \mathbf{P}^{(m)}\mathbf{P}^{(n)}$

$$\text{Prova :} \quad P_{ij}^{(m+n)} = P(X_{m+n} = j|X_0 = i) = \frac{P(X_{m+n} = j|X_0 = i)}{P(X_0 = i)} \qquad (1)$$

$$\begin{aligned} P(X_{m+n} = j|X_0 = i) &= \sum_{z \in S} P(X_{m+n} = j, X_m = z, X_0 = i) \\ &= \sum_{z \in S} P(X_0 = i)P(X_m = z|X_0 = i)P(X_{m+n} = j, X_0 = i, X_m = z) \\ &= P(X_0 = i)\sum_{z \in S} P(X_m = z|X_0 = i)P(X_{m+n} = j, X_m = z) \\ &= P(X_0 = i)\sum_{z \in S} P_{iz}^m P_{zj}^n \qquad (2) \end{aligned}$$

Substituindo (2) em (1):

$$P_{ij}^{(m+n)} = \frac{P(X_0 = i)\sum_{z \in S} P_{iz}^m P_{zj}^n}{P(X_0 = i)} = \sum_{z \in S} P_{iz}^m P_{zj}^n, \qquad \forall\, m,n \geq 0 \quad \forall\, i,j \in S.$$

Uma consequência importante das equações de Chapman-Kolmogorov é que $\mathbf{P}^{(m)} = \mathbf{P}^m$, ou seja, a matriz de transição a m passos da cadeia é igual a m-ésima potência da matriz de transição a um passo. Para ver que isso é verdade, basta usar indução matemática. Veja que se definimos $m = 1, n = 1$ na equação (2.2) temos

$$P_{ij}^{(2)} = \sum_{z \in S} P_{iz} P_{zj}.$$

Claramente $P_{ij}^{(2)}$ é a matriz produto $\mathbf{P}.\mathbf{P} = \mathbf{P}^2$. Baseado neste resultado, por indução assumimos que

$$\mathbf{P}^{(m)} = \mathbf{P}^m, \qquad m = 1, 2, ..., m-1$$

se agora fixamos $m = m-1, n = 1$ na equação (2.2), temos

$$P_{ij}^{(m)} = \sum_{z \in S} P_{iz}^{(m-1)} P_{zj},$$

que agora pode ser visto como o (i, j)-ésimo elemento da matriz produto $\mathbf{P}^m = \mathbf{P}^{m-1}\mathbf{P}$, que prova que $\mathbf{P}^{(m)} = \mathbf{P}^m$.

2.2.1 Classificação dos estados

Definição 2.5 *Dizemos que o estado j é acessível a partir do estado i se $P_m(i, j) > 0$ para algum $m \geq 0$. Para dizer que j é acessível a partir de i, usaremos a notação $i \to j$.*

Isto implica que o estado j é acessível a partir do estado i se, e somente se, começando em i, é possível que o processo, em um número finito de passos, chegue no estado j.

Definição 2.6 *Se dois estados i e j são acessíveis um a partir do outro então dizemos que i e j se comunicam e denotaremos $i \leftrightarrow j$.*

Uma consequência das definições (2.5) e (2.6) e que a relação de comunicação define uma relação de equivalência isto é:

1. Reflexibilidade. $i \leftrightarrow i$, pois

$$P_{ij}^{0} = \delta_{ij} = \begin{cases} 1 & \text{se} \quad i = j \\ 0 & \text{se} \quad i \neq j \end{cases}$$

2. Simetria. Se $i \leftrightarrow j$, implica que, $j \leftrightarrow i$.

3. Transitividade. Se $i \leftrightarrow j$ e $j \leftrightarrow z$, então $i \leftrightarrow z$.

Essa relação de equivalência divide o espaço de estados em classes de equivalência. E assim, se dois estados se comunicam, então eles pertencem à mesma classe. E como temos uma relação de equivalência, se tivermos duas classes, elas serão ou disjuntas ou idênticas.

Definição 2.7 *Uma cadeia de Markov é dita ser irredutível se todos os seus estados pertence a uma mesma classe de equivalência ou seja se todos os seus estados se comunicam entre si.*

Seja

$$f_{ii}^{(m)} = P[X_m = i, X_r \neq i(r = 1, 2, ..., m-1)|X_0 = i]$$

e

$$f_{ii}^{*} = \sum_{m=1}^{\infty} f_{ii}^{(m)}.$$

Logo, $f_{ii}^{(m)}$ é a probabilidade de que começando no estado i, o processo retorne a i pela primeira vez em m passos e f_{ii}^{*} é a probabilidade que começando no estado i, o processo retorne ao estado i em algum tempo f_{ii}^{*}.

Definição 2.8 *Um estado i é dito recorrente se $f_{ii}^{*} = 1$. Se $f_{ii}^{*} < 1$, é chamado transiente ou transitório.*

Teorema 2.2 *Um estado i é recorrente ou transiente de acordo com que,*

$$\sum_{m=1}^{\infty} P_{ii}^{(m)} = \infty \qquad ou \; < \infty.$$

Uma consequência deste teorema é que ele será usado para obter as seguintes propriedades.

Teorema 2.3 *Se $i \leftrightarrow j$ se o estado i e recorrente, então j também é recorrente. (ou seja a propriedade de ser recorrente é uma propriedade de classe).*

Definição 2.9 *Um estado i é dito ser absorvente se e somente se $P_{ii} = 1$. Ou seja, depois que a cadeia entra nele não consegue mais sair.*

Definição 2.10 *Se $C \subset S$ é um conjunto finito fechado irredutível de estados, então todo estado de C é recorrente.*

Uma consequência dos resultados acima é que os estados recorrentes são os estados em que a cadeia visita infinitas vezes enquanto para os estados transitórios sempre existir um instante de tempo a partir do qual a cadeia nunca mais volta neles. Como resultado disto, uma cadeia de Markov com espaço de estados finito tem que possuir pelo menos um estado recorrente. Além disso se a cadeia tiver espaço de estados finito e for irredutível, então todos os seus estados são recorrentes.

Definição 2.11 *Seja i um estado recorrente. Se começando em i, o tempo esperado até que o processo retorne a i for finito, i é dito recorrente positivo. Caso contrário, o estado i é dito recorrente nulo.*

Em uma cadeia de Markov com espaço de estados finito, todos os estados recorrentes são recorrentes positivos. Além disso se $i \leftrightarrow j$ e i é recorrente positivo, então j também é recorrente positivo.

Definição 2.12 *Seja $i \in S$ um estado da cadeia de Markov tal que $P^m(i,i) > 0$ para algum $m \geq 1$. Definimos seu período d_i como o máximo divisor comum de todos os inteiros m tais que $P^m(i,i) > 0$. Isto é,*

$$d_i = mdc\{m \geq 1 : P^m(i,i) > 0\}.$$

Então diremos que:

1. *Se $P(i,i) > 0$, então $d_i = 1$.*
2. *Se $d_i = d > 1$, então dizemos que o estado tem período d.*
3. *Se $d_i = 1$, então o estado é aperiódico.*
4. *Se $i \leftrightarrow j$, segue que $d_i = d_j$.*

Definição 2.13 *Um estado i é dito ergódico se ele é aperiódico e recorrente positivo.*

Na classificação dos estados, vimos que há três tarefas principais:

1. Identificação de classes de equivalência irredutíveis.
2. Classificação das classes de equivalência como sendo recorrente ou transitória.
3. Determinação do período para cada classe.

2.2.2 Representação canônica da matriz de probabilidade de transição

Segundo Bhat e Miller (2002), o primeiro procedimento é reorganizar a matriz de transição $\mathbf{P}$ de modo que todos os estados que pertencem à mesma classe de equivalência, são posicionadas continuamente. Em seguida, organizar estas classes de equivalência de tal maneira que o processo pode ir de um dado estado para outro na mesma ou em uma classe precedente, mas não a um estado em uma classe seguinte. Isto é feito colocando as classes recorrentes antes das transitórias na hierarquia. Se houver mais do que uma classe transiente, que podem ser dispostos com a mesma restrição.

Sejam $\mathbf{C}_1, \mathbf{C}_2, ..., \mathbf{C}_k, \mathbf{C}_{k+1}, ..., \mathbf{C}_n$ as classes de equivalência onde $\mathbf{C}_1, \mathbf{C}_2, ..., \mathbf{C}_k$ são recorrentes e $\mathbf{C}_{k+1}, ..., \mathbf{C}_n$ são transitórias. A matriz de probabilidades de transição pode ser representada da seguinte forma:

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_1 & & & & & \\ & \mathbf{P}_2 & & & \bigcirc & \\ \bigcirc & & \ddots & & & \\ \mathbf{R}_{k+1,1} & \mathbf{R}_{k+1,2} & \dots & \mathbf{R}_{k+1,k} & \mathbf{Q}_{k+1} & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \\ \mathbf{R}_{n1} & \mathbf{R}_{n2} & \dots & \mathbf{R}_{nk} & \mathbf{R}_{n,k+1} & \dots & \mathbf{Q}_n \end{bmatrix},$$

sendo que $\mathbf{P}_1, \mathbf{P}_2, ..., \mathbf{P}_k, \mathbf{Q}_{k+1}, ..., \mathbf{Q}_n$ são as submatrizes de probabilidades de transições correspondentes às classes de equivalência e $\mathbf{R}_{i,j}$ é a submatriz que dá as probabilidades de transição de uma classe de equivalência para as classes anteriores. As regiões com 0 consistem inteiramente de zeros. A matriz de probabilidades de transição disposta desta maneira é dito estar na forma canônica.

Portanto se consideremos uma cadeia de Markov finita o procedimento é reorganizar a matriz de transição $\mathbf{P}$ de modo que os $\mathbf{P}$ estados recorrentes e $\mathbf{Q}$ estados transientes sejam agrupados de modo a se criar uma matriz da forma que se convencionou chamar forma canônica. A matriz $\mathbf{P}$ na sua forma canônica tem a seguinte forma:

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_1 & \mathbf{0} \\ \mathbf{R} & \mathbf{Q} \end{bmatrix},$$

$\mathbf{P}_1$: submatriz com probabilidades de transição entre os estados recorrentes para os seus elementos. Alguns desses estados recorrentes pode até ser absorvente.
$\mathbf{Q}$: representa os elementos da submatriz de probabilidades de transição entre os estados transitórios.
$\mathbf{R}$: submatriz cujos elementos são as probabilidades de transição de um só passo a partir de $\mathbf{Q}$ estados transientes para os estados recorrentes.
$\mathbf{0}$: a submatriz com regiões que consistem inteiramente de zeros.

Vimos que $P^m(i,j)$ é a probabilidade da cadeia ir de um estado i a um estado j em m passos. Neste caso podemos representar $\mathbf{P}^m$ na forma canônica da seguinte maneira.

$$\mathbf{P}^m = \begin{bmatrix} \mathbf{P}_1^m & 0 \\ \mathbf{R}_m & \mathbf{Q}^m \end{bmatrix}.$$

Tomando limite quando $m \to \infty$ temos,

$$\lim_{m\to\infty} \mathbf{P}^m = \begin{bmatrix} \mathbf{P}_1^\infty & \mathbf{0} \\ \mathbf{R}_\infty & \mathbf{Q}^\infty \end{bmatrix}.$$

Das três submatrizes $\mathbf{P}_1^\infty, \mathbf{R}_\infty$ e $\mathbf{Q}^\infty$. $\mathbf{Q}^\infty$ pode ser demonstrado ser uma matriz nula, como indicado no seguinte teorema.

Teorema 2.4 *Numa cadeia de Markov finita, como o número de passos tende para o infinito, a probabilidade de que o processo esteja em um estado transiente tende para zero, independentemente do estado em que o processo começa.*

De acordo com o Teorema 2.4 observa-se que, este é um resultado bastante intuitivo, porque os elementos da submatriz $\mathbf{Q}$ representam probabilidades de transição entre os estados transitórios. A propriedade de transitoriedade de um estado implica que, eventualmente, o processo não vai voltar a ele. Por consequência, na presença de um estado recorrente acessível a partir de um estado transitório, claramente o processo vai acabar no recorrente, em vez de no estado transitório. Usando este argumento para cada um dos estados, correspondente aos elementos de $\mathbf{Q}$. Logo, conclui-se que quando $m \to \infty$ a cadeia de Markov não irá executar a transição entre os estados transitórios, mas teria se mudado para uma das classes recorrentes.

Corolario 2.1 *Se i e j são estados transitórios então $P^{(m)}(i,j) \to 0$ geometricamente.*

Corolario 2.2 *Em uma cadeia de Markov finita, nem todos os estados podem ser transiente.*

Este resultado decorre diretamente do Teorema (2.4) quando observamos que em uma cadeia de Markov de S-estados

$$\sum_{j=1}^{S} P_{ij}^{(m)} = 1, \qquad \forall\, m.$$

2.2.3 Cadeias de Markov finitas com estados transitórios

Aqui o principal interesse é o comportamento da cadeia principal na presença dos estados transitórios. Consideremos uma cadeia de Markov finita com uma classe transitória e um conjunto de estados recorrentes que podem ser agrupados em uma ou mais classes de equivalência. Uma maneira conveniente de representar a matriz de probabilidades de transição é a forma canônica descrito anteriormente.

Seja a cadeia de Markov de S-estados consistem em r estados recorrentes e $(S-r)$ estados transitórios com este último pertencente a uma única classe de equivalência. Seja T o conjunto desses estados transitórios e T^c o conjunto de estados recorrentes. A matriz de probabilidade de transição $\mathbf{P}$ agora pode ser apresentada sob a forma

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_1 & \mathbf{0} \\ \mathbf{R} & \mathbf{Q} \end{bmatrix},$$

onde $\mathbf{P}_1$ é $r \times r$, $\mathbf{R}$ é $(s-r) \times r$ e $\mathbf{Q}$ é $(s-r) \times (s-r)$.

Estamos interessados em responder a algumas das perguntas a seguir: Dado que o processo começa no estado transitório i. Qual é o número esperado de visitas a outro estado transitório j antes que eventualmente entre em um estado recorrente? Qual é a variância desse número de visitas? Qual é a média e a variância do número de passos necessários para deixar a classe transiente por completo? Começando no estado $i \in T$ qual a probabilidade de eventualmente, entrar em um estado recorrente j?

2.2.4 A matriz fundamental

A matriz $\mathbf{M} = (\mathbf{I} - \mathbf{Q})^{-1}$ desempenha um papel útil na determinação das médias e variâncias mencionados acima. Chamaremos $\mathbf{M}$ de matriz fundamental. A existência da inversa da matriz $\mathbf{I} - \mathbf{Q}$ é estabelecida no teorema abaixo.

Teorema 2.5 *A inversa* $(\mathbf{I} - \mathbf{Q})^{-1}$ *existe, e além disso*

$$(\mathbf{I} - \mathbf{Q})^{-1} = \mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + ... = \sum_{r=0}^{\infty} \mathbf{Q}^r. \tag{2.3}$$

Demonstração: por multiplicação direta tem-se

$$(\mathbf{I} - \mathbf{Q})(\mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + ... + \mathbf{Q}^{m-1}) = \mathbf{I} - \mathbf{Q}^m \qquad \text{pois,} \tag{2.4}$$

$$\mathbf{I} - \mathbf{Q} + \mathbf{Q} - \mathbf{Q}^2 + \mathbf{Q}^2 - \mathbf{Q}^3 + \mathbf{Q}^3 + ... - \mathbf{Q}^{m-1} + \mathbf{Q}^{m-1} - \mathbf{Q}^m = \mathbf{I} - \mathbf{Q}^m,$$

pelo Teorema (2.4) e corolário 2.1, $\mathbf{Q}^m$ converge a uma matriz com elementos zero. Portanto o determinante

$$|\mathbf{I} - \mathbf{Q}^m| \to 1, \qquad \text{quando } m \to \infty$$

Portanto para m suficientemente grande, $|\mathbf{I} - \mathbf{Q}^m| \neq 0$ onde,

$$|\mathbf{I} - \mathbf{Q}||\mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + ... + \mathbf{Q}^{m-1}| \neq 0,$$

de onde $|\mathbf{I} - \mathbf{Q}| \neq 0$ ou seja $\mathbf{I} - \mathbf{Q}$ tem inversa. Agora multiplicando ambos os lados da equação (2.4) por $(\mathbf{I} - \mathbf{Q})^{-1}$ tem-se,

$$\mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + ... + \mathbf{Q}^{m-1} = (\mathbf{I} - \mathbf{Q})^{-1}(\mathbf{I} - \mathbf{Q}^m),$$

tomando limite quando $m \to \infty$ temos

$$\mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + ... = (\mathbf{I} - \mathbf{Q})^{-1}.$$

Vamos agora tentar responder as perguntas feitas anteriormente. Todos os resultados são dados em termos da matriz $\mathbf{M} = (\mathbf{I} - \mathbf{Q})^{-1}$.

Seja $N_{ij}(i, j \in T)$ a variável aleatória que denota o número de vezes que o processo visita j antes que, eventualmente entre em um estado recorrente, dado que o processo começa no estado i. Seja $\mu_{ij} = E[N_{ij}]$.

Teorema 2.6

$$||\mu_{ij}|| = \mathbf{M} \qquad\qquad \textit{para } i,j \in T.$$

Demonstração: Inicialmente, a cadeia de Markov está em um estado $i \in T$. Se em um passo entra em um estado recorrente (com probabilidade $\sum_{k\in T^c} P_{ik}$), o número de visitas a j é zero, a menos que $j = i$. Se δ_{ij} é a função Kronecker tal que $\delta_{ij} = 1$ se $i = j$ e 0 se $i \neq j$, podemos escrever $N_{ij} = \delta_{ij}$ com probabilidade $\sum_{k\in T^c} P_{ik}$. Por outro lado, suponha que a cadeia de Markov move-se para um estado $k \in T$ no primeiro passo (com probabilidade P_{ik}). A partir dessa posição em diante, o número de visitas a j é N_{kj}. No entanto, se $i = j$, o número total de visitas a j seria $N_{kj} + \delta_{ij}$. Assim, temos

$$N_{ij} = \begin{cases} \delta_{ij} & \text{com probabilidade } \sum\limits_{k\in T^c} P_{ik} \\ N_{kj} + \delta_{ij} & \text{com probabilidade } \sum\limits_{k\in T} P_{ik} \end{cases}$$

logo,

$$\begin{aligned} E(N_{ij}) &= \sum_{k\in T^c} P_{ik}\delta_{ij} + \sum_{k\in T} P_{ik} E(N_{kj} + \delta_{ij}) \\ &= \sum_{k\in T} P_{ik}\mu_{kj} + \sum_{k\in T^c} P_{ik}\delta_{ij} + \sum_{k\in T} P_{ik}\delta_{ij} \\ &= \sum_{K\in T} P_{ik}\mu_{kj} + \delta_{ij}\left[\sum_{k\in T^c} P_{ik} + \sum_{k\in T} P_{ik}\right] \end{aligned}$$

que dá

$$\mu_{ij} = \delta_{ij} + \sum_{k\in T} P_{ik}\mu_{kj},$$

usando todos os elementos da matriz $||\mu_{ij}||$, temos

$$\begin{aligned} ||\mu_{ij}|| &= \mathbf{I} + \mathbf{Q}||\mu_{ij}|| \\ ||\mu_{ij}|| - \mathbf{Q}||\mu_{ij}|| &= \mathbf{I} \\ \mathbf{I}||\mu_{ij}|| - \mathbf{Q}||\mu_{ij}|| &= \mathbf{I} \\ (\mathbf{I} - \mathbf{Q})||\mu_{ij}|| &= \mathbf{I}. \end{aligned}$$

Multiplicando ambos os lados por $(\mathbf{I} - \mathbf{Q})^{-1}$ tem-se

$$||\mu_{ij}|| = (\mathbf{I} - \mathbf{Q})^{-1} = \mathbf{M}.$$

Ou seja $\mathbf{M} = (\mathbf{I} - \mathbf{Q})^{-1}$ é o número médio de vezes que o processo visita o estado transitório j antes de entrar numa classe recorrente, dado que o processo começou no estado i.

Teorema 2.7

$$||\sigma^2_{ij}|| = \mathbf{M}(2\mathbf{M}_D - \mathbf{I}) - \mathbf{M}_2, \qquad\qquad i,j \in T.$$

Demonstração: por definição seja $\sigma_{ij}^2 = V(N_{ij})$ então

$$V(N_{ij}) = E(N_{ij}^2) - [E(N_{ij})]^2.$$

A partir do Teorema (2.6) fica claro que

$$||[E(N_{ij})]^2|| = \mathbf{M}_2.$$

Assumindo que $E(N_{ij}^2)$ é finita vamos proceder da mesma forma que antes

$$N_{ij}^2 = \begin{cases} \delta_{ij}^2 & \text{com probabilidade } \sum_{k\in T^c} P_{ik} \\ (N_{kj} + \delta_{ij})^2 & \text{com probabilidade } \sum_{k\in T} P_{ik} \end{cases}$$

$$\begin{aligned} E(N_{ij}^2) &= \sum_{k\in T^c} P_{ik}\delta_{ij}^2 + \sum_{k\in T} P_{ik} E(N_{kj} + \delta_{ij})^2 \\ &= \sum_{k\in T^c} P_{ik}\delta_{ij}^2 + \sum_{k\in T} P_{ik} E(N_{kj}^2 + 2N_{kj}\delta_{ij} + \delta_{ij}^2) \\ &= \sum_{k\in T^c} P_{ik}\delta_{ij}^2 + \sum_{k\in T} P_{ik} E(N_{kj}^2) + 2\sum_{k\in T} P_{ik} E(N_{kj})\delta_{ij} + \sum_{k\in T} P_{ik}\delta_{ij}^2 \\ &= \sum_{k\in T} P_{ik} E(N_{kj}^2) + 2\sum_{k\in T} P_{ik} E(N_{kj})\delta_{ij} + \delta_{ij}^2 \left[\sum_{k\in T^c} P_{ik} + \sum_{k\in T} P_{ik}\right] \\ &= \sum_{k\in T} P_{ik} E(N_{kj}^2) + 2\sum_{k\in T} P_{ik} E(N_{kj})\delta_{ij} + \delta_{ij}. \end{aligned}$$

Usando todos os elementos da matriz $||\mu_{ij}^2||$ tem-se

$$\begin{aligned} ||E(N_{ij}^2)|| &= \mathbf{Q}||E(N_{ij}^2)|| + 2(\mathbf{QM})_D + \mathbf{I} \\ \mathbf{I}||E(N_{ij}^2)|| - \mathbf{Q}||E(N_{ij}^2)|| &= 2(\mathbf{QM})_D + \mathbf{I}. \end{aligned}$$

Mas nota-se que

$$\begin{aligned} \mathbf{QM} &= \mathbf{Q}(\mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \mathbf{Q}^3 + ...) \\ &= \mathbf{Q} + \mathbf{Q}^2 + \mathbf{Q}^3 + ... \\ &= \mathbf{M} - \mathbf{I}. \end{aligned}$$

e

$$\begin{aligned} (\mathbf{QM})_D &= (\mathbf{M} - \mathbf{I})_D \\ &= \mathbf{M}_D + \mathbf{I}, \end{aligned}$$

então

$$(\mathbf{I} - \mathbf{Q})||E(N_{ij}^2)|| = 2(\mathbf{M}_D - \mathbf{I}) + \mathbf{I}.$$

Multiplicando ambos os lados por $(\mathbf{I}-\mathbf{Q})^{-1}$ isto dá

$$\begin{aligned} ||E(N_{ij}^2)|| &= (\mathbf{I}-\mathbf{Q})^{-1}[2(\mathbf{M}_D-\mathbf{I})+\mathbf{I}] \\ &= \mathbf{M}[2(\mathbf{M}_D-\mathbf{I})+\mathbf{I}]. \end{aligned}$$

Portanto

$$||\sigma_{ij}^2|| = \mathbf{M}[2(\mathbf{M}_D-\mathbf{I})+\mathbf{I}] - \mathbf{M}_2 \qquad i,j\in T.$$

Seja $N_i = \sum\limits_{j\in T} N_{ij}$ a variável aleatória que indica o número total de passos passados em estados transitórios antes de entrar em um estado recorrente.

Usando os resultados obtidos pelos teoremas (2.6) e (2.7) podemos obter média e variância de $N_i (i \in T)$. Por definição,

$$\mathbf{M}_\rho = ||\sum_{j\in T}\mu_{ij}||.$$

Um vetor coluna, cujo k-ésimo componente é a soma dos elementos da k-ésima linha de $\mathbf{M}$, e

$$\mathbf{M}_{\rho^2} = ||(\sum_{j\in T}\mu_{ij})^2||.$$

Um vetor coluna cujo k-ésimo componente é o quadrado do k-ésimo componente em $\mathbf{M}_\rho$

Teorema 2.8

$$||E(N_i)|| = \mathbf{M}_\rho \qquad i\in T.$$

Demonstração: claramente

$$\begin{aligned} E(N_i) &= E(\sum_{j\in T}N_{ij}) \\ &= \sum_{j\in T}E(N_{ij}) \qquad \text{usando todos os elementos } ||\mu_{ij}|| \text{ tem-se} \\ ||E(N_i)|| &= ||\sum_{j\in T}\mu_{ij}|| \\ &= \mathbf{M}_\rho. \end{aligned}$$

Teorema 2.9

$$||V(N_i)|| = (2\mathbf{M}-\mathbf{I})\mathbf{M}_\rho - \mathbf{M}_{\rho^2}.$$

Demonstração: procedendo como no Teorema (2.6), temos

$$E(N_i^2) = \sum_{k\in T^c}P_{ik}\times 1 + \sum_{k\in T}P_{ik}E(N_k+1)^2$$

Note que uma vez que estamos interessados no número total de passos no lugar de δ_{ij} temos $\sum\limits_{k\in T}\delta_{ij}=1$. Temos

$$\begin{aligned}
E(N_i^2) &= \sum_{k\in T^c} P_{ik} + \sum_{k\in T} P_{ik}E(N_k+1)^2 \\
&= \sum_{k\in T^c} P_{ik} + \sum_{k\in T} P_{ik}E(N_k^2+2N_k+1) \\
&= \sum_{k\in T^c} P_{ik} + \sum_{k\in T} P_{ik}E(N_k^2) + 2\sum_{k\in T} P_{ik}E(N_k) + \sum_{k\in T} P_{ik} \\
&= \sum_{k\in T} P_{ik}E(N_k^2) + 2\sum_{k\in T} P_{ik}E(N_k) + 1 \qquad \text{Dando} \\
||E(N_i^2)|| &= \mathbf{Q}||E(N_i^2)|| + 2\mathbf{Q}\mathbf{M}_\rho + \epsilon,
\end{aligned}$$

em que ϵ é um vetor coluna que consiste de apenas $1'$s. Logo

$$\begin{aligned}
||E(N_i^2)|| - \mathbf{Q}||E(N_i^2)|| &= 2\mathbf{Q}\mathbf{M}_\rho + \epsilon \\
(\mathbf{I}-\mathbf{Q})||E(N_i^2)|| &= 2\mathbf{Q}\mathbf{M}_\rho + \epsilon \\
||E(N_i^2)|| &= (\mathbf{I}-\mathbf{Q})^{-1}[2\mathbf{Q}\mathbf{M}_\rho + \epsilon] \\
&= 2\mathbf{M}\mathbf{Q}\mathbf{M}_\rho + \mathbf{M}_\rho \\
&= 2(\mathbf{M}-\mathbf{I})\mathbf{M}_\rho + \mathbf{M}_\rho \\
&= 2\mathbf{M}\mathbf{M}_\rho - 2\mathbf{M}_\rho + \mathbf{M}_\rho \\
&= 2\mathbf{M}\mathbf{M}_\rho - \mathbf{M}_\rho \\
&= (2\mathbf{M}-\mathbf{I})\mathbf{M}_\rho.
\end{aligned}$$

Procedendo como no Teorema (2.8) tem-se que

$$||[E(N_i)]^2|| = \mathbf{M}_{\rho^2},$$

então a variância de N_i é

$$\begin{aligned}
||V(N_i)|| &= ||E(N_i^2)|| - ||[E(N_i)]^2|| \\
&= (2\mathbf{M}-\mathbf{I})\mathbf{M}_\rho - \mathbf{M}_{\rho^2}.
\end{aligned}$$

Agora considere $f_{ij}^{(m)}$ como sendo a probabilidade de que começando a partir do estado transitório i, o processo entre em um estado recorrente j em m passos. Nota-se que, quando $i\in T$ e $j\in T^c$, a transição $j\to i$ não será possível. O número de passos N_{ij} é, de fato, o tempo necessário para uma primeira passagem de transição $i\to j$. Se definimos T_{ij} para representar esta variável aleatória, podemos escrever $f_{ij}^{(m)}$, como a sua distribuição dada por

$$P(T_{ij}=m)=f_{ij}^{(m)}, \qquad i\in T,\ j\in T^c.$$

Seja

$$f_{ij}=\sum_{m=1}^{\infty} P_{ij}^{(m)}.$$

A probabilidade de uma eventual passagem para j. Em notação matricial também temos que

$$\mathbf{F}^{(m)} = ||f_{ij}^{(m)}|| \qquad \text{e} \qquad \mathbf{F} = ||f_{ij}||.$$

Teorema 2.10

$$\mathbf{F}^{(m)} = \mathbf{Q}^{m-1}\mathbf{R} \qquad e \qquad \mathbf{F} = \mathbf{MR}.$$

Demonstração: tem-se que

$$f_{ij}^{(1)} = P_{ij}, \qquad i \in T,\ j \in T^c$$

$$f_{ij}^{(m)} = \sum_{k \in T} P_{ik} f_{kj}^{(m-1)},$$

que é

$$\mathbf{F}^{(1)} = \mathbf{R}$$

$$\mathbf{F}^{(m)} = \mathbf{QF}^{(m-1)},$$

que em iteração com $m = 1, 2, ...,$ dando

$$\mathbf{F}^{(m)} = \mathbf{Q}^{(m-1)}\mathbf{R}.$$

Também

$$\begin{aligned}
\mathbf{F} &= \sum_{m=1}^{\infty} \mathbf{F}^{(m)} \\
&= \mathbf{R} + \sum_{m=2}^{\infty} \mathbf{Q}^{(m-1)}\mathbf{R} \\
&= (\mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + ...)\mathbf{R} \\
&= \mathbf{MR}.
\end{aligned}$$

Está completa a demonstração.

Corolario 2.3 *Seja $\mathbf{R}_j$ a j-ésima coluna de $\mathbf{R}$. Então a probabilidade que começando em um estado qualquer $i \in T$, o processo entre em um estado $j, j \in T^c$, é dado por $\mathbf{MR}_j$.*

2.3 Cadeias de Markov irredutíveis com estados ergódicos

Foi visto anteriormente que, quando uma cadeia de Markov é irredutível e aperiódica e recorrente positiva, é ergódica no sentido de que o seu comportamento limite pode ser determinado como um conjunto de médias ou como um tempo médio.

2.3.1 Comportamento transiente

Usando as equações de Chapmam-kolmogorov foi mostrado que as probabilidades de transições em m-passos $P_{ij}^{(m)}$ são obtidas como elementos da enésima potência da matriz de probabilidade de transição $\mathbf{P}$. Porém soluções analíticas estão disponíveis em alguns casos especiais. Portanto, com o avanço da computação que, fez com que a multiplicação de matrizes fosse realizável, este procedimento parece fornecer a solução mais eficiente. Uma vantagem da multiplicação de matrizes na determinação das probabilidades de $P_{ij}^{(m)}$ é que ela pode ser aplicada sempre sem estabelecer classes de equivalência. Mas observa-se, a única informação que é a magnitude de $P_{ij}^{(m)}$, mas não outras propriedades da cadeia de Markov que pode derivar a partir da classificação de estados. São dado a seguir dois casos especiais onde soluções analíticas estão disponíveis na determinação de $P_{ij}^{(m)}$.

2.3.2 Cadeias de Markov de dois estados

Cadeias de Markov de dois estados é simples de conceituar e apropriado como um modelo matemático quando há apenas dois resultados ou modelo iniciador como em situações mais gerais. Estas cadeias são bastantes úteis para modelar sistemas que alternam entre dois estados como por exemplo ON ou OFF. Seja os dois estados, 0 e 1 e suponha que a matriz de probabilidade de transição P é dada por

$$\mathbf{P} = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}, \quad 0 \leq a,\ b \leq 1.$$

Para as probabilidades de transição em m passos, expressões explícitas são dadas, como

$$\mathbf{P}^m = \begin{bmatrix} P_{00}^{(m)} & P_{01}^{(m)} \\ P_{10}^{(m)} & P_{11}^{(m)} \end{bmatrix},$$

$$\mathbf{P}^m = \begin{bmatrix} \dfrac{b}{a+b} + a\dfrac{(1-a-b)^m}{a+b} & \dfrac{a}{a+b} - a\dfrac{(1-a-b)^m}{a+b} \\ \dfrac{b}{a+b} - b\dfrac{(1-a-b)^m}{a+b} & \dfrac{a}{a+b} + b\dfrac{(1-a-b)^m}{a+b} \end{bmatrix}.$$

Demonstração: usando as equações de Chapman-Kolmogorov, temos

$$\mathbf{P}^m = \mathbf{P}^{m-1}\mathbf{P}$$

$$\mathbf{P}^m = \begin{bmatrix} P_{00}^{(m-1)} & P_{01}^{(m-1)} \\ P_{10}^{(m-1)} & P_{11}^{(m-1)} \end{bmatrix} \times \begin{bmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{bmatrix} \tag{2.5}$$

logo

$$
\begin{aligned}
P_{00}^{m} &= P_{00}^{(m-1)}P_{00} + P_{01}^{(m-1)}P_{10} \\
&= P_{00}^{(m-1)}(1-a) + (1-P_{00}^{(m-1)})b \\
&= P_{00}^{(m-1)} - aP_{00}^{(m-1)} + b - P_{00}^{(m-1)}b \\
&= b + (1-a-b)P_{00}^{(m-1)}
\end{aligned}
$$

Com $P_{00} = 1 - a$
Fazendo, $P_{00}^{(m)} = (1-a-b)^{(m-1)}y_m$, temos

$$
P_{00} = (1-a-b)^0 y_1 \Rightarrow P_{00} = y_1 = (1-a) \qquad \text{e}
$$

$$
\begin{aligned}
P_{00}^{m} &= b + (1-a-b)P_{00}^{(m-1)} \\
(1-a-b)^{(m-1)}y_m &= b + (1-a-b)(1-a-b)^{(m-2)}y_{m-1} \\
(1-a-b)^{(m-1)}y_m &= (1-a-b)^{(m-1)}y_{m-1} + b \\
y_m &= y_{m-1} + \frac{b}{(1-a-b)^{m-1}}
\end{aligned}
$$

Ou seja,

$$
\begin{aligned}
y_2 &= y_1 + \frac{b}{(1-a-b)^1} \\
y_3 &= y_2 + \frac{b}{(1-a-b)^2} \\
&\vdots \\
y_m &= y_{m-1} + \frac{b}{(1-a-b)^{m-1}}
\end{aligned}
$$

Donde

$$
\begin{aligned}
y_2 + y_3 + \cdots + y_m &= y_1 + \frac{b}{(1-a-b)} + y_2 + \frac{b}{(1-a-b)^2} + \cdots + y_{m-1} + \frac{b}{(1-a-b)^{m-1}} \\
y_m &= y_1 + b\left[\frac{1}{(1-a-b)} + \cdots + \frac{1}{(1-a-b)^{m-1}}\right] \\
y_m &= (1-a) + b\left[\frac{1}{(1-a-b)} \times \left[\frac{1-\left(\frac{1}{1-a-b}\right)^{m-1}}{1-\frac{1}{(1-a-b)}}\right]\right] \\
&= (1-a) + b\left[\frac{1-\frac{1}{(1-a-b)^{m-1}}}{(1-a-b)-1}\right] \\
&= (1-a) + \frac{b}{a+b} \times \left[\frac{1}{(1-a-b)^{m-1}} - 1\right]
\end{aligned}
$$

Consequentemente

$$\begin{aligned}
P_{00}^m &= (1-a-b)^{m-1}\left[\frac{(1-a)(a+b)-b}{a+b} + \frac{b}{(a+b)} \times \frac{1}{(1-a-b)^{m-1}}\right] \\
&= (1-a-b)^{m-1}\left[\frac{a(1-a-b)}{a+b} + \frac{b}{(a+b)} \times \frac{1}{(1-a-b)^{m-1}}\right] \\
&= \frac{a(1-a-b)^m}{a+b} + \frac{b}{a+b}.
\end{aligned}$$

Para P_{11}^m tem-se que,

$$\begin{aligned}
P_{11}^m &= P_{10}^{(m-1)}P_{01} + P_{11}^{(m-1)}P_{11} \\
&= (1-P_{11}^{(m-1)})a + P_{11}^{(m-1)}(1-b) \\
&= a - P_{11}^{(m-1)}a + P_{11}^{(m-1)} - P_{11}^{(m-1)}b \\
&= a + (1-a-b)P_{11}^{(m-1)},
\end{aligned}$$

pois $P_{11} = 1-b$. Fazendo, $P_{11}^{(m)} = (1-a-b)^{(m-1)}y_m$, obtemos

$$P_{11} = (1-a-b)^0 y_1 \Rightarrow P_{11} = y_1 = (1-b) \qquad \text{e}$$

$$\begin{aligned}
P_{11}^m &= a + (1-a-b)P_{11}^{(m-1)} \\
(1-a-b)^{(m-1)}y_m &= a + (1-a-b)(1-a-b)^{(m-2)}y_{m-1} \\
(1-a-b)^{(m-1)}y_m &= (1-a-b)^{(m-1)}y_{m-1} + a \\
y_m &= y_{m-1} + \frac{a}{(1-a-b)^{m-1}}
\end{aligned}$$

Ou seja,

$$\begin{aligned}
y_2 &= y_1 + \frac{a}{(1-a-b)^1} \\
y_3 &= y_2 + \frac{a}{(1-a-b)^2} \\
&\vdots \\
y_m &= y_{m-1} + \frac{a}{(1-a-b)^{m-1}}
\end{aligned}$$

Dessa forma, segue que

$$\begin{aligned}
y_2 + y_3 + \cdots + y_m &= y_1 + \frac{a}{(1-a-b)} + y_2 + \frac{a}{(1-a-b)^2} + \cdots + y_{m-1} + \frac{a}{(1-a-b)^{m-1}} \\
y_m &= y_1 + a\left[\frac{1}{(1-a-b)} + \cdots + \frac{1}{(1-a-b)^{m-1}}\right] \\
y_m &= (1-b) + a\left[\frac{1}{(1-a-b)} \times \left[\frac{1 - \left(\dfrac{1}{1-a-b}\right)^{m-1}}{1 - \dfrac{1}{(1-a-b)}}\right]\right] \\
&= (1-b) + a\left[\frac{1 - \dfrac{1}{(1-a-b)^{m-1}}}{(1-a-b)-1}\right] \\
&= (1-b) + \frac{a}{a+b} \times \left[\frac{1}{(1-a-b)^{m-1}} - 1\right]
\end{aligned}$$

logo

$$\begin{aligned}
P_{11}^m &= (1-a-b)^{m-1}\left[\frac{(1-b)(a+b)-a}{a+b} + \frac{a}{(a+b)} \times \frac{1}{(1-a-b)^{m-1}}\right] \\
&= (1-a-b)^{m-1}\left[\frac{b(1-a-b)}{a+b} + \frac{a}{(a+b)} \times \frac{1}{(1-a-b)^{m-1}}\right] \\
&= \frac{b(1-a-b)^m}{a+b} + \frac{a}{a+b}.
\end{aligned}$$

Usando a propriedade complementar encontramos as probabilidades de transições P_{01}^m e P_{10}^m, isto é

$$\begin{aligned}
P_{01}^m &= 1 - P_{00}^m \\
&= 1 - \left[\frac{b}{a+b} + \frac{a(1-a-b)^m}{a+b}\right] \\
&= \frac{a}{a+b} - \frac{a(1-a-b)^m}{a+b}. \\
P_{10}^m &= 1 - P_{11}^m \\
&= 1 - \left[\frac{a}{a+b} + \frac{b(1-a-b)^m}{a+b}\right] \\
&= \frac{b}{a+b} - \frac{b(1-a-b)^m}{a+b}.
\end{aligned}$$

Está completa a demonstração, para a e b não simultaneamente zero.

2.3.3 O método de autovalor para cadeias de Markov finitas irredutíveis

Se $\mathbf{P}$ é uma matriz $m \times m$, os autovalores de $\mathbf{P}$ são aqueles números λ para o qual a equação característica.

$$\mathbf{PX} = \lambda \mathbf{X}.$$

Tem uma solução $\mathbf{X} \neq 0$. O vetor coluna $\mathbf{X}$ é chamado de autovetor direito pertencente ao autovalor λ. De modo análogo, defini-se um vetor linha $\mathbf{Y}$ como um autovetor esquerdo pertencente ao autovalor λ se existir $\mathbf{Y} \neq 0$ tal que.

$$\mathbf{YP} = \lambda \mathbf{Y}.$$

Logo, notar-se que os autovalores são obtidos resolvendo a equação determinante.

$$|\lambda \mathbf{I} - \mathbf{P}| = 0.$$

Além do mais, quando os m autovalores de $\mathbf{P}$ são todos distintos, podemos obter m autovetores direitos linearmente independentes e m autovetores esquerdos linearmente independentes pertencentes a esses autovalores distintos.

Seja $\mathbf{Q}$ a matriz não singular $(\mathbf{X}_1, ..., \mathbf{X}_m)$, onde $\mathbf{X}_i$ é o autovetor direito (coluna) pertencente ao autovalor $\lambda_i, (i = 1, 2, ..., m)$. Então podemos escrever

$$\mathbf{PX}_j = \lambda_j \mathbf{X}_j,$$

de modo que

$$\mathbf{PQ} = \mathbf{Q\Lambda}$$

e

$$\mathbf{Q}^{-1}\mathbf{PQ} = \mathbf{\Lambda}, \tag{2.6}$$

Em que $\mathbf{\Lambda}$ é a matriz diagonal

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_m \end{bmatrix}.$$

Quando realizamos (2.6) a matriz $\mathbf{P}$ é dita ser diagonalizável. Verifica-se imediatamente que se tem

$$\begin{aligned} \mathbf{P} &= \mathbf{Q\Lambda Q}^{-1} \\ \mathbf{P}^2 &= (\mathbf{Q\Lambda Q}^{-1})(\mathbf{Q\Lambda Q}^{-1}) = \mathbf{Q\Lambda}^2\mathbf{Q}^{-1} \\ &\vdots \\ \mathbf{P}^m &= \mathbf{Q\Lambda}^m\mathbf{Q}^{-1}, \end{aligned}$$

Onde

$$\mathbf{\Lambda}^m = \begin{bmatrix} \lambda_1^m & & & 0 \\ & \lambda_2^m & & \\ & & \ddots & \\ 0 & & & \lambda_m^m \end{bmatrix}.$$

Este método dá uma forma conveniente para $\mathbf{P}^m$ para fins computacionais. Os vários passos desse processo pode ser resumido como segue:

1. Determinar os autovalores da matriz de probabilidade de transição $\mathbf{P}$ resolvendo o determinante da equação
$$|\lambda\mathbf{I} - \mathbf{P}|.$$
se os autovalores $\lambda_i (i = 1, 2, ..., m)$ são distintos proceder, caso contrário para. (consulte livros avançados sobre o assunto)

2. Se $\lambda_i' s$ são distintos, obter o vetor coluna m $(\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_m)$ correspondente aos m autovalores por resolver a equação do tipo $\mathbf{PX} = \lambda\mathbf{X}$. Assim determine
$$\mathbf{Q} = ||\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_m||.$$

3. Obter $\mathbf{Q}^{-1}$. Isso também pode ser determinado resolvendo a equação do tipo $\mathbf{YP} = \lambda\mathbf{Y}$, mas em seguida, o multiplicativo constante deve ser escolhido de modo a tornar
$$\mathbf{QQ}^{-1} = \mathbf{I}.$$

4. Obter $\mathbf{P}^m$ a partir da relação
$$\mathbf{P}^m = \mathbf{Q}\mathbf{\Lambda}^m\mathbf{Q}^{-1}.$$

2.3.4 Medida invariante

Muitas vezes estamos interessados no comportamento de um processo de Markov após um longo período de tempo, particularmente se esse comportamento se "estabiliza" probabilisticamente. As propriedades de uma cadeia de Markov após de muitas transições estão relacionadas com a noção de distribuição invariante.

Definição 2.14 *Seja $\{X_n,\ n \geq 0\}$ uma cadeia de Markov irredutível com os estados ergódicos, as probabilidades limitantes $\{\pi_i\}_{i=0}^{\infty}$ satisfazem as equações*
$$\pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}, \qquad j = 0, 1, 2, ... \tag{2.7}$$
$$\sum \pi_j = 1,$$
onde π foi denominada de medida invariante para a cadeia de Markov.

Em notação matricial
$$\pi \ = \ \pi\mathbf{P}, \tag{2.8}$$
onde P é a matriz de transição da cadeia e $\pi = (\pi_{(0)}, \pi_{(1)}, \pi_{(2)}, ...)$. A distribuição obtida é invariante.

Teorema 2.11 *Se π é uma distribuição invariante, então* $\pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}^{(m)}, \qquad m \geq 1.$

Demonstração:

$$
\begin{aligned}
\pi_j &= \sum_{i=0}^{\infty} \pi_i P_{ij} \\
\pi_k &= \sum_j \pi_j P_{jk} = \sum_{i=0}^{\infty} \pi_i \sum_j P_{ij} P_{jk} = \sum_i \pi_i P_{ik}^2 \\
\vdots \quad & \quad \vdots \qquad\qquad \vdots \\
&= \sum_j \pi_j P_{jk}^{m-1} = \sum_{i=0}^{\infty} \pi_i \sum_j P_{ij} P_{jk}^{(m-1)} = \sum_i \pi_i P_{ik}^{(m)}.
\end{aligned}
$$

Dando

$$\pi_j = \sum_i \pi_i P_{ij}^{(m)}, \qquad m \geq 1.$$

além do mais, se $\pi_0 = \pi$, então $\pi_m = \pi$ para qualquer $m \geq 1$.

Teorema 2.10 Para uma cadeia de Markov irredutível ergódica, $\lim_{m\to\infty} P_{ij}^m$ existe e é independente de i. Além do mais

$$\pi_j = \lim_{m\to\infty} P_{ij}^m, \qquad j \geq 0$$

e π_j é a única medida invariante da cadeia. Após o processo está sendo executado por um longo período a probabilidade de encontrarmos a cadeia em um dado estado j é π_j.

2.3.5 Tempo de permanência em um estado

O tempo de permanência em um estado é a quantidade de tempo que um processo permanece em um estado durante uma visita. Suponha que a cadeia de Markov está em um estado i em algum momento. Seja α_i ser o número de novos períodos de tempo que ele permanece no estado i até que ele se move fora do estado. A distribuição de probabilidade de α_i é obtida considerando as transições como ensaios repetidos com resultados $i \to i$ e $i \nrightarrow i$ com probabilidades p_{ii} e $1 - p_{ii}$, respectivamente. Assim temos uma distribuição geométrica para α_i, dado por

$$P(\alpha_i = m) = P_{ii}^m (1 - P_{ii}),$$

também tem-se que

$$E(\alpha_i) = \frac{P_{ii}}{1 - P_{ii}},$$

$$V(\alpha_i) = \frac{P_{ii}}{(1 - P_{ii})^2}.$$

Segundo Bhat e Miller (2002) estes resultados não devem ser interpretados, que em qualquer cadeia de Markov estados da cadeia podem ser agrupados de qualquer maneira. Tudo o que temos feito aqui é considerar um sentido único de transição $i \nrightarrow i$, a fim de obter a distribuição do número de etapas que uma cadeia de Makov ficar em um estado sem mover fora dele.

2.4 Processos de ramificação e outros tópicos especiais

2.4.1 Processos de ramificação

Uma classe especial de cadeias de Markov é um processo de ramificação. Usando a terminologia biológica considere uma situação em que cada organismo de uma geração produz um número aleatório de descendentes para a próxima geração. Dada a distribuição de probabilidade do número de descendentes produzidos por um organismo, estamos interessados em características tais como a distribuição (média e variância em especial) do tamanho da população para diferentes gerações e a probabilidade de extinção da população.

Exemplos de fenômenos naturais que podem ser modelados como um processo de ramificação são comuns. Um exemplo clássico refere-se à sobrevivência dos nomes de família. O fluxo de informação verbal também pode ser modelado como um processo de ramificação. No entanto, aqui usaremos a terminologia biológica de organismo e descendente.

Suponha que até o final de sua vida útil um organismo produz um número aleatório de descendentes S com uma distribuição de probabilidade,

$$P(S = j) = p_j, \qquad j = 0, 1, 2, ... \tag{2.9}$$

De tal forma que $p_j \geq 0$ e $\sum_j p_j = 1$. Seja $S_1, S_2, ..., S_i, ...$ variáveis aleatórias independentes identicamente distribuídas como em (2.9). Considere organismos que podem gerar novos organismo do mesmo tipo. O conjunto inicial de organismo é dito pertencer à geração 0. Organismos gerados a partir da n-ésima geração são ditos pertencer à $(n + 1)$-ésima geração. Então, definindo, para $n \geq 0$, X_n como sendo o tamanho da população na n-ésima geração, tem-se uma cadeia $\{X_n, n \geq 0\}$ que é markoviana e que se chama cadeia de ramificação ou processo de ramificação. Tem-se

$$X_{n+1} = \sum_{k=1}^{X_n} S_k$$

e

$$\begin{aligned} P_{ij} &= P(X_{n+1} = j | X_n = i) \\ &= P\left(\sum_{k=1}^{i} S_k = j\right). \end{aligned} \tag{2.10}$$

Em que S_k é o número de descendentes produzidos pelo K-ésimo membro da n-ésima geração. Deve notar-se que X_{n+1} é a soma de variáveis aleatórias, e uma vez que

$S_k, k = 1, 2, ...$ são independentes e identicamente distribuídas, logo

$$\begin{aligned} E[X_{n+1}] &= E[X_n]E[S_k] \\ V[X_{n+1}] &= E[X_n]V[S_k] + V[X_n]\{E[S_k]\}^2 \end{aligned} \tag{2.11}$$

Suponhamos que a população começa com um organismo (isto é, $X_0 = 1$). Seja

$$E[S_k] = \mu, \qquad \text{e} \qquad V[S_k] = \sigma^2 \qquad k = 1, 2, 3, ...$$

Uma vez que $X_1 = S_1$,

$$E[X_1] = \mu \qquad \text{e} \qquad V[X_1] = \sigma^2$$

Temos também que

$$E[X_n] = E[X_{n-1}]\mu, \qquad n > 1$$

Portanto,

$$E[X_n] = \mu^n.$$

Pois recursivamente temos que

$$\begin{aligned} E(X_n) &= \mu E(X_{n-1}) \qquad n > 1 \\ E(X_2) &= \mu E(X_1) \\ E(X_3) &= \mu E(X_2) \\ &\vdots \\ E(X_n) &= \mu E(X_{n-1}) \\ E(X_2) \times E(X_3) \times, ..., \times E(X_n) &= \mu E(X_1) \times \mu E(X_2) \times, ..., \times \mu E(X_{n-1}) \\ E(X_n) &= \mu^{n-1} \times E(X_1) \\ E(X_n) &= \mu^{n-1} \times \mu \\ E(X_n) &= \mu^n. \end{aligned}$$

Segundo Bhat e Miller (2002) a variância da formula em (2.11) pode ser aplicada recursivamente para produzir

$$V[X_n] = \begin{cases} \sigma^2 \mu^{n-1} \dfrac{[1 - \mu^n]}{1 - \mu} & \text{se } \mu \neq 1 \\ n\sigma^2 & \text{se } \mu = 1 \end{cases}$$

Para $n \geq 1$.

A partir da $E[X_n]$ e $V[X_n]$, observa-se que a média e a variância de X_n aumenta ou diminuir geometricamente conforme $\mu > 1$ ou $\mu < 1$. Também usando a desigualdade Chebyshev, que afirma que

$$P[|X_n - E(X_n)| > \epsilon] \leq \frac{V[X_n]}{\epsilon^2}.$$

Como $n \to \infty$, quando $\mu < 1$, logo

$$P[|X_n - 0| > \epsilon] \to 0,$$

$$P(X_n = 0) \to 1.$$

Portanto, pelo fato de μ ser o número médio de descendentes de cada organismo se $\mu < 1$, as chances da população sobreviver são pequenas, ou seja para n suficientemente grande a população vai ser extinta. Quando $\mu \geq 1$, a probabilidade da população ser extinta não é obtida facilmente. Neste caso, vamos usar a função geradora de probabilidade (f.g.p.), que também será útil na determinação das probabilidades de transição n-passo do processo de ramificação. Ou seja, consideremos a função ϕ definida por

$$\phi_{(z)} = \sum_{j=0}^{\infty} P_j z^j, \qquad |z| \leq 1.$$

Notando que a f.g.p da soma de variáveis aleatórias independentes é o produto de f.g.p's de variáveis aleatórias individuais, logo

$$\sum_{j=0}^{\infty} P_{ij} z^j = [\phi(z)]^i. \tag{2.12}$$

Agora permita que o tamanho da população inicial exceder 1 e considere as probabilidades de transição n-passos do processo de ramificação $\{X_n, n = 0, 1, ...\}$. Podemos escrever

$$P_{ij}^{(n)} = P[X_n = j | X_0 = i]$$

e

$$\phi_n(z) = \sum_{j=0}^{\infty} P_{ij}^{(n)} z^n, \qquad |z| \leq 1.$$

Quando $X_0 = i$ os processos de ramificação gerados pelos organismos i são independentes um do outro e portanto, o número total de descendentes na n-ésima geração é a soma de descendentes de cada organismo. Novamente usando a propriedade f.g.p acima mencionada, tem-se

$$\sum_{j=0}^{\infty} P_{ij}^{(n)} z^j = [\phi_n(z)]^i.$$

Assim, sem perda de generalidade, podemos novamente assumir que $i = 1$. Condicionado sobre o tamanho da $(n-1)$-ésima população, tem-se

$$\begin{aligned} \phi_n(z) &= \sum_{j=0}^{\infty} P(X_n = j) z^j \\ &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} P(X_n = j | X_{n-1} = k) P(X_{n-1} = k) z^j, \end{aligned}$$

usando (2.10) isto dá

$$\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} P(S_1 + S_2 + \cdots + S_k = j)P(X_{n-1} = k)z^j$$
$$= \sum_{k=0}^{\infty} P(X_{n-1} = k) \sum_{j=0}^{\infty} P(S_1 + S_2 + \cdots + S_k = j)z^j,$$

a partir de (2.12) a última linha pode ser escrita como

$$\sum_{k=0}^{\infty} P(X_{n-1} = k)[\phi(z)]^k,$$

que dá

$$\phi_n(z) = \phi_{n-1}[\phi(z)], \qquad n > 1. \tag{2.13}$$

Segundo Bhat e Miller (2002) um resultado semelhante a (2.13) pode ser derivado pelo condicionamento do tamanho da primeira geração, por oposição a $(n-1)$-ésima população. A relação resultante é

$$\phi_n(z) = \phi[\phi_{n-1}(z)], \qquad n > 1 \tag{2.14}$$

Estendendo os argumentos que levam a (2.13) e (2.14), pode-se derivar o resultado geral

$$\phi_r(z) = \phi_{n-r}[\phi_r(z)], \qquad 1 \leq r < n \tag{2.15}$$

De acordo com Bhat e Miller (2002) as relações da função geradora (2.13) - (2.15) são úteis na obtenção de solução iterativa.

Com o objetivo de determinar a probabilidade de extinção da população final, definirmos $z = 0$ em $\phi_n(z)$. Temos

$$\phi_n(0) = P[X_n = 0|X_0 = 1]. \tag{2.16}$$

Que é de fato a probabilidade de que a extinção ocorre na n-ésima geração ou antes. A partir de (2.14) também temos

$$\phi_n(0) = \phi(\phi_{n-1}(0)). \tag{2.17}$$

Como $n \to \infty$, seja $\zeta = \lim_{n\to\infty} \phi_n(0)$. Tomando limites em (2.17), desse modo ζ satisfaz a relação funcional

$$x = \phi(x). \tag{2.18}$$

Esta equação funcional é uma relação bem conhecida na teoria de processos estocásticos e na discussão analítica das suas raízes nos leva ao seguinte resultado (Karlin e Taylor (1975)):

$$\zeta = \begin{cases} 1, & \text{se } \mu \leq 1 \\ < 1, & \text{se } \mu > 1 \end{cases}$$

Onde ζ é a menor raiz positiva da equação (2.18). Logo, juntamente com a informação sobre a média e variância de X_n, podemos, portanto concluir que quando $\mu > 1$, existe uma probabilidade positiva de crescimento indefinido na população.

2.4.2 Cadeias de Markov de ordem maior que 1

No inicio da seção 2 definimos cadeias de makov de ordem maior que 1 com a seguinte propriedade. Um processo estocástico $\{X_n, n \geq 0\}$ é dito ser markoviano de ordem k se para $t_0 < t_1 ... < t_{n+1}$

$$P(X_{t_n} = i_{t_n} | X_{t_0} = i_{t_0}, X_{t_1} = i_{t_1}, ..., X_{t_{n-1}} = i_{t_{n-1}}) = P(X_{t_n} = i_{t_n} | X_{t_{n-k}} = i_{t_{n-k}}, ..., X_{t_{n-1}} = i_{t_{n-1}}),$$

para qualquer escolha $i_{t_0}, i_{t_1}, ..., i_{t_{n+1}}$ em S e qualquer n. Isto quer dizer que o estado atual da cadeia é influenciado pelas k observações mais recentes do processo. De acordo com Bhat e Miller (2002) a análise de uma cadeia de Markov de ordem maior do que 1 com grandes valores de k é bastante difícil. No entanto, para pequenos valores de k, tal como $k = 2$ ou 3, quando o espaço de estado é pequeno, cadeias de markov de ordem maior do que 1 pode ser convertida em cadeias de Markov de primeira ordem e pode ser analisado de maneira usual.

Considere uma cadeia de Markov de segunda ordem $\{X_n, n = 0, 1, 2, ...\}$ com os estados $\{1, 2\}$. Seja

$$P_{ijk} = P(X_n = k | X_{n-1} = j, X_{n-2} = i), \qquad (i, j, k = 1, 2).$$

Probabilidades de transição possíveis em um passo são:

$$P_{111}\ P_{112}\ P_{121}\ P_{122}\ P_{211}\ P_{212}\ P_{221}\ P_{222}$$

Que pode ser organizada em uma matriz na forma

$$\begin{array}{c|cc} ij/k & 1 & 2 \\ \hline 11 & P_{111} & P_{112} \\ 12 & P_{121} & P_{122} \\ 21 & P_{211} & P_{212} \\ 22 & P_{221} & P_{222} \end{array} \tag{2.19}$$

Portanto, para uma representação completa dos estados do sistema, em qualquer época, precisamos agora de informação sobre as duas épocas anteriores. Uma vez que a representação matricial (2.19) não é conveniente para operações matriciais podemos representa-lo sob a forma de uma matriz 4×4.

$$\begin{array}{c|cccc} ij/jk & 11 & 12 & 21 & 22 \\ \hline 11 & P_{111} & P_{112} & 0 & 0 \\ 12 & 0 & 0 & P_{121} & P_{122} \\ 21 & P_{211} & P_{212} & 0 & 0 \\ 22 & 0 & 0 & P_{221} & P_{222} \end{array} \tag{2.20}$$

A matriz de probabilidades de transição (2.20) é a de uma cadeia de Markov de primeira ordem cujos estados são os compostos iniciais $(11, 12, 21, 22)$, e a sua análise segue pela maneira usual.

Logo, esta técnica pode se estendida para cadeias de Markov de ordem maior que 1. No entanto, deve notar-se que o número de elementos de probabilidade (que pode ser diferente de zero) na matriz de probabilidades de transição da cadeia de Markov de ordem k vai ser de $2^{(k+1)}$. Quando o número de estados é S, o número correspondente é $S^{(k+1)}$, organizado em numa matriz de tamanho $S^k \times S^k$.

2.4.3 Cadeias de Markov inversa

Seja $\{X_n, n \geq 0\}$ uma cadeia de Markov com probabilidades de transição definida como

$$P_{ij}^{(n)} = P[X_n = j | X_0 = i].$$

Com uma matriz de probabilidade de transição a um-passo P cujos elementos são denotados $P_{ij} = (i, j = 1, 2, \ldots, S)$, e um vetor de probabilidade invariante $\pi = (\pi_1, \pi_2, \ldots, \pi_S)$. Agora, considere a probabilidade

$$R_{ij}^n = P[X_0 = j | X_n = i].$$

Usando argumentos de probabilidade condicionais, podemos escrever

$$P[X_0 = j | X_n = i] P[X_n = i] = P[X_n = i | X_0 = j] P[X_0 = j].$$

Partindo do princípio de que o processo seja em estado estacionário, tem-se

$$\begin{aligned} R_{ij}^{(n)} \pi_i &= P_{ji}^{(n)} \pi_j \\ R_{ij}^{(n)} &= \frac{\pi_j}{\pi_i} P_{ji}^{(n)}. \end{aligned}$$

Em particular, para $n = 1$,

$$R_{ij} = \frac{\pi_j}{\pi_i} P_{ji}.$$

Claramente, $0 \leq R_{ij} \leq 1$ e $\displaystyle\sum_j R_{ij} = \frac{\sum_j \pi_j P_{ji}}{\pi_i} = \frac{\pi_i}{\pi_i} = 1.$

Assim, a matriz $\mathbf{R}$, cujo elementos são R_{ij}, é uma matriz estocástica. Portanto, a matriz $\mathbf{R}$ representa o inverso do processo original. Logo, a matriz $\mathbf{R}$ fornecer informações para o seu comportamento passado, sabendo seu estado atual. Em investigações geológicas, onde o objetivo é compreender o passado sabendo o estado atual, cadeias de Markov inversas podem ser modelos úteis.

Importantes propriedades da cadeia de Markov inversa são obtidas a partir das propriedades da cadeia original:

1.

$$\begin{aligned} \sum_i \pi_i R_{ij} &= \sum_i \pi_i \frac{\pi_j P_{ji}}{\pi_i} \\ &= \sum_i \pi_j P_{ji} \; = \; \pi_j. \end{aligned}$$

Mostrando que o vetor probabilidade $\pi = \{\pi_1, \pi_2, \ldots, \pi_S\}$ é também o vetor de probabilidades de estado invariante para a cadeia inversa.
2. também

$$\sum_n R_{ij}^{(n)} = \sum_n \frac{\pi_j P_{ji}^{(n)}}{\pi_i}.$$

Os termos em ambos os lados da equação convergir ou divergir em conjunto, portanto, as propriedades de recorrência e transição são idênticos em ambos os processos. Um caso especial de cadeia inversa é obtido quando

$$R_{ij} = P_{ij}.$$

Isto é,

$$P[X_0 = j|X_1 = i] = P[X_1 = j|X_0 = i].$$

Este processo é chamado de cadeia de Markov reversível, e as probabilidades de transição são simétricas no tempo. A cadeia de Markov é reversível se suas probabilidades de transição em um passo exibem a seguinte propriedade

$$P_{ij} = \frac{\pi_j}{\pi_i} P_{ji}.$$

2.5 Inferência estatística para cadeias de Markov

Segundo Bhat e Miller (2002), nas últimas décadas, a área de inferência estatística para processos estocásticos tem visto um crescimento explosivo. Na maioria das vezes o tratamento de modelos estocásticos é de um ponto de vista probabilístico. Porém a estimativa e testes para os parâmetros desses modelos também são importantes. Por esta razão nesta seção discutimos a inferência para cadeias de Markov de tempo discreto.

2.5.1 Estimativa dos elementos de uma matriz de probabilidade de transição

Consideremos a estimativa que se refere à cadeias de Markov finitas com matrizes de probabilidades de transição estacionárias. Além disso, discutimos a estimativa brevemente para as cadeias tendo um espaço de estado infinito contável. Aqui apresenta-se apenas resultados simples para as estimativas de máxima verossimilhança dos elementos da matriz de probabilidade de transição.

Para fins de inferência, uma cadeia de Markov é frequentemente observada em uma de duas maneiras:

1. Uma observação (longa) de uma realização da cadeia.

2. Observações (mais curta) de várias realizações da mesma cadeia.

Considere observação de uma cadeia de Markov finita com S estados $(1, 2, \ldots, S)$ até n transições ter ocorrido. Seja n_{ij} o número de transições de i para j $(i, j = 1, 2, \ldots, S)$. Seja $\sum_{j=1}^{S} n_{ij} = n_i$. Estes valores podem ser representados como

$$\begin{array}{c|cccccc} & 1 & 2 & 3 & \ldots & S & \\ \hline 1 & n_{11} & n_{12} & n_{13} & \ldots & n_{1S} & n_1 \\ 2 & n_{21} & n_{22} & n_{23} & \ldots & n_{2S} & n_2 \\ 3 & n_{31} & n_{32} & n_{33} & \ldots & n_{3S} & n_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ S & n_{S1} & n_{S2} & n_{S3} & \ldots & n_{SS} & n_S \\ \hline & & & & & & n \\ \hline \end{array} \tag{2.21}$$

Seja a matriz de probabilidade de transição estacionária da cadeia de Markov $\mathbf{P}$, dada por

$$\mathbf{P} = \begin{bmatrix} P_{11} & P_{12} & \ldots & P_{1S} \\ P_{21} & P_{22} & \ldots & P_{2S} \\ \vdots & \vdots & & \vdots \\ P_{S1} & P_{S2} & \ldots & P_{SS} \end{bmatrix}.$$

Estamos interessados nas estimativas dos elementos P_{ij}; denotamos suas estimativas por $\hat{P}_{ij}$ $(i, j = 1, 2, \ldots, S)$.

Para um dado estado inicial i e um certo número de ensaios n_i, a amostra de transição conta $(n_{i1}, n_{i2}, \ldots, n_{iS})$ pode ser considerada como uma amostra de tamanho n_i a partir de uma distribuição multinomial com probabilidades $(P_{i1}, P_{i2}, \ldots, P_{iS})$, tal que $\sum_{j=1}^{S} P_{ij} = 1$. A probabilidade deste resultado pode, portanto, ser dado como

$$\frac{n!}{n_{i1}! n_{i2}! \ldots n_{iS}!} P_{i1}^{n_{i1}} P_{i2}^{n_{i2}} \ldots P_{iS}^{n_{iS}}.$$

De tal modo que $\sum_{j=1}^{S} n_{ij} = n_i$ e $\sum_{j=1}^{S} P_{ij} = 1$.

De acordo com Bhat e Miller (2002), estendendo este argumento para o estado inicial i $(1, 2, 3, \ldots, S)$, quando a repartição do número total de ensaios n em $(n_1, n_2, ..., n_S)$ é dado, a probabilidade da realização de transição conta como em (2.21) será dada por

$$\prod_{i=1}^{S} \frac{n!}{n_{i1}! n_{i2}! \ldots n_{iS}!} P_{i1}^{n_{i1}} P_{i2}^{n_{i2}} \ldots P_{iS}^{n_{iS}}. \tag{2.22}$$

No entanto, os valores de $n_i (i = 1, 2, ..., S)$ não são constante. Ou seja, as somas de linha em (2.21) são variáveis aleatórias. Assim $f(\mathbf{p})$ é o produto de (2.22) e outro fator que representa a contribuição dessas variáveis aleatórias para a função de verossimilhança.

Este fator adicional é independente de P_{ij}. Assim, denotamos este fator adicional como A. Isso leva à função de verossimilhança $f(\mathbf{p})$ e seu natural logaritmo $L(\mathbf{p})$ a ser expressa como

$$f(\mathbf{P}) = A\prod_{i=1}^{S}\frac{n!}{n_{i1}!n_{i2}!\dots n_{iS}!}P_{i1}^{n_{i1}}P_{i2}^{n_{i2}}\dots P_{iS}^{n_{iS}}. \tag{2.23}$$

Aplicando ln em $f(\mathbf{P})$ temos

$$\begin{aligned} L(\mathbf{P}) &= lnA + ln\left[\prod_{i=1}^{S}\frac{n!}{n_{i1!}n_{i2!}\dots n_{is!}}P_{i1}^{n_{i1}}P_{i2}^{n_{i2}}\dots P_{is}^{n_{is}}\right] \\ &= lnA + \sum_{i=1}^{S}\left[ln\left(\frac{n!}{n_{i1!}n_{i2!}\dots n_{is!}}\right)\right] + \sum_{i=1}^{S}\sum_{j=1}^{S}lnP_{ij}^{n_{ij}} \\ &= lnB + \sum_{i=1}^{S}\sum_{j=1}^{S}n_{ij}lnP_{ij}. \end{aligned} \tag{2.24}$$

Em que lnB contém todos os termos independentes dos Pij'S.

Além disso, neste caso temos um problema de maximização com restrições. Sobre a condição $\sum_{j=1}^{S}P_{ij} = 1(i = 1, 2, 3, \dots)$ incorporando está condição em (2.24) podemos escrever

$$L(\mathbf{P}) = lnB + \sum_{i=1}^{S}\sum_{j=1}^{S-1}n_{ij}P_{ij} + \sum_{i=1}^{S-1}n_{is}ln(1 - P_{i1} - P_{i2} - \dots - P_{i,S-1}).$$

Observa-se que a partir da estrutura da função $L(P)$, as estimativas podem ser obtidas separadamente para os valores de S de $i = 1, 2, 3, \dots S$. Para um valor específico de i temos

$$L_i(\mathbf{P}) = lnB + \sum_{j=1}^{S-1}n_{ij}lnP_{ij} + n_{is}ln(1 - P_{i1} - P_{i2} - \dots - P_{i,S-1}). \tag{2.25}$$

De acordo com Bhat e Miller (2002), para um valor fixo de i, a Equação (2.25) pode ser diferenciada em relação à P_{ij} para $J = 1, 2, \dots, S-1$. Definir as resultantes $S-1$ derivadas iguais a zero leva a um sistema de equações que, quando resolvido para um particular P_{ij} resulta em

$$\hat{P}_{ij} = \frac{n_{ij}}{n_i} \qquad i, j = 1, 2, \dots, S. \tag{2.26}$$

Portanto, observa-se que a estimativa dos elementos de uma matriz de probabilidade de transição $(\hat{P}_{ij})$ é obtido como a razão entre o número de transições do estado i para o estado j e o número de vezes em que a cadeia esteve no estado i, na realização da cadeia.

2.6 Teste de hipótese para cadeias de Markov

2.6.1 Teste de hipótese para uma matriz de probabilidade de transição

Para verificar se uma observação realizada veio de uma cadeia de Markov com uma dada matriz de probabilidade de transição, considera-se a hipótese nula,

$$H_0 : \mathbf{P} = \mathbf{P}^0.$$

De acordo com Bhat e Miller (2002), para n suficientemente grande n_{ij} são assintoticamente normalmente distribuídos e que a estatística $n_i^{1/2}P_{ij}(1-P_{ij})$ tem uma distribuição normal assintótica com média 0 e variância $P_{ij}(1-P_{ij})$. Baseado nessas informações um teste estatístico equivalente com a bondade de ajustar a estatística pode ser obtido. Essa estatística pode ser escrita como

$$\sum_{i=1}^{S}\sum_{j=1}^{S}\frac{n_i(\hat{P}_{ij}-P^0_{ij})^2}{P^0_{ij}}. \tag{2.27}$$

Onde a soma de (2.27) é tomada somente sobre (i, j) para os quais $P_{ij} > 0$. Sob a hipótese nula H_0 a estatística em (2.27) tem uma distribuição χ^2 com $S(S-1)-d$ graus de liberdade, em que d é o número de zeros em $\mathbf{P}^0$. Logo se a hipótese nula H_0 não for rejeitada pode-se concluir, a um nível de significância α, que a realização veio de uma cadeia de Markov. Por outro lado, quando a hipótese nula H_0 é rejeitada pode-se concluir, a um nível de significância α que a realização não veio de uma cadeia de Markov.

Segundo Bhat e Miller (2002), um teste estatístico assintoticamente equivalente é obtido pelo critério da razão de verossimilhança com base no lema de Neyman-Pearson. O critério da razão de verossimilhança para H_0, pode ser obtido como,

$$\Lambda = \frac{f(\mathbf{P}^0)}{f(\hat{\mathbf{P}})}.$$

Onde $f(\hat{\mathbf{P}})$ é o valor maximizado da função de verossimilhança (2.23) obtidos por substituição das estimativas em (2.26). Da teoria estatística sabe-se que, quando H_0 é verdadeiro, $-2\ln\Lambda$ tem uma distribuição χ^2 com $S(S-1)$ graus de liberdade. Neste caso, temos

$$\begin{aligned} -2\ln\Lambda &= 2[L(\hat{\mathbf{P}}) - L(\mathbf{P}^0)] \\ &= 2\sum_{i=1}^{S}\sum_{j=1}^{S} n_{ij}\ln\frac{n_{ij}}{n_i P^0_{ij}}. \end{aligned} \tag{2.28}$$

2.6.2 Teste para estacionariedade da matriz de transição de probabilidade

Agora vamos usar uma estatística de teste da razão de verossimilhança para testar a hipótese nula de que uma matriz de probabilidade de transição é estacionaria. Seja P^t_{ij}

ser as probabilidades de transição em um passo de um processo dependente do tempo $X(t)$ tal que

$$P_{ij}^{t} = P[X(t+1) = j | X(t) = i].$$

Suponhamos que S indica o número de realizações observadas, cada uma de tamanho T. Observa-se que para obter uma amostra adequada para este teste devemos observar mais de uma realização do processo. Portanto o plano de amostragem para este teste é observar várias realizações da mesma cadeia ao longo de um período mais curto de tempo.

Seja n_{ij} o número de transições $i \to j$ durante a t-ésima transição de uma realização. Isto é, S realizações, n_{ij} é o número de transições $i \to j$ na t-ésima etapa de uma realização $(t = 1, 2, \ldots, T)$. Para qualquer estado particular dado i, as n_{ij} contagens de transição pode ser representado como

$$\begin{array}{c|ccccc}
t/j & 1 & 2 & 3 & \ldots & S \\
\hline
1 & n_{i1}^{1} & n_{i2}^{1} & n_{i3}^{1} & \ldots & n_{i,S}^{1} \\
2 & n_{i1}^{2} & n_{i2}^{2} & n_{i3}^{2} & \ldots & n_{i,S}^{2} \\
3 & n_{i1}^{3} & n_{i2}^{3} & n_{i3}^{3} & \ldots & n_{i,S}^{3} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
T & n_{i1}^{T} & n_{i2}^{T} & n_{i3}^{T} & \ldots & n_{i,T}^{T}
\end{array} \tag{2.29}$$

Argumentando como antes, as estimativas de máxima verossimilhança de P_{ij}^{t} pode ser obtida:

$$\hat{P}_{ij}^{t} = \frac{n_{ij}^{t}}{n_{i}^{t-1}}.$$

Onde $n_i^{t-1} = \sum_{j=1}^{S} n_{ij}^{t}$. Aqui, n_i^{t-1} representa o número de realizações para o qual o processo foi no estado i no momento $t-1$.

Para a hipótese $H_0 : P_{ij}^{t} = P_{ij}(t = 1, 2, \ldots, T)$, a função de verossimilhança maximizada é dada por $f(\hat{\mathbf{P}}^{t})$, e assim, a estatística de teste da razão de verossimilhança é dada por,

$$\Lambda = \frac{f(\mathbf{P})}{f(\hat{\mathbf{P}}^{t})}.$$

Sobre H_0, $-2\ln\Lambda$ possui uma distribuição χ^2 com $(T-1)[S(S-1)]$ graus de liberdade. Nesse caso

$$\begin{aligned}
-2\ln\Lambda &= 2[L(\hat{\mathbf{P}}^{t}) - L(\mathbf{P})] \\
&= 2\sum_{t=1}^{T}\sum_{i=1}^{S}\sum_{j=1}^{S} n_{ij}^{t} \ln \frac{n_{ij}^{t}}{n_i^{t-1} P_{ij}}.
\end{aligned} \tag{2.30}$$

Assim (2.30) é usado para testar a hipótese H_0 de que uma matriz de probabilidade de transição é estacionária, contra a hipótese H_1 de que a matriz de probabilidade de transição não é estacionária.

2.6.3 Teste para independência versus a dependência de primeira ordem

O interesse agora é um teste para a hipótese nula de que observações coletadas são independentes, contra a alternativa que o processo observado é uma cadeia de Markov de primeira ordem. Considere,

$$H_0 : \mathbf{P} = \mathbf{P}^0.$$

Onde P^0 tem linhas idênticas sob a hipótese de independência. Mais especificamente, se P consistem de S linhas idênticas $\pi = (\pi_1, \pi_2, \cdots, \pi_S)$. Quando estas probabilidades não são conhecidas, as suas estimativas de máxima verossimilhança podem ser determinada como segue.

Seja $n_{.j} = \sum_{i=1}^{S} n_{ij}$. A função de log verossimilhança correspondente a (2.24) pode agora ser escrita como

$$L(\pi) = \ln B + \sum_{j=1}^{S} n_{.j} \ln \pi_j. \tag{2.31}$$

Usando (2.31) conduz a estimativa de máxima verossimilhança

$$\hat{\pi}_j = \frac{n_{.j}}{n}, \qquad j = 1, 2, \cdots, S.$$

A estatística χ^2 para o teste de independência contra a dependência de primeira ordem assume a forma

$$\sum_{i=1}^{S} \sum_{j=1}^{S} \frac{(n_{ij} - n_i n_{.j}/n)^2}{n_i n_{.j}/n}.$$

Com $(S-1)^2$ graus de liberdade.

A estatística da razão de verossimilhança correspondente a (2.28) tomar a forma

$$2 \sum_{i=1}^{S} \sum_{j=1}^{S} n_{ij} \ln \left(\frac{n_{ij}}{n_i n_{.j}/n} \right). \tag{2.32}$$

Com $(S-1)^2$ graus de liberdade.

Assim (2.32) é usado para testar a hipótese nula de que observações coletadas são independentes, contra a hipótese alternativa de que o processo observado é uma cadeia de Markov de primeira ordem.

2.6.4 Teste para a ordem da cadeia de Markov

O teste para a independência pode ser generalizado para permitir um teste para a ordem (possivelmente maior do que 1) de uma cadeia de Markov. Apresenta-se esse procedimento usando uma cadeia de Markov de segunda ordem com probabilidades de transição estacionárias. Seja

$$P_{ijk} = P(X_n = k | X_{n-1} = j, X_{n-2} = i).$$

Sendo n_{ijk} a contagem de transição correspondente. Além disso, seja $n^*_{ij} = \sum_k n_{ijk}$. Agora, procedendo como antes, a estimativa de P_{ijk} é obtida como

$$\hat{P}_{ijk} = \frac{n_{ijk}}{n^*_{ij}}.$$

A hipótese nula de que a cadeia de Markov é uma cadeia de Markov de primeira ordem contra à hipótese alternativa que a cadeia de Markov é uma cadeia de Markov de segunda ordem pode ser dada como

$$H_0 : P_{ijk} = P_{jk}, \qquad i, j, k = 1, 2, \cdots, S.$$

A estatística χ^2 para testar a hipótese H_0 acima pode ser escrita como

$$\frac{\sum_{i=1}^{S}\sum_{j=1}^{S}\sum_{k=1}^{S} n^*_{ij}(\hat{P}_{ijk} - \hat{P}_{jk})^2}{\hat{P}_{jk}}.$$

Que, tem uma distribuição χ^2 com $S(S-1)^2$ graus de liberdade.

Quando tal procedimento é estendido a um teste sobre uma cadeia de Markov de ordem r (a hipótese nula é que a cadeia é de ordem $r-1$ contra a hipótese alternativa que é de ordem r) a correspondente estatística χ^2 tem $S^{r-1}(S-1)^2$ graus de liberdade.

2.6.5 Inferência estatística para processos de ramificação

Na seção (2.4) discutimos sobre o processo de ramificação de tempo discreto. Agora, nesta seção, damos alguns resultados relativos à estimativa para esses processos. Observa-se que aqui, continuamos a usar a terminologia biológica de organismo e descendentes quando se refere ao processo.

Mais uma vez, supor que até o fim de sua vida útil um organismo produz um número aleatório de descendentes S com uma distribuição de probabilidade.

$$P(S = j) = p_j, \qquad j = 0, 1, 2, \cdots \tag{2.33}$$

Com $p_j \geq 0$ e $\sum p_j = 1$. Se $S_1, S_2, \cdots$ são variáveis aleatórias independentes e identicamente distribuídas como (2.33) e denotar X_n como sendo o tamanho da população na n-ésima geração, então

$$X_{n+1} = \sum_{k=1}^{X_n} S_k.$$

Aqui, S_k é o número de descendentes produzidos pelo k-ésimo membro da n-ésima geração. Seja $E[S_k] = \mu$ e $V[S_k] = \sigma^2,\ k = 1, 2, ...$ denotar a descendência média e a variância, respectivamente. Portanto, aqui vamos discutir a estimativa das quantidades p_j, μ e σ^2.

Primeiro consideremos que todos os tamanhos individuais de organismos nas primeiras n gerações têm sido observados. Em seguida, pode-se mostrar que o estimador de máxima verossimilhança de p_j é a frequência relativa

$$\hat{P}_j = \frac{n_j}{(X_0 + X_1 + \cdots + X_{n-1})}, \qquad j = 0, 1, 2, ... \tag{2.34}$$

Onde n_j é o número de vezes que j descendentes são produzidos.

Uma vez que $\mu = \sum_j \hat{P}_j$ podemos usar a Equação (2.34) para sugerir que $\hat{\mu} = \sum_j \hat{P}_j$ é um estimador razoável para μ, a média de descendentes. No entanto

$$\hat{\mu} = \sum_j \hat{P}_j = \frac{(X_1 + X_2 + \cdots + X_n)}{(X_0 + X_1 + \cdots + X_{n-1})},$$

logo, $\hat{\mu}$ dado acima é o estimador de máxima verossimilhança para μ quando apenas os tamanhos das gerações $X_0, X_1, \cdots, X_n$ são observados.

A teoria assintótica relativa a $\hat{\mu}$ pode ser explorado sob diferentes estruturas. Pode-se considerar um aumento do número de ancestrais ou um aumento do número de gerações. Em ambas as estruturas resultados de normalidade assintótica para $\hat{\mu}$ são disponíveis. Esses resultados podem auxiliar na capacidade de realizar testes de hipóteses para μ com grandes amostras.

A fim de implementar estes resultados para o propósito de testes de hipóteses ou intervalos de confiança para μ, precisaríamos de um estimador consistente para σ^2. Um tal estimador é,

$$\hat{\sigma}^2 = n^{-1} \sum_{k=0}^{n-1} X_k[(X_{k+1}/X_k) - \hat{\mu}]^2.$$

2.7 Processos de Markov simples

No inicio da seção 2 estudamos o processo de Markov com espaços de estados S discreto e o conjunto de índices T enumerável (processo a tempo discreto). Vimos que a matriz de probabilidades de transição determina o comportamento de uma cadeia de Markov de tempo discreto e investigamos suas propriedades e aplicações. Aqui abordaremos os processos estocásticos com parâmetro de tempo contínuo e espaços de estados discretos ou seja, $T = [0, \infty)$ e $S = \{0, 1, \dots\}$. Muitos fenômenos são modelados dessa forma tais como: a ocorrência de uma falha de um hardware ou software em um computador. A chegada de uma mensagem em um centro de computação operando on-line.

2.7.1 Processos de Markov: Propriedades gerais

Seja $\{X(t), t \in T\}$ um processo de Markov homogêneo no tempo com espaços de estados discretos designados por S. Então definimos,

$$P_{ij}(t) = P_{ij}(0, t) = P[X(t) = j | X(0) = i], \qquad i, j \in S$$

As seguintes propriedades de $P_{ij}(t)$ seguem da definição:

1. $P_{ij}(t) \geq 0,\, t > 0;$
2. $\sum\limits_{j \in S} P_{ij}(t) = 1,\, t > 0;$
3. $P_{ij}(t+s) = \sum\limits_{k \in S} P_{ik}(t)P_{kj}(s),\, t, s > 0;$
4. $P_{ij}(t)$ é contínuo;
5. $\lim\limits_{t \to 0} P_{ij}(t) = \begin{cases} 1, & \text{se } i = j \\ 0, & \text{se } \; i \neq j. \end{cases}$

No entanto, quando o parâmetro T é contínuo em um processo de Markov, análogo à matriz de probabilidade de transição, definimos uma matriz geradora (também conhecida como uma matriz de taxa de transição infinitesimal) que define completamente um processo de Markov. Os geradores (taxas de transição infinitesimais) são determinados a seguir.

Usando a série de Taylor, $P_{ij}(t, t + \Delta t)$ pode ser expandida nas proximidades de t do seguinte modo

$$P_{ij}(t, t+\Delta t) = P_{ij}(t) + \Delta t P'_{ij}(t) + \frac{\Delta t^2}{2!} P''_{ij}(t) + \ldots$$

Como o processo é homogêneo no tempo, estabelecemos $t = 0$; Usando as propriedades dadas acima, podemos escrever

$$P_{ij}(\Delta t) = \Delta t P'_{ij}(0) + \frac{\Delta t^2}{2!} P''_{ij}(0) + \ldots, \qquad j \neq i$$

$$P_{ii}(\Delta t) = 1 + \Delta t P'_{ii}(0) + \frac{\Delta t^2}{2!} P''_{ii}(0) + \ldots$$

Reescrevendo essas equações, tem-se

$$\frac{P_{ij}(\Delta t)}{\Delta t} = P'_{ij}(0) + \frac{o(\Delta t)}{\Delta t},$$

$$\frac{P_{ii}(\Delta t) - 1}{\Delta t} = P'_{ii}(0) + \frac{o(\Delta t)}{\Delta t}\cdot$$

Onde $o(\Delta t)$ é tal que,

$$\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0\cdot$$

Dai,

$$\lim_{\Delta t \to 0} \frac{P_{ij}(\Delta t)}{\Delta t} = P'_{ij}(0) = \lambda_{ij},$$

$$\lim_{\Delta t \to 0} \frac{P_{ii}(\Delta t) - 1}{\Delta t} = P'_{ii}(0) = -\lambda_{ii}.$$

Observa-se que a taxa de transição λ_{ij} é finita e λ_{ii} pode ser infinita. A propriedade (2) acima mostra que

$$\sum_{j \neq i} \lambda_{ij} - \lambda_{ii} = 0 \cdot \tag{2.35}$$

Assim, podemos considerar a equação (2.35) como sendo análoga à propriedade que em uma cadeia de Markov $\sum\limits_{j \in S} P_{ij} = 1$. Em vez de uma matriz de probabilidade de transição, temos

$$\mathbf{A} = \begin{bmatrix} -\lambda_{00} & \lambda_{01} & \lambda_{02} & \dots \\ \lambda_{10} & -\lambda_{11} & \lambda_{12} & \dots \\ \lambda_{20} & \lambda_{21} & -\lambda_{22} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

que é conhecido como a matriz geradora. Considere a equação de Chapman-kolmogorov Propriedade (3) acima, para o processo de Markov, definimos $s = \Delta t$, então temos

$$P_{ij}(t + \Delta t) = \sum_{k \in S} P_{ik}(t) P_{kj}(\Delta t)$$

Subtraindo $P_{ij}(t)$ em ambos os lados e dividindo por Δt, calculando a derivada de $P(t)$. Usando a definição de $\mathbf{A}$. Procedendo tem-se

$$\begin{aligned} P'_{ij}(t) &= \lim_{\Delta t \to 0} \frac{P_{ij}(t + \Delta t) - P_{ij}(t)}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} (P\{X(t + \Delta t) = j | X(0) = i\} - P\{X(t) = j | X(0) = i\}) \\ &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} (\sum_{k \in S} P\{X(t + \Delta t) = j | X(t) = k, X(0) = i\} P\{X(t) = k | X(0) = i\} \\ &- P\{X(t) = j | X(0) = i\}) \end{aligned}$$

No entanto,

$$\begin{aligned} &\sum_{k \in S} P\{X(t + \Delta t) = j | X(t) = k, X(0) = i\} P\{X(t) = k | X(0) = i\} \\ &= P\{X(t + \Delta t) = j | X(t) = j, X(0) = i\} P\{X(t) = j | X(0) = i\} \\ &+ \sum_{k \neq j} P\{X(t + \Delta t) = j | X(t) = k, X(0) = i\} P\{X(t) = k | X(0) = i\} \\ &= (1 - \lambda_{jj}(\Delta t) P_{ij}(t) + \sum_{k \neq j} \lambda_{kj}(\Delta t) P_{ik}(t) + o(\Delta t)), \end{aligned}$$

E assim,

$$\begin{aligned} P'_{ij}(t) &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left((1 - \lambda_{jj}(\Delta t) - 1) P_{ij}(t) + \sum_{k \neq j} \lambda_{kj} P_{ik}(t)(\Delta t) + o(\Delta t) \right) \\ &= -\lambda_{jj} P_{ij}(t) + \sum_{k \neq j} \lambda_{kj} P_{ik}(t). \end{aligned}$$

Logo,

$$P'_{ij}(t) \;= -\lambda_{jj}P_{ij}(t) + \sum_{k\neq j}\lambda_{kj}P_{ik}(t).$$

Esta equação é conhecida como equação de kolmogorov para frente (do inglês, *forward kolmogorov equations*), com i fixo.

Da mesma forma, defina $t = \Delta s$ na propriedade (3) acima, e escreva

$$P_{ij}(\Delta s + s) = \sum_{k\in S} P_{ik}(\Delta s)P_{kj}(s).$$

Subtraindo, $P_{ij}(s)$, dividindo ambos os lados por Δs, e com $\Delta s \to 0$, obtemos (usando t no lugar de s)

$$P'_{ij}(t) = -\lambda_{ii}P_{ij}(t) + \sum_{k\neq i}\lambda_{ik}P_{kj}(t).$$

Esta equação é conhecida como equação de Kolmogorov para trás (do inglês *backward kolmogorov equation*), com j fixo.

Seja $\mathbf{P}(t)$ e $\mathbf{P}'(t)$ representando matrizes cujos elementos são $P_{ij}(t),\ i,j,\ \in S$ e $P'_{ij}(t),\ i,j \in S$, respectivamente. Em notação matricial, usando a matriz geradora $\mathbf{A}$, as equações de kolmogorov para frente e para trás podem ser representadas, respectivamente, como

$$\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{A}$$

e

$$\mathbf{P}'(t) = \mathbf{A}\mathbf{P}(t).$$

As condições iniciais para ambos os conjuntos de equações são

$$\mathbf{P}(0) = \mathbf{I}.$$

3 Aplicação

Foi visto que as cadeias de Markov de tempo discreto tem larga aplicabilidade, inclusive na área da economia. Por isto, nesta seção é feita uma aplicação nessa área. O banco de dados é relacionado às altas e baixas da cotação do dólar no Brasil no começo deste ano. Câmbio é a operação de troca entre moedas de diferentes países. O câmbio flutuante (que é o usado no Brasil) representa o preço de uma moeda expressa em outra unidade monetária, isto é, quantas moedas de um país são necessárias para obter a moeda de um outro país. A sua cotação é diária, e é determinada pelos mercados que são afetados e por fatores como a procura e a oferta de moeda. Foi usado o $US\$$ (dólar) como referência. Os conceitos, no entanto, valem para todas as moedas.

Os dados foram obtidos do site https://economia.uol.com.br/cotacoes/cambio/dolar-comercial-estados-unidos/?historico. Os dados são sobre o histórico do cambio, o período consultado foi de 02/01/2017 a 03/07/2017. O banco de dados é constituído de um total de 123 observações dos quais representam a cotação do dólar.

Seja $\{X_n; n \geq 0\}$ uma cadeia de Markov onde X_n representa a cotação do dólar no n-ésimo dia. Com espaço de estados $\{0, 1\}$, vamos considerar,

$$X_n = \begin{cases} 0, & \text{se no } n\text{-ésimo dia a cotação do dólar baixou} \\ 1, & \text{se no } n\text{-ésimo dia a cotação do dólar teve alta} \end{cases}$$

Observação: Como X_n representa a cotação do dólar no n-ésimo dia, o X_0 é o ponto de referencia isso é, a observação anterior ao dia 02/01/2017 onde se iniciou o período consultado, a observação X_0 não esta contabilizada na amostra.

A cadeia visitou o estado zero 59 vezes e o estado um 64 vezes, nas 123 observações. Verifica-se também que,

$$n_{00} = 22; \qquad n_{01} = 37; \qquad n_{10} = 37; \qquad n_{11} = 26,$$

onde $n_{00} = 22$ foram as vezes que a cadeia estava no estado zero e permaneceu no estado zero, $n_{01} = 37$ foram 37 transições que a cadeia fez do estado zero para o estado um, $n_{10} = 37$ foram 37 transições que a cadeia fez do estado um para o estado zero, $n_{11} = 26$ foram as vezes que a cadeia estava no estado um e permaneceu no estado um. No último estado visitado na realização da cadeia tivemos n_i visitas e $n_i - 1$ transições neste caso o estado um foi o último estado visitado então o denominador de $\hat{P}_{ij}$ vai ser igual a $n_i - 1$, quando o tamanho da amostra e grande não há diferença se usarmos denominador n_i ou $n_i - 1$. Logo temos que

$$\hat{P}_{00} = \frac{22}{59} = 0,37; \qquad \hat{P}_{01} = \frac{37}{59} = 0,63; \qquad \hat{P}_{10} = \frac{37}{63} = 0,59; \qquad \hat{P}_{11} = \frac{26}{63} = 0,41.$$

Então a matriz de probabilidade de transição estimada $\hat{\mathbf{P}}$ é dada por.

$$\hat{\mathbf{P}} = \begin{bmatrix} 0,37 & 0,63 \\ 0,59 & 0,41 \end{bmatrix}.$$

Observa-se que a cadeia é irredutível e aperiódica e recorrente positiva, logo a cadeia é ergódica, portanto existe a distribuição invariante e ela é única. Vamos calcular a distribuição invariante da cadeia resolvendo o sistema

$$[\pi_0, \pi_1] \cdot \begin{bmatrix} 0,37 & 0,63 \\ 0,59 & 0,41 \end{bmatrix} = [\pi_0, \pi_1]. \tag{3.1}$$

Calculando as operações matriciais acima e com a condição que $\pi_0 + \pi_1 = 1$ obtemos o seguinte sistema de equações:

$$\begin{cases} -0,63\pi_0 + 0,59\pi_1 & = 0 \\ \pi_0 + \pi_1 & = 1 \end{cases}$$

Resolvendo o sistema, obtemos que $\pi_0 = 0,48$ e $\pi_1 = 0,52$, além do mais temos $\pi_0 = \pi$, isto é, a cadeia começa com uma distribuição inicial que é invariante. Isto significa que independentemente da cotação do dólar ser inicialmente alta ou baixa, há uma probabilidade equivalente a 52% da cotação do dólar ter alta, ou seja, uma valorização do dólar e uma desvalorização do real neste período. Consequentemente caem as importações e aumentam as exportações. Por outro lado conclui-se também que independentemente da cotação do dólar ser inicialmente alta ou baixa, há uma probabilidade equivalente a 48% da cotação do dólar ter baixa, ou seja o dólar se desvaloriza então podemos dizer que o real se valoriza. Consequentemente as importações aumentam e as exportações diminuem. O resultado desses dois movimentos contrários, e conduzir o câmbio ao equilíbrio.

Agora vamos usar um teste χ^2 como uma maneira de estabelecer a validade do modelo de Markov. Este teste é o teste de independência versus a dependência de primeira ordem. Considere a hipótese,

$$H_0 : \mathbf{P} = \mathbf{P}^0.$$

Onde sobre a hipótese H_0 as observações podem ser atribuídas a um processo de ensaios independentes, neste caso, o sistema é considerado “desmemoriado” (do inglês *memoryless*), o que corresponde a uma cadeia de Markov de ordem zero. A hipótese H_1 é que o processo observado é uma cadeia de Markov de primeira ordem, ou seja, a memória do sistema somente enxerga o dia anterior. A estatística χ^2 para o teste de independência contra a dependência de primeira ordem assume a forma:

$$\chi^2_{\text{Calc}} = \sum_{i=1}^{S} \sum_{j=1}^{S} \frac{(n_{ij} - n_i n_{.j}/n)^2}{n_i n_{.j}/n} \qquad \sim \chi^2_{(S-1)^2}.$$

Para auxiliar nos cálculos usaremos a tabela a seguir:

		Hoje		
		0	1	n_i
Ontem	0	22	37	59
	1	37	26	63
	$n_{.j}$	59	63	122

Tabela 1 – Matriz de contagem de transições

Usando os dados da matriz de contagem de transições obtemos o resultado a seguir:

$$\chi^2_{\text{Calc}} = \frac{(22 - 59 \times 59/122)^2}{59 \times 59/122} + \frac{(37 - 59 \times 63/122)^2}{59 \times 63/122} + \frac{(37 - 63 \times 59/122)^2}{63 \times 59/122} + \frac{(26 - 63 \times 63/122)^2}{63 \times 63/122} = 5,6.$$

Regra de decisão: Rejeitamos a hipótese H_0, se o valor calculado da estatística de teste $\chi^2_{\text{Calc}} > \chi^2_{(S-1)^2}$. Valor crítico do teste:

$$\chi^2_{(S-1)^2} = \chi^2_{(2-1)^2} = \chi^2_{(1;\,5\%)} = 3,841.$$

Conclusão do teste, como $\chi^2_{\text{Calc}} = 5,6 > \chi^2_{(1;\,5\%)} = 3,841$ rejeitamos a hipótese H_0 ao nível de 5% de significância e concluímos que o processo observado é uma cadeia de Markov de primeira ordem. Portanto concluímos que o modelo de cadeia de Markov de primeira ordem será o modelo adequado para descrever o comportamento dos dados.

4 Conclusão

No que se refere ao estudo teórico da teoria de processos markovianos, pode-se dizer que foi muito importante, visto que foi possível solidificar alguns conceitos vistos superficialmente na graduação, bem como aprender coisas novas. Os cálculos desenvolvidos, requeriam um conhecimento bom de cálculo (derivadas, integrais e séries) e também álgebra e probabilidade, o que resultou num crescimento da maturidade matemática, o que é fundamental para aqueles que queiram fazer uma pós-graduação na área de Estatística.

Quanto à aplicação, foi muito gratificante aplicar a teoria vista, a dados reais. Notamos que, o modelo de cadeias de Markov de primeira ordem foi o modelo adequado para descrever o comportamento da cotação do dólar no Brasil no começo deste ano. Para a estimabilidade de uma cadeia de ordem maior que 1 é preciso de uma amostra de tamanho maior. Verificou-se que independentemente da cotação do dólar ser inicialmente alta ou baixa há uma probabilidade equivalente a 52% da cotação do dólar ter alta. Por outro lado conclui-se também que independentemente da cotação do dólar ser inicialmente alta ou baixa há uma probabilidade equivalente a 48% da cotação do dólar ter baixa neste período estudado.

Conclui-se também que a importância de se estudar a teoria de processos Markovianos discretos esta no fato que ela tem larga aplicabilidade em diversas áreas como física atômica, teoria quântica, biologia, genética, comportamento social, economia e finança. Logo esses modelos tem notável importância e amplo uso teórico e prático.

Referências

BHAT, U. N.; MILLER, G. K. *Elements of Applied Stochastic Processes*. [S.l.]: Wiley Series in Probability and Statistics, 2002. Citado 10 vezes nas páginas 11, 17, 32, 34, 36, 37, 39, 40, 41 e 42.

BRÉMAUD, P. *Markov Chains - Gibbs Fields, Monte Carlo Simulation, and Queues*. [S.l.]: Springer, 1999. Citado na página 11.

HOEL, P. G.; PORT, S. C.; STONE, C. J. *Introduction to Stochastic Process*. [S.l.]: Houghton Mifflin Company, 1972. Citado na página 11.

KARLIN, S.; TAYLOR, H. *A First Course in Stochastic Process*. [S.l.]: Academic Press, 1975. Citado 2 vezes nas páginas 11 e 36.

ROSS, S. M. *Stochastic Processes*. [S.l.]: John Wiley & Sons, 1996. Citado na página 11.

ROSS, S. M. *Introduction to Probability models*. [S.l.]: Academic Press, 2007. Citado na página 11.

STEWART, W. J. *Probability, Markov Chains, queues, and simulation*. 3a. ed. [S.l.]: Princeton University Press, 2009. Citado na página 11.