



**Universidade do Estado do Rio de Janeiro**

Centro de Tecnologia e Ciências

Instituto de Matemática e Estatística

Michelle Lau de Almeida

**Modelagem epidemiológica com um modelo SIR estocástico  
utilizando Cadeia de Markov de tempo contínuo**

Rio de Janeiro

2020

Michelle Lau de Almeida

**Modelagem epidemiológica com um modelo SIR estocástico utilizando  
Cadeia de Markov de tempo contínuo**



Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Ciências Computacionais, da Universidade do Estado do Rio de Janeiro.

Orientadora: Prof.<sup>a</sup> Dra. Zochil González Arenas

Rio de Janeiro

2020

CATALOGAÇÃO NA FONTE  
UERJ / REDE SIRIUS / BIBLIOTECA CTC-A

A447 Almeida, Michelle Lau de.  
Modelagem epidemiológica com um modelo SIR estocástico utilizando Cadeia de Markov de tempo contínuo/ Michelle Lau de Almeida – 2020.  
75 f. : il.

Orientadora: Zochil González Arenas.  
Dissertação (Mestrado em Ciências Computacionais) - Universidade do Estado do Rio de Janeiro, Instituto de Matemática e Estatística.

1. Modelos matemáticos - Teses. 2. Epidemiologia – Métodos estatísticos – Teses. 3. Processo estocástico - Teses. 4. Markov, Processos de – Teses. I. González Arenas, Zochil. II. Universidade do Estado do Rio de Janeiro. Instituto de Matemática e Estatística. III. Título.

CDU 576.851

Patricia Bello Meijinhos CRB7/5217 -Bibliotecária responsável pela elaboração da ficha catalográfica

Autorizo, apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta dissertação, desde que citada a fonte

---

Assinatura

---

Data

Michelle Lau de Almeida

**Modelagem epidemiológica com um modelo SIR estocástico utilizando  
Cadeia de Markov de tempo contínuo**

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Ciências Computacionais, da Universidade do Estado do Rio de Janeiro.

Aprovada em 27 de novembro de 2020.

Banca Examinadora:

---

Prof.<sup>a</sup> Dra. Zochil González Arenas (Orientadora)  
Instituto de Matemática e Estatística – UERJ

---

Prof. Dr. Americo Barbosa da Cunha Junior  
Instituto de Matemática e Estatística – UERJ

---

Prof. Dr. Daniel Adrian Stariolo  
Universidade Federal Fluminense – UFF

---

Prof. Dr. Flávio Codeço Coelho  
Fundação Getúlio Vargas – FGV

---

Prof.<sup>a</sup> Dra. Patricia Nunes da Silva  
Instituto de Matemática e Estatística – UERJ

Rio de Janeiro

2020

## AGRADECIMENTOS

Agradeço, primeiramente, a minha orientadora linda, Profa. Zochil, por sua confiança, sua dedicação, seu carinho e sua paciência para realizarmos esse trabalho. Agradeço por além de ser a minha professora ser uma amiga muito especial na minha vida.

Agradeço também a profa. Roseli que me inspirou com suas aulas e me ajudou em alguns dos momentos difíceis durante essa caminhada. Muito obrigada pelo carinho e amizade.

Em especial, a minha mãe e a minha família, que me apoiaram, ajudaram e sempre torceram por mim. Agradeço por todo amor e carinho que dedicaram apesar da enorme distância que essa caminhada nos forçou a ter e dos momentos que não pude estar presencialmente com cada um.

Não poderia deixar de agradecer aos meus amigos que me ajudaram durante esse tempo, em especial ao Igor, Luis, Raphael e Yuri pela ajuda imprescindível na realização dos códigos. Agradeço pela amizade, pelo carinho que vocês tem por mim e também por todo o apoio e dedicação.

Aos meus amigos do laboratório (OFF-6145) que foram fundamentais para o meu crescimento pessoal e profissional. Nossas conversas, risadas e nossos momentos naquela sala foram muito especiais e que serviam, muitas das vezes, como uma “válvula de escape” quando estávamos muito pressionados com entrega de trabalhos, disciplinas, entre outros. Muito obrigada pela companhia de cada um.

Agradeço também as minhas amigas de república, da nossa querida casa (Yellow Submarine), Andreia, Brenda, Jennifer, Lanna, Rafaela e, em especial a Letícia e Thaís, minhas companheiras de quarto, pela ajuda, pelo apoio, pelas brincadeiras e conversas, e acima de tudo, pela nossa amizade. Vocês foram muito importantes pra mim durante essa caminhada. Muito obrigada!

Agradeço, ainda, a todos os amigos do Instituto de Matemática e Estatística da UERJ que estiveram comigo durante esta caminhada e me ajudaram de alguma forma mesmo que inconscientemente.

Agradeço aos professores que compuseram a minha banca examinadora pela dedicada avaliação e pelos conselhos que me auxiliaram a melhorar o trabalho.

Por fim, pelo apoio financeiro da Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro - FAPERJ e da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES.

Compreender as coisas que nos rodeiam é a melhor preparação para compreender o que há mais além. - *Hipátia*

## RESUMO

LAU DE ALMEIDA, M. *Modelagem epidemiológica com um modelo SIR estocástico utilizando Cadeia de Markov de tempo contínuo*. 2020. 75 f. Dissertação (Mestrado em Ciências Computacionais) – Instituto de Matemática e Estatística, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2020.

Na modelagem epidemiológica, a implementação de modelos matemáticos e suas respectivas simulações computacionais são ferramentas vantajosas para estimar características acerca da propagação de doenças infecciosas. Um dos modelos amplamente utilizados nessa modelagem é o Suscetível-Infetado-Recuperado (SIR), proposto por Kermack e McKendrick em 1927, utilizando uma abordagem determinística. Nesta dissertação, é apresentado o modelo SIR com uma abordagem estocástica, utilizando Cadeia de Markov de Tempo Contínuo. A modelagem estocástica permite inserir aleatoriedades nos parâmetros do modelo e, dessa forma, analisar situações não previstas pelo modelo determinístico. Isto traz algumas vantagens como o acesso a distribuições de probabilidade de algumas propriedades que são exclusivas do modelo estocástico. No presente trabalho, é feita uma revisão bibliográfica atualizada e apresenta-se o embasamento teórico da modelagem estocástica usando Cadeias de Markov. Uma adaptação do modelo SIR estocástico usando Cadeia de Markov de Tempo Contínuo (SIR CTMC) é considerada para duas situações epidemiológicas, um surto de Sarampo na Ilha Grande, Brasil, em 1976, e a pandemia de COVID-19 no estado do Rio de Janeiro, Brasil, em 2020. São realizadas simulações computacionais usando a linguagem Python, a partir das quais é efetuada uma comparação entre os modelos determinístico e estocástico. Nas simulações são utilizados valores dos parâmetros reportados na literatura. Desta forma, tem sido possível validar o modelo estocástico e algumas das suas vantagens ficam em evidência.

Palavras-chave: Modelagem epidemiológica estocástica. Modelos compartimentais.

Cadeias de Markov. Modelo SIR CTMC.

## ABSTRACT

LAU DE ALMEIDA, M. *Epidemiological Modeling with a Stochastic SIR Model using Continuous Time Markov Chain*. 2020. 75 f. Dissertação (Mestrado em Ciências Computacionais) – Instituto de Matemática e Estatística, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2020.

In epidemiological modeling, the implementation of mathematical models and their respective computer simulations are advantageous tools for estimating characteristics about the spread of infectious diseases. A well-known and widely used model for this aim is the Susceptible-Infected-Recovered (SIR), proposed by Kermack and McKendrick in 1927, by using a deterministic approach. In this dissertation, the SIR model is presented with a stochastic approach, using Continuous Time Markov Chain. Stochastic modeling allows to insert some randomness into the model's parameters and, thus, analyze situations that are not foreseen by the deterministic model. It brings some advantages such as the access to probability distributions for some properties that are exclusive to the stochastic model. In the present work, an updated bibliographic review is made and the theoretical background of stochastic modeling using Markov Chains is given. An adaptation of the stochastic SIR model using Continuous Time Markov Chain (SIR CTMC) is considered for two epidemiological situations, a measles outbreak in Ilha Grande, Brazil, in 1976, and the COVID-19 pandemic in the state of Rio de Janeiro, Brazil, in 2020. Computer simulations are performed using the Python language, from which a comparison is made between the deterministic and stochastic models. In the simulations, values of the parameters reported in the literature are used. In this way, it has been possible to validate the stochastic model and some of its advantages are in evidence.

Keywords: Epidemiological stochastic models. Compartmental models. Markov Chains. CTMC SIR Model.



## LISTA DE ILUSTRAÇÕES

Figura 1 - Grafo direcionado representando as relações entre os estados de uma Cadeia de Markov. . . . .	28
Figura 2 - Grafo direcionado representando as classificações dos estados em uma Cadeia de Markov. . . . .	29
Figura 3 - Ilustração do tempo entre eventos. . . . .	30
Figura 4 - Diagrama do modelo compartimental <i>SIR</i> . . . . .	41
Figura 5 - Diferentes cenários na modelagem estocástica. . . . .	58
Figura 6 - Média do pico da epidemia no surto de sarampo. . . . .	59
Figura 7 - Média do dia de pico da epidemia no surto de sarampo. . . . .	60
Figura 8 - Distribuição de Probabilidade do tamanho final da epidemia . . . . .	61
Figura 9 - Média do tamanho final da epidemia . . . . .	62
Figura 10 - Média da duração da Epidemia . . . . .	62
Figura 11 - Comparação dos modelos estocástico (linha cheia) e determinístico (linha tracejada) . . . . .	64
Figura 12 - Comparação dos modelos estocástico e determinístico (não ocorrência da epidemia). . . . .	65
Figura 13 - Média do dia de pico da epidemia para a epidemia da COVID-19 . . . . .	65
Figura 14 - Diagrama do modelo <i>SIQR</i> . . . . .	72
Figura 15 - Comparação da simulação dos modelos SIQR determinístico e estocástico . . . . .	75

## LISTA DE TABELAS

Tabela 1 - Probabilidades de transição para o modelo SIR CTMC. . . . .	46
Tabela 2 - Incidência de sarampo de acordo com a idade. . . . .	57
Tabela 3 - Probabilidades de transição para o modelo SIQR CTMC. . . . .	74

## SUMÁRIO

	<b>INTRODUÇÃO</b> . . . . .	10
1	<b>EPIDEMIOLOGIA MATEMÁTICA</b> . . . . .	12
1.1	Origem da Epidemiologia Matemática . . . . .	12
1.2	Conceitos Básicos da Epidemiologia . . . . .	13
1.3	Teorema do Limiar e Princípio de Ação de Massas . . . . .	14
1.4	Modelos Compartimentais . . . . .	16
2	<b>PROBABILIDADE E PROCESSOS ESTOCÁSTICOS</b> . . . . .	17
2.1	Conceitos Básicos de Probabilidade . . . . .	17
2.2	Variáveis Aleatórias . . . . .	19
2.3	Introdução aos Processos Estocásticos . . . . .	23
3	<b>CADEIAS DE MARKOV DE TEMPO CONTÍNUO</b> . . . . .	25
3.1	Classificação de Estados . . . . .	27
3.2	Tempo Entre Eventos . . . . .	29
3.3	Matriz Geradora $G$ . . . . .	33
3.4	Equações Diferenciais de Kolmogorov . . . . .	36
3.5	Primeiro Tempo de Passagem . . . . .	38
4	<b>MODELO COMPARTIMENTAL SIR</b> . . . . .	41
4.1	Modelo SIR . . . . .	41
4.2	Modelo Determinístico . . . . .	42
4.3	Modelo Estocástico . . . . .	45
5	<b>SIMULAÇÕES E RESULTADOS</b> . . . . .	56
5.1	Surto de Sarampo na Ilha Grande . . . . .	56
5.2	O Grande Desafio: COVID-19 no Rio de Janeiro . . . . .	61
	<b>CONCLUSÃO</b> . . . . .	67
	<b>REFERÊNCIAS</b> . . . . .	69
	<b>APÊNDICE</b> – Modelo SIQR . . . . .	72

## INTRODUÇÃO

A epidemiologia pode ser definida como a ciência que é voltada para o estudo das causas, da evolução e da distribuição das doenças em populações [1]. Inicialmente, a epidemiologia preocupava-se somente com doenças infecciosas, porém, seus estudos foram ampliados, incluindo também outros agravos à saúde. Através desses estudos podem-se implementar estratégias de prevenção e controle das doenças, sendo o objetivo geral da epidemiologia obter medidas para a melhoria da saúde da população [2].

Na epidemiologia matemática o interesse em modelar doenças infecciosas tem como objetivo geral melhorar as estratégias de prevenção e controle, assim como auxiliar governos e agentes de saúde na tomada de decisão. Os modelos matemáticos são ferramentas importantes para realizar a análise da propagação e de controle de doenças infecciosas. São usados para comparar, planejar, implementar e otimizar programas de detecção, prevenção e controle. Estimar a duração de uma epidemia e o número de pessoas que pode adoecer é um dos objetivos mais importantes na modelagem epidemiológica matemática. Além dos modelos, simulações computacionais também são úteis para construir e testar teorias e determinar a sensibilidade a mudanças nos valores dos parâmetros. A modelagem epidemiológica pode contribuir para fazer previsões gerais e estimar as incertezas nas previsões [3].

Os modelos para predição e análise da propagação de doenças podem ser determinísticos ou estocásticos. Nos modelos determinísticos o processo epidêmico é representado por equações diferenciais ordinárias, que são utilizadas para descrever a evolução temporal da doença, dada uma condição inicial, e é possível controlar os fatores que modelam esse processo, como, por exemplo, o termo de transmissão da doença, definido na Seção 1.3. Com isso, é possível prever resultados precisos de acordo com esses fatores. Uma forma de modelagem determinística extensivamente usada são os modelos compartimentais. Esses modelos dividem a população em classes (ou compartimentos), que representam os vários estados da doença de acordo com o modelo utilizado. Os estados podem ser, entre outros, de (S)uscetíveis, (I)nfectados, (R)ecuperados, (Q)uarentenados, usando-se a primeira letra da palavra para representar um compartimento específico. Maiores detalhes serão apresentados na próxima seção.

Por outra parte, a modelagem estocástica fornece uma abordagem importante para o estudo de fenômenos naturais. Quando essa modelagem é aplicada à descrição de problemas da biologia, em particular à modelagem epidemiológica, permite estudar comportamentos e fazer conclusões que não seriam possíveis com uma modelagem determinística, pois estimam as distribuições de probabilidade dos resultados possíveis. Os modelos estocásticos são ferramentas que permitem inserir uma certa aleatoriedade nos parâmetros estudados. Por exemplo, através dos modelos estocásticos pode-se falar da probabilidade

de surto de uma epidemia ou probabilidade de extinção, ao invés de tratar essas situações como totalmente certas a partir de modelos determinísticos. Pode-se também considerar efeitos sobre os quais não temos conhecimentos suficiente ou que podem ser aleatórios, permitindo, dessa forma, uma modelagem mais realista [4]. Nessa categoria, encontram-se fundamentalmente dois tipos de modelos: os modelos compartimentais, considerando cada classe antes indicada como uma variável aleatória e, modelos baseados em equações diferenciais estocásticas. Geralmente são usados em populações pequenas devido ao alto custo computacional para tratá-los. [4].

O objetivo geral do trabalho é o estudo de modelos compartimentais estocásticos, do ponto de vista analítico e computacional.

Será considerado o modelo SIR (Suscetíveis-Infetados-Recuperados), proposto por Kermack e McKendrick [5] e muito utilizado em modelagem epidemiológica [6, 7, 8], visando fazer uma introdução à teoria da modelagem estocástica e analisar o comportamento desse modelo na descrição da evolução de epidemias. No trabalho será apresentado o modelo SIR determinístico, modelado por equações diferenciais. O modelo servirá de base para a compreensão da modelagem e da extensão ao modelo estocástico. Será apresentada a sua versão estocástica através do uso de Cadeias de Markov [9].

Uma outra prática comum nas simulações em dinâmica estocástica é baseada em Simulações de Monte Carlo [10]. O presente trabalho tem também por objetivo o estudo e desenvolvimento de estratégias computacionais para realizar a simulação dessa dinâmica estocástica usando a metodologia dos modelos compartimentais. Para verificação dos resultados, vamos estabelecer uma comparação com aqueles obtidos pela simulação dos modelos compartimentais determinísticos. Em geral, a simulação da dinâmica de sistemas estocásticos tem um custo computacional muito elevado, dada a necessidade de múltiplas repetições para realizar uma estatística válida. No caso dos modelos compartimentais, por outra parte, o custo computacional é ainda maior, pelas razões acima descritas. Nesse trabalho, vamos formular e implementar rotinas para a simulação da dinâmica estocástica na linguagem Python.

A dissertação está organizada em cinco capítulos, sendo apresentada, no primeiro, a origem da epidemiologia matemática, bem como alguns conceitos e termos epidemiológicos utilizados no decorrer do trabalho. No Capítulo 2, uma revisão de conceitos básicos de probabilidade, variáveis aleatórias e uma introdução aos processos estocásticos será realizada. Na sequência, será feito um estudo sobre Cadeia de Markov de Tempo Contínuo, processo no qual foi baseada a construção do modelo estocástico. No Capítulo 4 apresentam-se conceitos importantes das versões determinística e estocástica do modelo SIR. No Capítulo 5, são apresentados os resultados do trabalho realizado, seguido pelas Conclusões. Por fim, uma abordagem preliminar do modelo compartimental SIQR é apresentada no Apêndice.

## 1 EPIDEMIOLOGIA MATEMÁTICA

A epidemiologia é definida como o estudo de fatores que geram a frequência e a distribuição das doenças em populações [11]. O estudo é realizado em grupos de pessoas, podendo ser um grupo pequeno mas, na maioria das vezes, em uma população numerosa. A aplicação desse estudo é utilizada no controle de doenças, visando propor medidas específicas de prevenção, controle e erradicação das doenças, fornecendo indicadores que servem para planejar, administrar e avaliar ações realizadas para promover a saúde da população [12].

### 1.1 Origem da Epidemiologia Matemática

A história da epidemiologia acompanha a história da medicina, especialmente da medicina preventiva, então não há certeza de quando e quem foi o primeiro a definir essa ciência. Devido a esse fato, considera-se que Hipócrates (460 – 377 a.C), conhecido como o “pai da medicina”, deu início às principais bases para os estudos epidemiológicos. Para ele, as doenças poderiam ser descritas através de perspectivas racionais, como fatores climáticos, raciais e do meio onde as pessoas viviam, contrastando com as ideias que existiam naquela época, que explicavam as doenças por argumentos sobrenaturais [1, 13].

Já na Idade Moderna, John Graunt (1620 – 1674) identificou diferenças na mortalidade de grupos populacionais de acordo com o sexo e com o lugar de residência. Com isso, conduziu os primeiros estudos analíticos de estatística vital [14]. Um marco importante na valorização da matemática como suporte para o estudo e entendimento no campo científico da saúde se deve a Daniel Bernoulli (1700 – 1782) que estabeleceu a primeira modelagem matemática da propagação de doenças infecciosas. Bernoulli deduziu fórmulas para estimar os anos de vida ganhos pela vacinação contra a varíola e para avaliar o custo-benefício em intervenções clínicas. Os parâmetros definidos por ele são utilizados ainda hoje para examinar rumos de epidemias [1, 15].

Estimado por sua pesquisa minuciosa sobre cólera, John Snow (1813 – 1858) é considerado o “pai da epidemiologia moderna”. Sua pesquisa sobre a cólera, em Londres, se baseou na teoria de que a doença havia sido transmitida a partir de água contaminada. Ele analisou a água que era distribuída por duas companhias e observou que as pessoas que recebiam água por elas eram contaminadas com a doença. Esse trabalho ficou conhecido como um experimento natural [14]. Muitos conceitos, abordagens e métodos utilizados por Snow são úteis em estudos epidemiológicos recentes [16].

Ronald Ross (1857 – 1932) começou seu estudo sobre a dinâmica da transmissão da malária em 1892 e realizou um experimento para justificar que a propagação da doença

era dada através de mosquitos, o que foi concretizado em 1897. Também formulou a hipótese de que ocorreria a extinção natural da doença se existisse um valor limite para a densidade de mosquitos [17]. Essa hipótese foi um prenúncio do Teorema do Limiar, proposto por Kermack e McKendrick em 1927 [5].

No início do século XX foram formuladas teorias específicas sobre a transmissão de doenças infecciosas. William Hamer (1862 – 1936), em 1906, propôs a curva epidêmica do sarampo e desenvolveu a expressão matemática tendo por base a sua “teoria mecânica de números e densidade”. Hamer definiu que uma epidemia se desenvolve através do número de suscetíveis e de infectados e da taxa de contato entre eles, o que resultou em um dos conceitos mais importantes na epidemiologia matemática, o princípio de ação de massas [18]. Por sua parte, Wade Hampton Frost (1880 – 1938) foi quem introduziu o elemento estocástico nas curvas epidêmicas, possibilitando com isso uma excelente diversificação dos objetos de investigação da epidemiologia e uma impactante ampliação da aplicabilidade de seus conhecimentos [18]. Frost ficou bastante conhecido também pelo desenvolvimento do modelo Reed-Frost para transmissão de doenças infecciosas, em parceria com Lowell Jacob Reed (1886 – 1966), e também pelo estudo do desenvolvimento da imunidade por rebanho [14].

No entanto, foi a partir do final da Segunda Guerra Mundial que a metodologia epidemiológica obteve um intenso desenvolvimento, com o auxílio da estatística, propiciado pelo aparecimento de computadores. Dessa forma, a aplicação da epidemiologia consegue se estender a diversos agravos à saúde.

## 1.2 Conceitos Básicos da Epidemiologia

Na dinâmica populacional de doenças infecciosas, os objetivos são impedir o progresso de uma epidemia, compreender quais razões levaram à ocorrência dessa epidemia e também adotar medidas preventivas para evitar que novas ocorrências aconteçam.

A seguir, relacionam-se os principais termos epidemiológicos que serão usados no decorrer do trabalho.

**agente infeccioso:** agente biológico capaz de produzir infecção ou doença infecciosa.

**endemia:** doença infecciosa que ocorre habitualmente.

**epidemia:** propagação de uma doença infecciosa acima do que seria teoricamente esperado.

**hospedeiro:** organismo simples ou complexo, incluindo o homem, capaz de ser infectado por um agente específico.

**incidência:** número de novos casos surgidos em uma determinada população em

um determinado intervalo de tempo.

**infecção:** reprodução de um agente em seu hospedeiro podendo haver ou não doença.

**infectado:** hospedeiro que tem uma infecção.

**pandemia:** propagação mundial de uma doença infecciosa. Termo utilizado quando uma epidemia ocorre em diferentes continentes.

**período latente:** é o intervalo de tempo que transcorre desde que se produz a infecção até que a pessoa se torne infecciosa.

**suscetível:** indivíduo acessível ou capaz de ser infectado.

**transmissão:** processo pelo qual um agente passa de uma fonte de infecção para um novo hospedeiro.

**vetor:** hospedeiro de parasitas com ciclos indiretos de vida ou qualquer coisa que transmite parasitas ou um invertebrado transmissor de vírus para vertebrados.

Na epidemiologia, os determinantes de doenças, que são fatores que interferem na transmissão de uma doença infecciosa, podem ser relacionados ao meio ambiente, ao hospedeiro ou ao agente infeccioso. Os fatores geográficos e climáticos são aqueles determinantes relacionados ao meio ambiente. Os fatores relacionados ao hospedeiro são o genótipo (composição genética de um indivíduo), idade e sexo. Já os determinantes relacionados ao agente infeccioso são a virulência (capacidade do agente infeccioso gerar manifestações graves e fatais), a patogenicidade (capacidade do agente produzir doença em pessoas infectadas), e o resultado da infecção [19]. Esses fatores acabam influenciando no espalhamento da doença em populações. A transmissão de doenças infecciosas pode acontecer de forma direta, através do contato direto entre um hospedeiro infectado e um hospedeiro suscetível, e pode ser de forma indireta, onde o contato entre esses hospedeiros é intermediado por um vetor.

### 1.3 Teorema do Limiar e Princípio de Ação de Massas

Uma das definições fundamentais na epidemiologia é o Teorema do Limiar, que está relacionado à ideia de que se a densidade de indivíduos suscetíveis de uma população estiver abaixo de um determinado limiar, então a epidemia de uma doença infecciosa não se propaga nessa população. Ao contrário, se a densidade estiver acima desse limiar, então pode favorecer a propagação dessa doença.

Relacionado ao Teorema do Limiar tem-se o conceito de *número de reprodução básico*,  $\mathcal{R}_0$ . Esse conceito tem um papel fundamental na modelagem epidemiológica. Ele é utilizado para medir o potencial de transmissão de uma doença.  $\mathcal{R}_0$  representa o número médio de casos secundários que um indivíduo infeccioso pode produzir, considerando uma



população totalmente suscetível. Pode-se determinar quando há uma epidemia através do valor do número de reprodução básico. Quando  $\mathcal{R}_0 > 1$ , há a ocorrência de epidemia; se  $\mathcal{R}_0 < 1$ , a doença é facilmente eliminada da população e não ocorre epidemia; por fim, quando  $\mathcal{R}_0 = 1$ , a doença é endêmica, isto é, há um equilíbrio endêmico.

No caso de já estar ocorrendo uma epidemia em uma população, ou seja, há indivíduos doentes e, portanto, a população não é suscetível na sua totalidade, define-se o *número de reprodução efetiva*  $\mathcal{R}$ , por

$$\mathcal{R} = \mathcal{R}_0 S, \tag{1}$$

onde  $S$  é a densidade de indivíduos suscetíveis a contrair a doença. O número de reprodução efetiva representa quantos casos secundários ocorrem, em média, a cada indivíduo que é infectado. Dessa forma, para que a doença não seja propagada em uma população, o número de reprodução efetiva definido pela equação (1) tem que ser menor que um. Daqui, obtem-se, como relação alternativa,

$$S < \frac{1}{\mathcal{R}_0}. \tag{2}$$

Esta relação indica qual o valor limite de indivíduos suscetíveis para que a doença não seja propagada.

Sejam  $S$  e  $I$  as densidades de indivíduos suscetíveis e infectados, respectivamente, numa população dada, e seja  $\beta$  a taxa de contato entre esses indivíduos (suscetíveis e infectados). O Princípio de Ação de Massas estabelece então a existência de um termo de transmissão, representado por  $\beta SI$ .

A ideia do Princípio de Ação de Massas utilizado aqui é semelhante ao Princípio de Ação de Massas utilizado na química, onde massas de diferentes reagentes se misturam e interagem e resultam em uma certa concentração de produtos. No caso da epidemiologia, têm-se massas de suscetíveis que interagem com massas de infectados de uma maneira homogênea. A partir do termo de transmissão, podem-se elaborar algumas estratégias de prevenção e controle da doença. Por exemplo, para controlar uma doença é necessário reduzir o termo de transmissão. Por sua vez, para reduzir o termo de transmissão pode-se reduzir qualquer um dos seus elementos. Uma estratégia possível para reduzir os indivíduos suscetíveis é a vacinação. Por outra parte, para reduzir a taxa de contato pode-se pensar na estratégia de isolar os indivíduos infectados. Dessa forma, é possível traçar estratégias que controlem uma epidemia.

## 1.4 Modelos Compartimentais

Os modelos matemáticos são ferramentas muito utilizadas na epidemiologia a fim de simular e analisar a propagação de doenças infecciosas. Além disso, através da modelagem é possível prever impactos de intervenções e otimizar ações de controle. Nesta área, os modelos mais utilizados são os modelos compartimentais, que recebem esse nome pois dividem a população em classes ou compartimentos que representam os diversos estados da doença. Esses compartimentos são definidos de acordo com as particularidades da doença e o modelo proposto para o seu estudo. Além disso, cada compartimento é representado pela primeira letra, em maiúscula, do estado da doença que está sendo considerado.

Alguns dos modelos compartimentais mais empregados são:

O modelo **SI**  $\Rightarrow$  *Suscetível - Infectado* é utilizado para modelar doenças onde os indivíduos suscetíveis tornam-se infectados e permanecem nesse compartimento até a morte, ou seja, não há cura nesse modelo. Um exemplo de doença que pode ser modelada através do modelo *SI* é a AIDS.

O modelo **SIS**  $\Rightarrow$  *Suscetível - Infectado - Suscetível* é utilizado para modelar doenças onde os indivíduos suscetíveis tornam-se infectados e após se recuperarem voltam ao compartimento de suscetíveis. Esses indivíduos não possuem imunidade. Um exemplo de doença utilizado para essa modelagem é a gripe comum.

O modelo **SIR**  $\Rightarrow$  *Suscetível - Infectado - Recuperado* é utilizado para modelar doenças onde os indivíduos suscetíveis tornam-se infectados e após se recuperarem tornam-se imunes à doença. Esse modelo é muito utilizado na modelagem de doenças como, por exemplo, sarampo, rubéola e dengue.

A partir do modelo SIR diversas variações têm sido propostas, bem como variações das variações. Trata-se de adaptar o modelo a situações diferentes, acrescentando novos compartimentos e/ou relações de vários tipos entre esses compartimentos. Alguns exemplos clássicos são os modelos:

- **SIRS**  $\Rightarrow$  *Suscetível - Infectado - Recuperado - Suscetível*
- **SIQR**  $\Rightarrow$  *Suscetível - Infectado - Quarentenado - Recuperado*
- **SEIR**  $\Rightarrow$  *Suscetível - Exposto - Infectado - Recuperado*

## 2 PROBABILIDADE E PROCESSOS ESTOCÁSTICOS

São encontradas na natureza inúmeras situações que geram incertezas, nas quais a utilização de um modelo determinístico é inadequada. Tais situações podem ser mais apropriadamente descritas utilizando modelos probabilísticos ou estocásticos. As relações e variáveis consideradas nesses modelos são baseadas na teoria das probabilidades. Dessa forma, quando tais modelos são empregados, sabe-se que o fenômeno é relacionado a uma imprecisão. A essas situações denominamos *experimentos aleatórios*.

Existem variadas abordagens para considerar modelos estocásticos. Na presente dissertação, vamos trabalhar com modelos compartimentais estocásticos, cuja teoria será apresentada no Capítulo 4. Nesta seção, são oferecidos conceitos básicos de probabilidade que ajudam na compreensão do próximo capítulo, assim como do trabalho em geral.

### 2.1 Conceitos Básicos de Probabilidade

Experimentos aleatórios são fenômenos que, avaliados repetidas vezes em um processo, geram resultados imprevisíveis. O conjunto de todos esses resultados possíveis chama-se de *espaço amostral* e denota-se por  $\Omega$ . O espaço amostral pode ser finito ou infinito. No caso do espaço amostral ser infinito não enumerável ele é denominado espaço amostral contínuo. Cada resultado possível desse conjunto é denominado elemento do espaço e representado por  $\omega$ . Subconjuntos do espaço amostral  $\Omega$  são representados por letras do alfabeto maiúsculas.

O *conjunto das partes* do espaço amostral é formado por todos os subconjuntos de  $\Omega$  e denotado por  $\mathcal{P}\{\Omega\}$ . Se  $\Omega$  for um conjunto finito de  $n$  elementos, o conjunto  $\mathcal{P}\{\Omega\}$  terá  $2^n$  elementos.

#### **Definição 2.1.** *$\sigma$ -álgebra*

Uma classe de subconjuntos de  $\Omega$ , representada por  $\mathcal{F}$ , é denominada  *$\sigma$ -álgebra* se satisfaz as seguintes condições:

- $\Omega \in \mathcal{F}$
- Se  $A \in \mathcal{F}$ , então  $A^c \in \mathcal{F}$
- Se  $A_i \in \mathcal{F}, i \geq 1$  então  $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

Se apenas a união finita está em  $\mathcal{F}$ , então tem-se uma classe menos restrita, denominada somente *álgebra*. Uma  $\sigma$ -álgebra é interpretada como uma coleção de eventos de  $\Omega$  que podem ser atribuídos a probabilidades.

A definição mais simples de probabilidade refere-se a conjuntos unitários equiprováveis. Considerando o caso de um espaço amostral finito ou infinito enumerável, temos a seguinte definição:

$$Prob(A) = \frac{\text{número de elementos de } A}{\text{número total de elementos de } \Omega} . \quad (3)$$

Se o número de elementos em  $\Omega$  for infinito, essa definição é dada através do uso de limites.

Para o caso de  $\Omega$  não ser enumerável, esse conceito de probabilidade será aplicado ao comprimento de intervalos, dando origem ao que chamamos de *probabilidade geométrica*.

Essas definições são casos particulares da definição de probabilidade dada por Andrei Nikolaevich Kolmogorov (1903–1987), importante matemático russo que desenvolveu uma axiomática para a teoria de probabilidade [20]. A partir dessa axiomática, pode ser considerada a seguinte definição.

**Definição 2.2.** *Uma função  $\mathcal{P}$ , definida na  $\sigma$ -álgebra  $\mathcal{F}$  de subconjuntos de  $\Omega$  e com valores no intervalo  $[0, 1]$ , é uma probabilidade se satisfaz os seguintes Axiomas de Kolmogorov:*

- $Prob(\Omega) = 1$
- Para todo subconjunto  $A \in \mathcal{F}$ ,  $Prob(A) \geq 0$
- Para toda sequência  $A_1, A_2, \dots \in \mathcal{F}$ , mutuamente exclusivos, temos:

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

A trinca  $(\Omega, \mathcal{F}, \mathcal{P})$  é denominada *espaço de probabilidade*.

A probabilidade possui propriedades que são importantes para algumas demonstrações realizadas no decorrer do trabalho. Algumas propriedades são destacadas abaixo.

Dada a trinca  $(\Omega, \mathcal{F}, \mathcal{P})$  e considerando que os conjuntos  $A$  e  $B$  são eventos desse espaço de probabilidade, tem-se que

1.  $Prob(A) = 1 - Prob(A^c)$
2. Sendo  $A$  e  $B$  dois eventos quaisquer,  $Prob(B) = Prob(B \cap A) + Prob(B \cap A^c)$
3. Se  $A \subset B$  então  $Prob(A) \leq Prob(B)$
4.  $Prob(A \cup B) = Prob(A) + Prob(B) - Prob(A \cap B)$
5. Para quaisquer eventos  $A_1, A_2, \dots$ ,  $Prob\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} Prob(A_i)$

**Definição 2.3.** Consideremos os eventos  $A$  e  $B$  no seguinte espaço de probabilidade  $(\Omega, \mathcal{F}, \mathcal{P})$ . Sendo  $\text{Prob}(B) > 0$ , a probabilidade condicional de ocorrer  $A$  dado que anteriormente ocorreu  $B$ , é dada por

$$\text{Prob}(A|B) = \frac{\text{Prob}(A \cap B)}{\text{Prob}(B)}. \quad (4)$$

A seguir, vamos deduzir uma propriedade importante, que segue da definição 2.3. Para a probabilidade condicional de ocorrer  $A \cap B$  dado que ocorreu  $C$ , tem-se

$$\text{Prob}(A \cap B|C) = \frac{\text{Prob}((A \cap B) \cap C)}{\text{Prob}(C)} = \frac{\text{Prob}(A \cap (B \cap C))}{\text{Prob}(C)}. \quad (5)$$

Usando novamente a expressão (4), pode-se escrever

$$\text{Prob}(A \cap (B \cap C)) = \text{Prob}(A|B \cap C)\text{Prob}(B \cap C), \quad (6)$$

e, da mesma forma,

$$\text{Prob}(B \cap C) = \text{Prob}(B|C)\text{Prob}(C). \quad (7)$$

Substituindo agora as expressões (6) e (7) no membro da direita da equação (5), obtem-se

$$\text{Prob}(A \cap B|C) = \frac{\text{Prob}(A|B \cap C)\text{Prob}(B|C)\text{Prob}(C)}{\text{Prob}(C)}, \quad (8)$$

chegando, por fim, na seguinte propriedade da probabilidade condicional,

$$\text{Prob}(A \cap B|C) = \text{Prob}(A|B \cap C)\text{Prob}(B|C). \quad (9)$$

## 2.2 Variáveis Aleatórias

Muitas das vezes é conveniente representar numericamente o resultado de uma experiência realizada. A representação numérica dessa experiência é dada através do conceito de variável aleatória.

**Definição 2.4.** Seja  $(\Omega, \mathcal{F}, \mathcal{P})$  um espaço de probabilidade. Denomina-se variável aleatória, qualquer função  $X : \Omega \rightarrow \mathbb{R}$  tal que

$$X^{-1}(I) = \{\omega \in \Omega : X(\omega) \in I\} \in \mathcal{F}, \quad (10)$$

para todo intervalo  $I \subset \mathbb{R}$ . Ou seja,  $X$  é tal que a sua imagem inversa de intervalos

$I \subset \mathbb{R}$  pertence à  $\sigma$ -álgebra  $\mathcal{F}$ .

**Definição 2.5.** *Seja  $X$  uma variável aleatória em  $(\Omega, \mathcal{F}, \mathcal{P})$ . Sua função de distribuição acumulada é definida por*

$$F_X(x) = \text{Prob}(X \in (-\infty, x]) = \text{Prob}(X \leq x) \quad (11)$$

com  $x$  percorrendo todos os reais.

Informações sobre a variável aleatória podem ser obtidas através do conhecimento da função de distribuição. Como  $F_X$  é uma probabilidade, ela deve satisfazer os Axiomas de Probabilidade. Dessa forma, precisa cumprir as seguintes propriedades,

1.  $\lim_{x \rightarrow -\infty} F_X(x) = 0$  e  $\lim_{x \rightarrow \infty} F_X(x) = 1$ ;
2.  $F_X$  é contínua à direita;
3.  $F_X$  é não decrescente, ou seja,  $F_X(x) \leq F_X(y)$  sempre que  $x \leq y$ ,  $\forall x, y \in \mathbb{R}$ .

Uma variável aleatória associa a todo elemento do espaço amostral um único número real e pode ser discreta ou contínua. Essa classificação tem estreita relação com o comportamento da função de distribuição. Uma variável aleatória é discreta quando assume um número enumerável de valores (finito ou infinito). A função que atribue probabilidade a cada um dos possíveis valores que a variável aleatória pode assumir é denominada função de probabilidade.

**Definição 2.6.** *A Função de Probabilidade é a função que associa a cada valor assumido pela variável aleatória a probabilidade do evento correspondente, ou seja,*

$$p_X(x_i) = \text{Prob}(X = x_i) = \text{Prob}(A_i), \quad i = 1, 2, \dots$$

A função de probabilidade de  $X$  em  $(\Omega, \mathcal{F}, \mathcal{P})$  deve satisfazer as seguintes propriedades,

1.  $0 \leq p_X(x_i) \leq 1$ , para todo  $i = 1, 2, \dots$ ;
2.  $\sum_i p_X(x_i) = 1$ .

O conjunto dos pares possíveis,  $\{(x_i, p_X(x_i))\}$ , com  $i = 1, 2, \dots$ , de valores que a variável aleatória discreta  $X$  pode assumir e suas respectivas probabilidades formam a distribuição de probabilidade da variável aleatória  $X$ .

Por outro lado, uma variável aleatória é contínua quando assume qualquer valor numérico em um determinado intervalo ou série de intervalos.

**Definição 2.7.** Uma variável aleatória  $X$  em  $(\Omega, \mathcal{F}, \mathcal{P})$ , com função de distribuição  $F_X$ , é classificada como contínua, se existe uma função real não negativa  $f_X$  tal que,

$$F_X(x) = \int_{-\infty}^x f_X(\bar{x})d\bar{x}, \quad (12)$$

para todo  $x \in \mathbb{R}$ . A função  $f_X$  é denominada função densidade de probabilidade (*pdf*<sup>1</sup>, pelas siglas em inglês) ou simplesmente densidade da variável aleatória contínua  $X$ .

A função densidade de  $X$  em  $(\Omega, \mathcal{F}, \mathcal{P})$  deve satisfazer as seguintes propriedades,

1.  $f_X(x) \geq 0$ , para todo  $x \in \mathbb{R}$ ;
2.  $\int_{-\infty}^{\infty} f_X(\bar{x})d\bar{x} = 1$ .

Para se obter a probabilidade de uma variável aleatória contínua estar num determinado intervalo  $[x, y]$ , calcula-se a integral da função densidade nesse intervalo. Dessa forma,

$$Prob(x \leq X \leq y) = \int_x^y f_X(\bar{x})d\bar{x} = F_X(y) - F_X(x). \quad (13)$$

Em muitos problemas é interessante estudar mais de um resultado de um experimento aleatório simultaneamente, que pode ser analisado através da utilização de variáveis aleatórias bidimensionais.

**Definição 2.8.** Seja  $X$  uma variável aleatória que assume os valores  $x_1, x_2, \dots, x_n$ , e  $Y$  uma variável aleatória que assume os valores  $y_1, y_2, \dots, y_m$ . A função de probabilidade conjunta associa a cada par  $(x_i, y_j)$ , com  $i = 1, \dots, n$  e  $j = 1, \dots, m$ , a probabilidade

$$Prob(X = x_i, Y = y_j) = p(x_i, y_j). \quad (14)$$

O conjunto dos pares  $\{(x_i, y_j), (p(x_i), p(y_j))\}$ , com  $i = 1, \dots, n$  e  $j = 1, \dots, m$ , é chamado de *distribuição conjunta de probabilidades* da variável bidimensional  $(X, Y)$ .

Observa-se que:

$$\sum_{i=1}^n \sum_{j=1}^m Prob(X = x_i, Y = y_j) = 1. \quad (15)$$

Um conceito relevante ao se trabalhar com variáveis aleatórias, necessário na obtenção de resultados importantes, é o conceito de independência.

---

<sup>1</sup> probability density function

**Definição 2.9.** *Sejam  $X = x_1, \dots, x_n$  e  $Y = y_1, \dots, y_m$  duas variáveis aleatórias tais que  $\text{Prob}(X = x_i) = p(x_i)$  para  $i = 1, \dots, n$  e  $\text{Prob}(Y = y_j) = p(y_j)$  para  $j = 1, \dots, m$ .  $X$  e  $Y$  são independentes se, e somente se,*

$$\text{Prob}(X = x_i, Y = y_j) = \text{Prob}(X = x_i)\text{Prob}(Y = y_j), \quad (16)$$

para todo par  $(x_i, y_j)$ .

Para tentar resumir o comportamento de uma variável aleatória existe uma medida que estuda a tendência central da variável aleatória, chamada de esperança ou valor esperado ou valor médio de uma variável aleatória.

**Definição 2.10.** *Seja  $X$  uma variável aleatória discreta, tal que  $X = x_1, x_2, \dots, x_n$  com função de probabilidade  $\text{Prob}(X = x_i) = p_X(x_i)$  para  $i = 1, \dots, n$ . A esperança matemática de uma variável aleatória é a média aritmética ponderada do valor que  $X$  pode assumir e suas respectivas probabilidades.*

$$\mathbb{E}(X) = \sum_{i=1}^n x_i p(x_i) \quad (17)$$

**Definição 2.11.** *Seja  $X$  uma variável aleatória contínua com função densidade de probabilidade  $f_X(x)$ . O valor esperado de  $X$  é dado por,*

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x) dx \quad (18)$$

Por fim, definimos uma distribuição de probabilidade muito importante para estabelecer alguns conceitos no decorrer do trabalho, que é a distribuição exponencial. As variáveis aleatórias com distribuição exponencial têm uma propriedade muito utilizada, conhecida como falta de memória.

**Definição 2.12.** *A variável aleatória  $X$  tem distribuição exponencial com parâmetro  $\lambda$ ,  $\lambda > 0$ , se tiver a sua pdf dada por:*

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}, \quad (19)$$

onde o parâmetro  $\lambda$  é a taxa da distribuição. A função de distribuição acumulada  $F_X(x)$  é dada por,

$$F_X(x) = \int_0^x f_X(\bar{x}) d\bar{x} = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}. \quad (20)$$



### 2.3 Introdução aos Processos Estocásticos

A palavra *estocástico* tem origem grega (“stochastikós”) e significa “processos que não estão submetidos senão às leis do acaso”, aleatórios. O contrário de estocástico é o que é certo ou determinístico. No modelo determinístico é previsto um único resultado de um determinado conjunto de circunstâncias. Já o modelo estocástico prevê um conjunto de possíveis resultados ponderados por suas probabilidades [21]. Dessa forma, pode-se descrever um processo estocástico, matematicamente, como um fenômeno/experimento aleatório que muda com o tempo. A teoria dos processos estocásticos trata sistemas que se desenvolvem em tempo ou espaço de acordo com as leis probabilísticas [22]. O estudo do desenvolvimento dessa teoria teve início a partir de problemas biológicos e físicos e, atualmente, são usados para modelar fenômenos em diferentes áreas como química, física, biologia, engenharia e economia, entre outras [9]. Em muitos estudos, [23, 24], a utilização de processos estocásticos em modelagem nos dá uma percepção mais realista que os modelos determinísticos.

**Definição 2.13.** *Um processo estocástico é uma coleção de variáveis aleatórias  $\{X_t(\omega) : t \in T, \omega \in \Omega\}$ , onde  $T$  é o conjunto de índices e  $\Omega$  é o espaço amostral.*

Em um processo estocástico, para cada  $t$  fixado,  $X_t(\omega)$  representa uma única variável aleatória definida em  $\Omega$ . De forma semelhante, para cada  $\omega \in \Omega$  fixado,  $X_t(\omega)$  corresponde a uma função definida em  $T$ , denominada *caminho de amostra* ou *realização estocástica* do processo. Comumente, usa-se apenas a notação  $X_t$  ou  $X(t)$  para se referir a um processo estocástico, onde o índice  $t$  é geralmente interpretado como um parâmetro de tempo.  $X(t)$  é referido como o estado do processo no instante  $t$ . Dessa forma, o espaço  $\Omega$ , que é o conjunto de todos os possíveis valores que a variável aleatória  $X(t)$  pode tomar, é conhecido como espaço de estados.

Um processo estocástico pode ser classificado de acordo com os valores que ele assume ou com os valores que o seu parâmetro assume. Ele é discreto quando a variável aleatória toma valores em um conjunto discreto de pontos do eixo real, e contínuo quando a variável aleatória toma valores na união de vários intervalos de  $\mathbb{R}$ . Também podem ser classificados de acordo com o valor que o seu parâmetro assume. Se o parâmetro for discreto, ele assume valores em um subconjunto discreto de pontos reais, ou seja, um subconjunto contável. Se o parâmetro for contínuo ele assume valores em um subconjunto contínuo de  $\mathbb{R}$ , de tal forma que  $t \geq 0, t \in \mathbb{R}$ . Portanto, pode-se classificar um processo estocástico em quatro tipos: Processo Estocástico Discreto com parâmetro Discreto, Processo Estocástico Discreto com parâmetro Contínuo, Processo Estocástico Contínuo com parâmetro Discreto e Processo Estocástico Contínuo com parâmetro Contínuo [25].

Os métodos e técnicas utilizados para formular, analisar e simular computacionalmente as soluções de processos estocásticos dependem dos valores assumidos pelas

variáveis aleatórias e pelo conjunto de índices.

Existem vários processos estocásticos reconhecidos, como por exemplo o Processo de Wiener e Processo de Itô. Neste trabalho serão abordados os processos estocásticos denominados *Processos Markovianos*.

**Definição 2.14.** *Seja  $X(t)$  um processo estocástico e considere  $T$  o seu conjunto de índices. Um Processo Markoviano é um processo estocástico  $X(t)$  que satisfaz*

$$\begin{aligned} \text{Prob}(X(t_{k+1}) \leq x_{k+1} | X(t_k) = x_k, X(t_{k-1}) = x_{k-1}, \dots, X(t_0) = x_0) \\ = \text{Prob}(X(t_{k+1}) \leq x_{k+1} | X(t_k) = x_k), \end{aligned} \quad (21)$$

para  $t_j \in T$ , com  $j = 0, 1, 2, \dots, k+1$ , tal que  $t_0 < t_1 < \dots < t_k < t_{k+1}$  e  $x_j$  representa o estado que a variável aleatória pode assumir.

A expressão (21) caracteriza um processo no qual, para determinar a probabilidade de qualquer estado futuro é suficiente conhecer apenas a probabilidade do seu estado anterior. O seu passado não oferece nenhuma informação. Essa propriedade de poder desprezar a informação de todos os estados passados define um processo markoviano como sendo um *processo sem memória* (*memoryless process* em inglês).

Como visto anteriormente, um processo estocástico pode ser discreto ou contínuo, e pode possuir parâmetro discreto ou contínuo. Dado um processo markoviano  $X(t)$ , se as variáveis aleatórias  $X(t_j)$  assumem valores discretos, então o processo é denominado *Cadeia de Markov*. A depender de se o tempo é considerado discreto ou contínuo, uma Cadeia de Markov é classificada como Cadeia de Markov de Tempo Discreto, (*DTMC*<sup>2</sup>, pelas siglas em inglês), ou como Cadeia de Markov de Tempo Contínuo, (*CTMC*<sup>3</sup>).

---

<sup>2</sup> Discrete Time Markov Chain

<sup>3</sup> Continuous Time Markov Chain

### 3 CADEIAS DE MARKOV DE TEMPO CONTÍNUO

Serão definidas neste trabalho as Cadeias de Markov de Tempo Contínuo (*CTMC*), utilizadas para o modelo proposto. As definições aqui consideradas são análogas às que correspondem às Cadeias de Markov de Tempo Discreto.

**Definição 3.1.** *Seja o processo estocástico  $\{X_t(\omega) : t \in [0, \infty)\}, \omega \in \Omega$ , com  $\Omega$  um espaço discreto.  $X_t$  é uma Cadeia de Markov de Tempo Contínuo se, para alguma sequência de números reais  $0 \leq t_0 < t_1 < \dots < t_n < t_{n+1}$ , satisfaz a propriedade de Markov dada pela equação (21).*

Dado o processo  $X(t)$ , que pode tomar valores  $x \in \mathbb{Z}^+$ , vamos denotar por  $p_x(t)$  a probabilidade de se ter para  $X(t)$  no estado  $x$ , ou seja,

$$p_x(t) = \text{Prob}\{X(t) = x\}. \quad (22)$$

Dessa forma,  $\{p_x(t)\}_{x=0}^{\infty}$  é a distribuição de probabilidade dos estados acessíveis à variável aleatória  $X(t)$ , ou seja, a expressão (22) representa a probabilidade dos estados possíveis que a variável aleatória  $X(t)$  pode assumir durante o processo. A representação dessa distribuição de probabilidade pode ser feita através do vetor  $p(t) = (p_0(t), p_1(t), \dots)^T$ , onde  $T$  denota a transposta. No estudo do sistema, é possível relacionar os estados do processo para diferentes instantes de tempo. Essa relação é dada pela probabilidade condicional, também conhecida como *probabilidade de transição* entre estados.

**Definição 3.2.** *A probabilidade de transição de uma etapa (one-step transition probability), denotada por  $p_{yx}(t)$  é definida como a seguinte probabilidade condicional:*

$$p_{yx}(t) = \text{Prob}\{X(s+t) = y | X(s) = x\}, \quad s < s+t. \quad (23)$$

A expressão (23) representa a probabilidade da variável aleatória  $X$  no instante de tempo  $t$  encontrar-se no estado  $y$ , dado que no instante de tempo  $s$ , estava no estado  $x$ , para  $x, y = 0, 1, 2, \dots$ . As probabilidades de transição são importantes para determinar a probabilidade das mudanças possíveis de um estado para outro no sistema no decorrer do tempo. Em uma *CTMC*, se as probabilidades de transição  $p_{yx}$  não dependem explicitamente dos tempos  $s+t$  ou  $s$ , mas sim da diferença entre eles,  $s+t-s=t$ , elas são definidas como *probabilidades estacionárias* ou *homogêneas*.

As probabilidades de transição estacionárias são utilizadas quando o processo está em determinado estado no tempo  $s$  e deseja-se determinar a probabilidade do processo mover-se para outro estado no tempo  $s+t$ , sem que essa mudança de estado dependa do tempo mas dependa somente do estado atual do processo. Por conveniência, usaremos as

probabilidades de transição estacionárias no decorrer do trabalho. Ou seja,

$$p_{yx}(t) = Prob\{X(t) = y | X(0) = x\}. \quad (24)$$

As probabilidades de transição em uma etapa também podem ser expressas em forma matricial.

**Definição 3.3.** A matriz de transição ou matriz das probabilidades de transição de uma Cadeia de Markov de tempo contínuo  $X(t)$ , com  $t \in T$ ,  $T = [0, \infty)$  e espaço de estados  $\Omega = \{1, 2, \dots\}$ , e probabilidades de transição  $\{p_{yx}(t)\}_{x,y=1}^{\infty}$  é

$$P(t) = \begin{pmatrix} p_{11}(t) & p_{12}(t) & \cdots & p_{1y}(t) & \cdots \\ p_{21}(t) & p_{22}(t) & \cdots & p_{2y}(t) & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{y1}(t) & p_{y2}(t) & \cdots & p_{yy}(t) & \cdots \\ \vdots & \vdots & \cdots & \vdots & \ddots \end{pmatrix}. \quad (25)$$

Alternativamente, como notação matricial pode ser usada  $P(t) = (p_{yx}(t))$ .

Se o conjunto de estados possíveis do sistema for finito,  $\{1, 2, \dots, N\}$ , então a matriz de transição tem tamanho  $N \times N$ . A soma dos elementos de cada coluna da matriz é igual a 1. Essa característica define a matriz  $P(t)$  como uma *Matriz Estocástica*. Se além das colunas, a soma de cada linha também for igual a 1, então a matriz é denominada *Matriz Duplamente Estocástica*. A matriz de transição, dada pela Definição 3.3 é uma matriz estocástica para todo  $t \geq 0$ .

Vamos analisar algumas das propriedades características das probabilidades de transição,  $0 \leq p_{yx}(t) \leq 1$ . Dado que a matriz de transição definida em (25) é uma matriz estocástica, cada coluna  $x$ -ésima da matriz representa uma distribuição de probabilidade,

$$\sum_{y=1}^{\infty} p_{yx}(t) = 1, \quad t \geq 0. \quad (26)$$

A expressão (26) indica que há a probabilidade do processo em qualquer estado  $x$  ser transferido para um outro estado  $y$ , com  $x \neq y$  ou permanecer no mesmo estado  $x$  no próximo intervalo de tempo. Quando essa propriedade não é satisfeita tem-se um caso específico, denominado processo explosivo. Um processo explosivo pode ocorrer quando o espaço de estados do sistema é infinito. Em particular, vamos considerar que todas as CTMC's possuem espaço de estados finito, ou seja, são não explosivas.

Além da propriedade representada pela expressão (26), tem-se que as probabilidades de transição são soluções das equações de Chapman-Kolmogorov. Essas equações indicam que a probabilidade de transição de um estado para outro pode ser obtida através das probabilidades de transição de estados intermediários.

Suponha que a probabilidade de transição do estado  $x$  para o estado  $y$ , com  $s, t \in [0, \infty)$ , é definida da seguinte forma:

$$\begin{aligned} p_{yx}(t+s) &= \text{Prob}\{X(t+s) = y | X(0) = x\}, \\ &= \sum_{k=0}^{\infty} \text{Prob}\{X(t+s) = y, X(t) = k | X(0) = x\}. \end{aligned} \quad (27)$$

Usando a propriedade  $\text{Prob}(A \cap B | C) = \text{Prob}(A | B \cap C) \text{Prob}(B | C)$ , dada pela equação (9), tem-se que

$$p_{yx}(t+s) = \sum_{k=0}^{\infty} \text{Prob}\{X(t+s) = y | X(t) = k, X(0) = x\} \text{Prob}\{X(t) = k | X(0) = x\}, \quad (28)$$

e considerando a condição do processo ser de Markov,

$$p_{yx}(t+s) = \sum_{k=0}^{\infty} p_{yk}(s) p_{kx}(t). \quad (29)$$

A expressão (29) é conhecida como equações de Chapman-Kolmogorov. Em forma matricial, podem ser escritas como

$$P(s)P(t) = P(t+s). \quad (30)$$

### 3.1 Classificação de Estados

Os estados de uma CTMC podem ser classificados de acordo com as relações possíveis entre esses estados e, conseqüentemente, fornecem uma classificação para a Cadeia de Markov.

**Definição 3.4.** Um estado  $y$  é **alcançável** por um estado  $x$  se existe uma probabilidade não negativa, tal que  $p_{yx}(t) > 0$ , para algum  $t \geq 0$ . Essa relação é denotada por  $x \rightarrow y$  ( $x$  alcança  $y$ ).

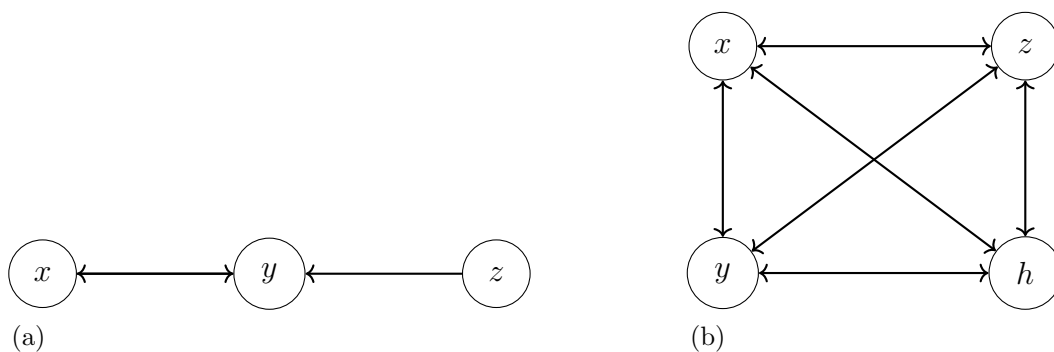
Ou seja, um estado  $y$  é alcançável a partir de um estado  $x$  se é possível o sistema chegar ao estado  $y$  dado que em algum momento anterior estava no estado  $x$ .

**Definição 3.5.** Se o estado  $y$  é alcançável pelo estado  $x$ , então  $x \rightarrow y$ , e se o estado  $x$  também é alcançável pelo estado  $y$ , então  $y \rightarrow x$ . Esses estados  $x$  e  $y$  são chamados de

**comunicantes.** Essa relação é denotada por  $x \leftrightarrow y$ . Ou seja, existem  $t \geq 0$  e  $s \geq 0$  tais que  $p_{yx}(t) > 0$  e  $p_{xy}(s) > 0$ .

O conjunto de todos os estados que são comunicantes é chamado de classe de comunicação. Pode-se ter mais de uma classe de comunicação em uma Cadeia de Markov. As relações entre os estados em uma Cadeia de Markov podem ser representadas por um grafo direcionado, onde os nós desse grafo representam os estados e as arestas representam a transição possível entre esses estados. Se todos os estados do sistema podem ser alcançados partindo-se de qualquer outro estado, a cadeia de Markov é denominada **irredutível**, ou seja todos os estados do sistema são comunicantes entre si, caso contrário, é denominada **redutível**. Na Figura 1, ilustram-se exemplos de cadeias de Markov redutível e irreductível e as relações entre os estados.

Figura 1 - Grafo direcionado representando as relações entre os estados de uma Cadeia de Markov.



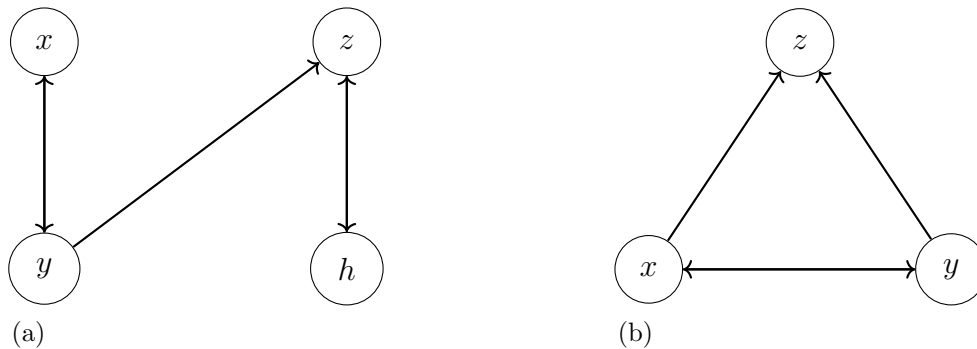
Legenda: (a) Existem  $t, t', r$  e  $s \geq 0$  tais que  $p_{zy}(s) > 0$ ,  $p_{yx}(t) > 0$ ,  $p_{xy}(t') > 0$ ,  $p_{zx}(r) > 0$ . Os estados  $x$  e  $y$  são comunicantes e são alcançáveis pelo estado  $z$  que, por sua vez, não é alcançável por nenhum outro estado. Esse é um exemplo de Cadeia de Markov redutível. (b) Neste exemplo todos os estados são comunicantes entre si; portanto, a Cadeia de Markov é irreductível.

Fonte: A autora, 2020.

Um conjunto de estados  $A$  é denominado *fechado* se for impossível alcançar qualquer estado fora desse conjunto partindo de um estado dentro dele. Dessa forma, se  $A$  for um conjunto fechado,  $p_{yx}(t) = 0$  para  $t \geq 0$  se  $x \in A$  e  $y \notin A$ .

Os estados também podem ser especificados como *transientes*, *recorrentes* e *absorventes*. Um estado é *transiente* quando há uma probabilidade não nula de que, após sair desse estado, não é mais possível voltar a ele. Em contraste, no estado *recorrente* sempre é possível voltar a esse mesmo estado após ter passado por ele. Por fim, um estado  $x$  é *absorvente* quando, ao entrar nesse estado, a probabilidade de sair dele é igual a zero, ou seja,  $p_{yx}(t) = 0$  para todo  $y \neq x$ , com  $t \geq 0$ . Exemplos de tais estados são ilustrados na

Figura 2 - Grafo direcionado representando as classificações dos estados em uma Cadeia de Markov.



Legenda: (a) Os estados  $x$  e  $y$  são estados transitientes e os estados  $z$  e  $h$  são recorrentes. (b) Os estados  $x$  e  $y$  são transitientes e o estado  $z$  é absorvente.

Fonte: A autora, 2020.

Figura 2.

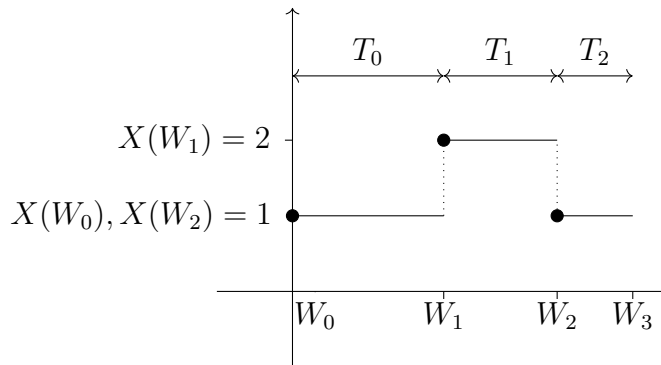
### 3.2 Tempo Entre Eventos

Em uma *CTMC*, o tempo que o processo leva para ir de um estado a outro, pode tomar um valor real  $t \geq 0$  qualquer, diferentemente de uma *DTMC*, onde esse “salto” para um novo estado ocorre em instantes discretos  $(1, 2, \dots)$ . O *tempo de salto*, também denominado *tempo de espera* é uma variável aleatória, denotando-se por  $\{W_n\}_{n=0}^{\infty}$ , onde  $W_n$  é o tempo do  $n$ -ésimo salto. Como exemplo, vamos considerar uma Cadeia de Markov de Tempo Contínuo, com estado inicial  $X(W_0) = x$ , onde  $W_0 = 0$ . O processo permanece no estado inicial  $x$  por um período de tempo aleatório  $W_1$  até passar para um novo estado  $X(W_1) = y$ . Em seguida, ele permanece no estado  $y$  por um novo período de tempo aleatório  $W_2$  até alcançar um novo estado  $X(W_2) = z$ , e assim sucessivamente.

Para realizar simulações estocásticas em uma *CTMC* é necessário conhecer a distribuição do tempo entre sucessivos eventos, denominado *tempo de permanência* ou *tempo entre eventos*. Temos então uma outra variável aleatória para o tempo entre eventos, denotada por  $T_n$ , contínua e não-negativa, de modo que  $T_n = W_{n+1} - W_n \geq 0$ , onde  $W_n$  é o tempo do  $n$ -ésimo salto. Na Figura 3, pode-se observar graficamente essa relação entre o tempo de espera  $W_n$  e o tempo entre eventos  $T_n$ .

Supondo que o processo esteja no estado  $x$  no  $n$ -ésimo salto, ou seja,  $X(W_n) = x$ , define-se a probabilidade do processo mover-se para um estado diferente de  $x$ , em um

Figura 3 - Ilustração do tempo entre eventos.



Legenda: Na figura observa-se o tempo de espera de um estado para outro  $W_n$  e o tempo entre eventos  $T_n$ , para  $n = 0, 1, 2, 3$ , onde  $X(W_0) = 1$ ,  $X(W_1) = 2$  e  $X(W_2) = 1$ .

Fonte: A autora, 2020.

período curto de tempo  $\Delta t$ , por  $\alpha(x)\Delta t + o(\Delta t)$ . Isto é,

$$\sum_{y \neq x} p_{yx}(\Delta t) = \alpha(x)\Delta t + o(\Delta t) \quad (31)$$

Dessa forma, a probabilidade de que o processo não altere do estado  $x$  para qualquer outro estado do sistema, ou seja, que permaneça no estado  $x$ , é dada por:

$$p_{xx}(\Delta t) = 1 - \alpha(x)\Delta t + o(\Delta t). \quad (32)$$

A probabilidade do processo permanecer no estado  $x$  por um período de tempo  $t$  (que também pode ser representado por  $[W_n, W_n + t]$ ), denotada por  $S_n(t)$  pode ser expressa por meio do tempo entre eventos  $T_n$  da seguinte forma,

$$S_n(t) = \text{Prob}\{W_{n+1} > W_n + t\} = \text{Prob}\{W_{n+1} - W_n > t\} = \text{Prob}\{T_n > t\}. \quad (33)$$

Considerando que o estado  $x$  não é um estado absorvente, ou seja, existe uma probabilidade para que o processo mude de estado, para  $t = 0$  tem-se

$$S_n(0) = \text{Prob}\{T_n > 0\} = 1. \quad (34)$$

Para um  $\Delta t$  suficientemente pequeno, utilizando a equação (32), tem-se

$$S_n(t + \Delta t) = S_n(t)p_{xx}(\Delta t) = S_n(t)(1 - \alpha(x)\Delta t + o(\Delta t)). \quad (35)$$

Assume-se  $S_n(t + \Delta t) = S_n(t)p_{xx}(\Delta t)$  a partir da propriedade de “perda de me-



mória”, inerente da distribuição exponencial. Por este fato, os processos Markovianos possuem um tempo entre eventos que é exponencial. Subtraindo  $S_n(t)$  em ambos os lados da expressão (35) e dividindo por  $\Delta t$ ,

$$\frac{S_n(t + \Delta t) - S_n(t)}{\Delta t} = \frac{S_n(t)(1 - \alpha(x)\Delta t + o(\Delta t)) - S_n(t)}{\Delta t}. \quad (36)$$

Fazendo o limite com  $\Delta t \rightarrow 0$ ,

$$\lim_{\Delta t \rightarrow 0} \frac{S_n(t + \Delta t) - S_n(t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{S_n(t)(1 - \alpha(x)\Delta t + o(\Delta t)) - S_n(t)}{\Delta t}, \quad (37)$$

chega-se à equação

$$\frac{dS_n(t)}{dt} = -\alpha(x)S_n(t). \quad (38)$$

Resolvendo a equação (38), tem-se

$$\int \frac{dS_n(t)}{S_n(t)} = \int -\alpha(x)dt \implies \ln(S_n(t)) = -\alpha(x)t + C. \quad (39)$$

Utilizando agora a condição inicial, definida pela expressão (34), é possível, por fim, obter a seguinte expressão para a probabilidade  $S_n(t)$ ,

$$Prob\{T_n > t\} = S_n(t) = e^{-\alpha(x)t}. \quad (40)$$

Com isso, a probabilidade para  $T_n \leq t$ , com  $t \geq 0$ , é dada por

$$Prob\{T_n \leq t\} = 1 - S_n(t) = 1 - e^{-\alpha(x)t}. \quad (41)$$

Dessa forma, a função de distribuição acumulada para o tempo entre eventos  $T_n$  é definida por  $F_n(t) = 1 - e^{-\alpha(x)t}$ , e pode-se concluir que  $T_n$  é uma variável aleatória com distribuição exponencial, cujo parâmetro é dado por  $\alpha(x)$ .

Uma relação probabilística muito importante acerca do conjunto dos tempos entre eventos,  $\{T_n\}_{n=0}^{\infty}$ , é que esses tempos são independentes se condicionados aos estados sucessivos que a Cadeia de Markov visita. Em um dado estado  $X(W_n) = x$ , o tempo entre eventos  $T_n$  é independente do tempo  $T_{n-1}$ .

A variável aleatória  $T_n$  pode ser expressa em termos da função de distribuição acumulada  $F_n(t)$  juntamente com uma variável aleatória uniforme, que denotaremos por  $U$ .

**Teorema 3.1.** *Seja  $U$  uma variável aleatória uniforme definida em  $[0, 1]$  e  $T$  uma variável aleatória contínua definida em  $[0, \infty)$ . Então  $T = F^{-1}(U)$ , onde  $F$  é a distribuição acumulada da variável aleatória  $T$ .*

*Demonstração.* Seja  $F(t) = \text{Prob}\{T \leq t\}$ . Deseja-se mostrar que  $F(t) = \text{Prob}\{F^{-1}(U) \leq t\}$ . A função  $F$  é estritamente crescente,  $F : [0, \infty) \rightarrow [0, 1)$ , portanto sua inversa  $F^{-1}$  existe. Além disso, tem-se que  $t \in [0, \infty)$  e

$$\text{Prob}\{F^{-1}(U) \leq t\} = \text{Prob}\{F(F^{-1}(U)) \leq F(t)\} = \text{Prob}\{U \leq F(t)\}. \quad (42)$$

Como  $U$  é uma variável aleatória uniforme, tem-se que  $\text{Prob}\{U \leq t\} = t$  para  $t \in [0, 1]$ . Portanto,  $\text{Prob}\{U \leq F(t)\} = F(t)$ .  $\square$

Como  $T_n$  é independente de outro tempo entre eventos,  $\alpha$  é independente do estado do processo. Portanto, o tempo entre eventos é o mesmo para todo salto  $n$ ,  $T_n \equiv T_E$ . A seguir, vamos procurar uma expressão para o tempo entre eventos  $T_E$ .

Vamos expressar o tempo entre eventos  $T_E$  em função da variável aleatória uniforme  $U$ , como visto no Teorema 3.1. Como  $F(T_E) = 1 - e^{-\alpha(x)T_E}$ , calcula-se a inversa de  $F(T_E)$ , e assim obtem-se:

$$\begin{aligned} e^{-\alpha(x)T_E} &= 1 - F(T_E), \\ \ln e^{-\alpha(x)T_E} &= \ln(1 - F(T_E)), \\ -\alpha(x)T_E &= \ln(1 - F(T_E)). \end{aligned}$$

Isolando a variável  $T_E$  e fazendo  $F(T_E) = U$ , tem-se:

$$T_E = -\frac{\ln(1 - F(T_E))}{\alpha(x)} = -\frac{\ln(1 - U)}{\alpha(x)}. \quad (43)$$

Como  $U$  é uma variável aleatória uniforme no intervalo  $[0, 1]$ , o mesmo acontece com  $1 - U$  e, portanto,

$$T_E = F^{-1}(U) = -\frac{\ln U}{\alpha(x)}, \quad (44)$$

obtendo-se, por fim, uma expressão para o tempo entre eventos  $T_E$ .

O algoritmo de simulação estocástica usado para calcular os caminhos de amostra de processos simples aplica o que é conhecido como método direto em aplicações bioquímicas. No método direto, duas variáveis aleatórias uniformes são necessárias por iteração; a primeira para simular o tempo até o próximo evento e a segunda para escolher o evento que ocorrerá. [9].

### 3.3 Matriz Geradora $G$

Para representar se houve ou não mudança de um estado para outro são utilizadas as taxas de transição  $g_{yx}$ , que são obtidas através das derivadas das probabilidades de transição  $p_{yx}$ . Essas taxas de transição também podem ser expressas em forma matricial, e essa forma é conhecida como *Matriz Geradora Infinitesimal*  $G$ .

Pode-se definir uma relação entre essas taxas de mudança das probabilidades de transição a partir da matriz  $G$ . Inicialmente, assume-se que as probabilidades de transição são contínuas e diferenciáveis para  $t \geq 0$ . Também considera-se que no instante inicial não há probabilidade de transição para outro instante, ou seja, quando  $t = 0$ , as probabilidades de transição são  $p_{yx}(0) = 0$ ,  $y \neq x$  e  $p_{xx}(0) = 1$ . Esta última representa que o estado permanece nele mesmo.

Dessa forma, pode-se definir a taxa de transição de um estado para outro como sendo

$$g_{yx} = \lim_{\Delta t \rightarrow 0^+} \frac{p_{yx}(\Delta t) - p_{yx}(0)}{\Delta t} = \lim_{\Delta t \rightarrow 0^+} \frac{p_{yx}(\Delta t)}{\Delta t}, \quad x \neq y, \quad (45)$$

e também a taxa de transição para permanecer no mesmo estado,

$$g_{xx} = \lim_{\Delta t \rightarrow 0^+} \frac{p_{xx}(\Delta t) - p_{xx}(0)}{\Delta t} = \lim_{\Delta t \rightarrow 0^+} \frac{p_{xx}(\Delta t) - 1}{\Delta t}. \quad (46)$$

Pela equação (26) das probabilidades de transição,  $\sum_{y=1}^{\infty} p_{yx}(t) = 1$ , tem-se que

$$p_{xx}(\Delta t) + \sum_{y=0, y \neq x}^{\infty} p_{yx}(\Delta t) = 1, \quad (47)$$

e utilizando a expressão (45) chega-se a

$$1 - p_{xx}(\Delta t) = \sum_{y=0, y \neq x}^{\infty} p_{yx}(\Delta t) = \sum_{y=0, y \neq x}^{\infty} [g_{yx} \Delta t + o(\Delta t)]. \quad (48)$$

Simplificando,

$$g_{xx} = \lim_{\Delta t \rightarrow 0^+} \frac{-\sum_{y=0, y \neq x}^{\infty} [g_{yx} \Delta t + o(\Delta t)]}{\Delta t} = -\sum_{y=0, y \neq x}^{\infty} g_{yx}, \quad (49)$$

onde assumiu-se que  $\sum_{y \neq x} o(\Delta t) = o(\Delta t)$ .

Considerando  $P(\Delta t) = (p_{yx}(\Delta t))$  a matriz de transição infinitesimal e  $I$  uma matriz da mesma dimensão de  $P(\Delta t)$  com os elementos da diagonal iguais a 1 e o restante igual a zero (uma matriz identidade no caso finito), pode-se definir a matriz  $G$  da seguinte forma

$$G = \lim_{\Delta t \rightarrow 0^+} \frac{P(\Delta t) - I}{\Delta t}. \quad (50)$$

Utilizando a expressão (49), pode-se representar a matriz  $G$  na forma matricial

$$G = \begin{pmatrix} g_{00} & g_{01} & g_{02} & \cdots \\ g_{10} & g_{11} & g_{12} & \cdots \\ g_{20} & g_{21} & g_{22} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} = \begin{pmatrix} -\sum_{y=1}^{\infty} g_{y0} & g_{01} & g_{02} & \cdots \\ g_{10} & -\sum_{y=0, y \neq 1}^{\infty} g_{y1} & g_{12} & \cdots \\ g_{20} & g_{21} & -\sum_{y=0, y \neq 2}^{\infty} g_{y2} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

que coincide com o limite tomado na expressão (50).

A matriz  $G$  também pode ser denominada como matriz de taxas de transição, matriz infinitesimal ou simplesmente matriz geradora. Ela possui as seguintes propriedades: cada coluna tem a soma igual a zero e o  $y$ -ésimo elemento da diagonal é o oposto da soma dos elementos não diagonais da sua coluna.

A matriz geradora é uma ferramenta muito importante em uma *CTMC* pois através dela podem-se definir as equações avançadas de Kolmogorov, que serão definidas mais a frente. Através das equações de Kolmogorov é possível expressar a matriz de transição  $P(t)$  por meio de equações diferenciais e encontrar a matriz de transição de uma cadeia de Markov incorporada. Esta última é utilizada para definir a transição de um estado para outro sem depender do tempo, apenas das taxas de transição entre eles.

**Definição 3.6.** *Seja  $Y_n$  a variável aleatória para o estado de uma CTMC  $\{X(t) : t \in [0, \infty)\}$  no  $n$ -ésimo salto, tal que*

$$Y_n = X(W_n), \quad n = 0, 1, 2, \dots$$

*O conjunto de variáveis aleatórias  $\{Y_n\}_{n=0}^{\infty}$  é denominado Cadeia de Markov Incorporada (EMC<sup>4</sup>, pelas siglas em inglês).*

Seja  $\{Y_n\}_{n=0}^{\infty}$  uma cadeia de Markov Incorporada e sua matriz de transição definida por  $P_Y = (\tilde{p}_{yx})$ , onde  $\tilde{p}_{yx} = Prob\{Y_{n+1} = y | Y_n = x\}$ . A matriz de transição  $P_Y = (\tilde{p}_{yx})$  também pode ser definida por meio da matriz geradora  $G$ .

---

<sup>4</sup> Embedded Markov Chain

Um estado  $x$ , em uma cadeia de tempo contínuo, é absorvente se  $g_{xx} = 0$ , ou seja, a taxa de mudança é zero pois o estado não muda. Portanto,

$$\tilde{p}_{xx} = \begin{cases} 0, & \text{se } g_{xx} \neq 0 \\ 1, & \text{se } g_{xx} = 0 \end{cases}. \quad (51)$$

Para definirmos a probabilidade de transição  $\tilde{p}_{yx}$  utilizamos as expressões (45) e (46),

$$-\frac{g_{yx}}{g_{xx}} = \lim_{\Delta t \rightarrow 0^+} \frac{p_{yx}(\Delta t)}{1 - p_{yx}(\Delta t)}. \quad (52)$$

A expressão (52) é a probabilidade de transferência do estado  $x$  para o estado  $y$ , dado que o processo não permanece no estado  $x$  ( $g_{xx} \neq 0$ ). Portanto, define-se a probabilidade de transição  $\tilde{p}_{yx}$  como

$$\tilde{p}_{yx} = \begin{cases} \frac{g_{yx}}{\sum_{k=0, k \neq x}^{\infty} g_{kx}} = -\frac{g_{yx}}{g_{xx}}, & \text{se } g_{xx} \neq 0 \\ 0, & \text{se } g_{xx} = 0 \end{cases}. \quad (53)$$

A representação matricial da EMC  $P_Y = (\tilde{p}_{yx})$ , onde os elementos são definidos pelas expressões (51) e (53) e, em particular, para  $g_{xx} \neq 0$ ,  $x = 0, 1, 2, \dots$ , é

$$P_Y = \begin{pmatrix} 0 & -\frac{g_{01}}{g_{11}} & -\frac{g_{02}}{g_{22}} & \dots \\ -\frac{g_{10}}{g_{00}} & 0 & -\frac{g_{12}}{g_{22}} & \dots \\ -\frac{g_{20}}{g_{00}} & -\frac{g_{21}}{g_{11}} & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (54)$$

No caso de  $g_{xx} = 0$ , o elemento  $(x, x)$  da matriz  $P_Y$  é igual a 1 e os elementos restantes da coluna são iguais a 0, por essa razão a matriz  $P_Y$  é uma matriz estocástica.

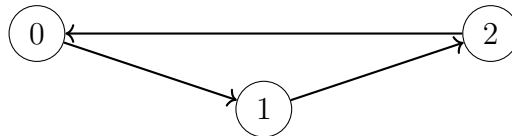
**Exemplo 3.3.1.** *Suponha que uma CTMC que possui os estados 0, 1 e 2 tenha a seguinte matriz geradora  $G$ :*

$$G = \begin{pmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}$$

A matriz de transição correspondente à EMC é dada por:

$$P = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

Dessa forma, por meio do grafo direcionado, pode-se observar como os estados se comunicam nessa CTMC.



■

As probabilidades de transição em uma EMC são homogêneas, ou seja, independem do salto  $n$ . Além disso,  $P_Y^n = (\tilde{p}_{yx}^{(n)})$ , onde  $\tilde{p}_{yx}^{(n)} = Prob\{Y_n = y | Y_0 = x\}$ .

### 3.4 Equações Diferenciais de Kolmogorov

Expressões para a taxa de variação das probabilidades de transição podem ser encontradas através das equações diferenciais de Kolmogorov avançadas (*forward*) e regressivas (*backward*). As equações avançadas, também conhecidas como “equações mestras”, na física e na química, são utilizadas para prever a dinâmica futura do sistema, enquanto as equações regressivas são utilizadas para estimar a probabilidade de atingir um estado de absorção [26]. A continuação, vamos deduzir essas equações.

A probabilidade de transição  $p_{yx}(t + \Delta t)$  pode ser expandida aplicando as equações de Chapman-Kolmogorov. Subtraindo  $p_{yx}(t)$  em ambos os lados da expressão (29), obtém-se

$$p_{yx}(t + \Delta t) - p_{yx}(t) = \sum_{k=0}^{\infty} p_{yk}(\Delta t)p_{kx}(t) - p_{yx}(t). \quad (55)$$

Manipulando a equação (55), chega-se à expressão,

$$p_{yx}(t + \Delta t) - p_{yx}(t) = \sum_{k \neq y}^{\infty} p_{yk}(\Delta t)p_{kx}(t) + p_{yy}(\Delta t)p_{yx}(t) - p_{yx}(t), \quad (56)$$

$$p_{yx}(t + \Delta t) - p_{yx}(t) = \sum_{k \neq y}^{\infty} p_{yk}(\Delta t)p_{kx}(t) - p_{yx}(t)(1 - p_{yy}(\Delta t)). \quad (57)$$

Dividindo por  $\Delta t$  a expressão (57), e fazendo  $\Delta t \rightarrow 0$ ,

$$\lim_{\Delta t \rightarrow 0} \frac{p_{yx}(t + \Delta t) - p_{yx}(t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \left( \frac{\sum_{k \neq x}^{\infty} p_{yk}(\Delta t) p_{kx}(t)}{\Delta t} - \frac{p_{yx}(t)(1 - p_{yy}(\Delta t))}{\Delta t} \right). \quad (58)$$

Por fim, usando-se as expressões (45) e (46), obtém-se

$$\frac{dp_{yx}(t)}{dt} = \sum_{k \neq x}^{\infty} g_{yk} p_{kx}(t) - p_{yx}(t) g_{yy}, \quad x, y = 0, 1, \dots \quad (59)$$

**Definição 3.7.** *O sistema de equações (59) é conhecido como equações diferenciais avançadas de Kolmogorov. Expressando em forma matricial, tem-se*

$$\frac{dP(t)}{dt} = GP(t), \quad (60)$$

onde  $P(t) = (p_{yx}(t))$  é a matriz das probabilidades de transição e  $G = (g_{yx})$  é a matriz geradora.

No caso em que a distribuição inicial do processo é  $X(0) = k$  ( $p_x(0) = \delta_{xk}$ ), então a probabilidade de transição  $p_{xk}(t)$  é a mesma que as probabilidades de estado  $p_x(t) = Prob\{X(t) = x | X(0) = k\}$ . Portanto, nesse caso, as probabilidades de estado são soluções das equações diferenciais avançadas de Kolmogorov

$$\frac{dp(t)}{dt} = Gp(t), \quad (61)$$

onde  $p(t) = (p_0(t), p_1(t), \dots)^T$ .

As equações diferenciais de Kolmogorov regressivas podem ser derivadas de maneira semelhante às equações de Kolmogorov avançadas, usando as equações de Chapman-Kolmogorov e considerando a seguinte expressão,

$$p_{yx}(\Delta t + t) = \sum_{k=0}^{\infty} p_{yk}(t) p_{kx}(\Delta t) = \sum_{k \neq x}^{\infty} p_{yk}(t) p_{kx}(\Delta t) - p_{yx}(t)(1 - p_{xx})(\Delta t). \quad (62)$$

Através de manipulações e simplificações semelhantes às empregadas acima na derivação das equações avançadas, produzem o seguinte sistema de equações diferenciais

$$\frac{dp_{yx}(t)}{dt} = \sum_{k \neq x}^{\infty} p_{yk}(t) g_{kx} - p_{yx}(t) g_{xx}, \quad x, y = 0, 1, \dots \quad (63)$$

**Definição 3.8.** *O sistema de equações expressas em (63) é conhecido como equações*

diferenciais de Kolmogorov regressivas. Expressando em forma matricial, tem-se

$$\frac{dP(t)}{dt} = P(t)G, \quad (64)$$

onde  $P(t) = (p_{yx}(t))$  é a matriz das probabilidades de transição e  $G = (g_{yx})$  é a matriz geradora.

As equações diferenciais de Kolmogorov regressivas são úteis em problemas de tempo de primeira passagem, como a distribuição do tempo para atingir um estado específico ou a probabilidade de atingir um estado específico pela primeira vez. Esses dois problemas dependem do estado inicial do processo.

Essas equações diferenciais dependem da existência da matriz geradora  $G$ . Para sistemas de dimensão finita ou cadeias de Markov finitas,  $G$  sempre existe. A solução  $P(t)$  pode ser encontrada através das equações avançadas ou regressivas.

### 3.5 Primeiro Tempo de Passagem

O tempo até o processo atingir um determinado estágio pela primeira vez é conhecido como o “primeiro tempo de passagem”. Suponha que o processo esteja no estado  $a$  e queremos encontrar o tempo que leva até atingir o estado  $b$ , onde  $a < b$  ou  $b < a$ . Os problemas de tempo de primeira passagem estão relacionados à equação diferencial regressiva de Kolmogorov.

Seja  $T_{x+1,x}$  a variável aleatória para o tempo que leva para ir do estado  $x$  para  $x+1$ . Derivando o tempo entre eventos, sabemos que a pdf para ele tem uma distribuição exponencial com o parâmetro  $\alpha(x)$ , ( $X(0) = x$ ). Dessa forma,

$$f_x(t) = \alpha(x)e^{-\alpha(x)t}. \quad (65)$$

Assim, o tempo esperado para ir do estado  $x$  para  $x+1$  ou  $x-1$  é a média da distribuição exponencial,  $1/(\alpha(x))$ . Se  $a < b$ , o tempo que leva para ir do estado  $a$  para  $b$  é

$$T_{b,a} = T_{a+1,a} + T_{a+2,a+1} + \cdots + T_{b,b-1}, \quad (66)$$

e o tempo esperado para ir do estado  $a$  para  $b$  é

$$\mathbb{E}(T_{b,a}) = \mathbb{E}(T_{a+1,a}) + \mathbb{E}(T_{a+2,a+1}) + \cdots + \mathbb{E}(T_{b,b-1}). \quad (67)$$

Expressões semelhantes podem ser derivadas para  $a > b$ . Estes são tempos médios de primeira passagem. Expressões gerais para  $\mathbb{E}(T_{x,x+1})$  e  $\mathbb{E}(T_{x+1,x})$  são derivadas a seguir.



Suponha que o processo esteja no estado  $x$ . Após um período exponencial de tempo, o processo salta de  $x$  para  $x + 1$  com probabilidade  $c_x/(\alpha(x))$  e para o estado  $x - 1$  com probabilidade  $d_x/(\alpha(x))$ , considerando  $c_x + d_x = \alpha(x)$ . Para encontrar o tempo esperado de ir do estado  $x$  para  $x + 1$ , devemos considerar que o processo pode saltar para  $x - 1$ . Então, o tempo esperado para voltar a  $x + 1$  deve ser adicionado a esse tempo.

$$\mathbb{E}(T_{x+1,x}) = \frac{c_x}{\alpha(x)} \left( \frac{1}{\alpha(x)} \right) + \frac{d_x}{\alpha(x)} \left( \frac{1}{\alpha(x)} + \mathbb{E}(T_{x,x-1}) + \mathbb{E}(T_{x+1,x}) \right). \quad (68)$$

Simplificando a expressão (68), obtem-se

$$\mathbb{E}(T_{x+1,x}) = \frac{1}{c_x} + \frac{d_x}{c_x} \mathbb{E}(T_{x,x-1}). \quad (69)$$

Pode-se obter uma relação semelhante para  $\mathbb{E}(T_{x-1,x})$ :

$$\mathbb{E}(T_{x-1,x}) = \frac{d_x}{\alpha(x)} \left( \frac{1}{\alpha(x)} \right) + \frac{c_x}{\alpha(x)} \left( \frac{1}{\alpha(x)} + \mathbb{E}(T_{x,x+1}) + \mathbb{E}(T_{x-1,x}) \right). \quad (70)$$

Simplificando (70),

$$\mathbb{E}(T_{x-1,x}) = \frac{1}{d_x} + \frac{c_x}{d_x} \mathbb{E}(T_{x,x+1}). \quad (71)$$

Suponha  $c_x = 0 = d_x$  e  $\lim_{t \rightarrow \infty} p_0(t) = 1$ . Considerando  $\mathbb{E}(T_{0,x}) = \tau_x$  como o tempo esperado para a absorção a partir do estado  $x$ , é possível obter uma fórmula para calcular esse tempo  $\tau_x$ . Para um intervalo de tempo pequeno  $\Delta t$ , tem-se

$$\tau_x = c_x \Delta t (\tau_{x+1} + \Delta t) + d_x \Delta t (\tau_{x-1} + \Delta t) + [1 - (\alpha(x)) \Delta t] (\tau_x + \Delta t) + o(\Delta t). \quad (72)$$

Dividindo (72) por  $\Delta t$ , e fazendo  $\Delta t \rightarrow 0$  obtem-se o seguinte resultado

$$\tau_x = \frac{1}{\alpha(x)} + \frac{c_x}{\alpha(x)} \tau_{x+1} + \frac{d_x}{\alpha(x)} \tau_{x-1}. \quad (73)$$

A expressão (73) pode ser expressa da seguinte forma,

$$d_x \tau_{x-1} + c_x \tau_{x+1} - (\alpha(x)) \tau_x = -1 \quad x = 1, 2, \dots \quad (74)$$

Se a *CTMC* é finita, com estados  $\{0, 1, 2, \dots, N\}$ , de modo que  $c_N = 0$ , já que não há mais transição após o estado  $N$ , então para  $x = 1, 2, \dots, N$ , o sistema de equações (74) pode ser representado pela seguinte forma matricial,

$$\tau^T \tilde{G} = -C^T, \quad (75)$$

onde  $C$  é um vetor tal que  $C = [1, 1, \dots, 1]^T$  e a matriz  $\tilde{G}$  é a matriz geradora  $G$  eliminando as linhas e colunas dos estados que são absorventes.

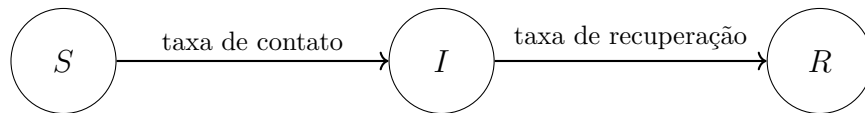
## 4 MODELO COMPARTIMENTAL SIR

A estratégia usual na modelagem é considerar inicialmente o modelo mais simples possível, mesmo que algumas características do contágio real não sejam levadas em conta. A depender do resultado, novos elementos podem ser introduzidos controladamente no modelo e uma avaliação do ganho em função da complexidade deverá ser considerada. Embora simples, o modelo SIR [5], que divide a população nos estados de (S)uscetíveis, (I)nfectados e (R)ecuperados, tem sido usado amplamente no estudo de diversas doenças [27, 28, 29], inclusive, da COVID-19 [30, 31, 32].

### 4.1 Modelo SIR

O modelo utilizado neste trabalho é o modelo *SIR*, proposto por Kermack e McKendrick [5], um modelo compartimental que relaciona os seguintes compartimentos: *S* para os indivíduos suscetíveis, *I* para os infectados e *R* para os recuperados. O esquema da Figura 4 representa a sua dinâmica.

Figura 4 - Diagrama do modelo compartimental *SIR*.



Legenda: *S* representa o compartimento dos indivíduos suscetíveis, *I*, dos infectados e *R*, dos recuperados.

Fonte: A autora, 2020.

Um indivíduo do compartimento suscetível (**S**), através da taxa de contato, torna-se um indivíduo do compartimento infectado (**I**). Um indivíduo infectado (**I**) passa ao compartimento recuperado (**R**) através de uma taxa de recuperação. A taxa de recuperação representa o inverso do período que a doença pode ser transmitida.

## 4.2 Modelo Determinístico

O modelo determinístico é apresentado através das seguintes equações diferenciais,

$$\frac{dS}{dt} = -\frac{\beta}{N}SI, \quad (76)$$

$$\frac{dI}{dt} = \frac{\beta}{N}SI - \gamma I, \quad (77)$$

$$\frac{dR}{dt} = \gamma I. \quad (78)$$

Será considerado que não há nascimentos ou mortes na população estudada. Dessa forma, a população permanece constante, ou seja,  $S(t) + I(t) + R(t) = N$  considerando que, inicialmente,  $S(0), I(0) > 0$  e  $R(0) = 0$ . A taxa de contato, também conhecida como taxa de infecção, é representada por  $\beta$ . A taxa de recuperação de um indivíduo infectado é representada por  $\gamma$ . Vamos considerar que não há período latente nesse modelo, dessa forma tem-se que um indivíduo infectado é infeccioso. A equação (76) representa a variação de indivíduos suscetíveis que ao ter contato com indivíduos infectados tornam-se infectados. Em (77) é representada a variação de indivíduos infectados, ou seja, os indivíduos que eram suscetíveis e tornaram-se infectados menos os indivíduos que se recuperaram (tornaram-se imunes ou morreram). Na equação (78) tem-se a variação dos indivíduos infectados que se recuperaram.

Combinando as equações (76) e (77) pode-se encontrar uma solução analítica para determinar a quantidade de indivíduos infectados. Dessa forma,

$$\frac{dI}{dS} = \frac{\frac{\beta}{N}SI - \gamma I}{-\frac{\beta}{N}SI}, \quad (79)$$

$$\frac{dI}{dS} = \frac{\frac{\beta}{N}SI}{-\frac{\beta}{N}SI} + \frac{\gamma I}{\frac{\beta}{N}SI} = -1 + \frac{\gamma N}{\beta S}. \quad (80)$$

Fazendo a separação de variáveis em (80), temos:

$$dI = \left(-1 + \frac{\gamma N}{\beta S}\right) dS, \quad (81)$$

Resolvendo a Equação (81), obtem-se o seguinte resultado,

$$\int dI = \int \left(-1 + \frac{\gamma N}{\beta S}\right) dS, \quad (82)$$

$$I(t) = -S(t) + \frac{\gamma N}{\beta} \ln S(t) + C. \quad (83)$$

Considerando que inicialmente tem-se, pelo menos, um indivíduo infectado dentro da população total de tamanho  $N$ , ou seja,  $I(0) = 1$  e  $S(0) = N - 1$ , utilizamos essas condições iniciais para resolver a equação (83). Temos, então, que

$$I(0) = -S(0) + \frac{\gamma N}{\beta} \ln(S(0)) + C, \quad (84)$$

de onde,

$$1 = -(N - 1) + \frac{\gamma N}{\beta} \ln(N - 1) + C. \quad (85)$$

Dessa forma, podemos encontrar o valor da constante de integração  $C$ ,

$$C = N - \frac{\gamma N}{\beta} \ln(N - 1). \quad (86)$$

Substituindo esse resultado na equação (83), obtem-se a solução analítica para o compartimento dos indivíduos infectados no instante de tempo  $t$ ,

$$I(t) = -S(t) + \frac{\gamma N}{\beta} \ln(S(t)) + N - \frac{\gamma N}{\beta} \ln(N - 1). \quad (87)$$

Simplificando esse resultado, chega-se a

$$I(t) = -S(t) + N + \frac{\gamma N}{\beta} \ln \left( \frac{S(t)}{N - 1} \right). \quad (88)$$

A dinâmica do modelo SIR considera que, com o decorrer do tempo, a quantidade de infectados irá diminuir de acordo com o limiar epidêmico e, conseqüentemente a epidemia irá terminar. Dessa forma, quando  $t \rightarrow \infty$ ,  $I(\infty) = 0$ . Substituindo na equação (88),

$$I(\infty) + S(\infty) = N + \frac{\gamma N}{\beta} \ln \left( \frac{S(\infty)}{N - 1} \right), \quad (89)$$

de onde,

$$S(\infty) = N + \frac{\gamma N}{\beta} \ln \left( \frac{S(\infty)}{N - 1} \right). \quad (90)$$

A equação (90) representa uma solução genérica para a classe de indivíduos suscetíveis com o decorrer do tempo.

A gravidade da epidemia pode ser determinada pelo número total de casos que ocorreu, denominado tamanho final da epidemia. Ele é calculado através do  $R(\infty)$  pois representa todos os indivíduos que foram infectados e se recuperaram. Dessa forma, pode-se considerar

$$R(\infty) = N - S(\infty). \quad (91)$$

Para analisarmos se ocorrerá epidemia ou não em determinada população, utilizamos o número de reprodução básico,  $\mathcal{R}_0$ , que é uma medida do instante inicial da epidemia.  $\mathcal{R}_0$  é a taxa de formação de novos indivíduos infectados dividido pela taxa de recuperação desses mesmos indivíduos. A taxa de novos infectados é dada por  $\beta SI/N$ , e a taxa de recuperação desses indivíduos é  $\gamma I$ . Dessa forma, tem-se que no instante inicial,

$$\mathcal{R}_0 = \frac{\beta S(0)I(0)/N}{\gamma I(0)} = \frac{\beta S(0)I(0)}{\gamma NI(0)} = \frac{\beta S(0)}{\gamma N}. \quad (92)$$

No início de uma epidemia o número de suscetíveis iniciais  $S(0)$  é aproximadamente o total da população  $N$ , considerando  $I(0)$  muito pequeno, ou seja,  $S(0) = N - I(0) \approx N$ . Utilizando essa característica é possível simplificar (92), de modo que

$$\mathcal{R}_0 = \frac{\beta}{\gamma}. \quad (93)$$

O número de reprodução básico é importante para analisar a velocidade do crescimento dos indivíduos infectados. Por outra parte, o número de reprodução efetivo,  $\mathcal{R}$ , pode ser determinado através do número de reprodução básico. Considerando  $S/N = S(t)$ ,  $I/N = I(t)$  e  $R/N = R(t)$  as frações de indivíduos, com valores entre 0 e 1, referentes a cada compartimento no decorrer do tempo,  $\mathcal{R}$  é definido como,

$$\mathcal{R} = \frac{\beta S(t)I(t)}{\gamma I(t)} = \frac{\beta S(t)}{\gamma} = \mathcal{R}_0 S(t). \quad (94)$$

$\mathcal{R}$  é o número de reprodução da doença ao longo do tempo. Refere-se a quantos casos, em média, cada indivíduo da população vai infectar no decorrer do tempo. É interessante determinar quando  $\mathcal{R} < 1$ , pois representa quando não ocorrerá epidemia, como visto na Seção 1.3. Dessa forma, a partir da desigualdade

$$\mathcal{R} = \mathcal{R}_0 S(t) < 1, \quad (95)$$

encontra-se

$$S(t) < \frac{1}{\mathcal{R}_0}. \quad (96)$$

A expressão (96) representa o limiar epidêmico que é o responsável por fazer a curva dos indivíduos infectados diminuir.

### 4.3 Modelo Estocástico

O modelo SIR estocástico apresentado nesta seção tem como referência um estudo realizado na Ref. [9], utilizando Cadeias de Markov de Tempo Contínuo (*CTMC*). Por essa razão, denomina-se o modelo como SIR CTMC. Os compartimentos do modelo,  $S(t)$ ,  $I(t)$  e  $R(t)$ , para um determinado instante de tempo  $t$ , agora são considerados variáveis aleatórias. Seguindo o esquema descrito para o modelo determinístico, não são considerados nascimentos ou mortes na população, de forma que  $S(t) + I(t) + R(t) = N$ . Considera-se também que não há período latente. Tem-se que  $R(t) = N - S(t) - I(t)$ , tendo-se apenas duas das variáveis independentes,  $S(t)$  e  $I(t)$ . Portanto, a probabilidade de transição considerada nesse modelo é uma probabilidade conjunta dessas duas variáveis.

As probabilidades de transição são definidas a seguir em termos das taxas do modelo determinístico, considerando a mudança dos estados  $S(t)$  e  $I(t)$  durante um intervalo de tempo pequeno  $\Delta t$  como  $\Delta S = S(t + \Delta t) - S(t)$  e  $\Delta I = I(t + \Delta t) - I(t)$ ,

$$\begin{aligned} \text{Prob}\{(\Delta S(t), \Delta I(t)) | (S(t), I(t))\} = \\ = \begin{cases} \frac{\beta}{N} S(t) I(t) \Delta t + o(\Delta t), & (\Delta S(t), \Delta I(t)) = (-1, 1) \\ \gamma I(t) \Delta t + o(\Delta t), & (\Delta S(t), \Delta I(t)) = (0, -1) \\ 1 - \left[ \frac{\beta}{N} S(t) I(t) + \gamma I(t) \right] \Delta t, & (\Delta S(t), \Delta I(t)) = (0, 0) \\ o(\Delta t), & \textit{outro} \end{cases} \quad (97) \end{aligned}$$

Os estados do sistema,  $S$ ,  $I$  e  $R$  representam a quantidade de indivíduos que pertencem ao compartimento correspondente. Dessa forma, a probabilidade condicional dada pelo sistema (97) representa a transição da variação de suscetíveis e infectados considerando que a evolução do sistema é dada através da variação de somente um indivíduo em um pequeno intervalo de tempo  $\Delta t$  e, para cada caso de ocorrência do sistema há uma probabilidade associada.

O par  $(\Delta S(t), \Delta I(t)) = (-1, 1)$  representa a condição de um indivíduo ter sido infectado,  $(\Delta S(t), \Delta I(t)) = (0, -1)$  ocorre quando há um indivíduo infectado que foi recuperado e, por fim  $(\Delta S(t), \Delta I(t)) = (0, 0)$  representa que não houve alteração nos compartimentos, ou seja, nenhum indivíduo foi infectado e nenhum se recuperou. As equações do modelo são resumidas na Tabela 1 no intuito de facilitar a visualização dessas situações e das equações correspondentes.

Tabela 1 - Probabilidades de transição para o modelo SIR CTMC.

$(\Delta S, \Delta I)$	Probabilidade
$(-1, 1)$	$\frac{\beta}{N}S(t)I(t)\Delta t + o(\Delta t)$
$(0, -1)$	$\gamma I(t)\Delta t + o(\Delta t)$
$(0, 0)$	$1 - \left[ \frac{\beta}{N}S(t)I(t) + \gamma I(t) \right] \Delta t + o(\Delta t)$

Legenda: Resumo das equações do modelo (97).

Fonte: A autora, 2020.

Inicialmente, consideraremos a distribuição do sistema como  $(S(0), I(0)) = (s_0, i_0)$ , onde  $s_0 + i_0 = N$  e  $R(0) = 0$ , pois considera-se que neste instante não há indivíduos recuperados e  $s_0 \geq 0$  e  $i_0 > 0$ .

A probabilidade de estado é dada por  $p_{(s,i)}(t) = \text{Prob}\{S(t) = s, I(t) = i\}$  e pode ser obtida através das soluções das equações avançadas de Kolmogorov,

$$\frac{dp_{(s,i)}(t)}{dt} = \frac{\beta}{N}(s+1)(i-1)p_{(s+1,i-1)}(t) + \gamma(i+1)p_{(s,i+1)}(t) - \left[ \frac{\beta}{N}si + \gamma i \right] p_{(s,i)}(t), \quad (98)$$

onde  $s \in [0, N]$ ,  $i \in [0, N-s]$  e  $s+i \leq N$ . Se o estado  $(s, i)$  estiver fora desse intervalo, a probabilidade é igual a zero. Para o caso onde  $s = 0, 1, 2, \dots, N-1$ , tem-se

$$\frac{dp_{(s,0)}}{dt} = \gamma p_{(s,1)}. \quad (99)$$

Por outro lado, para  $s = N$

$$\frac{dp_{(N,0)}}{dt} = 0. \quad (100)$$

Nos  $N+1$  estados  $(s, 0)$ , considerando  $s = 0, 1, 2, \dots, N$ , não ocorrem transições para qualquer outro estado do sistema pois não há indivíduos infectados e, portanto, esses estados são fechados. Além disso, esses estados são denominados absorventes. A epidemia acaba quando um estado absorvente é alcançado.

No modelo SIR CTMC é possível obter algumas propriedades que o modelo determinístico não oferece. A probabilidade de surto de uma doença em determinada população e a distribuição de probabilidade tanto para o tamanho final como para a duração esperada de uma epidemia, são propriedades acessíveis através do modelo SIR CTMC.

Como visto no modelo determinístico, através do  $\mathcal{R}_0$  é possível determinar se ocorrerá epidemia ou não no instante inicial e, por meio do  $\mathcal{R}$ , se ocorrerá epidemia no decorrer do tempo. No modelo SIR CTMC, pode-se definir uma propriedade para calcular a probabilidade de surto a depender do  $\mathcal{R}_0$  e do número de infectados inicial,  $I(0)$ . O número



de reprodução básico no modelo estocástico é o mesmo do modelo determinístico, definido pela equação (93). Dessa forma, considerando o início de uma epidemia com a condição inicial  $I(0) = i$  sendo um valor pequeno, a probabilidade de ocorrer um surto da doença [33],  $P_{surto}$ , é dada por,

$$P_{surto} = \begin{cases} 0 & , \mathcal{R}_0 \leq 1 \\ 1 - \left(\frac{1}{\mathcal{R}_0}\right)^i & , \mathcal{R}_0 > 1 \end{cases} . \quad (101)$$

**Exemplo 4.3.1.** *Suponha que em determinada população  $N = 100$ ,  $\mathcal{R}_0 = 2$  e que, inicialmente,  $I(0) = 3$ . Para analisar a probabilidade de ocorrer uma epidemia utilizamos a equação (101). Substituindo os valores dados, obtem-se  $P_{surto} = 1 - (1/2)^3 = 0.875$ . Logo, a probabilidade de ocorrer uma epidemia nessa população é de 0.875. ■*

O tamanho final da epidemia no modelo estocástico é determinado da mesma forma que no modelo determinístico, ou seja,  $R(\infty) = N - S(\infty)$ . Porém, no modelo SIR CTMC pode-se obter uma distribuição de probabilidade para esse tamanho final e assim, analisar características específicas do sistema. Para calcular essa distribuição de probabilidade para o tamanho final da epidemia, é necessário determinar a matriz geradora  $G$  e a matriz de transição da Cadeia de Markov Incorporada do sistema,  $P_Y$ . A matriz geradora  $G$  e a matriz de transição  $P_Y$  dependem de como os estados do sistema são ordenados. No modelo SIR CTMC trabalhamos com as probabilidades conjuntas dos compartimentos  $S$  e  $I$  e, portanto, existem  $(N + 1)(N + 2)/2$  pares de estados, que é o número das combinações possíveis desses dois compartimentos. Os pares de estado serão ordenados da seguinte forma:

$$(N, 0), (N - 1, 0), (N - 2, 0), \dots, (0, 0), (N - 1, 1), (N - 2, 1), \dots, (0, 1), \dots, (0, N) \quad (102)$$

Usando essa ordenação, pode-se definir o conjunto das probabilidades de estado como o vetor  $p(t) = (p_{(N,0)}, p_{(N-1,0)}, \dots, p_{(0,N)})^T$  e, assim, obtemos a matriz geradora  $G$ . A matriz  $G$  é uma matriz quadrada de dimensão  $(N + 1)(N + 2)/2$ , que pode ser encontrada através do sistema dado pela expressão (61),  $dp/dt = Gp$ . Para facilitar a notação na construção da matriz  $G$ , vamos associar cada estado com um número inteiro da seguinte forma,

$$\begin{array}{ll} (N, 0) & \rightarrow 1 \\ (N - 1, 0) & \rightarrow 2 \\ \vdots & \dots \\ (0, N) & \rightarrow N_F = \frac{(N + 1)(N + 2)}{2} \end{array} . \quad (103)$$

**Exemplo 4.3.2.** *Considerando um modelo epidêmico SIR CTMC com população  $N = 3$ ,*

tem-se  $\frac{(N+1)(N+2)}{2} = \frac{(3+1)(3+2)}{2} = \frac{20}{2} = 10$  possíveis pares de estados. Associando cada par de estado a um número real, definido em (103), tem-se:

$$\begin{array}{ll} (3,0) \rightarrow 1 & (1,1) \rightarrow 6 \\ (2,0) \rightarrow 2 & (0,1) \rightarrow 7 \\ (1,0) \rightarrow 3 & (1,2) \rightarrow 8 \\ (0,0) \rightarrow 4 & (0,2) \rightarrow 9 \\ (2,1) \rightarrow 5 & (0,3) \rightarrow 10 \end{array}$$

Então, a matriz de transição  $P(\Delta t)$  é dada por,

$$P(\Delta t) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & \gamma i & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & \gamma i & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & \gamma i & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 - C & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 - C & 0 & \gamma i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 - C & 0 & \gamma i & 0 & 0 \\ 0 & 0 & 0 & 0 & \beta si/N & 0 & 0 & 1 - C & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \beta si/N & 0 & 0 & 1 - C & \gamma i & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \beta si/N & 0 & 1 - C & \gamma i \end{pmatrix} \quad (104)$$

Dessa forma, pode-se definir a matriz  $G$  da seguinte maneira,

$$G = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \gamma i & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \gamma i & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \gamma i & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -C & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -C & 0 & \gamma i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -C & 0 & \gamma i & 0 & 0 \\ 0 & 0 & 0 & 0 & \beta si/N & 0 & 0 & -C & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \beta si/N & 0 & 0 & -C & \gamma i & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \beta si/N & 0 & -C & \gamma i \end{pmatrix} \quad (105)$$

onde  $s$  e  $i$  correspondem ao valor de  $S(t)$  e  $I(t)$  e  $C = \beta si/N + \gamma i$  é o coeficiente de  $p_{(s,i)}(t)$  no sistema de equações (98). ■

A ordenação definida pela expressão (102) foi utilizada por conveniência, mas não é única. Se os estados forem ordenados de outra forma, o conjunto das probabilidades de estado e a matriz geradora tem que ser modificados de acordo com essa outra ordenação.

A partir da matriz  $G$ , pode-se obter a matriz de transição da Cadeia de Markov Incorporada,  $P_Y$ . Esta matriz é muito útil no cálculo do tamanho final da epidemia. Utilizando a notação (103) pode-se construir a matriz de transição  $P_Y$ , definida em (54), da seguinte forma,

$$P_Y = \begin{pmatrix} \tilde{p}_{11} & \tilde{p}_{12} & \tilde{p}_{13} & \cdots & \tilde{p}_{1N_F} \\ \tilde{p}_{21} & \tilde{p}_{22} & \tilde{p}_{23} & \cdots & \tilde{p}_{2N_F} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{p}_{N_F1} & \tilde{p}_{N_F2} & \tilde{p}_{N_F3} & \cdots & \tilde{p}_{N_FN_F} \end{pmatrix}, \quad (106)$$

onde  $\tilde{p}$  representa a probabilidade de transição de um estado para outro, definidas pela expressão (53). A matriz  $P_Y$  é uma matriz quadrada de dimensão  $(N+1)(N+2)/2$ , assim como a matriz geradora  $G$ .

Na matriz  $G$ , os elementos das primeiras  $N+1$  colunas são iguais a zero. Esses  $N+1$  estados são classificados como estados absorventes, pois não há transição em qualquer um desses estados já que não há indivíduos infectados, ou seja,  $(s, i) = (s, 0)$ . Isso significa que a matriz de transição  $P_Y = (\tilde{p}_{kl})$  da cadeia de Markov incorporada satisfaz  $\tilde{p}_{ll} = 1$ , para  $l = 1, 2, \dots, N+1$ .

**Exemplo 4.3.3.** Utilizando a matriz geradora do Exemplo (4.3.2), pode-se determinar a matriz de transição  $P_Y$ , de acordo com as expressões (51) e (53),

$$P_Y = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & \gamma i/C & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & \gamma i/C & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & \gamma i/C & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \gamma i/C & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \gamma i/C & 0 & 0 \\ 0 & 0 & 0 & 0 & \beta si/NC & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \beta si/NC & 0 & 0 & 0 & 0 & \gamma i/C \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \beta si/NC & 0 & 0 & 0 \end{pmatrix}, \quad (107)$$

onde  $s$  e  $i$  correspondem ao valor de  $S(t)$  e  $I(t)$  e  $C = \beta si/N + \gamma i$ , como no exemplo anterior. ■

Sendo  $(S(t), I(t)) = (s, i)$ , sabe-se que no modelo há a possibilidade de duas transições já que ele analisa a variação de somente um indivíduo a cada instante de tempo. A transição do estado  $a = (s, i)$  para o estado  $b = (s-1, i+1)$  representa a infecção de um indivíduo suscetível, ao passo que a transição do estado  $a = (s, i)$  para o estado  $c = (s, i-1)$  significa a recuperação de um indivíduo infectado. Assim, podem-se

definir as probabilidades da matriz  $P_Y$ , de acordo com a definição da matriz geradora  $G$ .

$$-\frac{g_{ba}}{g_{aa}} = \frac{\gamma^i}{\beta si/N + \gamma i} = \frac{\gamma N}{\beta s + \gamma N} = \tilde{p}_s. \quad (108)$$

Como a matriz  $P_Y$  é estocástica, a soma dos elementos da coluna tem que ser igual a 1. Dessa forma, a probabilidade de que um indivíduo suscetível seja infectado é dada por

$$1 - \tilde{p}_s = 1 - \frac{\gamma N}{\beta s + \gamma N} = \frac{\beta s}{\beta s + \gamma N}. \quad (109)$$

**Exemplo 4.3.4.** *Um modelo epidêmico SIR CTMC com população total  $N = 3$  possui 10 pares de estados. A relação entre os estados desse modelo, correspondente à EMC, pode ser representada através de um grafo orientado. O agrupamento dos estados é feito em quatro conjuntos, de acordo com o esquema (103). Dessa forma, tem-se*

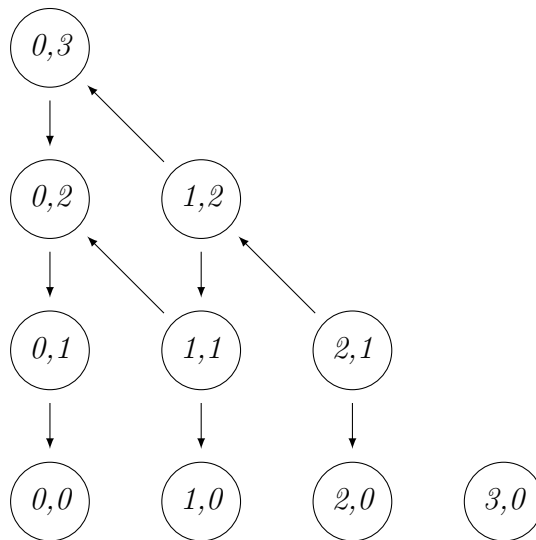
I:  $(3,0), (2,0), (1,0), (0,0)$

II:  $(2,1), (1,1), (0,1)$

III:  $(1,2), (0,2)$

IV:  $(0,3)$

Representando a relação desses conjuntos através de um grafo orientado, é possível visualizar melhor as relações entre os estados e algumas das suas características,



A matriz de transição da EMC, definida pelas expressões (108) e (109), pode ser

expressa em forma de matriz em blocos, na forma a seguir,

$$P_Y = \left( \begin{array}{cccc|ccc|cc|c} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & \tilde{p}_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & \tilde{p}_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & \tilde{p}_0 & 0 & 0 & 0 & 0 \\ - & - & - & - & - & - & - & - & - & - & - \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \tilde{p}_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \tilde{p}_0 & 0 & 0 \\ - & - & - & - & - & - & - & - & - & - & - \\ 0 & 0 & 0 & 0 & 1 - \tilde{p}_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 - \tilde{p}_1 & 0 & 0 & 0 & 0 & \tilde{p}_0 \\ - & - & - & - & - & - & - & - & - & - & - \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 - \tilde{p}_1 & 0 & 0 & 0 \end{array} \right), \quad (110)$$

onde

$$P_Y = \begin{pmatrix} I & A_1 & 0 & 0 \\ 0 & 0 & A_2 & 0 \\ 0 & B_1 & 0 & A_3 \\ 0 & 0 & B_2 & 0 \end{pmatrix}. \quad (111)$$

O vetor de probabilidade de estado também pode ser dividido nos quatro conjuntos em que foram separados os estados, de modo que  $p = (p_{(s,i)}) = (p_I, p_{II}, p_{III}, p_{IV})^T$ , onde  $T$  denota a transposta. Dessa forma,

$$p_I = (p_{(3,0)}, p_{(2,0)}, p_{(1,0)}, p_{(0,0)})^T \quad (112)$$

e assim, sucessivamente para os outros grupos. ■

Como visto no Exemplo (4.3.4), a matriz  $P_Y$  pode ser dividida em blocos. Cada bloco representa as transições possíveis entre esses conjuntos. Além disso, cada bloco possui tamanho diferente um do outro. O bloco  $I$  é uma matriz identidade de tamanho  $(N + 1) \times (N + 1)$ , representando o conjunto dos estados absorventes. Os blocos denotados por  $A_i$ , representam as transições de recuperação de indivíduos infectados, ou seja transições de  $i$  indivíduos infectados para  $i - 1$ , com  $i = 1, 2, \dots, N$ . Por sua parte, os blocos  $B_i$  caracterizam a infecção de um indivíduo suscetível, isto é, a transição de  $i$  indivíduos infectados para  $i + 1$  sendo, nesse caso,  $i = 1, 2, \dots, N - 1$ .

Para calcular a distribuição do tamanho final da epidemia é necessário conhecer as probabilidades de absorção. Considerando que inicialmente tem-se um indivíduo infec-

tado, ou seja,  $p_{(N-1,1)}(0) = 1$ , em uma população de tamanho total igual a  $N$ , o número máximo de transições para chegar à absorção é  $2N - 1$ .

**Exemplo 4.3.5.** *Utilizando o grafo orientado do Exemplo (4.3.4), pode-se observar que se o estado inicial for  $(2, 1)$ , o número máximo de transições até a absorção será 5. Se seguirmos o caminho das setas representadas no grafo orientado, é fácil ver que há cinco transições.* ■

Considerando que inicialmente há apenas um indivíduo infectado,  $I(0) = 1$ , e não há indivíduos recuperados,  $R(0) = 0$ , a quantidade de indivíduos suscetíveis é dada por  $S(0) = N - 1$ . Calculando as probabilidades de absorção, obtemos a probabilidade associada ao tamanho final da epidemia,  $\{p_i^f\}$ ,  $i = 1, 2, \dots, N$ . Como os  $N + 1$  estados sem indivíduos infectados são estados absorventes e a quantidade de infectados inicialmente é  $I(0) = 1$ , tem-se

$$\lim_{t \rightarrow \infty} \sum_{s=0}^{N-1} p_{(s,0)}(t) = 1. \quad (113)$$

Com isso, as probabilidades para a distribuição do tamanho final satisfazem, para  $s = 0, 1, \dots, N - 1$ ,

$$\lim_{t \rightarrow \infty} p_{(s,0)}(t) = p_{N-s}^f. \quad (114)$$

Quando o número de infectados atinge o valor zero,  $I(t) = 0$ , há um total de  $s$  indivíduos que foram infectados,  $S(t) = s$ , e, portanto, o tamanho final da epidemia é dado por  $N - s$ . Utilizando a matriz  $P_Y$  encontramos as probabilidades de absorção e, conseqüentemente, a distribuição de probabilidade para o tamanho final da epidemia, da seguinte forma,

$$\lim_{t \rightarrow \infty} p(t) = p(2N - 1) = (P_Y)^{2N-1} p(0), \quad (115)$$

onde  $p(0)$  representa o vetor das probabilidades de estado no instante inicial.

**Exemplo 4.3.6.** *Para uma população total  $N = 3$  é possível determinar a distribuição do tamanho final da epidemia através da expressão (115). A matriz  $P_Y$  foi definida no*

Exemplo 4.3.4 e  $p(0)$  foi definido para a condição inicial de  $I(0) = 1$ . Dessa forma,

$$\lim_{t \rightarrow \infty} p(t) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & \tilde{p}_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & \tilde{p}_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & \tilde{p}_0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \tilde{p}_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \tilde{p}_0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 - \tilde{p}_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 - \tilde{p}_1 & 0 & 0 & 0 & \tilde{p}_0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 - \tilde{p}_1 & 0 & 0 & 0 \end{pmatrix}^5 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

O resultado desse produto nos dá a distribuição de probabilidade para o tamanho final da epidemia. ■

A última propriedade do modelo SIR CTMC, referente à duração esperada de uma epidemia, pode ser calculada usando o método do primeiro tempo de passagem, introduzido na Seção 3.5. A duração de uma epidemia corresponde ao tempo que leva até chegar à absorção, ou seja, o tempo  $t$  até  $I(t) = 0$ . Para o modelo SIR CTMC, a probabilidade de absorção é igual a um, independentemente do valor do número de reprodução básico,  $R_0$ . No entanto, dependendo do número inicial de infectados,  $i$ , do tamanho da população  $N$  e do valor de  $R_0$ , o tempo até a absorção pode ser muito curto ou muito longo. Dessa forma, temos um sistema de equações que pode ser resolvido, de modo que seja possível encontrar o tempo esperado até a absorção. Como o modelo é analisado para o par  $(S(t), I(t)) = (s, i)$  representando o estado possível do sistema, cada elemento da matriz geradora  $G$  é uma taxa de transição entre um par de estados  $(s_1, i_1) \rightarrow (s_2, i_2)$ . Para calcular o tempo esperado para a absorção  $\tau$ , definido pela expressão (75), utilizamos a matriz geradora  $G$  excluindo as transições para os estados absorventes, denotada por  $\tilde{G}$ . A partir da matriz  $G$ , pode-se escrever a matriz quadrada  $\tilde{G}$ , de dimensão  $N(N+1)/2$  como sendo, para o Exemplo (4.3.2)

$$\tilde{G} = \begin{pmatrix} \beta si/N + \gamma i & 0 & 0 & 0 & 0 & 0 \\ 0 & \beta si/N + \gamma i & 0 & \gamma i & 0 & 0 \\ 0 & 0 & \beta si/N + \gamma i & 0 & \gamma i & 0 \\ \beta si/N & 0 & 0 & \beta si/N + \gamma i & 0 & 0 \\ 0 & \beta si/N & 0 & 0 & \beta si/N + \gamma i & \gamma i \\ 0 & 0 & 0 & \beta si/N & 0 & \beta si/N + \gamma i \end{pmatrix}. \quad (116)$$

A matriz  $\tilde{G}$  também depende da ordenação dos estados. Nesse caso, vamos considerar a ordenação feita em (103). O vetor para a duração esperada da epidemia, ou seja, o tempo para extinção da doença,  $\tau$ , depende dos pares  $(s, i)$ . Considerando  $T_{(s,i)}$  a variável aleatória para o tempo até a absorção, tem-se que

$$\tau_{(s,i)} = \mathbb{E}(T_{(s,i)}). \quad (117)$$

Usando a equação (67), a expressão (117) pode ser reescrita da seguinte forma,

$$\begin{aligned} \tau_{(s,i)} = & \left[ 1 - \left( \frac{\beta}{N} si + \gamma i \right) \Delta t \right] (\tau_{(s,i)} + \Delta t) + \left( \frac{\beta}{N} si \Delta t \right) (\tau_{(s-1,i+1)} + \Delta t) \\ & + (\gamma i \Delta t) (\tau_{(s,i-1)} + \Delta t), \end{aligned} \quad (118)$$

$$\begin{aligned} = & \tau_{(s,i)} + \Delta t - \left( \frac{\beta}{N} si + \gamma i \right) \Delta t \tau_{(s,i)} - \left( \frac{\beta}{N} si + \gamma i \right) \Delta t \Delta t \\ & + \left( \frac{\beta}{N} si \Delta t \right) \tau_{(s-1,i+1)} + \left( \frac{\beta}{N} si \Delta t \right) \Delta t + \gamma i \Delta t \tau_{(s,i-1)} + \gamma i \Delta t \Delta t. \end{aligned} \quad (119)$$

Reescrevendo e simplificando a equação (119), obtem-se

$$\Delta t - \left( \frac{\beta}{N} si + \gamma i \right) \Delta t \tau_{(s,i)} + \left( \frac{\beta}{N} si \Delta t \right) \tau_{(s-1,i+1)} + \gamma i \Delta t \tau_{(s,i-1)} = 0. \quad (120)$$

Dividindo por  $\Delta t$ ,

$$1 - \left( \frac{\beta}{N} si + \gamma i \right) \tau_{(s,i)} + \left( \frac{\beta}{N} si \right) \tau_{(s-1,i+1)} + \gamma i \tau_{(s,i-1)} = 0, \quad (121)$$

de onde,

$$- \left( \frac{\beta}{N} si + \gamma i \right) \tau_{(s,i)} + \left( \frac{\beta}{N} si \right) \tau_{(s-1,i+1)} + \gamma i \tau_{(s,i-1)} = -1. \quad (122)$$

A equação (122) pode ser reescrita utilizando as expressões (108) e (109), obtendo-se, por fim,

$$(\gamma i + (\beta/N) si) (p_s \tau_{(s,i-1)} - \tau_{(s,i)} + (1 - p_s) \tau_{(s-1,i+1)}) = -1. \quad (123)$$

O sistema de equações dado pela expressão (123) corresponde à duração esperada para a epidemia estudada, que também pode ser representada pela forma matricial (75).

Para calcular o tempo entre eventos no modelo SIR CTMC, utiliza-se a soma das probabilidades do processo mover-se para um estado diferente, definida pela ex-



pressão (31), correspondente a

$$\sum_{y \neq x} p_{yx}(\Delta t) = \left( \frac{\beta}{N} S(t) I(t) + \gamma I(t) \right) \Delta t + o(\Delta t), \quad (124)$$

com  $\alpha(x) = (\beta/N)S(t)I(t) + \gamma I(t)$ .

O tempo entre eventos, definido na equação (44), pode ser então calculado utilizando essa expressão para  $\alpha(x)$ ,

$$T_E = -\frac{\ln U}{\frac{\beta}{N} S(t) I(t) + \gamma I(t)}, \quad (125)$$

onde,  $U$  é uma variável aleatória uniforme definida no intervalo  $[0, 1]$ .

## 5 SIMULAÇÕES E RESULTADOS

Nesse capítulo, serão apresentados e discutidos os resultados das simulações computacionais. O modelo SIR CTMC, apresentado na Seção 4.3, foi implementado computacionalmente e usado na modelagem de duas doenças. A primeira delas, um caso de sarampo acontecido na Ilha Grande, Rio de Janeiro, Brasil, no ano de 1976. O surto de sarampo foi estudado e documentado na Referência [34]. Desse trabalho foram extraídos os valores dos parâmetros usados na modelagem. Esse exemplo foi usado com a intenção de validar a implementação computacional do modelo SIR CTMC, assim como ampliar a compreensão da modelagem estocástica. A segunda situação epidemiológica abordada foi a atual pandemia de coronavírus e a sua evolução no Estado do Rio de Janeiro, Brasil. Em relação a esse caso, fica evidenciada a limitação, em termos de recursos computacionais, da modelagem estocástica através de modelos compartimentais para a análise de epidemias em populações muito grandes. Contudo, foram obtidos alguns resultados úteis e geradas novas linhas importantes de trabalho, motivadas pelas dificuldades supracitadas.

### 5.1 Surto de Sarampo na Ilha Grande

As simulações realizadas tiveram como base para os parâmetros, dados identificados em um estudo sobre um surto epidêmico de sarampo em uma vila de pescadores na região de Ilha Grande, no estado do Rio de Janeiro, em 1976 [34]. Vale ressaltar que a modelagem realizada não tem a função de descrever um surto de sarampo em qualquer lugar ou condição. Para isso, uma equipe interdisciplinar teria que trabalhar em função de definir os parâmetros apropriados para esse caso particular. O sarampo é uma doença altamente contagiosa. Contudo, as condições de contágio e propagação vão depender de múltiplos fatores, como densidade populacional, faixa etária, políticas de vacinação e outras [35].

Na região da Ilha Grande considerada, a população total de moradores na época era de 453 indivíduos. A doença começou a se propagar por um indivíduo que viajou para esse local e acabou transmitindo a doença para os moradores. Dessa forma, tem-se que, como condição inicial,  $I(0) = 1$ . Na Tabela (2), são relacionados os valores de tamanho da população da região por faixa etária. Pode-se observar também a incidência da doença de acordo com a idade, evidenciando que crianças e adolescentes são os mais afetados. Observa-se que a propagação da doença aconteceu em indivíduos entre 1 a 19 anos. Foi considerado que indivíduos com mais de 19 anos possuíam imunidade natural ou haviam sido vacinados no surto de sarampo que ocorrera alguns anos antes, e crianças menores de 1 ano possuíam imunidade materna, pela transmissão de anticorpos da mãe. Por esse motivo, para realizar a modelagem e mostrar a usabilidade do modelo SIR CTMC,

Tabela 2 - Incidência de sarampo de acordo com a idade.

Faixa etária (anos)	Total de Casos	População Residente
< 1	-	11
1 a 5	16	66
6 a 10	20	60
11 a 15	11	53
16 a 19	3	41
> 19	-	222

Legenda: Distribuição populacional e número de infectados por Sarampo em surto acontecido na Ilha Grande, RJ, em 1976.

Fonte: [34]

consideramos a população do compartimento de indivíduos suscetíveis, no instante inicial, como  $S(0) = 220$ , o de infectados,  $I(0) = 1$  e, por fim, o de recuperados como  $R(0) = 233$ .

A epidemia do sarampo nessa localidade atingiu um total de 50 indivíduos, sem a ocorrência de mortes, com duração de 60 dias.

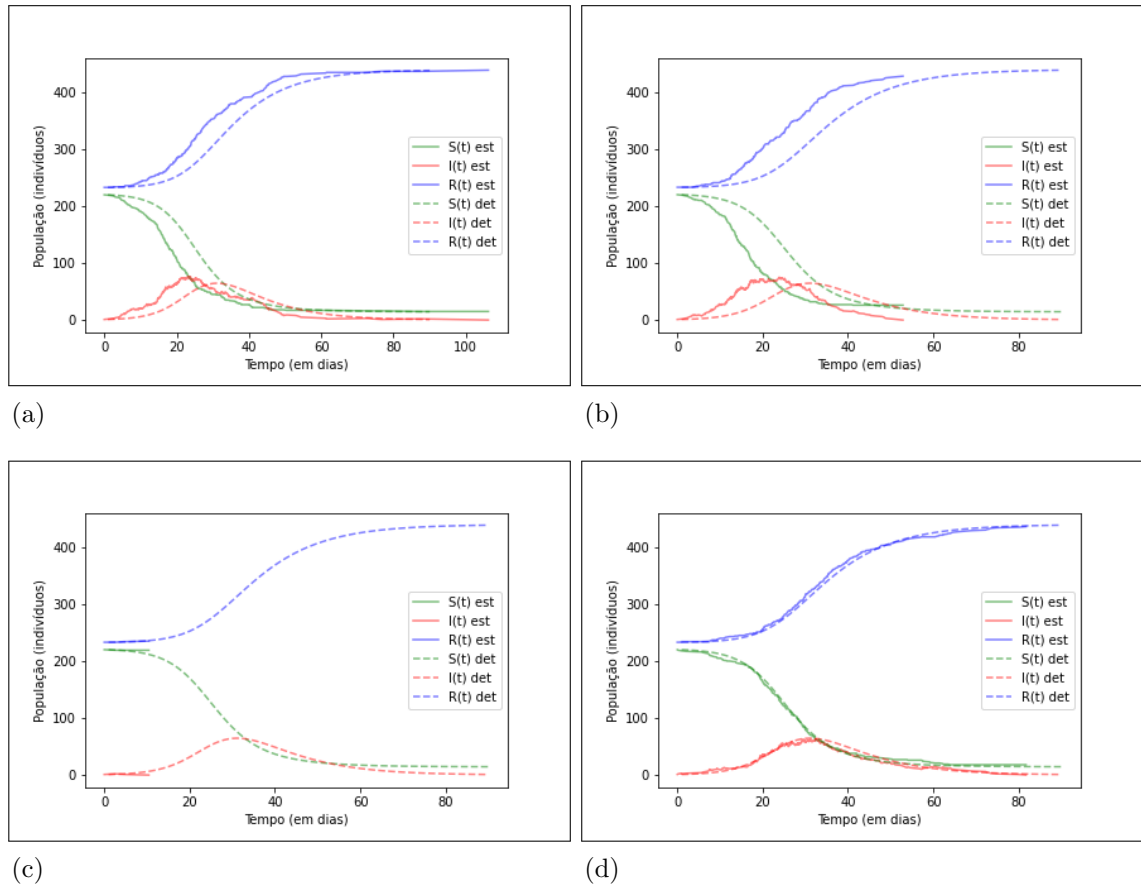
A duração média em que um indivíduo permaneceu infectado foi de 9,5 dias. Essa duração representa o tempo médio que um indivíduo leva para se recuperar, e será utilizado para definir a taxa de recuperação,  $\gamma = 2/19$ .

No caso do sarampo, pelas considerações acima indicadas sobre contágio e propagação, as variações possíveis na evolução da doença são muito grandes [35]. Assim, o valor do número de reprodução básico,  $R_0$ , deve ser cuidadosamente estimado. Na Ref. [34], não é oferecido um valor para esse parâmetro. Dessa forma, analisando as taxas de contágio por faixa etária, documentadas no referido trabalho, determinamos o valor  $R_0 = 6$ . Em consequência, a taxa de contato empregada foi  $\beta = R_0\gamma \approx 0,6315$ .

Foram realizadas 1000 simulações do modelo estocástico. Na Figura 5, são mostrados vários dos cenários possíveis para a evolução da doença através desta modelagem. Em cada sub-figura mostra-se a evolução de uma única simulação. Para efeitos de visualização das variações, compara-se com o resultado da modelagem determinística, que produz uma única evolução. Pode-se observar como o pico da doença, que representa o número máximo de indivíduos infectados ao mesmo tempo, varia no tamanho e no tempo para as diferentes simulações, em contraste com o resultado único que provê a evolução determinística. Em particular, na Figura 5c, mostra-se um cenário no qual não se desenvolve uma epidemia.

De acordo com a definição do número básico de reprodução, sabe-se que há ocorrência da epidemia na população na condição de que  $R_0 > 1$ . Para o modelo estocástico pode-se determinar a probabilidade de ocorrência do surto da doença na população, utilizando a equação (101). Essa probabilidade foi calculada com o  $R_0$  utilizado nas simulações,

Figura 5 - Diferentes cenários na modelagem estocástica.



Legenda: Em cada sub-figura mostra-se o resultado de uma única simulação do modelo estocástico (linha cheia). Mostra-se também a evolução do modelo determinístico (linha tracejada) para ilustrar as variações do modelo estocástico. A seguir, indicam-se os valores de alguns resultados da modelagem. (a) Pico da Epidemia: 77, Dia do Pico: 23, 19, Duração da epidemia: 106, 20 dias e Tamanho Final da Epidemia: 206. (b) Pico da Epidemia: 75, Dia do Pico: 23, 88, Duração da epidemia: 52, 86 dias e Tamanho Final da Epidemia: 195. (c) Pico: 2, Dia do Pico: 2, 72, Duração: 10, 42 dias e Tamanho Final: 2. (d) Pico da Epidemia: 63, Dia do Pico: 29, 54, Duração da epidemia: 81, 71 dias e Tamanho Final da Epidemia: 203.

Fonte: A autora, 2020.

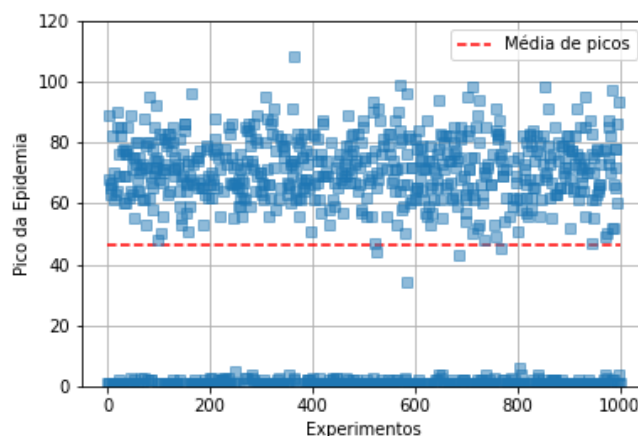
obtendo-se o valor

$$P_{surto} = 1 - \left(\frac{1}{6}\right)^1 \approx 0,83. \quad (126)$$

Dessa forma, há uma probabilidade de, aproximadamente, 83% da doença se propagar na população. É importante observar que, ao se trabalhar com modelos estocásticos, todos os cenários possíveis têm uma certa probabilidade de ocorrência, inclusive a situação de não se ter epidemia. Portanto, para se ter resultados confiáveis utilizando o modelo estocástico, múltiplas simulações devem ser realizadas. Nesse intuito, 1000 simulações do modelo SIR CTMC foram realizadas com o objetivo de fazer estatística sobre os resultados obtidos. Dentre todas as simulações, houve também a ocorrência do caso em que não ocorreria a epidemia na população, como foi ilustrado na Figura 5c.

Também na Figura 5, observa-se que os valores encontrados nas simulações, muitas vezes diferem do caso real do surto de sarampo [34], no qual baseamos esse exemplo. Isso é normal. Por uma parte, como foi acima explicado, a modelagem estocástica oferece uma ampla gama de cenários, todos os quais têm uma certa probabilidade de ocorrência. Por outra parte, na previsão inicial de uma epidemia descrita pelo modelo compartimental SIR, não está se considerando a influência de políticas de saúde e controle do contágio, para o qual talvez seja necessário um modelo mais complexo. Contudo, na modelagem estocástica, não faz sentido analisar simulações isoladas. É preciso realizar uma análise estatística dos resultados, para o qual é indispensável a realização de um grande número de simulações.

Figura 6 - Média do pico da epidemia no surto de sarampo.

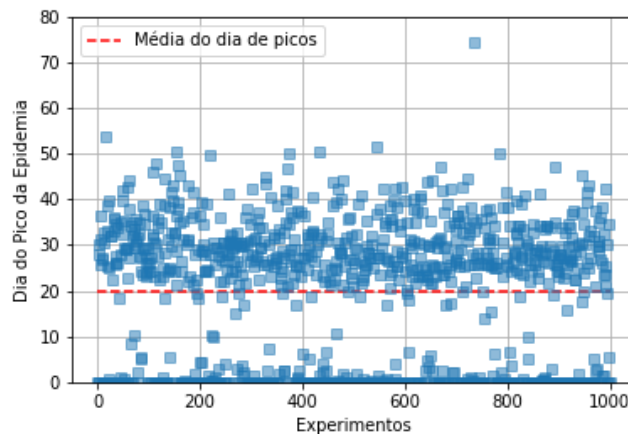


Legenda: Média do pico da epidemia para 1000 simulações do modelo estocástico. Em média, acumularam-se 46 indivíduos contagiados, aproximadamente, no mesmo dia.

Fonte: A autora, 2020.

Usando-se os resultados das 1000 simulações realizadas, foram calculadas as médias das grandezas de interesse na modelagem. Dessa forma, encontrou-se que o pico da epidemia aconteceria, em média, com a infecção de, aproximadamente, 46 indivíduos ao redor do vigésimo dia. Esse é um valor razoável na descrição do surto de sarampo, considerando que foram contagiados 50 indivíduos no total, ao longo de 60 dias que durou a epidemia. Em contraste, o modelo SIR determinístico forneceu uma estimativa de, aproximadamente, 65 indivíduos contagiados no dia de pico, que foi o 30º dia. Nas Figuras 6 e 7, mostram-se estes resultados.

Figura 7 - Média do dia de pico da epidemia no surto de sarampo.



Legenda: Média do dia de pico da epidemia para 1000 simulações do modelo estocástico em, aproximadamente, o vigésimo dia.

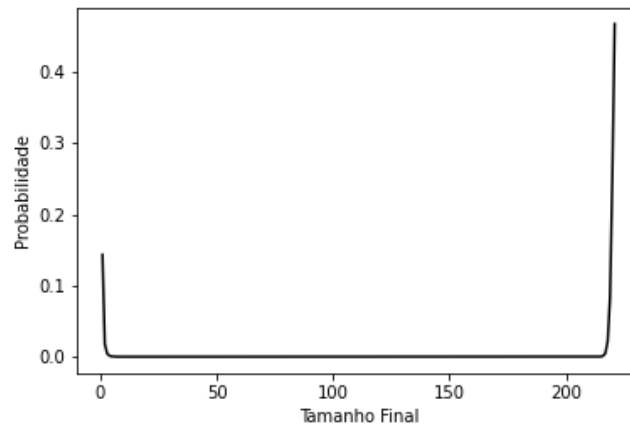
Fonte: A autora, 2020.

Uma das vantagens da modelagem estocástica em relação à determinística é a possibilidade de gerar distribuições de probabilidades para algumas propriedades interessantes no estudo de uma doença. Por exemplo, pode-se encontrar uma distribuição de probabilidade para o tamanho final da epidemia. Através dessa distribuição tem-se acesso à probabilidade de se chegar a um número qualquer de infectados pela doença, não apenas ao resultado previsto de tamanho final. O tamanho final da epidemia, como visto no Capítulo 4, representa a força da epidemia, ou seja, quantos indivíduos ela infectou no total. O tamanho final é dado pela expressão (91),  $R(\infty) = N - S(\infty)$ .

No caso em estudo, a distribuição de probabilidade para o tamanho final da epidemia pode ser observada na Figura 8. Analisando essa distribuição de probabilidade, pode-se observar um pico inicial, que representa a probabilidade de não ocorrer um surto epidêmico na população. Esta probabilidade é de, aproximadamente, 0,17. A probabilidade de surto é de 0,833, calculada usando-se a expressão (126). Esta probabilidade de surto encontra-se distribuída para todos os casos possíveis de tamanho final da epi-

ma, com a maior probabilidade associada ao tamanho final de, aproximadamente, 221 contágios. Consideramos, nesse caso, que um surto da epidemia ocorreu quando mais de 10 indivíduos foram infectados. Esses resultados são ilustrados nas Figuras 8 e 9.

Figura 8 - Distribuição de Probabilidade do tamanho final da epidemia



Legenda: O pico inicial corresponde à probabilidade de não se ter surto, enquanto que a soma das probabilidades restantes corresponde à probabilidade de surto dada pela expressão (126).

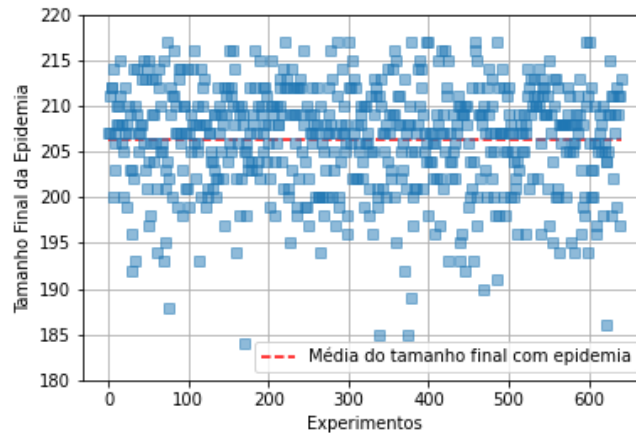
Fonte: A autora, 2020.

Por fim, outra propriedade que o modelo SIR CTMC oferece é a duração esperada da epidemia. A duração da epidemia representa o tempo  $t$  que se leva até  $I(t) = 0$ , como visto no Capítulo 4. Por meio da expressão (123) pode-se obter o valor da duração esperada de acordo com o estado inicial do sistema. Nesse caso, utilizamos o estado  $(s, i) = (220, 1)$  que representa o compartimento de suscetíveis para a condição inicial do problema, com 220 indivíduos suscetíveis e apenas 1 indivíduo infectado. A duração esperada iniciando por esse estado foi de, aproximadamente, 63,33 dias. A média da duração da epidemia, calculada através das 1000 simulações, foi de 61,32 dias. O valor encontrado é bem próximo da duração que teve realmente o surto de sarampo, que foi de 60 dias. Esse resultado é mostrado na Figura 10.

## 5.2 O Grande Desafio: COVID-19 no Rio de Janeiro

A emergência do novo Coronavírus (SARS-CoV-2) no fim de 2019, gerando uma pandemia sem precedentes na história recente da humanidade, catalisou a pesquisa na modelagem matemática aplicada à epidemiologia. Como foi visto no Capítulo 4, enquanto os modelos determinísticos descrevem uma dinâmica entre compartimentos, os modelos es-

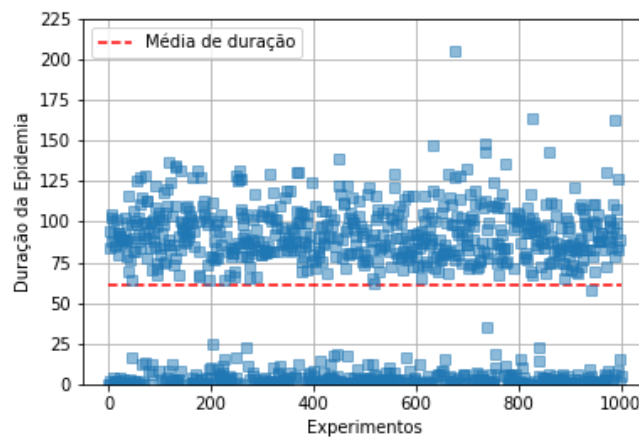
Figura 9 - Média do tamanho final da epidemia



Legenda: Valor médio calculado para o tamanho final considerando todas as situações em que ocorreu epidemia na simulação do modelo SIR CTMC.

Fonte: A autora, 2020.

Figura 10 - Média da duração da Epidemia



Legenda: O valor médio encontrado para a duração da epidemia, através da modelagem estocástica, foi de 61,32 e é muito próximo do valor real da duração do surto de sarampo, que foi de 60 dias.

Fonte: A autora, 2020.



tocásticos consideram a evolução de cada um dos indivíduos na população. Dessa forma, na modelagem estocástica, para populações muito grandes, calcular esses valores tem um custo computacional muito alto [36]. Combinar a capacidade computacional com precisão numérica e um código otimizado continua a ser um desafio nessa área de pesquisa. A continuação, são apresentados alguns resultados do uso do modelo SIR CTMC estudado, na modelagem da atual situação epidêmica no estado de Rio de Janeiro produzida pelo vírus SARS-CoV-2. São apresentados resultados preliminares e definidas as linhas nas quais continuamos a trabalhar para a obtenção da distribuição do tamanho final da epidemia, visando ampliar o tamanho da população estudada [37]. Dessa forma, procura-se contribuir no desenvolvimento dos métodos estocásticos na modelagem epidemiológica e não se faz uma avaliação real da situação pandêmica no Rio de Janeiro.

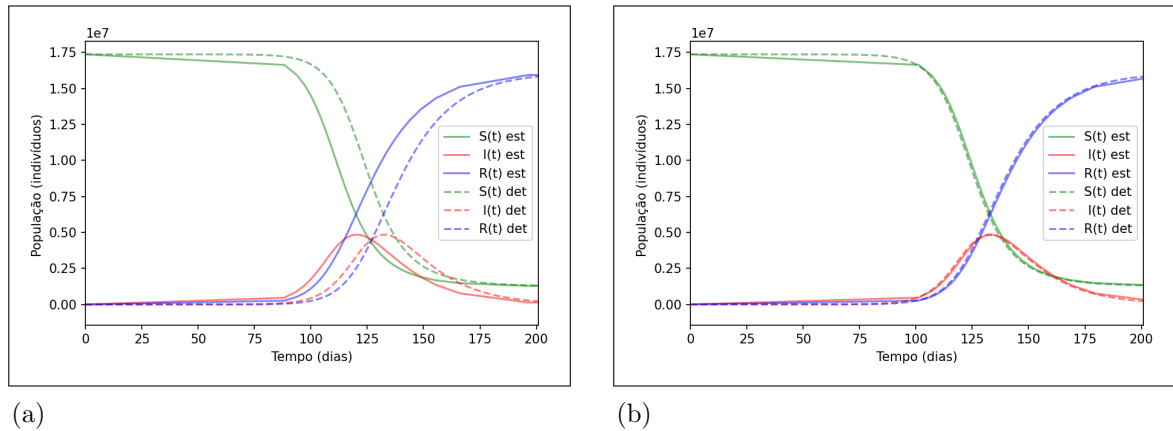
Para realizar as simulações foi considerada a população total do Estado do Rio de Janeiro,  $N = 17.366.189$ , de acordo com os dados do Instituto Brasileiro de Geografia e Estatística (IBGE) [38]. Para o número básico de reprodução, foi utilizado o valor  $\mathcal{R}_0 = 2,83$ , previamente encontrado em [39]. A taxa de recuperação foi definida por  $\gamma = 14^{-1}$ , em consonância com os 14 dias estabelecidos como tempo médio de recuperação da doença. Por fim, usando a expressão (93),  $\mathcal{R}_0 = \beta/\gamma$ , obtem-se a taxa de infecção  $\beta = 0,202$ .

Considerando inicialmente apenas um indivíduo infectado na população, ou seja, com a condição inicial  $I(0) = 1$ , foi realizado um conjunto de simulações para analisar os possíveis cenários gerados a partir do modelo estocástico em comparação ao modelo determinístico. Como o valor da população total é muito alto, foi utilizado um código otimizado para realizar as simulações, exposto em [40, 41]. Empregando o modelo determinístico, para os valores acima considerados dos parâmetros, foi observado que o pico da epidemia teria ocorrência no 132º dia. Em contraste, o modelo estocástico não gera uma evolução única, de modo que para cada simulação encontra-se um valor específico para o pico da epidemia. O modelo estocástico gera uma coleção de resultados possíveis para a dinâmica evolutiva da doença, incluindo o resultado previsto do determinístico, que é ilustrado na Figura 11b. Nas Figuras 11a e 11b mostram-se dois cenários diferentes, ilustrando-se duas possíveis situações geradas pelo modelo estocástico. Pode-se observar que o comportamento das simulações realizadas do modelo estocástico é semelhante ao do modelo determinístico, porém não idêntico. O pico da epidemia nessas simulações estocásticas ocorre nos dias 121 e 134, respectivamente.

Através da equação (101) pode-se determinar a probabilidade de ocorrência de surto da doença na população, como feito na seção anterior. Portanto, aplicando o  $R_0$  utilizado nas simulações, obtem-se

$$P_{surto} = 1 - \left( \frac{1}{2,83} \right)^1 \approx 0,647. \quad (127)$$

Figura 11 - Comparação dos modelos estocástico (linha cheia) e determinístico (linha tracejada)



Legenda: Diferentes cenários produzidos pela modelagem estocástica. (a) Modelo Estocástico com pico da epidemia no 121º dia. (b) Modelo Estocástico com pico da epidemia no 134º dia.

Fonte: A autora, 2020.

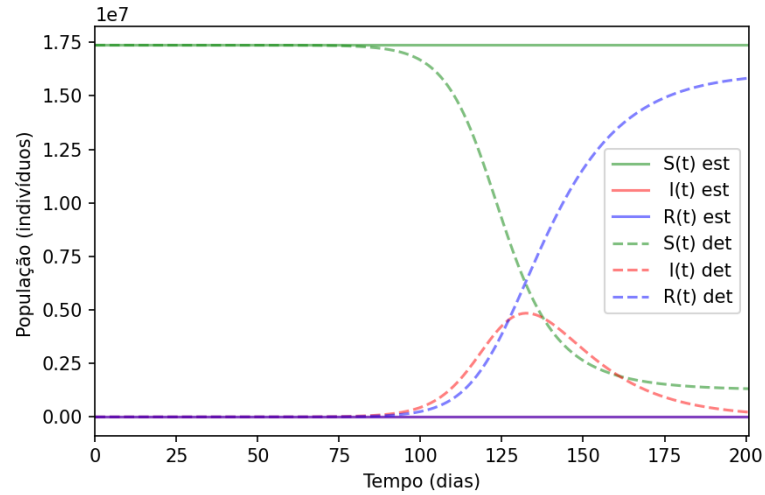
Na Figura 12 pode-se observar uma trajetória do modelo estocástico que não gerou um comportamento epidêmico, em contraste com as curvas obtidas no modelo determinístico.

Com o objetivo de verificar os dias de pico que o modelo estocástico ofereceria, foram realizadas 1000 simulações do modelo SIR CTMC, como feito na seção anterior. A partir dos resultados numéricos, foi obtida uma probabilidade de 0,653 para a ocorrência da epidemia, ou seja, um valor muito próximo do esperado, calculado em (127). Na Figura 13, pode-se observar o resultado dos experimentos onde houve a ocorrência de epidemia e seus respectivos dias de pico. O valor médio dos dias de pico foi de 135, muito próximo do pico da epidemia segundo o modelo determinístico. Esse resultado é ilustrado também na Figura 13.

Para o tamanho final da epidemia, usando o modelo SIR determinístico com os valores estabelecidos para os parâmetros, chega-se a  $I(t) = 0$ , para  $t \approx 417$ . Nesse caso, o tamanho final da epidemia é  $R(417) \approx 16.108.178$ . Considerando que o primeiro caso confirmado no estado do Rio de Janeiro ocorreu em 05 de março de 2020 [42], o tamanho final encontrado ocorreria em 26 de abril de 2021. Note-se que esse número,  $R(417)$ , é extremamente elevado porque nesse modelo não temos considerado nenhuma estratégia de limitação da epidemia.

As propriedades até aqui analisadas correspondem a uma abordagem preliminar no uso do modelo SIR CTMC. Já as propriedades específicas da modelagem estocástica, como a probabilidade de duração ou de tamanho final da epidemia, que foram ilustradas

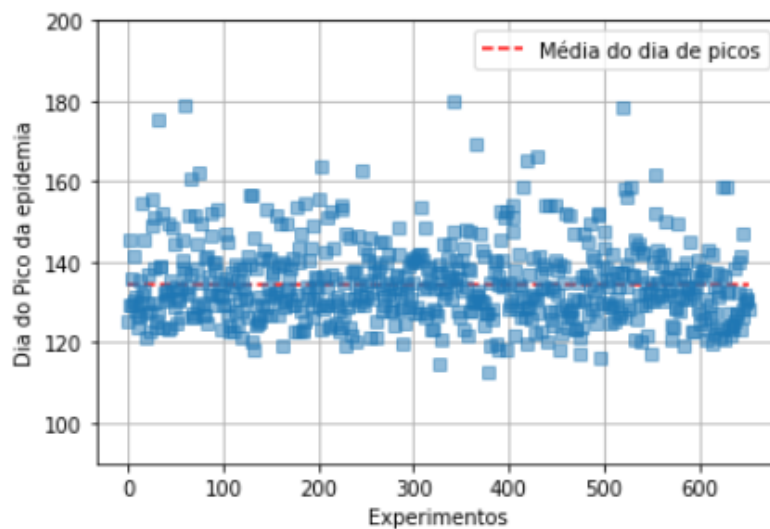
Figura 12 - Comparação dos modelos estocástico e determinístico (não ocorrência da epidemia).



Legenda: Evolução do modelo determinístico (linhas tracejadas), em contraste com a não ocorrência de surto da doença na população segundo o modelo estocástico.

Fonte: A autora, 2020.

Figura 13 - Média do dia de pico da epidemia para a epidemia da COVID-19



Legenda: Ocorrência de epidemia em, aproximadamente, 65.3% dos experimentos realizados, com média do dia de pico da epidemia no valor de 134,59 dias.

Fonte: A autora, 2020.

no caso do surto de sarampo, na seção anterior, não puderam ainda ser calculadas para a do surto da COVID-19. Para encontrar essas distribuições é necessário se ter acesso à matriz geradora  $G$  e à matriz de transição  $P_Y$ , respectivamente. Essas matrizes possuem dimensão  $(N + 1)(N + 2)/2$ . Considerando o valor  $N$  tão alto da população do Rio de Janeiro, o cálculo possui um alto custo computacional que impossibilitou chegar a resultados. Essa é uma limitação da modelagem estocástica, que geralmente é usada para população não muito grandes. Para um tamanho de população na ordem de  $10^4$  ou mais, são necessários métodos alternativos para o cálculo do tempo entre eventos. Tais métodos são necessários pois o tempo entre eventos é tão pequeno que cada caminho de amostra levaria um tempo extremamente longo para ser calculado. Um método numérico mais eficiente para essa situação é conhecido como  $\tau$ -leaping [43]. Neste método, é considerado que mais de um evento ocorra durante o “salto” de um evento a outro, e assim, melhorar o custo computacional que é gasto com populações muito grandes e conseguir gerar as propriedades restantes do modelo estocástico.

## CONCLUSÃO

No presente trabalho, apresentamos uma versão estocástica do modelo compartimental SIR através do uso de Cadeia de Markov de Tempo Contínuo. O modelo SIR CTMC permite a obtenção de informações relacionadas à possível evolução de uma epidemia, às quais não se tem acesso através dos modelos determinísticos. Essas propriedades são, basicamente, a probabilidade de ocorrência de surto da doença, a distribuição de probabilidade para o tamanho final e a duração esperada da epidemia. Foi realizada uma ampla pesquisa sobre o tema, fazendo uma apresentação neste trabalho dos conceitos mais importantes. O trabalho trouxe uma abordagem simples, como uma primeira aproximação ao tema. Dessa forma, foi baseado em estimativas iniciais para os parâmetros e desconsidera os efeitos de medidas (farmacológicas ou não) estabelecidas para controlar a propagação da doença.

Foram considerados dois casos de doenças infecciosas para realizar as simulações computacionais. O primeiro, um surto de sarampo ocorrido em Ilha Grande, no Rio de Janeiro, em 1976; e o segundo, a atual pandemia do coronavírus (SARS-CoV-2) no estado do Rio de Janeiro. A programação foi feita completamente em Python, permitindo a reusabilidade do código. Nos dois exemplos foram realizadas 1000 simulações do modelo estocástico, assim como a simulação do modelo SIR determinístico, para fins de validação e comparação dos resultados.

No caso do surto de sarampo foi possível estimar todas as propriedades específicas da modelagem estocástica, chegando a resultados interessantes quando comparados com os valores reais da referida epidemia. Nesse caso, foi possível mostrar uma superioridade da modelagem estocástica em relação à modelagem determinística.

No caso da COVID-19, foi possível realizar uma modelagem estocástica inicial da evolução esperada para os compartimentos de Suscetíveis, Infectados e Recuperados, em consonância com os resultados da modelagem determinística [37]. O modelo SIR determinístico tem sido usado no estudo da evolução da atual pandemia do coronavírus. Entretanto, dado que a população do estado do Rio de Janeiro é muito grande, o alto custo computacional da modelagem impediu o cálculo de todas as propriedades inerentes ao modelo estocástico. Neste caso, foi implementado um código otimizado para realizar as simulações estocásticas [40, 41]. O tamanho final da epidemia, no modelo estocástico, tem sua propriedade vinculada à matriz de transição  $P_Y$ , onde pode-se determinar a distribuição de probabilidade para esse conceito. Como foi indicado na Seção 4.3, para se chegar nessa distribuição, é necessário trabalhar com matrizes de dimensão da ordem  $N^2$  e calcular uma potência maior que  $N$  dessa matriz [Equação (115)]. O custo computacional, devido ao número total da população utilizada neste caso, é extremamente alto e não foi possível chegar a resultados conclusivos. Temos uma pesquisa em andamento relacionada

à otimização do código, de modo a possibilitar o trabalho com populações muito grandes.

Também como uma extensão do trabalho durante esse período, foi realizada uma generalização estocástica de uma variação do modelo SIR, o modelo SIQR. O modelo SIQR tem sido também utilizado para estudar a propagação da COVID-19. Nesse intuito, elaboramos uma versão estocástica do modelo, também utilizando Cadeia de Markov de Tempo Contínuo. Algumas das propriedades oferecidas por esse modelo ainda estão em fase de desenvolvimento. Por ser ainda um trabalho em andamento, esse modelo foi colocado em Apêndice.

Como trabalhos futuros, para contornar o problema do custo computacional e conseguir realizar uma modelagem estocástica para populações grandes, estamos desenvolvendo duas linhas de trabalho. Por uma parte, estamos estudando métodos alternativos para efetuar o cálculo do tempo entre eventos, como por exemplo, o  $\tau$ -*leaping*, e vamos otimizar o código na linguagem Python para encontrar a distribuição de probabilidade do tamanho final da epidemia; por outra, estamos tentando encontrar formas analíticas recorrentes que permitam calcular os elementos não nulos da matriz final. Ademais, pretendemos desenvolver as propriedades que podem ser oferecidas pelo modelo SIQR estocástico.

## REFERÊNCIAS

- 1 FILHO, N. A.; BARRETO, M. L. *Epidemiologia e saúde: Fundamentos, métodos e aplicações*. Rio de Janeiro: Guanabara Koogan, 2011.
- 2 WALDMAN, E. A. *Vigilância em Saúde Pública*. 7<sup>a</sup>. ed. São Paulo: Fundação Peirópolis, 1998. 267 p.
- 3 HETHCOTE, H. W. The mathematics of infectious diseases. *SIAM Review*, p. 599–653, 2000. Vol. 42, No. 4.
- 4 TROTTIER, H.; PHILIPPE, P. Deterministic modeling of infectious diseases: Theory and methods. *The Internet Journal of Infectious Diseases*, 2000. Vol. 1, No. 2.
- 5 KERMACK, W. O.; MCKENDRICK, A. G. A contribution to the mathematical theory of epidemics. *The Royal Society*, p. 700–721, 1927. Vol. 115, No. 772.
- 6 TUCKWELL, H. C.; WILLIAMS, R. J. Some properties of a simple stochastic epidemic model of SIR type. *Mathematical Biosciences*, p. 76–97, 2007. Vol. 208.
- 7 WEISS, H. The SIR model and the foundations of public health. *MATerials MATemàtics*, p. 17, 2013. Vol. 2013, No. 3.
- 8 BALL, F. et al. A stochastic SIR network epidemic model with preventive dropping of edges. *Journal of Mathematical Biology*, Springer, v. 78, n. 6, p. 1875–1951, 2019.
- 9 ALLEN, L. *An Introduction to Stochastic Processes with Applications to Biology*. Texas, USA: Globo, 2010. 486 p.
- 10 JORGENSEN, E.; SOLLESTED, T. A.; TOFT, N. Evaluation of SIR epidemic models in slaughter pig units using monte carlo simulation. p. 10, 2013. Vol. 2013, No. 3.
- 11 RODRIGUES, V. P. et al. Conceitos e ferramentas de epidemiologia: epidemiologia e os determinantes da saúde. 2014.
- 12 MARTINS, A. A. B. et al. *Epidemiologia*. [S.l.]: Sagah Educação S.A., 2018.
- 13 PEREIRA, C.; VEIGA, N. A epidemiologia. de hipócrates ao século xxi. *Millenium-Journal of Education, Technologies, and Health*, n. 47, p. 129–140, 2016.
- 14 ROTHMAN, K. J. *Epidemiology: An Introduction*. 2. ed. New York: Oxford University Press, 2012. 281 p.
- 15 DIETZ, K.; HEESTERBEEK, J. Daniel Bernoulli's epidemiological model revisited. *Mathematical biosciences*, Elsevier, v. 180, n. 1-2, p. 1–21, 2002.
- 16 MERRIL, R. *Introduction to epidemiology*. [S.l.]: Jones & Bartlett Learning, 2006.
- 17 ROSS, R. – Biographical. 2020. <<https://www.nobelprize.org/prizes/medicine/1902/ross/biographical>>. Acesso em: 04-11-2020.
- 18 AYRES, J. R. d. C. M. Epidemiologia, promoção da saúde e o paradoxo do risco. *Revista Brasileira de Epidemiologia*, SciELO Public Health, v. 5, p. 28–42, 2002.

- 19 BARBOSA, J. et al. *Módulos de princípios de epidemiologia para o controle de enfermidades*. [S.l.]: Módulo, 2010.
- 20 KOLMOGOROV, A. N. *Foundations of the Theory of Probability*. [S.l.]: Second English Edition, Dover Publications Inc., 2018.
- 21 TAYLOR, H. M.; KARLIN, S. *An introduction to stochastic modeling*. 3. ed. Londres: Academic press, 1998.
- 22 COX, D.; MILLER, H. *The Theory of Stochastic Processes*. 1. ed. [S.l.]: Chapman and Hall/CRC, 1965.
- 23 MANTEGNA, R.; STANLEY, H. *An Introduction to Econophysics: Correlations and Complexity in Finance*. [S.l.]: Cambridge University Press: Cambridge, 1999.
- 24 FREUND, J. A.; POSCHEL, T. *Stochastic Process in Physics, Chemistry and Biology*. 1. ed. New York: Springer-Verlag Berlin Heidelberg, 2000. 499 p.
- 25 ALBUQUERQUE, J. P. d. A.; FORTES, J. M. P.; FINAMORE, W. A. *Probabilidade, Variáveis Aleatórias e Processos Estocásticos*. 1. ed. [S.l.]: PUC-Rio/Interciência, 2008.
- 26 ALLEN, L. J. A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. *Infectious Disease Modelling*, Elsevier, v. 2, n. 2, p. 128–142, 2017.
- 27 CHANPRASOPCHAI, P.; TANG, I. M.; PONGSUMPUN, P. SIR model for dengue disease with effect of dengue vaccination. *Computational and mathematical methods in medicine*, Hindawi, v. 2018, 2018.
- 28 KUDDUS, A. et al. A modified SIR model to study on physical behaviour among smallpox infective population in bangladesh. *American Journal of Mathematics and Statistics*, v. 4, n. 5, p. 231–239, 2014.
- 29 SIDE, S.; NOORANI, M. A SIR model for spread of dengue fever disease (simulation for south sulawesi, indonesia and selangor, malaysia). *World Journal of Modelling and Simulation*, v. 9, n. 2, p. 96–105, 2013.
- 30 COMUNIAN, A.; GABURRO, R.; GIUDICI, M. Inversion of a SIR-based model: A critical analysis about the application to COVID-19 epidemic. *Physica D: Nonlinear Phenomena*, v. 413, p. 132674, 2020. ISSN 0167-2789.
- 31 COOPER, I.; MONDAL, A.; ANTONOPOULOS, C. A SIR model assumption for the spread of COVID-19 in different communities. *Chaos, Solitons & Fractals*, v. 139, p. 110057, 2020. ISSN 0960-0779.
- 32 NEVES, A.; GUERRERO, G. Predicting the evolution of the COVID-19 epidemic with the A-SIR model: Lombardy, Italy and São Paulo state, Brazil. *Physica D: Nonlinear Phenomena*, v. 413, p. 132693, 2020. ISSN 0167-2789.
- 33 DALEY, D. J.; GANI, J. *Epidemic Modelling: An Introduction*. [S.l.]: Cambridge University Press, 1999. (Cambridge Studies in Mathematical Biology).

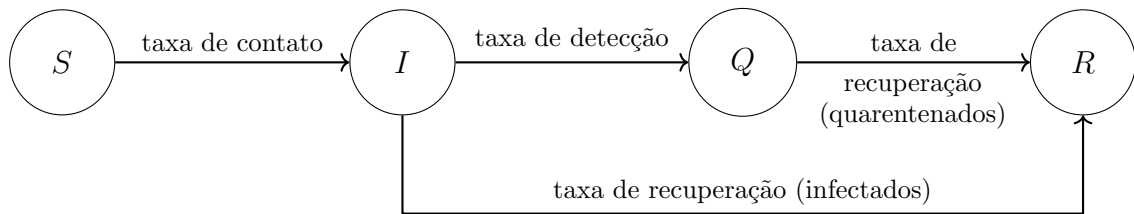


- 34 FILHO, N. A. A.; COURA, J. R. Surto epidêmico de sarampo na Ilha Grande, Rio de Janeiro, Brasil. *Revista da Sociedade Brasileira de Medicina Tropical*, SciELO Brasil, v. 13, n. 1, p. 147–155, 1980.
- 35 SZUSZ, E.; GARRISON, L.; BAUCH, C. A review of data needed to parameterize a dynamic model of measles in developing countries. *BMC Research Notes*, PMC, v. 3, n. 75, 2010.
- 36 GREENWOOD, P.; GORDILLO, L. Stochastic epidemic modeling. In: CHOWELL, G. et al. (Ed.). *Mathematical and Statistical Estimation Approaches in Epidemiology*. [S.l.]: Springer, Dordrecht, 2009. p. 31–52.
- 37 LAU, M.; MORGADO, I.; ARENAS, Z. G. Modelo SIR estocástico na transmissão da COVID-19 no Rio de Janeiro - abordagem preliminar. In: *Anais do XXIII Congresso Brasileiro de Automática*. [S.l.: s.n.], 2020.
- 38 IBGE. *População estimada: IBGE, Estimativas da população residente com data de referência de 1º de julho de 2019*. 2019.
- 39 VOLPATTO, D. et al. Spreading of COVID-19 in Brazil: Impacts and uncertainties in social distancing strategies. *medRxiv*, p. 2020.05.30.20117283, 2020.
- 40 MORGADO, I.; LAU, M. *Biblioteca Python Modelo SIR estocástico*. 2020. Disponível em: <https://github.com/igormorgado/sir>. Acesso em: 15/09/2020.
- 41 MORGADO, I.; LAU, M.; ARENAS, Z. G. Otimização computacional de cadeias de markov em tempo contínuo aplicadas ao modelo SIR utilizando sub amostragem. In: *Anais do XXIII ENMC – Encontro Nacional de Modelagem Computacional e XI ECTM – Encontro de Ciências e Tecnologia de Materiais*. [S.l.: s.n.], 2020.
- 42 SES-RJ. *Painel Coronavirus COVID-19, Secretaria de Saúde do Estado do Rio de Janeiro*. 2020. <http://painel.saude.rj.gov.br/monitoramento/covid19.html>. [Online; acesso em 21-Setembro-2020].
- 43 GILLESPIE, D. T. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of chemical physics*, American Institute of Physics, v. 115, n. 4, p. 1716–1733, 2001.
- 44 HETHCOTE, Herbert; ZHIEN, Ma; SHENGBING, Liao. Effects of quarantine in six endemic models for infectious diseases. *Mathematical biosciences*, Elsevier, v. 180, n. 1-2, p. 141–160, 2002.
- 45 CROKIDAKIS, Nuno et al. Modeling the early evolution of the COVID-19 in Brazil: Results from a susceptible–infectious–quarantined–recovered (SIQR) model. *International Journal of Modern Physics C (IJMPC)*, World Scientific Publishing Co. Pte. Ltd., v. 31, n. 10, p. 1–7, 2020.

## APÊNDICE – Modelo SIQR

O modelo compartimental SIQR é uma variação do modelo compartimental SIR, onde há um compartimento extra,  $Q$ , denominado *quarentenado*. Esse compartimento é composto pelos indivíduos da classe infecciosa que são identificados com a doença e são isolados voluntariamente ou coercivamente. Para algumas doenças mais leves, indivíduos em quarentena podem ser pessoas que optam por ficar em casa porque estão doentes. Para outras doenças mais graves, as pessoas em quarentena podem ser aquelas que são forçadas ao isolamento. Presume-se que esses indivíduos em quarentena não se misturem com outros, de modo que eles não transmitam a doença para indivíduos suscetíveis [44].

Figura 14 - Diagrama do modelo *SIQR*.



Legenda: Diagrama do modelo SIQR mostrando as relações entre os compartimentos e as respectivas taxas.

Fonte: A autora, 2020.

Neste modelo também consideramos a população constante, a qual é denotada por  $N$ , onde  $N = S(t) + I(t) + Q(t) + R(t)$ . A partir desse modelo foram estudados o modelo determinístico e sua contraparte estocástica.

### A.1 Modelo Determinístico

O modelo *SIQR*, em sua versão determinística, é descrito pelas seguintes equações diferenciais [45],

$$\frac{dS}{dt} = \frac{-\beta}{N}SI, \quad (128)$$

$$\frac{dI}{dt} = \frac{\beta}{N}SI - (\alpha + \eta)I, \quad (129)$$

$$\frac{dQ}{dt} = \eta I - \gamma Q, \quad (130)$$

$$\frac{dR}{dt} = \gamma Q + \alpha I. \quad (131)$$

Inicialmente, considera-se que os compartimentos  $Q(0) = R(0) = 0$ , pois ainda não foram isolados os indivíduos infectados e não há indivíduos recuperados, e também que  $S(0) = N - I(0)$ . Nas equações acima,  $\beta$  é a taxa de contato ou taxa de infecção,  $\alpha$  é a taxa de recuperação de indivíduos assintomáticos ou que não se isolaram,  $\eta$  é a taxa de detecção de novos casos, ou seja, dos que testaram positivo para a doença e se isolaram e, por fim,  $\gamma$  é a taxa de recuperação de um indivíduo quarentenado.

Considerando que no início da epidemia  $I(0)$  é muito pequeno, então  $S(0) \approx N$ . Nesse caso, a expressão (129) pode ser simplificada, tal que,

$$\frac{dI}{dt} = (\beta - (\alpha + \eta))I. \quad (132)$$

Resolvendo a expressão (132),

$$I(t) = I(0)e^{(\beta - (\alpha + \eta))t}. \quad (133)$$

A expressão (133) nos dá a variação dos indivíduos infectados no decorrer do tempo, e assim pode-se obter uma solução para os outros compartimentos.

O número de reprodução básico desse modelo pode ser obtido através da razão entre taxa de novos infectados e a taxa de indivíduos removidos do compartimento de infectados. Ou seja,

$$\mathcal{R}_0 = \frac{\beta SI/N}{(\alpha + \eta)I} = \frac{\beta}{\alpha + \eta}. \quad (134)$$

Conseqüentemente, o número de reprodução efetivo é dado por,

$$\mathcal{R} = \frac{\beta}{\alpha + \eta} \frac{S(t)}{N}. \quad (135)$$

## A.2 Modelo Estocástico

No modelo SIQR CTMC os compartimentos  $S, I, Q$  e  $R$  são considerados como variáveis aleatórias. As probabilidades de transição são definidas a seguir em termos das taxas do modelo determinístico, considerando a mudança dos estados  $S, I$  e  $Q$  durante um intervalo de tempo pequeno  $\Delta t$  como  $\Delta S = S(t + \Delta t) - S(t)$ ,  $\Delta I = I(t + \Delta t) - I(t)$  e  $\Delta Q = Q(t + \Delta t) - Q(t)$ :

Tabela 3 - Probabilidades de transição para o modelo SIQR CTMC.

$(\Delta S, \Delta I, \Delta Q)$	Probabilidade
$(-1, 1, 0)$	$\frac{\beta}{N} S(t)I(t)\Delta t + o(\Delta t)$
$(0, -1, 0)$	$\alpha I(t)\Delta t + o(\Delta t)$
$(0, -1, 1)$	$\eta I(t)\Delta t + o(\Delta t)$
$(0, 0, -1)$	$\gamma Q(t)\Delta t + o(\Delta t)$
$(0, 0, 0)$	$1 - \left[ \frac{\beta}{N} S(t)I(t) + \alpha I(t) + \eta I(t) + \gamma Q(t) \right] \Delta t$

Legenda: Representação das probabilidades de transição utilizadas no modelo SIQR.

Fonte: A autora, 2020.

Os estados do sistema,  $S, I, Q$  e  $R$  representam a quantidade de indivíduos que pertencem ao compartimento correspondente. Dessa forma, a probabilidade condicional acima representa a transição da variação de suscetíveis e infectados considerando que a evolução do sistema é dada através da variação de somente um indivíduo em um pequeno intervalo de tempo  $\Delta t$  e, para cada caso de ocorrência do sistema há uma probabilidade associada. Por esse motivo, não há probabilidade de transição de dois compartimentos simultaneamente.

Cada  $(\Delta S, \Delta I, \Delta Q) = (s, i, q)$  representa uma transição possível do sistema. O compartimento  $R$  não é necessário para avaliar essas transições, uma vez que não há uma variação decrescente em seu compartimento. Portanto, quando  $(s, i, q) = (-1, 1, 0)$  consideramos que houve uma infecção de um indivíduo suscetível. Quando  $(s, i, q) = (0, -1, 0)$  consideramos que um indivíduo do compartimento infectado se recuperou. Se  $(s, i, q) = (0, -1, 1)$  significa que um indivíduo infectado entrou para o compartimento de quarentena. E, quando  $(s, i, q) = (0, 0, -1)$  significa que um indivíduo que estava em quarentena se recuperou.

Inicialmente, consideraremos a distribuição do sistema como  $(S(0), I(0), Q(0)) = (s_0, i_0, q_0)$ , onde  $s_0 + i_0 = N$  e  $Q(0) = R(0) = 0$ , pois considera-se que neste instante não há indivíduos quarentenados nem recuperados e  $s_0 \geq 0$  e  $i_0 > 0$ .

No modelo SIQR CTMC também é possível obter algumas propriedades que o modelo determinístico não oferece, como no modelo SIR CTMC. A probabilidade de

surto de uma doença em determinada população, a distribuição de probabilidade para o tamanho final da epidemia e para a duração esperada de uma epidemia, são propriedades desses modelos.

A propriedade para determinar se ocorrerá um surto na população é dada pela mesma equação do modelo SIR CTMC. Dessa forma, pode-se obter informações dessa propriedade pela expressão (101), utilizando o  $R_0$  definido em (134).

O tempo entre eventos para realizar as simulações computacionais nesse modelo, também é definido utilizando a expressão (44), onde o denominador dessa expressão é dado por  $\beta S(t)I(t)/N + \alpha I(t) + \eta I(t) + \gamma Q(t)$ , portanto

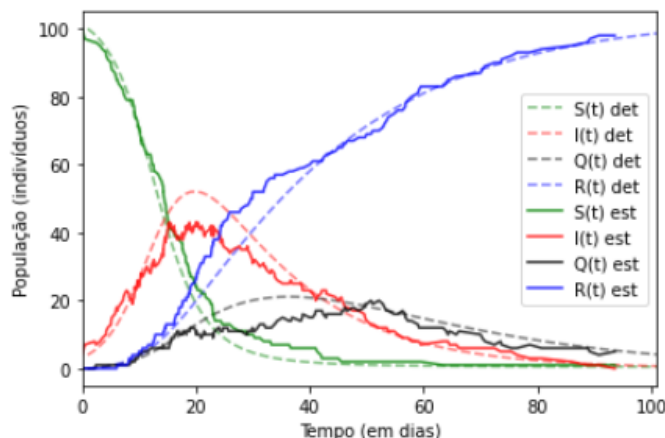
$$T_E = F^{-1}(U) = -\frac{\ln U}{\beta S(t)I(t)/N + \alpha I(t) + \eta I(t) + \gamma Q(t)} \quad (136)$$

sendo  $U$  uma variável aleatória uniforme no intervalo  $[0, 1]$ .

O maior desafio no modelo SIQR CTMC é o cálculo da distribuição do tamanho final da epidemia e da duração esperada da epidemia. Essas propriedades estão atreladas ao conhecimento da matriz geradora  $G$  e da matriz de transição  $P_Y$ , porém, ambas as matrizes dependem da ordenação dos estados do modelo. Essa ordenação ainda não foi determinada de modo que possamos encontrar um padrão para realizar as operações futuras. Por esse motivo, ainda não há resultados acerca dessas propriedades.

Na Figura (15) pode-se observar que o comportamento do modelo SIQR estocástico, ou seja, a forma como as curvas referentes a cada compartimento evoluem ao decorrer do tempo, é semelhante ao modelo SIQR determinístico, como esperado.

Figura 15 - Comparação da simulação dos modelos SIQR determinístico e estocástico



Legenda: Ilustração de uma única simulação do modelo estocástico (linha cheia) em comparação ao modelo determinístico (linhas tracejada).

Fonte: A autora, 2020.