

# Modelo SIR Estocástico na Transmissão da COVID-19 no Rio de Janeiro — Abordagem preliminar<sup>\*</sup>

Michelle Lau de Almeida<sup>\*</sup> Igor Morgado<sup>\*\*</sup>  
Zochil González Arenas<sup>\*\*\*</sup>

<sup>\*</sup> Pós-graduação em Ciências Computacionais (PPG-CCOMP),  
Instituto de Matemática e Estatística (IME), Universidade do Estado  
do Rio de Janeiro (UERJ), RJ (e-mail: michelle.lau@ime.uerj.br).

<sup>\*\*</sup> PPG-CCOMP, IME/UERJ, RJ (e-mail: igor.morgado@ime.uerj.br)

<sup>\*\*\*</sup> Departamento de Matemática Aplicada, IME, UERJ, RJ  
(e-mail: zochil@ime.uerj.br)

---

**Abstract:** The recent Coronavirus (COVID-19) spread quickly around the world, boosting the research of mathematical modeling applied to epidemiology. One of the commonly used models is the deterministic version of the Susceptible-Infected-Recovered (SIR) model. In this work, we present the stochastic version of the SIR model, using Continuous Time Markov Chain (CTMC). Some numerical simulations were performed in order to compare the two models using public data on COVID-19 and parameter values reported in the literature. Computing some properties of stochastic models demands many computational resources, limiting its usability in large populations. Preliminary results as well as a discussion of future work are presented.

**Resumo:** O novo Coronavirus (COVID-19) espalhou-se rapidamente pelo mundo e dinamizou a pesquisa da modelagem matemática aplicada a epidemiologia. Um dos modelos utilizados na modelagem dessa doença é o Suscetível-Infetado-Recuperado (SIR) utilizando uma abordagem determinística. Neste trabalho, apresentamos o modelo SIR com uma abordagem estocástica, utilizando Cadeia de Markov de Tempo Contínuo (CTMC). Foram realizadas simulações a fim de comparação entre os dois modelos utilizando dados públicos sobre a COVID-19 e valores de parâmetros reportados na literatura. Algumas das propriedades dos modelos estocásticos possuem um custo computacional elevado que limitam o seu uso em populações grandes. Resultados preliminares são apresentados assim como uma discussão dos trabalhos futuros.

*Keywords:* Stochastic epidemic modeling; Compartmental models; CTMC SIR model; COVID-19; Python

*Palavras-chaves:* Modelagem epidemiológica estocástica; Modelos compartimentais; Modelo SIR CTMC; COVID-19; Python

---

## 1. INTRODUÇÃO

A emergência do novo Coronavirus (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. Nessa área, 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. 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 (Hethcote, 2000).

Largamente usados na simulação e estudo de doenças infecciosas, os modelos compartimentais dividem a população em classes (ou compartimentos), que representam os vários estados da doença. Esses modelos podem ser determinísticos ou estocásticos. Nos primeiros, o processo epidêmico é representado por um conjunto de equações diferenciais ordinárias que, dada uma condição inicial, descrevem a evolução temporal da doença. Por outro lado, por meio dos modelos estocásticos é possível inserir uma certa aleatoriedade nos parâmetros. Esses modelos estimam as distribuições de probabilidade dos resultados possíveis e, dessa forma, oferecem diferentes cenários aos quais não se tem acesso através dos modelos determinísticos. Uma diferença importante entre esses modelos é sua dinâmica assintótica. Um modelo determinístico é utilizado para prever um único resultado de um determinado conjunto de circunstâncias. Já um modelo estocástico prevê um

---

<sup>\*</sup> O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 e da Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ).

conjunto de possíveis resultados ponderados por suas probabilidades.

Os modelos epidemiológicos estocásticos possuem algumas vantagens exclusivas, como o acesso a informação em relação à probabilidade de um surto, bem como à duração esperada e à distribuição do tamanho final de uma epidemia (Allen, 2008). Podem ser também considerados efeitos sobre os quais não se tem conhecimento suficiente ou que podem ser aleatórios. Dessa forma, é possível fazer uma modelagem mais realista. Entretanto, existe uma limitação importante para o uso de modelos estocásticos que é relacionada ao tamanho da população em estudo. Enquanto os modelos determinísticos descrevem uma dinâmica entre compartimentos, os modelos estocásticos consideram a evolução de cada indivíduo da população. Assim, os modelos determinísticos são convenientemente usados na epidemiologia matemática com populações grandes, como acontece especificamente no caso da COVID-19 (Crokidakis, 2020; Manrique-Abril et al., 2020; Volpatto et al., 2020). Na modelagem estocástica, para populações muito grandes, calcular esses valores tem um custo computacional muito alto (Greenwood and Gordillo, 2009). Combinar a capacidade computacional com precisão numérica e um código otimizado continua a ser um desafio nessa área de pesquisa.

Estimar a duração de uma epidemia e o número de pessoas que pode adoecer é um dos objetivos mais importantes na modelagem epidemiológica matemática. 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 (Kermack and McKendrick, 1927), que divide a população nos estados de (S)uscetíveis, (I)nfectados e (R)ecuperados, tem sido usado amplamente no estudo de diversas doenças e, em particular, da COVID-19 (Comunian et al., 2020; Cooper et al., 2020; Neves and Guerrero, 2020).

Nesse trabalho, faz-se uma introdução à modelagem estocástica usando um modelo SIR estocástico. O modelo será usado no estudo do contágio pela COVID-19. A evolução dos indivíduos suscetíveis, infectados e recuperados no estado do Rio de Janeiro será obtida através de simulação numérica em Python e comparada a resultados prévios obtidos por modelos determinísticos. Na obtenção do tamanho final da epidemia, estamos trabalhando para ampliar o tamanho da população a ser estudada. Resultados preliminares serão apresentados. Esse trabalho não pretende fazer uma avaliação real da situação pandêmica no Rio de Janeiro e sim contribuir no desenvolvimento dos métodos estocásticos na modelagem epidemiológica.

O trabalho está organizado da seguinte forma. Na próxima seção, faz-se uma introdução a conceitos importantes na modelagem estocástica. Na Seção 3 é definido o modelo compartimental SIR utilizado, nas versões determinística e estocástica. Após, na Seção 4, são apresentados os resultados das simulações realizadas e, por fim, discute-se acerca do trabalho futuro.

## 2. CONCEITOS PRELIMINARES

A estrutura natural com a qual se trabalha nos modelos compartimentais estocásticos são os processos de Markov.

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, denominado espaço de estados. Para cada  $t$  fixado,  $X_t(\omega)$  denota uma única variável aleatória definida em  $\Omega$ . Para cada  $\omega \in \Omega$  fixado,  $X_t(\omega)$  corresponde a uma função definida em  $T$  que é chamada de caminho de amostra ou realização estocástica do processo.

Um processo estocástico markoviano caracteriza-se por não ter memória, isto é, o conhecimento do estado futuro do processo depende só do seu estado atual. Qualquer conhecimento prévio não aporta mais informação. Matematicamente, isto é representado da seguinte forma,

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

para  $t_0 < t_1 < \dots < t_k < t_{k+1}$ ,  $t_i \in T$ , com  $i = 0, 1, 2, \dots, k+1$ , onde  $x_i$  representa o estado que a variável aleatória pode assumir.

Se as variáveis aleatórias  $X_t(\omega)$  assumem valores discretos, então o processo markoviano 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* pelas siglas em inglês (Discrete Time Markov Chain), ou como Cadeia de Markov de Tempo Contínuo, *CTMC* (Continuous Time Markov Chain). Os modelos compartimentais são CTMC, pois suas variáveis aleatórias assumem valores discretos e o tempo é analisado em um intervalo contínuo.

Cada variável aleatória  $X(t)$  do processo tem uma distribuição de probabilidade  $\{p_i(t)\}_{i=0}^{\infty}$  associada, com

$$p_i(t) = \text{Prob}\{X_t = i\}. \quad (2)$$

A expressão (2) representa a probabilidade dos estados possíveis que a variável aleatória  $X(t)$  pode assumir. Essa distribuição de probabilidade pode ser representada através do vetor  $p(t) = (p_0(t), p_1(t), \dots)^T$ , onde  $T$  denota a transposta.

Em uma cadeia de Markov é possível relacionar as variáveis aleatórias do processo, por exemplo, as variáveis  $X_s$  e  $X_t$ , considerando  $s < t$  para obter informações acerca do sistema. Essa relação entre elas é conhecida como *probabilidade de transição* entre estados.

A probabilidade de transição, denotada por  $p_{ji}(t, s)$  é definida como a seguinte probabilidade condicional:

$$p_{ji}(t, s) = \text{Prob}\{X_t = j | X_s = i\}, \quad s < t \quad (3)$$

A expressão definida em (3) representa a probabilidade da variável aleatória  $X$  no instante de tempo  $t$  encontrar-se no estado  $j$  dado que no instante de tempo  $s$ , estava no estado  $i$ , para  $i, j = 0, 1, 2, \dots$ . Por outro lado, as probabilidades de transição são soluções das equações de Chapman-Kolmogorov

$$\sum_{k=0}^{\infty} p_{jk}(s)p_{ki}(t) = p_{ji}(t+s). \quad (4)$$

Se as probabilidades de transição não dependem do tempo  $s$  ou  $t$ , mas sim do tamanho do intervalo  $t - s$ , elas são chamadas de probabilidades de transição estacionárias. Por conveniência, as probabilidades de transição tratadas neste trabalho são assumidas como estacionárias. Portanto, tem-se,

$$p_{ji}(t - s) = \text{Prob}\{X(t - s) = j | X(0) = i\}, \quad s < t \quad (5)$$

Define-se a matriz das probabilidades de transição ou, simplesmente, matriz de transição, a matriz que tem a probabilidade  $p_{ji}$  como elemento na posição  $ji$ , *i.e.*,

$$P(t) = (p_{ji}(t)).$$

A matriz de transição possui algumas propriedades particulares. Todos os seus elementos são não-negativos,  $p_{ji}(t) \geq 0$ , por se tratarem de probabilidades, e a soma dos elementos por coluna é igual a 1,

$$\sum_{j=0}^{\infty} p_{ji}(t) = 1, \quad t \geq 0 \quad (6)$$

Essas propriedades definem a matriz de transição como uma matriz estocástica para todo  $t \geq 0$ .

Além de determinar as possíveis mudanças de um estado para outro no sistema estudado, as probabilidades de transição são também usadas para derivar as taxas de transição  $g_{ji}$ . As taxas de transição são úteis para representar se houve mudança ou não de um estado para outro. Elas podem ser expressas em forma matricial, sendo conhecida como *Matriz Geradora Infinitesimal G*.

Para encontrar os elementos da matriz  $G$ , vamos assumir que as probabilidades de transição são contínuas e diferenciáveis para  $t \geq 0$  e considerar os seus valores iniciais, para  $t = 0$ , como sendo  $p_{ji}(0) = 0$ , para  $j \neq i$  e  $p_{ii}(0) = 1$ . Dessa forma, as taxas de mudança das probabilidades de transição são

$$g_{ji} = \lim_{\Delta t \rightarrow 0^+} \frac{p_{ji}(\Delta t) - p_{ji}(0)}{\Delta t} = \lim_{\Delta t \rightarrow 0^+} \frac{p_{ji}(\Delta t)}{\Delta t}, \quad i \neq j \quad (7)$$

$$g_{ii} = \lim_{\Delta t \rightarrow 0^+} \frac{p_{ii}(\Delta t) - p_{ii}(0)}{\Delta t} = \lim_{\Delta t \rightarrow 0^+} \frac{p_{ii}(\Delta t) - 1}{\Delta t} \quad (8)$$

Reescrevendo apropriadamente a expressão (6) e usando a definição de  $g_{ji}$ , equação (7), segue que

$$1 - p_{ii}(\Delta t) = \sum_{j=0, j \neq i}^{\infty} p_{ji}(\Delta t) = \sum_{j=0, j \neq i}^{\infty} [g_{ji} \Delta t + o(\Delta t)]. \quad (9)$$

Substituindo esse resultado agora na definição de  $g_{ii}$ , equação (8),

$$g_{ii} = \lim_{\Delta t \rightarrow 0^+} \frac{-\sum_{j=0, j \neq i}^{\infty} [g_{ji} \Delta t + o(\Delta t)]}{\Delta t} \quad (10)$$

$$= -\sum_{j=0, j \neq i}^{\infty} g_{ji}. \quad (11)$$

Se  $g_{ii}$  é finito, então  $\sum_{j=0}^{\infty} g_{ji} = 0$  e usando as expressões (7) e (8) chega-se à versão infinitesimal das probabilidades de transição,

$$p_{ji}(\Delta t) = \delta_{ji} + g_{ji} \Delta t + o(\Delta t), \quad (12)$$

onde  $\delta_{ji}$  é a delta do Kronecker.

Seja  $P(\Delta t) = (p_{ji}(\Delta t))$ . Então, a matriz  $G$  é dada por

$$G = \lim_{\Delta t \rightarrow 0^+} \frac{P(\Delta t) - I}{\Delta t}, \quad (13)$$

onde  $I$  é uma matriz identidade da mesma dimensão de  $P(\Delta t)$  (no caso finito). A matriz geradora é importante por dois motivos; ela é usada para definir as equações avançadas de Kolmogorov, as quais expressam a matriz de transição  $P(t)$  através de equações diferenciais,  $dP/dt = GP$ ; e também é usada para encontrar a matriz de transição de uma Cadeia de Markov Incorporada (*EMC*) (*Embedded Markov Chain*) ou *processo de salto*, que será definido a seguir.

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

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

onde  $W_n$  é o tempo no qual o processo permanece no estado  $n$ . O conjunto de variáveis aleatórias  $\{Y_n\}_{n=0}^{\infty}$  é denominado Cadeia de Markov Incorporada e é uma *DTMC*. Em geral, as cadeias de Markov incorporada e de tempo contínuo têm a mesma dinâmica a longo prazo, inclusive o tamanho final da epidemia. As primeiras são muito apropriadas para o trabalho numérico, razão pela qual são usadas na classificação de estados da *CTMC* correspondente (Daley and Gani, 1999).

Uma diferença entre uma *DTMC* e uma *CTMC* é que em cadeias de tempo discreto há um “salto” para um novo estado no tempo discreto  $1, 2, \dots$ , mas em cadeias de tempo contínuo o “salto” pode ocorrer em qualquer tempo  $t \geq 0$ .

Em uma cadeia de Markov de tempo contínuo, com estado inicial  $X(0) = i$ , o processo permanece no estado inicial  $i$  por um período aleatório de tempo  $W_1$  até passar para um novo estado  $j$ , ou seja,  $X(W_1) = j$ . Em seguida, ele permanece no estado  $j$  por um novo período aleatório de tempo  $W_2$  até alcançar o novo estado  $X(W_2) = k$ , e assim sucessivamente. Portanto, define-se  $W_i$  como a variável aleatória para expressar o tempo do  $i$ -ésimo salto. Considera-se, inicialmente,  $W(0) = 0$ . O conjunto de variáveis aleatórias  $\{W_i\}_{i=0}^{\infty}$  é denominado como *tempo de salto* ou *tempo de espera* do processo.

Seja  $\{Y_n\}_{n=0}^{\infty}$  uma *EMC*. Sua matriz de transição é dada por  $P_Y = (\tilde{p}_{ji})$ , onde  $\tilde{p}_{ji} = \text{Prob}\{Y_{n+1} = j | Y_n = i\}$ . A matriz de transição  $P_Y = (\tilde{p}_{ji})$  pode ser definida usando a matriz geradora  $G$ .

Um estado  $i$ , em uma *DTMC*, é *absorvente* se  $g_{ii} = 0$ , ou seja, a taxa de mudança é zero pois o estado não muda. Portanto,

$$\tilde{p}_{ii} = \begin{cases} 0, & \text{se } g_{ii} \neq 0 \\ 1, & \text{se } g_{ii} = 0 \end{cases} \quad (14)$$

Para definirmos a probabilidade de transição  $\tilde{p}_{ji}$  utilizamos as expressões (7) e (8),

$$-\frac{g_{ji}}{g_{ii}} = \lim_{\Delta t \rightarrow 0^+} \frac{p_{ji}(\Delta t)}{1 - p_{ii}(\Delta t)} \quad (15)$$

A expressão (15) é a probabilidade de uma transição do estado  $i$  para o estado  $j$ , dado que o processo não permanece no estado  $i$  ( $g_{ii} \neq 0$ ). Define-se então a probabilidade de transição  $\tilde{p}_{ji}$  como

$$\tilde{p}_{ji} = \begin{cases} \frac{g_{ji}}{\sum_{k=0, k \neq i}^{\infty} g_{ki}} = -\frac{g_{ji}}{g_{ii}}, & \text{se } g_{ii} \neq 0 \\ 0, & \text{se } g_{ii} = 0 \end{cases} \quad (16)$$

No caso de  $g_{ii} = 0$ , o elemento  $(i, i)$  da matriz  $P_Y$  é igual a 1 e os elementos restantes da coluna são iguais a 0.

A matriz  $P_Y$  também é uma matriz estocástica pois possui a soma dos elementos de cada coluna igual a 1. As probabilidades de transição são homogêneas, ou seja, independem de  $n$ . Além disso, se cumpre que  $(P_Y)^n = \tilde{p}_{ji}^{(n)}$ , onde  $\tilde{p}_{ji}^{(n)} = \text{Prob}\{Y_n = j | Y_0 = i\}$ .

Para realizar simulações estocásticas em uma CTMC é necessário conhecer a distribuição para o tempo entre sucessivos eventos, também denominado, *tempo entre eventos*. Vamos definir uma variável aleatória para o tempo entre eventos, denominada  $T_i$ , contínua e não-negativa, de modo que  $T_i = W_{i+1} - W_i \geq 0$ , onde  $W_i$  é o tempo do  $i$ -ésimo salto.

Por outro lado, para calcular numericamente um caminho de amostra de um modelo CTMC, é necessário usar o fato de que o tempo entre eventos tem uma distribuição exponencial. Isso decorre da propriedade de Markov. A distribuição exponencial tem o que é chamado de propriedade sem memória, *memoryless property*. Portanto, a variável aleatória  $T_i$  do tempo entre eventos é uma variável aleatória exponencial.

Considere que no  $i$ -ésimo salto o processo esteja no estado  $n$ , ou seja,  $X(W_i) = n$ . Seja  $\alpha(n)\Delta t + o(\Delta t)$  a probabilidade do processo mover-se para um estado diferente de  $n$  no período de tempo  $\Delta t$ , ou seja,

$$\sum_{j \neq n} p_{jn}(\Delta t) = \alpha(n)\Delta t + o(\Delta t). \quad (17)$$

Considerando que as mudanças de estado do sistema ocorrem com probabilidade  $\alpha(n)\Delta t + o(\Delta t)$ , num intervalo de tempo pequeno  $\Delta t$ , a função de distribuição para o tempo entre eventos é

$$F_i(t) = \text{Prob}\{T_i \leq t\} = 1 - e^{-\alpha(n)t}. \quad (18)$$

Como  $\alpha(n)$  é independente do estado do processo, o tempo entre eventos é o mesmo para todo salto  $i$ . Vamos definir  $T_i \equiv T_E$  para todo  $i$ . Então, o tempo entre eventos,  $T_E$ , é uma variável aleatória contínua definida em  $[0, \infty)$ , e pode ser expressa em função de uma outra variável aleatória uniforme,  $U$ , definida no intervalo  $[0, 1]$ . A partir da expressão (18), pode-se definir o tempo entre eventos,  $T_E$ , como

$$T_E = F^{-1}(U) = -\frac{\ln U}{\alpha(n)}, \quad (19)$$

onde  $F$  é a função de distribuição acumulada da variável aleatória  $T_E$ .

### 3. MODELO COMPARTIMENTAL SIR

#### 3.1 Determinístico

O modelo compartimental considerado neste trabalho é o modelo (S)uscetível-(I)nfetado-(R)ecuperado, também denominado SIR, proposto por Kermack and McKendrick

(1927). Vamos considerar o número total de indivíduos na população dado por  $N$ , assumindo que esse valor é constante, ou seja, não há nascimentos ou mortes durante o estudo da dinâmica evolutiva da doença, e também que não ocorre o período latente, período onde os infectados ainda não transmitem a doença. O modelo determinístico pode ser descrito através das seguintes equações diferenciais,

$$\frac{dS}{dt} = -\frac{\beta}{N}SI, \quad (20)$$

$$\frac{dI}{dt} = \frac{\beta}{N}SI - \gamma I, \quad (21)$$

$$\frac{dR}{dt} = \gamma I, \quad (22)$$

com valores iniciais  $S(0), I(0), R(0) \geq 0$ , tais que  $S(0) + I(0) + R(0) = N$ . Nas equações acima, define-se  $\beta$  como a taxa de contato ou taxa de infecção e  $\gamma$  como a taxa de recuperação de um indivíduo infectado. A equação (20) representa a variação de indivíduos suscetíveis que ao ter contato com indivíduos infectados tornam-se infectados. Na equação (21) 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 (22), tem-se a variação dos indivíduos infectados que se recuperaram.

O conceito do *número básico de reprodução*,  $\mathcal{R}_0$  tem um papel fundamental na modelagem epidemiológica. Ele é utilizado para medir o potencial de transmissão de uma doença e representa o número médio de casos secundários que um indivíduo infeccioso pode produzir, considerando uma população totalmente suscetível. É definido como

$$\mathcal{R}_0 = \frac{\beta}{\gamma} \quad (23)$$

e permite estimar quando ocorrerá uma epidemia. Se  $\mathcal{R}_0 > 1$ , então há um surto da doença na população; se  $\mathcal{R}_0 < 1$ , o contágio da doença é controlado e não ocorre uma epidemia; por fim, para  $\mathcal{R}_0 = 1$ , a doença é endêmica, isto é, há um equilíbrio endêmico.

Combinando as equações (20) e (21) pode-se encontrar uma solução analítica para a classe de indivíduos infectados dada por

$$I(t) = -S(t) + \frac{\gamma N}{\beta} \ln(S(t)) + C. \quad (24)$$

Utilizando como condições iniciais  $I(0) = 1$  e  $S(0) = N - 1$ , chega-se à seguinte solução para a classe dos indivíduos infectados no tempo  $t$ ,

$$I(t) = -S(t) + N + \frac{\gamma N}{\beta} \ln\left(\frac{S(t)}{N-1}\right). \quad (25)$$

Sabe-se que com o decorrer do tempo, a quantidade de infectados irá diminuir e a epidemia encerrará. Assumindo essa condição, para  $t \rightarrow \infty$ ,  $I(\infty) = 0$  e, então

$$S(\infty) = N + \frac{\gamma N}{\beta} \ln\left(\frac{S(\infty)}{N-1}\right) \quad (26)$$

A equação (26) nos dá uma solução genérica para a classe de indivíduos suscetíveis a longo prazo.

O tamanho final da epidemia, utilizando o modelo SIR determinístico, é dado por  $R(\infty) = N - S(\infty)$  pois representa todos os indivíduos que se infectaram e foram recuperados (ou mortos).

### 3.2 Estocástico

No modelo estocástico, denominado SIR CTMC, os compartimentos  $S(t)$ ,  $I(t)$  e  $R(t)$  são considerados como variáveis aleatórias, sendo  $S(t) + I(t) + R(t) = N$ . Também será considerado que não há período latente, assim como no modelo determinístico.

Nesse modelo, apenas duas das variáveis aleatórias são independentes,  $S$  e  $I$ . A probabilidade conjunta dessas variáveis será utilizada para determinar a probabilidade de transição de um estado para outro (Allen, 2010).

Dessa forma, o modelo é definido através das probabilidades de transição indicadas na Tabela 1. A mudança dos estados  $S$  e  $I$  durante um intervalo de tempo pequeno  $\Delta t$  é definida como  $\Delta S = S(t + \Delta t) - S(t)$  e  $\Delta I = I(t + \Delta t) - I(t)$ .

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

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

Considerando que inicialmente não há indivíduos recuperados, define-se a distribuição inicial do sistema como  $(S(0), I(0)) = (s_0, i_0)$ , sendo  $i_0 > 0$  e  $s_0 + i_0 = N$ .

No início de uma epidemia, ou seja, para  $I(0) = i$  pequeno, o número básico de reprodução  $\mathcal{R}_0$  é usado para conhecer a probabilidade de ocorrência de um surto da doença na população. Essa probabilidade é obtida da seguinte forma (Whittle, 1955),

$$\text{probabilidade de surto} = \begin{cases} 0 & , \mathcal{R}_0 \leq 1 \\ 1 - \left( \frac{1}{\mathcal{R}_0} \right)^i & , \mathcal{R}_0 > 1 \end{cases} \quad (27)$$

Para analisar as transições possíveis, os estados são organizados em pares, onde o par  $(S, I)$  representa o estado possível que o modelo pode assumir, com  $0 \leq S + I \leq N$ . Essas transições são organizadas da seguinte forma,

$$(N, 0), (N-1, 0), (N-2, 0), \dots, (0, 0), (N-1, 1), (N-2, 1), \dots, (0, 1), \dots, (0, N) \quad (28)$$

De acordo com isso, existem  $(N+1)(N+2)/2$  pares de estados possíveis. Considerando os estados ordenados dessa maneira, temos o vetor das probabilidades de estado dado por  $p(t) = (p_{(N,0)}, p_{(N-1,0)}, \dots, p_{(0,N)})^T$ . A matriz geradora  $G$  é dada por  $dp/dt = Gp$ . Assim, tanto o vetor  $p$  como a matriz  $G$  dependem dessa ordenação dos estados. Se os estados são organizados de forma diferente,  $p$  e  $G$  também serão modificados.

A partir da Matriz Geradora, pode-se determinar a Matriz de Transição da Cadeia de Markov Incorporada do sistema,  $P_Y$ . Por conveniência, será utilizada a notação da expressão (28) e, assim, tem-se

$$P_Y = \begin{pmatrix} \tilde{p}_{11} & \tilde{p}_{12} & \dots & \tilde{p}_{1 \frac{(N+1)(N+2)}{2}} \\ \tilde{p}_{21} & \tilde{p}_{22} & \dots & \tilde{p}_{2 \frac{(N+1)(N+2)}{2}} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{p}_{\frac{(N+1)(N+2)}{2} 1} & \tilde{p}_{\frac{(N+1)(N+2)}{2} 2} & \dots & \tilde{p}_{\frac{(N+1)(N+2)}{2} \frac{(N+1)(N+2)}{2}} \end{pmatrix}$$

A matriz de transição  $P_Y$  e a matriz geradora  $G$  são matrizes quadradas de dimensão  $(N+1)(N+2)/2$ . Na matriz  $G$ , os elementos das primeiras  $N+1$  colunas são iguais a zero, pois não há transição entre quaisquer desses estados, já que não há nenhum indivíduo infectado. Isso significa que a matriz de transição  $P_Y = (\tilde{p}_{kl})$  satisfaz  $\tilde{p}_{ll} = 1$ , para  $l = 1, 2, \dots, N+1$ , que são os primeiros  $N+1$  estados que correspondem a  $(s, 0)$ , onde  $s = 0, 1, \dots, N+1$ . Esses  $N+1$  estados são os estados absorventes, aqueles dos quais, uma vez que são alcançados pelo sistema, não há mais possibilidade de transição para outro estado.

Considerando  $(S(t), I(t)) = (s, i)$ , podem-se determinar os elementos da matriz  $P_Y$  a partir das probabilidades de transição dos estados possíveis. No caso em que ocorre transição do estado  $(s, i)$  para o estado  $(s, i-1)$ , um indivíduo infectado é recuperado e sua probabilidade é dada por

$$p_s = \frac{\gamma i}{\gamma i + (\beta/N)si} = \frac{\gamma}{\gamma + (\beta/N)s}. \quad (29)$$

Quando ocorrer transição do estado  $(s, i)$  para o estado  $(s-1, i+1)$ , é considerada a infecção de um indivíduo suscetível e sua probabilidade é

$$1 - p_s = 1 - \frac{\gamma}{\gamma + (\beta/N)s} = \frac{(\beta N)s}{\gamma + (\beta/N)s}. \quad (30)$$

A matriz de transição  $P_Y$  da EMC é muito útil no cálculo do tamanho final da epidemia. Em geral, para qualquer população de tamanho  $N$ , começando com um indivíduo infectado ou  $p_{(N-1,1)}(0) = 1$ , o número máximo de transições até ocorrer a absorção, ou seja, quando não há mais indivíduos infectados, é  $2N-1$ .

Supondo que, inicialmente, o número de indivíduos infectados é 1 e não haja indivíduos recuperados, tem-se  $S(0) = N-1$ ,  $I(0) = 1$  e  $R(0) = 0$ . A probabilidade associada ao tamanho final da epidemia  $\{p_j^f\}$ ,  $j = 1, 2, \dots, N$  pode ser determinada calculando-se as probabilidades de absorção. Como os estados sem indivíduos infectados estão absorvendo e  $I(0) = 1$ , tem-se que

$$\lim_{t \rightarrow \infty} \sum_{s=0}^{N-1} p_{(s,0)}(t) = 1. \quad (31)$$

Se houver  $s$  indivíduos suscetíveis quando o número de indivíduos infectados atingir zero, o tamanho final da epidemia é  $N-s$ . Dessa forma, as probabilidades para a distribuição do tamanho final da epidemia satisfazem

$$\lim_{t \rightarrow \infty} p_{(s,0)}(t) = p_{N-s}^f \quad (32)$$

para  $s = 0, 1, \dots, N-1$ . Portanto, podem-se encontrar as probabilidades de absorção usando a matriz de transição  $P_Y$ . Em particular,

$$\lim_{t \rightarrow \infty} p(t) = p(2N-1) = (P_Y)^{2N-1} p(0). \quad (33)$$

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 em (17), correspondente a

$$\sum_{j \neq n} p_{jn}(\Delta t) = \left( \frac{\beta}{N} S(t)I(t) + \gamma I(t) \right) \Delta t + o(\Delta t) \quad (34)$$

sendo  $\alpha(n) = (\beta/N)S(t)I(t) + \gamma I(t)$ .

O tempo entre eventos, definido na Equação (19), pode ser então calculado utilizando essa expressão para  $\alpha(n)$ ,

$$T_E = -\frac{\ln U}{\frac{\beta}{N}S(t)I(t) + \gamma I(t)} \quad (35)$$

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

#### 4. RESULTADOS

Nessa seção, são apresentados alguns resultados do uso do modelo SIR CTMC antes descrito, na modelagem da atual situação epidêmica no estado de Rio de Janeiro produzida pelo vírus SARS-CoV-2. Foi usada a linguagem Python nas simulações realizadas. Como foi indicado na introdução, a obtenção da matriz de transição para uma população muito grande, como é o caso de Rio de Janeiro, é um grande desafio em termos computacionais. Esse é um resultado que ainda não temos e que deve ser resolvido futuramente. Contudo, apresentamos o resultado de algumas simulações que validam o modelo usado e comparamos com resultados prévios.

Para realizar as simulações foi considerada a população total do Estado do Rio de Janeiro,  $N = 17.366.189$ , de acordo com o IBGE (2019). Para o número básico de reprodução, foi utilizado o valor  $\mathcal{R}_0 = 2.83$ , encontrado em Volpatto et al. (2020). 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 (23), obtem-se a taxa de infecção  $\beta = 0.202$ .

Dada 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. Com os valores considerados dos parâmetros, foi observado que o pico da epidemia teria ocorrência no 132º dia no modelo determinístico. Nas Figuras 1 e 2 mostra-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, devido à natureza estocástica do modelo. 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 2. O pico da epidemia nessas simulações estocásticas ocorre nos dias 121 e 134, respectivamente.

De acordo com a definição do número básico de reprodução, sabe-se que haverá ocorrência da epidemia na população na condição de que  $R_0 > 1$ . Para o modelo estocástico pode-se determinar a probabilidade de ocorrência da doença na população, utilizando a equação (27). Essa probabilidade foi calculada com o  $R_0$  utilizado nas simulações, obtendo-se o valor

$$1 - \left(\frac{1}{2.83}\right)^1 \approx 0.647 \quad (36)$$

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. Essa situação é representada na Figura (3), onde pode-se observar uma trajetória do modelo estocástico que

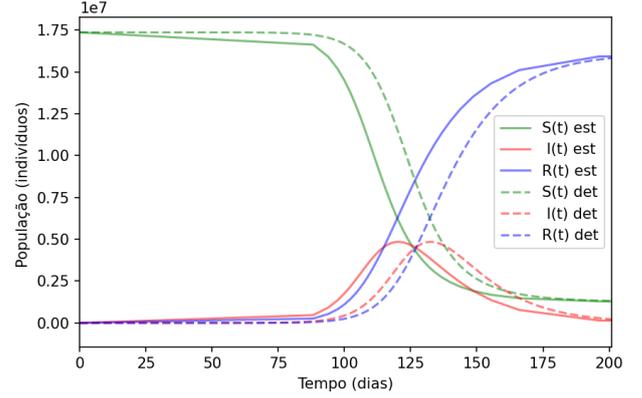


Figura 1. Comparação dos modelos estocástico (linha cheia) e determinístico (linha tracejada). Modelo Estocástico com pico da epidemia no 121º dia.

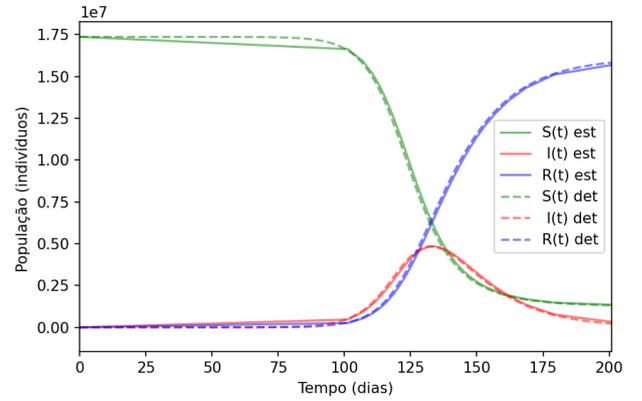


Figura 2. Comparação dos modelos estocástico (linha cheia) e determinístico (linha tracejada). Modelo Estocástico com pico da epidemia no 134º dia.

não gerou um comportamento epidêmico, em contraste com as curvas obtidas do modelo determinístico. Por isso,

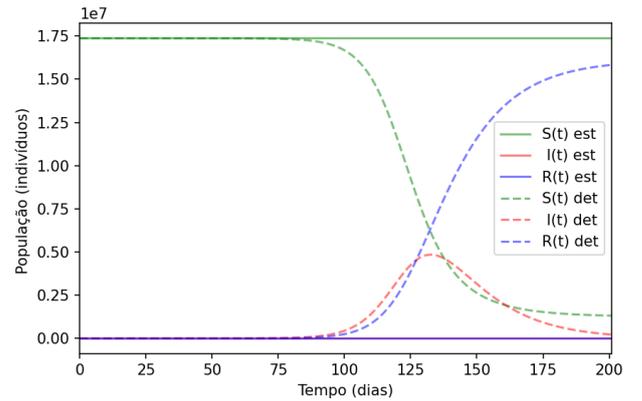


Figura 3. Comparação dos modelos estocástico e determinístico. É representada a evolução do modelo determinístico com linhas tracejadas, em contraste com a não ocorrência de surto da doença na população segundo o modelo estocástico.

para se ter resultados confiáveis, múltiplas corridas devem ser feitas. Nesse sentido, foram realizadas 100 simulações

do modelo estocástico com o objetivo de verificar os dias de pico que o modelo estocástico ofereceria. Das 100 simulações, houve também a ocorrência do caso em que não ocorreria a epidemia na população. A partir dos resultados numéricos, foi obtida uma probabilidade de 0,63 para a ocorrência da epidemia, ou seja, um valor muito próximo ao esperado, calculado em (36). Na Figura 4, 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.

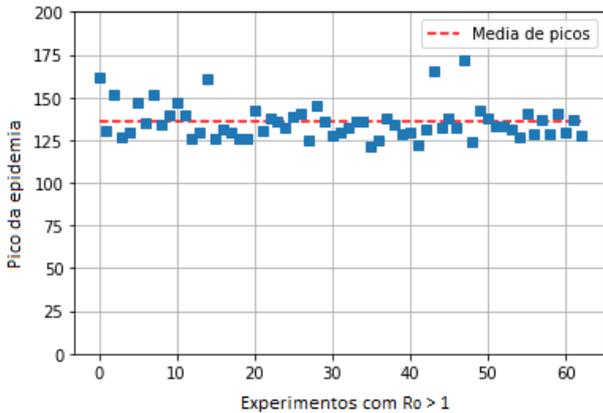


Figura 4. Média do dia de pico da epidemia com uma amostra de 100 simulações do modelo estocástico, onde ocorreu epidemia em, aproximadamente, 63% dos experimentos realizados.

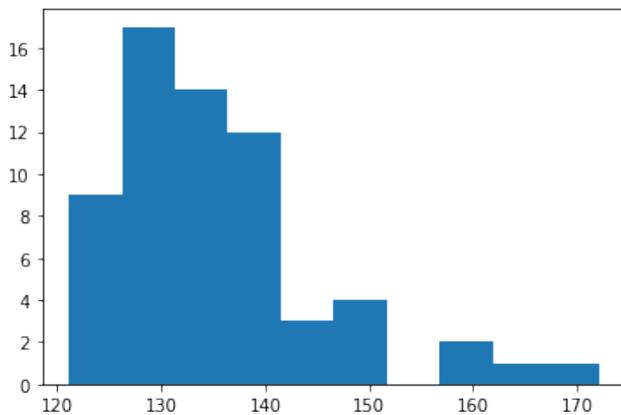


Figura 5. Histograma do dia de pico da epidemia para as 100 simulações do modelo estocástico. A maior parte dos resultados concentra-se ao redor do dia 135, valor médio encontrado nas simulações.

Como visto na seção anterior, o tamanho final da epidemia pode ser determinado quando  $I(t) = 0$ , através de  $R(t) = N - S(t)$ , no modelo determinístico. Com isso, 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 (SES-RJ, 2020), 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. O tamanho final no modelo estocástico é calculado através da matriz de transição que nos dá a distribuição final da epidemia, e como o  $N$  utilizado é muito alto, o cálculo possui um alto custo computacional que impossibilitou chegar a resultados. Na próxima seção discutimos sobre as alternativas que estamos desenvolvendo para contornar essa dificuldade.

## 5. CONCLUSÃO E TRABALHOS FUTUROS

Como pode ser observado no trabalho, o modelo compartimental estocástico permite a obtenção de informações no desenvolvimento de uma epidemia, que se correspondem com os resultados de modelos determinísticos, para populações grandes. O trabalho foi baseado numa estimativa inicial para o número básico de reprodução,  $\mathcal{R}_0$ , e desconsidera os efeitos de medidas (farmacológicas ou não) estabelecidas no controle da epidemia. Foi feito assim, objetivando a verificação do modelo. Uma vez verificado, uma análise mais atualizada será realizada, atendendo principalmente à evolução do número efetivo de reprodução,  $\mathcal{R}(t)$ .

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 3.2, 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 33). O custo computacional, devido ao número total da população utilizada neste trabalho, é extremamente alto e não contamos com recursos computacionais para chegar nesse resultado. Para contornar o problema tecnológico e conseguir realizar uma modelagem estocástica para populações grandes estamos desenvolvendo duas linhas de trabalho. Por uma parte, estamos trabalhando na otimização do código, na linguagem Python, e, por outra, estamos tentando encontrar formas analíticas recorrentes que permitam calcular os elementos não nulos da matriz final. Esperamos reportar resultados em breve.

## REFERÊNCIAS

- Allen, L. (2010). *An Introduction to Stochastic Processes with Applications to Biology*. Globo, Texas, USA.
- Allen, L.J. (2008). An introduction to stochastic epidemic models. In *Mathematical epidemiology*, 81–130. Springer.
- Comunian, A., Gaburro, R., and Giudici, M. (2020). Inversion of a SIR-based model: A critical analysis about the application to COVID-19 epidemic. *Physica D: Nonlinear Phenomena*, 413, 132674.
- Cooper, I., Mondal, A., and Antonopoulos, C. (2020). A SIR model assumption for the spread of COVID-19 in different communities. *Chaos, Solitons & Fractals*, 139, 110057.
- Crokidakis, N. (2020). 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*, 2050135.
- Daley, D. and Gani, J. (1999). *Epidemic modelling : an introduction*. Cambridge studies in mathematical biology, vol 14, Cambridge University Press.

- Greenwood, P. and Gordillo, L. (2009). Stochastic epidemic modeling. In G. Chowell, J. Hyman, L. Bettencourt, and C. Castillo-Chavez (eds.), *Mathematical and Statistical Estimation Approaches in Epidemiology*, 31–52. Springer, Dordrecht.
- Hethcote, H. (2000). The mathematics of infectious diseases. *SIAM Review*, 42, 599–653.
- IBGE (2019). População estimada: IBGE, Estimativas da população residente com data de referência de 1º de julho de 2019.
- Kermack, W. and McKendrick, A. (1927). A contribution to the mathematical theory of epidemics. *The Royal Society*, 115, 700–721.
- Manrique-Abril, F., Agudelo-Calderon, C., González-Chordá, V., Gutiérrez-Lesmes, O., Téllez-Piñerez, C., and Herrera-Amaya, G. (2020). Modelo SIR de la pandemia de Covid-19 en Colombia. *Rev. Salud Pública*, 22(1), e185977.
- Neves, A. and Guerrero, G. (2020). Predicting the evolution of the COVID-19 epidemic with the A-SIR model: Lombardy, Italy and São Paulo state, Brazil. *Physica D: Nonlinear Phenomena*, 413, 132693.
- SES-RJ (2020). Painel Coronavirus COVID-19, Secretaria de Saúde do Estado do Rio de Janeiro. <http://painel.saude.rj.gov.br/monitoramento/covid19.html>. [Online; acesso em 21-Setembro-2020].
- Volpatto, D., Resende, A., Anjos, L., Silva, J., Dias, C., Almeida, R., and Malta, S. (2020). Spreading of COVID-19 in Brazil: Impacts and uncertainties in social distancing strategies. *medRxiv*, 2020.05.30.20117283.
- Whittle, P. (1955). The outcome of a stochastic epidemic – A note on Bailey’s paper. *Biometrika*, 42(1/2), 116–122.