

Aplicação de um modelo SIR estocástico na transmissão do sarampo

Michelle Lau de Almeida¹
CCOMP/UERJ, Rio de Janeiro, RJ
Zochil González Arenas²
IME/UERJ, Rio de Janeiro, RJ

Resumo 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 determinísticos utilizados nessa modelagem é o Suscetível-Infetado-Recuperado (SIR), proposto por Kermack e McKendrick, em 1927. Neste trabalho, apresentamos o modelo SIR com uma abordagem estocástica, utilizando Cadeia de Markov de Tempo Contínuo (Continuous Time Markov Chain - CTMC). Foram realizadas simulações a fim de comparação entre os dois modelos utilizando dados de referência sobre a propagação do sarampo na Ilha Grande - Rio de Janeiro, em 1976. Resultados das propriedades do modelo são apresentados. Como vantagens exclusivas da modelagem estocástica tem-se o acesso a distribuições de probabilidade de certas propriedades, permitindo observar cenários aos quais não se tem acesso utilizando o modelo determinístico. Essas vantagens ficaram em evidência na análise da evolução do surto de sarampo.

Palavras-chave. Modelagem epidemiológica estocástica, Modelos compartimentais, Modelo SIR CTMC

1 Introdução

A modelagem de 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. Para analisar a propagação e tentar controlar tais doenças, é fundamental a utilização de ferramentas, tais como os modelos matemáticos. Além dos modelos, simulações computacionais também são úteis para construir e testar teorias e determinar a sensibilidade a mudanças nos valores dos parâmetros.

Os modelos compartimentais são muito utilizados no estudo e simulação de doenças infecciosas. Eles dividem a população estudada em classes (compartimentos), que representam os possíveis estados da doença de acordo com o modelo utilizado e podem ser determinísticos ou estocásticos. Os modelos compartimentais determinísticos representam o processo epidêmico através de um conjunto de equações diferenciais ordinárias e, dada uma condição inicial, descrevem a evolução temporal da doença. São utilizados para prever um único resultado de um determinado conjunto de circunstâncias. Nos modelos estocásticos, pode-se inserir uma certa aleatoriedade nos parâmetros e definir as distribuições de probabilidade dos resultados possíveis para a evolução da doença. Dessa forma, podem-se obter diversos cenários, ponderados por suas probabilidades, aos quais não temos acesso com a utilização dos modelos determinísticos.

¹michelle.lau@ime.uerj.br

²zochil@ime.uerj.br

Obter informação em relação à probabilidade de ocorrência de um surto e conhecer a duração esperada e a distribuição do tamanho final de uma epidemia são algumas possibilidades exclusivas dos modelos epidemiológicos estocásticos [1]. Dessa forma, é possível fazer uma modelagem mais realista ao considerar efeitos sobre os quais não se tem conhecimento suficiente ou que podem ser aleatórios.

Nesse trabalho, apresenta-se uma versão estocástica usando Cadeias de Markov de Tempo Contínuo, em inglês, Continuous Time Markov Chain (CTMC³) [2] do modelo (S)uscetível-(I)nfetado-(R)ecuperado, também denominado SIR [5]. O modelo será usado na análise da propagação de um surto de Sarampo em Ilha Grande, Rio de Janeiro, em 1976. A evolução dos indivíduos suscetíveis, infectados e recuperados será obtida através de simulação numérica em Python e comparada com os resultados do modelo determinístico.

A seguir, define-se o modelo SIR, nas versões determinística e estocástica. Após, apresentam-se os resultados das simulações realizadas, as conclusões do trabalho e principais referências.

2 Modelo Compartimental SIR

No modelo compartimental SIR [5] considera-se que o número total de indivíduos na população é dado por N . Vamos assumir que esse valor é constante, ou seja, não há nascimentos ou mortes durante o estudo da dinâmica evolutiva da doença. Também será considerado que não há a ocorrência de período latente, período no qual os infectados ainda não transmitem a doença.

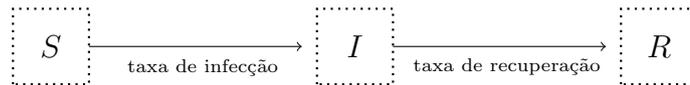


Figura 1: Diagrama do modelo SIR

A Figura 1 representa o diagrama do funcionamento do modelo. A classe de indivíduos suscetíveis se comunica com a classe de indivíduos infectados por meio de uma taxa de infecção, e a classe de indivíduos infectados se comunica com os recuperados por uma taxa de recuperação. Essas taxas serão definidas na próxima seção.

O conceito do número de reprodução básico, \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,

$$\mathcal{R}_0 = \frac{\text{taxa de infecção}}{\text{taxa de recuperação}}$$

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

2.1 Determinístico

Através das seguintes equações diferenciais, pode-se descrever o modelo SIR determinístico,

$$\frac{dS}{dt} = -\frac{\beta}{N}SI, \quad \frac{dI}{dt} = \frac{\beta}{N}SI - \gamma I, \quad \frac{dR}{dt} = \gamma I, \quad (1)$$

³Sigla definida pelo nome em inglês e, será utilizada para definir o nome do modelo estocástico.

com valores iniciais $S(0), I(0) > 0, R(0) \geq 0$, tais que $S(0) + I(0) + R(0) = N$. Define-se β como a taxa de contato ou taxa de infecção e γ como a taxa de recuperação de um indivíduo infectado.

Combinando as primeiras duas equações no sistema (1) e utilizando como condições iniciais $I(0) = 1, S(0) = N - 1$ e $R(0) = 0$, chega-se à seguinte solução para a classe dos indivíduos infectados no tempo t ,

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

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

$$S(\infty) = N + \frac{\gamma N}{\beta} \ln \left(\frac{S(\infty)}{N-1} \right). \tag{2}$$

A equação (2) 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).

2.2 Estocástico

A modelagem estocástica foi realizada utilizando Cadeia de Markov de Tempo Contínuo. No modelo, denominado SIR CTMC, consideramos os compartimentos $S(t), I(t)$ e $R(t)$ como variáveis aleatórias, de tal forma que $S(t) + I(t) + R(t) = N$.

Utilizando o conjunto de equações do modelo determinístico como base, temos que apenas duas das variáveis aleatórias do modelo estocástico são independentes. Considerando S e I como variáveis independentes, utilizaremos a sua probabilidade conjunta para determinar a probabilidade de transição de um estado para outro [2]. O modelo é definido, então, através das probabilidades de transição indicadas na Tabela 1. Considera-se a mudança/variação dos estados S e I , durante um intervalo de tempo pequeno Δt , 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
$(-1, 1)$	$\frac{\beta}{N} S(t)I(t)\Delta t + o(\Delta t)$
$(0, -1)$	$\gamma I(t)\Delta t + o(\Delta t)$
$(0, 0)$	$1 - \left[\frac{\beta}{N} S(t)I(t) + \gamma I(t) \right] \Delta t + o(\Delta t)$

Assumindo que inicialmente não há indivíduos recuperados, define-se a distribuição inicial do modelo como $(S(0), I(0)) = (s_0, i_0)$, sendo $i_0 > 0$ e $s_0 + i_0 = N$. Como foi indicado antes, através do modelo estocástico podem-se obter propriedades importantes para análise da propagação de uma doença. A primeira propriedade é a probabilidade de ocorrência de um surto da doença na população, calculada no início de uma epidemia, quando $I(0) = i$ é pequeno, que pode ser obtida da seguinte forma [7]:

$$\text{probabilidade de surto} = \begin{cases} 0 & , \mathcal{R}_0 \leq 1 \\ 1 - \left(\frac{1}{\mathcal{R}_0} \right)^i & , \mathcal{R}_0 > 1 \end{cases} \tag{3}$$

Neste modelo SIR, os estados são organizados em pares (S, I) , indicando o número de indivíduos no compartimento S e no compartimento I para cada estado, de modo que $0 \leq S + I \leq N$. Para

analisarmos as transições do sistema, ordenam-se todos os estados possíveis da seguinte forma: $(N, 0), (N - 1, 0), (N - 2, 0), \dots, (0, 0), (N - 1, 1), (N - 2, 1), \dots, (0, 1), \dots, (0, N)$. Assim, existem $(N + 1)(N + 2)/2$ estados possíveis e podem-se definir o vetor das probabilidades de cada estado do sistema por $p(t) = (p_{(N,0)}, p_{(N-1,0)}, \dots, p_{(0,N)})^T$ e determinar as taxas de transição entre estados.

Considerando $(S(t), I(t)) = (s, i)$, determinam-se os elementos da matriz de transição da Cadeia de Markov Incorporada P_Y a partir das probabilidades de transição entre estados. Quando ocorrer transição do estado (s, i) para o estado $(s, i - 1)$, um indivíduo infectado é recuperado e sua probabilidade é

$$\tilde{p}_s = \frac{\gamma i}{\gamma i + (\beta/N)si} = \frac{\gamma}{\gamma + (\beta/N)s}.$$

Por outro lado, no caso em que ocorre transição do estado (s, i) para o estado $(s - 1, i + 1)$, é considerada a infecção de um indivíduo suscetível e sua probabilidade é

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

A matriz de transição P_Y é muito útil no cálculo do tamanho final da epidemia. O tamanho final representa a força da epidemia, ou seja, quantos indivíduos foram infectados no total. 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, ou seja, $S(0) = N - 1, I(0) = 1$ e $R(0) = 0$, a probabilidade associada ao tamanho final da epidemia pode ser obtida calculando-se as probabilidades de absorção,

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

Se houver s indivíduos suscetíveis quando o número de indivíduos infectados atingir zero, o tamanho final da epidemia é $N - s$. Portanto, é possível encontrar as probabilidades de absorção usando a matriz de transição P_Y . Em particular, $\lim_{t \rightarrow \infty} p(t) = p(2N - 1) = (P_Y)^{2N-1}p(0)$.

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. No modelo, a duração de uma epidemia corresponde ao tempo que leva até chegar à absorção, ou seja, o tempo t até $I(t) = 0$. 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. Denotando por $\tau_{s,i}$ a duração da epidemia com condição inicial (s, i) , a duração esperada para a epidemia calcula-se a partir do sistema de equações

$$(\gamma i + (\beta/N)si) (\tilde{p}_s \tau_{(s,i-1)} - \tau_{(s,i)} + (1 - \tilde{p}_s) \tau_{(s-1,i+1)}) = -1. \tag{4}$$

Para realizarmos simulações estocásticas em uma CTMC, precisamos conhecer a distribuição do tempo entre sucessivos eventos, ou seja, o tempo que leva para ir de um estado a outro. Esse tempo é denominado tempo entre eventos, e é definido como uma variável aleatória. No modelo SIR CTMC, pode ser calculado da seguinte forma, $T_E = -N \ln U / [\beta S(t)I(t) + \gamma NI(t)]$, sendo U uma variável aleatória uniforme definida no intervalo $[0, 1]$.

3 Resultados

Os valores usados para os parâmetros nas simulações realizadas, foram tomados de 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 [4].

A população total de moradores era de 453 indivíduos. A doença teve início a partir de um indivíduo que viajou para esse local e transmitiu a doença para os moradores. Dessa forma, consideramos como condição inicial, $I(0) = 1$.

Na análise de um surto de sarampo, considera-se que indivíduos com mais de 19 anos possuem imunidade natural ou foram vacinados anteriormente, e que crianças menores de 1 ano possuem imunidade materna, razão pela qual esses grupos etários não apresentam contágio. Consequentemente, como população suscetível foi considerada a população residente com idade entre 1 ano e 19 anos, a qual totalizava 220 indivíduos. Assim, no instante inicial, foi considerado $S(0) = 220$, $I(0) = 1$ e, por fim, $R(0) = 233$. A epidemia do sarampo nessa localidade durou 60 dias e atingiu um total de 50 indivíduos, sem a ocorrência de mortes.

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

Apesar do sarampo ser uma doença altamente contagiosa, as condições de contágio e propagação dependem de vários fatores, como densidade populacional, faixa etária, políticas de vacinação e outras [6]. Assim, o valor do número de reprodução básico, \mathcal{R}_0 , deve ser cuidadosamente estimado. Na referência [4], que serviu de base para o trabalho, \mathcal{R}_0 não foi definido. Analisando as taxas de contágio por faixa etária, determinamos o valor $\mathcal{R}_0 = 6$. Em consequência, a taxa de infecção utilizada foi $\beta = \mathcal{R}_0\gamma \approx 0,6315$.

Na Figura 2, são mostrados alguns dos possíveis cenários obtidos na evolução do sarampo. Em cada sub-figura mostra-se a evolução de uma única simulação do modelo estocástico (linha cheia) em comparação com o resultado da modelagem determinística (linha tracejada). 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 2(b), pode-se observar um cenário no qual não se desenvolve uma epidemia, enquanto na Figura 2(c), os dois modelos seguem o mesmo perfil.

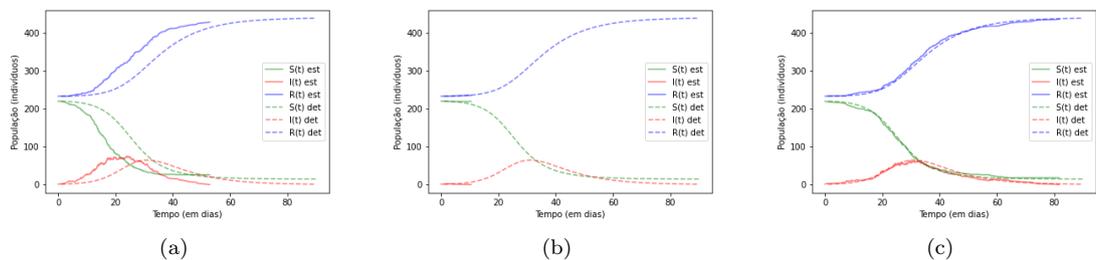


Figura 2: Diferentes cenários na modelagem estocástica.

Utilizando a equação (3), é possível determinar a probabilidade de ocorrência do surto da doença na população. Com o valor de \mathcal{R}_0 definido anteriormente, obtemos um resultado de 0,83 para essa propriedade, ou seja, a probabilidade da doença se propagar na população é de 83%. Na modelagem estocástica, todos os cenários possíveis têm uma determinada probabilidade de ocorrência, inclusive a de não ocorrer epidemia. Dessa forma, para se ter resultados válidos utilizando o modelo estocástico, devemos realizar um grande conjunto de simulações. Com o objetivo de fazer uma análise estatística sobre os resultados obtidos, realizamos 1000 simulações do modelo SIR CTMC, nas quais também houve a ocorrência do cenário no qual não há epidemia, como foi ilustrado na Figura 2(b). A partir dos resultados das 1000 simulações realizadas, as médias das grandezas de interesse foram calculadas. Dessa forma, encontrou-se que o pico da epidemia acon-

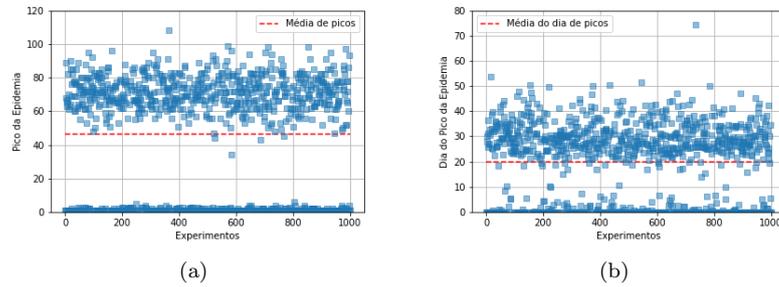


Figura 3: (a) Média do pico da epidemia no surto de sarampo. (b) Média do dia de pico da epidemia no surto de sarampo.

teceria, em média, com a infecção de 46 indivíduos ao redor do vigésimo dia, um valor razoável na descrição do surto de sarampo, considerando que foram contagiados 50 indivíduos no total. Em contraste, o modelo SIR determinístico forneceu uma estimação de 65 indivíduos contagiados no dia de pico, o 30º dia. Nas Figuras 3(a) e 3(b), mostram-se estes resultados.

Outra vantagem importante na modelagem estocástica é a possibilidade de gerar uma distribuição de probabilidade para o tamanho final da epidemia. Essa distribuição, no caso em estudo, pode ser observada na Figura 4(a). Analisando essa distribuição de probabilidade, pode-se observar um pico inicial, que representa a probabilidade de não ocorrer um surto epidêmico na população. Fora isso, a probabilidade de surto para todos os casos possíveis de tamanho final da epidemia encontra-se distribuída, com a maior probabilidade associada ao tamanho final de, aproximadamente, 221 contágios. Esses resultados são ilustrados nas Figuras 4(a) e 4(b).

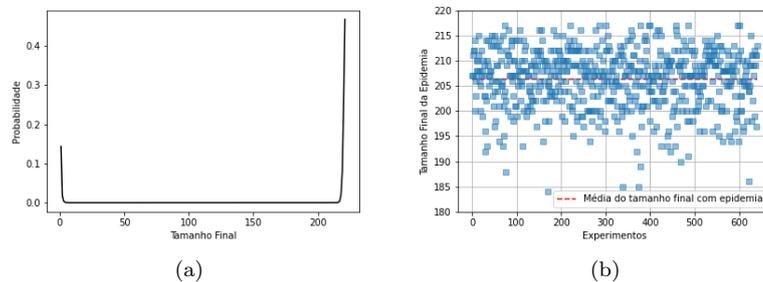


Figura 4: (a) Distribuição de Probabilidade do tamanho final da epidemia. (b) Média do tamanho final da epidemia.

Por fim, outra propriedade que o modelo SIR CTMC oferece é a duração esperada da epidemia, dada pela equação (4). A duração esperada iniciando pelo estado $(s, i) = (220, 1)$ foi de, aproximadamente, 63.33 dias. Na Figura 5 pode-se observar que a média da duração também é muito próxima do valor real da duração do surto de sarampo.

4 Conclusão

No presente trabalho, apresentamos uma versão estocástica do modelo compartimental SIR determinístico através do uso de Cadeia de Markov de Tempo Contínuo. O trabalho trouxe a abordagem mais simples, como uma primeira aproximação ao tema. Dessa forma, foi baseado em estimativas iniciais para os parâmetros e desconsidera os efeitos de medidas (farmacológicas

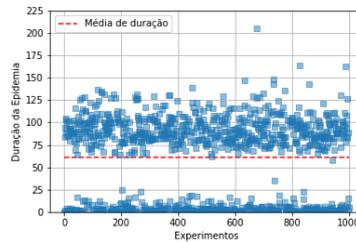


Figura 5: Valor médio da duração da epidemia igual a 61,35.

ou não) estabelecidas para controlar a propagação da doença. Consideramos para modelagem um surto de sarampo que ocorreu em Ilha Grande, no Rio de Janeiro, em 1976. A programação foi feita completamente em Python, permitindo a reusabilidade do código. Na aplicação, foram realizadas 1000 simulações do modelo estocástico, assim como a simulação do modelo SIR determinístico, para fins de validação e comparação dos resultados. Obtivemos resultados interessantes quando comparados com os valores reais da epidemia do sarampo, utilizando as propriedades específicas do modelo. Nesse caso, foi possível mostrar uma superioridade da modelagem estocástica em relação à modelagem determinística.

Agradecimentos

O presente trabalho foi realizado com apoio da FAPERJ e da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Referências

- [1] Allen, L. J. S., *An introduction to stochastic epidemic models*, Mathematical epidemiology, Springer, pages 81-130, 2008.
- [2] Allen, L. J. S., *An introduction to stochastic processes with applications to biology*, 2a. edição, CRC Press, Lubbock, 2010.
- [3] Almeida, M. L., Modelagem Epidemiológica com um Modelo SIR Estocástico utilizando Cadeia de Markov de Tempo Contínuo, Dissertação de Mestrado, UERJ, 2020.
- [4] Araújo Filho, N. A., Coura, J. R., *Surto epidêmico de sarampo na Ilha Grande, Rio de Janeiro, Brasil*, Revista da Sociedade Brasileira de Medicina Tropical, 13(1):147–155, SciELO Brasil, 1980. DOI: 10.1590/S0037-86821980000100018
- [5] Kermack, W. O., McKendrick, A. G., *A Contribution to the Mathematical Theory of Epidemics*, The Royal Society, 115:700–721, 1927.
- [6] Szusz, E. K., Garrison, L. P., Bauch, C. T., *A review of data needed to parameterize a dynamic model of measles in developing countries*, BMC Research Notes, 3(1):1–17, 2010.
- [7] Whittle, P., *The Outcome of a Stochastic Epidemic - A Note on Bailey's Paper*, In Biometrika, 42(1/2):116–120, 1955.