

Expressão fechada para o vetor de tamanho final da epidemia para um modelo SIR estocástico

Michelle Lau¹

PPG-CCOMP/UERJ, Rio de Janeiro, RJ

Zochil Gonzalez Arenas²

IME/UERJ, Rio de Janeiro, RJ

Resumo. Na modelagem epidemiológica, modelos estocásticos permitem estudar propriedades às quais não se tem acesso através de modelos deterministas. Tal é o caso da distribuição de probabilidade para o tamanho final de uma epidemia, uma propriedade muito importante na descrição da dinâmica evolutiva de uma doença. Uma abordagem comum para se obter esta distribuição, é dada pelo cálculo de potências de matrizes que podem ser excessivamente grandes a depender do tamanho da população. Com isso, a modelagem torna-se inviável por conta do custo computacional. Como uma forma alternativa, neste trabalho, deduzimos uma expressão para encontrar a forma fechada do vetor de tamanho final da epidemia no caso de um modelo SIR estocástico definido por meio de Cadeia de Markov de Tempo Contínuo, eliminando a necessidade do cálculo matricial.

Palavras-chave. Modelagem epidemiológica estocástica, Modelo SIR CTMC, Tamanho final da epidemia

1 Introdução

A modelagem epidemiológica auxilia na análise da propagação de doenças e na identificação de medidas que podem ser adotadas para contê-las. Com a utilização de modelos estocásticos epidemiológicos, pode-se inserir uma certa aleatoriedade nos parâmetros e encontrar a distribuição de probabilidade dos resultados possíveis para a evolução da doença. Dessa forma, temos acesso a diversos cenários para a dinâmica evolutiva, ponderados por suas probabilidades. A probabilidade de ocorrência de um surto, a duração esperada, assim como a distribuição de probabilidade do tamanho final de uma epidemia são algumas propriedades exclusivas dos modelos epidemiológicos estocásticos [1]. Assim, uma modelagem mais realista é realizada quando são considerados fenômenos sobre os quais não há conhecimento suficiente ou que podem ser aleatórios.

Neste trabalho, considera-se uma versão estocástica do modelo (S)uscetível-(I)nfetado-(R)ecuperado, conhecido como modelo SIR [7], usando Cadeias de Markov de Tempo Contínuo (CTMC³) [2, 5]. Os modelos epidemiológicos estocásticos formulados a partir de Cadeias de Markov são de grande utilidade pois incorporam o aspecto de aleatoriedade que é de vital importância para a modelagem precisa de muitos processos, ao mesmo tempo que são manipuláveis graças à propriedade de Markov. Ao fornecer uma compreensão quantitativa de vários fenômenos eles podem dar uma contribuição útil para a ciência [8].

Para muitos modelos, a simulação estocástica tem sido a única maneira de calcular as propriedades antes referidas, mas isso tem a desvantagem de exigir uma análise estatística sobre muitas

¹michelle.lau@ime.uerj.br

²zochil@ime.uerj.br

³Sigla definida pelo nome em inglês, *Continuous Time Markov Chain*, que será utilizada neste trabalho.

realizações. Com o aumento dos recursos computacionais, métodos numéricos de solução são hoje uma proposta atraente para esse tipo de modelo [3, 6]. No entanto, o elevado custo computacional continua a ser uma dificuldade para ampliar o seu uso. Dessa forma, o desenvolvimento ou aprimoramento de estratégias para melhorar a eficiência computacional, bem como a precisão, são um fator primordial nesta área de pesquisa.

Em epidemiologia, uma propriedade muito importante a ser estabelecida, tanto para inferência sobre uma epidemia, quanto para a tomada de decisões aplicadas à saúde pública, refere-se ao tamanho final da epidemia. O tamanho final é o número total de indivíduos que contraíram a doença ao longo do tempo de duração da epidemia. A estimação do tamanho final é um assunto de pesquisa atual e de amplo interesse, enfrentado através de diversas abordagens [3, 4, 6, 9]. Usando cadeias de Markov, pode-se encontrar a distribuição do tamanho final, detalhando a probabilidade de cada um dos resultados possíveis de uma epidemia.

Nesse trabalho, deduzimos a forma fechada para o vetor de tamanho final da epidemia para um modelo SIR CTMC. A expressão obtida permite acessar esta informação sem a necessidade do cálculo de potências de alta ordem de matrizes de grande porte, que é um método bem estabelecido para o cálculo desta propriedade para modelos descritos por cadeias de Markov [2, 5].

O trabalho é organizado da seguinte forma. Na próxima seção, apresenta-se brevemente o modelo SIR CTMC. A seguir, define-se a propriedade de tamanho final da epidemia e descreve-se o método matricial comumente usado para obter a sua distribuição de probabilidade. Na Seção 4 é apresentada a principal contribuição desse trabalho e, por fim, são oferecidas as conclusões e principais referências.

2 Modelo SIR CTMC

O modelo compartimental SIR (suscetíveis-infectados-recuperados) [7], é um modelo que caracteriza o estado de cada indivíduo da população ao longo da evolução de uma doença. Neste modelo, considera-se que não há nascimentos ou mortes no decorrer da dinâmica evolutiva da doença. Com isso, o número total de indivíduos na população é dado por um valor constante N . Considera-se também que não há período latente, no qual os infectados ainda não transmitem a doença. A dinâmica do modelo SIR ocorre com a interação do compartimento de indivíduos suscetíveis (S) com o compartimento dos indivíduos infectados (I) através de uma taxa de infecção, denotada por β , e de indivíduos infectados que tornam-se recuperados (R) a uma taxa γ , denominada taxa de recuperação.

Neste trabalho é usada uma versão estocástica do modelo SIR, como foi indicado acima, denominada SIR CTMC. Nesta modelagem, os compartimentos definidos por $S(t)$, $I(t)$ e $R(t)$ são variáveis aleatórias, tais que $S(t) + I(t) + R(t) = N$. Este modelo é bivariado, de modo que será utilizada a probabilidade conjunta das variáveis S e I para determinar a probabilidade de transição de um estado para outro [2]. Estas probabilidades de transição definem o modelo e são apresentadas na Tabela 1. Considera-se a 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
$(-1, 1)$	$\frac{\beta}{N} S(t)I(t)\Delta t + o(\Delta t)$
$(0, -1)$	$\gamma I(t)\Delta t + o(\Delta t)$
$(0, 0)$	$1 - \left[\frac{\beta}{N} S(t)I(t) + \gamma I(t) \right] \Delta t + o(\Delta t)$

A distribuição inicial do modelo é dada por $(S(0), I(0)) = (s_0, i_0)$, sendo $i_0 > 0$ e $s_0 + i_0 = N$, pois assume-se que inicialmente não há indivíduos recuperados.

3 Tamanho Final de uma Epidemia

Para calcular a distribuição de probabilidade do tamanho final no modelo SIR CTMC, inicialmente, determinamos a Matriz de Transição da Cadeia de Markov Incorporada, também denominada *Embedded Markov Chain* (EMC). Neste modelo, 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 $0 \leq S + I \leq N$.

No modelo SIR CTMC, as combinações possíveis entre os compartimentos, utilizando uma população de tamanho N , nos dá o valor total de pares de estados desse sistema, que corresponde a $(N + 1)(N + 2)/2$. Esses pares de estado são ordenados de uma forma conveniente para a construção da EMC, na seguinte forma,

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

Considerando os estados ordenados em (1), temos que o conjunto das probabilidades de estado é dado por:

$$p(t) = (p_{(N,0)}, p_{(N-1,0)}, \dots, p_{(0,N)})^T, \quad (2)$$

onde T denota a transposta.

Para organizarmos os estados na construção da matriz de transição da EMC, cada um deles será associado a um número inteiro. Dessa forma,

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

Utilizando a notação definida pela ordenação (1), é feita a construção da matriz de transição da EMC, denominada P_N , da seguinte forma:

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

A matriz de transição $P_N = (\tilde{p}_{kl})$ é uma matriz estocástica e quadrada de dimensão $(N + 1)(N + 2)/2$, que satisfaz $\tilde{p}_{ll} = 1$, para $l = 1, 2, \dots, N + 1$. Os primeiros $N + 1$ estados correspondem a $(s, 0)$, com $s = 0, 1, \dots, N + 1$. Esses $N + 1$ estados são os estados absorventes, aqueles desde os 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_N 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

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

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

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

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,

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

3.1 Forma Matricial

Para explicitarmos a utilização da forma matricial para calcular a distribuição para o tamanho final de uma epidemia, considerando o modelo SIR CTMC, utilizaremos um exemplo de modelo epidêmico para uma população com três indivíduos, $N = 3$.

Para $N = 3$, existem 10 possíveis pares de estados, listados a seguir, onde cada par é associado a um número inteiro.

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

A Figura 1 mostra o grafo que representa as possíveis transições entre os estados. A partir dessas transições, pode-se determinar a matriz de transição da EMC desse sistema.

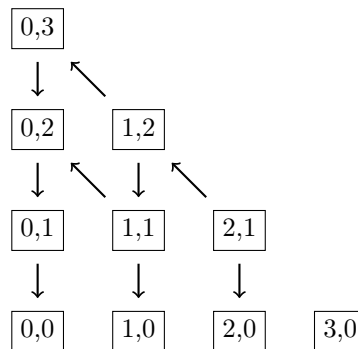


Figura 1: Possíveis transições entre os estados, considerando $N = 3$.

A matriz de transição da EMC é definida da seguinte forma,

$$P_3 = \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \end{matrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & \tilde{p}_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & \tilde{p}_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & \tilde{p}_0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \tilde{p}_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \tilde{p}_0 & 0 \\ 0 & 0 & 0 & 0 & 1 - \tilde{p}_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 - \tilde{p}_1 & 0 & 0 & 0 & \tilde{p}_0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 - \tilde{p}_1 & 0 & 0 \end{pmatrix}.$$

Os números à esquerda representam os pares de estados definidos na ordenação dada em (3). As transições são realizadas dos estados da coluna para os estados da linha. Para encontrar as probabilidades de absorção que determinam a distribuição para o tamanho final, utiliza-se a Eq. (3). Então, calculam-se as potências da matriz P_3 até chegar à potência $(P_N)^{2N-1}$. Neste caso, calculamos $(P_3)^5$, obtendo,

$$(P_3)^5 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & \tilde{p}_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & \tilde{p}_1^2(1 - \tilde{p}_2) & \tilde{p}_1 & 0 & 0 & \tilde{p}_1^2 & 0 & 0 \\ 0 & 0 & 0 & 1 & \tilde{p}_0^2(1 - \tilde{p}_2)(1 - \tilde{p}_1)(\tilde{p}_0 + \tilde{p}_1) & \tilde{p}_0^2(1 - \tilde{p}_1) & \tilde{p}_0 & (\tilde{p}_0^3 + \tilde{p}_0^2\tilde{p}_1)(1 - \tilde{p}_1) & \tilde{p}_0^2 & \tilde{p}_0^3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Quando tratamos uma situação onde a epidemia começa com apenas um indivíduo, temos que $p_{(N-1,1)}(0) = 1$ pois $(N - 1, 1)$ é aquele par de estados, de acordo com a ordenação estabelecida em (1), onde há o infectado inicial. Portanto, o vetor $p(0)$ é um vetor nulo, de tamanho $(N + 1)(N + 2)/2$, com exceção da posição $N + 2$ (posição do estado $(N - 1, 1)$) onde o valor é igual a 1. Nesse caso, tem-se

$$p(0) = (0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0)^T.$$

De acordo com a Eq.(3), faz-se o produto da matriz $(P_3)^5$ pelo vetor $p(0)$ e chega-se ao seguinte resultado,

$$(P_3)^5 \cdot p(0) = \begin{pmatrix} 0 \\ \tilde{p}_2 \\ \tilde{p}_1^2(1 - \tilde{p}_2) \\ \tilde{p}_0^2(1 - \tilde{p}_1)(1 - \tilde{p}_2)(\tilde{p}_0 + \tilde{p}_1) \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Para concluir esta abordagem matricial, definiremos um vetor contendo apenas os elementos não nulos. Denominamos esse vetor por $(vetor_final)_N$, obtendo

$$(vetor_final)_3 = \begin{pmatrix} \tilde{p}_2 \\ \tilde{p}_1^2(1 - \tilde{p}_2) \\ \tilde{p}_0^2(1 - \tilde{p}_1)(1 - \tilde{p}_2)(\tilde{p}_0 + \tilde{p}_1) \end{pmatrix}. \quad (4)$$

A expressão (4) nos dá o vetor referente à distribuição de probabilidade do tamanho final da epidemia. Através do cálculo das potências da matriz P_Y tem-se acesso as dinâmicas temporais do sistema. Entretanto, pode-se notar que para tamanhos de população muito grandes, a matriz de transição da EMC vai ser, correspondentemente, de grande porte. Além disso, será necessário calcular potências de uma ordem muito alta, gerando um elevado custo computacional. Isso inviabiliza o uso desta modelagem no estudo epidemiológico para populações muito grandes, ao tempo que torna como um objeto de estudo a procura de métodos de cálculo eficientes. As limitações do tamanho do sistema para essas abordagens estão relacionadas aos recursos computacionais disponíveis para armazenar e processar grandes matrizes esparsas [6].

4 Forma fechada do vetor de tamanho final da epidemia

Como foi indicado na seção anterior, a depender do tamanho N da população, o cálculo realizado na forma matricial pode ser muito custoso do ponto de vista computacional. Nesta seção, deduzimos uma forma fechada, analítica, para encontrar de forma direta os elementos do vetor final referente à distribuição do tamanho final da epidemia.

Como visto na seção 3.1, o vetor final, considerando $N = 3$, é dado pela Eq.(4). Ao considerarmos uma população de tamanho $N = 4$, após realizarmos as etapas da forma matricial para encontrar o vetor final, obtemos o seguinte resultado,

$$(vetor_final)_4 = \begin{pmatrix} \tilde{p}_3 \\ \tilde{p}_2^2(1 - \tilde{p}_3) \\ \tilde{p}_1^2(1 - \tilde{p}_2)(1 - \tilde{p}_3)(\tilde{p}_1 + \tilde{p}_2) \\ \tilde{p}_0^2(1 - \tilde{p}_1)(1 - \tilde{p}_2)(1 - \tilde{p}_3)(\tilde{p}_0(\tilde{p}_0 + \tilde{p}_1 + \tilde{p}_2) + \tilde{p}_1(\tilde{p}_1 + \tilde{p}_2)) \end{pmatrix}.$$

A partir da análise do vetor final para diversos tamanhos de população, encontramos um padrão que permite identificar cada elemento do vetor final, de acordo com o valor de N .

Seja k o índice que identifica a posição de um dado elemento no vetor final. O primeiro elemento do vetor final coincide sempre com a probabilidade

$$\tilde{p}_{N-1} \quad (5)$$

que corresponde a \tilde{p}_{N-k} para $k = 1$.

O segundo elemento do vetor final, ou seja, para $k = 2$, é dado por

$$\tilde{p}_{N-2}^2(1 - \tilde{p}_{N-1}). \quad (6)$$

Este elemento é obtido ao substituir o valor $k = 2$ na forma geral $\tilde{p}_{N-k}^2(1 - \tilde{p}_{N-k+1})$.

O valor do k -ésimo elemento, para $2 < k \leq N$, é obtido através da expressão

$$\tilde{p}_{N-k}^2 \left(\sum_{i_1=N-k}^{N-(k-1)} \tilde{p}_{i_1} \left(\sum_{i_2=i_1}^{N-(k-2)} \tilde{p}_{i_2} \left(\cdots \left(\sum_{i_{N-2}=i_{N-3}}^{N-2} \tilde{p}_{i_{N-2}} \right) \right) \right) \right) \left(\prod_{i=(N-k)+1}^{N-1} (1 - \tilde{p}_i) \right). \quad (7)$$

Com as expressões (5), (6) e (7) temos acesso ao vetor final, que fornece a distribuição de probabilidade do tamanho final da epidemia, sem a necessidade de calcularmos a potência $2N - 1$ da matriz P_Y , como feito na Seção 3.1. A expressão encontrada para este vetor em forma fechada é um resultado inédito.

5 Considerações Finais

No presente trabalho, foi desenvolvida uma nova forma de se obter o vetor em forma fechada que oferece a informação completa da distribuição de probabilidade do tamanho final de uma epidemia para um modelo SIR CTMC. Esta abordagem permite evadir o cálculo matricial descrito na Seção 3, que pode ter um custo excessivamente elevado para populações muito grandes, tornando impraticável o modelo apresentado.

A partir da expressão encontrada para o vetor final, através dos seus elementos (5), (6) e (7), o próximo passo será estabelecer um algoritmo eficiente para o cálculo destes elementos. Nesse sentido, estão sendo desenvolvidas duas estratégias, uma com foco em incorporar elementos de recorrência e outra, com o objetivo de paralelizar o cálculo.

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] L. J. S. Allen. “An introduction to stochastic epidemic models”. Em: **Mathematical epidemiology**. Springer, 2008. Cap. 3, pp. 81–130.
- [2] L. J. S. Allen. **An introduction to stochastic processes with applications to biology**. 2a. ed. Lubbock: CRC Press, 2010. ISBN: 9781439894682.
- [3] Andrew J Black e Joshua V Ross. “Computation of epidemic final size distributions”. Em: **Journal of theoretical biology** 367 (2015), pp. 159–165.
- [4] Fred Brauer. “A final size relation for epidemic models of vector-transmitted diseases”. Em: **Infectious Disease Modelling** 2.1 (2017), pp. 12–20.
- [5] Daryl J Daley e Joe Gani. **Epidemic modelling: an introduction**. 15. Cambridge University Press, 2001.
- [6] Thomas House, Joshua V Ross e David Sirl. “How big is an outbreak likely to be? Methods for epidemic final-size calculation”. Em: **Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences** 469.2150 (2012), p. 22.
- [7] William Ogilvy Kermack e A. G. McKendrick. “A contribution to the mathematical theory of epidemics”. Em: **Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences** 115.772 (1927), pp. 700–721.
- [8] James Robert Norris. **Markov chains**. Cambridge Series in Statistical and Probabilistic Mathematics, 2. Cambridge University Press, 1998. ISBN: 9780511810633.
- [9] Yu Tsubouchi, Yasuhiro Takeuchi e Shinji Nakaoka. “Calculation of final size for vector-transmitted epidemic model”. Em: **Mathematical Biosciences and Engineering** 16.4 (2019), pp. 2219–2232.