

**Métodos para estimação de
parâmetros em um modelo
SIR**

Thiago Bonfim Quintas

TESE APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TÍTULO
DE
BACHAREL EM MATEMÁTICA APLICADA COMPUTACIONAL

Orientadora: Dra Larissa Marques Sartori
Coorientador: Prof. Dr. Sérgio Muniz Oliva Filho

São Paulo, Fevereiro de 2019

Resumo

A dengue é atualmente a doença viral humana mais comum, com cerca de 40% da população mundial em risco de contrair a infecção, é transmitida pelo mosquito do gênero *Aedes*. O presente trabalho tem como objetivo estudar o modelo matemático SIR aplicado à dengue, que consiste em dividir a população em suscetível, infectada e recuperada. O foco principal é o estudo dos parâmetros do modelo para reproduzir os dados de incidência da dengue. Foram estudados três métodos de estimação de parâmetros: Método de mínimos quadrados - Uma estimativa pura e direta, método de varredura de Monte Carlo e método de Monte Carlo usando estatística de probabilidade log-negativa de Poisson. Os resultados mostraram que o método de varredura de parâmetros de Monte Carlo resultou em uma melhor aproximação, uma vez que retornou o menor erro quadrático.

Palavras chave: Dengue, ajuste de parâmetros, SIR.

Abstract

Dengue is currently the most common human viral disease, with about 40% of the world population at risk of contracting the infection, it is transmitted by the mosquito of the genus *Aedes*. The present work aims to study the SIR mathematical model applied to dengue, which consists of dividing the population into susceptible, infected and recovered. The main focus is on the study of the parameters of the model so that it reproduces the incidence data of dengue. We studied three methods of parameter estimation: Least-squares method - A pure and direct estimation, Monte Carlo parameter scanning method and Monte Carlo method using Poisson log-negative probability statistics. The results showed that the Monte Carlo parameter scanning method resulted in a better approximation, since it returned the smallest quadratic error.

Key words: Dengue, parameter adjustment, SIR.

Sumário

Introdução	1
1. Dengue.....	2
2. Modelo compartimental SIR	4
2.1 Força de infecção λ	6
2.2 Ponto de equilíbrio S^* , I^* , R^*	7
2.3 Razão de reprodutividade basal R_0	8
3. Métodos.....	10
3.1 Método dos mínimos Quadrados aplicados ao modelo SIR – Uma aproximação pura e direta.	10
3.2 Método de Monte Carlo.....	12
3.2.1 Método de varredura de parâmetros de Monte Carlo.	13
3.2.2 Método de Monte Carlo via log-probabilidade negativa de Poisson.....	14
4. Resultados e métodos	16
4.1 Estudo 1.....	16
4.2 Estudo 2.....	20
4.3 Estudo 3.....	21
5. Conclusão	22
6. Referências	24

Introdução

Estudos epidemiológicos por meio de modelos matemáticos possibilitam entender o desenvolvimento de epidemias e buscar medidas eficientes para a prevenção ou mesmo erradicação de determinada doença. Dessa maneira se faz necessário o uso de modelos matemáticos coerentes para que possamos entender e explicar resultados, que por sua vez terão como finalidade fazer previsões e auxiliar na tomada de decisões (KEELING E ROHANI, 2008; MARTCHEVA, 2015).

Modelos muito utilizados para esses estudos são os chamados “modelos compartimentais”, que como o próprio nome sugere, dividem a população por meio de compartimentos. Neste trabalho, focaremos no modelo SIR, que possui três compartimentos: Suscetíveis, Infectados e Recuperados, estruturado por Kermack W.O e Mckendrik A. G. (KEELING E ROHANI, 2008).

O modelo SIR possui parâmetros em suas equações que devem ser estimados, de maneira que sejam suficientemente consistentes para retornar valores próximos aos dados reais de determinada doença, que no nosso estudo será a Dengue.

Para estimar os parâmetros do modelo SIR, faremos o uso de três métodos: **Método dos mínimos quadrados**, **Método de varredura de parâmetros de Monte Carlo** e **Método de Monte Carlo usando a estatística log-probabilidade negativa de Poisson**.

Capítulo 1

Dengue

A dengue é uma doença viral, transmitida pela fêmea do mosquito *Aedes aegypti*, reconhecido pela primeira vez no Egito, daí o seu nome, chegou ao Brasil colonial por meio de navios negreiros originários da África, estudos indicam que os mosquitos se reproduziam nos depósitos de água dos barcos. A primeira epidemia registrada se deu em Salvador no ano de 1872, causando 2.000 mortes ([Portal Educação, 2018](#)).

Antonio Gonçalves Peryassú, pesquisador do então Instituto Soroterápico Federal, que ganharia o nome de Instituto Oswaldo Cruz (IOC), em 1908, fez descobertas sobre o ciclo de vida, os hábitos e a biologia do *Aedes aegypti*. Os estudos de Peryassú foram fundamentais para a erradicação do mosquito em território nacional nas décadas seguintes e ainda hoje norteiam as pesquisas sobre o controle do vetor. Entre as descobertas do pesquisador estão também a relação do mosquito com a temperatura e a densidade populacional, destacando a influência da temperatura no desenvolvimento e reprodução, que abaixo de 20 graus se reduz drasticamente, levando a uma redução dos casos ([Portal Educação, 2018](#)).

Peryassu, através de suas descobertas, foi fundamentalmente importante para a campanha de eliminação do mosquito, coordenada por Oswaldo Cruz, médico sanitário. Segundo relatos, a dengue teria sido controlada em 1920 no Rio de Janeiro, e considerada erradicada no Brasil na década de 1950 ([Portal Educação, 2018](#)).

O Brasil é considerado um país tropical em sua totalidade devido ao seu clima quente e úmido que proporciona um ambiente receptivo e altamente favorável para a proliferação do vetor de dengue ([FARES et al. 2015](#)). As mudanças climáticas, urbanização, ineficiência no combate ao vetor gerada pela falta de recursos para a saúde pública e pesquisas, são alguns dos fatores responsáveis pela forte emergência da dengue e pelo ressurgimento desta epidemia ([MACIEL et al 2008](#)).

Entre 1981 e 2006, foram registrados no Brasil 4.243.049 casos de dengue, com 338 mortes. A maioria dos casos ocorreram nas regiões Nordeste e Sudeste. No ano de 2012, houve uma queda no número de casos em relação a 2011, entretanto, se comparado ao ano de 2013, há um aumento de 190%. Em 2013 foram registrados 1.468.873 casos, sendo 6.969 casos graves e 545 vítimas fatais (FARES et al. 2015).

Atualmente a dengue, uma doença viral transmitida por um mosquito, pode ser definida como a de maior número de casos no mundo. Destacando-se pelo aumento gradual no número de casos, no Brasil, passou a causar grande preocupação aos órgãos públicos, por ser uma doença com grande dificuldade em dois principais pontos: atendimento adequado aos indivíduos infectados e dificuldade no controle de epidemias. (BARRETO et al. 2008).

Os dados utilizados no presente trabalho são os casos de incidência de dengue na cidade do Rio de Janeiro no ano de 2012. A figura abaixo contém o número de casos observados semanalmente (MINISTÉRIO DA SAÚDE/SVS, 2018) :

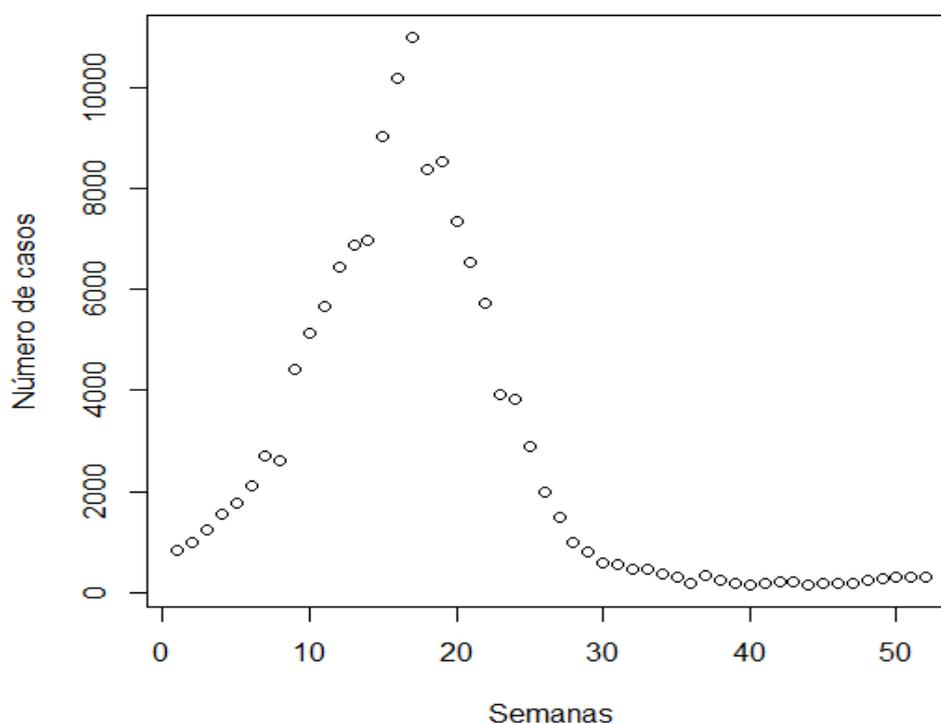


Figura 1.1 - Gráfico do número de casos de dengue por semana na cidade do Rio de Janeiro no ano de 2012.

Capítulo 2

Modelo compartimental SIR

O modelo epidemiológico SIR foi proposto por Kermack e McKendrick, 1927, é atualmente um dos modelos mais utilizados para representação de doenças infecciosas, utilizando estratégia de compartimentos. O modelo divide a população em três compartimentos ou classes, cada compartimento indica o estado atual da doença nos indivíduos da população, dessa maneira, cada indivíduo pode pertencer apenas a um compartimento por unidade de tempo. Apesar da infecção da dengue se dar através do mosquito, neste trabalho vamos supor transmissão direta entre humanos como forma de simplificação do modelo devido a falta de informações sobre os vetores transmissores.

Assim, a fim de gerar o modelo com as interações entre os compartimentos, assumimos que :

1. O tamanho N da população é constante
2. Os indivíduos nascem e morrem a uma taxa μ constante, esta hipótese é considerada para manter o tamanho total da população fixo.
3. A interação entre as populações dos compartimentos se dá de forma homogênea.
4. Após se recuperar o indivíduo se torna imune.
5. O contágio se dá através do contato efetivo entre humanos suscetíveis e infectados.

Dessa forma, a divisão é feita da seguinte maneira:

S = Suscetíveis: Indivíduos que podem contrair a doença

I = Infectados: Já possuem a doença e ainda podem propagá-la, ou seja, podem fazer com que os suscetíveis se tornem infectados.

R = Recuperados: São aqueles que já contraíram a doença, mas se recuperaram e tornaram-se imunes por toda a vida.
De forma que:

$S(t)$: indivíduos suscetíveis no tempo t , com as seguintes características:

- Recebe μN : nascimento de novos indivíduos, a uma taxa μ .
- Perde $\mu S(t)$: morte natural de indivíduos suscetíveis no tempo t a uma taxa μ .
- Perde $\lambda(t)S(t)$: indivíduos suscetíveis se tornam infectados a uma taxa $\lambda(t)$.

$I(t)$: indivíduos infectados no tempo t , com as seguintes características :

- Recebe $\lambda(t)S(t)$: indivíduos suscetíveis se tornam infectados a uma taxa $\lambda(t)$.
- Perde $\mu I(t)$: morte natural de indivíduos infectados no tempo t , a uma taxa μ .
- Perde $\gamma I(t)$: indivíduos infectados que se recuperam a uma taxa γ .

$R(t)$: indivíduos recuperados no tempo t , com as seguintes características :

- Recebe $\gamma I(t)$: indivíduos infectados que se recuperam a uma taxa γ .
- Perde $\mu R(t)$: morte natural de indivíduos recuperados no tempo t , a uma taxa μ .

Podemos representar o fluxo de indivíduos que vão da classe dos Suscetíveis para a classe dos Infectados e diretamente para a classe dos Recuperados, segundo o esquema da figura abaixo:

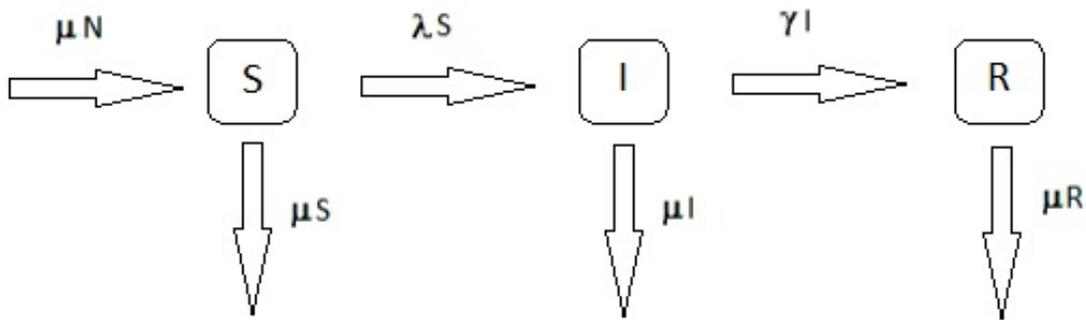


Figura 2.1 – Esquemática do modelo SIR com as relações entre os compartimentos e suas respectivas entradas e saídas.

O modelo compartimental SIR, com as observações acima é descrito pelo sistema de equações diferenciais ordinárias, dadas por:

$$\begin{cases} \frac{d}{dt}S(t) = \mu N - \lambda(t)S(t) - \mu S(t) \\ \frac{d}{dt}I(t) = \lambda(t)S(t) - \mu I(t) - \gamma I(t) \\ \frac{d}{dt}R(t) = \gamma I(t) - \mu R(t) \end{cases} \quad (2.1)$$

em que $S(t) + I(t) + R(t) = N$, com valores iniciais: $S(0) \geq 0, I(0) \geq 0$ e $R(0) \geq 0 \in \mathbb{R}_+$ (KEELING e ROHANI, 2008).

Para uma determinada doença com população específica, podemos trabalhar as funções do modelo SIR com o objetivo de prever possíveis surtos de modo a controlá-los (WEISSTEIN, 2010). A maioria dos modelos tem como parâmetros relevantes a força de infecção e a razão de reprodutividade Basal. A força de infecção representa o risco de um indivíduo suscetível tornar-se infectado; ela determina o tamanho da propagação da doença e o esforço necessário para combatê-la.

2.1 Força de infecção λ

Seja N o número total de indivíduos de uma população, e β a taxa de contatos infectantes, se $I(t)/N(t)$ é a fração de infectados, concluímos que

$\beta S(t) I(t)/N(t)$ é o número de contatos efetivos. Portanto λ é dado por:

$$\lambda = \beta \frac{I(t)}{N(t)} \quad (2.2)$$

A partir das definições acima, podemos reescrever o sistema de equações SIR, substituindo λ na equação:

$$\begin{cases} \frac{d}{dt} S(t) = \mu N - \beta \frac{I(t)}{N(t)} S(t) - \mu S(t) \\ \frac{d}{dt} I(t) = \beta I(t) \frac{S(t)}{N(t)} - \mu I(t) - \gamma I(t) \\ \frac{d}{dt} R(t) = \gamma I(t) - \mu R(t) \end{cases} \quad (2.3)$$

Com $S(t) + I(t) + R(t) = N$.

2.2 Ponto de equilíbrio S^*, I^*, R^*

Estudaremos agora os pontos de equilíbrio para o modelo SIR.

Com o objetivo de encontrar os pontos de equilíbrio do sistema (2.3), dividimos todas as equações por N de modo a simplificar para:

$$\begin{cases} \frac{d}{dt} s(t) = \mu - \beta i(t)s(t) - \mu s(t) \\ \frac{d}{dt} i(t) = \beta i(t)s(t) - \mu i(t) - \gamma i(t) \end{cases} \quad (2.4)$$

Assim temos $s(t) + i(t) + r(t) = 1$, sendo $s(t)$ e $i(t)$ a fração de suscetíveis e infectados na população, com $r(t) = 1 - s(t) - i(t)$, a fração de recuperados, obtida a partir de $s(t)$ e $i(t)$.

Igualando o lado direito do sistema a zero e resolvendo, encontramos os pontos de equilíbrio.

Analisando a segunda equação do sistema 2.4 temos $\beta i(t)s(t) - \mu i(t) - \gamma i(t) = 0$

Colocando $i(t)$ em evidência temos que:

$$i(t)(\beta s(t) - \mu - \gamma) = 0 \quad (2.5)$$

De forma que se $i(t) = 0$, então temos o ponto de equilíbrio livre de doença:

- $P_1(s^*, i^*) = (1, 0)$; população sem nenhum infectado pela doença

No caso em que $i(t) \neq 0$, então obtemos:

$$s(t) = \frac{(\mu + \gamma)}{\beta} \quad (2.6)$$

substituindo $s(t)$ na primeira equação do sistema 2.4 obtemos o valor de $i(t) = \frac{\mu}{\beta} - \frac{\mu}{\mu + \gamma}$, assim o ponto de equilíbrio endêmico é dado por:

- $P_2(s^*, i^*) = \left(\frac{\mu + \gamma}{\beta}, \frac{\mu}{\beta} - \frac{\mu}{\mu + \gamma} \right)$; que corresponde a uma situação de coexistência de suscetíveis e infectados.

2.3 Razão de reprodutividade basal R_0

Um modelo epidemiológico tem como objetivo determinar se uma doença vai se espalhar ou não pela população. Para isso usamos informações sobre a reprodutividade basal R_0 (PEIXOTO e BARROS, 2004).

A razão de reprodutividade basal R_0 representa o número médio de infecções secundárias quando um indivíduo infectado é introduzido em uma população de hospedeiros completamente suscetíveis.

Para:

- $R_0 = 1$, temos uma situação de equilíbrio, isto é, o número de infectados em uma população permanecerá igual.
- $R_0 > 1$, o número de casos aumentará.
- $R_0 < 1$, o número de casos diminuirá até alcançar o índice de zero infectados.

A reprodutividade efetiva R_{0e} é o número médio de casos secundários em uma população não totalmente suscetível, produzidos por um indivíduo infectado.

Seja $s^* = S(t)/N$, a fração dos indivíduos suscetíveis na população, temos que:

$$R_{0e} = R_0 s^*(t) \quad (2.7)$$

Em situação de equilíbrio $R_{0e} = 1$, portanto:

$$R_0 = \frac{1}{s^*(t)} \quad (2.8)$$

Usando $s^*(t)$ obtido na equação (2.6) concluímos que para o modelo **SIR**:

$$R_0 = \frac{\beta}{\mu + \gamma} \quad (2.9)$$

No caso do modelo SIR sem considerar demografia, isto é, sem considerarmos nascimentos e mortes, teremos que:

$$R_0 = \frac{\beta}{\gamma} \quad (2.10)$$

Capítulo 3

Métodos

Neste capítulo estudaremos três métodos que buscarão encontrar os melhores valores para os parâmetros estudados, a fim de alcançar o objetivo de encontrar boas aproximações para os casos de dengue observados; **Método dos mínimos quadrados, Método de varredura de parâmetros de Monte Carlo e Método de Monte Carlo usando log-probabilidade negativa de Poisson.** .

3.1 Método dos mínimos Quadrados aplicados ao modelo SIR – Uma aproximação pura e direta.

É grande o número de pessoas que ao lerem “mínimos quadrados” já pensam em resolver o problema por meio de regressão linear, até mesmo por ser uma função pronta em todos os pacotes de programação estatística, mas nesse caso não podemos usá-la por alguns fatores, são eles (POLIMATHEYA):

- Regressão linear, como o próprio nome já indica, exige um modelo linear, o que não ocorre na maioria das vezes nos modelos epidemiológicos.
- As equações que descrevem os modelos compartimentais, como é o caso do modelo SIR, em sua maioria, são resolvidas numericamente, por não possuir solução analítica. Para a resolução numérica, são usados algoritmos como por exemplo o método de Euler ou o método Runge-Kutta.

- Os dados medem a incidência semanal da doença, que por diversas vezes são apenas observações parciais. Com base no modelo SIR, buscamos estimar os parâmetros.

O método dos mínimos quadrados é uma técnica que se aplica para determinar os parâmetros de uma relação entre duas ou mais grandezas. Corresponde a encontrar de certo modelo, os parâmetros que minimizam a soma das distâncias quadradas entre as previsões do modelo e os dados reais, ponto a ponto, com o objetivo de diminuir ao máximo a distância entre elas.

Se temos M dados simulados \bar{y}_i e dados reais y_i , calculamos os mínimos quadrados (MQ) da seguinte maneira:

$$MQ = \sum_{n=1}^M (y_i - \bar{y}_i)^2 \quad (3.2)$$

A figura abaixo exemplifica o erro quadrático entre os dados reais (reta) e os dados simulados (pontos):

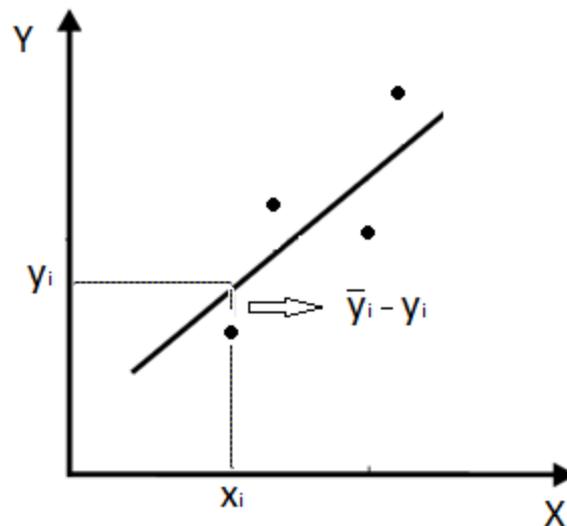


Figura 3.1 – Exemplificação do erro quadrático entre dados reais e dados gerados por algum modelo linear.

Por ser uma maneira simples e intuitiva de observar se o seu modelo fornece uma descrição razoável, muito se usa a estatística de Mínimos Quadrados para avaliar se os parâmetros estimados do modelo SIR (Suscetível – Infectado – Recuperado) se adequam bem ou não aos dados reais.

O nosso objeto de estudo é encontrar parâmetros para o modelo SIR que sejam capazes de realizar uma boa aproximação a dados reais de dengue. O método deverá comparar os valores presentes no compartimento I com os dados reais de infectados.

Seja I_r o vetor que representa o número de infectados de acordo com os dados reais, e I_e o vetor que representa o número de infectados, de acordo com a resolução do modelo SIR, calculamos a estatística dos mínimos quadrados substituindo os dados da equação (3.2) da seguinte maneira:

$$MQ = \sum_{n=1}^M (I_r - I_e)^2 \quad (3.3)$$

Para a aplicação no modelo SIR, o método precisa de um palpite inicial para os parâmetros. No nosso caso vamos estimar t_0 e R_0 mantendo γ fixo e vamos observar graficamente como a aproximação se comporta, calcular o erro quadrático e repetir o processo até que os parâmetros sejam relativamente bons. Uma boa maneira de começar é buscar em referências os valores os valores iniciais para os parâmetros e aplicar computacionalmente na resolução do modelo; observar como a função que aproxima os dados se comporta, e partir para um novo “palpite”, até que seja razoável a visualização gráfica e o erro quadrático.

3.2 Método de Monte Carlo

As simulações por esse método estão baseadas em eventos que ocorrem aleatoriamente, similares aos jogos de azar. Na Segunda Guerra Mundial, durante o Projeto Manhattan, Ulam e Von Neumann o denominaram método “Monte Carlo”, fazendo referência à famosa cidade de Mônaco conhecida mundialmente como a capital dos jogos de azar (ZAIDI e SGOUROS, 2003).

Trata-se de um método muito importante, quando o assunto é estimar parâmetros de uma dinâmica. Depende de amostragem aleatória repetida para

encontrar resultados numéricos (POLIMATHEYA).

Podemos sintetizar o algoritmo em alguns passos:

- Definir um intervalo de valores para cada parâmetro.
- Aleatoriamente gerar entradas seguindo determinada distribuição de probabilidade, dentro do intervalo pré estabelecido.
- Executar o modelo com as entradas geradas e observar o erro quadrático.
- Repetir o processo muitas vezes sempre colocando como valor ótimo aquele cujo erro é o menor.
- Devolver a entrada com os valores ótimos, isto é, aquela cujo erro quadrático foi o menor possível dentro de todo o processo.

Dentro do método de Monte Carlo, em nosso estudo de estimação de parâmetros do modelo SIR para a dengue, vamos estudar duas variações:

Método de varredura de parâmetros de Monte Carlo e Método de Monte Carlo usando log-probabilidade negativa de Poisson.

3.2.1 Método de varredura de parâmetros de Monte Carlo.

Ao invés de ficar alterando os parâmetros manualmente e observar tanto graficamente como através do erro quadrático, se os dados do modelo se aproximam aos dados reais, como é feito no método dos mínimos quadrados que vimos no tópico anterior, uma maneira mais fácil de realizar o processo é escrever um algoritmo que faça um loop, de modo a observar a amostragem de várias hipóteses diferentes para os parâmetros e que calcule a estatística de mínimos quadrados, mantendo o controle de como a estatística dos mínimos quadrados depende dos parâmetros. Com o objetivo determinado, podemos utilizar o Método de varredura de parâmetros de Monte Carlo (POLIMATHEYA).

O método tem como característica o uso de muitas iterações, e dentro de cada uma delas, aleatoriamente, observar uma amostra de hipóteses para

os parâmetros a partir de distribuições uniformes com intervalos apropriados que esperamos que estejam na estimativa do valor real do parâmetro. O próximo passo é, para cada um desses conjuntos de parâmetros, calcular a estimativa do modelo da distribuição de dados e comparar para ver se é um bom ajuste, dessa maneira, o método retornará o melhor valor, aquele com o menor erro quadrático.

É comum perguntar como saber quais intervalos usar para a varredura de parâmetros de Monte Carlo. A ideia é fazer algumas varreduras preliminares e verificar se os intervalos que serão usados parecem conter os valores que minimizam a estatística de mínimos quadrados, dessa maneira, o algoritmo que deverá ser produzido poderá se aproximar do ponto ideal, e gerará uma aproximação bem consistente (POLIMATHEYA).

3.2.2 Método de Monte Carlo via log-probabilidade negativa de Poisson.

Trata-se de uma estatística de ajuste que leva em consideração que os dados observados são distribuídos por Poisson. Isso é obtido otimizando a estatística de log-probabilidade negativa de Poisson.

Ao lidar com dados de contagem, como por exemplo os dados da dengue utilizados nesse trabalho, a estatística de probabilidade de Poisson é utilizada para avaliar quão bem um modelo particular descreve os dados. A estatística de probabilidade de Poisson pode, de fato, ser aplicada a casos em que alguns dos dados têm contagem baixa.

A probabilidade de observar k conta nos dados, com a distribuição de Poisson, quando o valor predito pelo modelo é λ , é dada por:

$$p(k|\lambda) = \frac{\lambda^k \cdot e^{-\lambda}}{k!} \quad (3.4)$$

Seja um conjunto de dados k_i (com $i = 1, \dots, N$), a probabilidade de observar esses pontos de dados com previsões de modelo para cada ponto, λ_i , é :

$$\mathcal{L} = \prod_{i=1}^N \frac{\lambda_i^{k_i} \cdot e^{-\lambda_i}}{k_i!} \quad (3.5)$$

esta probabilidade é maximizada com os parâmetros ótimos.

Tendo em mente a otimização do ajuste dos dados, é padrão lidar com a probabilidade negativa de log, que no caso da distribuição de Poisson, após os devidos ajustes, é dada pela seguinte equação:

$$-\log \mathcal{L} = \sum_{i=1}^N -k_i \cdot \log \lambda_i + \lambda_i \quad (3.6)$$

A qual deve ser minimizada com o objetivo de encontrar os melhores parâmetros para o modelo (POLYMATHEIA).

Capítulo 4

Resultados e métodos

Neste capítulo de resultados, vamos supor que uma pessoa infectada é introduzida em uma população completamente suscetível no tempo t_0 e vamos supor que o período infeccioso para a dengue é $1 / \gamma = 7$ dias. Dessa maneira teremos que $\gamma = 1/7$, e precisaremos estimar os valores de R_0 e t_0 , uma vez que a taxa β é calculada por (Equação (3.10)):

$$\beta = R_0 \cdot \gamma \quad (4.1)$$

Escolhemos para estudo os casos de dengue reportados na cidade do Rio de Janeiro no ano de 2012, sendo 1 de janeiro a 31 de dezembro. Para ajustarmos os modelos a estes dados, utilizaremos os três métodos apresentados anteriormente a fim de observar como cada um deles se apresenta, para efeito de comodidade chamaremos o estudo da seguinte maneira:

Estudo 1: Método de mínimos quadrados- estimacão pura e direta.

Estudo 2: Método de varredura de parâmetros de Monte Carlo.

Estudo 3: Método de Monte Carlo via log-probabilidade negativa de Poisson.

4.1 Estudo 1

Como vimos, o método de mínimos quadrados aplicado ao modelo SIR se trata de uma estimacão pura e direta, dessa maneira, “daremos um palpite” de algum valor para os parâmetros que desejamos estimar e observaremos como o gráfico que retrata os valores reais e os valores estimados pelo modelo se comporta, além de calcular o erro quadrático.

Como primeiro palpite, tomaremos valores pequenos para os dois

parâmetros a fim de observar como o gráfico se comporta:

Nas figuras abaixo, os pontos representam os valores reais de pessoas infectadas com dengue por semana, na cidade do Rio de Janeiro no ano de 2012, enquanto as linhas representam os valores gerados pelo modelo.

- Para $R_0 = 1.3$, $t_0 = 12$ (semana 2), $\gamma = 1/7$, o modelo gera o seguinte gráfico:

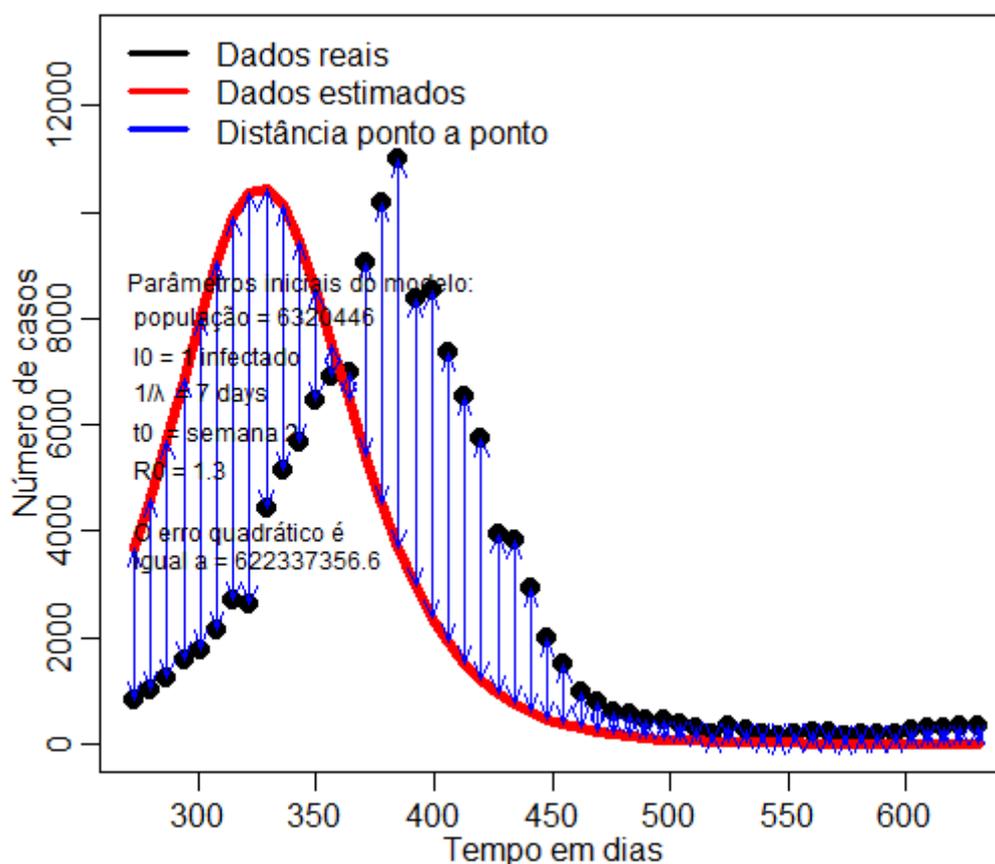


Figura 4.1 - Comparação entre os dados gerados pelo modelo SIR, usando $R_0 = 1.3$, $t_0 = 12$ (equivalente a semana 2) e $\gamma = 1/7$, e os dados reais.

Visualmente os dados estão muito distantes, além disso, o erro quadrático é igual a 622.337.356,6. Por isso, experimentaremos valores mais altos para os parâmetros.

- Para $R_0 = 1.22$, $t_0 = 34$ (semana 5), $\gamma = 1/7$ temos o seguinte gráfico:

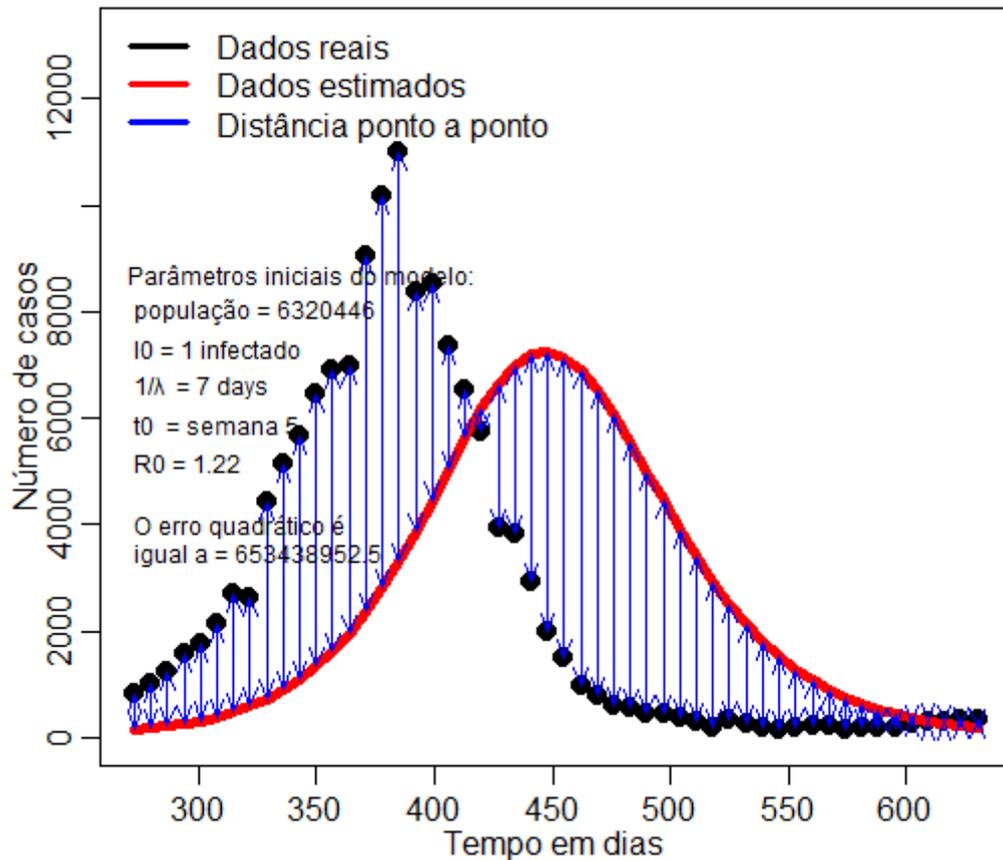


Figura 4.2 - Comparação entre os dados gerados pelo modelo SIR, usando $R_0 = 1.22$, $t_0 = 34$ (equivalente a semana 5) e $\gamma = 1/7$, e os dados reais.

Notamos visualmente e através do erro quadrático (igual a 653.438.952.5) que a segunda tentativa foi melhor do que a primeira. Agora busquemos um meio termo.

- Para $R_0 = 1.1$, $t_0 = 152$ (semana 22), $\gamma = 1/7$ que produzem o seguinte gráfico:

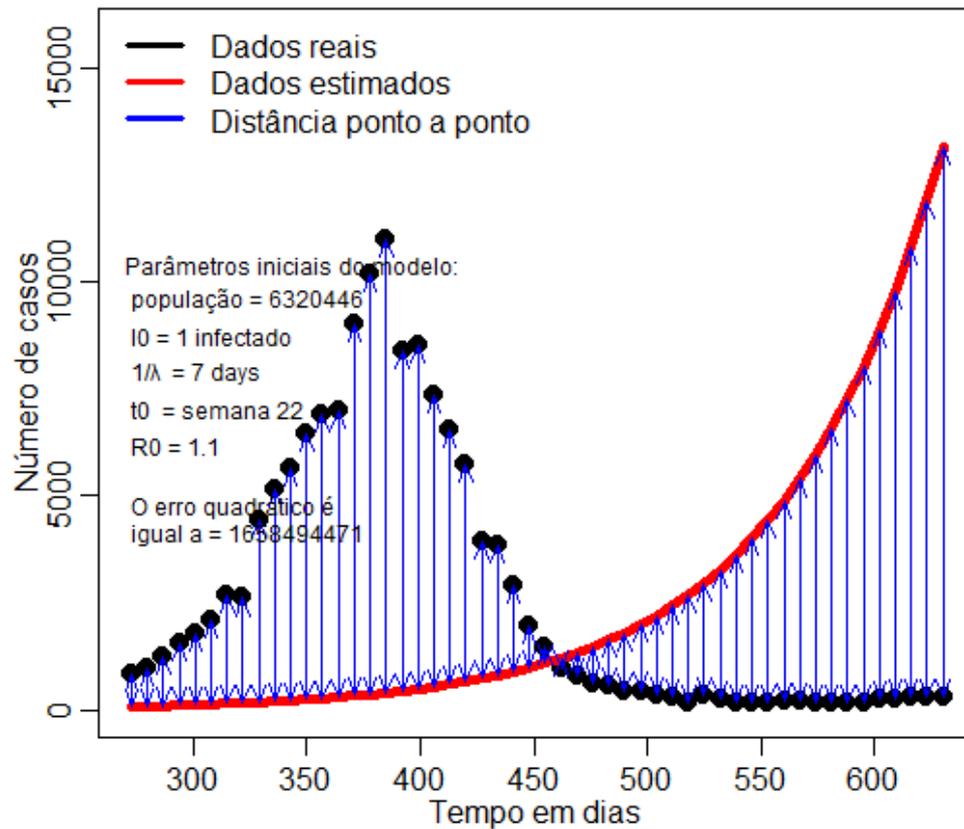


Figura 4.3 - Comparação entre os dados gerados pelo modelo SIR, usando $R_0 = 1.1$, $t_0 = 152$ (equivalente a semana 22) e $\gamma = 1/7$, e os dados reais.

Visualmente e através do erro quadrático(igual a 1.658.494.471), podemos concluir do **ESTUDO 1** que o teste 1 foi aquele que obteve a melhor aproximação entre os dados reais e os dados gerados pelo modelo, o que nos faz entender o quanto demoraria para chegar em um valor satisfatório. Por esse motivo que precisamos de outras maneiras para estimar bons parâmetros que possam nos devolver boas aproximações.

4.2 Estudo 2

Através da dificuldade encontrada no Estudo1, onde testamos parâmetros aleatoriamente, percebemos que se fosse possível testar milhares de parâmetros, em algum momento encontraríamos uma solução ótima, capaz de gerar uma aproximação satisfatória. É pensando nisso que usaremos no Estudo 2 o **Método de varredura de parâmetros de Monte Carlo**, um método que permite gerar valores aleatórios, testar o modelo para os valores dos parâmetros gerados e devolver os parâmetros ótimos através da observação do erro quadrático entre os dados gerados e os dados reais.

Rodamos o algoritmo com $\gamma = 1/7$ fixo, que retornou os seguintes valores ótimos para os parâmetros testados:

$$R_0 = 1.298, t_0 = 61 \text{ (semana 9)}$$

Observe abaixo o resultado que o algoritmo produz, com erro quadrático igual a 12.948.908,8.

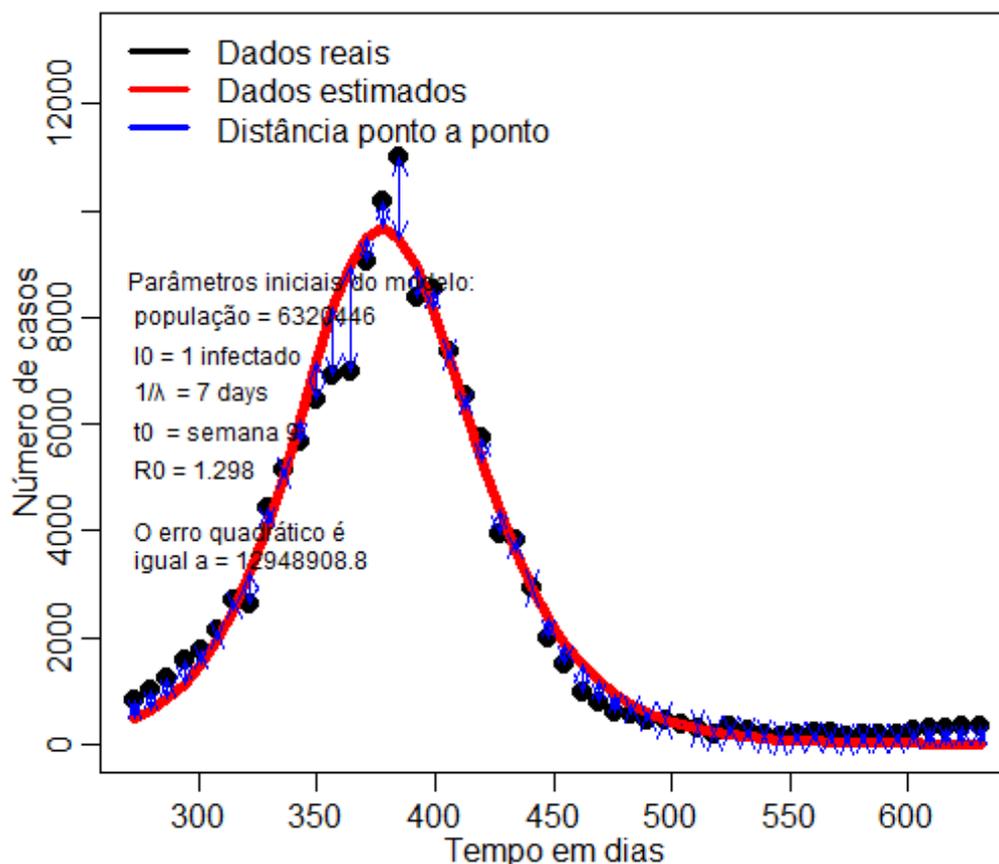


Figura 4.4 - Comparação entre os dados gerados pelo modelo SIR, usando $R_0 = 1.298$, $t_0 = 61$ (equivalente a semana 9) e $\gamma = 1/7$, e os dados reais.

Visualmente e através do erro quadrático, podemos observar que esta é a melhor aproximação entre os dados reais e os dados gerados pelo modelo.

O Estudo 2 calcula 10.000 iterações e plota o gráfico acima. É um estudo mais eficiente do que o Estudo 1, pois é necessário somente definir o intervalo dos parâmetros e deixar que o algoritmo retorne os valores ótimos.

4.3 Estudo 3

Trata-se do método de Monte Carlo (Figura 4.5) usando a estatística log-probabilidade negativa de Poisson. Calculam-se 10.000 iterações. Obtemos os parâmetros $R_0 = 1.265$, $t_0 = 27$ (semana 4), com erro quadrático = 16.526.598.

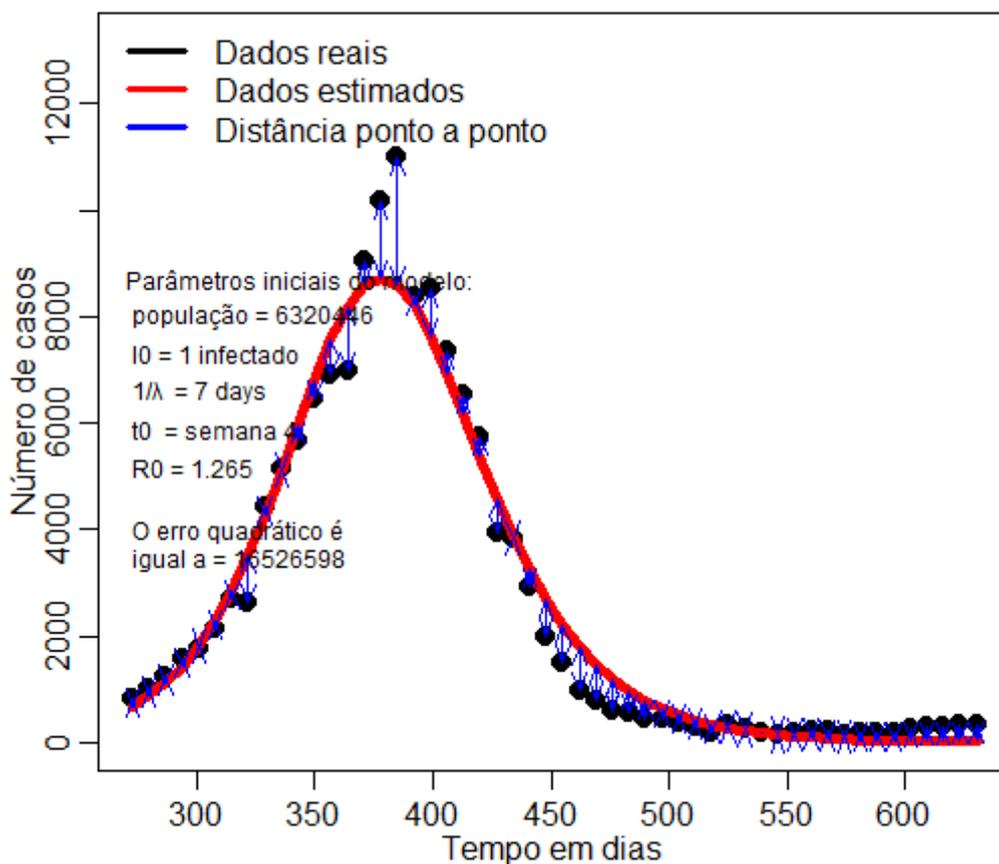


Figura 4.5 – Aproximação gerada pelo Método de Monte Carlo usando a estatística log-negativa de Poisson..

Conclusão

Fizemos um estudo para estimar parâmetro para o modelo SIR aplicado para descrever uma epidemia de dengue ocorrida na cidade do Rio de Janeiro em 2012. Para tal finalidade aplicamos três métodos estatísticos.

O Método de mínimos quadrados, é aquele cuja implementação é a mais simples com o menor tempo de rodagem do programa. Contudo, trata-se do método mais impreciso por ser necessário “chutar” valores dos parâmetros e observar como a aproximação reage.

O Método de varredura de parâmetros de Monte Carlo é um método de média complexidade para implementação com tempo de rodagem médio, é aquele que consegue solucionar o problema do primeiro método e encontrar dentro de um intervalo de valores em que os parâmetros podem variar, a melhor aproximação comparando os dados reais e os dados estimados através do erro quadrático.

O Método de Monte Carlo usando a estatística log-probabilidade negativa de Poisson também é um método de média complexidade para implementação e possui tempo de rodagem equivalente ao Método de varredura de parâmetros de Monte Carlo.

Após realizar os testes para os três métodos, concluímos que aquele que retornou o menor erro quadrático, ou seja, a menor diferença entre os dados para dengue na cidade do Rio de Janeiro no ano de 2012 e os dados gerados foi o **Método de varredura de parâmetros de Monte Carlo**, com erro quadrático igual a 12.948.908,8, como pode ser observado na Figura 4.4. Esta constatação foi feita através da observação do erro quadrático, mas era também facilmente percebida através da observação dos gráficos.

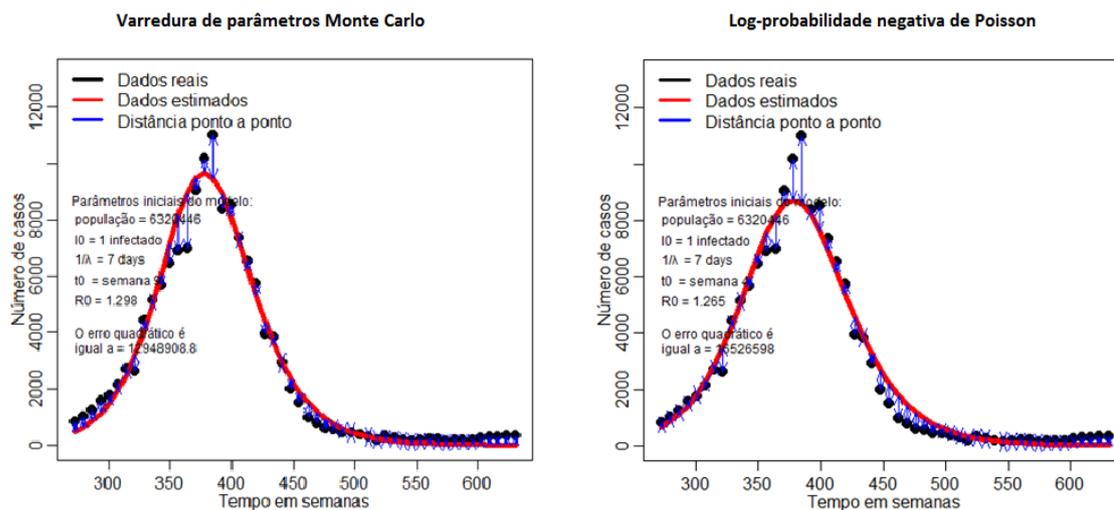


Figura 5.1 - Comparação entre os gráficos gerados pelos métodos de varredura de parâmetros e log-probabilidade negativa de Poisson.

Referências

BARRETO et al. (2008) Barreto, Maurício L.; Teixeira, Maria Glória. Dengue no Brasil: situação epidemiológica e contribuições para uma agenda de pesquisa. *Estud. av., São Paulo*, v. 22, n. 64, p. 53-72, dez. 2008. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-40142008000300005&lng=pt&nrm=iso>. acesso em 19 jan. 2019.

FARES et al. (2015) R. C. G. Fares, K. P. R Souza, G. Añez e M. Rios. Epidemiological scenario of dengue in Brazil. *BioMed research international*, 2015.

KEELING e ROHANI (2008) M. J. Kelling e P. Rohani. *Modeling infectious disease in humans and animals*. Princeton University.

KERMACK, W. O.; MCKENDRICK, A. G. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. A.*, 1927.

MACIEL et al. (2008) Maciel, I. J.; Siqueira, J. J. B.; Martelli, C. M. T. Epidemiologia e Desafios no controle do Dengue. [S.l.], v.37, n.2, p.111-130, mai-jun 2008.

MARTCHEVA (2015) Maia Martcheva. An introduction to mathematical epidemiology, volume 61. Springer

MINISTÉRIO DA SAÚDE/SVS(2018) Ministério da Saúde/SVS(2018).DATASUS- Sistema de informações de agravos de notificações – Sinan Net. <http://tabnet.datasus.gov.br/cgi/deftohtm.exe?idb2012/d0203.def>, 2018. Último acesso em 10/12/18.

POLYMATHEIA.,. [http:// sherrytowers.com](http://sherrytowers.com). Ultimo acesso em 19/01/2019.

PORTAL EDUCAÇÃO. Histórico da Dengue, <https://www.portaleducacao.com.br/conteudo/artigos/medicina/historico-da-dengue/38336>. Ultimo acesso em 17/01/2019

PEIXOTO, M. S. e BARROS, L. C. (2004). Um estudo de autômatos celulares para o espalhamento geográfico de epidemias com parâmetro fuzzy, XXXVI CNMAC: *Tend. Mat. Apl. Comput* 5(1): 125–133

WEISSTEIN(2010) Weisstein, E. W. SIR Model. 2010. Disponível em: <<http://mathworld.wolfram.com/SIRModel.html>>.

ZAIDI e SGOUROS (2003) Zaidi H, Sgouros G. *Therapeutic Applications of Monte Carlo Calculations in Nuclear Medicine*. London: Institute of Physics Publishing; 2003.

