



Universidade Federal do Rio Grande do Sul
Instituto de Matemática e Estatística
Programa de Pós-Graduação em Estatística

**Estimação para o modelo compartimental
SIRS via filtragem iterada para processos
de Markov parcialmente observados
utilizando o pacote pomp do R: uma análise
dos casos de influenza A**

Rafaela Gomes de Jesus

Porto Alegre, Dezembro de 2021.

Dissertação submetida por Rafaela Gomes de Jesus como requisito parcial para a obtenção do título de Mestre em Estatística pelo Programa de Pós-Graduação em Estatística da Universidade Federal do Rio Grande do Sul.

Orientador:

Dra. Gabriela Bettella Cybis (UFRGS)

Coorientador:

Dr. Guilherme Pumi (UFRGS)

Comissão Examinadora:

Dr. José Luiz Padilha da Silva (UFPR)

Dr. Rodrigo Citton Padilha dos Reis (UFRGS)

Dra. Taiane Schaedler Prass (UFRGS)

Data de Apresentação: 23 de Dezembro de 2021

AGRADECIMENTOS

- Aos meus pais, Adriana e Jeferson, e à minha irmã Júlia por todo amor e suporte que sempre me deram;
- Ao meu namorado Geovane, que esteve do meu lado me apoiando e incentivando;
- Aos amigos de longa data e aos amigos que ganhei durante o mestrado, especialmente ao Artur, ao Felipe e à Janaína;
- À professora Gabriela e ao professor Guilherme, pela orientação neste trabalho, pela paciência e por todo o conhecimento transmitido;
- Aos demais professores e funcionários do Programa de Pós-graduação em Estatística da UFRGS;
- Ao ensino público, gratuito e de qualidade.

CIP - Catalogação na Publicação

de Jesus, Rafaela Gomes
Estimação para o modelo compartimental SIRS via
filtragem iterada para processos de Markov
parcialmente observados utilizando o pacote pomp do R:
uma análise dos casos de influenza A / Rafaela Gomes
de Jesus. -- 2021.
48 f.
Orientadora: Gabriela Bettella Cybis.

Coorientador: Guilherme Pumi.

Dissertação (Mestrado) -- Universidade Federal do
Rio Grande do Sul, Instituto de Matemática e
Estatística, Programa de Pós-Graduação em Estatística,
Porto Alegre, BR-RS, 2021.

1. Influenza A. 2. Modelos POMP. 3. Modelo SIRS. 4.
Filtro de partículas. 5. Filtragem iterada. I. Cybis,
Gabriela Bettella, orient. II. Pumi, Guilherme,
coorient. III. Título.

RESUMO

A gripe é uma doença que se espalha globalmente todos os anos e, até antes da pandemia do coronavírus em 2019, infectava entre 10% a 20% da população mundial, causando aproximadamente 500.000 mortes anuais. Uma forma de acompanhar a progressão da gripe no Brasil é através do número de casos de infecção reportados na plataforma INFOgripe, que faz a vigilância da Síndrome Respiratória Aguda Grave (SRAG) no país. Entretanto, devido à expressiva sub-notificação, esses dados representam apenas um retrato parcial da dinâmica da epidemia.

Para estudar a dinâmica do vírus influenza A, a série temporal do número de casos positivos pela infecção deste patógeno foi extraída e tratada como um processo de Markov parcialmente observado. Para modelar a série e fazer a inferência dos parâmetros que determinam a dinâmica da doença, fez-se uso do modelo compartimental SIRS (Suscetível - Infectado - Recuperado - Suscetível) que, ao dividir a população em compartimentos, fornece uma estrutura base para compreender a transmissão do vírus entre os indivíduos.

Este trabalho teve como objetivo estimar os parâmetros que governam a intensidade do fluxo de indivíduos entre os compartimentos do modelo SIRS utilizando o método da filtragem iterada, baseada na função de verossimilhança. A partir destas estimativas, valores para o período médio de infecção (μ_i), o período médio de perda de imunidade (μ_r) e o número reprodutivo básico (R_0) da doença foram avaliados. O parâmetro ρ , que representa a taxa de reporte da doença, também foi estimado e uma parametrização alternativa para o modelo foi proposta com o objetivo de aumentar a eficiência do algoritmo que realiza a estimação. O pacote `pomp`, implementado no software R, foi utilizado para realizar toda a parte computacional.

Os resultados apresentados indicam que a utilização da parametrização alternativa produziu valores de parâmetros que aparentam representar um máximo global para a função de verossimilhança. Dentre os resultados obtidos, observamos que quando a magnitude de ρ está na ordem de 10^{-4} , o conjunto de parâmetros resultantes são compatíveis com valores de referência na literatura. Este estudo deve ser compreendido como um passo intermediário na modelagem do vírus influenza A. Trabalhos futuros podem considerar o ajuste de modelos mais complexos aos dados, introduzindo, por exemplo, covariáveis para os parâmetros de interesse.

Palavras-chave: Influenza A, modelos POMP, modelo SIRS, filtro de partículas, filtragem iterada

ABSTRACT

The Influenza virus spreads globally every year and, until before the coronavirus pandemic in 2019, infected from 10% to 20% of the world's population, causing approximately 500.000 deaths annually. A way of keeping up with the progress of Influenza in Brazil is through the number of reported cases on the INFOgripe, which surveys Acute Respiratory Distress Syndrome (ARDS) in the country. However, due to high subnotification levels, these data represent only a partial picture of the disease's dynamic.

In order to study the dynamic of the virus, we extracted the time series of positive cases reported for Influenza A and treated it as a partially observed Markov process. To draw inference on the parameters that determine the dynamic of the disease, we employed the SIRS (Susceptible - Infected - Recovered - Susceptible) model that splits the population into compartments, providing a framework for studying the transmission of the virus between individuals.

The main objective of this study is to estimate the parameters that determine the flow of individuals between SIRS model compartments using the iterated filtering method. From these estimates, values for the average infection period (μ_i), average loss of immunity period (μ_r) and the basic reproductive number (R_0) of the disease were evaluated. The ρ parameter, which represents the disease's report rate was also estimated, and an alternative parametrization for the model was proposed with the aim of increasing efficiency in the estimating algorithm. The `pomp` package, implemented in the R software, was used for all computations.

Our findings indicate that the alternative parameterization yields parameter values close to a global maximum for the likelihood function. Our study reveals that when the magnitude of ρ is about 10^{-4} , the resulting set of parameters are compatible with reference values found in the literature. The present study should be understood as an intermediate step in modeling influenza A virus. This dissertation represents an attempt at modeling the influenza A and future work may consider fitting more complex models to the data, with the introduction, for example, of covariates for the parameters of interest.

Keywords: Influenza A, POMP models, SIRS model, particle filter, iterated filtering

ÍNDICE

1	Introdução	2
2	Revisão Bibliográfica	5
2.1	Modelos compartimentais	5
2.2	Modelo Suscetível-Infected-Recuperado (SIR)	6
2.3	Modelo Suscetível-Infected-Recuperado-Suscetível (SIRS)	7
2.4	Número básico de reprodução (R_0)	9
2.5	Processo de Markov parcialmente observado	9
2.6	Função de verossimilhança	11
2.7	Filtro de partículas	12
2.8	Filtragem iterada	14
3	O banco de dados	16
4	O pacote pomp	19
4.1	Implementação do modelo SIRS no pacote pomp	20
5	Resultados	24
5.1	Estimação para o modelo completo	24
5.2	Estimação com valores fixos para a taxa de detecção ρ	28
5.3	Estimação com valores fixos para a taxa de recuperação γ	31
6	Conclusões e trabalhos futuros	33

CAPÍTULO 1

INTRODUÇÃO

A influenza é uma infecção respiratória viral cujos sintomas mais comuns são calafrios, febre, dores de garganta, musculares e de cabeça, tosse, fadiga e sensação geral de desconforto (Eccles, 2005). Estimativas anteriores à pandemia do coronavírus em 2019 mostravam que, a cada ano, a doença infectava entre 10% e 20% da população mundial e causava em torno de 500.000 mortes anuais (Stöhr, 2002; Neher and Bedford, 2015). A gravidade da doença é ainda maior quando ocorrem pandemias, o que eleva o número de óbitos a milhões. A mortalidade ocorre principalmente entre pacientes considerados de alto risco - pessoas com mais de 50 anos, adultos e crianças com problemas pulmonares, cardiovasculares, disfunção renal, doenças metabólicas e imunossupressoras, crianças que fazem tratamento com uso prolongado de aspirina e grávidas. (Forleo-Neto et al., 2003; Taubenberger and Morens, 2008).

Com altas taxas de mutação, o vírus está constantemente modificando seu material genético, eventualmente gerando variantes resistentes às vacinas vigentes e, com isso, infectando a população que não possui imunidade para combatê-las. Por conta da evolução do vírus, a vacina contra a influenza deve ser administrada anualmente. Existem três tipos de vírus influenza que infectam humanos: influenza A, B e C. Dos três tipos, apenas os dois primeiros estão associados com morbidade e mortalidade sazonal significativa. Neste trabalho, escolhemos estudar apenas os vírus do tipo influenza A, cujos subtipos H1N1 e H3N2 foram responsáveis pelos maiores números de casos nas últimas décadas.

Segundo a Dra. Wenqing Zhang, gerente do Programa Global de Influenza da OMS, “outra pandemia causada por um novo vírus da gripe é uma certeza. Mas não sabemos quando isso vai acontecer, que cepa de vírus será e quão grave será a doença. Esta incerteza torna a gripe muito diferente de muitos outros patógenos” (Zhang, 2021). Portanto, é de grande importância acompanhar o processo de evolução do vírus ao longo do tempo pois, dessa forma, pode-se tomar medidas mais assertivas para interferir em seu processo de propagação.

Uma forma de estudar a dinâmica de doenças infecciosas é através da modelagem matemática. Os modelos compartimentais são uma classe de modelos matemáticos amplamente utilizados para descrever o curso de uma doença infecciosa numa determinada população. A base dos modelos compartimentais é conceber que a população possa ser dividida em compartimentos ou classes. Cada compartimento indica o estado atual da doença nos indivíduos que pertencem a ela, sendo assim, cada indivíduo pode estar apenas em um compartimento por unidade de tempo.

Um dos modelos mais simples para descrever a evolução de uma doença infecciosa é o modelo compartimental SIR, sigla para os compartimentos Suscetível, Infectado e Recuperado.

O compartimento S representa o número de indivíduos suscetíveis à doença, I o número de indivíduos que tenham sido infectados com a doença e que são capazes de transmiti-la aos indivíduos da categoria suscetível e R os indivíduos que foram infectados e, em seguida, removidos do processo de transmissão da doença, quer devido à imunização ou devido à morte. No modelo SIR, os indivíduos da categoria R não são capazes de ser infectados novamente ou de transmitir a infecção para outras pessoas.

O modelo SIR serve de ponto de partida para generalizações com o objetivo de acomodar séries temporais apresentando comportamentos mais complexos, que são geralmente incorporadas à dinâmica do modelo através da introdução de novos compartimentos (ou interações entre os compartimentos) ao modelo SIR base. No caso do influenza A, quando uma nova variante do vírus surge e ainda não se tem imunidade para combatê-la, o modelo SIR pode ser estendido para acomodar a capacidade de um indivíduo recuperado se tornar suscetível novamente. Estes são chamados de modelos SIRS, em que o último compartimento S representa o retorno de um indivíduo recuperado para o estado Suscetível.

Uma limitação dos modelos compartimentais em suas versões originais é que eles são determinísticos, o que significa que, dadas as condições iniciais, o modelo sempre irá fornecer os mesmos resultados, característica que vai contra a natureza estocástica de uma epidemia. Além disso, não observamos o número total de casos de influenza, mas sim os casos reportados da doença. Assumimos que este processo observado ao longo do tempo consiste em uma série temporal de medições incompletas e ruidosas de um processo de Markov latente não observado. Para contornar a primeira limitação, pode-se adaptar o modelo compartimental determinístico à uma versão estocástica e, para contornar a segunda, utilizam-se os modelos de processos de Markov Parcialmente Observado (POMP), que tratam o processo observado como uma variável aleatória cuja distribuição depende de um processo latente não-observado. Dessa forma, se torna possível estimar parâmetros de interesse do processo latente com base no processo observado.

Objetivo

Este projeto tem como objetivo estudar a metodologia do filtro iterado para realizar a estimação dos parâmetros do modelo epidemiológico SIRS para o vírus influenza A; investigar os efeitos de uma parametrização alternativa do modelo na eficiência da estimação; abordar a estimação dos parâmetros dentro dos valores de referência.

Organização do trabalho

A dissertação está organizada como segue: no Capítulo 2 apresentamos uma revisão bibliográfica, onde introduzimos conceitos básicos dos modelos compartimentais determinísticos SIR e SIRS, do modelo de processo de Markov parcialmente observado e da metodologia estatística utilizada para inferência dos parâmetros. No Capítulo 3 descrevemos o banco de dados modelado e, no Capítulo 4, descrevemos o pacote pomp (versão 3.6, [King et al., 2021](#)) do R (versão 4.0.3, [R Core Team, 2020](#)), utilizado na pesquisa, juntamente com o código executado para realizar as análises. Em seguida, no Capítulo 5, apresentamos os resultados obtidos. Por fim, a conclusão acerca do trabalho desenvolvido e alguns possíveis caminhos futuros são discutidos no Capítulo 6.

O banco de dados

O banco de dados utilizado para a modelagem realizada neste trabalho foi extraído da plataforma INFOGripe. As observações do banco de dados iniciam no ano de 2009 e estão organizadas por semana epidemiológica, o que faz com que a primeira semana de 2009 seja a primeira observação do banco. Para cada semana epidemiológica se observa o número de casos positivos de influenza A registrados no Brasil no decorrer da respectiva semana.

Embora o banco de dados tenha sido extraído no ano de 2020, resultando em mais de 600 semanas epidemiológicas, utilizamos apenas as informações dos anos de 2016 e 2017 (104 semanas epidemiológicas). Além disso, contabilizamos apenas os casos de influenza A registrados no Rio Grande do Sul. Informações sobre a plataforma INFOGripe, os dados que estão disponíveis para extração e as justificativas para o recorte do banco de dados estão detalhados no [Capítulo 3](#).

CAPÍTULO 2

REVISÃO BIBLIOGRÁFICA

2.1 Modelos compartimentais

Os modelos compartimentais são modelos matemáticos frequentemente aplicados à modelagem de doenças infecciosas. Um dos objetivos destes modelos é descrever a transmissão de um patógeno em uma população. Os modelos compartimentais também podem ser utilizados para estimar parâmetros de uma epidemia, como por exemplo, a taxa de transmissão. Para tal, dividem a população em compartimentos, os quais indicam o estado de cada indivíduo em relação à doença de interesse.

A primeira modelagem matemática de propagação de doenças infecciosa foi atribuída ao matemático e físico holandês Daniel Bernoulli (1700-1782). Bernoulli demonstrou a eficácia da técnica de inoculação preventiva contra a varíola, que poderia ampliar a expectativa de vida de 26 anos e 7 meses para 29 anos e 9 meses (Graunt, 1939). O modelo de Bernoulli dividiu a população em dois grupos: os suscetíveis, que ainda não haviam sido infectados, e os imunes, que já tinham adquirido proteção contra a doença depois de a terem contraído.

Em 1927, tomando como base o modelo de Bernoulli, o bioquímico William Ogilvy Kermack (1898-1970) e o epidemiologista Anderson Gray McKendrick (1876-1943) elaboraram uma teoria sobre a transmissão de doenças infecciosas, que ficou conhecido por modelo SIR (Suscetível - Infectado - Recuperado), no qual os indivíduos são classificados como “suscetível”, “infectado” ou “recuperado” (Kermack and McKendrick, 1927).

O SIR é um dos modelos frequentemente utilizados em aplicações envolvendo a evolução de epidemias. Em Bjørnstad et al. (2002), por exemplo, o SIR é utilizado para descrever a dinâmica do sarampo e estimar a taxa de transmissão da doença e em Kudryashov et al. (2021), para investigar os parâmetros de interesse para a infecção por COVID-19. O modelo SIR serve de ponto de partida para diversas generalizações, aproveitando sua dinâmica base e acrescentando outras dinâmicas evolucionárias ao modelo. Entre alguns modelos desenvolvidos a partir do modelo SIR estão os modelos SIS, SIRS, SEIS e SEIR, dos quais tratamos abaixo.

O modelo SIS (Suscetível - Infectado - Suscetível) se diferencia do SIR pelo fato de que os indivíduos se recuperam sem imunidade à doença, retornando imediatamente ao estado suscetível. Em Zhou and Liu (2003), os autores descrevem o comportamento do modelo SIS. Já o modelo SIRS (Suscetível - Infectado - Recuperado - Suscetível) é simplesmente uma extensão do SIR, permitindo que indivíduos recuperados retornem para o estado suscetível após perderem imunidade.

Para modelar adequadamente as doenças que têm uma fase latente, durante a qual o in-

divíduo já contraiu a doença mas ainda não está infeccioso, criaram-se os modelos com o compartimento E (exposto), como é o caso do SEIS (Suscetível - Exposto - Infectado - Suscetível) e do SEIR (Suscetível - Exposto - Infectado - Recuperado). Definições e mais detalhes sobre estes dois modelos podem ser encontradas em [Korobeinikov \(2004\)](#). Quando considerado relevante, pode-se também agregar outros fatores que interferem no curso da infecção, como a natalidade e a mortalidade da população ([Brauer et al., 2019](#)).

A adição de novos compartimentos visa enriquecer e flexibilizar a dinâmica do modelo. Entretanto, ao adicionar compartimentos no modelo, adiciona-se também complexidade. Apesar de haver diversos outros fatores que contribuem para a transmissão de doenças, muitos modelos omitem a maioria dos detalhes e são projetados apenas para destacar o comportamento geral da dinâmica de interesse. Isso acontece pois excesso de detalhes contribuem pouco para modelar as principais tendências da dinâmica. Um modelo matemático simplificado ajuda a concentrar a atenção nos processos essenciais de uma doença infecciosa e a revelar os parâmetros que são mais influentes e passíveis de controle ([Brauer et al., 2019](#)).

2.2 Modelo Suscetível-Infectado-Recuperado (SIR)

O modelo compartimental SIR é comumente utilizado para estudar epidemias a nível populacional. Nele, assume-se que o processo epidêmico é determinístico, o tamanho da população é fixo e cada indivíduo deve ser classificado em um, e apenas um, compartimento. Assume-se também que o processo de nascimento e morte na população é muito mais lento que a taxa de infecção e recuperação podendo ser, portanto, ignorado. No modelo SIR, os possíveis compartimentos são três:

- $S(t)$, representando o número de indivíduos suscetíveis à doença no tempo t ;
- $I(t)$, representando o número de indivíduos que tenham sido infectados com a doença e que são capazes de transmiti-la aos indivíduos da categoria suscetível, no tempo t ;
- $R(t)$, representando os indivíduos que foram infectados e removidos do processo de transmissão da doença no tempo t , quer devido à imunização ou devido à morte. Os que estão nesta categoria não são capazes de serem infectados novamente ou de transmitir a infecção para outras pessoas.

A Figura 2.1 mostra uma representação esquemática do modelo SIR, onde as setas entre os compartimentos representam o fluxo de indivíduos. Cada fluxo está associado a uma taxa.

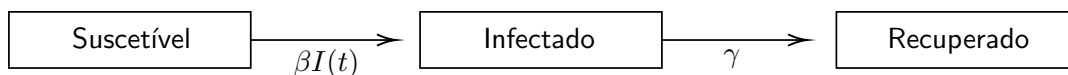


Figura 2.1: Fluxograma do modelo determinístico SIR.

O fluxo de indivíduos que saem do compartimento suscetível para o infectado está associado a uma taxa β , que é a taxa de infecção da doença. Isto significa que um indivíduo na população, quando entra em contato com um indivíduo infectado, contrai a doença a uma taxa β . O fluxo do compartimento infectado para o recuperado acontece a uma taxa γ , que representa a taxa média de recuperação ou morte do infectado. A fração $1/\gamma$ é chamada de período médio de infecção (μ_i). O modelo está definido para taxas γ e β maiores que zero ($\gamma > 0$ e $\beta > 0$).

A partir das definições anteriores, o modelo SIR pode ser formalizado matematicamente em termos de um sistema de equações diferenciais ordinárias. Relembrando que a população total

P é fixa, temos, em qualquer instante de tempo t ,

$$P = S(t) + I(t) + R(t).$$

O modelo SIR é definido através do sistema de equações

$$\begin{aligned}\frac{\partial S(t)}{\partial t} &= -\beta S(t)I(t), \\ \frac{\partial I(t)}{\partial t} &= \beta S(t)I(t) - \gamma I(t), \\ \frac{\partial R(t)}{\partial t} &= \gamma I(t).\end{aligned}$$

A Figura 2.2 exemplifica a dinâmica de uma população através do modelo determinístico SIR. No gráfico, a abcissa representa a semana epidemiológica e as ordenadas, a proporção da população em cada compartimento. Na semana 0, toda a população se encontra como suscetível. Com o passar do tempo, os indivíduos contraem a doença e, por fim, se recuperam. No exemplo, ao final da semana 50 é atingido o ponto de equilíbrio do modelo no qual toda a população é considerada recuperada da infecção. Nesse modelo é possível observar apenas um pico para a curva dos infectados.

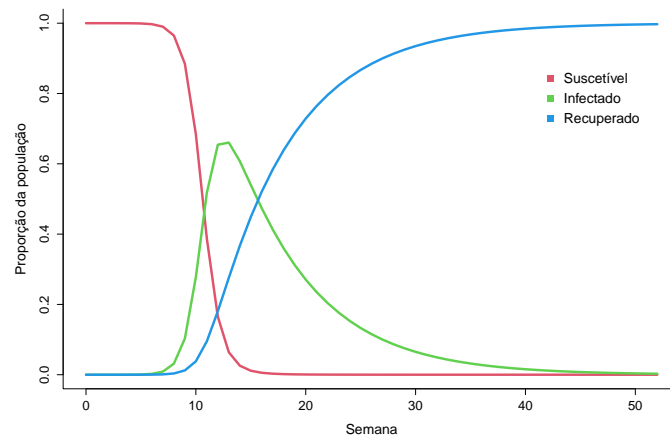


Figura 2.2: Exemplo da dinâmica do modelo determinístico SIR.

2.3 Modelo Suscetível-Infetado-Recuperado-Suscetível (SIRS)

Com base nas ideias básicas da estrutura SIR, vários outros modelos foram desenvolvidos para incorporar mais detalhes do processo de transmissão e da dinâmica das doenças infecciosas. Quando a imunidade a uma doença infecciosa é apenas temporária, o modelo SIR é estendido para acomodar a capacidade de um indivíduo recuperado se tornar suscetível novamente. Estes são chamados de modelos SIRS e são formulados como um sistema de equações diferenciais

ordinárias escritas como

$$\begin{aligned}\frac{\partial S(t)}{\partial t} &= -\beta S(t)I(t) + \alpha R(t), \\ \frac{\partial I(t)}{\partial t} &= \beta S(t)I(t) - \gamma I(t), \\ \frac{\partial R(t)}{\partial t} &= \gamma I(t) - \alpha R(t),\end{aligned}\tag{2.1}$$

onde as variáveis $S(t)$, $I(t)$ e $R(t)$ e as taxas β e γ possuem os mesmos significados e interpretações do modelo SIR, apresentado na Seção 2.2. O parâmetro α , por sua vez, representa a taxa de indivíduos migrando do compartimento R para o S , ou seja, é a taxa de indivíduos que retornam ao estado suscetível após infecção, por perder imunidade à doença. Analogamente, $1/\alpha$ representa o período médio de perda da imunidade (μ_r). O modelo está definido para os parâmetros β , γ e α assumindo valores maiores que zero ($\beta > 0$, $\gamma > 0$ e $\alpha > 0$). A Figura 2.3 ilustra o esquema do modelo SIRS. As setas, acompanhadas por suas respectivas taxas, indicam a direção e a intensidade do fluxo de indivíduos entre compartimentos.

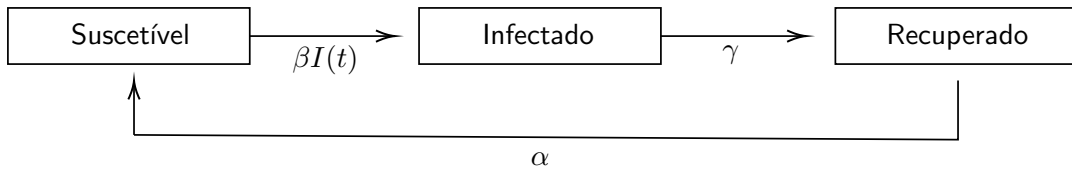


Figura 2.3: Fluxograma do modelo determinístico SIRS.

O modelo SIRS é amplamente utilizado em aplicações. Por exemplo, em [Sun and Yang \(2010\)](#), os autores utilizam o modelo SIRS para investigar e generalizar o impacto da vacinação e do isolamento de indivíduos infectados. Em [Weidemann et al. \(2014\)](#) os autores utilizam o modelo SIRS para analisar dados relativos à infecção por Rotavírus na Alemanha. Já em [Chen et al. \(2014\)](#), os autores utilizam o modelo SIRS para estudar o impacto da idade no processo de transmissão de doenças infecciosas.

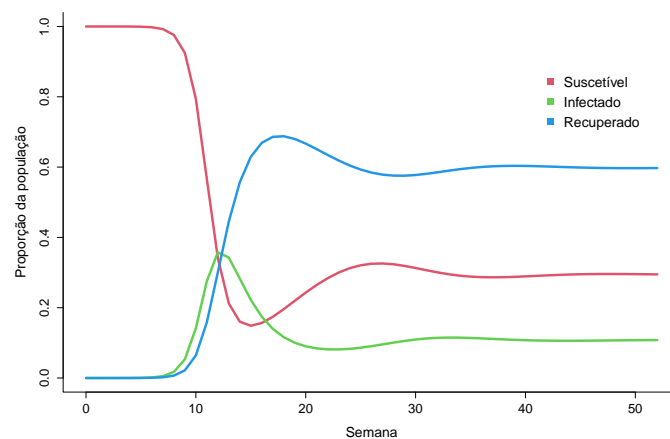


Figura 2.4: Exemplo da dinâmica do modelo determinístico SIRS.

A Figura 2.4 exemplifica a dinâmica de uma população através do modelo SIRS. A interpretação do gráfico é a mesma da Figura 2.2, mas a dinâmica entre os compartimentos ao longo

do tempo é diferente. Para o modelo SIRS, indivíduos recuperados retornam ao estado suscetível após perder a imunidade contra a doença e podem contraí-la novamente. Essa característica faz com que seja possível observar mais de uma onda de infecção.

2.4 Número básico de reprodução (R_0)

O número básico de reprodução (R_0), também conhecido por número reprodutivo básico, é uma métrica epidemiológica usada para descrever a transmissibilidade de agentes infecciosos e que foi muito discutido pela mídia desde o início da pandemia do coronavírus, em 2019. O valor de R_0 é utilizado por cientistas e tomadores de decisão para explicar e justificar políticas públicas de controle de uma doença infecciosa. Pode-se definir R_0 como o número esperado de casos secundários produzidos por um único indivíduo infectado durante todo o período de infecciosidade da doença quando a população está completamente suscetível (Diekmann et al., 1990).

Suponha que temos o interesse em determinar se uma doença contagiosa pode ou não se alastrar em uma população estável, onde todos os indivíduos são suscetíveis à doença. Podemos inferir isto a partir do valor de R_0 : se $R_0 > 1$ a doença possui potencial de invadir a população e causar uma epidemia, ao passo que se $R_0 < 1$ isto tende a não ocorrer. Quando $R_0 = 1$, então uma pessoa é capaz de transmitir a doença para apenas uma outra pessoa, em média.

O R_0 é afetado por vários fatores biológicos, sócio-comportamentais e ambientais que governam a transmissão de patógenos. É um valor que raramente é medido diretamente, pois depende das estruturas e suposições do modelo epidemiológico (Delamater et al., 2019). Por isso, a estimativa de R_0 é obtida mais facilmente de maneira indireta, através dos parâmetros definidos no modelo.

Tomando como base o modelo SIRS, explicado na Seção 2.3, definimos três parâmetros que governam a intensidade do fluxo de indivíduos entre os compartimentos: β , γ e α . Os parâmetros β e γ são definidos como a taxa de infecção da doença e a taxa média de recuperação, respectivamente. A relação entre β e γ define o grau de invasão de uma doença infecciosa, isto é, a velocidade que uma doença infecciosa pode se alastrar na população. O valor de R_0 pode ser obtido através da expressão (Mohd and Sulayman, 2020)

$$R_0 = \frac{\beta}{\gamma},$$

de onde concluímos que $R_0 > 1$ quando $\beta > \gamma$ e $R_0 < 1$ quando $\beta < \gamma$. Na literatura, a estimativa para o R_0 para o vírus influenza A H1N1 na pandemia de 2009 está entre 1,2 e 2,3, com mediana de 1,5 (Boëlle et al., 2011). Para o surto de vírus ebola em 2014, que ocorreu na África, as estimativas de R_0 são de 1,51 para Guiné, 2,53 para Serra Leoa e 1,59 para a Libéria (Althaus, 2014). Resultados indicando uma potencial pandemia do Covid-19 foram apresentadas em Zhao et al. (2020), em que R_0 foi estimado variando de 2,24 a 3,58. Em uma revisão sistemática apresentada em Guerra et al. (2017), estimativas de R_0 para o surto de sarampo ocorrido na Inglaterra em 2012 variaram entre 6,2 e 9,5.

2.5 Processo de Markov parcialmente observado

O processo de Markov parcialmente observado é também conhecido como modelo de espaço de estados, modelo oculto de Markov ou também por POMP, que é a sigla derivada do acrônimo

em inglês: *Partially Observed Markov Processes*. Um POMP consiste em medições incompletas e ruidosas de um processo Markoviano latente e não observado, conectado aos dados observados por meio de um processo de medição, que é definido por uma função que produz as observações a partir do processo latente (King et al., 2016). As observações obtidas através do processo de medição são consideradas como uma variável proxy para os valores dos estados ocultos do processo latente.

A razão pela qual um POMP é um bom candidato para a modelagem de doenças infecciosas é que o verdadeiro número de indivíduos suscetíveis, infectados e recuperados na população é desconhecido. Neste caso, o número relatado de casos de uma doença é vista como uma observação parcial do número total de pessoas infectadas, podendo ser tratado como uma taxa a ser estimada no modelo.

Seja $X(t)$ o processo latente e $Y(t)$ o processo observado definidos para todos os tempos $t \in \mathbb{R}$. Como estamos particularmente interessados no valor obtido do processo em cada tempo de observação, escrevemos $X_n = X(t_n)$. Com isto, definimos que os estados do processo latente e do processo observado são representados por $X_{0:N} = (X_0, \dots, X_N)$ e $Y_{1:N} = (Y_1, \dots, Y_N)$, respectivamente.

No caso particular do modelo SIRS, os estados ocultos gerados pelo processo latente são o número de indivíduos em cada compartimento $X_n = (S_n, I_n, R_n)$ nos momentos $n \in \{0, \dots, N\}$ e o processo observado é a coleção do número de casos relatados Y_n nos instantes $n \in \{1, \dots, N\}$. Neste trabalho, a taxa de reporte da doença foi chamada de ρ , assume valores em $[0, 1]$ e foi tratada como o parâmetro do processo de medição do modelo parcialmente observado SIRS. Com isso, a distribuição do processo observado Y dado o processo latente X ($Y|X$) foi definida como uma distribuição binomial com número de repetições correspondendo ao número de infectados no processo latente e com probabilidade de sucesso (notificação) ρ .

Um POMP é caracterizado por duas propriedades, que, no caso em que X_n e Y_n são variáveis aleatórias discretas, podem ser escritas como:

- O processo latente, X_n , é Markoviano de primeira ordem e, dado o seu passado, é independente do passado do processo observado. Esta condição pode ser expressa como

$$\begin{aligned} \mathbb{P}[X_n = k | X_0 = x_0, \dots, X_{n-1} = x_{n-1}, Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}] \\ = \mathbb{P}[X_n = k | X_{n-1} = x_{n-1}]. \end{aligned}$$

- A observação Y_n é condicionalmente independente de todas as outras variáveis dado X_n , isto é,

$$\mathbb{P}[Y_n = k | X_0 = x_0, \dots, X_n = x_n, Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}] = \mathbb{P}[Y_n = k | X_n = x_n],$$

para todo $n \in \{1, \dots, N\}$. No caso em que X_n e Y_n são contínuas, propriedades análogas são válidas. A Figura 2.5 ilustra a relação entre o processo latente X_n , em vermelho, e o processo observado Y_n , em azul.

As setas horizontais representam as transições condicionais do processo latente e os compartimentos ligados por elas representam os estados do processo latente nos tempos $n \in \{0, \dots, N\}$. As setas verticais que conectam os estados dos processos X_n e Y_n representam o processo de medição. As observações Y_1, \dots, Y_N resultantes do processo de medição são os estados do processo observado nos tempos $n \in \{1, \dots, N\}$. Percebe-se que não há

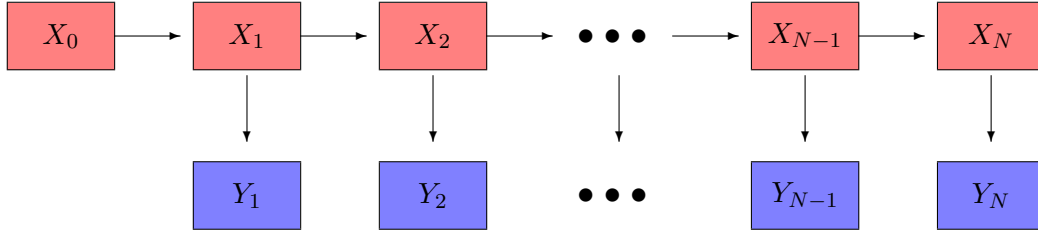


Figura 2.5: Fluxograma de um modelo POMP.

observações emitidas pelo estado oculto X_0 pois este depende apenas de sua densidade¹ f_{X_0} , que corresponde ao estado inicial do processo latente que será inferido. Quando não há escolha cientificamente motivada de f_{X_0} , o valor de X_0 pode ser tratado como um parâmetro do modelo a ser estimado.

Matematicamente, cada modelo é determinado por uma distribuição de probabilidade. Com base na notação empregada em King et al. (2016), definimos que o modelo do processo latente é determinado pela densidade de transição $f_{X_n|X_{n-1}}(x_n|x_{n-1}; \theta)$ e pela densidade inicial $f_{X_0}(x_0; \theta)$, onde θ é o vetor de parâmetros do modelo, a ser estimado. O processo observado é determinado pela densidade do processo de medição $f_{Y_n|X_n}(y_n|x_n; \theta)$. Essas duas densidades determinam o modelo POMP completo, ou seja, a densidade conjunta $f_{X_{0:N}, Y_{1:N}}(x_{0:N}, y_{1:N}; \theta)$, que pode ser escrita como

$$f_{X_{0:N}, Y_{1:N}}(x_{0:N}, y_{1:N}; \theta) = f_{X_0}(x_0; \theta) \prod_{n=1}^N f_{X_n|X_{n-1}}(x_n|x_{n-1}; \theta) f_{Y_n|X_n}(y_n|x_n; \theta).$$

2.6 Função de verossimilhança

A função de verossimilhança é uma das bases principais para inferência frequentista e bayesiana e, heurísticamente, representa a informação sobre os parâmetros do modelo contida nos dados (Pawitan, 2001). Um modelo probabilístico é representado por uma função de densidade, denotada por $f_{Y_{1:N}}(y_{1:N}; \theta)$, que define a distribuição de probabilidade do vetor $Y_{1:N} = (Y_1, \dots, Y_N)$ para cada valor do vetor de parâmetros θ . Para realizar inferência, buscamos os valores de θ que melhor explicam $y_{1:N}$ como uma observação aleatória de $f_{Y_{1:N}}(y_{1:N}; \theta)$. A função de verossimilhança é dada por

$$\mathcal{L}(\theta) = f_{Y_{1:N}}(y_{1:N}; \theta),$$

onde $\mathcal{L}(\theta)$ é uma função de θ , ou seja, para cada particular valor do parâmetro, digamos $\theta = \theta_1$, a função de verossimilhança retorna a densidade conjunta do modelo em θ_1 , levando em conta os valores observados $y_{1:N}$. Na estimação por máxima verossimilhança buscamos o valor de θ que maximiza $\mathcal{L}(\theta)$. Por razões analíticas e computacionais, muitas vezes é mais conveniente trabalhar com o logaritmo da função de verossimilhança, isto é, $l(\theta) = \log(\mathcal{L}(\theta))$.

A função de verossimilhança para um modelo POMP pode ser obtida integrando-se a den-

¹Por questão de simplicidade de apresentação, utilizaremos sempre a expressão densidade e a notação matemática para variáveis aleatórias contínuas na exposição teórica. Entretanto, deve ser compreendido que as variáveis X_n e Y_n por princípio podem ser tanto contínuas quanto discretas. No caso de variáveis discretas, essa notação deve ser compreendida como função massa, e as devidas alterações (como substituição de integrais por somatórios) devem ser feitas.

sidade conjunta $f_{X_{0:N}, Y_{1:N}}(x_{0:N}, y_{1:N}; \theta)$ em relação a $x_{0:N}$, resultando em

$$\begin{aligned} \mathcal{L}(\theta) &= f_{Y_{1:N}}(y_{1:N}; \theta) = \int_{\mathcal{X}} f_{X_{0:N}, Y_{1:N}}(x_{0:N}, y_{1:N}; \theta) dx_{0:N} \\ &= \int_{\mathcal{X}} f_{X_0}(x_0; \theta) \prod_{n=1}^N f_{X_n|X_{n-1}}(x_n|x_{n-1}; \theta) f_{Y_n|X_n}(y_n|x_n; \theta) dx_{0:N}, \end{aligned} \quad (2.2)$$

onde \mathcal{X} denota a imagem de $X_{0:N}$. Note que a integral (2.2) é calculada na imagem de $X_{0:N}$, cuja alta dimensionalidade dificulta, ou até mesmo impossibilita, seu cálculo. Uma forma de contornar este problema é utilizando filtro de partículas, detalhado na Seção 2.7.

2.7 Filtro de partículas

Conforme visto na Seção 2.6, integrar a distribuição conjunta do processo observado e do processo latente para obter a função de verossimilhança pode ser uma tarefa inactível por conta da alta dimensionalidade dos dados. Uma forma de obter uma estimativa não viesada da função de verossimilhança é através do filtro de partículas, que utiliza técnicas de Monte Carlo sequencial. A estimação da função de verossimilhança se dá de forma sequencial, recursiva, utilizando fórmulas de predição e filtragem, que serão apresentadas nesta seção.

Simulações de Monte Carlo são comumente utilizadas para a estimação de integrais de alta dimensionalidade impossíveis de calcular analiticamente, como no caso de (2.2). Para ver como podemos utilizar este método para aproximar a integral, observe que, partindo de (2.2), podemos escrever $\mathcal{L}(\theta)$ como uma esperança,

$$\mathcal{L}(\theta) = \int_{\mathcal{X}} \prod_{n=1}^N f_{Y_n|X_n}(y_n|x_n; \theta) f_{X_{0:N}}(x_{0:N}; \theta) dx_{0:N} = \mathbb{E} \left(\prod_{n=1}^N f_{Y_n|X_n}(y_n|X_n; \theta) \right),$$

onde a esperança acima é calculada em relação à distribuição de $X_{0:N}$. O próximo passo é gerar J amostras independentes de realizações de $f_{Y_n|X_n}(\cdot|\cdot, \theta)$ a partir das quais utilizamos a lei dos grande números para aproximar $\mathcal{L}(\theta)$ como

$$\mathcal{L}(\theta) \approx \frac{1}{J} \sum_{j=1}^J \prod_{n=1}^N f_{Y_n|X_n}(y_n|x_n^j; \theta), \quad (2.3)$$

onde x_n^j é a n -ésima observação da j -ésima amostra da distribuição do processo $X_{0:N}$.

Apesar da equação (2.3) parecer ser uma boa solução para estimar $\mathcal{L}(\theta)$, novamente encontramos problemas. Um deles é que, para J pequeno, se muitas trajetórias simuladas divergirem dos dados, estas prejudicarão a estimação da função de verossimilhança. O outro problema é que, para J grande (necessário para obter boas estimativas), o custo computacional de se obter realizações do trajeto J vezes é enorme.

Uma outra abordagem para o problema é a utilização do filtro de partículas. A ideia heurística do filtro de partículas é que a cada passo do algoritmo, as trajetórias com baixa probabilidade de ocorrência sejam “filtradas” impedindo essas de influenciarem na estimação.

Essa abordagem se baseia em uma fatoração alternativa de $\mathcal{L}(\theta)$, a saber

$$\begin{aligned}\mathcal{L}(\theta) &= f_{Y_{1:N}}(y_{1:N}; \theta) = \prod_{n=1}^N f_{Y_n|Y_{1:(n-1)}}(y_n|y_{1:(n-1)}; \theta) \\ &= \prod_{n=1}^N \int_{\mathcal{X}_n} f_{Y_n|X_n}(y_n|x_n; \theta) f_{X_n|Y_{1:(n-1)}}(x_n|y_{1:(n-1)}; \theta) dx_n,\end{aligned}$$

onde \mathcal{X}_n denota a imagem de X_n . Por convenção escrevemos, para $n = 1$, $f_{X_1|Y_0} = f_{X_1}$. Obtemos uma fórmula de predição, que é dada por

$$f_{X_n|Y_{1:(n-1)}}(x_n|y_{1:(n-1)}; \theta) = \int_{\mathcal{X}_{n-1}} f_{X_n|X_{n-1}}(x_n|x_{n-1}; \theta) f_{X_{n-1}|Y_{1:(n-1)}}(x_{n-1}|y_{1:(n-1)}; \theta) dx_{n-1},$$

e, pelo teorema de Bayes, construímos a fórmula de filtragem

$$\begin{aligned}f_{X_n|Y_{1:n}}(x_n|y_{1:n}; \theta) &= f_{X_n|Y_n, Y_{1:(n-1)}}(x_n|y_n, y_{1:(n-1)}; \theta) \\ &= \frac{f_{Y_n|X_n}(y_n|x_n; \theta) f_{X_n|Y_{1:(n-1)}}(x_n|y_{1:(n-1)}; \theta)}{\int_{\mathcal{X}_n} f_{Y_n|X_n}(y_n|u_n; \theta) f_{X_n|Y_{1:(n-1)}}(u_n|y_{1:(n-1)}; \theta) du_n}.\end{aligned}$$

Com isso, podemos calcular iterativamente as distribuições de predição $f_{X_n|Y_{1:(n-1)}}(x_n|y_{1:(n-1)})$ e de filtragem $f_{X_n|Y_{1:n}}(x_n|y_{1:n})$ para cada instante de tempo n . Essas duas fórmulas geram uma recursão de dois passos:

- A fórmula de predição gera a distribuição de predição no instante n , utilizando a distribuição de filtragem obtida em $n - 1$.
- A fórmula de filtragem gera a distribuição de filtragem no instante n , utilizando a distribuição da predição em n .

Para exemplificar o funcionamento do filtro de partículas, suponha que $X_{n-1,j}^F$, para $j \in \{1, \dots, J\}$, representam as J amostras da distribuição de filtragem no instante de tempo $n - 1$. No primeiro passo obtemos uma amostra da distribuição de predição $X_{n,j}^P$ no instante de tempo n simulando,

$$X_{n,j}^P \sim (X_{n-1,j}^F; \theta),$$

para cada $j \in \{1, \dots, J\}$, utilizando a distribuição de transição do processo latente $f_{X_n|X_{n-1}}(x_n|x_{n-1}; \theta)$. No segundo passo, obtemos uma coleção de observações da distribuição de filtragem no instante de tempo n reamostrando de $\{X_{n,j}^P; j \in \{1, \dots, J\}\}$, com pesos $w_{n,j} = f_{Y_n|X_n}(y_n|X_{n,j}^P; \theta)$. Os pesos $w_{n,j}$ são definidos pela probabilidade de observar y_1, \dots, y_n dado a distribuição de predição. Esta operação é repetida até a última observação, onde $n = N$. A função de verossimilhança condicional em n , dado todas as observações anteriores pode ser escrita como

$$\mathcal{L}_n(\theta) = f_{Y_n|Y_{1:(n-1)}}(y_n|y_{1:(n-1)}; \theta) = \int_{\mathcal{X}_n} f_{Y_n|X_n}(y_n|x_n; \theta) f_{X_n|Y_{1:(n-1)}}(x_n|y_{1:(n-1)}; \theta) dx_n,$$

e é aproximada por

$$\hat{\mathcal{L}}_n(\theta) = \frac{1}{J} \sum_{j=1}^J f_{Y_n|X_n}(y_n|X_{n,j}^P; \theta),$$

já que os $X_{n,j}^P$, como construídos acima, representam amostras de $f_{X_n|Y_{1:(n-1)}}(x_n|y_{1:(n-1)}; \theta)$. A log-verossimilhança é aproximada por

$$l(\theta) = \log(\mathcal{L}(\theta)) = \sum_{n=1}^N \log(\mathcal{L}_n(\theta)) \approx \sum_{n=1}^N \log(\hat{\mathcal{L}}_n(\theta)) = \hat{l}(\theta).$$

Para uma abordagem mais detalhada sobre métodos de Monte Carlo e filtro de partículas, veja [Arulampalam et al. \(2002\)](#) e [Doucet et al. \(2001\)](#).

2.8 Filtragem iterada

A filtragem iterada é um método que permite estimar a máxima verossimilhança para processos de Markov parcialmente observados fazendo uso iterado do filtro de partículas, adicionando a este perturbações nos parâmetros. Enquanto o algoritmo do filtro de partículas estima o valor de verossimilhança para um conjunto de parâmetros fixos, a filtragem iterada faz inferência sobre os parâmetros do modelo. Ao final das iterações, espera-se obter a estimativa de máxima verossimilhança aplicando no filtro de partículas o conjunto de parâmetros estimados na última iteração da filtragem iterada.

O algoritmo de filtragem iterada inicia em um ponto inicial fornecido pelo usuário e tenta se mover em direção ao ponto de máximo mais próximo que puder encontrar. O movimento a partir da estimativa inicial é provocado por perturbações estocásticas nos parâmetros desconhecidos para que seus espaços sejam mais explorados. Estas perturbações no espaço de parâmetros suavizam a superfície de probabilidade, permitem que a busca dos parâmetros escape de máximos locais e ajudam a superar as dificuldades numéricas que possam surgir durante o Monte Carlo sequencial ([Ionides et al., 2011](#)).

Para resolver o problema de máximos locais e tentar encontrar o máximo global é aconselhável executar o algoritmo iniciando em diferentes pontos iniciais. Espera-se que os resultados do algoritmo sejam as estimativas dos parâmetros que maximizam a verossimilhança de forma global e também forneçam uma compreensão geral do perfil da função de verossimilhança. Cada iteração da filtragem iterada consiste em uma execução do filtro de partículas com a adição de perturbação nos parâmetros e, ao final de uma iteração, a coleção de valores dos parâmetros estimados pelo filtro de partículas é reciclada como parâmetros iniciais para a próxima iteração.

O pseudocódigo do algoritmo de filtragem iterada é apresentado no Algoritmo 1. As linhas 4–10 (laço em n) representam o algoritmo de filtro de partículas com perturbações estocásticas aplicadas nos parâmetros. O laço em m representa as repetições do filtro de partícula com perturbações sucessivamente diminuídas, para que as partículas não se afastem da região de máxima verossimilhança.

Para iniciar o algoritmo é necessário determinar o número de partículas a ser utilizado no filtro de partículas (J) e um conjunto de valores iniciais para o vetor de parâmetros, Θ_j^0 , $j \in \{1, \dots, J\}$ (para cada j haverá uma coleção de parâmetros). Com esses valores iniciais, determinamos a densidade inicial de X_0 , a densidade condicional de transição do processo latente $f_{X_n|X_{n-1}}(x_n|x_{n-1}; \theta)$ e do processo de medição dado o processo latente, $f_{Y_n|X_n}(y_n|x_n; \theta)$, que gera o processo observado $y_{1:N}$. Deve-se ainda definir o número de iterações M , a função densidade de perturbação inicial dos parâmetros h_0 e o vetor de escala de perturbação $\sigma_{1:M}$, que é inteiramente definido antes do algoritmo iniciar e é composto por valores de perturbação decrescente. Dessa forma, a cada laço M , ocorre o “resfriamento” das perturbações dos parâmetros. A função densidade h_n é a atualização de h_0 durante o laço em n , conforme o novo conjunto

Algorithm 1 Pseudocódigo do algoritmo da Filtragem Iterada

Require: $f_{X_0}(\cdot; \theta)$; $f_{X_n|X_{n-1}}(\cdot|x_{n-1}; \theta)$; $f_{Y_n|X_n}(\cdot|x_n; \theta)$; $y_{1:N}$; M ; J ; $\Theta_j^0, j = 1, \dots, J$ and

h_0 ; $\sigma_{1:M}$.

- 1: **for** m in $\{1, \dots, M\}$ **do**
 - 2: $\Theta_{0,j}^{F,m} \sim h_0(\cdot | \Theta_j^{m-1}; \sigma_m)$, for $j \in \{1, \dots, J\}$;
 - 3: $X_{0,j}^{F,m} \sim f_{X_0}(x_0; \Theta_{0,j}^{F,m})$, for $j \in \{1, \dots, J\}$;
 - 4: **for** n in $\{1, \dots, N\}$ **do**
 - 5: $\Theta_{n,j}^{P,m} \sim h_n(\cdot | \Theta_{n-1,j}^{F,m}; \sigma_m)$, for $j \in \{1, \dots, J\}$;
 - 6: $X_{n,j}^{P,m} \sim f_{X_n|X_{n-1}}(\cdot | X_{n-1}^{F,m}; \Theta_{n,j}^{P,m})$, for $j \in \{1, \dots, J\}$;
 - 7: $w_{n,j}^m = f_{Y_n|X_n}(y_n | X_{n,j}^{P,m}; \Theta_{n,j}^{P,m})$, for $j \in \{1, \dots, J\}$;
 - 8: Draw $k_{1:J}$ with $\mathbb{P}[k_j = i] = \frac{w_{n,i}^m}{\sum_{u=1}^J w_{n,u}^m}$;
 - 9: $\Theta_{n,j}^{F,m} = \Theta_{n,k_j}^{P,m}$ and $X_{n,j}^{F,m} = X_{n,k_j}^{P,m}$, for $j \in \{1, \dots, J\}$;
 - 10: **end for**
 - 11: Set $\Theta_j^m = \Theta_{N,j}^{F,m}$, for $j \in \{1, \dots, J\}$;
 - 12: **end for**
-

de valores para o vetor de parâmetros do modelo definido pela probabilidade de observação do processo $y_{1:N}$.

No laço em n , J amostras Monte Carlo são selecionadas sequencialmente para estimar a verossimilhança através das distribuições de filtragem $X_{n,j}^F$ e predição $X_{n,j}^P$. A cada passo do algoritmo, os pesos $w_{n,j}$ são reajustados com base na probabilidade de observação do processo $y_{1:N}$ considerando a densidade da fórmula de predição. Com base nos novos pesos, a densidade de filtragem é atualizada e assim o algoritmo sucede até o final da série $y_{1:N}$ (isto é, até $n = N$). Ao final do laço em n , o laço em m é incrementado em uma unidade, e a intensidade da perturbação dos parâmetros para o próximo laço em n é diminuída.

Os parâmetros estimados ao final da filtragem iterada serão utilizados como pontos de partida para outra busca com o filtro de partículas. O resultado final da filtragem iterada será o vetor Θ_j^M , para $j \in \{1, \dots, J\}$, que contém as estimativas dos parâmetros após M iterações do filtro de partícula de tamanho J . Para mais detalhes sobre o método de filtragem iterada, veja [lonides et al. \(2011\)](#).

CAPÍTULO 3

O BANCO DE DADOS

Uma forma de acompanhar a dinâmica de uma doença é através do seu número de casos reportados a sistemas de vigilância epidemiológica. Apesar da sub-notificação, o reporte de casos positivos é uma informação relevante quando se deseja detectar surtos, estimar a incidência, avaliar a disseminação e o progresso da doença. Com base nas inferências obtidas através dos reportes, pode-se planejar medidas de prevenção e controle para tentar conter o avanço dos casos.

Dessa forma, os dados utilizados para as análises deste estudo foram retirados do site do INFOGripe (info.gripe.fiocruz.br), que é uma iniciativa para monitorar os casos reportados de Síndrome Respiratória Aguda Grave (SRAG) no Sistema de Informação de Agravos de Notificação (SINAN), do ministério da saúde. Além disso, no site é possível encontrar níveis de alerta para a população, que podem ser obtidos por estados, por capitais ou por macrorregiões de saúde. O conteúdo apresentado no site do INFOGripe é o produto da parceria entre pesquisadores do Programa de Computação Científica da Fundação Oswaldo Cruz (Fiocruz, PROCC), da Escola de Matemática Aplicada da Fundação Getúlio Vargas (FGV, EMAp), no Rio de Janeiro, e do GT-Influenza da Secretaria de Vigilância Sanitária do Ministério da Saúde (GT-Influenza, SVS, MS).

No banco de dados extraído, cada linha representa uma semana epidemiológica e o número de casos pode ser estratificado por: tipo de localidade (país, estado ou região); tipo de SRAG (causado por influenza ou COVID-19); tipo de escala (incidência por 100 mil habitantes ou número absoluto de casos); situação do dado (estável ou incompleto); sexo (masculino, feminino ou ignorado); semana do ano de primeiros sintomas; faixa etária (< 2 anos, 2–4 anos, 0–4 anos, 5–9 anos, 10–19 anos, 20–29 anos, 30–39 anos, 40–49 anos, 50–59 anos ou > 60 anos); situação do caso (testes positivos, testes negativos, casos aguardando resultado, casos sem informação laboratorial, casos sem teste laboratorial); tipo de vírus (Influenza A, Influenza B, SARS-CoV-2, Vírus sincicial respiratório (VSR), Parainfluenza 1, Parainfluenza 2, Parainfluenza 3 ou Adenovírus). Para obtermos o banco final para o presente estudo, utilizamos os seguintes filtros:

- Tipo de localidade: Rio Grande do Sul;
- Tipo de SRAG: SRAG por influenza;
- Tipo de escala: Número absoluto de casos;
- Situação do dado: Estável;

- Ano da semana de primeiros sintomas: Entre a primeira semana de 2016 até última semana de 2017;
- Situação do caso: Testes positivos;
- Tipo de vírus: Influenza A.

Os filtros por sexo e idade não foram utilizados por não serem considerados importantes para o tipo de análise feita neste trabalho. O filtro por semana do ano de primeiros sintomas não foi utilizado porque todas as semanas são importantes para obter os dados completos em série temporal. Ao aplicar todos os filtros, restaram 104 observações no banco, cada uma delas representando o número de casos positivos de influenza A por semana epidemiológica dos anos 2016 e 2017 para o estado do Rio Grande do Sul. A análise descritiva dos dados extraídos está contida na tabela 3.1. A série de dados completa extraída entre os anos de 2009 e 2020 é apresentada na Figura 3.1, enquanto que a série entre os anos de 2016 e 2017 é apresentada Figura 3.2.

Tabela 3.1: Análise descritiva dos casos positivos reportados de influenza A extraídos do INFOgripe.

Ano	Casos	Média	Mediana	Mínimo	Máximo	1º Quartil	3º Quartil
2009	1409,00	27,10	0,00	0,00	331,00	0,00	8,75
2010	2,00	0,04	0,00	0,00	1,00	0,00	0,00
2011	182,00	3,50	1,00	0,00	35,00	0,00	2,00
2012	501,00	9,63	0,00	0,00	85,00	0,00	5,25
2013	386,00	7,42	1,00	0,00	49,00	0,00	9,00
2014	143,00	2,70	0,00	0,00	20,00	0,00	2,00
2015	50,00	0,96	0,00	0,00	5,00	0,00	1,25
2016	1011,00	19,44	3,00	0,00	122,00	1,00	16,25
2017	260,00	5,00	0,00	0,00	32,00	0,00	4,00
2018	422,00	8,12	1,00	0,00	50,00	0,00	10,25
2019	332,00	6,38	2,00	0,00	27,00	0,00	12,00
2020	5,00	0,20	0,00	0,00	2,00	0,00	0,00
Total geral	4703,00	7,86	0,00	0,00	331,00	0,00	4,00

Segundo Marcelo Gomes, atual coordenador da plataforma INFOgripe e pesquisador da Fiocruz, o início da coleta dos casos de SRAG iniciou em 2009 por conta da epidemia de influenza A H1N1. Desde então, a plataforma faz a vigilância epidemiológica de doenças respiratórias, tendo como foco principal a circulação sazonal de vírus influenza e a detecção oportuna de novos subtipos com alta patogenicidade, como novas variantes de influenza aviária ou suína com transmissão em humanos, ou de novos vírus respiratórios (portal.fiocruz.br).

Apesar da série temporal iniciar em 2009 e ser atualizada semanalmente, o período escolhido para a análise neste estudo foi a partir da primeira semana de 2016 até a última semana de 2017, totalizando 104 semanas epidemiológicas. A escolha de trabalhar com apenas 2 anos da série foi feita baseada em características do modelo SIRS (ver capítulo 6) e no fato de que este trabalho tem caráter exploratório com enfoque em estudar a metodologia para ajuste do modelo. Com isso, a escolha dos 2 anos a serem modelados foi feita pois 2016 foi um ano com um pico importante de casos reportados seguido de um ano com comportamento regular (comportamento descrito pelo modelo). O ano de 2009 (maior pico da série), por outro lado, foi seguido pelo ano de 2010 quase sem casos relatados devido à imunização em massa da população (89 milhões de pessoas) após a pandemia do influenza A H1N1 em 2009 ([Domingues](#)

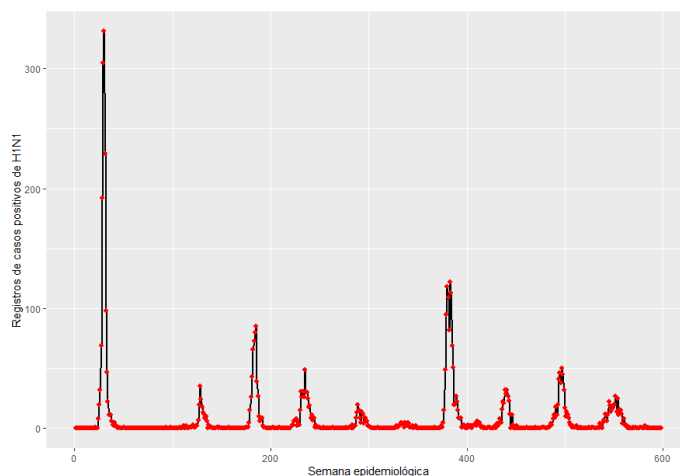


Figura 3.1: Número de casos positivos de influenza A registrados no Rio Grande do Sul entre 2009 e 2020.

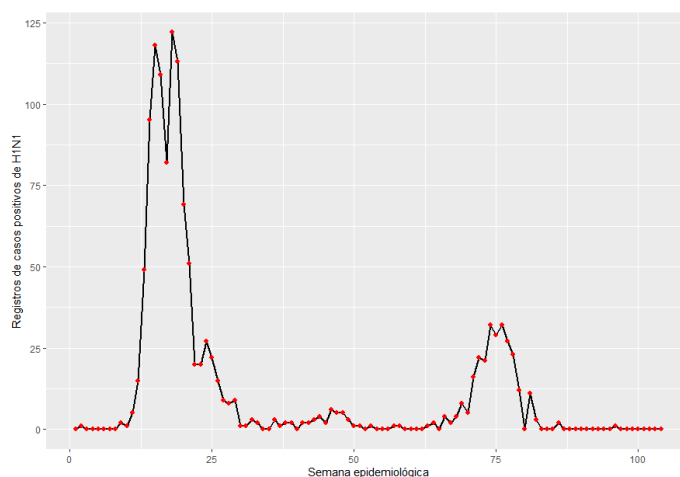


Figura 3.2: Número de casos positivos de influenza A registrados no Rio Grande do Sul entre 2016 e 2017.

et al., 2012). Além disso, o processo de transmissão do vírus é diferente em cada estado, pois depende de muitas variáveis exógenas à doença, como por exemplo, o clima, políticas públicas de prevenção, perfil socio-econômico da população. Por essa razão, escolhemos analisar somente os dados do estado do Rio Grande do Sul.

O modelo compartimental SIRS foi utilizado para modelar a série temporal restrita ao período de 2016 e 2017. Como será visto no Capítulo 4, os parâmetros a serem estimados são: β , γ , α e ρ . Na literatura, os valores reportados para o parâmetro $\mu_i = 1/\gamma$ estão entre 2 e 7 dias, como descrito em Shaman and Karspeck (2012). O valor de R_0 (relembrando que R_0 pode ser obtido por β/γ) foi estimado entre 1,2 e 2,3 em Boëlle et al. (2011). Os valores para os parâmetros α e ρ não estão estabelecidos na literatura. A partir das referências encontradas, podemos comparar os achados deste estudo apresentados no Capítulo 5 com as características já conhecidas da infecção por influenza.

CAPÍTULO 4

O PACOTE POMP

O `pomp` (King et al., 2021) é um pacote desenvolvido para o software R (R Core Team, 2020) por Aaron A. King, Dao Nguyen e Edward L. Ionides, da Universidade de Michigan e fornece um conjunto de ferramentas especialmente concebido para a análise de processos de Markov parcialmente observados não lineares. Com o `pomp`, pode-se implementar um modelo especificando seu processo não observado e os componentes de medição desse processo. Além disso, com uma gama variada de métodos implementados, pode-se simular, analisar, ajustar, inferir e avaliar os modelos especificados (King et al., 2016).

O pacote `pomp` aceita contribuições na forma de códigos, exemplos, melhorias na documentação, solicitações de recursos e de ajuda pelo site github.com/kingaa. Novos usuários do pacote podem buscar material de apoio no link github.com/gettingstarted onde há informações sobre como instalá-lo adequadamente, como simular um POMP e como utilizar as ferramentas de análise disponíveis no pacote. Além disso, com o intuito de melhorar o desempenho computacional ao realizar as estimativas de parâmetros e outras inferências, este tutorial demonstra como codificar um modelo usando C snippets.

As funções do pacote `pomp` utilizadas neste trabalho foram `pomp`, `pfilter` e `mif2`. A função `pomp` simula um objeto POMP a partir dos argumentos passados para a função, são elas: as distribuições condicionais que definem o POMP (o processo latente $f_{X_n|X_{n-1}}(x_n|x_{n-1}; \theta)$ e o processo de medição $f_{Y_n|X_n}(y_n|x_n; \theta)$); a série temporal dos dados observados ($y_{1:N}$, no caso da aplicação deste trabalho representados pelos casos positivos reportados para influenza); os nomes das variáveis do processo latente (no caso do modelo SIRS visto na Seção 2.3: $X_n = (S_n, I_n, R_n)$ são representados por S , I e R no código); os valores iniciais dos estados do processo latente ($S = P$, $I = 1$ e $R = 0$, que corresponde a toda a população suscetível no instante 0, e 1 novo indivíduo adicional que introduz a doença); os parâmetros do modelo $\theta = (\beta, \gamma, \alpha, \rho)$, conforme as Seções 2.3 e 2.5); os valores iniciais dos parâmetros (ver na Seção 4.1); o esqueleto do POMP (o sistema de equações que definem o modelo, conforme a Seção 2.3); as transformações utilizadas para os parâmetros.

Dentre os vários métodos de inferência disponíveis no pacote `pomp`, para a análise apresentada neste trabalho, utilizamos o algoritmo de filtro de partículas, que está implementado na função `pfilter` e o algoritmo de filtragem iterada, implementado na função `mif2`. Os argumentos necessários para a função `pfilter` são apenas o objeto POMP retornado pela função `pomp` e o número de partículas (J), ou seja, o tamanho da amostra de Monte Carlo. Como discutido na Seção 2.7, o resultado do filtro de partículas será a estimativa da função de verossimilhança para um determinado conjunto de parâmetros fixos usando a técnica de Monte Carlo Sequencial.

Para o algoritmo da filtragem iterada, os argumentos necessários para utilizar a função `mif2` são o objeto POMP, o vetor Θ_0 que contém os valores iniciais dos parâmetros a serem estimados, o número de vezes que o filtro de partículas será iterado (M) e replicado (J) e o esquema de perturbação dos parâmetros (h_0 , como visto na Seção 2.8). A filtragem iterada retornará, ao final de M iterações, qual o conjunto de parâmetros, dentre os explorados, que resulta o maior valor encontrado para a função de verossimilhança.

4.1 Implementação do modelo SIRS no pacote `pomp`

O objetivo desta seção é apresentar a ligação entre as metodologias de análise para um POMP e as funcionalidade do pacote `pomp`, apresentadas ao longo do capítulo 2 e no início deste capítulo, respectivamente. Para isso, o código usado para fazer a implementação do modelo é apresentado e comentado. Por conta do desempenho computacional, parte do código foi escrito em linguagem C, via `Csnippet`. Além disso, foi necessário paralelizar o código com o pacote `doParallel` do R (Microsoft Corporation and Weston, 2020).

O primeiro passo para utilizar as funcionalidade do pacote `pomp` é definir o sistema de equações que regem o comportamento da série. Como estamos interessados na dinâmica do modelo SIRS, já apresentado na Seção 2.3, e também em adicionar estocasticidade ao modelo definimos o objeto `sirs_step` que contém o que é chamado de “esqueleto” do modelo. O `sirs_step` contém a informação de como a estrutura probabilística do processo latente progride passo a passo em uma população fechada.

```
sirs_step <- Csnippet("
  double dN_SI = rbinom(S, 1 - exp(-beta*I/N*dt));
  double dN_IR = rbinom(I, 1 - exp(-gamma*dt));
  double dN_RS = rbinom(R, 1 - exp(-alpha*dt));
  S -= dN_SI - dN_RS;
  I += dN_SI - dN_IR;
  R += dN_IR - dN_RS;
")
```

Os compartimentos do modelo são definidos como S (suscetível), I (infectado) e R (recuperado). A direção do fluxo de indivíduos entre compartimentos acontece de forma semelhante ao discutido na Seção 2.3, entretanto adotamos um modelo estocástico para a transição entre compartimentos. Assim, ao invés das tradicionais equações diferenciais a intensidade do fluxo entre compartimentos é definida por variáveis com distribuição binomial. Portanto `dN_SI`, `dN_IR` e `dN_RS` representam os indivíduos saindo do compartimento S e ingressando no I com probabilidade $1 - \exp(-\beta I/N)$, saindo de I e ingressando no R com probabilidade $1 - \exp(-\gamma)$ e saindo de R e retornando ao S com probabilidade $1 - \exp(-\alpha)$, respectivamente.

Após a implementação do esqueleto do modelo SIRS, definimos os estados iniciais do modelo, onde assumimos que toda a população (P) está suscetível no início do processo, e com isso, o número inicial de recuperados é zero. A quantidade de infectados, por sua vez, deve ser pequena no estado inicial ($I = 1$), para reproduzir a invasão de um vírus em população suscetível. O objeto `sirs_rinit` contém a informação da quantidade de indivíduos da população em cada compartimento. O valor para P foi definido em 10.693.929 com base no total da população do Rio Grande do Sul no último censo de 2010 (cidades.ibge.gov.br).

```
sirs_rinit <- Csnippet("S = P; I = 1; R = 0;")
```

Além disso, o modelo observado foi definido como uma distribuição binomial com número de repetições correspondendo ao número de infectados no processo latente I e probabilidade de notificação rho.

```
sirs_dmeas <- Csnippet("lik = dbinom(cases, I, rho, give_log=TRUE);")
```

```
sirs_rmeas <- Csnippet("cases = rbinom(I, rho);")
```

Com os objetos construídos acima, temos todos os argumentos necessários para a construção do objeto pomp.

```
pomp(
  data=flu,
  times = "week", t0 = 0,
  rprocess = euler(sirs_step, delta.t = 1),
  rinit = sirs_rinit,
  rmeasure = sirs_rmeas,
  dmeasure = sirs_dmeas,
  statenames = c("S", "I", "R"),
  paramnames = c("beta", "gamma", "alpha", "rho", "P")
) -> SIRSflu
```

O objeto SIRSflu guarda as informações do POMP definido e, a partir dele, pode-se estimar o valor da verossimilhança para qualquer conjunto fixo de parâmetros. Para isso, utiliza-se o filtro de partículas, implementado na função *pfilter*. Definimos o tamanho da amostra de Monte Carlo (*J*) em 5.000 e os valores dos parâmetros do modelo definidos no argumento *paramnames* da função *pomp*. A escolha dos parâmetros foi arbitrária, com exceção do P, já definido anteriormente.

```
SIRSflu %>%
  pfilter(Np = 5000,
    params = c(beta = 1.85,
               gamma = 1.55,
               alpha = 1/20,
               rho = 0.001,
               P = 10693929)) -> pfrick
```

O objeto *pfrick* contém o valor estimado da verossimilhança para os valores de parâmetros definidos em *params*. Como o filtro de partículas é um algoritmo de Monte Carlo, obtemos apenas uma estimativa ruidosa da verossimilhança. Ao aumentar o número de partículas no algoritmo, reduz-se o ruído da estimativa. Pode-se também estimar a magnitude do erro de Monte Carlo executando alguns filtros de partículas independentes.

```
replicate(10, pfrick %>% pfilter() %>% logLik()) -> loglik
logmeanexp(loglik, se = TRUE) -> loglik_se
```


O objeto `loglik` é um vetor que armazena os valores das estimativas da verossimilhança para as 10 replicações independentes do filtro de partículas criadas no exemplo, enquanto o objeto `loglik.se` contém a estimativa do erro de Monte Carlo baseado nas 10 estimativas contidas no vetor `loglik`.

Passamos agora ao foco principal do trabalho: a maximização da função de verossimilhança pela filtragem iterada. Para executar o algoritmo é necessário informar o objeto `pomp` que contém todas as informações do POMP (`SIRSflu`); os valores iniciais dos parâmetros (`params`); o número de amostras Monte Carlo desejadas (5.000); o número de iterações do algoritmo (300); o tipo de transformação dos parâmetros (log e logit); o nome dos parâmetros do modelo (`beta`, `gamma`, `alpha`, `rho`, `P`); a intensidade de perturbação dos parâmetros (0,02); o resfriamento das perturbações (0,5). O resfriamento das perturbações é definido no argumento `cooling.fraction.50`, que determina que a cada 50 iterações, a perturbação dos parâmetros tem sua magnitude reduzida pela metade. Note que as transformações dos parâmetros utilizadas levam os parâmetros transformados em domínios restritos à reta.

```
SIRSflu %>%
  mif2(
    params = params,
    Np = 5000,
    Nmif = 300,
    partrans = parameter_trans(
      log = c("beta", "gamma", "alpha"),
      logit = c("rho")
    ),
    paramnames = c("beta", "gamma", "alpha", "rho", "P"),
    rw.sd = rw.sd(beta = 0.02,
                  gamma = 0.02,
                  alpha = 0.02,
                  rho = 0.02
    ),
    cooling.fraction.50 = 0.5
  ) -> mf
```

O objeto `mf` contém a trajetória das 300 iterações da filtragem iterada. Como simulamos o algoritmo para um único conjunto de parâmetros (`params`), a busca pela máxima verossimilhança ocorrerá de forma local. Como o objetivo deste trabalho é realizar a busca pela máxima verossimilhança de forma global, definimos 500 conjuntos de valores iniciais de parâmetros com a função `sobol_design`, na qual definimos o espaço onde cada parâmetro será explorado. No argumento `lower` informamos os valores mínimos da busca e em `upper`, os valores máximos. Aqui, os valores foram escolhidos com base na literatura, conforme detalhado no Capítulo 3.

```
sobol_design(
  lower = c(beta = 1,
            gamma = 1,
            alpha = 1/60,
            rho = 0.00005,
            P = 10693929),
  upper = c(beta = 10,
            gamma = 1/0.15,
```

```
      alpha = 1/40,  
      rho = 0.0002,  
      P = 10693929),  
  nseq = 500  
) -> guesses
```

Para iniciar a busca global, podemos reciclar as informações contidas em `mf` e apenas substituir `params` pelos valores contidos no objeto `guesses` obtido acima. Para realizar a paralelização das simulações, a função `foreach` foi utilizada.

```
foreach (guess = iter(guesses,"row"),  
        .combine = c, .packages = c("pomp"),  
        .errorhandling = "remove", .inorder = FALSE) %dopar% {  
  mf %>% mif2(params = guess)  
} -> mifs
```

O objeto `mifs` contém as trajetórias das 300 iterações da filtragem iterada para os 500 conjuntos de valores iniciais de parâmetros.

CAPÍTULO 5

RESULTADOS

Este capítulo descreve os resultados das análises feitas com o pacote `pomp` para realizar inferência sobre os dados reportados de influenza A nos anos de 2016 e 2017 no estado do Rio Grande do Sul. Em todas as simulações foram utilizadas 5.000 partículas para o filtro de partículas e 300 iterações do algoritmo de filtragem iterada. Os resultados estão organizados em tabelas que apresentam, de maneira ordenada, as combinações de parâmetros que obtiveram os melhores valores de log-verossimilhança de acordo com a parametrização utilizada, quando aplicável.

Os valores estimados dos parâmetros β , γ , α , definidos na Seção 2.3, e ρ , definido na seção 2.5, o valor de $l(\theta)$, onde θ é o vetor de parâmetros $\theta = (\beta, \gamma, \alpha, \rho)$, e o seu erro padrão $sd(l(\theta))$ são apresentados em todas as tabelas. Para facilitar a interpretação das estimativas, as três últimas colunas das tabelas foram calculadas a partir dos valores de β , γ e α . Os valores de R_0 , μ_i e μ_r representam o número reprodutivo básico, o período médio de infecção em dias e o período médio de perda de imunidade em dias, respectivamente. O parâmetro ρ representa a proporção de casos de infecção por influenza A que são reportados.

Devido ao custo computacional alto para realizar a otimização da verossimilhança e à aparente correlação entre os parâmetros nas iterações do filtro iterado, conjecturamos que uma parametrização alternativa poderia aumentar a eficiência da estimação. Assim, propusemos uma outra parametrização para o modelo e a comparamos com a parametrização original. A parametrização 1, que estima β , γ , α e ρ de maneira direta, é chamada de P1 nas tabelas. A parametrização alternativa, chamada de P2, ajusta o modelo equivalente introduzindo dois novos parâmetros q e k , de modo que os parâmetros γ e α da Parametrização 1 são obtidos através das transformações $\gamma = \beta \times q$ e $\alpha = \beta \times k$. Os parâmetros q e k assumem apenas valores maiores que zero ($q > 0$ e $k > 0$).

5.1 Estimação para o modelo completo

Nesta primeira análise apresentamos os resultados da estimação pelo filtro iterado para o modelo completo, estimando todos os parâmetros livremente, considerando as duas parametrizações. A Tabela 5.1, apresenta os conjuntos de parâmetros estimados com os maiores valores de log-verossimilhança. Para esta análise, o algoritmo da filtragem iterada iniciou em 500 pontos de partida distintos para ambas parametrizações.

Note que a parametrização alternativa, mesmo considerando valores maiores de $sd(l(\theta))$, se mostrou vantajosa em relação aos valores de $l(\theta)$, aparentando representar um máximo global para a função de verossimilhança pelo fato das melhores estimativas listadas serem consistentes,

resultando em valores muito próximos para $l(\theta)$ e para todos os parâmetros estimados. A Figura 5.1 apresenta alguns caminhos simulados do modelo (em vermelho) com o conjunto de parâmetros que resultou no maior valor de $l(\theta)$ e compara com a série temporal de casos reportados (em azul).

Tabela 5.1: Melhores estimativas baseado no valor de log-verossimilhança.

	β	γ	$\alpha (\times 10^{-2})$	$\rho (\times 10^{-5})$	$l(\theta)$	$sd(l(\theta))$	R_0	μ_i	μ_r
P1	1,64	0,35	1,10	2,67	-428,44	0,13	4,57	19,52	634,82
	1,59	0,38	1,19	2,82	-428,62	0,17	4,13	18,12	588,80
	1,62	0,35	1,11	2,60	-428,69	0,17	4,59	19,74	632,11
	1,63	0,36	1,12	2,63	-429,05	0,22	4,52	19,41	625,37
	1,71	0,33	1,03	2,56	-429,48	0,18	5,03	20,59	680,05
P2	1,63	0,36	1,12	2,63	-426,47	1,22	4,51	19,38	622,68
	1,61	0,37	1,14	2,73	-426,56	1,70	4,37	18,89	611,60
	1,60	0,35	1,12	2,61	-426,86	1,50	4,54	19,77	624,19
	1,65	0,36	1,10	2,67	-427,50	0,86	4,61	19,48	636,23
	1,60	0,38	1,19	2,85	-427,53	0,27	4,18	18,33	589,86

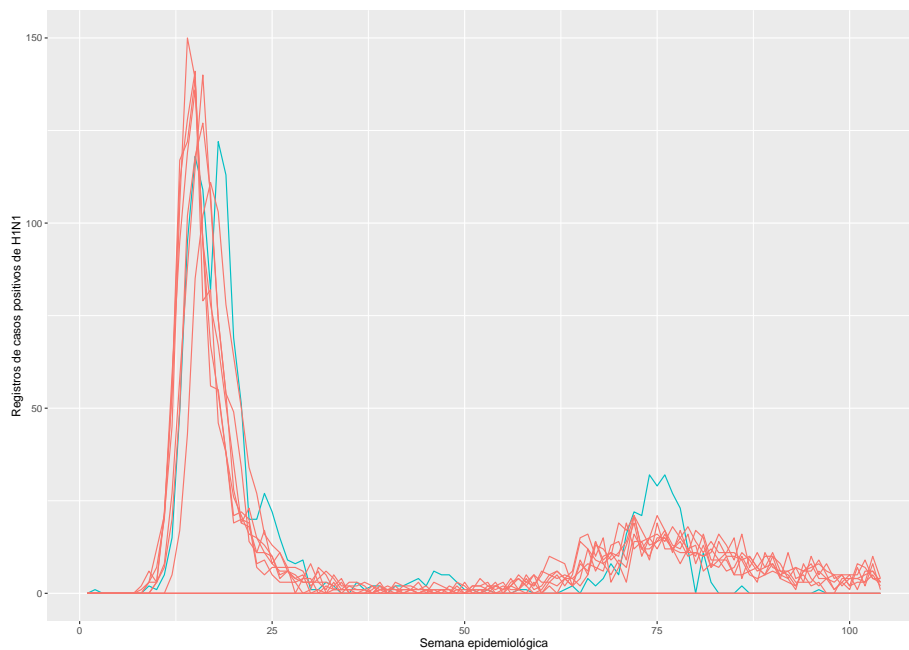


Figura 5.1: Caminhos simulados do modelo com o maior valor de $l(\theta)$ na parametrização alternativa.

Em ambas parametrizações, as estimativas de R_0 e μ_i não são compatíveis com os valores de referência discutidos no Capítulo 3. Uma possível explicação para resultados tão díspares em relação aos valores da literatura é que a otimização irrestrita da função de verossimilhança talvez não seja objeto de tanto interesse epidemiológico e outras especificações do modelo que restrinjam o espaço de busca podem acabar trazendo ganhos na interpretação biológica dos parâmetros.

A Figura 5.2 apresenta o gráfico de dispersão dos valores dos parâmetros junto com os valores da log-verossimilhança obtidos utilizando a parametrização 1. Cada ponto no gráfico representa os valores dos parâmetros na última iteração da filtragem iterada (há 500 pontos

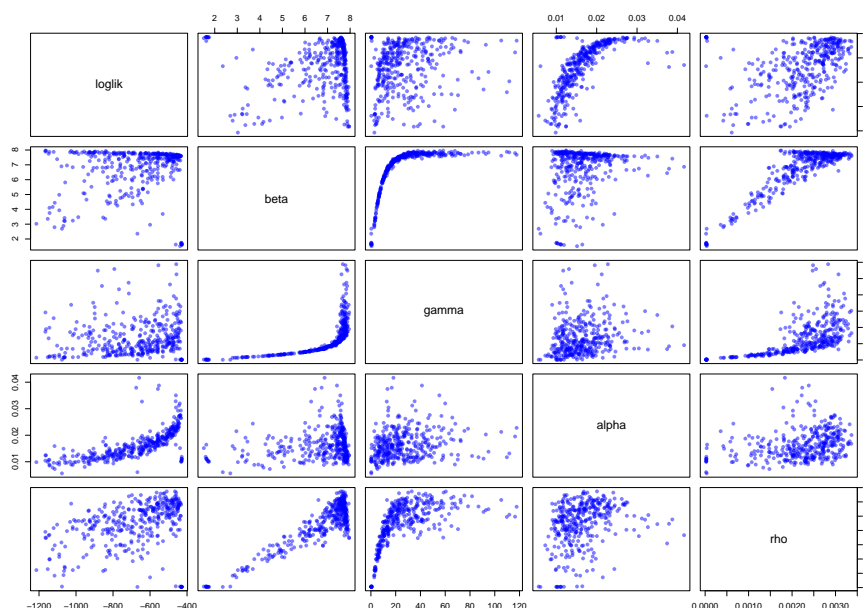


Figura 5.2: Gráficos de dispersão dos parâmetros e log-verossimilhança estimados com a parametrização 1.

pois iniciamos o algoritmo em 500 pontos distintos). Os gráficos podem ser utilizados como uma ferramenta para analisar a correlação entre os parâmetros.

Percebe-se uma clara relação entre os parâmetros β e γ enquanto o gráfico de dispersão entre β e α parece revelar um comportamento aproximadamente linear. Por esse motivo, a parametrização alternativa foi proposta de maneira que os parâmetros γ e α fossem obtidos de β , via $\gamma = \beta \times q$ e $\alpha = \beta \times k$. Os gráficos de dispersão entre os parâmetros e a log-verossimilhança obtidos a partir da nova parametrização são apresentados na Figura 5.3. Apesar de q e β na parametrização alternativa seguirem fortemente ligados, a relação entre os outros parâmetros parece ter sido reduzida.

O parâmetro R_0 é uma das quantidades fundamentais para a determinação da dinâmica em modelos SIRS. Para entender o seu efeito na estimação dos outros parâmetros e tendo em vista os valores grandes (relativos ao reportado na literatura) obtidos na análise reportada na Tabela 5.1, onde R_0 foi estimado junto com os demais parâmetros de forma livre, uma segunda análise foi conduzida. Nesta análise consideramos a estimação dos parâmetros como no cenário anterior, porém restringindo R_0 ao valores tipicamente reportados na literatura, a saber $1,2 < R_0 < 2,3$ (Boëlle et al., 2011). As melhores estimativas dessa seleção estão apresentados na Tabela 5.2. Neste caso também há um ganho em relação ao valor de $l(\theta)$ ao utilizar a parametrização alternativa. Note que, quando restringimos o domínio de R_0 , os valores de μ_i ficam próximos aos valores tipicamente reportados na literatura (Shaman and Karspeck, 2012).

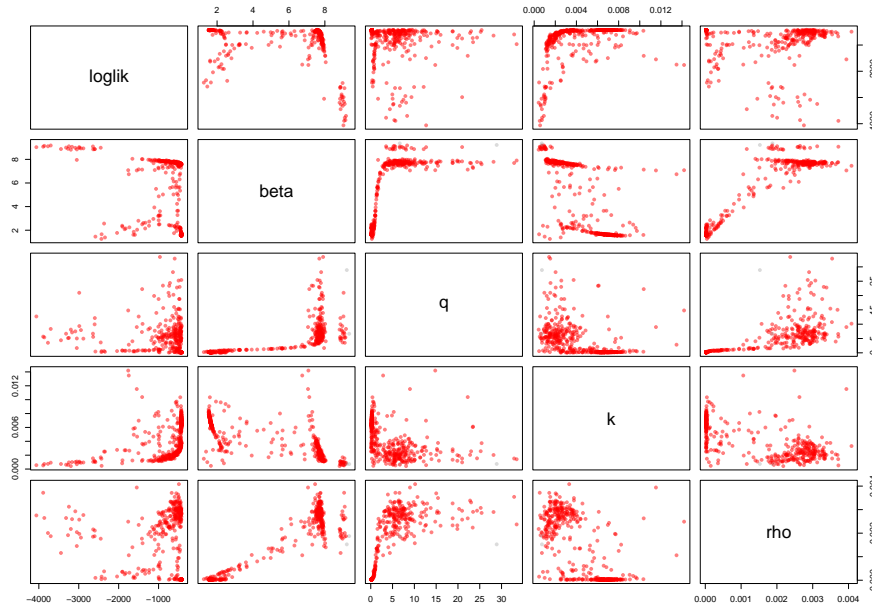


Figura 5.3: Gráficos de dispersão dos parâmetros e log-verossimilhança estimados com a parametrização alternativa.

Tabela 5.2: Melhores estimativas baseado no valor de log-verossimilhança com a restrição $1,2 < R_0 < 2,3$.

	β	γ	$\alpha (\times 10^{-2})$	$\rho (\times 10^{-4})$	$l(\theta)$	$sd(l(\theta))$	R_0	μ_i	μ_r
P1	2,96	2,67	2,15	3,57	-599,87	5,80	1,11	2,62	325,43
	3,20	3,15	1,18	6,90	-1060,18	2,26	1,02	2,21	589,83
	3,22	3,19	1,23	5,83	-1064,65	1,75	1,01	2,19	568,02
	2,88	2,80	1,49	4,37	-1067,63	9,42	1,03	2,49	469,62
	2,79	2,66	1,59	4,17	-1116,16	13,43	1,05	2,63	439,96
P2	2,65	2,20	2,39	2,29	-522,19	5,20	1,20	3,17	292,52
	3,28	3,21	2,39	4,40	-526,10	4,01	1,02	2,17	292,28
	1,93	0,97	1,88	7,41	-548,65	0,88	1,98	7,16	371,18
	2,47	1,86	2,31	1,80	-548,89	1,57	1,32	3,75	301,95
	2,20	1,45	2,11	1,23	-570,46	3,54	1,51	4,80	330,41

5.2 Estimação com valores fixos para a taxa de detecção ρ

Na otimização irrestrita, apresentada na Tabela 5.1, encontramos valores de $l(\theta)$ que parecem representar um máximo global para a função de verossimilhança. Apesar disso, os valores obtidos para R_0 e γ estão bastante altos relativamente aos valores reportados na literatura, o que pode ser consequência de valores subestimados para ρ (10^{-5}). Esta relação fica clara quando comparamos os resultados da Tabela 5.1 com a Tabela 5.2, a qual apresenta resultados bem próximos aos valores reportados na literatura. Neste caso, a estimativa de ρ é da ordem de 10^{-4} .

Com o objetivo de investigar o comportamento dos parâmetros para diferentes magnitudes dos valores de ρ , realizamos três análises fixando este parâmetro em $1,5 \times 10^{-3}$, $1,5 \times 10^{-4}$ e $1,5 \times 10^{-5}$. Os resultados desta seção foram obtidos utilizando 250 pontos de partida para a filtragem iterada. No primeiro bloco da Tabela 5.3, onde $\rho = 1,5 \times 10^{-3}$, todas as estimativas para R_0 ficaram abaixo de 1. Levando em consideração a interpretação epidemiológica desse parâmetro, esses valores de R_0 não caracterizam a invasão da doença na população. Os valores de μ_i também estão abaixo de 1, o que representa um período médio de infecção de menos de 1 dia. Ambas estimativas estão subestimadas se considerarmos sua interpretação epidemiológica e valores de referência para influenza. Uma possível explicação para os parâmetros subestimados pode ser valores altos de ρ .

Na sequência, uma outra análise com um valor menor de ρ foi realizada. Neste caso ρ foi fixado em $1,5 \times 10^{-4}$ e os resultados são apresentados no segundo bloco da Tabela 5.3. Neste caso, os valores de R_0 ficaram próximos às estimativas encontradas na literatura (Boëlle et al., 2011). Os valores de μ_i e μ_r também estão mais próximos dos valores tipicamente encontrados na literatura (Shaman and Karspeck, 2012), e os valores da log-verossimilhança foram semelhantes para ambas as parametrizações.

Para completar nossa investigação relativa aos valores de ρ , consideramos ainda o caso em que o valor de ρ é ainda menor, fixado em $\rho = 1,5 \times 10^{-5}$. Os resultados deste cenário podem ser encontrados no terceiro bloco da Tabela 5.3. Notamos que neste caso as estimativas obtidas estão acima dos valores de referência epidemiológica para R_0 e μ_i , podendo ser um indício de que a verdadeira taxa de reporte ρ assume valores em escala maior que 10^{-5} .

Comparando os três cenários para ρ fixo, observamos que as estimativas mais próximas dos valores de referência epidemiológica foram obtidas fixando-se $\rho = 1,5 \times 10^{-4}$. Por outro lado, as estimativas com maior log-verossimilhança foram obtidas para $\rho = 1,5 \times 10^{-3}$. As tabelas da Seção 5.1 mostram padrão análogo, com estimativas dos valores dos parâmetros mais próximas dos valores de referência epidemiológica quando ρ está na casa de 10^{-4} .

É intuitivamente esperado que a verdadeira taxa de notificação deva estar associada a um modelo epidemiologicamente motivado. Baseado nos resultados obtidos, conjecturamos que o valor de ρ esteja na ordem de 10^{-4} , considerando que o modelo utilizado de fato reflita a realidade. Nas análises restritas conduzidas, quando a taxa de notificação ρ é tratada como conhecida, a parametrização 1 se mostrou superior, rendendo valores de log-verossimilhança sensivelmente maiores que os competidores. A Figura 5.4 apresenta alguns caminhos simulados do modelo com os parâmetros que resultaram no maior valor de $l(\theta)$ da parametrização 1 com $\rho = 1,5 \times 10^{-4}$, comparado com os dados observados. Percebe-se que os caminhos simulados apresentam defasagem de algumas semanas em relação aos dados observados, e esta característica pode ser consequência da escolha fixa da quantidade inicial de indivíduos da população em cada compartimento ($S = P$, $I = 1$ e $R = 0$) para todas as simulações.

Tabela 5.3: Melhores estimativas baseado no valor de log-verossimilhança.

$\rho = 1,5 \times 10^{-3}$	β	γ	$\alpha (\times 10^{-2})$	$l(\theta)$	$sd(l(\theta))$	R_0	μ_i	μ_r
P1	5,91	9,75	2,89	-457,56	2,35	0,61	0,71	241,95
	5,70	9,24	2,90	-458,58	6,98	0,62	0,75	241,11
	5,90	10,03	3,03	-464,85	1,41	0,59	0,69	230,48
	5,98	10,01	2,73	-465,29	2,49	0,59	0,69	255,48
	5,95	10,10	2,91	-465,46	1,44	0,58	0,69	239,75
P2	5,76	9,27	2,68	-463,35	1,79	0,62	0,75	260,97
	5,63	8,61	2,14	-528,69	7,65	0,65	0,81	326,97
	5,31	7,73	2,02	-573,98	2,17	0,68	0,90	346,20
	5,57	8,42	1,94	-589,98	6,26	0,66	0,83	360,60
	5,45	8,12	1,80	-622,22	0,81	0,67	0,86	388,25
$\rho = 1,5 \times 10^{-4}$	β	γ	$\alpha (\times 10^{-2})$	$l(\theta)$	$sd(l(\theta))$	R_0	μ_i	μ_r
P1	2,12	1,55	2,42	-492,65	1,88	1,37	4,51	288,61
	2,10	1,50	2,37	-496,60	1,30	1,39	4,64	294,37
	2,15	1,55	2,37	-496,97	0,82	1,38	4,50	295,19
	2,18	1,57	2,37	-498,44	1,40	1,38	4,43	294,38
	2,11	1,49	2,35	-499,15	1,49	1,41	4,67	297,27
P2	2,16	1,54	2,40	-494,72	1,79	1,39	4,52	291,28
	2,19	1,60	2,42	-496,65	1,06	1,36	4,36	288,49
	2,21	1,63	2,45	-497,86	0,60	1,35	4,27	284,95
	2,19	1,61	2,41	-498,19	1,07	1,36	4,33	289,38
	2,21	1,64	2,41	-499,35	1,85	1,34	4,25	289,61
$\rho = 1,5 \times 10^{-5}$	β	γ	$\alpha (\times 10^{-3})$	$l(\theta)$	$sd(l(\theta))$	R_0	μ_i	μ_r
P1	1,65	0,24	8,97	-501,33	2,01	6,98	29,64	780,55
	1,62	0,24	9,23	-501,96	1,69	6,81	29,52	758,29
	1,61	0,23	9,16	-502,38	1,83	6,91	30,05	763,91
	1,58	0,24	9,57	-502,86	1,35	6,59	29,13	731,32
	1,63	0,24	9,30	-503,07	0,82	6,84	29,32	752,47
P2	1,65	0,24	9,23	-502,64	1,76	6,88	29,21	758,29
	1,60	0,24	9,37	-503,21	1,19	6,68	29,26	746,73
	1,66	0,23	8,82	-503,38	0,89	7,16	30,17	793,49
	1,62	0,24	9,20	-503,39	0,63	6,80	29,34	760,45
	1,61	0,24	9,38	-503,59	0,97	6,69	29,07	746,30

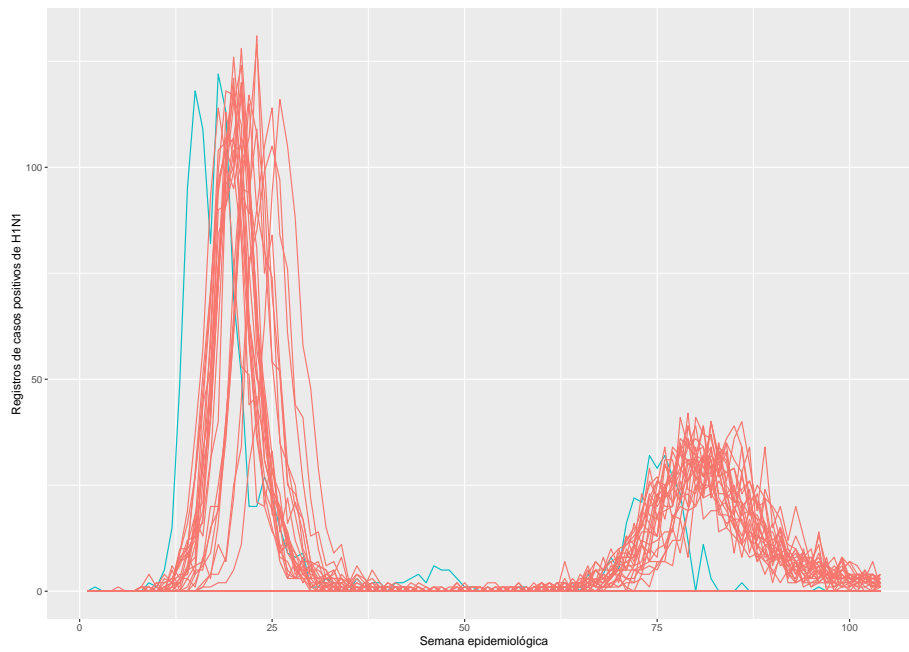


Figura 5.4: Caminhos simulados do modelo com o maior valor de $l(\theta)$ na parametrização 1, considerando $\rho = 1,5 \times 10^{-4}$.

5.3 Estimação com valores fixos para a taxa de recuperação γ

Nesta seção apresentamos uma análise análoga à da seção anterior, porém fixando a taxa de recuperação γ ao invés de ρ . Como a parametrização alternativa substituiu γ por $\beta \times q$, apresentamos aqui apenas os resultados da parametrização 1. Os valores fixos escolhidos para γ se basearam em valores reportados na literatura, a saber, $\gamma \in \{3,5; 2,33; 1,75; 1,4; 1,167\}$. Considerando o modelo determinístico, estes valores implicam períodos infecciosos médios de $\mu_i \in \{2, 3, 4, 5, 6\}$ dias, respectivamente.

A Tabela 5.4 apresenta os valores estimados dos parâmetros que resultaram os maiores valores de $l(\theta)$ para cada valor de γ fixo. Para $\gamma = 3,5$, correspondendo a $\mu_i = 2$ dias, o valor de R_0 é menor que 1, o que indicaria a tendência da infecção não invadir a população. Para μ_i fixado em 3, 4, 5 e 6 dias, as estimativas de R_0 foram condizentes com valores da literatura. A escala de ρ foi estimada em $\approx 10^{-4}$ em todos os cenários, variando basicamente entre 10^{-4} e 5×10^{-4} conforme γ cresce. Além disso, os valores de $l(\theta)$ foram semelhantes em todos os cenários e, ignorando os resultados para $\gamma = 3,5$ cujos valores de $R_0 < 1$, o cenário $\gamma = 1,4$ ($\mu_i = 5$ dias) apresenta o maior valor da log-verossimilhança. Nesse caso, ρ foi estimado em $12,62 \times 10^{-5}$, indicando que aproximadamente 0,0126% dos casos da doença na população seriam reportados. A Figura 5.5 mostra a série temporal correspondente à alguns caminhos relativos aos parâmetros que resultaram no maior valor de $l(\theta)$ reportado na Tabela 5.4, desconsiderando o caso em que $\gamma = 3,5$.

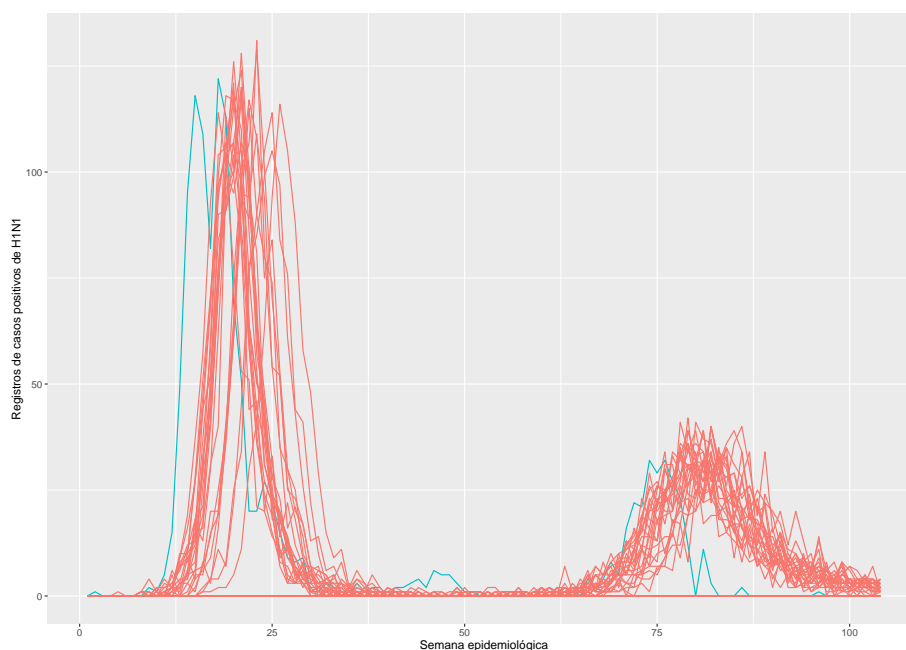


Figura 5.5: Caminhos simulados do modelo considerando $\gamma = 1,4$ na parametrização 1.

Tabela 5.4: Melhores estimativas baseado no valor da log-verossimilhança para γ fixo.

	β	$\alpha (\times 10^{-2})$	$\rho (\times 10^{-5})$	$l(\theta)$	$sd(l(\theta))$	R_0	μ_i	μ_r
$\gamma = 3,5$	3,44	2,56	47,66	-499,56	4,32	0,98	2,00	273,11
	3,41	2,55	47,47	-504,11	0,57	0,97	2,00	274,24
	3,44	2,56	46,12	-504,31	0,61	0,98	2,00	272,52
	3,43	2,34	50,04	-539,74	1,31	0,98	2,00	298,03
	3,44	2,25	50,47	-555,42	10,36	0,98	2,00	310,07
$\gamma = 2,33$	2,73	2,48	25,12	-517,48	1,28	1,17	3,00	281,99
	2,69	2,41	28,61	-521,59	1,23	1,15	3,00	289,73
	2,74	2,47	24,05	-525,25	3,41	1,17	3,00	283,09
	2,71	2,39	27,34	-528,06	0,39	1,16	3,00	292,31
	2,76	2,41	24,21	-536,93	0,89	1,18	3,00	290,44
$\gamma = 1,75$	2,32	2,38	17,03	-511,45	1,49	1,33	4,00	294,68
	2,33	2,40	16,28	-513,25	3,40	1,33	4,00	291,55
	2,32	2,40	16,08	-515,68	0,94	1,32	4,00	291,66
	2,34	2,40	16,12	-515,77	3,53	1,34	4,00	291,42
	2,34	2,38	15,87	-523,10	2,65	1,34	4,00	294,40
$\gamma = 1,4$	2,08	2,29	12,62	-504,70	1,93	1,49	5,00	305,66
	2,10	2,31	11,50	-521,71	2,47	1,50	5,00	303,19
	2,10	2,24	12,23	-521,76	1,11	1,50	5,00	312,43
	2,09	2,27	10,96	-530,98	3,64	1,49	5,00	308,48
	2,18	2,15	10,79	-574,65	4,53	1,56	5,00	325,50
$\gamma = 1,167$	1,91	2,14	9,57	-506,76	1,06	1,64	6,00	325,80
	1,93	2,15	9,58	-507,28	1,67	1,65	6,00	324,13
	1,91	2,16	9,03	-510,28	0,53	1,64	6,00	323,26
	1,94	2,18	9,18	-511,63	1,53	1,67	6,00	320,84
	1,93	2,16	9,18	-512,22	0,87	1,66	6,00	323,29

CAPÍTULO 6

CONCLUSÕES E TRABALHOS FUTUROS

Nesse trabalho estudamos a modelagem por meio de um modelo POMP baseado no modelo SIRS para o vírus influenza A. Para tanto, primeiro apresentamos a motivação para o estudo da dinâmica deste patógeno e a plataforma INFOgripe que faz a vigilância dos casos de SRAG no Brasil e foi a fonte dos dados desse trabalho. Também trouxemos uma introdução aos modelos compartimentais, explicando mais detalhadamente os modelos determinísticos SIR e SIRS. Adicionalmente, como fundamentação metodológica, foram apresentados os modelos POMP, assim como o funcionamento do filtro de partículas para a estimação da função de verossimilhança e a filtragem iterada para a busca dos parâmetros que maximizam a função de verossimilhança. Além disso, apresentamos o pacote `pomp` e descrevemos as funções utilizadas para realizar as análises apresentadas no trabalho.

O principal foco desta dissertação foi realizar um estudo metodológico sobre o ajuste de um POMP com estrutura do modelo SIRS estocástico. Avaliações preliminares mostraram que o modelo SIRS é capaz de reproduzir a estrutura de múltiplos picos encontrada na curva dos casos de influenza. Entretanto, após um ano mais forte de introdução da doença, esses picos diminuem gradativamente em intensidade tendendo a desaparecer em um estado de endemia com número de infectados constante. Assim o modelo foi capaz de se ajustar bem a dados de dois anos, com um primeiro ano de elevado número de casos e um segundo ano com pico menos pronunciado, mas não à série completa. Foi com base nessas observações que selecionamos uma janela de observações adequada ao presente estudo.

Com base nesse conjunto de dados restrito foi possível estudar o ajuste do modelo por meio da filtragem iterada e avaliar a relação entre os parâmetros. Notamos que, sem qualquer conhecimento prévio da função de verossimilhança, os esforços computacionais necessários para estimação paramétrica são grandes. Motivada por um possível aumento na eficiência da estimação, uma parametrização alternativa foi proposta e resultou em valores maiores para a log-verossimilhança, aparentando ser um ponto de máximo global da função. Em relação ao tempo computacional para realizar as estimativas, as duas parametrizações obtiveram performance semelhante para números comparáveis de partículas e replicações. Para as simulações considerando todos os parâmetros livres, ambas parametrizações levaram aproximadamente 60 horas para serem executadas. Uma análise para identificar quantas iterações e partículas até convergência ao ótimo não foi realizada.

Em outra análise com foco na interpretabilidade dos parâmetros, ainda com o parâmetro ρ livre, e selecionando apenas as estimativas com valores de R_0 restritos a intervalos compatíveis com a literatura, vê-se que, novamente, a parametrização alternativa foi superior em relação à otimização da verossimilhança. Com menores valores de R_0 , os parâmetros μ_i e μ_r também

diminuíram, ficando dentro de uma faixa de valores esperados para infecções por influenza. O valor de ρ , por sua vez, mudou sua magnitude de 10^{-5} para 10^{-4} .

Para investigar o comportamento dos parâmetros para diferentes escalas de ρ , fixamos este parâmetro em 3 valores: $1,5 \times 10^{-3}$, $1,5 \times 10^{-4}$ e $1,5 \times 10^{-5}$. Considerando os cenários avaliados, a parametrização 1 apresentou os melhores valores de verossimilhança. Para $\rho = 1,5 \times 10^{-4}$, os valores para R_0 , μ_i e μ_r estão próximos com o que já se conhece sobre a infecção do influenza. Além disso, fizemos uma análise análoga para diversos valores de γ fixados compatíveis com a literatura. O cenário em que $\gamma = 1,4$ ($\mu_i = 5$ dias) apresenta maior log-verossimilhança dentre as estimativas cujos parâmetros estão condizentes com a literatura e ρ foi estimado na escala 10^{-4} .

Foi possível utilizar as ferramentas de inferência com sucesso para inferir parâmetros do modelo. A busca pelo valor de máxima verossimilhança global aparentemente obteve sucesso com a parametrização alternativa, apesar dos valores dos parâmetros estimados não estarem próximos aos reportados na literatura. Estimativas dos parâmetros para $\rho = 1,5 \times 10^{-4}$ geraram valores de R_0 e μ_i compatíveis com a literatura. Quando fixamos o parâmetro $\gamma \in \{2,33; 1,75; 1,4; 1,167\}$ também obtemos valores de $\rho = 1,5 \times 10^{-4}$, podendo ser um indício de que a taxa de reporte da doença está realmente na escala 10^{-4} .

Esse estudo deve ser compreendido como um passo intermediário na modelagem do vírus influenza A no estado do Rio Grande do Sul nos anos de 2016 e 2017. A etapa subsequente teria como objetivo captar a estrutura completa de vários anos da série, com seus picos anuais. Para tanto, pode-se considerar a introdução de covariáveis para as taxas de infecção β ou perda da imunidade α . Uma outra abordagem possível é considerar um modelo dinâmico para os parâmetros do modelo permitindo que um ou mais variem com o tempo de forma adaptativa.

REFERÊNCIAS BIBLIOGRÁFICAS

- Althaus, C. L. (2014). Estimating the reproduction number of ebola virus (ebov) during the 2014 outbreak in west africa. *PLoS currents*, 6.
- Arulampalam, M. S., Maskell, S., Gordon, N., and Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188.
- Bjørnstad, O. N., Finkenstädt, B. F., and Grenfell, B. T. (2002). Dynamics of measles epidemics: estimating scaling of transmission rates using a time series SIR model. *Ecological monographs*, 72(2):169–184.
- Boëlle, P.-Y., Ansart, S., Cori, A., and Valleron, A.-J. (2011). Transmission parameters of the A/H1N1 (2009) influenza virus pandemic: a review. *Influenza and other respiratory viruses*, 5(5):306–316.
- Brauer, F., Castillo-Chavez, C., and Feng, Z. (2019). *Mathematical models in epidemiology*, volume 32. Springer.
- Chen, Y., Yang, J., and Zhang, F. (2014). The global stability of an sirs model with infection age. *Mathematical Biosciences & Engineering*, 11(3):449.
- Delamater, P. L., Street, E. J., Leslie, T. F., Yang, Y. T., and Jacobsen, K. H. (2019). Complexity of the basic reproduction number (R_0). *Emerging Infectious Diseases*, 25(1):1.
- Diekmann, O., Heesterbeek, J. A. P., and Metz, J. A. (1990). On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28(4):365–382.
- Domingues, C. M. A. S., de Oliveira, W. K., Team, B. P. I. V. E., et al. (2012). Uptake of pandemic influenza (H1N1)-2009 vaccines in Brazil, 2010. *Vaccine*, 30(32):4744–4751.
- Doucet, A., De Freitas, N., and Gordon, N. J. (2001). *Sequential Monte Carlo methods in practice*. Springer.
- Eccles, R. (2005). Understanding the symptoms of the common cold and influenza. *The Lancet Infectious Diseases*, 5(11):718–725.
- Forleo-Neto, E., Halker, E., Santos, V. J., Paiva, T. M., and Toniolo-Neto, J. (2003). Influenza. *Revista da Sociedade Brasileira de Medicina Tropical*, 36(2):267–274.
- Graunt, J. (1939). *Natural and political observations made upon the bills of mortality*. Johns Hopkins Press.

- Guerra, F. M., Bolotin, S., Lim, G., Heffernan, J., Deeks, S. L., Li, Y., and Crowcroft, N. S. (2017). The basic reproduction number (r_0) of measles: a systematic review. *The Lancet Infectious Diseases*, 17(12):e420–e428.
- Ionides, E. L., Bhadra, A., Atchadé, Y., and King, A. (2011). Iterated filtering. *The Annals of Statistics*, 39(3):1776–1802.
- Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, containing papers of a mathematical and physical character*, 115(772):700–721.
- King, A. A., Ionides, E. L., Bretó, C. M., Ellner, S. P., Ferrari, M. J., Funk, S., Johnson, S. G., Kendall, B. E., Lavine, M., Nguyen, D., O’Dea, E. B., Reuman, D. C., Wearing, H., and Wood, S. N. (2021). *pomp: Statistical Inference for Partially Observed Markov Processes*. R package, version 3.6.
- King, A. A., Nguyen, D., and Ionides, E. L. (2016). Statistical inference for partially observed Markov processes via the R package pomp. *Journal of Statistical Software, Articles*, 69(12):1–43.
- Korobeinikov, A. (2004). Lyapunov functions and global properties for SEIR and SEIS epidemic models. *Mathematical medicine and biology: a journal of the IMA*, 21(2):75–83.
- Kudryashov, N. A., Chmykhov, M. A., and Vigdorowitsch, M. (2021). Analytical features of the SIR model and their applications to COVID-19. *Applied Mathematical Modelling*, 90:466–473.
- Microsoft Corporation and Weston, S. (2020). *doParallel: Foreach Parallel Adaptor for the ‘parallel’ Package*. R package version 1.0.16.
- Mohd, M. H. and Sulayman, F. (2020). Unravelling the myths of r_0 in controlling the dynamics of COVID-19 outbreak: A modelling perspective. *Chaos, Solitons & Fractals*, 138:109943.
- Neher, R. A. and Bedford, T. (2015). nextflu: Real-time tracking of seasonal influenza virus evolution in humans. *Bioinformatics*, 31(21):3546–3548.
- Pawitan, Y. (2001). *In all likelihood: statistical modelling and inference using likelihood*. Oxford University Press.
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Shaman, J. and Karspeck, A. (2012). Forecasting seasonal outbreaks of influenza. *Proceedings of the National Academy of Sciences*, 109(50):20425–20430.
- Stöhr, K. (2002). Influenza — WHO cares. *The Lancet Infectious Diseases*, 2(9):517.
- Sun, C. and Yang, W. (2010). Global results for an SIRS model with vaccination and isolation. *Nonlinear Analysis: Real World Applications*, 11(5):4223–4237.
- Taubenberger, J. K. and Morens, D. M. (2008). The pathology of influenza virus infections. *Annual Review of Pathology: Mechanisms of Disease*, 3:499–522.
- Weidemann, F., Dehnert, M., Koch, J., Wichmann, O., and Höhle, M. (2014). Bayesian parameter inference for dynamic infectious disease modelling: rotavirus in Germany. *Statistics in Medicine*, 33(9):1580–1599.

-
- Zhang, W. (2021). World Health Organization. Influenza: are we ready?
- Zhao, S., Lin, Q., Ran, J., Musa, S. S., Yang, G., Wang, W., Lou, Y., Gao, D., Yang, L., He, D., et al. (2020). Preliminary estimation of the basic reproduction number of novel coronavirus (2019-ncov) in china, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *International journal of infectious diseases*, 92:214–217.
- Zhou, Y. and Liu, H. (2003). Stability of periodic solutions for an SIS model with pulse vaccination. *Mathematical and Computer Modelling*, 38(3-4):299–308.