



SERGIO LUIS MERCADO LONDOÑO

ESTIMAÇÃO DO NÚMERO DE REPRODUÇÃO BASAL EM
MODELOS COMPARTIMENTAIS

CAMPINAS
2014



UNICAMP

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE MATEMÁTICA, ESTATÍSTICA
E COMPUTAÇÃO CIENTÍFICA

SERGIO LUIS MERCADO LONDOÑO

ESTIMAÇÃO DO NÚMERO DE REPRODUÇÃO BASAL EM MODELOS COMPARTIMENTAIS

Dissertação apresentada ao Instituto de Matemática,
Estatística e Computação Científica da Universidade
Estadual de Campinas como parte dos requisitos exigidos
para a obtenção do título de Mestre em Estatística.

Orientador: Luiz Koodi Hotta

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA
DISSERTAÇÃO DEFENDIDA PELO ALUNO
SERGIO LUIS MERCADO LONDOÑO, E ORIENTADA PELO
PROF. LUIZ KOODI HOTTA.

Assinatura do Orientador

CAMPINAS
2014

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Ana Regina Machado - CRB 8/5467

M533e Mercado Londoño, Sergio Luis, 1981-
Estimação do número de reprodução basal em modelos compartimentais /
Sergio Luis Mercado Londoño. – Campinas, SP : [s.n.], 2014.

Orientador: Luiz Koodi Hotta.

Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de
Matemática, Estatística e Computação Científica.

1. Modelos compartimentais (Matemática). 2. Epidemiologia - Modelos
matemáticos. 3. Epidemiologia - Métodos estatísticos. 4. Biomatemática. I. Hotta,
Luiz Koodi, 1952-. II. Universidade Estadual de Campinas. Instituto de Matemática,
Estatística e Computação Científica. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Estimation of the basic reproduction number in compartmental
models

Palavras-chave em inglês:

Compartmental models (Mathematics)

Epidemiology - Mathematical models

Epidemiology - Statistical methods

Biomathematics

Área de concentração: Estatística

Titulação: Mestre em Estatística

Banca examinadora:

Luiz Koodi Hotta [Orientador]

Hyun Mo Yang

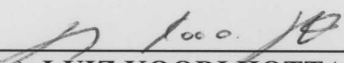
Luzia Aparecida Trinca

Data de defesa: 24-02-2014

Programa de Pós-Graduação: Estatística

Dissertação de Mestrado defendida em 24 de fevereiro de 2014 e aprovada

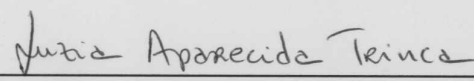
Pela Banca Examinadora composta pelos Profs. Drs.



Prof(a). Dr(a). LUIZ KOODI HOTA



Prof(a). Dr(a). HYUN MO YANG



Prof(a). Dr(a). LUZIA APARECIDA TRINCA

Abstract

The basic reproduction number, usually denoted by R_0 , is one of the most important quantities defined in epidemiology and is associated with the potential of an infectious disease to spread through a population. It provides a measure of the intensity needed to control the epidemic interventions. At the same time, the compartmental epidemiological models SIR and SEIR, both in the deterministic and in the stochastic approach, have been very helpful for understanding the mechanisms of infectious diseases transmission. This paper considers the four methods presented by Chowell et al. (Mathematical Biosciences, 2007, v. 208, p. 571-589) to estimate R_0 . All methods are based on compartmental epidemiological models. The first method is based on the epidemic (initial) exponential growth rate. Given an estimate of the exponential growth rate and an underlying compartmental model we have an estimate of R_0 . The second method is based on fitting SIR or SEIR compartmental models, and the third method in fitting an extended SEIR model. The fourth method uses a Bayesian approach to a stochastic SIR model. The aim of this work is to study the properties of estimators based on methods 1, 2 and 4. The bias, the mean squared errors, the coverage and the widths of the confidence intervals are estimated through simulation. The methods are studied when the true data generating processes are the stochastic SIR or SEIR models. Initially the methods, as presented by Chowell et al. (2007), were studied and then presented modifications to improve the performance of the estimators. The dissertation is organized as follows: Chapter 2 consists of the presentation of compartmental SIR and SEIR models, the deterministic and stochastic approaches for analysis of infectious diseases. This chapter also presents the basic reproduction number. Chapter 3 explains the four estimation methods presented in Chowell et al. (2007) to estimate the basic reproduction number. Chapter 4 discusses and compares three of the four methods by simulation when the data generating process is a SIR or SEIR model. In this chapter the modifications of the methods are also considered. The final conclusion and suggestions for future work are presented in Chapter 5.

Keywords: Compartmental models, Basic reproduction number, Exponential growth rate, SIR and SEIR models.

Resumo

Uma das quantidades mais importante definida na epidemiologia é o número de reprodução basal, ou básico, associado com a pandemia e denotado por R_0 . Ele proporciona uma medida da intensidade das intervenções necessárias para o controle da epidemia. Ao mesmo tempo, os modelos

epidemiológicos compartimentais SIR, SEIR, tanto no enfoque determinístico quanto no estocástico, têm sido de grande ajuda para a compreensão dos mecanismos de transmissão de doenças infecciosas em todo o mundo. Esta dissertação apresenta alguns métodos para estimar esta quantidade através da utilização dos modelos epidemiológicos compartimentais. São considerados os quatro métodos apresentados por Chowell et al. (Mathematical Biosciences, 2007, v. 208, p. 571-589). O primeiro método é baseado na taxa de crescimento (inicial) exponencial da epidemia. Dada a taxa de crescimento exponencial e o modelo subjacente temos uma estimativa de R_0 . No caso dos métodos 2 e 3 o processo de estimação do R_0 baseia-se nos modelos compartimentais, modelos SIR e SEIR no método 2, e em um modelo SEIR estendido no método 3. O método 4 utiliza uma abordagem bayesiana do modelo SIR estocástico.

O objetivo da dissertação é estudar as propriedades dos estimadores baseados nos métodos 1, 2 e 4. Através de simulações são estimados os vícios, os erros quadráticos médios, as coberturas e as larguras dos intervalos de confiança. Os métodos são estudados quando os verdadeiros processos geradores de dados são os modelos SIR ou SEIR estocásticos. Inicialmente foram estudados os métodos, como apresentados por Chowell et al. (2007), e depois apresentadas modificações para melhorar o desempenho dos estimadores.

A dissertação está organizada da seguinte forma: o Capítulo 2 consiste na apresentação dos modelos compartimentais, SIR e SEIR para análise das doenças infecciosas; tanto na abordagem determinística quanto estocástica. Este capítulo apresenta também o número de reprodução basal. O Capítulo 3 apresenta os quatro métodos de estimação apresentados em Chowell et al. (2007) para estimação do número de reprodução basal. O Capítulo 4 apresenta uma comparação de três dos quatro métodos através de simulação, quando o processo gerador de dados é um modelo SIR ou SEIR estocásticos. Neste capítulo também são apresentadas as modificações dos métodos. A conclusão final e as sugestões de trabalhos futuros são apresentadas no Capítulo 5.

Palavras-chave: Modelos compartimentais, Número de reprodução basal, Taxa de crescimento exponencial, Modelos SIR e SEIR.

Sumário

1	Introdução	1
2	Modelos Compartmentais	4
2.1	Modelagem determinística ou estocástica	4
2.2	Modelo SIR	5
2.2.1	Modelo SIR determinístico simples	6
2.2.2	Modelo SIR estocástico	7
2.3	Modelo SEIR	10
2.4	Modelo SEIR estendido	11
2.5	O número de reprodução basal R_0	13
2.5.1	Uma breve revisão sobre o cálculo do R_0	14
2.5.2	R_0 nos modelos compartmentais	18
3	Métodos Considerados para a Estimação do R_0	20
3.1	Método baseado na taxa de crescimento exponencial	20
3.1.1	Estimação da taxa de crescimento exponencial	21
3.1.2	Relação entre r e R_0 no modelo SIR determinístico simples	22
3.1.3	Relação entre r e R_0 no modelo SEIR determinístico simples	23
3.1.4	Estimação por intervalos de confiança	24
3.2	Método baseado nos modelos SEIR e SIR	25
3.3	Método baseado no modelo SEIR estendido	25
3.4	Método bayesiano baseado no modelo SIR estocástico	26
4	Simulações	30
4.1	Algoritmo para geração dos dados	31
4.2	Processo gerador de dados: SIR	32
4.2.1	Análise inicial dos estimadores	32
4.2.2	Estudo comparativo dos estimadores propostos	39
4.3	Processo gerador de dados: SEIR	45
5	Comentários Finais	59
	Bibliografia	61

Dedicada a minha mãe Luz Mary Londoño Legarda. . . .
por seu apoio espiritual

Não é o conhecimento
mas o ato de aprender
não a posse
mas o ato de chegar lá
que concede o maior prazer

Carl Friedrich Gauss

Agradecimentos

Agradeço ao Professor Luiz Koodi Hotta que me orientou e me deu a oportunidade de trabalhar um tema interessante na área de epidemiologia.

Agradeço à CAPES pelo financiamento do curso.

Agradeço ao Instituto de Matemática, Estatística e Computação Científica, IMECC, por minha formação como estatístico.

Agradeço ao Laboratório Epifisma (IMECC) e ao projeto temático da FAPESP *Avaliando controle de epidemias utilizando modelos matemáticos e computacionais* (Processo No 2009/15098-0) pelo apoio ao projeto.

Agradeço a Nilmar Bispo Santana que compartilhou seus conhecimentos sobre os modelos epidemiológicos.

Lista de Ilustrações

2.1	Dinâmica do fluxo dos indivíduos segundo o modelo SIR.	6
2.2	Dinâmica do fluxo dos indivíduos segundo o modelo SEIR simples.	10
2.3	Dinâmica dos indivíduos na população segundo o modelo determinístico SEIR es- tendido.	11
4.1	Número de casos novos diários observados na epidemia de influenza na Califórnia em 1918- 1919 e dadas pelo modelo determinístico SIR com os valores dos parâmetros utilizados para gerar as séries simuladas pelo modelo SIR estocástico.	33
4.2	Frequência da fase inicial da epidemia selecionada pelos critérios T_1 e T_{2b} nas 1000 séries geradas pelo modelo SIR estocástico.	35
4.3	Box plot das estimativas de R_0 pelo método 1, segundo a fase inicial da epidemia selecionada pelos critérios T_1 e T_{2b} . Baseados em 1000 séries geradas pelo modelo SIR estocástico. A linha cheia indica o verdadeiro valor de R_0	36
4.4	Box plot das estimativas de R_0 pelo método 2, segundo a fase inicial da epidemia selecionada pelos critérios T_1 e T_{2b} . Baseados em 1000 séries geradas pelo modelo SIR estocástico. A linha horizontal cheia indica o verdadeiro valor de R_0	37
4.5	Box plot das estimativas de R_0 pelo método 4, segundo a fase inicial da epidemia selecionada pelos critérios T_1 e T_{2b} . Baseados em 1000 séries geradas pelo modelo SIR estocástico. A linha horizontal cheia indica o verdadeiro valor de R_0	38
4.6	Box plot das estimativas de R_0 pelo método 1 segundo a fase inicial da epidemia selecionada pelos critérios T_1 e T_{2b} , com número mínimo de observações igual a 10. Baseados em 1000 séries geradas pelo modelo SIR estocástico. A linha horizontal cheia indica o verdadeiro valor de R_0	41
4.7	Box plot das estimativas de R_0 pelo método 2 segundo a fase inicial da epidemia selecionada pelos critérios T_1 e T_{2b} , com número mínimo de observações igual a 10. Baseados em 1000 séries geradas pelo modelo SIR estocástico. A linha horizontal cheia indica o verdadeiro valor de R_0	42
4.8	Box plot das estimativas de R_0 pelo método 4, segundo a fase inicial da epidemia selecionada pelos critérios T_1 e T_{2b} , com número mínimo de observações igual a 10. Baseados em 1000 séries geradas pelo modelo SIR estocástico. A linha horizontal cheia indica o verdadeiro valor de R_0	43

4.9	Box plot das estimativas de R_0 pelos métodos 1, 2 e 4 com os critérios de seleção T_1 e T_{2b} , com número mínimo de observações igual a 10 e com 30 observações. Baseados em 1000 séries geradas pelo modelo SIR estocástico. A linha horizontal cheia indica o verdadeiro valor de R_0	44
4.10	Distribuições a priori e a posteriori para a estimação do R_0 para uma série SIR após a décima observação. As distribuições a priori são as de maior variabilidade e a linha vertical indica o verdadeiro valor do R_0	46
4.11	Número de casos novos diários observados na epidemia de influenza na Califórnia em 1918-1919 e dados pelo modelo determinístico SEIR com os valores dos parâmetros utilizados para gerar as séries simuladas pelo modelo SEIR estocástico.	47
4.12	Frequência da fase inicial da epidemia selecionada pelos critérios T_1 e T_{2a} nas 1000 séries geradas pelo modelo SEIR estocástico.	49
4.13	Box plot das estimativas de R_0 pelo método 1 segundo a fase inicial da epidemia selecionada pelos critérios T_1 e T_{2a} , com número mínimo de observações igual a 10. Baseados em 1000 séries geradas pelo modelo SEIR estocástico. A linha horizontal cheia indica o verdadeiro valor de R_0	50
4.14	Box plot das estimativas de R_0 pelo método 2 segundo a fase inicial da epidemia selecionada pelos critérios T_1 e T_{2a} , com número mínimo de observações igual a 10. Baseados em 1000 séries geradas pelo modelo SEIR estocástico. A linha horizontal cheia indica o verdadeiro valor de R_0	51
4.15	Box plot das estimativas de R_0 pelos métodos 1 e 2 com os critérios de seleção T_1 e T_{2a} , com número mínimo de observações igual a 10 e com 30 observações. Baseados em 1000 séries geradas pelo modelo SEIR estocástico. A linha horizontal cheia indica o verdadeiro valor de R_0	52

Lista de Tabelas

4.1	Resumo dos estimadores utilizados na simulação segundo o modelo subjacente, método utilizado na seleção do período inicial e parâmetros conhecidos e parâmetros a ser estimados	30
4.2	Resumo das estimativas de R_0 para o modelo SIR estocástico. Resultados baseados em 1000 replicações. Método 1: método da taxa de crescimento intrínseco; método 2: modelo SIR determinístico; método 4: estimador bayesiano do modelo SIR estocástico. Fase inicial da epidemia selecionada pelos critérios T_1 e T_{2b} . Resultado por intervalo de T_1 e T_{2b} em $[4; 9]$, $[10; 20]$, $[20; 30]$ e para todas os casos	54
4.3	Resumo das estimativas de R_0 para o modelo SIR estocástico. Resultados baseados em 1000 replicações, exceto para os intervalos de confiança para o método 2, que é baseado em 200 replicações. Método 1: método da taxa de crescimento intrínseco; método 2: modelo SIR determinístico; método 4: estimador bayesiano do modelo SIR estocástico. Resultados com a fase inicial da epidemia selecionada pelos critérios T_1 e T_{2b} , com número mínimo de 10 observações, e com fase inicial com 30 observações	55
4.4	Correlação (acima da diagonal) e covariância das estimativas de R_0 para o modelo SIR estocástico. Resultados baseados em 1000 replicações. Método 1: método da taxa de crescimento intrínseco; método 2: modelo SIR determinístico; método 4: estimador bayesiano do modelo SIR estocástico. Resultados com fase inicial da epidemia selecionada pelos critérios T_1 e T_{2b} , com número mínimo de 10 observações, e com 30 observações	56
4.5	Resumo das estimativas de R_0 para o modelo SEIR estocástico. Resultados baseados em 1000 replicações, exceto para os intervalos de confiança para o método 2, que é baseado em 200 replicações. Método 1: método da taxa de crescimento intrínseco; método 2: modelo SEIR determinístico. Resultados com a fase inicial da epidemia selecionada pelos critérios T_1 e T_{2a} , com número mínimo de 10 observações, e com fase inicial com 30 observações	57
4.6	Correlação (acima da diagonal) e covariância das estimativas de R_0 para o modelo SEIR estocástico. Resultados baseados em 1000 replicações. Método 1: método da taxa de crescimento intrínseco; método 2: modelo SEIR determinístico. Resultados com a fase inicial da epidemia selecionada pelos critérios T_1 e T_{2a} , com número mínimo de 10 observações, e com fase inicial com 30 observações	58

Símbolos

B Número de amostras bootstrap.

$E(0)$ Número de indivíduos latentes iniciais.

$I(0)$ Número de indivíduos infectantes iniciais.

$\overline{m}$ Média.

Med Mediana.

N Tamanho total da população.

r Taxa de crescimento exponencial inicial, ou intrínseca.

R_0 Número de reprodução basal ou básico.

R_t Número de reprodução efetivo.

T_1 Período de crescimento inicial da epidemia escolhido pelo método 1 considerando o processo gerador dos dados SIR.

T_{2a} Período de crescimento inicial da epidemia escolhido pelo método 2 utilizando o modelo SEIR determinístico considerando o processo gerador dos dados SEIR.

T_{2b} Período de crescimento inicial da epidemia escolhido pelo método 2 utilizando o modelo SIR determinístico considerando o processo gerador dos dados SIR.

β Taxa de transmissão.

γ Taxa de remoção ou recuperação.

δ Taxa de mortalidade da doença.

κ Taxa de infecção.

μ Taxa natural de mortalidade.

Abreviações

CFP Proporção de casos fatais da doença.

EDO Equações diferenciais ordinárias.

EQM Erro quadrático médio.

SARS Síndrome respiratória aguda severo.

SEIR Modelo epidemiológico composto com os compartimentos: suscetíveis, latentes, infectantes e recuperados.

SIDA/VIH Síndrome do imune deficiência adquirida.

SIR Modelo epidemiológico composto com os compartimentos: suscetíveis, infectantes e recuperados

Capítulo 1

Introdução

A primeira e maior epidemia registrada na história da humanidade foi a praga na cidade de Atenas na Grécia (430-428 BC), cujas descrições foram fornecidas pelo historiador grego Tucídides (460-400 BC), como é apresentado em Dagnino (2011). Tucídides descreve aspectos importantes como os sintomas, avanços da epidemia e o número de mortos. Hipócrates (459-337), anos depois, discute os fatores que influenciaram na propagação desta doença e sua forma de difusão no tempo. Há diversos trabalhos esclarecedores sobre estas discussões de Hipócrates. Veja, por exemplo, Suter (1988) e Margotta (1968, cap. 2 e 3). Séculos depois, especificamente no século XVI, propagou-se na Europa uma nova epidemia conhecida como a peste bubônica, que ocasionou aproximadamente 25 milhões de mortes humanas no período de 1347 a 1353 (Haensch et al., 2010). Este vírus não foi erradicado por completo e reapareceu na Grã-Bretanha no ano 1665 com o nome de praga negra, cujo nome vem de seus sintomas. O vírus da praga negra foi transmitido aos humanos por pulgas que vivem em ratos. O vírus se instalou na população e se manteve inicialmente num nível endêmico sem ocasionar casos fatais. Entretanto, uma mutação na sua estrutura genética tornou o vírus novamente letal e matou mais de 100 mil pessoas no período de 1665 e 1666 na cidade de Londres (Haensch et al., 2010). Outros desastres epidêmicos aconteceram nos séculos seguintes. Podemos citar a propagação da varíola que matou aproximadamente 35 milhões de pessoas no século XVI no México (Brooks, 1993; Fenn, 2001) e, mais recentemente no século XX, depois da primeira guerra mundial, a pandemia da gripe que matou aproximadamente 20 milhões de pessoas em todo o planeta (Barry, 2004, cap. 4; Jhonson, 2006).

Atualmente também temos a presença de muitos surtos de epidemia, os quais são, em sua maioria, endêmicos. Por exemplo, a praga de Bombay na África (1905-1906) (Murray, 2002, cap. 10), a SARS (Síndrome respiratória aguda) em Cingapura estudada em Donelly et al. (2003) e Lipsitch et al. (2003). Outro problema de grande importância na atualidade é a estrutura genética de um vírus. Hoje, a velocidade das mutações dos vírus é mais rápida do que em séculos passados e, em consequência, podem passar mais facilmente ao humanos (Röhm, S. et al., 1995, Mark, et al., 2004).

Embora as doenças infecciosas tenham uma longa história, o estudo matemático destas e sua divulgação tem pouco mais de 300 anos de idade.

A epidemiologia, a ciência que estuda as doenças infecciosas, surgiu em 1662 quando Jhon Graunt publicou seu livro *Natural and Political Observations Made upon the Bills of Mortality*.

Neste livro ele discute vários problemas demográficos do século dezessete na Inglaterra, faz observações sobre os relatos das mortes e calcula o risco de morte com relação a certas doenças. As análises de Graunt proporcionaram o primeiro método sistemático para estimar os riscos comparativos de morte por causas de doenças infecciosas.

Um século mais tarde, em 1760, Daniel Bernoulli mostrou uma abordagem mais teórica para o estudo dos efeitos virais de uma doença, publicando o primeiro modelo epidemiológico, com o objetivo de demonstrar que a inoculação com o vírus vivo obtida diretamente a partir de um paciente com um caso leve de varíola reduz a taxa de mortalidade. Depois, D'Alembert em 1761, desenvolveu um método alternativo para lidar com os riscos concorrentes de mortes, que é aplicável a doenças não-infecciosas e infecciosas. Uma discussão do método proposto por D'Alembert pode ser encontrada em Hens et al. (2012, cap. 1). Pouco depois, em meados de 1800, Louis Pasteur apresenta sua teoria germinal das doenças infecciosas, que diz que a maioria das doenças infecciosas são causadas por germes. Simultaneamente, Robert Koch tornou-se famoso pela descoberta dos bacilos de antraz (1877), tuberculose (1882) e do vibrião da cólera (1883) e por seu desenvolvimento do postulado de Koch.

Estes novos estudos construíram as bases para o início da biologia matemática moderna, que começou com Hamer. Ele foi o primeiro que aplicou o princípio de ação das massas no ano 1906 para um modelo epidêmico determinístico em tempo discreto. Logo, Ross em 1911 publicou um modelo epidemiológico simples que aplicou à malária (Bacaër, 2011, cap. 10), e mais tarde, Kermarck e McKendrick (1927) generalizaram este modelo.

Estes modelos clássicos têm sido construídos segundo a abordagem determinística e são amplamente utilizados. A partir destes, novos modelos foram criados levando em consideração vários fatores, como a migração, a vacinação e a perda gradual da sua eficiência, quimioterapia, quarentena, imunidade passiva, heterogeneidade genética, distribuição não uniforme da população, etc. Existe uma ampla bibliografia sobre estes modelos. Ver, por exemplo, Hethcote (2000) e Fraser et al. (2004).

Generalizações importantes foram feitas por Anderson e May (1991, p. 209-227) e, como uma consequência dos estudos destes modelos, Liu et al. (1987) mostrou resultados importantes sobre as taxas de incidência não lineares em suas equações e, anos depois, uma grande variedade de modelos foram formulados para o estudo de diferentes doenças infecciosas. Por exemplo, modelos especiais foram criados para doenças como a varíola, malária, SIDA, SARS, sarampo, rubéola, cólera, coqueluche, difteria, gonorréia, sífilis, herpes, e outras.

A estimação dos parâmetros para doenças infecciosas tem sido de grande ajuda para muitos epidemiologistas porque ela permite a avaliação de estratégias de intervenção através de enfoques matemáticos, que ajudam a controlar a propagação de pandemia da gripe e outras doenças infecciosas em todo o mundo (Flahault et al., 1988; Ferguson et al., 2006). Mais detalhes podem ser encontrados em Cauchemez (2004), Gani et al. (2005) e Keeling e Rohani (2007, cap. 1 e 2).

Uma das quantidades mais importante definida na epidemiologia é o número de reprodução basal, ou básico, associado com a pandemia. Ela proporciona uma medida da intensidade das intervenções necessárias para o controle da epidemia, denotado por R_0 (Anderson e May, 1991, p. 17). Em outras palavras, esta quantidade é o número médio de casos infectados produzidos pelo caso inicial infeccioso introduzido numa população completamente suscetível (Dietz, 1993).

No trabalho desenvolvido nesta dissertação estamos interessados em estimar esta quantidade

através da utilização dos modelos epidemiológicos compartimentais. Estudaremos os quatro métodos apresentados por Chowell et al. (2007), os quais utilizam alguns modelos epidemiológicos básicos, que têm sido usados em estudos de doenças infecciosas para estimar o R_0 . Existem vários métodos utilizados na literatura para estimação desta quantidade. Ver, por exemplo, Farrington et al. (2001), Heffernan et al. (2005), Li et al. (2011), Ma et al. (2014). O primeiro método é baseado na taxa de crescimento (inicial) exponencial da epidemia. Dada a taxa de crescimento exponencial e o modelo subjacente temos uma estimativa de R_0 . No caso do método 2 e 3 o processo de estimação do R_0 baseia-se no ajuste dos modelos SEIR (suscetível, latente (exposto), infectante, recuperado) e SIR (suscetível, infectante, recuperado), enquanto o método 3 baseia-se no ajuste do modelo SEIR estendido. O método 4 utiliza uma abordagem bayesiana do modelo SIR estocástico.

O objetivo da dissertação é estudar as propriedades de estimadores baseados nos métodos 1, 2 e 4. Através de simulações são estimados os vícios, os erros quadráticos médio, as coberturas e as larguras dos intervalos de confiança. Os métodos são estudados quando os verdadeiros processos geradores de dados são os modelos SIR estocástico e SEIR estocástico. Inicialmente foram estudados os métodos, como apresentados por Chowell et al. (2007), e depois apresentadas modificações para melhorar o desempenho dos estimadores.

A dissertação está organizada em cinco capítulos. O Capítulo 2 consiste na apresentação dos modelos compartimentais SIR e SEIR para análise das doenças infecciosas, tanto na abordagem determinística quanto estocástica. Este capítulo apresenta também o R_0 , enquanto o Capítulo 3 apresenta os quatro métodos de estimação apresentados em Chowell et al. (2007) para estimá-lo. O Capítulo 4 apresenta uma comparação de três dos quatro métodos através de simulação quando o processo gerador de dados é um modelo SIR ou SEIR. Neste capítulo também são apresentadas as modificações dos métodos. A conclusão final e sugestões de trabalhos futuros são apresentados no Capítulo 5.

Capítulo 2

Modelos Compartimentais

Este capítulo tem por objetivo apresentar os modelos compartimentais, principalmente quando eles são utilizados para modelar o espalhamento de doenças infecciosas. Serão apresentadas as versões determinísticas e estocásticas. Uma referência básica para a abordagem determinística é o livro de Anderson e May (1991), e para a abordagem estocásticas podemos citar Andersson e Britton (2000, cap. 1 e 2) e Lekone e Finkenstädt (2006).

O nome compartimental vem do fato de que os indivíduos da população em estudo são divididos em classes. A classificação dos indivíduos nestas classes depende das características dos indivíduos relevantes à infecção sob estudo. Desta forma, as classes consideradas dependerão da forma de transmissão da epidemia e dos objetivos do estudo. Quanto mais detalhado o estudo, maior o número de classes considerado. Por exemplo, um dos modelos mais simples, divide a população em apenas três classes ou compartimentos: suscetíveis (S), infectados (I) e removidos (R) (mortos ou imunes). Brauer e Van den Driessche (2001) e Xia et al. (2005) aplicaram este modelo a doenças infecciosas como o sarampo e rubéola. Este modelo é chamando de SIR, cujas letras vem do tipo de classe. O número de indivíduos nas classes S, I e R, no tempo t , são denotados por $S(t)$, $I(t)$ e $R(t)$, respectivamente. O modelo compartimental é completamente definido pelas regras que movem temporalmente um indivíduo de um compartimento a outro, isto é, definida pela dinâmica da população. Estas regras podem ser determinísticas ou estocásticas, definindo se o modelo é determinístico ou estocástico. Uma rápida discussão sobre a utilização da modelagem determinística ou estocástica é apresentada na Seção 2.1. Nas seções seguintes serão apresentados alguns modelos e algumas regras determinísticas e estocásticas. A apresentação partirá do modelo mais simples para os mais complexos, mas poderia também ser apresentado um modelo mais complexo e depois os mais simples como casos particulares. Na Seção 2.2 é apresentado o modelo SIR, na Seção 2.3 o modelo SEIR e na Seção 2.4 o modelo SEIR estendido. Finalmente, na Seção 2.5 temos uma rápida discussão sobre a estimação do R_0 .

2.1 Modelagem determinística ou estocástica

A modelagem matemática surgiu como uma ferramenta importante para possibilitar uma melhor compreensão da dinâmica da disseminação de doenças infecciosas. Nesta abordagem, os

modelos geralmente são descritos através de um sistema de equações diferenciais, sendo as taxas de fluxo entre os compartimentos determinados pelos parâmetros, especificando a trajetória da doença. Ver, por exemplo, Hethcote (1975) e Yang (2001, cap. 2).

Dentre suas características, estes modelos descrevem e explicam o que acontece em média em uma população, e se encaixam bem em grandes populações para explicar a dinâmica dos indivíduos em diferentes compartimentos. Ver, por exemplo, o livro de Diekmann et al. (2012). Comparado com a abordagem estocástica, a abordagem determinística geralmente necessita de menos observações, é relativamente mais fácil de estimar e analisar, e existe maior disponibilidade de programas de computador.

É também reconhecido que a modelagem estocástica é importante (Bailey, 1975; Mollison, 1977; Andersson, 1999; Anderson e Britton, (2000, cap. 1); Artalejo e Lopez-Herrero, 2013; Chowell et al., 2013). Os modelos estocásticos explicam as flutuações aleatórias dos riscos de exposição à doença e a outros fatores, e fornecem muito mais conhecimento em uma modelagem ao nível individual, levando em consideração o tamanho pequeno da população, em que cada indivíduo desempenha um papel importante no modelo. Especificamente, os modelos estocásticos permitem obter as variabilidades individuais, como prever a dinâmica de um indivíduo em relação à doença e determinam o período adequado de aplicação de um tratamento de controle. Os modelos estocásticos permitem também considerar os parâmetros variando estocasticamente ao longo do tempo, inclusive como função de covariáveis. No entanto, pode ser difícil trabalhar com estes modelos e ser necessário realizar muitas simulações para se obter previsões úteis. O primeiro modelo estocástico foi o modelo conhecido como *modelo de cadeia binomial*, proposto por En'ko (1889). Este trabalho foi o precursor do modelo proposto por Reed e Frost em uma palestra em 1928 (Siettos e Russo, 2013). Dietz (1988) analisou e aplicou este modelo para o estudo da doença de transmissão sexual VIH. Nesta abordagem assume-se que tanto os períodos latentes como os infecciosos são independentes e distribuídos exponencialmente. Como consequência, dada a propriedade de falta de memória da distribuição exponencial, não é necessário guardar as informações sobre o tempo de transição de cada indivíduo, mas somente o número de indivíduos em cada compartimento.

2.2 Modelo SIR

No modelo SIR mais simples apresentado a seguir consideramos uma população fechada sem nascimento e morte e que, após a infecção, os indivíduos recuperados ficam imunes a novas infecções. Isto é uma boa aproximação quando a mortalidade pela doença é baixa e/ou o tempo da epidemia é curto em termos demográficos. A imunidade total também pode ser utilizada quando o tempo da imunidade parcial é grande comparado com o tempo da epidemia. O diagrama deste modelo é dado na Figura 2.1, no qual as setas indicam o fluxo dos indivíduos. Os significados das letras utilizadas no diagrama serão ainda explicados ainda nesta seção. O diagrama é o mesmo para os modelos determinísticos e estocásticos, que serão apresentados nas subseções seguintes.

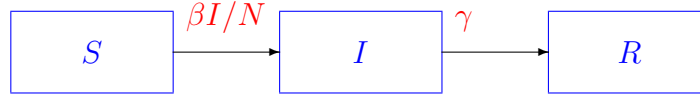


Figura 2.1: Dinâmica do fluxo dos indivíduos segundo o modelo SIR.

2.2.1 Modelo SIR determinístico simples

O modelo SIR mais simples é o modelo publicado por Kermarck e McKendrick (1927), que foi o ponto de partida para estudos das propriedades das epidemias.

Este modelo é dado pelo seguinte sistema de equações diferenciais:

$$\begin{aligned} S'(t) &= -\beta S(t)I(t)/N, \\ I'(t) &= \beta S(t)I(t)/N - \gamma I(t), \\ R'(t) &= \gamma I(t), \end{aligned} \tag{2.2.1}$$

onde N é o tamanho da população dado por $N = S(t) + I(t) + R(t)$; $S(t)$, $I(t)$, $R(t)$ são os números de indivíduos suscetíveis, infecciosos e recuperados, respectivamente, no tempo t ; $S'(t)$, $I'(t)$, $R'(t)$, denotam as derivadas de S , I , R no tempo t , respectivamente; β é a taxa de transmissão da doença; e γ é a taxa de recuperação.

Na formulação do modelo em função das derivadas do número de indivíduos em cada compartimento, assumiremos que o número de membros em um compartimento é uma função diferenciável no tempo, o que pode ser uma aproximação razoável se houver muitos membros em um compartimento, mas não é aceitável no caso contrário.

O modelo é baseado nas seguintes suposições:

- A população é fechada, isto é, não existe imigração e emigração de novos indivíduos, nem mortos nem nascimento, ou seja, o tamanho da população é constante.
- A população é homogênea, no sentido que, em média, os indivíduos têm intrinsecamente propriedades similares, independente de sua idade, constituição genética, hábitos sociais, localização geográfica.
- Um membro da população, em média, faz contato para transmitir a infecção para outro indivíduo a uma taxa $\beta I(t)/N$ por unidade de tempo (princípio de ação das massas).

- Assumimos que os indivíduos infectados são instantaneamente infectantes, isto é eles têm a capacidade de propagar a doença quando entram em contato com os indivíduos suscetíveis.
- Os indivíduos infectantes deixam a classe dos infectantes e passam à classe dos recuperados, onde permanecem imunes durante o tempo de duração da doença.

Existem outros tipos de modelos SIR, por exemplo o modelo SIR com taxa de nascimento e morte. Este modelo é similar ao modelo 2.2.1, com a diferença que ele tem um influxo de recém-nascidos na classe suscetível e saída devido à mortalidade em cada compartimento. A literatura sobre estes modelos é vasta, sendo que boas referências são encontradas; por exemplo, Hethcote (2000) e Gordillo e Greendwood (2007, 2008).

2.2.2 Modelo SIR estocástico

Existem várias formas de se introduzir a estocasticidade em modelos compartimentais. Nesta seção iremos apresentar três formas, sem maiores detalhes, inclusive a forma através de equações diferenciais estocásticas, que será a utilizada nesta dissertação. Como a generalização para modelos mais complexos do que os modelos SIR é quase que direta, não discutiremos estes modelos.

A decisão sobre qual modelo utilizar depende dos objetivos, e das características da epidemia e da população em estudo.

Equações diferenciais estocásticas

O modelo estocástico baseado em equações diferenciais estocásticas é muito parecido com o modelo determinístico. A parte estocástica entra na derivada através de um movimento Browniano. Apenas para ilustrar, no modelo SIR estocástico temos o seguinte sistema de equações estocásticas:

$$\begin{aligned} S'(t) &= -\beta S(t)I(t)/N(t) + a_1 W_1(t), \\ I'(t) &= \beta S(t)I(t)/N(t) - \gamma I(t) + a_2 W_2(t), \\ R'(t) &= \gamma I(t) + a_3 W_3(t), \end{aligned} \tag{2.2.2}$$

onde $W_i(t)$, $i = 1, 2, 3$ são movimentos brownianos independentes. Assim como nos modelos determinísticos, nestes modelos estocásticos o número de indivíduos não tem a restrição de ser inteiro. Esta suposição torna-se menos importante quando tratamos de grandes populações em todos os compartimentos. Outra crítica é que, no modelo, o número de indivíduos em cada compartimento pode ser negativo, exceto se forem colocadas restrições adequadas. Há toda uma discussão sobre estes modelos em Allen (2011, cap. 7 e 8) e Lahrouz et al. (2011).

Modelo de Reed-Frost

Reed e Frost, em um trabalho não publicado utilizado em suas aulas na Universidade de Johns Hopkins, modificaram um modelo em tempo discreto, originalmente proposto por Soper (1929).

Soper considerou uma população com as seguintes suposições:

- Todos os indivíduos são igualmente suscetíveis à doença.
- Os indivíduos tem a mesma capacidade de transmitir a doença.
- Os indivíduos passam ao compartimento dos removidos quando termina seu período de transmissão.
- Os períodos de incubação (latência) e infecção são relativamente curtos.
- A lei de ação das massas também é aplicado no processo de transmissão da doença.

Reed e Frost modificaram este modelo considerando que o contato entre um suscetível e um ou mais infecciosos produzirá só um novo infectado. Este modelo proposto por Reed e Frost, nomeado como modelo Reed-Frost, é o modelo epidemiológico mais simples para o estudo de algumas doenças infecciosas num grupo pequeno de indivíduos. Este modelo é um exemplo de um modelo de cadeia binomial, um modelo simplificado e iterativo que ajuda a analisar como se comportará uma epidemia ao longo do tempo (Matthew, 2005).

Este modelo baseia-se nas seguintes suposições:

- A infecção é transmitida de um indivíduo para outro, se acontece um contato entre um suscetível e um infectado.
- Qualquer indivíduo não imune na população depois de ter contato com um indivíduo infectante em um período de tempo tornar-se-à infectante só no período de tempo seguinte, e só, em períodos seguintes, poderá passar para o compartimento de recuperado.

O modelo Reed-Frost geralmente é especificado utilizando uma dinâmica em tempo discreto. Neste contexto consideramos que o período infeccioso médio é curto e é precedido por um período médio de latência, que se considera maior que o período infectantes médio. Então a nova geração de indivíduos infectantes, que se origina quando finaliza o período de latência, é separada da geração infectada anteriormente por um período de latência.

As probabilidades dos eventos em uma geração dependem somente do estado da doença na geração anterior (modelo markoviano). Estes eventos são especificados por probabilidades binomiais.

Seja p a probabilidade de contato entre quaisquer dois indivíduos na população em um intervalo de tempo dado. Então, $q = 1 - p$ é a probabilidade de um indivíduo suscetível não ter contato potencialmente infectante.

O contato potencialmente infeccioso depende:

- da suscetibilidade ou resistência do indivíduo,
- do potencial infeccioso da doença,
- do tempo de exposição de um indivíduo suscetível à infecção, e
- das condições do meio ambiente.

Reed e Frost consideram que uma epidemia não está determinada somente pelas condições iniciais porque, em cada geração há variações ao acaso dos eventos. Assim, se consideramos S_j e I_j o número de indivíduos suscetíveis e infecciosos, respectivamente no tempo (ou geração) j , então o modelo Reed e Frost pode ser formulado através das seguintes probabilidades condicionais:

$$\begin{aligned} P(I_{j+1} = i_{j+1} \mid S_0 = s_0, I_0 = i_0, \dots, S_j = s_j, I_j = i_j) &= P(I_{j+1} = i_{j+1} \mid S_j = s_j, I_j = i_j) \\ &= \binom{s_j}{i_{j+1}} (1 - q^{i_j})^{i_{j+1}} (q^{i_j})^{s_j - i_{j+1}}, \end{aligned}$$

e $S_{j+1} = S_j - I_{j+1}$.

Isto significa que, dado um indivíduo suscetível na geração j , este indivíduo permanecerá suscetível na próxima geração se ele não entra em contato com nenhum indivíduo infectante na geração j . Todos estes eventos são independentes e acontecem com probabilidade q , sendo que os diferentes indivíduos suscetíveis em uma geração j serão infectados de forma independente um do outro, e esses indivíduos infectantes serão removidos nas gerações seguintes.

Modelo baseado em agentes

Nos modelos estocásticos os indivíduos entram e saem em cada compartimento de forma estocástica. Os modelos baseados em agentes associam a cada indivíduo um modelo de probabilidade para o tempo de transição. O modelo probabilístico mais simples, que está estreitamente ligado ao modelo determinístico, é o modelo que considera que os tempos de transição são independentes, com distribuição exponencial, e com taxas de falhas iguais às taxas de transição dadas no modelo determinístico. Ver, por exemplo, Gordillho e Greenwood (2007, 2008).

Considere o modelo SIR determinístico e a mesma notação dada na Seção 2.2.1. Assim, o tempo até o primeiro indivíduo passar de suscetível a infectante tem uma distribuição exponencial com taxa de falha $\lambda_{SI}(t) = \beta S(t)I(t)/N(t)$, e o tempo até o próximo indivíduo infeccioso passar a recuperado tem uma distribuição exponencial com taxa de falha $\lambda_{IR}(t) = \gamma I(t)$. Dada a suposição de independência e as propriedades da distribuição exponencial¹, no tempo t , a distribuição do tempo até a ocorrência da próxima transição tem distribuição exponencial com taxa de falha $\lambda(t) = \lambda_{SI}(t) + \lambda_{IR}(t)$, com a probabilidade da transição ser de suscetível a infectante igual a $\lambda_{SI}(t)/\lambda(t)$.

Na prática, temos observações apenas no final dos períodos, geralmente com espaçamentos iguais; por exemplo, no final do dia. Como só temos informações dos números de transições entre períodos, a abordagem mais comum é considerar que, durante o intervalo, a probabilidade de transição de cada indivíduo é constante, sendo as taxas de falhas dadas pelos números de indivíduos no início do período. Assim temos que a probabilidade de um indivíduo suscetível ser infectado no intervalo $(t, t+1)$ é dada por $1 - \exp(-\lambda_{SI}(t))$, e de um indivíduo infectado ser recuperado no intervalo $(t, t+1)$ é dada por $1 - \exp(-\lambda_{IR}(t))$. Considerando independência temos que o número de indivíduos que passam de um compartimento a outro tem distribuição binomial. No caso de

¹ A distribuição do mínimo entre variáveis aleatórias independentes com distribuição exponencial é uma distribuição exponencial com taxa de falha dada pela soma das taxas de falhas

termos um número grande de indivíduos no compartimento de origem e probabilidade pequena de sair deste compartimento, a distribuição binomial pode ser substituída pela distribuição de Poisson.

A exigência de que os tempos de transição sejam exponencialmente distribuídos não é essencialmente restritiva, existem formas para generalizar estas suposições sem perder as propriedades markovianas do processo. Por exemplo, podemos adicionar algumas etapas, tal que o tempo de transição seja a soma de tempos exponenciais independentes.

2.3 Modelo SEIR

O modelo SEIR considera um compartimento adicional conhecido como classe dos latentes e constituídos por indivíduos já infectados, mas ainda não infectantes. Muitas vezes, a classe dos latentes é denominado de expostos, daí a utilização da letra E. Assim, segundo este modelo, na sua versão mais simples, a população é dividida em quatro compartimentos. As suposições sobre a população são as mesmas que as consideradas para o modelo SIR definido na Seção 2.2.1.

Como ilustração nós apresentamos na Figura 2.2 o diagrama que descreve a dinâmica da população segundo este modelo. Assim, como no modelo SIR não consideramos a entrada e saída de indivíduos no sistema.



Figura 2.2: Dinâmica do fluxo dos indivíduos segundo o modelo SEIR simples.

O modelo SEIR determinístico consiste do seguinte sistema de equações diferenciais ordinárias não lineares:

$$\begin{aligned}
 S'(t) &= -\beta S(t)I(t)/N(t), \\
 E'(t) &= \beta S(t)I(t)/N(t) - \kappa E(t), \\
 I'(t) &= \kappa E(t) - \gamma I(t), \\
 R'(t) &= \gamma I(t),
 \end{aligned} \tag{2.3.1}$$

onde $E(t)$ é o número de indivíduos latentes no tempo t ; $S(t)$, $I(t)$, $R(t)$, $S'(t)$, $E'(t)$, $I'(t)$ e $R'(t)$ são definidos como em 2.2; e $N(t) = S(t) + E(t) + I(t) + R(t)$. Os parâmetros do modelo SEIR

são definidos como no modelo SIR, com o parâmetro adicional taxa de infecção κ . O parâmetro é definido como a frequência por unidade de tempo em que os indivíduos saem do compartimento dos latentes e entram no compartimento dos infectantes.

Existem outros tipos de modelos SEIR. Ver, por exemplo, Li et al. (1999). A formulação do modelo SEIR estocástico é dada da mesma forma que o modelo SIR estocástico, apenas com a inclusão de mais um compartimento, o compartimento dos latentes. Ver Seção 4.1

2.4 Modelo SEIR estendido

Existem várias formas de se estender o modelo SEIR apresentado. Dependendo da epidemia e dos objetivos das análises, os compartimentos podem considerar fatores como o sexo e a faixa de idade dos indivíduos, ou outro tipo de heterogeneidade. Outras modificações podem ser o efeito de imunização ou, no caso em que o vetor da infecção não é o ser humano, o ciclo de vida do parasita. Nesta seção será apresentada apenas a extensão utilizada por Chowell et al. (2007) no seu terceiro método de estimação do número de reprodução basal, e também por Chowell e Nishiura (2008). O diagrama de fluxo deste modelo é dado na Figura 2.3.

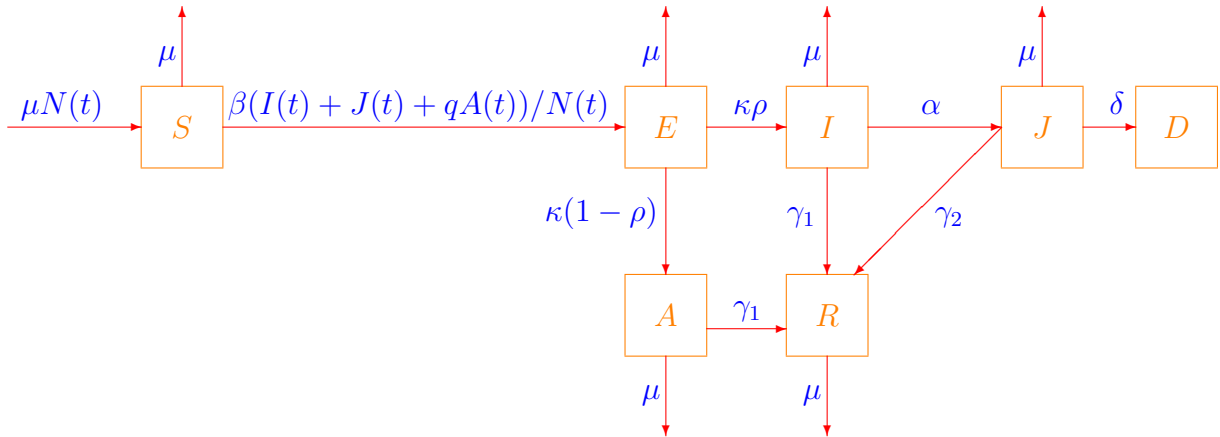


Figura 2.3: Dinâmica dos indivíduos na população segundo o modelo determinístico SEIR estendido.

O modelo apresentado nesta seção foi desenvolvido originalmente para estudar a transmissibilidade e o efeito de intervenções hipotéticas para a pandemia de gripe em Genebra, Suíça (Chowell et al., 2006). Neste modelo, os indivíduos são classificados como suscetíveis (S), latentes (E), clinicamente doentes e infecciosos (I), assintomáticos e parcialmente infecciosos (A), diagnosticados e informados (J), recuperados (R) e mortos (D).

O processo de transmissão descrito anteriormente pode ser modelado através do seguinte sistema de equações diferenciais não lineares:

$$\begin{aligned}
S'(t) &= \mu N(t) - \beta S(t)(I(t) + J(t) + qA(t))/N(t) - \mu S(t), \\
E'(t) &= \beta S(t)(I(t) + J(t) + qA(t))/N(t) - (\kappa + \mu)E(t), \\
A'(t) &= \kappa(1 - \rho)E(t) - (\gamma_1 + \mu)A(t), \\
I'(t) &= \kappa\rho E(t) - (\alpha + \gamma_1 + \mu)I(t), \\
J'(t) &= \alpha I(t) - (\gamma_2 + \delta + \mu)J(t), \\
R'(t) &= \gamma_1(A(t) + I(t)) + \gamma_2 J(t) - \mu R(t), \\
D'(t) &= \delta J(t), \\
C'(t) &= \alpha I(t),
\end{aligned} \tag{2.4.1}$$

onde:

- $S'(t), E'(t), A'(t), I'(t), J'(t), R'(t), D'(t), C'(t)$ denotam as derivadas de S, E, A, I, J, R, D, C no tempo t , respectivamente.
- O tamanho da população é dado por $N(t) = S(t) + E(t) + I(t) + A(t) + J(t) + R(t)$.
- μ é a taxa natural de mortalidade. No modelo foi considerado que a taxa natural de mortalidade é igual a taxa de nascimento. Isto torna o tamanho da população constante, facilitando bastante o tratamento do modelo.
- Os indivíduos suscetíveis entram em contato com o grupo dos infectantes numa taxa $\beta(I(t) + J(t) + qA(t))/N(t)$, em que β é a taxa de transmissão da classe portadora (indivíduos que são portadores de uma doença ou infecção e tem a capacidade de infectar, mas assintomática, isto é, não experimentam sintomas). Esta classe é denotada por A , $\beta \geq 0$.
- O parâmetro q , $0 < q < 1$ é o fator de redução na transmissibilidade da classe A .
- Os indivíduos saem da classe dos latentes a uma taxa κ , $\kappa \geq 0$. Uma proporção $0 < \rho < 1$ destes indivíduos passam à classe I (indivíduos clinicamente doentes e infectantes) à mesma taxa κ e a proporção restante $(1 - \rho)$ entram no compartimento A .
- Os indivíduos da classe A se recuperam a uma taxa γ_1 .
- Os indivíduos infectantes passam ao compartimento J (indivíduos diagnosticados e notificados) a uma taxa α .
- Os indivíduos na classe J morrem a uma taxa adicional δ , ou se recuperam a uma taxa γ_2 .
- Os indivíduos da classe I , são diagnosticados e observados a uma taxa α ou se recuperam sem serem diagnosticados (exemplo, infecções leves) a uma taxa γ_1 .
- Os indivíduos diagnosticados se recuperam a uma taxa $\gamma_2 = 1/(1/\gamma_1 - 1/\alpha)$ e tem uma taxa uma taxa adicional de mortalidade δ .

- A taxa de mortalidade da doença é ajustada de acordo com a proporção de casos fatais CFP , tal que $\delta = [CFP/(1 - CFP)](\mu + \gamma_2)$, em que $CFP \neq 1$.
- Todos os parâmetros são positivos.

As suposições deste modelo são:

- A população é totalmente suscetível no início da pandemia.
- As taxas de mortalidade em cada compartimento, exceto, no compartimento J , são consideradas iguais a taxa de natalidade.
- Os indivíduos que são hospitalizados e diagnosticados (J), são assumidos igualmente infectantes.
- A população é homogênea; portanto, o quociente $(I(t) + J(t) + qA(t))/N(t)$ é a probabilidade de um indivíduo suscetível ter um contato aleatório com um indivíduo infectante.

2.5 O número de reprodução basal R_0

O número de reprodução basal é um conceito chave na epidemiologia e é, indiscutivelmente, uma das mais avançadas e valiosas ideias que o pensamento matemático trouxe à teoria epidemiológica. Este conceito é fundamental para o estudo da dinâmica entre o hospedeiro e o patógeno. Mais importante ainda, R_0 serve como um parâmetro limite que prevê se uma infecção se espalhará. Originalmente este parâmetro foi utilizado para estudos em demografia (Sharpe e Lotka, 1911; Dublin, 1925; Kuczynski, (1928, cap. 2 e 3)). De forma independente, também foi estudado por Kermarck e McKendrick (1927) para infecções que podem ser transmitidas para humanos. MacDonald et al. (1968) também utilizou este conceito para estudar a dinâmica das doenças transmitidas por vetores, como a malária, e Dietz (1975) continuou com as ideias de Ross sobre o estudo das doenças infecciosas. Este conceito é amplamente utilizado no estudo de doenças infecciosas e, mais recentemente, em modelos de dinâmica de populações. R_0 é definido como o número esperado de indivíduos secundários produzidos por um indivíduo em seu vida, quando introduzido em uma população totalmente suscetível. A interpretação de "secundários" depende do contexto; por exemplo, em demografia e em ecologia, R_0 significa o sucesso reprodutivo de um membro da mesma espécie na população em seu tempo de vida, mas em epidemiologia significa o número médio de indivíduos infectados por um único membro infectado durante seu período infeccioso. Se $R_0 \leq 1$, então, em média, os indivíduos infectados não são substituídos e a doença não se propagará. No caso em que $R_0 > 1$, em média, o número de infectados aumentará com cada geração, e portanto, a epidemia poderá se propagar. Desta forma, o número de reprodução é um parâmetro com grande potencial de utilização como índice de vigilância epidemiológica. Quando uma doença inicia sua propagação existem condições extrínsecas e intrínsecas que favorecerão sua evolução e o número de reprodução pode ser uma boa medida do potencial de transmissão da doença. Em muitos modelos sobre doenças de transmissão, a prevalência (proporção de indivíduos que apresentam uma doença durante um período de tempo, em relação ao tamanho total da

população) máxima do número de indivíduos infectados e o tamanho final da epidemia é uma função crescente do R_0 , indicando que R_0 pode ser considerado um parâmetro que mede a difusão do agente transmissor. A propriedade de controle é o aspecto mais importante e útil do conceito R_0 . Numa infecção endêmica podemos determinar quais medidas de controle e, em que magnitude, seriam mais eficazes para a redução de R_0 abaixo de um, fornecendo orientações importantes para gerar políticas de vigilância na saúde dos indivíduos.

A magnitude de R_0 também é usada para medir o risco de uma epidemia ou pandemia infecciosa emergente. Por exemplo, foi de grande importância na compreensão da propagação e o perigo potencial de epidemias como a síndrome respiratória aguda grave SARS (Choi e Pak, 2003; Lipsitch et al., 2003; Riley et al., 2003; Chowell et al., 2003, 2004a), a encefalite bovina (Woolhouse et al., 1997; Ferguson et al., 1999), a febre aftosa (Ferguson et al., 2001; Matthews et al., 2003), novas estirpes de gripe (Mills et al., 2004; Stegeman et al., 2004, Chowell et al., 2013a, b), o vírus do oeste do Nilo (Wonham et al., 2004), a dengue (Luz et al., 2003), a malária (Hagmann et al., 2003), o Ébola (Chowell et al., 2004b; Gravenor et al., 2004), o vírus do ar (Rudnick e Milton, 2003), bioterrorismo (Kaplan et al., 2002; Longini et al., 2004), o vírus de computador (Lloyd e May, 2001) e o vírus bovino (De Koeijer et al., 2008).

Estes trabalhos teóricos desenvolvidos através da história ampliaram o conceito do R_0 para uma grande variedade de modelos complexos, tais como modelos com estrutura espacial (Lloyd e May, 1996; Keeling, 1999), ou modelos com estrutura de idade (Anderson e May, 1991 p. 172-179; Diekmann e Heesterbeek, 2000, cap. 7; Hyman e Li, 2000), modelos para microparasitos (Anderson e May, 1991, cap. 1; Diekmann e Heesterbeek, 2000, cap. 1), incluindo também modelos estocásticos e sistemas finitos (Nasell, 1995).

2.5.1 Uma breve revisão sobre o cálculo do R_0

Nesta seção apresentaremos uma breve revisão sobre os métodos utilizados para o cálculo do R_0 . Para uma breve revisão histórica, ver Heesterbeek (2002). Além as várias referências de aplicações e discussões citadas anteriormente podemos acrescentar Vynnycky et al. (2007) e Van den Driessche e Watmough (2008).

Método função de sobrevivência

Vários estudos tem assumido padrões de transmissão homogêneos para analisar a propagação de doenças infecciosas em modelos compartimentais definidos através de equações diferenciais; ver, por exemplo, Pugliese (1990) e Thieme (1992). Além disso, vários estudos matemáticos sobre a dinâmica de uma população em crescimento com perda de imunidade tem sido considerados. Ver, por exemplo, Busenberg e Haderl (1990) e Derrick e Van den Driessche (1993). Outros estudos, como os de Inaba e Nishiura (2008) consideram a estrutura etária e o crescimento da população suscetível. O método conhecido como *função de sobrevivência* considera esta abordagem. Este método é também descrito em Heesterbeek e Dietz (1996) e Heffernan et al. (2005).

Considere uma população grande. Seja $F_d(a)$ a função de sobrevivência, ou seja, a probabilidade que um recém-infeccioso permaneça infectante até um tempo a , e seja $b(a)$ o número médio de indivíduos infectados que um indivíduo infectante produzirá, por unidade de tempo, quando estiver

infectado por um período total de tempo a .

A função $b(.)F_d(.)$ é conhecida como função de reprodução. R_0 é definido por:

$$R_0 = \int_0^\infty b(a)F_d(a)da. \quad (2.5.1)$$

Esta abordagem é apropriada para qualquer modelo que apresente expressões para a função de sobrevivência, $F(a)$, e a infectividade como função do tempo, $b(a)$. Assim, esta derivação de R_0 não está restrita só aos sistemas descritos por equações diferenciais ordinárias (EDO).

Este método também pode ser estendido de forma natural para descrever modelos em que uma série de estados estão envolvidos na reprodução de um indivíduo infectado. Considere, por exemplo, a modelagem da epidemia de malária, como descrita em Heffernan et al. (2005). Inicialmente um humano infectado pode transmitir a infecção para um mosquito, que por sua vez pode infectar mais seres humanos. Este ciclo completo deve ser levado em consideração para a derivação de R_0 , o qual fornecerá o número total de pessoas infectadas por um ser humano infectado inicialmente. Em geral, se apenas dois estados infecciosos distintos estão envolvidos neste ciclo de infecção, então $F(a)$ pode ser definida como a probabilidade de que um indivíduo no estado 1 no tempo zero produza um indivíduo infeccioso no estado 2 até um tempo a . Do mesmo modo, $b(a)$ é o número médio de novos indivíduos no estado 1 produzidos por um indivíduo que tem permanecido no estado 2 durante um tempo a . Na modelagem da malária, $F(a)$ pode ser a probabilidade de que um humano infectado no tempo zero produza um mosquito infectado, que permanece vivo até pelo menos um tempo a . Desta forma, $F(a)$ é a integral do seguinte produto:

$$\begin{aligned} F(a) = & \int_0^a \text{prob}(\text{humano infectado no tempo 0 exista ainda no tempo } t) \\ & \times \text{prob}(\text{humano infectado durante um tempo total } t \text{ infecte um mosquito}) \\ & \times \text{prob}(\text{o mosquito infectado viva até um tempo } a - t) dt, \end{aligned} \quad (2.5.2)$$

enquanto que $b(a)$ é simplesmente o número médio de humanos infectados por um mosquito que permanecer infectado por um período de tempo a . Note que também poderíamos tomar o mosquito infectado, no estado 1, e depois derivar uma expressão que daria o mesmo valor de R_0 .

Infelizmente, derivações como as apresentadas na equação 2.5.2 tornam-se cada vez mais complicadas quando este método é estendido para ciclos de infecção envolvendo três ou mais estados (Hethcote e Tudor, 1980; Lloyd, 2001; Huang et al, 2003). Uma vasta bibliografia é encontrada na literatura sobre a derivação de R_0 , quando se tem mais de uma classe de infecciosos envolvidos no processo; por exemplo, Rushton e Mautner (1955); Hethcote (1978), Nold (1980) e Hethcote e Thieme (1985).

Método da matriz da próxima geração

Este método é aplicável a uma ampla variedade de modelos compartimentais, em particular aos modelos compartimentais SIR e SEIR determinísticos para calcular o R_0 . O método da matriz da próxima geração foi introduzido por Diekmann et al. (1990). O método pode ser encontrado em Heffernan et al. (2005). Este é um método geral para derivar R_0 , em casos onde a população

é dividida em classes disjuntas, e ele pode ser utilizado para modelos com estrutura espacial, entre outras possibilidades. Para sua implementação, as variáveis contínuas dentro da população são aproximadas por um número discreto de classes. Este método é amplamente utilizado e aplicações importantes podem ser encontradas em Matthews et al. (1999), Van den Driessche e Watmough (2002), Hill e Longini (2003) e Wonham et al. (2004). A teoria está baseada na teoria de matrizes não negativas descrita em Berman and Plemmons (1974). Este método foi utilizado para determinar a formulação teórica do número de reprodução basal considerando os modelos SIR, SEIR e SEIR complexo apresentados na Subseção 2.5.2.

Como ilustração considere o modelo tipo SEIR dado por:

$$\begin{aligned} S'(t) &= \Pi - \mu S - \beta S(t)I(t), \\ E'(t) &= \beta S(t)I(t) - (\mu + \kappa)E(t), \\ I'(t) &= \kappa E(t) - (\mu + \alpha)I(t), \\ R'(t) &= \alpha I(t) - \mu R(t), \end{aligned} \tag{2.5.3}$$

Este modelo é um caso geral do modelo SEIR determinístico apresentado na Seção 2.3 porque, considerando $\Pi = 0, \mu = 0, \alpha = \gamma$, temos o modelo SEIR determinístico 2.3.1.

Este modelo foi utilizado para o estudo de doenças infecciosas em crianças; por exemplo, o sarampo. Os novos infectados no compartimento E surgem através de contatos entre os indivíduos suscetíveis e infectados que estão nos compartimentos S e I , respectivamente a uma taxa $\beta S(t)I(t)$. Depois, os indivíduos passam do compartimento E ao compartimento I a uma taxa κ e finalmente se tornam imunes a uma taxa α . Além disso, alguns indivíduos morrem a uma taxa natural μ . Para maior simplicidade, este modelo assume uma constante de recrutamento Π de indivíduos suscetíveis. β é a taxa de transmissão, com incidência βSI (lei de ação das massa). O sistema 2.5.3 tem um único ponto de equilíbrio $S_0 = \Pi/\mu$.

Considere os dois compartimentos de infectados E e I . Utilizando a teoria do método da matriz da próxima geração temos: $\mathfrak{S}_1 = \beta SI$, $\mathfrak{S}_2 = 0$, $\nu_1 = (\mu + \kappa)E$ e $\nu_2 = \kappa E - (\mu + \alpha)I$.

Assim:

$$\begin{aligned} \mathfrak{S} &= \begin{pmatrix} \mathfrak{S}_1 \\ \mathfrak{S}_2 \end{pmatrix}, \\ \nu &= \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix}; \end{aligned}$$

portanto,

$$F = \begin{pmatrix} \partial \mathfrak{S}_1(0, S_0)/\partial E & \partial \mathfrak{S}_1(0, S_0)/\partial I \\ \partial \mathfrak{S}_2(0, S_0)/\partial E & \partial \mathfrak{S}_2(0, S_0)/\partial I \end{pmatrix},$$

$$V = \begin{pmatrix} \partial \nu_1(0, S_0)/\partial E & \partial \nu_1(0, S_0)/\partial I \\ \partial \nu_2(0, S_0)/\partial E & \partial \nu_2(0, S_0)/\partial I \end{pmatrix}.$$

Assim, obtemos:

$$F = \begin{pmatrix} 0 & \beta S_0 \\ 0 & 0 \end{pmatrix}$$

$$V = \begin{pmatrix} \mu + \kappa & 0 \\ -\kappa & \mu + \alpha \end{pmatrix}.$$

Finalmente, a matriz de geração seguinte denotada por K , é dada por

$$\begin{aligned} K &= FV^{-1} \\ &= \begin{pmatrix} \kappa\beta S_0/((\mu + \kappa)(\mu + \alpha)) & \beta S_0/(\mu + \alpha) \\ 0 & 0 \end{pmatrix}. \end{aligned}$$

Como consequência o número de reprodução basal é dado pelo autovalor próprio não negativo da matriz K , isto é,

$$R_0 = \kappa\beta S_0/((\mu + \kappa)(\mu + \alpha)).$$

Método do Jacobiano

O método da matriz Jacobiana é um dos métodos mais utilizados para calcular o número de reprodução e é utilizado quando o modelo é descrito por um sistema de equações diferenciais ordinárias. O método baseia no estudo dos autovalores do jacobiano no ponto de equilíbrio do sistema quando livre da doença. O problema com este método é que o valor encontrado para R_0 pode não ter significado biológico. Este é um dos motivos que leva alguns pesquisadores a considerar inapropriado chamar o valor encontrado por este método de R_0 (Heffernan et al., 2005). O método consiste de 6 passos:

- Calcular o ponto de equilíbrio da doença.
- Calcular a matriz jacobiana.
- Calcular a matriz jacobiana no ponto de equilíbrio.
- Determinar os autovalores da matriz jacobiana.
- Se todos os autovalores são negativos então existe estabilidade; caso contrário, existe instabilidade.
- $R_0 = \lambda_{max}$.

Método do termo constante do polinômio característico

Este é um método simples, amplamente utilizado também quando o modelo é dado por um sistemas de equações diferenciais ordinárias. Considere o polinômio característico relacionado com o jacobiano anterior:

$$\lambda^n + a_{n-1}\lambda^{n-1} + \dots + a_1\lambda + a_0 = 0. \quad (2.5.4)$$

Este método é relativamente fácil de utilizar pois basta encontrar o maior valor próprio. Usando este método, o parâmetro R_0 é derivado da condição de que todos os valores próprios do Jacobiano devem ter uma parte real negativa. Isto pode facilmente ser feito utilizando o polinômio característico 2.5.4 e as condições de estabilidade de Routh-Hurwitz. Uma discussão completa deste método é dada em Brauer et al. (2008).

Todos os métodos descritos anteriormente formulam o parâmetro R_0 em termos de algum modelo determinístico; por exemplo, um modelo compartimental. Esta abordagem geralmente considera que alguns parâmetros são conhecidos para simplificar o processo de estimação. Para ter uma visão mais completa destas técnicas considere, por exemplo Hethcote (2000).

Outros métodos de grande importância são:

- Suscetíveis em equilíbrio endêmico (Mollison, 1995).
- Método da idade média da infecção que estima R_0 como o raio L/A , em que L é o tempo médio de vida e A é a idade média de aquisição da doença (Dietz, 1975). Uma derivação completa é dada em Hethcote (2000), e para uma maior discussão ver Anderson e May (1991, p. 72-74) e Brauer (2002).

Atualmente, um dos métodos mais utilizado na estimação do R_0 utiliza a taxa de crescimento constante no período inicial da epidemia. Esta taxa também é conhecida como taxa de crescimento intrínseca da epidemia. Este método é utilizado, por exemplo, em Lipsitch et al. (2003) e Chowell et al. (2007). Além disso, este método é discutido em detalhes na Seção 3.1 desta dissertação.

2.5.2 R_0 nos modelos compartimentais

Nesta seção apresentaremos o valor de R_0 apenas para os modelos compartimentais SIR, SEIR e SEIR estendido apresentados anteriormente.

Segundo o método da matriz da próxima geração, o número de reprodução segundo o modelo SIR determinístico apresentados na Seção 2.2.1 é dado por:

$$R_0 = \beta/\gamma. \quad (2.5.5)$$

Há diversos trabalhos esclarecedores sobre a formulação teórica do R_0 considerando o modelo determinístico SIR. Ver, por exemplo Hethcote (2000) e Allen (2011, cap. 7 e 8).

Segundo o método da matriz da próxima geração, o número de reprodução basal para o modelo SEIR determinístico apresentado na Seção 2.3 é dado por:

$$R_0 = \beta/(\gamma + \delta). \quad (2.5.6)$$

Existem diversas referências importantes sobre a formulação teórica do número de reprodução basal para o modelo SEIR simples. Ver, por exemplo Hethcote (2000), Chowell et al. (2007) e Chowell e Brauer (2009).

Aplicando o método da matriz da próxima geração (Diekmann e Heesterbeek, 2000, cap. 7) ao sistema de equações diferenciais ordinárias do modelo SEIR estendido 2.4.1, apresentado em Chowell et al. (2006), o número de reprodução basal é dado por:

$$R_0 = \frac{\beta k}{\kappa + \mu} \left\{ \rho \left(\frac{1}{\gamma_1 + \alpha + \mu} + \frac{\alpha}{(\gamma_1 + \alpha + \mu)(\gamma_2 + \delta + \mu)} \right) + (1 - \rho) \left(\frac{q}{\gamma_1 + \mu} \right) \right\}. \quad (2.5.7)$$

Capítulo 3

Métodos Considerados para a Estimação do R_0

Existem vários métodos citados na literatura para a estimação do R_0 , como apresentado brevemente na Seção 2.5. Neste capítulo, apresentaremos com maiores detalhes os quatro métodos utilizados no artigo de Chowell et al. (2007). Dentro do escopo da dissertação todos os métodos considerados são relacionados a modelos compartimentais, determinísticos ou estocásticos. A escolha do estimador depende, entre outras coisas, da informação disponível. Geralmente não se observa o número de indivíduos de todos os compartimentos e nem durante o início da epidemia. Assim, como em Chowell et al. (2007), consideraremos que observamos apenas o número de casos novos diários de infectados durante um certo período, e utilizaremos os modelos SIR, SEIR e SEIR estendido. Em cada um dos modelos consideraremos que alguns dos parâmetros são conhecidos. Serão analisados quatro métodos de estimação. O método 1, apresentado na Seção 3.1 é baseado na suposição de que, no início de uma epidemia a taxa de crescimento é exponencial, e depois na relação entre esta taxa e o R_0 . Na estimação da taxa de crescimento exponencial não é considerado nenhum modelo em particular, mas na relação entre a taxa e R_0 é necessário adotar um modelo. No caso, iremos considerar modelos SIR e SEIR determinísticos. Os métodos 2 e 3 utilizam os modelos SEIR e SEIR estendido determinísticos, respectivamente. Estes métodos são apresentados nas Seções 3.2 e 3.3, respectivamente. O método 4, apresentado na Seção 3.4, utiliza uma abordagem bayesiana do modelo SIR estocástico para estimar o R_0 .

3.1 Método baseado na taxa de crescimento exponencial

Esta seção apresenta o método 1 para estimar o número de reprodução basal R_0 . O método está baseado na natureza intrínseca da doença, a qual geralmente apresenta um crescimento exponencial no período inicial da epidemia, quando ela avança livremente sem intervenção. O modelo considera que a taxa de crescimento da população infectada é proporcional ao número de indivíduos infectados. Esta suposição é geralmente válida para períodos iniciais da doença e para populações grandes. Mais detalhes podem ser encontrados em Brauer e Castillo-Chavez (2001, p. 4 a 6). O conceito de crescimento exponencial das doenças infecciosas na sua fase inicial tem sido

amplamente discutido em Lipsitch et al., (2003), Heffernann et al. (2005), Chowell et al. (2007) e Gani et al. (2011).

Doenças infecciosas como a gripe de 1918 em San Francisco, a gripe A(H1N1) de 2009 no México (Yang, et al. 2009), a síndrome respiratória aguda severa (SARS) de 2002 em Cingapura, e muitas outras são exemplos de epidemias que mostram este comportamento exponencial na sua fase inicial.

Denote por T o tempo até o qual podemos aproximar o número total de infectados pelo crescimento exponencial. O valor de T deve ser selecionado segundo algum critério. Nesta dissertação serão utilizados dois critérios.

O método é dividido em duas partes. Na primeira parte é estimada a taxa de crescimento exponencial r utilizando a fase de crescimento exponencial. Nesta fase também é necessário selecionar o período inicial da epidemia para selecionar as observações utilizadas para estimar a taxa, isto é, é necessário selecionar o valor de T . Na segunda parte, dada a estimativa da taxa de crescimento exponencial r , utiliza-se um modelo subjacente para a epidemia em que é conhecida a relação entre R_0 e r para estimar R_0 . Esta relação é derivada na Seção 3.1.2 para o modelo SIR e SEIR. Na Seção 3.1.1 é apresentado o método de estimação de r .

3.1.1 Estimação da taxa de crescimento exponencial

Em todos os métodos vamos considerar que são observados apenas o número total de casos notificados da epidemia no final de cada dia. Denote por $C_o(t), t = 1, 2, \dots, n$, o número de casos notificados até o final do dia t , com n o período total de observações. Observe que $C_o(1)$ é o número de caso observados no primeiro dia, mas não é o número de casos de infectados. Durante a fase de crescimento inicial o número total de casos notificados é modelado por

$$f(a_0, r, t) = a_0 \exp(rt), t \leq T, \quad (3.1.1)$$

onde a_0 é a condição inicial e r é o parâmetro que mede a velocidade de propagação da doença, conhecida como taxa de crescimento exponencial, e consideramos que até a observação T podemos considerar um crescimento exponencial da epidemia e não houve ainda nenhuma intervenção. O valor de T deve ser selecionado pelo pesquisador e é no máximo igual ao número de observações. O valor mínimo depende de cada autor. Chowell et al. (2007) sugere utilizar o mínimo de quatro observações, de modo a ter pelo menos dois graus de liberdade¹. Após uma transformação linear do $f(a_0, r, t)$ temos um modelo linear, mas a estimativa por mínimos quadrados não é adequada porque a variância das perturbações não é constante. Considerando que a variância do erro deve ser proporcional a $f(a_0, r, t)$, sugere-se estimar o modelo através da minimização da função objetivo, ou função perda:

$$L_1(a_0, r; T) = \sum_{i=1}^T \left(\frac{f(a_0, r, i) - C_o(i)}{\sqrt{f(a_0, r, i)}} \right)^2, \quad (3.1.2)$$

¹Número de observações menos o número de parâmetros estimados

que pode ser considerada uma medida de bondade de ajuste. Denote os valores de a_0 e r que minimizam 3.1.2 de $\hat{a}_{0,T}$ e $\hat{r}_T$, respectivamente.

Como para cada valor de T temos uma estimativa de r é necessário escolher um valor de T . Um critério, utilizado por Favier et al. (2006), Chowell et al. (2006, 2007) e Nishiura (2006) baseia-se no fato de que, sendo o modelo adequado, a variável aleatória associada a $L_1(\hat{a}_{0,T}, \hat{r}_T; T)$ tem distribuição aproximadamente qui-quadrado com $T - 2$ graus de liberdade. Assim, o valor de T selecionado é o que minimiza

$$P[X_{T-2} \leq L_1(\hat{a}_{0,T}, \hat{r}_T; T)],$$

onde X_{T-2} é uma variável aleatória com distribuição qui-quadrado com $T - 2$ graus de liberdade e $L_1(\hat{a}_{0,T}, \hat{r}_T; T)$ é o valor observado.

Em algumas situações podemos dar a T um valor a priori T_0 , menor do que n . Por exemplo, se sabemos que o primeiro controle da epidemia ocorreu no vigésimo dia, podemos tomar $T_0 = 20$. Vamos denotar o valor do limite do período selecionado para a fase inicial da epidemia através deste método como T_1 . Na Seção 3.2 apresentaremos outro método de seleção do período. Assim, se os dois valores do período forem diferentes, teremos dois estimadores diferentes no método 1. Ma et al. (2014) discutem e comparam outras formas alternativas de se estimar a taxa de crescimento exponencial de epidemias.

3.1.2 Relação entre r e R_0 no modelo SIR determinístico simples

Considere o modelo epidemiológico SIR determinístico dado pelo sistema de equações diferenciais 2.2.1. Como $S(t) + I(t) + R(t) = N$, onde N é o tamanho da população que consideramos constante $\forall t \geq 0$, e supondo que $S(t) \approx N$ perto do ponto de equilíbrio, então podemos considerar o sistema:

$$\begin{aligned} S'(t) &= -\beta I(t) \\ I'(t) &= \beta I(t) - \gamma I(t). \end{aligned}$$

Portanto,

$$I'(t) = \beta I(t) - \gamma I(t),$$

que é uma equação diferencial ordinária com solução:

$$I(t) = I(0)e^{(\beta-\gamma)t}.$$

Pela definição da taxa de crescimento exponencial 3.1.1, temos que $r = \beta - \gamma$. Assim, utilizando o resultado 2.5.5, temos que

$$R_0 = 1 + r/\gamma. \tag{3.1.3}$$

3.1.3 Relação entre r e R_0 no modelo SEIR determinístico simples

Considere o modelo epidemiológico SEIR determinístico dado na Seção 2.3. Considerando que $S(t) \approx N$ perto do ponto de equilíbrio, o sistema de equações diferenciais 2.3.1 pode ser escrito como

$$\begin{aligned} S'(t) &= -\beta I(t), \\ E'(t) &= \beta I(t) - \kappa E(t), \\ I'(t) &= \kappa E(t) - (\gamma + \delta)I(t). \end{aligned} \tag{3.1.4}$$

Linearizando o sistema de equações diferenciais 3.1.4, omitindo o índice do tempo para facilidade de notação, obtemos:

$$\begin{bmatrix} E' \\ I' \end{bmatrix} = \begin{bmatrix} -\kappa & \beta \\ \kappa & -(\gamma + \delta) \end{bmatrix} \begin{bmatrix} E \\ I \end{bmatrix}.$$

Seja

$$J = \begin{bmatrix} -\kappa & \beta \\ \kappa & -(\gamma + \delta) \end{bmatrix}.$$

Segundo a definição dos autovalores para J , temos:

$$\begin{aligned} (-\kappa - x)(-g - x) - \kappa\beta &= 0, \\ \kappa g + \kappa x + xg + x^2 - \kappa\beta &= 0, \\ x^2 + (\kappa + g)x + \kappa(g - \beta) &= 0. \end{aligned}$$

Logo, a taxa de crescimento exponencial é dada por:

$$\begin{aligned} r &= \left((\kappa + g) + \sqrt{(\kappa + g)^2 - 4\kappa(g - \beta)} \right) / 2, \\ 2r &= (\kappa + g) + \sqrt{(\kappa + g)^2 - 4\kappa(g - \beta)}, \\ 4r^2 + 4r(\kappa + g) + (\kappa + g)^2 &= (\kappa + g)^2 - 4\kappa(g - \beta), \\ 4r^2 + 4r(\kappa + g) &= -4\kappa(g - \beta), \end{aligned}$$

$$4r^2 + 4r(\kappa + g) = 4\kappa g(\beta/g - 1). \tag{3.1.5}$$

Assim, substituindo 2.5.6 em 3.1.5, obtemos:

$$\begin{aligned} 4r^2 + 4r(\kappa + g) &= 4\kappa g(R_0 - 1), \\ R_0 &= 1 + r^2/\kappa g + r(\kappa + g)/\kappa g. \end{aligned}$$

Segundo este processo obtemos que

$$R_0 = 1 + Vr + f(1 - f)(Vr)^2, \quad (3.1.6)$$

onde $g = \gamma + \delta$, $V = 1/\kappa + 1/\gamma$, $f = (1/g)/V$. V é conhecido como intervalo serial (Wallinga e Lipsitch, 2007).

Dependendo do modelo epidemiológico escolhido e considerando a estimativa do parâmetro r na primeira parte do processo de estimação do número de reprodução basal R_0 , a estimativa deste parâmetro é obtida depois de substituir a estimativa do r na expressão 3.1.3, ou na expressão 3.1.6, que é o resultado da linearização do modelo epidemiológico SIR e SEIR determinístico, respectivamente.

Existe uma ampla bibliografia sobre o processo de linearização dos modelos compartimentais apresentados nesta seção. Ver, por exemplo, Chowell e Brauer (2001), Lipsitch et al. (2003), Li e Jin (2004, 2005), Heffernan et al. (2005), Allen (2011, cap.7 e 8).

3.1.4 Estimação por intervalos de confiança

Para quantificar as incertezas dos estimadores de R_0 baseados no método 1 utilizaremos a técnica bootstrap paramétrica de Efron e Tibshirani (1986). Nesta seção apresentaremos o método utilizado por Chowell et al. (2007). Algumas modificações serão apresentadas mais tarde na Seção 4.2.1. A mesma técnica é utilizada para os estimadores baseados nos métodos 2 e 3, que serão apresentados nas Seções 3.2 e 3.3, respectivamente. No método 4, que utiliza a abordagem bayesiana, os intervalos de credibilidade são obtidos das distribuições a posteriori.

Considere $C_o(t)$, $t = 1, \dots, n$ como sendo os dados observados, ou seja os casos acumulados de indivíduos infectados, e a partir de estas observações queremos construir B conjuntos de realizações de C_o utilizando bootstrap paramétrico. O método de se obter as amostras bootstraps foi utilizado, por exemplo, por Chowell et al. (2007) e Chowell e Nishiura (2008) para o mesmo problema. Na série observada temos que $C'_o(t) = C_o(t) - C_o(t - 1)$, é o número notificado de casos novos de indivíduos infectados no período $(t - 1, t)$. Dado os valores observados, a estimativa do parâmetro da distribuição de Poisson em um dado intervalo é dada pelo número de casos novos observados neste intervalo. Denote por $C'_{oi}(t)$, $t = 1, \dots, n$ a i -ésima amostra bootstrap do número de casos notificados no período $(t - 1, t)$. A amostra $C'_{oi}(t)$ é obtida de uma distribuição de Poisson com média $C'_o(t)$. Tomando-se $C'_{oi}(1) = C_o(1)$, $i = 1, 2, \dots, B$, a i -ésima amostra bootstrap é dada por $C_{oi}(t) = \sum_{j=1}^t C'_{oi}(j)$, $t = 1, \dots, n$. No estudo de simulação a ser apresentado no Capítulo 4 veremos que este procedimento não tem um bom desempenho e serão apresentadas outras alternativas.

Seja $\widehat{R}_{0i}^*$ a estimativa de R_0 para $i = 1, \dots, B$. Considerando a distribuição empírica de $\widehat{R}_{01}^*, \dots, \widehat{R}_{0B}^*$ como uma estimativa da verdadeira distribuição do estimador de R_0 , $\widehat{R}_0$, podemos construir intervalos de confiança e estimar a variância do estimador.

A suposição de Poisson pode não ser uma boa aproximação quando o número de indivíduos no compartimento anterior ao compartimento de infectantes não for grande e se a suposição de que a probabilidade de um indivíduo do compartimento anterior passar a infectante ser constante no período não for aproximadamente válida. Uma alternativa é a utilização da distribuição binomial negativa, que possui a vantagem adicional de não ter a restrição da variância ser igual a média, mas

possui a desvantagem de possuir mais um parâmetro. No Capítulo 4 com um estudo de simulação será verificado se este procedimento é capaz de construir intervalos com a cobertura correta.

3.2 Método baseado nos modelos SEIR e SIR

Nesta seção nós apresentaremos o método 2 para estimar o R_0 , que é baseado na escolha do modelo SEIR ou SIR. Inicialmente, como em Chowell et al. (2007) iremos apresentar o método quando ajustado o modelo SEIR determinístico dado pelas equações 2.3.1. A função objetivo, ou função perda considerada é:

$$L_2(\beta, I(0); T) = \sum_{i=1}^T \left(\frac{C(i) - C_o(i)}{\sqrt{C(i)}} \right)^2, \quad (3.2.1)$$

onde $C(t) = I(t) + R(t)$, ou seja $C(t)$ é o número de casos notificados totais no tempo t estimado depois de ajustar o modelo 2.3.1 aos dados observados, e $C_o(t)$ é o número de casos infecciosos totais observados no tempo t . Assim, como em Chowell et al. (2007), consideramos que os únicos valores desconhecidos são $I(0)$ e β , e suas estimativas são os valores destes parâmetros que minimizam 3.2.1; denote-os por $\hat{\beta}_T$ e $\hat{I}(0)_T$, respectivamente. Os outros valores são considerados conhecidos e iguais aos valores verdadeiros utilizados no modelo gerador de dados, que foi baseado em valores encontrados na literatura. Observe que é necessário selecionar o valor de T . Uma possibilidade é utilizar o valor de T_1 como descrito na Seção 3.1.1. Outra sugestão, é estimar a fase de crescimento exponencial, portanto, analogo como na Seção 3.1.1, a estimativa de T é aquele valor que minimiza o percentil de $L_2(\hat{\beta}_T, \hat{I}(0)_T, T)$ na distribuição qui-quadrado com $T - 2$ graus de liberdade. Denote este valor de T_{2a} . Dada uma estimativa de β , R_0 é estimado substituindo a estimativa do β na expressão 2.5.6. Assim, no modelo SEIR temos dois estimadores de R_0 , um com o período dado por T_1 e outro por T_{2a} . Podemos ajustar também o modelo SIR determinístico dado pelas equações 2.2.1. O processo de estimação é similar ao caso do modelo SEIR. Para evitar confusões denote o período selecionado pelo modelo SIR de T_{2b} .

O processo para quantificar a incerteza do parâmetro R_0 neste método está baseado na técnica bootstrap, cujas amostras são realizadas como na Seção 3.1.4

3.3 Método baseado no modelo SEIR estendido

Nesta seção nós apresentaremos o método 3 para estimar o R_0 , método este baseado no modelo SEIR estendido 2.4.1. O método é o mesmo que o apresentado na Seção 3.2, apenas com o modelo SEIR básico substituído pelo modelo SEIR estendido. A função objetivo, ou função perda considerada é:

$$L_3(\beta, q, \rho, \gamma_1, \alpha, E(0), I(0); T) = \sum_{i=1}^T \left(\frac{C(i) - C_o(i)}{\sqrt{C(i)}} \right)^2, \quad (3.3.1)$$

onde $C_o(t)$ é o número de casos infecciosos totais observados no tempo t , e $C(t)$ é o número de casos infecciosos totais no tempo t estimado depois de ajustar o modelo 2.4.1 ao dados observados. Consideramos que os valores desconhecidos são $\beta, q, \rho, \gamma_1, \alpha, E(0)$ e $I(0)$ e suas estimativas são os valores que minimizam 3.3.1; denote-os por $\hat{\beta}_T, \hat{q}_T, \hat{\rho}_T, \hat{\gamma}_{1T}, \hat{\alpha}_T, \hat{E}(0)_T$ e $\hat{I}(0)_T$, respectivamente. Os outros valores são considerados conhecidos e encontrados na literatura. Observe que é necessário selecionar o valor de T , em que $T \geq 7$. Uma possibilidade é utilizar o valor de T_1 que foi escolhido na Seção 3.1.1. Outra sugestão, análogo ao realizado na Seção 3.1.1, é selecionar como valor de T o valor que minimiza o percentil de $L_3(\hat{\beta}_T, \hat{q}_T, \hat{\rho}_T, \hat{\gamma}_{1T}, \hat{\alpha}_T, \hat{E}(0)_T, \hat{I}(0)_T; T)$ na distribuição qui-quadrado com $T - 7$ graus de liberdade. Denote este valor de T_3 . Dadas as estimativas dos parâmetros $\beta, q, \rho, \gamma_1, \alpha, E(0)$ e $I(0)$, relativas ao período inicial escolhido, R_0 é estimado substituindo estas estimativas na expressão 2.5.7. Assim, no modelo SEIR estendido temos um estimador de R_0 , com o período dado por T_3 .

O processo para quantificar a incerteza do parâmetro R_0 neste método, como no caso anterior, é baseado no método bootstrap, cujas amostras são realizadas como na Seção 3.1.4

3.4 Método bayesiano baseado no modelo SIR estocástico

Nesta seção é apresentada uma abordagem bayesiana baseado no modelo SIR estocástico. O método é utilizado para estimar o número de reprodução efetivo, denotado geralmente por R_t , no lugar do R_0 . O número de reprodução efetivo é discutido amplamente na literatura, e existem diversas referências onde nós podemos encontrar sua formulação. Ver, por exemplo, Haydon et al. (2003), Wallinga e Teunis (2004), Cauchemez et al. (2006), Nishiura (2006), Bettencourt e Ribeiro (2008), Cowling et al. (2008) e Coelho et al. (2011).

Este parâmetro é definido como o número médio de casos secundários produzidos por um caso infectante no tempo t para $t > 0$.

A diferença entre R_0 e R_t depende das mudanças na estrutura da população, evolução de patógenos, variações da doença, declínio de indivíduos suscetíveis (fator intrínseco) e à implementação de medidas de controle (fator extrínseco), como é discutido em Bettencourt e Ribeiro (2008) e Nishiura e Chowell (2009).

A formulação do R_t leva em consideração a natureza probabilística do processo de contágio e permite uma estimativa direta da distribuição de probabilidade do número de reprodução efetivo R_t dos dados em tempo real sem precisar a pesquisa de parâmetros e otimização como foi feito nos métodos anteriores (Cauchemez et al., 2006; Bettencourt e Ribeiro, 2008). Neste sentido, este método, junto com os três métodos apresentados nas três seções anteriores, ajudam na estimativa do número de reprodução de forma complementar, como é discutido em Chowell et al. (2007).

O método pode ser aplicado a qualquer modelo compartimental, mas, como em Chowell et al. (2007), iremos considerar o modelo SIR apresentado na Seção 2.2.1. Chowell et al. (2007) chama a atenção que a aplicação a modelos, como o SEIR, é mais complexa. Como nos métodos anteriores, este método considera que são observados apenas os números de casos infectantes, e a formulação probabilística é dada em função da série temporal do número de casos infectantes novos, que é definido por:

$$\Delta C(t) = C(t) - C(t - \tau), \quad (3.4.1)$$

com $C(t), t = 1, 2, \dots, n$ definido na Seção 3.1.1 e τ o intervalo de tempo entre observações.

Considere inicialmente o modelo na forma determinística dado pelo sistema de equações 2.2.1. Os casos novos de indivíduos infectantes são os aqueles que saem do compartimento de suscetíveis; portanto:

$$C'(t) = -S'(t) = -\beta S(t)I(t)/N, \quad (3.4.2)$$

e $I'(t)/I(t) = \beta S(t)/N - \gamma$ que tem como solução

$$\begin{aligned} I(t + \tau) &= I(t) \exp \left[\int_t^{t+\tau} (\beta S(s)/N - \gamma) ds \right], \\ I(t + \tau) &= I(t) \exp \left[\gamma \int_t^{t+\tau} \left(\frac{\beta S(s)}{\gamma N} - 1 \right) ds \right]. \end{aligned}$$

Considerando que $R_0 = \beta/\gamma$ e $R_t = R_0 S(t)/N(t)$ e que $S(t)/N(t)$ é aproximadamente constante no intervalo $[t, t + \tau]$ temos

$$I(t + \tau) \approx I(t) \exp [\gamma \tau (R_t - 1)] \quad (3.4.3)$$

$$= b(R_t)I(t), \quad (3.4.4)$$

onde $b(R_t) = \exp [\gamma \tau (R_t - 1)]$.

Da equação 3.4.2, e utilizando 3.4.4, obtemos

$$\begin{aligned} \frac{C(t + \tau) - C(t)}{\tau} &= \frac{\beta S(t + \tau)}{N} I(t + \tau) \\ &\approx \frac{\beta S(t)}{N} b(R_t) I(t), \end{aligned} \quad (3.4.5)$$

em que supomos que $S(t)$ é aproximadamente constante no intervalo $[t, t + \tau]$, porque geralmente em doenças infecciosas, o número de suscetíveis na população é muito maior que o número de notificações diárias. τ pode variar entre intervalos sucessivos, contribuindo com a variabilidade de R_t . Além disso:

$$\frac{C(t) - C(t - \tau)}{\tau} = \frac{\beta S(t)}{N(t)} I(t), \quad (3.4.6)$$

e substituindo 3.4.6 em 3.4.5 obtemos:

$$\Delta C(t + \tau) = b(R_t) \Delta C(t). \quad (3.4.7)$$

A equação apresentada em 3.4.7 dá a evolução do número de casos novos no tempo t no modelo determinístico, que pode ser considerado como a evolução da média no modelo estocástico. Note

que o número de casos novos de epidemia depende dos casos notificados no tempo imediatamente anterior e do número de reprodução efetivo R_t .

Um modelo probabilístico é crucial para a modelagem de casos novos notificados quando os dados observados tem uma grande variabilidade em relação ao tempo. Considere que $\Delta(t + \tau)$ é uma variável aleatória discreta gerada por uma distribuição de probabilidade com parâmetro igual ao número médio de casos notificados no tempo t , dado por $\Delta C(t)$, isto é, supomos que:

$$\Delta(t + \tau)|R(t), \Delta(t) \sim P[\lambda], \quad (3.4.8)$$

em que $\lambda = b(R_t)\Delta(t)$ e $P[\lambda]$ é uma distribuição de probabilidade discreta com média λ . Chowell et al. (2007) sugerem a distribuição de Poisson. Outra possibilidade é a distribuição binomial negativa, mas que exige o conhecimento de outras informações.

O cálculo das distribuições a posteriori de R_t é realizado recursivamente, à medida que são observados os casos novos de epidemia $\Delta(t)$. Por questão de facilidade de notação considere que τ é sempre igual a 1. No início considere $\Pi_1(\cdot)$ a distribuição a priori de R_2 e denote por c_t o valor observado de casos novos no período $(t - 1, t]$. Observado $\Delta(2) = c_2$, temos que a distribuição a posteriori de R_2 , denotado por $\Pi_2(\cdot)$ é dada por:

$$\begin{aligned} \Pi_2(x) &= \frac{\Pi_1(x)P[\Delta_2 = c_2|\Delta_1 = c_1, \Pi_1(x)]}{P[\Delta_1 = c_1, \Delta_2 = c_2]} \\ &= \frac{\Pi_1(x)e^{-(b(x)c_1)}(b(x)c_1)^{c_2}}{c_2! \text{const}}, \end{aligned} \quad (3.4.9)$$

onde a constante de proporcionalidade *const* no denominador é dada por:

$$\text{const} = \int \frac{e^{-(b(x)c_1)}(b(x)c_1)^{c_2}}{c_2!} \Pi_1(x) dx.$$

Esta recursão é realizada até o final da fase inicial da epidemia. Observe que $\Pi_t(\cdot)$ denota a distribuição a posteriori de R_t quando se observa os números diários de casos novos de infecciosos até o dia t , e é a distribuição a priori de R_{t+1} . No caso em que $S(t)/N(t)$ é praticamente constante, como $R_t = R_0 S(t)/N(t)$, temos que $R_0 \approx R_t$, e podemos utilizar a distribuição a posteriori de R_t para estimar R_0 , inclusive construindo intervalos de confiança. Nesta dissertação é utilizada como distribuição a priori de R_2 uma normal $N(2, 1)$ truncada no intervalo $[0, 4]$. Como estimativa pontual será utilizada a moda da distribuição a posteriori e como intervalo de credibilidade $100(1 - \alpha)\%$, o intervalo

$$\left(F^{-1}(\alpha/2), F^{-1}(1 - \alpha/2)\right), \quad (3.4.10)$$

em que F é a função de distribuição à posteriori de R_t . Outras estimativas pontuais, como a média e a mediana poderiam ser utilizados. Intervalos de credibilidade de mais alta densidade poderia também ser considerado, mas nada garante que iríamos ter um intervalo contínuo.

A distribuição à posteriori foi calculada da seguinte forma:

1. **Passo 1:** Calcule 4001 valores de x igualmente espaçados no intervalo $[0, 4]$, isto é espaçados de 0.001. As densidades a posteriori serão calculadas apenas nestes pontos.
2. **Passo 2:** Calcule os valores da densidade a priori $\Pi_1(x)$ nestes 4001 pontos. Eles são os valores da densidade da distribuição normal $N(2, 1)$ truncada no intervalo $[0, 4]$.
3. **Passo 3:** Calcule os valores da densidade a posteriori $\Pi_2(x)$ dada pela equação 3.4.9 sem a constante

$$num(x) = \frac{\Pi_1(x)e^{-(b(x)c_1)} (b(x)c_1)^{c_2}}{c_2!}, \quad x = 0,001k, k = 0, 1, \dots, 4000.$$

Calcule a constante do denominador através de

$$const = 0,001 \sum_{k=0}^{4000} \frac{\Pi_1(0,001k)e^{-(b(0,001k)c_1)} (b(0,001k)c_1)^{c_2}}{c_2!}.$$

4. **Passo 4:** Calcule as densidades a posteriori

$$\Pi_2(x) = num(x)/const.$$

5. **Passo 5:** Repita os passos 3 a 4 substituindo $\Pi_1(x)$ e c_1 por $\Pi_t(x)$ e c_t e $\Pi_2(x)$ e c_2 por $\Pi_{t+1}(x)$ e c_{t+1} para $t = 2, \dots, T - 1$, onde T é a última observação do período inicial da epidemia. No final teremos $\Pi_T(x)$. Note que $\Pi_t(\cdot)$ é utilizado como distribuição a priori de $\Pi_{t+1}(\cdot)$
6. **Passo 5:** A estimativa pontual vai ser o valor de x que dá o maior valor de $\Pi_T(x)$, $x = 0; 0,001; \dots, 4$, e os intervalos de credibilidade utilizando os valores de $\Pi_T(x)$.

Neste método, pode-se utilizar as distribuições a posteriori para verificar se houve mudanças na evolução da epidemia. Seja $\Pi_T(\cdot)$ a distribuição a posteriori de R_T no tempo T , a qual é a distribuição a priori no tempo $(T + 1)$. Seja c_T , o número de caso novos notificados no período $[T - 1, T]$. Considere que no intervalo $[T, T + 1]$ sejam observados k casos novos. Se $P[\Delta(T + 1) \in [k - a, k + b] | \Pi_T, \Delta(T) = c_T]$ for muito pequeno, temos indicações de que houve mudanças. Os valores de a e b podem ser escolhidos através da distribuição de $\Pi_T(\cdot)$.

Capítulo 4

Simulações

Neste capítulo utilizaremos simulação para comparar três dos quatro métodos para estimação da taxa de reprodução R_0 apresentados no Capítulo 3. Não será considerado no estudo das simulação o método 3, que é baseado no modelo SEIR estendido. Os métodos que serão utilizados para obter as estimativas podem ser classificados segundo os dados considerados na estimação, a suposição do modelo subjacente e suposições em relação a certos parâmetros.

A classificação dos métodos segundo:

Dados considerados: Os três métodos utilizados no processo de estimação do R_0 consideram apenas os dados relativos à primeira fase de crescimento da epidemia, onde supostamente ainda não tem intervenção e/ou está na fase inicial da epidemia. Como o término desta fase não é conhecida, estes métodos incluem um estágio inicial de escolha do período.

Modelos subjacentes: O método 4 considera um modelo SIR estocástico subjacente, enquanto os métodos 1 e 2 consideram o modelo SIR ou SEIR determinístico.

Suposição sobre o parâmetros: É bastante comum considerar alguns dos parâmetros como conhecidos. O modelo, bem como os valores dos parâmetros selecionados da literatura, dependerão da epidemia sob estudo. Por exemplo, Chowell et al. (2007), no método 2 considerou $\gamma = 1/(4,1 \text{ dias})$ e $\kappa = 1/(1,9 \text{ dias})$ e, no método 3, $\mu = 4,56621e - 05$, com γ , κ e μ definidos nas Seções 2.2, 2.3 e 2.4, respectivamente, com o tempo em dias.

Um resumo dos métodos estudados neste capítulo em relação a estes critérios é dado na Tabela 4.1. Nesta tabela não consta o método utilizado para estimação por intervalos.

Tabela 4.1: Resumo dos estimadores utilizados na simulação segundo o modelo subjacente, método utilizado na seleção do período inicial e parâmetros conhecidos e parâmetros a ser estimados

Método	Modelo subjacente	Seleção dados	Parâmetros estimados	Parâmetros conhecidos
1	SIR	T_1 e T_{2b}	r	γ
	SEIR	T_1 e T_{2a}	r	γ
2	SIR	T_1 e T_{2b}	$\beta, I(0)$	γ
	SEIR	T_1 e T_{2a}	$\beta, I(0)$	γ, κ
4	SIR	T_1 e T_{2b}	β	γ

O objetivo das simulações apresentadas neste capítulo é estudar as propriedades estatísticas dos estimadores dados pelos métodos 1, 2 e 4. Serão considerados apenas os casos em que o processo gerador dos dados é um modelo SIR estocástico simples ou SEIR estocástico simples, estes modelos são as versões estocásticas dos modelos SIR determinístico e SEIR determinísticos apresentados nas seções 2.2.1 e 2.3. O capítulo é organizado da seguinte forma. A Seção 4.1 apresenta os algoritmos de geração de dados. A Seção 4.2 discute o caso em que o processo gerador dos dados é um modelo SIR estocástico. A Seção 4.3 discute o caso em que o processo gerador dos dados é um modelo SEIR estocástico. Serão consideradas as seguintes propriedades dos estimadores: vício, erro quadrático médio, cobertura e largura média dos intervalos de confiança (de confiabilidade no caso bayesiano).

4.1 Algoritmo para geração dos dados

Nesta seção nós apresentamos os algoritmos para os processos geradores dos dados que serão utilizados para o estudo desenvolvido nesta dissertação. O modelo SIR é um caso particular do modelo SEIR. A justificativa teórica foi apresentada na Seção 2.2.2 e é baseado nas propriedades da distribuição exponencial e sua relação com processos de Poisson. Para simplificar consideramos que não existe mudança nos valores dos parâmetros. As notações são as mesmas das apresentadas na Seção 2.3, onde N é o tamanho da população e $S(t)$, $E(t)$, $I(t)$ e $R(t)$ é número de indivíduos suscetíveis, latentes, infectantes e recuperados no tempo t , respectivamente. Estes valores permanecem os mesmos até a próxima transição que ocorre no tempo $t + \Delta$, onde Δ é o comprimento de tempo para que a próxima transição ocorra, e os novos valores dos números de indivíduos em cada compartimento é dado pelos seguintes passos:

1. Seja $\lambda_{SE}(t) = \beta S(t)I(t)/N(t)$, $\lambda_{EI}(t) = \kappa E(t)$ e $\lambda_{IR}(t) = \gamma I(t)$ as taxas de ocorrência de uma transição de suscetível a latente, latente a infectante e de infectante a recuperado, respectivamente. Defina $\lambda(t) = \lambda_{SE}(t) + \lambda_{EI}(t) + \lambda_{IR}(t)$ a taxa de falha total. Estas taxas permanecem constantes até a próxima transição.
2. Gere Δ , o tempo até a próxima transição, através da distribuição exponencial com taxa de falha $\lambda(t)$.
3. Gere um valor a de uma distribuição uniforme $U(0, \lambda(t))$.
4. Se $a < \lambda_{SE}(t)$, então $S(t + \Delta) = S(t) - 1$, $E(t + \Delta) = E(t) + 1$, $I(t + \Delta) = I(t)$ e $R(t + \Delta) = R(t)$; caso contrário, se $a < \lambda_{SE}(t) + \lambda_{EI}(t)$, então $S(t + \Delta) = S(t)$, $E(t + \Delta) = E(t) - 1$, $I(t + \Delta) = I(t) + 1$ e $R(t + \Delta) = R(t)$; caso contrário, $S(t + \Delta) = S(t)$, $E(t + \Delta) = E(t)$, $I(t + \Delta) = I(t) - 1$ e $R(t + \Delta) = R(t) + 1$.

Esta recursão termina quando o número de indivíduos suscetíveis e infectantes é igual a zero, ou se atingiu o tempo máximo de observação da epidemia.

O algoritmo é iniciado com valores iniciais do número de indivíduos em cada compartimento. No final consideramos os valores que foram considerados observados, por exemplo, nas simulações apresentadas para o estudo consideramos que são observados apenas o número de casos acumulados ao final de cada dia, isto é, são observados apenas $C_o(t) = C(t - 1)$, $t = 1, 2, 3, \dots$

4.2 Processo gerador de dados: SIR

Consideramos como o primeiro processo gerador dos dados, uma abordagem estocástica em tempo discreto do modelo epidemiológico determinístico SIR dado pelo sistema de equações 2.2.1. Esta abordagem estocástica foi tratada de forma breve na Subseção 2.2.2 no modelo baseado em agentes.

A escolha dos parâmetros utilizou as observações da série de influenza observadas durante a epidemia de 1918-1919 na Califórnia e analisada por Chowell et al. (2007) como referência, além dos valores dos parâmetros utilizados no mesmo artigo. Assim, temos o tamanho da população $N = 550000$ e a taxa de remoção dada por $\gamma = 1/(4,1 \text{ dias})$. O valor da taxa de transmissão β e $I(0)$ foram escolhidos de forma que o número de casos novos dados pelo modelo SIR determinístico associado, no trigésimo dia, fosse próximo do número máximo de casos novos observados na série de influenza no trigésimo terceiro dia. Após alguns testes foi escolhido $\beta = 0,4865$ e $I(0) = 5$, que equivale a $R_0 = 1,995$. Após a intervenção nós consideramos $\beta = 0,17$. Este novo valor do β foi estabelecido de forma que o número de casos novos após a intervenção, dado pelo modelo SIR determinístico, decresçam de forma semelhante aos observados nos dados reais da série de influenza. O gráfico do número de notificações diárias de influenza, bem como o dado pelo modelo determinístico com os valores selecionados é apresentado na Figura 4.1.

Na simulação das séries consideramos que a epidemia é observada no máximo durante 60 dias, com intervenção no trigésimo dia. Foram consideradas apenas séries geradas que continham pelo menos 10 dias de observação. Para obter as 1000 séries de epidemia foi necessário gerar 1045 séries, ou seja, em 45 das 1045 séries simuladas, a epidemia durou menos do que 10 dias. Destas 1000 séries de observações simuladas, oito das séries de epidemias não haviam terminado até o sexagésimo dia.

Para este conjunto de observações, o parâmetro R_0 foi estimado inicialmente através dos seguintes métodos:

1. Método 1 com a fase inicial da epidemia selecionada pelos critérios T_1 (ver Seção 3.1.1) e T_{2b} (ver Seção 3.2), e a relação entre R_0 e a taxa de crescimento exponencial dada pela Equação 3.1.3. Neste método é estimado apenas a taxa de crescimento exponencial r , e considerado conhecido a taxa de remoção γ . Para ter no mínimo dois graus de liberdade, assim com Chowell et al. (2007), é utilizado T_1 e T_{2b} no mínimo igual a quatro.
2. Método 2 com modelo SIR determinístico com T_1 e T_{2b} selecionados como o período para estimação do modelo. Neste método são estimados a taxa de transmissão β e o número de infectantes iniciais, e considerado conhecido a taxa de remoção γ . Analogamente ao método 1 utilizou-se T_1 e T_{2b} no mínimo igual a quatro.
3. Método 4 utilizando as observações até os tempos T_1 e T_{2b} . Como distribuição a priori foi utilizada a mesma distribuição utilizada por Chowell et al. (2007), ver Seção 3.4.

4.2.1 Análise inicial dos estimadores

Primeiro iremos realizar uma análise dos estimadores da forma proposta inicialmente e, a partir destas análises propor modificações. A Figura 4.2 mostra a distribuição de frequências dos períodos

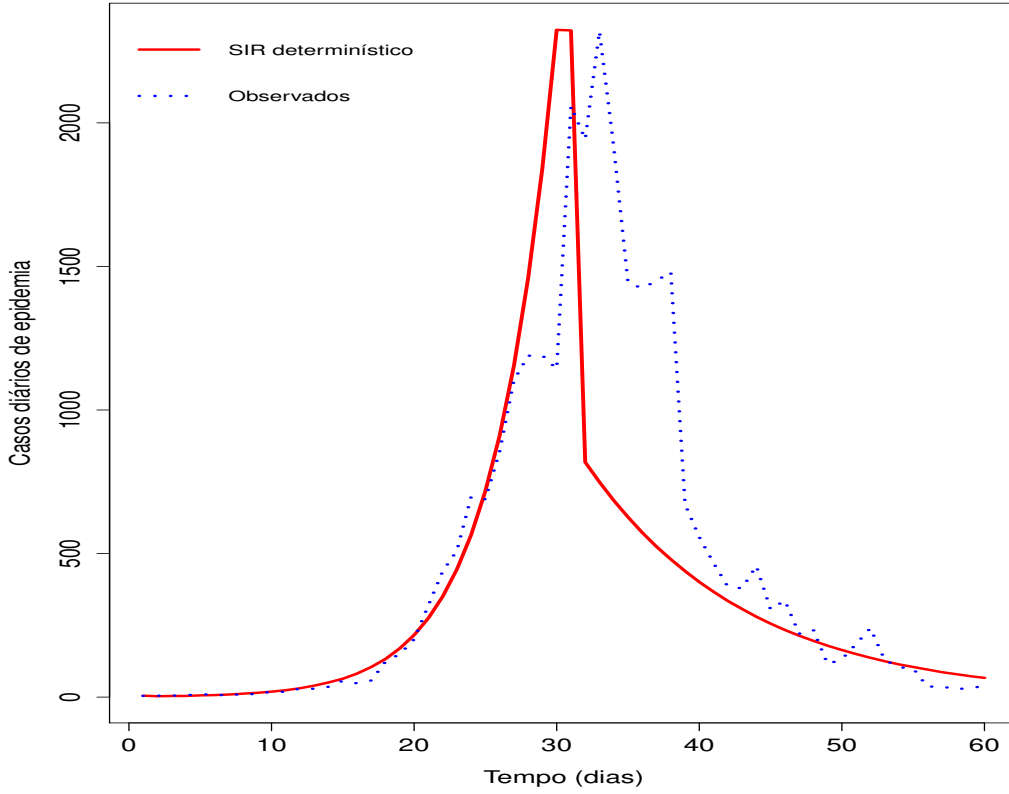


Figura 4.1: Número de casos novos diários observados na epidemia de influenza na Califórnia em 1918-1919 e dados pelo modelo determinístico SIR com os valores dos parâmetros utilizados para gerar as séries simuladas pelo modelo SIR estocástico.

iniciais selecionados pelos critérios T_1 e T_{2b} . Existe uma leve tendência dos valores de T_1 serem maiores do que T_{2b} , em 46,1% dos casos T_1 é maior do que T_{2b} ; em 26,4% dos casos os dois valores são iguais e nos restantes 27,1% dos casos T_1 é menor do que T_{2b} . A correlação entre os dois foi de 0,351. Em nenhum dos critérios de seleção foi escolhido um valor para a fase de crescimento exponencial maior do que 30 dias, quando ocorre a mudança dos valores dos parâmetros. As Figuras 4.3 a 4.5 mostram os box plots¹ das estimativas de R_0 pelos métodos 1, 2 e 4, respectivamente, segundo a fase inicial da epidemia selecionada pelos critérios T_1 e T_{2b} . A Tabela 4.2 apresenta as estatísticas descritivas das estimativas segundo o número de observações selecionadas para cada série. Nos três métodos as estimativas tendem a ser melhores para as séries onde foram selecionadas maiores fases iniciais da epidemia. A exceção ocorre no método 4, porque quando o período vai até 29 ou 30 dias o desempenho do estimador diminui, tendendo a subestimar R_0 . É esperado que R_t deve diminuir com o tempo, mas a diminuição é maior do que o previsto pela diminuição que ocorre no número de suscetíveis. Os resultados indicam claramente que as estimativas são piores para as séries para as quais o período de crescimento selecionado é pequeno, com clara super-estimação pelos métodos

¹Utilizaremos o nome em inglês, sem colocar em itálico porque o seu uso é bastante difundido

1 e 2 (no método 2 até período igual a 6), e maior variabilidade pelos três métodos. Isto vale tanto pelo critério T_1 quanto T_{2b} para a seleção da fase. Para evitar este problema as estimativas foram refeitas tomando como mínimo de 10 observações. Pelo critério T_1 são 46,9% dos casos com até nove observações, enquanto pelo critério T_{2b} são 61,8% dos casos. Para estas séries foram refeitas as estimações considerando o mínimo de 10 observações para o período inicial da epidemia. Nas 469 séries com $T_1 \leq 9$, para o método 1, os quartis das estimativas com a restrição $T_1 \geq 4$ eram iguais a (2,078; 2,421; 2,374) e passaram a (1,970; 2,135; 2,263) e o EQM diminuiu de 0,345 a 0,159, enquanto para o método 2, os quartis das estimativas com a restrição $T_1 \geq 4$ eram iguais a (1,735; 2,179; 2,396) e passaram a (1,894; 2,066; 2,202) e o EQM diminuiu de 0,345 a 0,089. Já, para as 618 séries com $T_{2b} \leq 9$, para o método 1, os quartis das estimativas com a restrição $T_1 \geq 4$ eram iguais a (2,190; 2,461; 2,694) e passaram a (1,976; 2,117; 2,235) e o EQM diminuiu de 0,346 a 0,049, enquanto para o método 2, os quartis das estimativas com a restrição $T_{2b} \geq 4$ eram iguais a (1,833; 2,163; 2,425) e passaram a (1,847; 2,020; 2,171) e o EQM diminuiu de 0,261 a 0,075. Estes resultados comprovam que, pelo menos quando a restrição de tempo mínimo igual a 10 não é inadequada, os resultados das estimativas melhoram bastante. Desta forma, a primeira modificação dos estimadores propostos por Chowell et al.(2007) é aumentar o número mínimo de observações de quatro para dez.

Na metodologia apresentada na Seção 3.1.4 foi considerado que em cada série bootstrap é utilizado o mesmo período inicial da epidemia. Inicialmente foi realizado um experimento com o método 1, com o critério T_1 para selecionar o período inicial da epidemia. Neste experimento foram simuladas 100 séries e os intervalos de confiança obtidos da forma tradicional, como descrito na Seção 3.1.4, com 1000 réplicas bootstrap para cada série simulada. A cobertura dos intervalos de confiança 90%, foi de apenas 20,0% considerando todas as 100 séries. Considerando as 80 séries onde $T_1 \leq 20$ a cobertura foi de 18,75%, e entre as 20 séries restantes onde $T_1 \geq 21$, a cobertura foi de 25,0%. Para aumentar a largura dos intervalos e melhorar a cobertura, com as mesmas 100 séries simuladas foram geradas 1000 réplicas bootstrap para cada série simulada, mas para cada réplica bootstrap foram escolhidos novamente os períodos iniciais da epidemia. Embora as coberturas ainda estejam longe do valor nominal de 90%, em todos os casos houve um aumento da cobertura. Os valores de 20,0%, 18,7% e 25,0% foram aumentados para 38,8%, 30,0% e 70,0%, respectivamente. Este aumento ainda está longe do adequado e, como o vício não é grande, a baixa cobertura deve-se à pouca variabilidade das estimativas bootstrap. Outro problema verificado foi a alta frequência de estimativas pontuais fora dos intervalos bootstrap. Por exemplo, no método 1, considerando o bootstrap tradicional, e considerando a fase de crescimento exponencial nos intervalos [10, 20], [21, 30] e [10, 30], as porcentagens encontradas foram de 34,0%, 33,0% e 30,0%, respectivamente, e na modificada foram de 37,5%, 40,0% e 38,0%, respectivamente. A utilização da média ou mediana das estimativas bootstrap como estimadores pontuais resolve este problema, mas nestas alternativas encontramos maiores vícios e EQM.

Para aumentar a cobertura dos intervalos de confiança foram testadas duas modificações. Como a distribuição da soma de variáveis aleatórias independentes com distribuição de Poisson também é uma Poisson, temos que, nas amostras bootstrap, $C_{oi}(t) - C_o(1)$ tem distribuição Poisson com média $C_o(t) - C_o(1)$ e desvio padrão $\sqrt{C_o(t) - C_o(1)}$, e tudo indica que é necessário aumentar a variabilidade. Para aumentar a variabilidade foram testadas duas alternativas. Na Seção 3.1.4 $C'_{oi}(t) = C_{oi}(t) - C_{oi}(t - 1)$ é obtida de uma distribuição de Poisson com média $C'_o(t) = C_o(t) -$

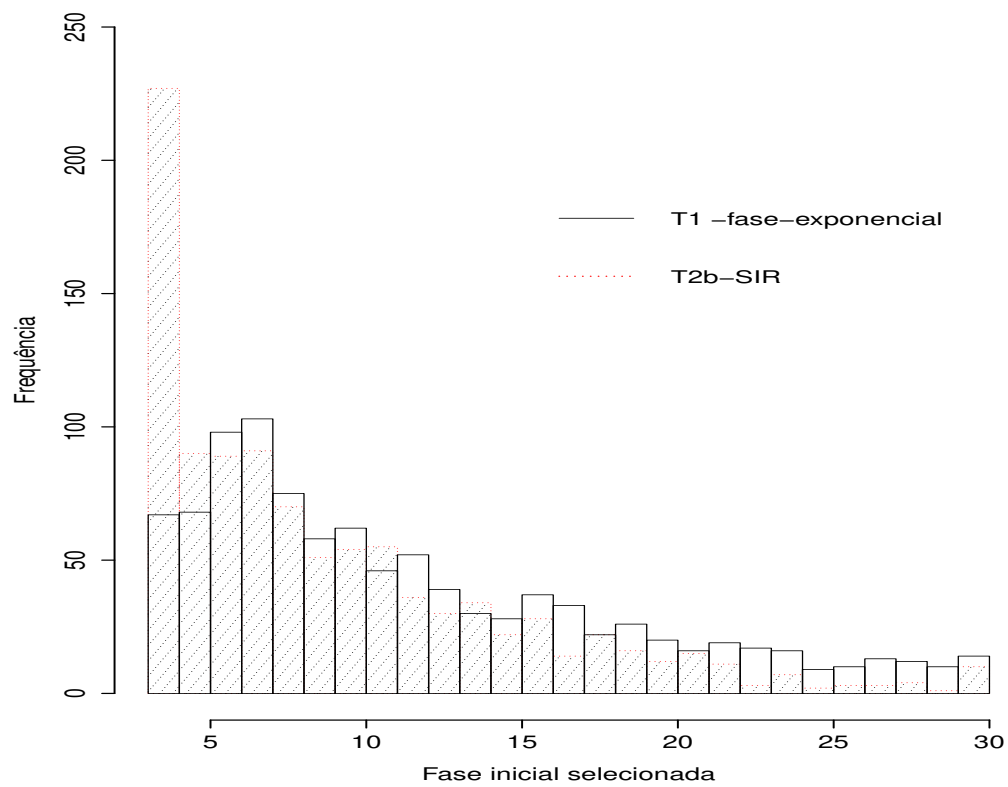


Figura 4.2: Frequência da fase inicial da epidemia selecionada pelos critérios T_1 e T_{2b} nas 1000 séries geradas pelo modelo SIR estocástico.

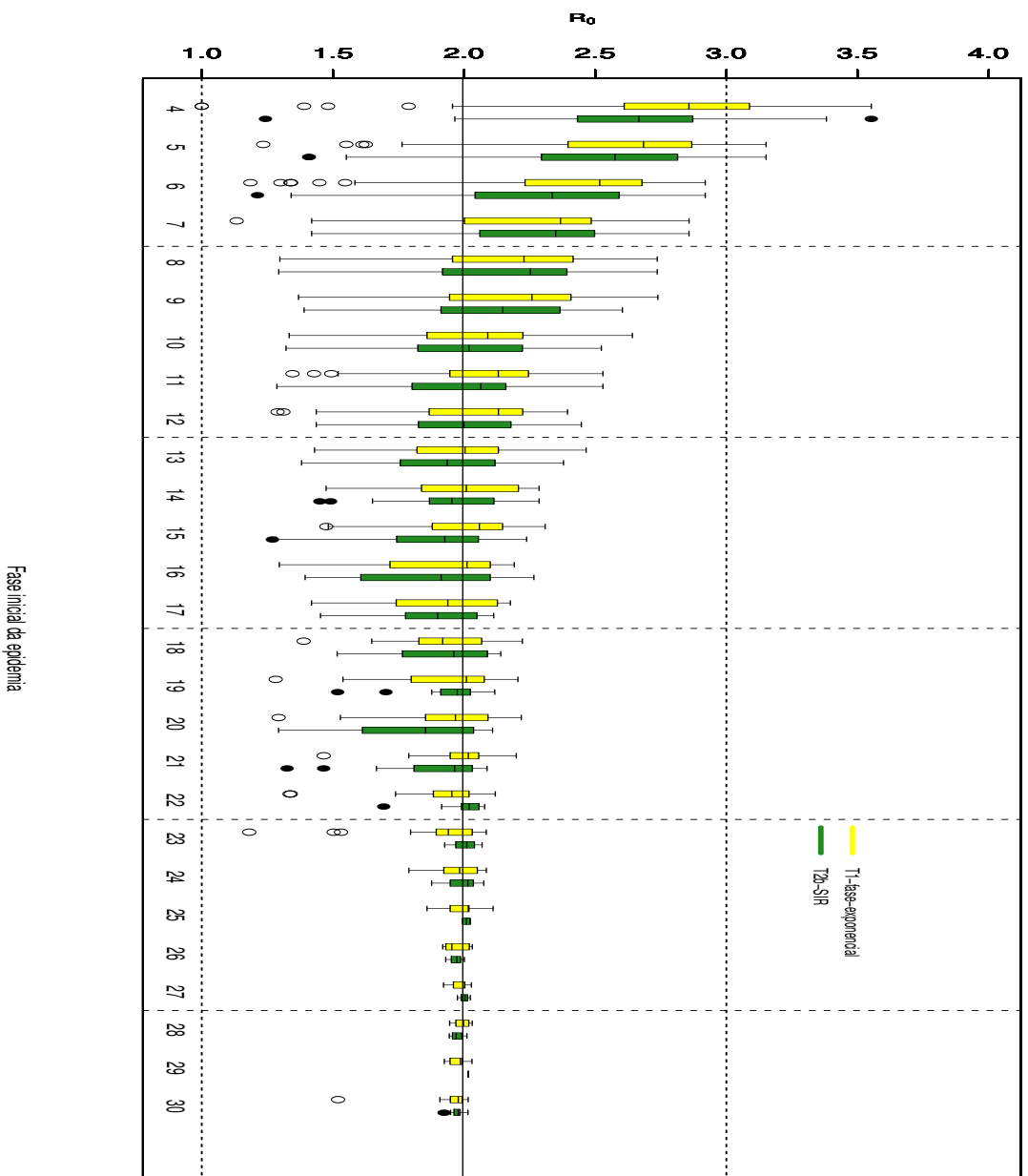
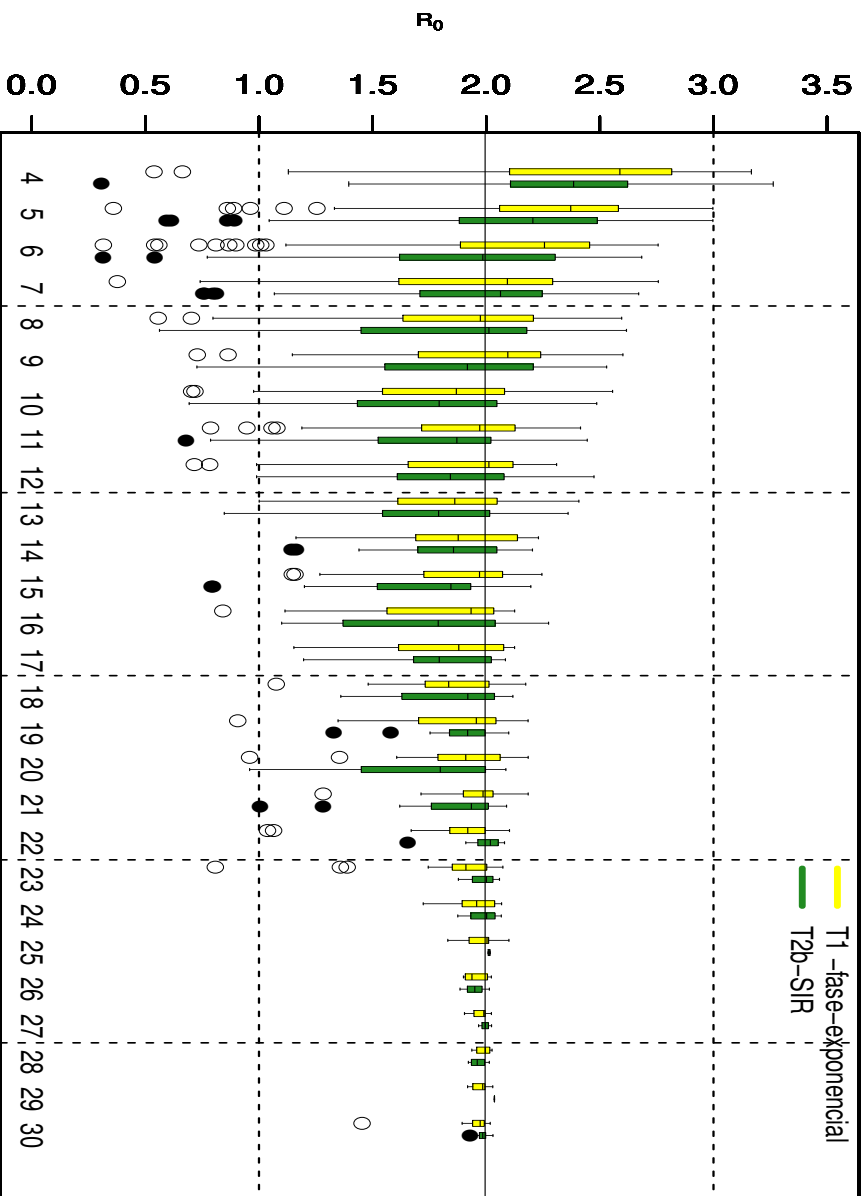


Figura 4.3: Box plot das estimativas de R_0 pelo método 1, segundo a fase inicial da epidemia selecionada pelos critérios T_1 e T_2a . Baseados em 1000 séries geradas pelo modelo SIR estocástico. A linha cheia indica o verdadeiro valor de R_0 .



Fase inicial da epidemia

Figura 4.4: Box plot das estimativas de R_0 pelo método 2, segundo a fase inicial da epidemia seleccionada pelos critérios T_1 e T_{2b} . Baseados em 1000 séries geradas pelo modelo SIR estocástico. A linha horizontal cheia indica o verdadeiro valor de R_0 .

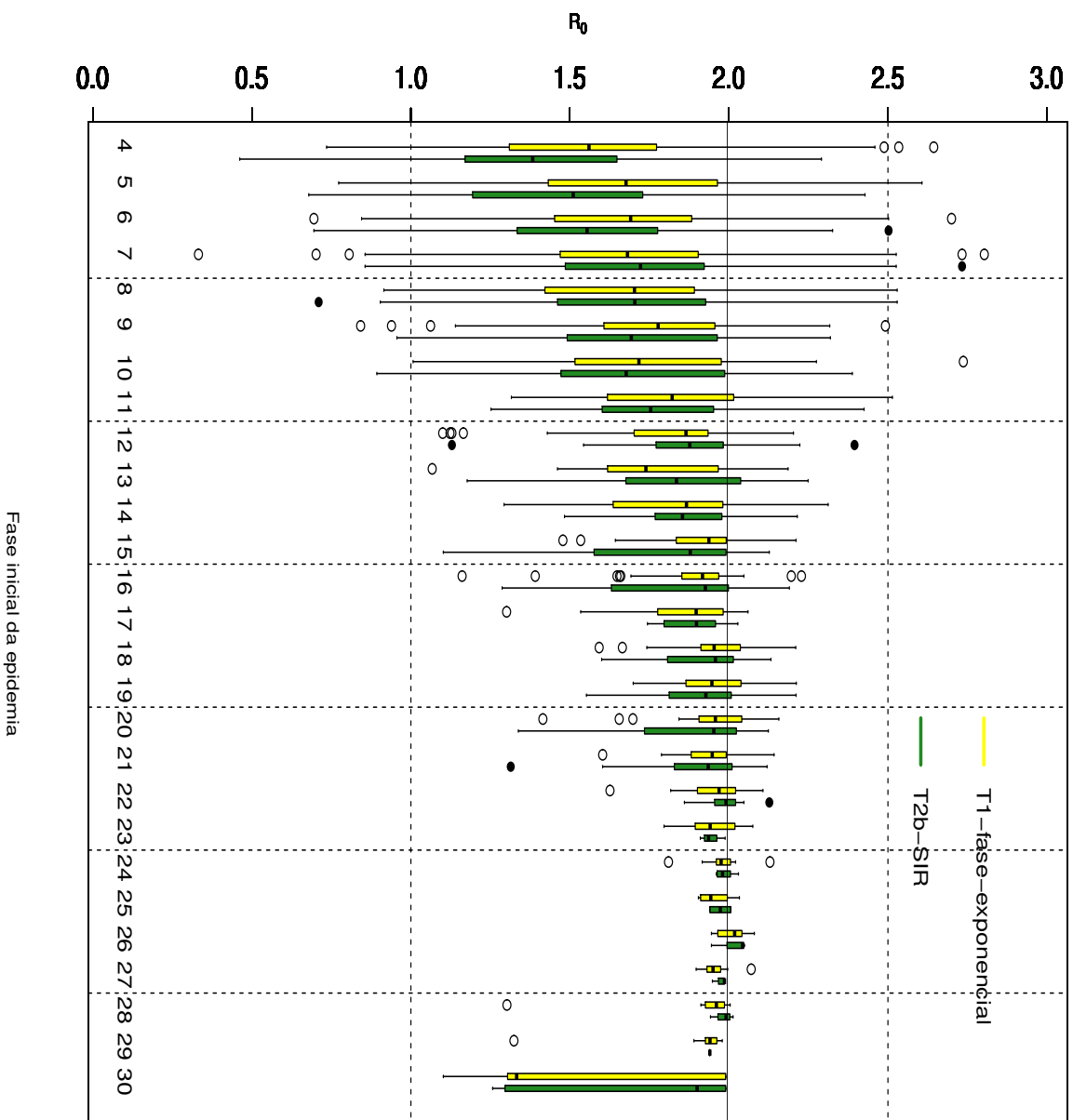


Figura 4.5: Box plot das estimativas de R_0 pelo método 4, segundo a fase inicial da epidemia selecionada pelos critérios T_1 e T_2 . Baseados em 1000 séries geradas pelo modelo SIR estocástico. A linha horizontal cheia indica o verdadeiro valor de R_0 .

$C_o(t - 1)$. A alternativa agora é gerar $C'_{oi}(t)$ como uma mistura de distribuições de Poisson, onde obteremos $C'_{oi}(t)$ de uma Poisson com média μ_{oi} , isto é, μ_{oi} não é mais um valor fixo e igual a $C'_o(t)$, mas amostrada de uma distribuição com média $C'_o(t)$. Foram testadas duas alternativas. Na primeira, $\mu_{oi}(t)$ é amostrada de uma distribuição de Poisson com média $C'_o(t)$, e na segunda alternativa de uma distribuição gama com média e desvio padrão $C'_o(t)$. Na primeira alternativa foram construídos intervalos de confiança do 90% para 200 séries com 200 réplicas bootstrap para cada série.

Para a fase de crescimento exponencial nos intervalos $[10, 20]$, $[21, 30]$ e $[10, 30]$, as percentagens encontradas nesta primeira alternativa foram de 42,8%, 60,9% e 49,5%, respectivamente, enquanto na segunda alternativa, considerando 1000 séries e 1000 réplicas bootstrap para cada série, estas percentagens aumentaram para 69,9%, 88,8% e 73,6%, respectivamente. Embora o resultado para o intervalo $[10, 20]$ para o crescimento exponencial não seja boa, esta modificação mostra-se promissora.

Ao mesmo tempo, na primeira alternativa 22,0% das estimativas pontuais encontram-se dentro dos intervalos de confiança bootstrap, enquanto na segunda alternativa esta percentagem diminuiu para 3,8%.

Resultados semelhantes ocorreram para o método 2 e para outros intervalos. Desta forma, nos métodos 1 e 2 as réplicas bootstrap são construídas utilizando a segunda alternativa e, em cada réplica bootstrap, é escolhido o período inicial da epidemia. Na seção seguinte são investigadas as propriedades destes novos estimadores e do estimador bayesiano dado pelo método 4.

4.2.2 Estudo comparativo dos estimadores propostos

Nesta subseção são apresentados os resultados da simulação utilizando os estimadores dado na Tabela 4.1 com a restrição de número mínimo de 10 observações no período inicial da epidemia selecionado, e com os intervalos de confiança construídos conforme dado na subseção anterior. Um sumário de todas as estimativas são apresentadas na Tabela 4.3 e os box plots nas Figuras 4.6 a 4.9. A classificação do tempo selecionado para o período inicial da epidemia é segundo o tempo das séries simuladas consideradas observadas. Observe que nas séries bootstrap os períodos podem ser diferentes. Na maioria das séries os critérios selecionaram, para as séries observadas, um período inicial da epidemia menor ou igual a 20. As percentagens foram iguais a 80,3% e 91,5% para os critérios T_1 e T_{2b} , respectivamente. Todos os resultados para estimativas pontuais e para os intervalos de credibilidade do método 4 são baseados em 1000 replicações. Na construção dos intervalos de confiança para as estimativas do método 1 foram consideradas 1000 replicações e para cada série simulada foram construídas 1000 réplicas bootstrap, enquanto no método 2, os resultados para intervalos de confiança são baseadas em 200 séries com 200 réplicas bootstrap para cada série. Como nas estimativas pontuais não foi selecionado nenhuma vez um período inicial maior do que 30, embora para cada série fossem testados todos os intervalos de 10 a 60 dias, por questão de tempo computacional, nas séries bootstrap foram testadas apenas intervalos de 10 a 30 dias.

Como sabemos que ocorrem mudanças dos valores dos parâmetros após o trigésimo dia, foram também obtidas as estimativas pelos três métodos considerando o período de estimação até o trigésimo dia. Neste caso, tanto para o método 1 como para o método 2 nas estimativas pontuais

e por intervalos foram consideradas 1000 replicações e, para cada série simulada foram construídas 1000 réplicas bootstrap para estimação por intervalos.

A Tabela 4.4 apresenta a matriz de covariância/correlação das estimativas dos métodos considerados.

A análise dos resultados indica que:

- Nos métodos 1 e 2 os melhores resultados ocorrem quando consideramos a fase inicial da epidemia até o trigésimo dia.
- O aumento do número mínimo de observações, de quatro para dez, melhorou bastante as estimativas.
- Quando é necessário escolher a fase inicial da epidemia, o critério T_1 , com valor mínimo 10, foi o que apresentou os melhores resultados.
- Quando é necessário escolher a fase inicial da epidemia, utilizando-se o o critério T_1 , com valor mínimo 10, os três métodos, em termos de EQM, são muito próximos, mas o método 1 tende a dar valores menos extremos.
- As diferenças entre a qualidade das estimativas utilizando como fase de crescimento exponencial, o trigésimo dia, indica a necessidade de se encontrar um critério melhor para selecionar o período.
- Quando as observações consideradas nas estimativas são as mesmas, existe uma grande correlação entre os métodos 1 e 2, com os critérios T_1 e T_{2b} de escolha do período inicial. Como T_1 e T_{2b} são próximos, temos indicação de que os métodos 1 e 2 dão resultados próximos quando as observações utilizadas são aproximadamente as mesmas. No método 1, as correlações considerando os períodos selecionados T_1 ou T_{2b} , com as estimativas considerando período até o trigésimo dia são baixas. Isto ocorre porque T_1 e T_{2b} são geralmente pequenos, levando a um pequeno número de observações utilizadas. Isto enfatiza a necessidade de encontrar um critério melhor para selecionar o período inicial da epidemia.
- Embora a moda da distribuição a priori utilizada no método 4 seja próxima do valor verdadeiro, seu desvio padrão, igual a 1,0, é quase três vezes maior do que a variabilidade das estimativas. Podemos considerar uma priori pouco informativa, mas é interessante realizar outros testes de sensibilidade utilizando outras distribuições a priori. Para a primeira série simulada foram também consideradas as seguintes distribuições a priori: normal $N(1; 1)$ truncada no intervalo $[0, 4]$, normal $N(3; 2^2)$ truncada no intervalo $[0, 6]$ e gama com média 2,0 e desvio padrão 2,0 truncada em $[0, 6]$. As distribuições priori e posteriori estão dadas na Figura 4.10. Embora as distribuições a priori sejam bastante diferentes, as distribuições a posteriori são bastante próximas mesmo considerando apenas 10 observações. Considerando 20 observações, as diferenças são imperceptíveis a olho nu em um gráfico com a mesma escala. Podemos considerar que, pelo menos para os valores dos parâmetros do modelo considerado, os resultados não são sensíveis à distribuição a priori, caso sejam consideradas pelo menos 10 observações.

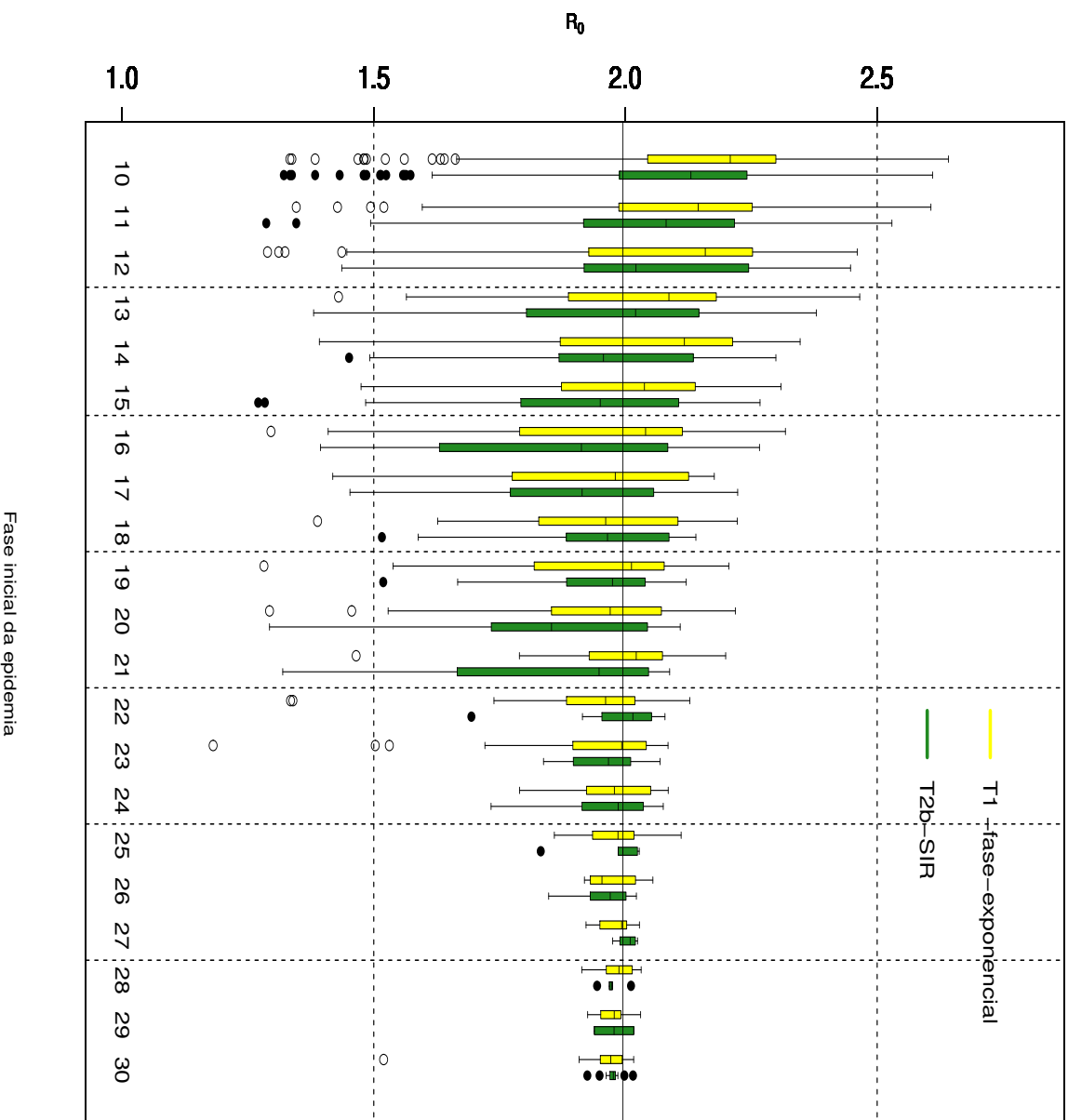


Figura 4.6: Box plot das estimativas de R_0 pelo método 1 segundo a fase inicial da epidemia selecionada pelos critérios T_1 e T_{2b} , com número mínimo de observações igual a 10. Baseados em 1000 séries geradas pelo modelo SIR estocástico. A linha horizontal cheia indica o verdadeiro valor de R_0 .

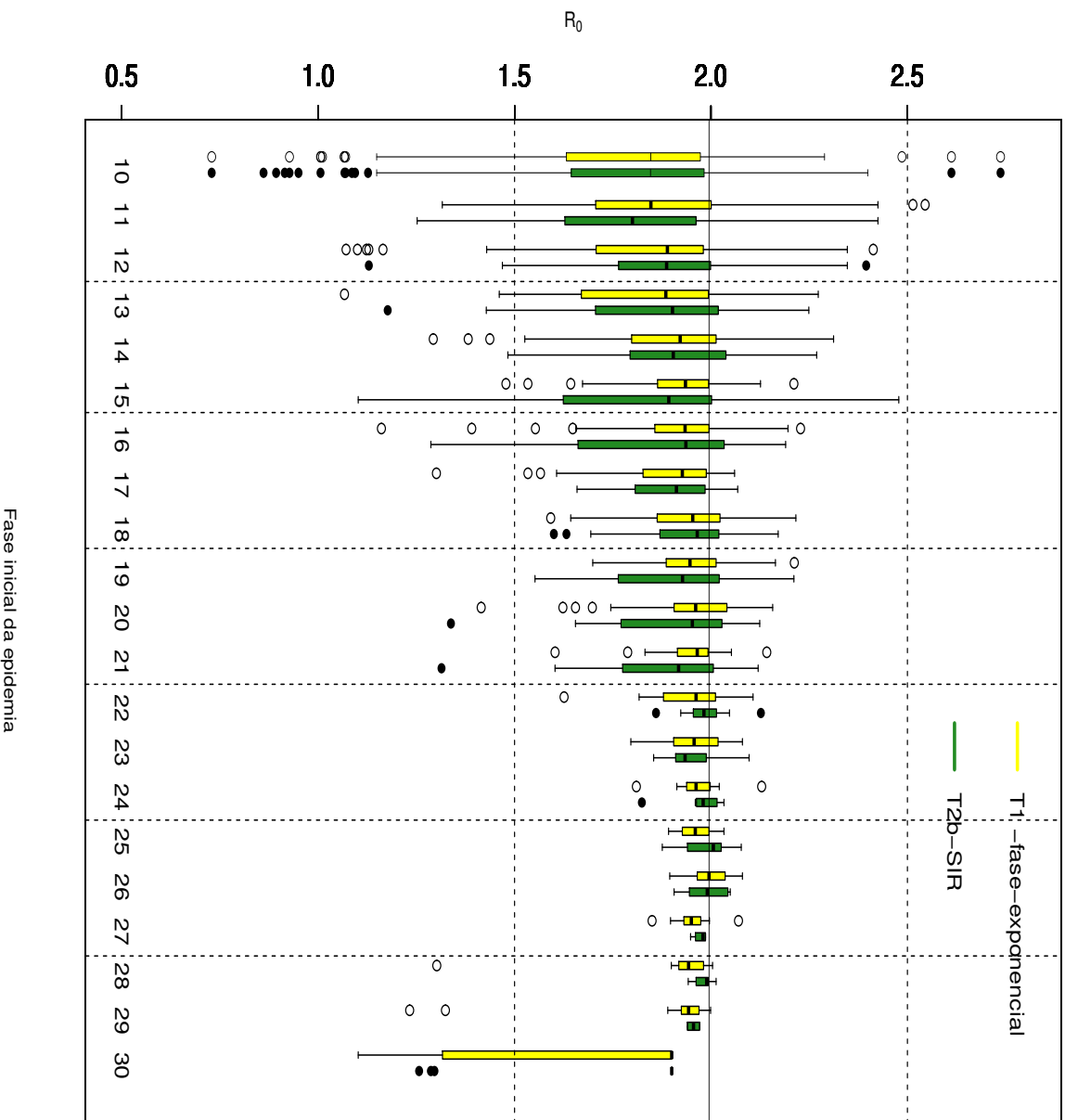


Figura 4.7: Box plot das estimativas de R_0 pelo método 2 segundo a fase inicial da epidemia selecionada pelos critérios T_1 e T_{2b} , com número mínimo de observações igual a 10. Baseados em 1000 séries geradas pelo modelo SIR estocástico. A linha horizontal cheia indica o verdadeiro valor de R_0 .

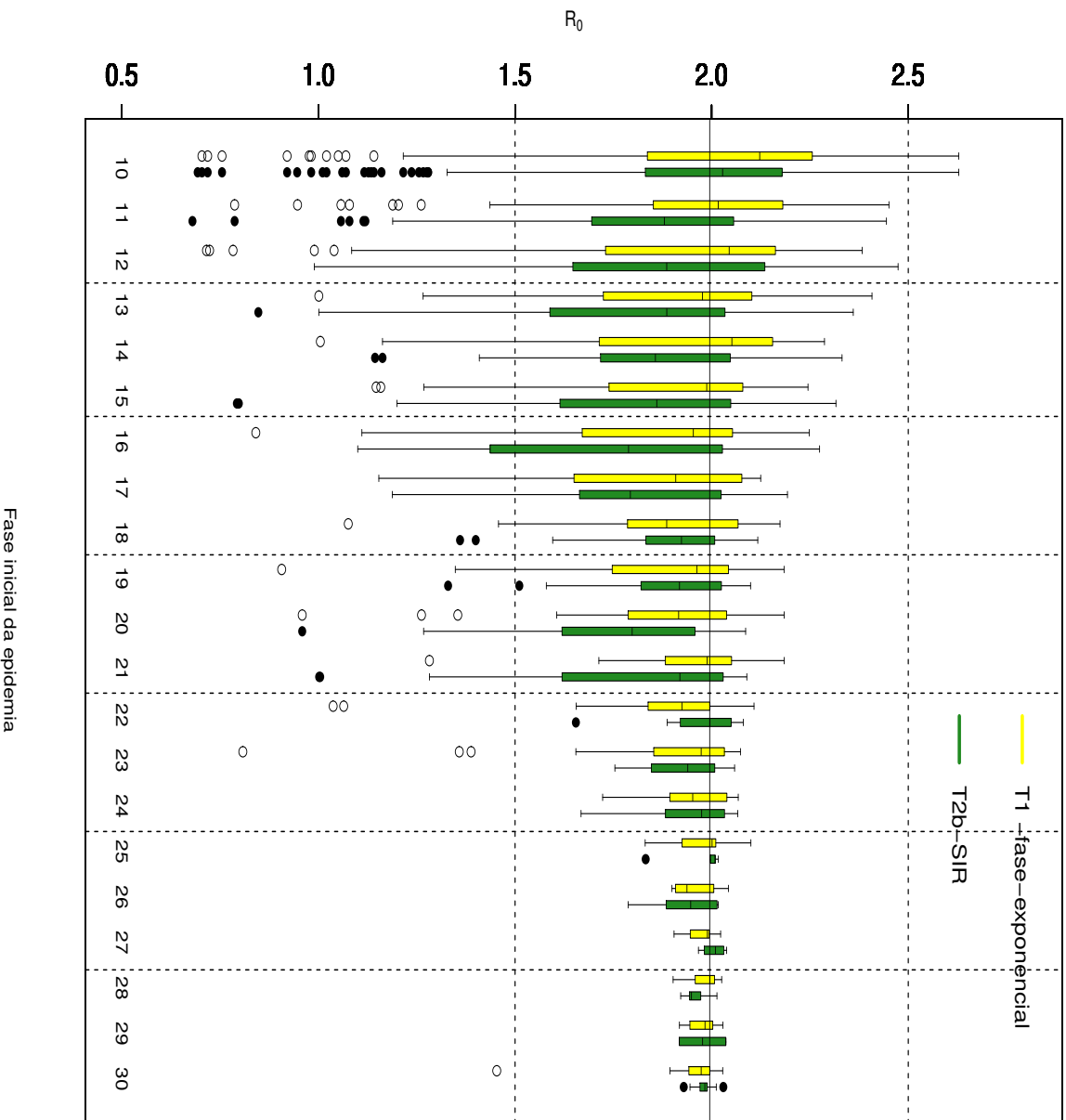


Figura 4.8: Box plot das estimativas de R_0 pelo método 4, segundo a fase inicial da epidemia selecionada pelos critérios T_1 e T_{2b} , com número mínimo de observações igual a 10. Baseados em 1000 séries geradas pelo modelo SIR estocástico. A linha horizontal cheia indica o verdadeiro valor de R_0 .

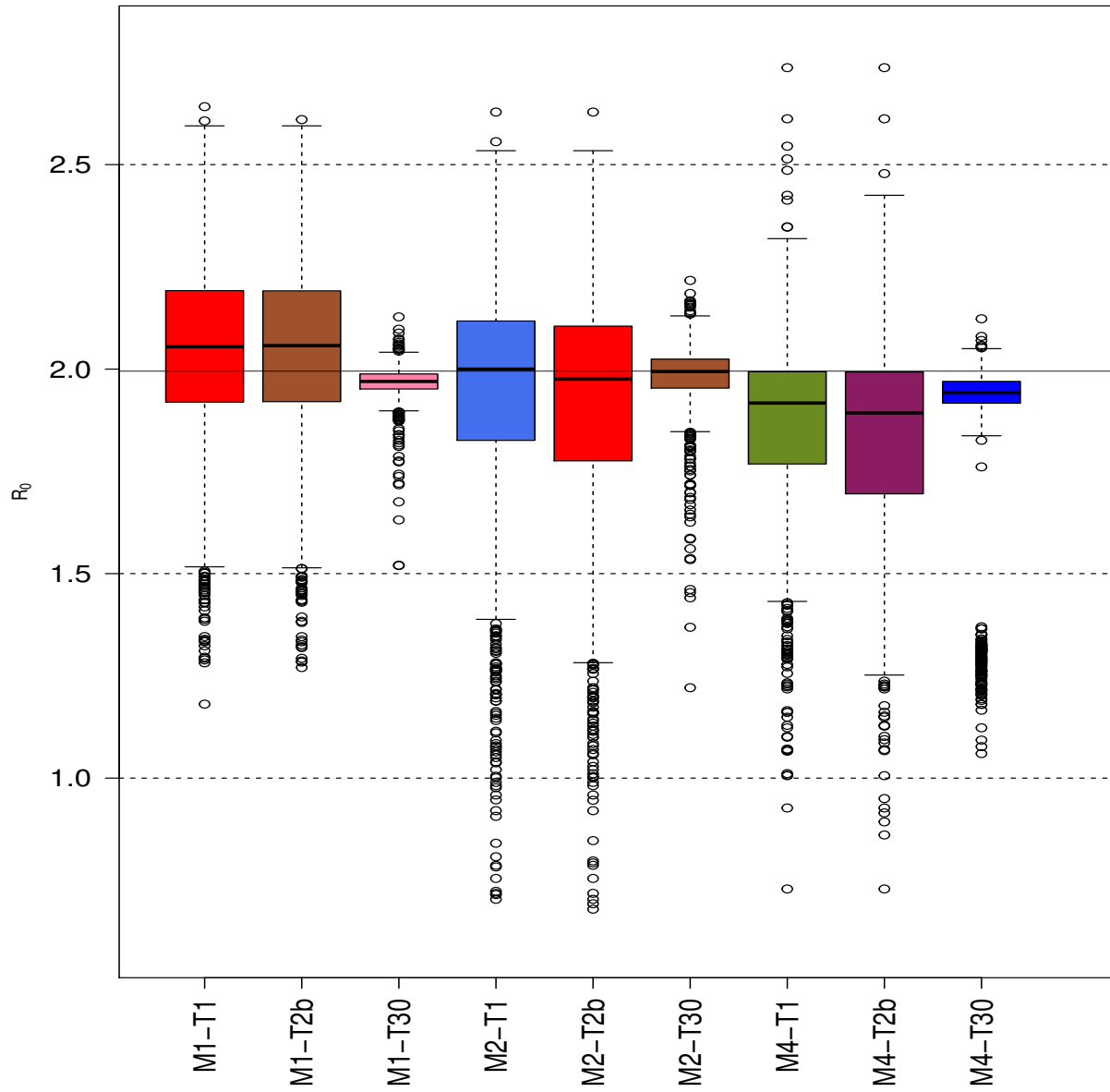


Figura 4.9: Box plot das estimativas de R_0 pelos métodos 1, 2 e 4 com os critérios de seleção T_1 e T_{2b} , com número mínimo de observações igual a 10 e com 30 observações. Baseados em 1000 séries geradas pelo modelo SIR estocástico. A linha horizontal cheia indica o verdadeiro valor de R_0 .

- A cobertura dos intervalos de credibilidade do método 4, quando utilizados os critérios T_1 e T_{2b} , são sempre maiores do que os valores nominais. As coberturas são próximas ao utilizarmos os critérios T_1 e T_{2b} , mas como o comprimento médio é menor utilizando o critério T_1 , sugere-se utilizar este critério. Observe que ao utilizar sempre todas as observações até o trigésimo dia, embora o comprimento dos intervalos sejam menores, sua cobertura é bem menor do que o valor nominal, o que nos leva à necessidade de escolher um período inicial da epidemia.
- A análise dos intervalos de confiança bootstrap segundo o método 2 deve ser feita com cuidado porque os resultados são relativos a apenas 200 séries e com apenas 200 séries bootstrap para cada série. Adicionalmente, considerando que são poucos os casos em que o período inicial selecionado foi maior do que 20 não é aconselhável analisar os resultados para cobertura 95% e nem para o intervalo [21, 30]. Os intervalos de confiança bootstrap nos métodos 1 e 2, construídos segundo o artigo de Chowell et al. (2007), mostraram-se péssimos. A sugestão para gerar séries bootstrap levou a uma melhora significativa nas coberturas dos intervalos de confiança 90% e 95%. Quando foi tomado o período até a trigésima observação a cobertura dos intervalos de confiança 95% é muito próximo do nominal, e para os intervalos de confiança 90%, a cobertura é inclusive maior do que o nominal. Isto demonstra que os resultados são promissores, e que talvez um melhor critério para selecionar o período inicial da epidemia junto com a metodologia sugerida de geração das séries bootstrap possam levar à construção de intervalos de confiança com a cobertura próxima do valor nominal. Os intervalos de confiança construídos pelo método M2- T_{2b} teve uma cobertura um pouco maior dos que os construídos pelos métodos M1- T_1 , M1- T_{2b} e M2- T_1 , mas a largura média dos seus intervalos é em torno de 25% a 50% maior do que os outros três métodos. Estes três métodos tiveram desempenhos melhores e muito parecidos em termos de cobertura e comprimento médio dos intervalos.

4.3 Processo gerador de dados: SEIR

Considere agora o modelo epidemiológico SEIR estocástico associado ao modelo determinístico apresentado na Seção 2.3 e descrito pelo sistema de equações 2.3.1, mas sem a taxa de mortalidade, isto é, com $\delta = 0$.

A escolha dos parâmetros foi realizada como no modelo SIR, isto é, utilizou as observações da série de influenza analisada por Chowell et al. (2007) como referência e os valores dos parâmetros utilizados no mesmo artigo. Além disso, como no modelo SIR, considerou-se o número de casos novos no trigésimo dia e o decaimento após a intervenção. Assim temos o tamanho da população $N = 550000$, a taxa de remoção é dada por $\gamma = 1/(4,1 \text{ dias})$ e o período de latência por $\kappa = 1/(1,9 \text{ dias})$. Antes da intervenção foi escolhido $\beta = 0,67$ e $I(0) = 11$, que equivale a $R_0 = 2,747$. Após a intervenção nós consideramos $\beta = 0,13$. O gráfico do número de notificações diárias de influenza, bem como o dado pelo modelo determinístico com os valores selecionados é apresentado na Figura 4.11. Assim como no modelo SIR, consideramos que a epidemia é observada no máximo durante 60 dias e foram escolhidas apenas as séries geradas que continham pelo menos 10 dias de

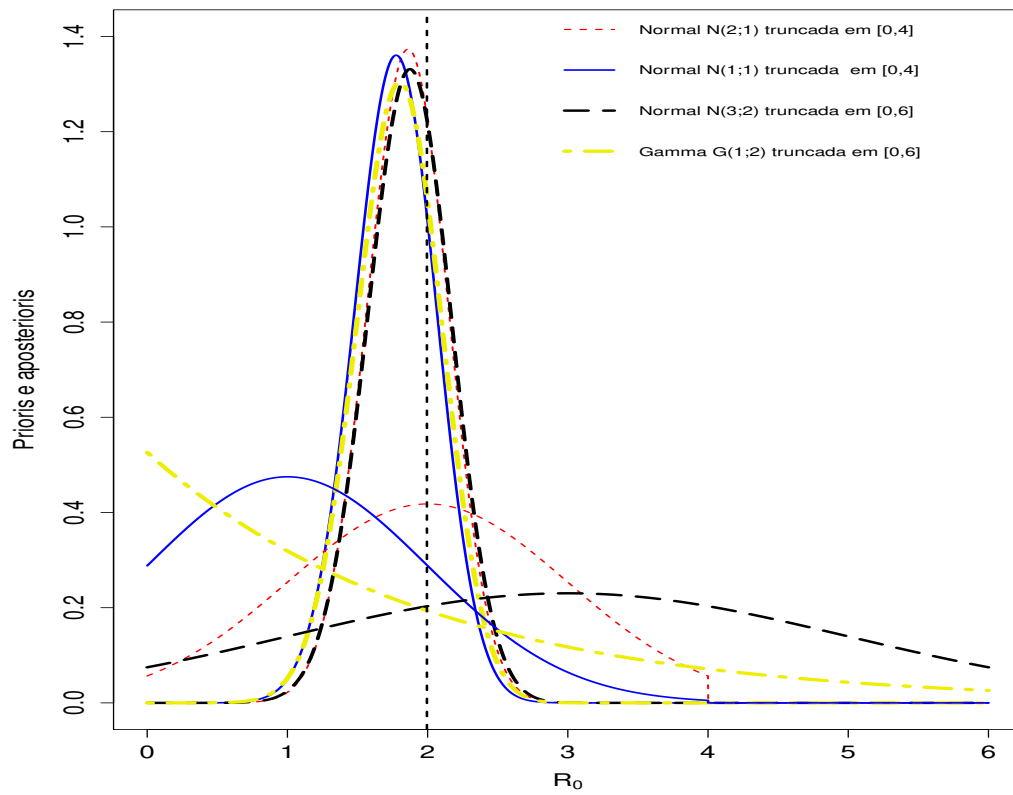


Figura 4.10: Distribuições a priori e a posteriori para a estimação do R_0 para uma série SIR após a décima observação. As distribuições a priori são as de maior variabilidade e a linha vertical indica o verdadeiro valor do R_0 .

observação. Para obter as 1000 séries de epidemia foi necessário gerar 1002 séries, ou seja, em duas replicações, a epidemia durou menos do que 10 dias. Destas 1000 séries de observações simuladas duas séries de epidemias não haviam terminado até o sexagésimo dia.

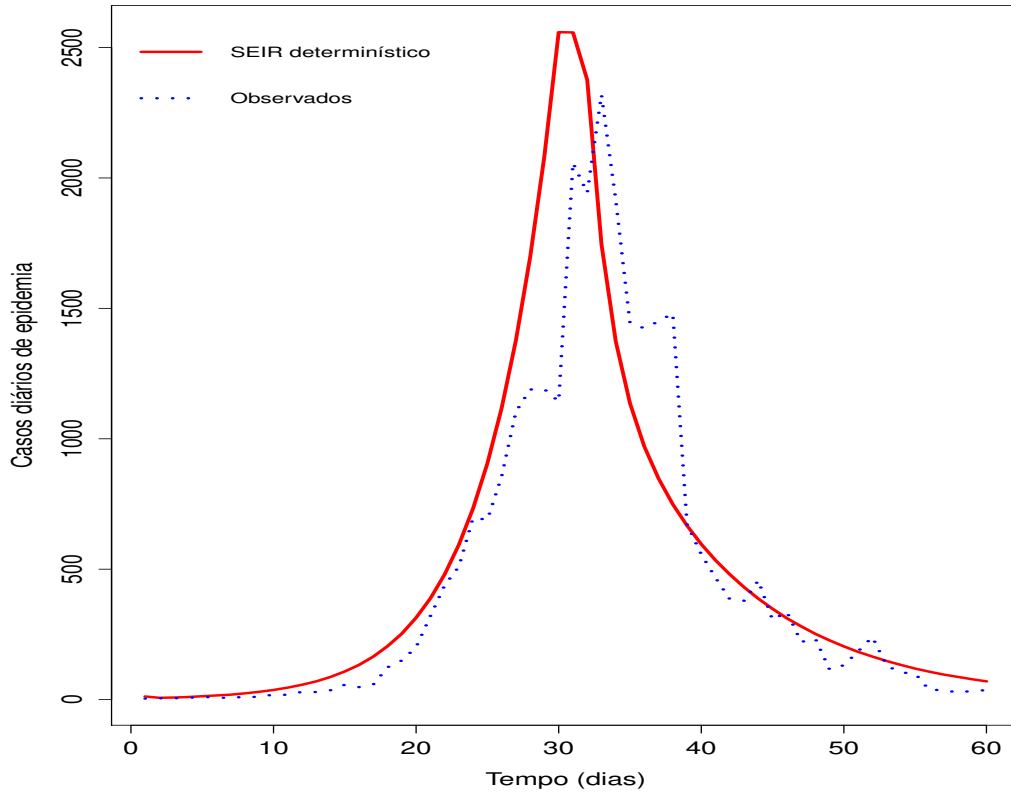


Figura 4.11: Número de casos novos diários observados na epidemia de influenza na Califórnia em 1918-1919 e dados pelo modelo determinístico SEIR com os valores dos parâmetros utilizados para gerar as séries simuladas pelo modelo SEIR estocástico.

Para evitar os problemas encontrados quando o processo gerador de dados era o modelo SIR foi considerado o mínimo de 10 observações para a estimação. Para o conjunto de observações obtido do processo gerador SEIR, o parâmetro R_0 foi estimado através dos seguintes métodos:

1. Método 1 com a fase inicial da epidemia sendo selecionada através dos critérios T_1 (ver Seção 3.1.1) e T_{2a} (ver Seção 3.2). Neste método é estimada apenas a taxa de crescimento exponencial r , considerando conhecida a taxa de remoção γ , sendo a estimativa de R_0 dada pelo método da substituição através da Equação 3.1.6, que dá a relação entre R_0 e a taxa de crescimento exponencial. São utilizados T_1 e T_{2a} com valores mínimos iguais a 10. As estimativas pontuais e por intervalos são baseadas em 1000 replicações. Assim como no caso SIR, na construção dos intervalos de confiança foram construídas 1000 réplicas bootstrap para cada série simulada.

2. Método 2 com modelo SEIR determinístico com T_1 e T_{2a} selecionados como o período para a estimação do modelo. Neste método são estimados a taxa de transmissão β e o número de infectantes iniciais, e considerado conhecido a taxa de remoção γ . Analogamente ao primeiro método utilizou-se T_1 e T_{2a} no mínimo igual a 10. As estimativas pontuais são baseadas em 1000 replicações. Como no caso SIR, os resultados para os intervalos de confiança referem-se a 200 séries simuladas e, para cada série simulada, foram construídas 200 réplicas bootstrap, exceto para o caso $T = 30$ quando os resultados para intervalo de confiança são baseadas em 1000 séries simuladas e, para cada série simulada, 1000 réplicas bootstrap

A Figura 4.12 mostra a distribuição de frequência dos períodos iniciais selecionados pelos critérios T_1 e T_{2a} . Existe uma tendência dos valores de T_1 serem menores do que T_{2a} , em 64,1% dos casos T_{2a} é maior do que T_1 ; em 11,5% dos casos os dois valores são iguais e nos restantes 24,4% dos casos T_{2a} é menor do que T_1 . A correlação entre os dois foi de $-0,040$. Em nenhum dos critérios de seleção foi escolhido um valor para a fase de crescimento exponencial maior do que 30 dias, quando ocorre a mudança dos valores dos parâmetros. Como sabemos que ocorre mudança dos valores dos parâmetros após o trigésimo dia, foi também realizadas as estimativas pelos dois métodos considerando o período de estimação até o trigésimo dia. Na maioria das séries os critérios selecionaram um período inicial da epidemia menor ou igual a 20. As percentagens foram iguais a 66,1% e 87,5% para os critérios T_1 e T_{2b} , respectivamente. As Figuras 4.13 e 4.14 mostram os box plots das estimativas de R_0 pelos métodos 1 e 2, respectivamente, segundo a fase inicial da epidemia selecionada pelos critérios T_1 e T_{2a} . A Tabela 4.5 apresenta as estatísticas descritivas das estimativas por intervalo do número de observações selecionadas e a Figura 4.15 os box plots das estimativas. A Tabela 4.6 apresenta a matriz de covariância/correlação das estimativas dos métodos considerados.

Como no caso do processo gerador SIR, para evitar os problemas de pouca variabilidade e baixa cobertura dos intervalos de confiança bootstrap foi utilizado a segunda alternativa de misturas de distribuições Poisson para gerar as séries bootstrap.

A análise dos resultados indica que:

- Nos métodos 1 e 2 os melhores resultados ocorrem utilizando a fase inicial da epidemia até o trigésimo dia. Neste caso, o método 1 apresentou o melhor desempenho.
- Para as séries para as quais a fase inicial da epidemia selecionada é maior do que 20 dias os resultados dos dois métodos são melhores.
- Quando é necessário escolher a fase inicial da epidemia, estranhamente o método 1 apresentou melhor desempenho com o critério T_{2a} e o método 2 com o critério T_1 .
- Quando é necessário escolher a fase inicial da epidemia, utilizando-se o o critério T_1 , os dois métodos, em termos de EQM, são muito próximos, sendo o EQM menor no método 1.
- As estimativas obtidas através do método 1, com os critérios T_1 , T_{2a} e $T = 30$ tem altas correlações. Quando o processo gerador de dados foi o SIR, não foi encontrada alta correlação entre as estimativas obtidas através do método 1 com $T = 30$, com as estimativas do mesmo método com os critérios T_1 e T_{2b} . Com o método 2 os resultados foram semelhantes, com exceção de baixa correlação entre as estimativas com os critérios T_1 e T_{2a} .

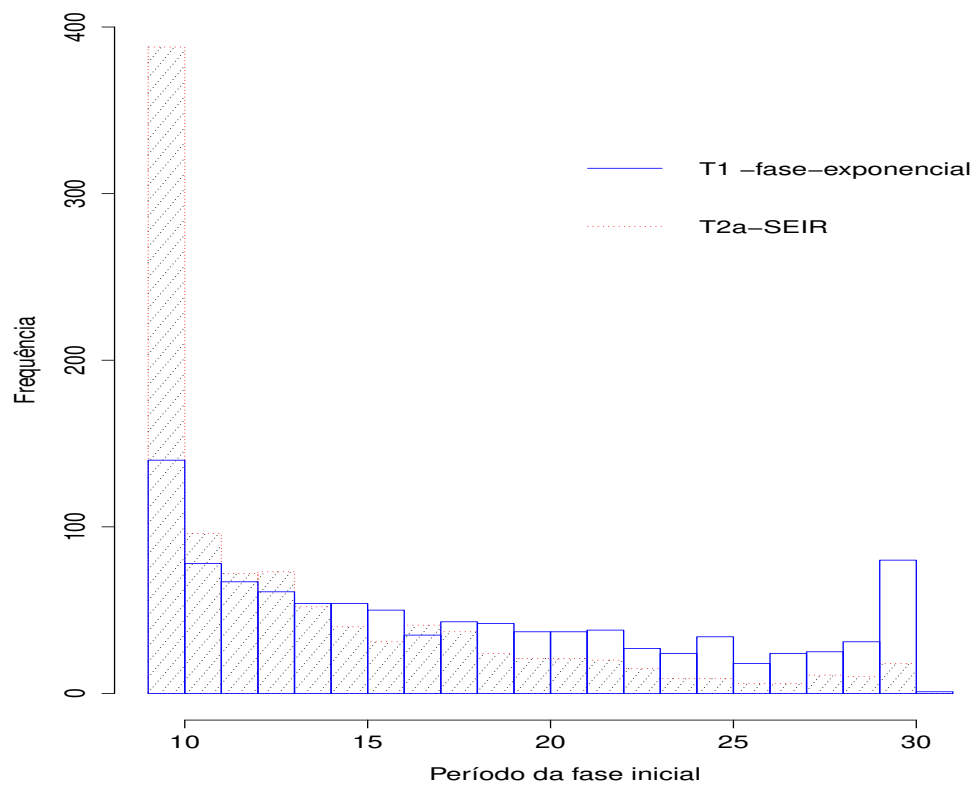


Figura 4.12: Frequência da fase inicial da epidemia selecionada pelos critérios T_1 e T_{2a} nas 1000 séries geradas pelo modelo SEIR estocástico.

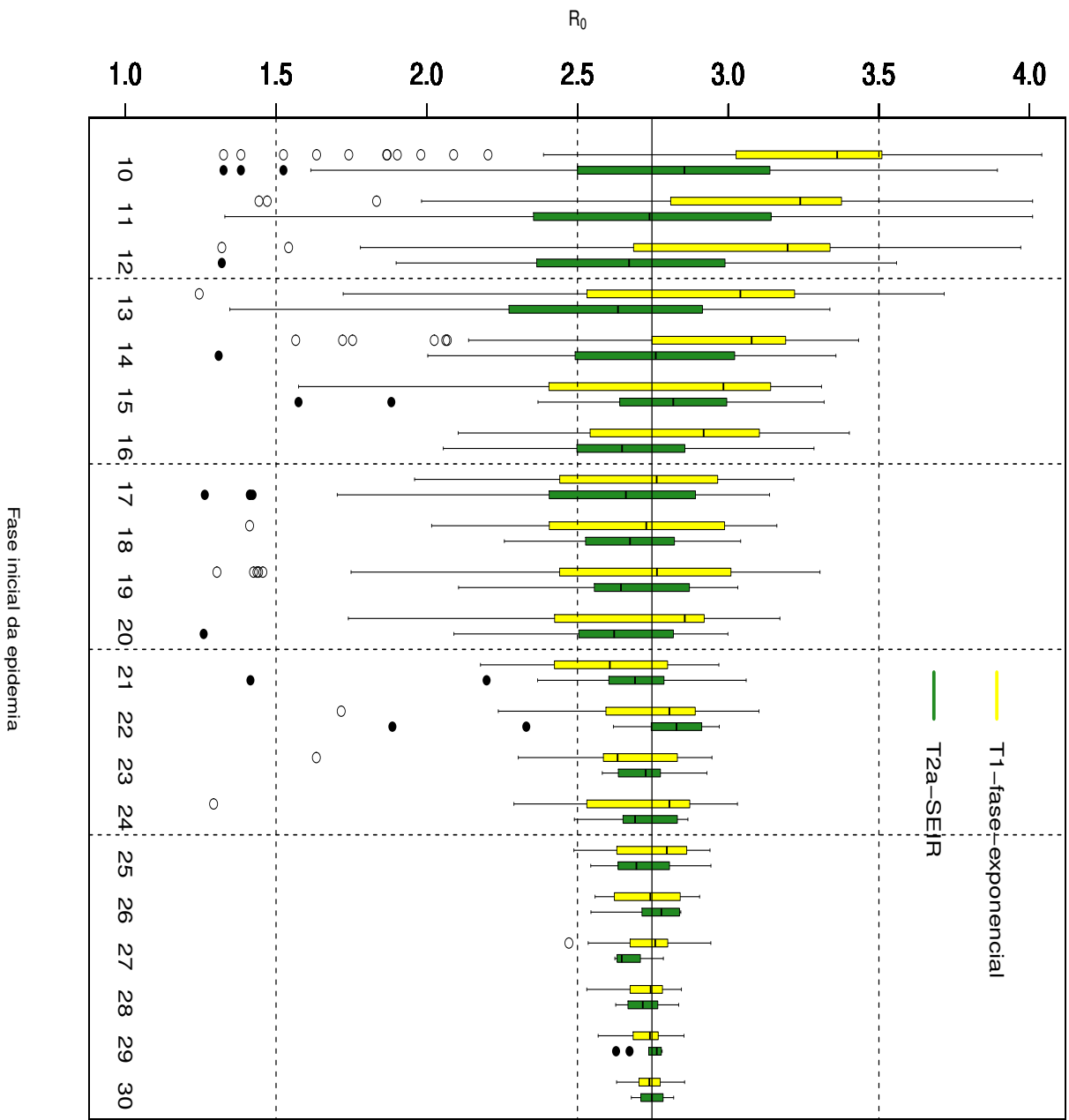


Figura 4.13: Box plot das estimativas de R_0 pelo método 1 segundo a fase inicial da epidemia selecionada pelos critérios T_1 e T_{2a} , com número mínimo de observações igual a 10. Baseados em 1000 séries geradas pelo modelo SEIR estocástico. A linha horizontal cheia indica o verdadeiro valor de R_0 .

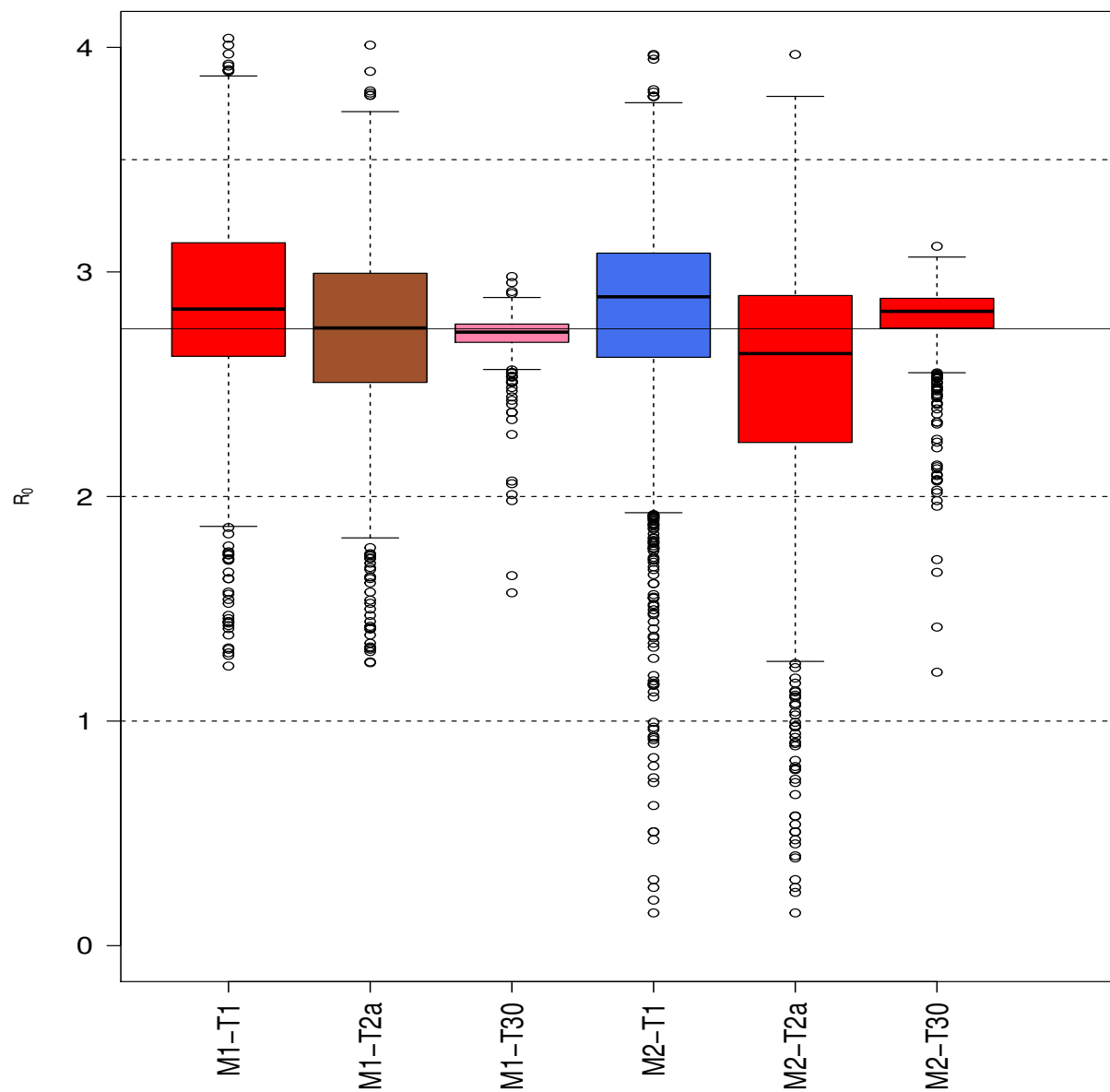


Figura 4.15: Box plot das estimativas de R_0 pelos métodos 1 e 2 com os critérios de seleção T_1 e T_{2a} , com número mínimo de observações igual a 10 e com 30 observações. Baseados em 1000 séries geradas pelo modelo SEIR estocástico. A linha horizontal cheia indica o verdadeiro valor de R_0 .

- A cobertura média dos intervalos de confiança construídos pelos métodos $M1-T_1$, $M1-T_{2a}$, $M2-T_1$ e $M2-T_{2a}$ foram sempre abaixo dos valores nominais. O método $M2-T_{2a}$ foi o que apresentou pior desempenho por ter menor cobertura e ter a maior média dos comprimentos dos intervalos. Os outros tiveram desempenhos muito próximos e coberturas próximos dos valores nominais.

Tabela 4.2: Resumo das estimativas de R_0 para o modelo SIR estocástico. Resultados baseados em 1000 replicações. Método 1: método da taxa de crescimento intrínseco; método 2: modelo SIR determinístico; método 4: estimador bayesiano do modelo SIR estocástico. Fase inicial da epidemia selecionada pelos critérios T_1 e T_{2b} . Resultado por intervalo de T_1 e T_{2b} em $[4; 9]$, $[10; 20]$, $[20; 30]$ e para todas os casos

Método	T_1 ou T_{2b}	Vício	1º quartil	Mediana	Média	3º Quartil	Desvio Padrão	EQM
M1- T_1	[4,9]	0,379	2,084	2,421	2,374	2,669	0,450	0,345
	[10,20]	-0,014	1,831	2,042	1,981	2,155	0,253	0,064
	[21,30]	-0,043	1,934	1,984	1,952	2,021	0,146	0,023
	todos	0,166	1,929	2,107	2,161	2,406	0,404	0,190
M1- T_{2b}	[4,9]	0,428	2,190	2,461	2,423	2,694	0,404	0,346
	[10,20]	-0,058	1,794	1,974	1,937	2,117	0,259	0,070
	[21,30]	-0,037	1,944	1,985	1,958	2,024	0,135	0,019
	todos	0,243	1,958	2,201	2,238	2,550	0,423	0,238
M2- T_1	[4,9]	0,057	1,735	2,179	2,052	2,396	0,551	0,306
	[10,20]	-0,167	1,659	1,913	1,828	2,076	0,341	0,143
	[21,30]	-0,072	1,903	1,974	1,923	2,008	0,193	0,042
	todos	-0,049	1,736	1,998	1,946	2,200	0,451	0,206
M2- T_{2b}	[4,9]	0,094	1,833	2,163	2,089	2,425	0,502	0,261
	[10,20]	-0,235	1,567	1,842	1,760	2,032	0,363	0,186
	[21,30]	-0,056	1,929	1,989	1,939	2,029	0,179	0,035
	todos	-0,001	1,740	2,008	1,994	2,259	0,472	0,223
M4- T_1	[4,9]	-0,319	1,443	1,686	1,675	1,903	0,370	0,239
	[10,20]	-0,162	1,692	1,893	1,832	1,985	0,244	0,086
	[21,30]	-0,085	1,912	1,961	1,909	1,996	0,191	0,044
	todos	-0,225	1,594	1,835	1,769	1,979	0,318	0,152
M4- T_{2b}	[4,9]	-0,447	1,296	1,556	1,547	1,788	0,375	0,341
	[10,20]	-0,176	1,671	1,859	1,819	1,988	0,255	0,096
	[21,30]	-0,081	1,905	1,963	1,914	2,003	0,189	0,042
	todos	-0,338	1,406	1,698	1,657	1,938	0,360	0,244

Tabela 4.3: Resumo das estimativas de R_0 para o modelo SIR estocástico. Resultados baseados em 1000 replicações, exceto para os intervalos de confiança para o método 2, que é baseado em 200 replicações. Método 1: método da taxa de crescimento intrínseco; método 2: modelo SIR determinístico; método 4: estimador bayesiano do modelo SIR estocástico. Resultados com a fase inicial da epidemia selecionada pelos critérios T_1 e T_{2b} , com número mínimo de 10 observações, e com fase inicial com 30 observações

Método	T_1 ou T_{2b}	Vício	1º quartil	Mediana	Média	3º Quartil	Desvio Padrão	EQM	Cobertura		Largura	
									90%	95%	90%	95%
M1- T_1	[10,20]	0,058	1,901	2,113	2,053	2,225	0,249	0,066	69,9	79,4	0,676	0,809
	[21,30]	-0,038	1,934	1,976	1,957	2,020	0,128	0,018	88,8	91,9	0,637	0,764
	todos	0,039	1,920	2,055	2,034	2,192	0,234	0,056	73,6	81,9	0,669	0,800
M1- T_{2b}	[10,20]	0,049	1,916	2,082	2,045	2,209	0,234	0,057	71,2	80,1	0,641	0,777
	[21,30]	-0,049	1,931	1,980	1,946	2,017	0,142	0,022	94,1	96,5	0,606	0,737
	todos	0,041	1,921	2,057	2,036	2,191	0,229	0,054	73,2	81,5	0,638	0,773
M1	T=30	-0,028	1,951	1,970	1,967	1,988	0,045	0,002	93,4	94,6	0,912	1,092
M2- T_1	[10,20]	-0,067	1,757	2,030	1,928	2,153	0,343	0,122	70,4	79,2	0,674	0,807
	[21,30]	-0,061	1,906	1,973	1,934	2,011	0,169	0,032	87,8	92,4	0,638	0,764
	todos	-0,066	1,826	2,000	1,929	2,118	0,316	0,104	73,8	81,8	0,667	0,799
M2- T_{2b}	[10,20]	-0,089	1,759	1,975	1,906	2,124	0,327	0,115	78,9	83,3	0,912	1,091
	[21,30]	-0,075	1,904	1,976	1,919	2,017	0,193	0,043	77,6	83,5	0,905	1,098
	todos	-0,088	1,776	1,976	1,907	2,105	0,318	0,108	78,8	83,4	0,912	1,092
M2	T=30	-0,015	1,953	1,994	1,979	2,024	0,088	0,008	93,8	94,8	0,370	0,437
M4- T_1	[10,20]	-0,148	1,719	1,900	1,847	1,994	0,241	0,080	94,1	98,4	0,999	1,185
	[21,30]	-0,080	1,903	1,953	1,914	1,993	0,169	0,035	93,4	97,5	0,304	0,361
	todos	-0,134	1,768	1,917	1,860	1,993	0,230	0,071	94,0	98,2	0,862	1,022
M4- T_{2b}	[10,20]	-0,191	1,650	1,870	1,804	2,010	0,331	0,146	93,4	97,9	1,134	1,346
	[21,30]	-0,029	1,950	1,990	1,965	2,020	0,125	0,016	97,6	97,6	0,337	0,400
	todos	-0,177	1,670	1,900	1,818	2,010	0,322	0,135	93,8	97,9	1,066	1,265
M4	T=30	-0,132	1,917	1,943	1,863	1,970	0,230	0,070	66,4	74,9	0,147	0,175

Cobertura: Cobertura média dos intervalos de confiança. Largura: Largura média dos intervalos de confiança

Tabela 4.4: Correlação (acima da diagonal) e covariância das estimativas de R_0 para o modelo SIR estocástico. Resultados baseados em 1000 replicações. Método 1: método da taxa de crescimento intrínseco; método 2: modelo SIR determinístico; método 4: estimador bayesiano do modelo SIR estocástico. Resultados com fase inicial da epidemia selecionada pelos critérios T_1 e T_{2b} , com número mínimo de 10 observações, e com 30 observações

Método	M1- T_1	M1- T_{2b}	M1-T=30	M2 - T_1	M2- T_{2b}	M2- $T = 30$	M4- T_1	M4- T_{2b}	M4-T=30
M1- T_1	0,055	0,949	0,144	0,950	0,903	0,014	0,345	0,469	0,474
M1- T_{2b}	0,051	0,053	0,141	0,912	0,927	0,019	0,384	0,444	0,524
M1-T=30	0,001	0,001	0,002	0,100	0,096	0,083	0,172	0,220	0,232
M2- T_1	0,070	0,066	0,001	0,0100	0,934	0,023	0,394	0,444	0,509
M2- T_{2b}	0,067	0,068	0,001	0,093	0,101	0,025	0,389	0,438	0,591
M2-T=30	0,0003	0,0004	0,0003	0,0006	0,0007	0,0081	0,049	-0,016	0,049
M4- T_1	0,018	0,020	0,002	0,029	0,028	0,001	0,053	0,538	0,216
M4- T_{2b}	0,027	0,026	0,002	0,035	0,035	-0,0003	0,031	0,063	0,246
M4-T=30	0,025	0,028	0,002	0,037	0,043	0,001	0,011	0,014	0,053

Tabela 4.5: Resumo das estimativas de R_0 para o modelo SEIR estocástico. Resultados baseados em 1000 replicações, exceto para os intervalos de confiança para o método 2, que é baseado em 200 replicações. Método 1: método da taxa de crescimento intrínseco; método 2: modelo SEIR determinístico. Resultados com a fase inicial da epidemia selecionada pelos critérios T_1 e T_{2a} , com número mínimo de 10 observações, e com fase inicial com 30 observações

Método	T_1 ou T_{2a}	Vício	1º quartil	Mediana	Média	3º Quartil	Desvio Padrão	EQM	Cobertura		Largura	
									90%	95%	90%	95%
M1- T_1	[10,20]	0,159	2,557	3,024	2,907	3,270	0,525	0,301	86,1	90,1	1,658	1,977
	[21,30]	-0,037	2,649	2,738	2,710	2,822	0,185	0,035	97,0	97,3	1,569	1,868
	todos	0,092	2,624	2,835	2,840	3,130	0,450	0,210	89,8	92,6	1,628	1,940
M1- T_{2a}	[10,20]	-0,014	2,474	2,768	2,732	3,042	0,456	0,208	85,8	90,2	1,658	1,980
	[21,30]	-0,033	2,662	2,736	2,714	2,798	0,186	0,036	96,7	97,9	1,563	1,867
	todos	-0,016	2,510	2,750	2,730	2,994	0,432	0,187	89,4	92,7	1,627	1,942
M1	T=30	-0,026	2,687	2,732	2,720	2,768	0,097	0,010	94,3	96,2	0,694	0,825
M2- T_1	[10,20]	0,017	2,455	2,976	2,764	3,168	0,642	0,412	87,3	91,5	1,604	1,926
	[21,30]	0,046	2,731	2,838	2,793	2,919	0,221	0,051	96,1	98,4	1,802	2,139
	todos	0,027	2,621	2,889	2,774	3,083	0,537	0,289	88,1	92,4	1,628	1,952
M2- T_{2a}	[10,20]	-0,276	2,170	2,572	2,471	2,892	0,613	0,452	80,1	83,7	2,008	2,310
	[21,30]	0,045	2,738	2,834	2,793	2,899	0,233	0,056	78,4	85,6	2,031	2,340
	todos	-0,236	2,241	2,636	2,511	2,894	0,589	0,403	79,8	83,1	2,011	2,314
M2	T=30	0,045	2,750	2,825	2,793	2,882	0,162	0,028	96,1	96,5	0,825	0,970

Cobertura: Cobertura média dos intervalos de confiança. Largura: Largura média dos intervalos de confiança

Tabela 4.6: Correlação (acima da diagonal) e covariância das estimativas de R_0 para o modelo SEIR estocástico. Resultados baseados em 1000 replicações. Método 1: método da taxa de crescimento intrínseco; método 2: modelo SEIR determinístico. Resultados com a fase inicial da epidemia selecionada pelos critérios T_1 e T_{2a} , com número mínimo de 10 observações, e com fase inicial com 30 observações

Método	M1- T_1	M1- T_{2a}	M1-T=30	M2- T_1	M2- T_{2a}	M2-T=30
M1- T_1	0,202	0,942	0,852	0,886	0,245	0,595
M1- T_{2a}	0,228	0,289	0,856	0,902	0,248	0,612
M1-T=30	0,166	0,199	0,187	0,945	0,286	0,627
M2- T_1	0,235	0,286	0,241	0,348	0,269	0,618
M2- T_{2a}	0,011	0,013	0,012	0,015	0,010	0,814
M2-T=30	0,044	0,054	0,044	0,059	0,013	0,027

Capítulo 5

Comentários Finais

Os resultados de simulação apresentados mostraram que as estimativas pontuais de R_0 são razoáveis mas a medida de incerteza dada pelas estimativas intervalares não são boas, exceto pelo método bayesiano quando o processo gerador de dados é o SIR estocástico. Para o caso em que é considerado o processo gerador SIR, o método bayesiano apresentou um melhor desempenho. Para o caso do processo gerador SEIR foram testados apenas os métodos 1 e 2 com os critérios de escolha T_1 e T_{2a} . Neste caso o método 1 com o critério T_{2a} foi o que apresentou o melhor desempenho. A análise dos resultados encontrados mostram que ainda existem vários pontos a serem esclarecidos em investigações futuras. Alguns destes pontos são:

- Os métodos exigem a escolha do modelo compartimental a ser utilizado. É necessário criar critérios para seleção do modelo e de técnicas de diagnóstico.
- As estimativas dependem da escolha da fase inicial da epidemia. Os critérios utilizados não se mostraram eficientes.
- Os intervalos de confiança construídos através do método bootstrap, como utilizado por Chowell et al. (2007) tiveram péssimo desempenho. Como o vício não é grande, o problema maior é a largura dos intervalos. Isto provavelmente ocorreu devido às amostras bootstrap construídas. A modificação proposta de geração das séries bootstrap melhoraram consideravelmente o desempenho dos métodos. No caso do processo gerador de dados ser um SEIR, as coberturas estiveram próximas das nominais. Este problema merece maiores investigações, mas a sugestão apresentada, acoplada a um melhor critério para seleção da fase inicial da epidemia parece ser bastante promissora.
- É necessário fazer um estudo de simulação mais completo. Por exemplo, no método 2 aumentando o número de replicações e de réplicas bootstrap.
- Alguns dos parâmetros foram considerados como conhecidos. Por exemplo, no método 1, no caso do modelo SIR temos $R_0 = 1 + r/\gamma$, em que γ foi considerado como conhecido. Isto significa que a estimativa de R_0 não é robusta em relação ao parâmetro γ . Assim, é necessário comparar com métodos onde o parâmetro γ é estimado para verificar se vale a pena considerar este parâmetro como conhecido.

- No método bayesiano é necessário discutir melhor a sensibilidade em relação à escolha da priori e analisar outras prioris. Por exemplo, no método implementado no pacote R_0 do R, a distribuição a priori por default é uma uniforme.
- Em muitos casos temos grandes conhecimentos sobre o valor do R_0 . Por exemplo, sabemos que é positivo e é possível colocar um limite superior. Estas informações poderiam também ser consideradas em uma versão bayesiana dos métodos 1 e 2.
- Os modelos que consideram os parâmetros variando no tempo merecem maiores estudos. Algumas aplicações deste modelo podem ser encontrados em Chowell et al. (2004b) e Lekone e Finkenstädt (2006).

Bibliografia

- ALLEN, L. J. S. *An Introduction to Stochastic Processes with Applications to Biology*. 2^a.ed. Prentice Hall, 2011.
- ANDERSON, R. M.; MAY, R. M. *Infectious Disease of Humans*. Nova Iorque: Oxford University Press, 1991.
- ANDERSSON, G. Epidemiological features of chronic low-back pain. *The Lancet*, v. 354, p. 581-585, 1999.
- ANDERSSON, H.; BRITTON, T. *Stochastic Epidemic Models and their Statistical Analysis*. Nova Iorque: Springer Verlag, 2000.
- ARTALEJO, J. R.; LOPEZ-HERRERO, M. J. On the exact measure of disease spread in stochastic epidemic models. *Bulletin of Mathematical Biology*, v. 75, p. 1-20, 2013.
- BACAËR, N. *A Short History of Mathematical Population Dynamics*. London: Springer Verlag, 2011.
- BAILEY, N. *The Mathematical Theory of Infectious Diseases and its Applications*. 2^a. ed. Nova Iorque: Oxford University Press, 1975.
- BARRY, J. *The Great Influenza: The Story of the Deadliest Pandemic in History*. Nova Iorque: Viking Penguin, 2004.
- BERMAN, A.; PLEMMONS, R. J. Inverses of nonnegative matrices. *Linear and Multilinear Algebra*, v. 2, p. 161-172, 1974.
- BETTENCOURT, L. M. A. ; RIBEIRO, M. Real time Bayesian estimation of the epidemic potential of emerging infectious diseases. *Plos One*, v. 5, e2185, 2008.
- BRAUER, F. Basic ideas of mathematical epidemiology. In CASTILLO-CHAVEZ, C.; DRIESSCHE, P.; KIRSCHNER, D.; YAKUBU, A. (eds.) *Mathematical Approaches for Emerging and Reemerging Infection Diseases: An Introduction*. The IMA Volumes in Mathematics and its Applications, vol. 125, p. 31-65. Nova Iorque: Springer, 2002.
- BRAUER, F. Compartmental models in epidemiology. In BRAUER, F.; VAN DEN DRIESSCHE, P.; WU, J. (eds.) *Mathematical Epidemiology*. Lecture Notes in Mathematics, vol. 1945, p. 19-79. Berlin: Springer Berlin Heidelberg, 2008.

- BRAUER, F.; CASTILLO-CHAVEZ, C. *Mathematical Models in Population Biology and Epidemiology*. Nova Iorque: Springer-Verlag, 2001.
- BRAUER, F.; VAN DEN DRIESSCHE, P. Models for transmission of disease with immigration of infectives. *Mathematical Biosciences*, v. 171, p. 143-154, 2001.
- BRAUER, F.; VAN DEN DRIESSCHE, P.; WANG, L. Oscillations in a patchy environment disease model. *Mathematical Biosciences*, v. 215, p. 1-10, 2008.
- BROOKS, F.J. Revising the conquest of Mexico: Smallpox, sources, and populations. *Journal of Interdisciplinary History*, v. 24, p 1-29, 1993.
- BUSENBERG, S.; HADELER, K. P. Demography and epidemics. *Mathematical Biosciences*, v. 101, p. 41-62, 1990.
- CAUCHEMEZ, C.; BOELLE, P. Y.; GUY THOMAS.; VALLERON, A. J. Estimating in real time the efficacy of measures to control emerging communicable diseases. *American Journal of Epidemiology*, v. 164, p. 591-597, 2006.
- CAUCHEMEZ, C.; CARRAT, F.; VIBOUD, V.; VALLERON, A. J.; BOELLE, P. Y. A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. *Statistics in Medicine*, v. 23. p. 3469-3487, 2004.
- CHOI, B. C. K.; PAK, A. W. P. A simple approximate mathematical model to predict the number of severe acute respiratory syndrome cases and deaths. *Journal of Epidemiology and Community Health*. v. 57, p. 831-835, 2003.
- CHOWELL, G.; AMMON, C. E.; HENGARTNER, N. W.; HYMAN, J. M. Transmission dynamics of the great influenza pandemic of 1918 in Geneva, Switzerland: Assessing the effects of hypothetical interventions. *Journal of Theoretical Biology*, v. 241, p. 193-204, 2006.
- CHOWELL, G.; BRAUER, F. The basic reproduction number of infectious diseases: Computational and estimation using compartmental epidemic models. In CHOWELL, G., HYMAN, J. M., BETTENCOURT, L. M. A.; CASTILLO-CHAVEZ, C. (eds.). *Mathematical and Statistical Estimation Approaches in Epidemiology*, p. 1-30. Londres: Springer, 2009.
- CHOWELL, G.; CASTILLO-CHAVEZ, C.; FENIMORE, P.W.; KRIBS-ZALETA, C.M.; ARRIOLA, L.; HYMAN, J. M. Model parameters and outbreak control for SARS. *Emerging Infectious Disease Journal*, v. 10, p. 1258-1263. 2004a
- CHOWELL, G.; DIAZ-DUEÑAS, P.; MILLER, J. C.; ALCAZAR, A.; HYMAN, J. M.; FENIMORE, P. W.; CASTILLO-CHAVEZ, C. Estimation of the reproduction number of dengue fever from spatial epidemic data. *Mathematical Biosciences*, v. 208, p. 571-589, 2007.
- CHOWELL, G.; FENIMORE, P. W.; CASTILLO-GARSON, M. A.; CASTILLO-CHAVEZ, C. SARS outbreaks in Ontario, Hong Kong and Singapore: The role of diagnosis and isolation as a control mechanism, *Journal of Theoretical Biology*, v. 224, p. 1-8, 2003.

- CHOWELL, G.; FUENTES, R.; OLEA, A.; AGUILERA, X.; NESSE, H.; HYMAN, J. M. The basic reproduction number R_0 and effectiveness of reactive interventions during dengue epidemics: The 2002 dengue outbreak in Easter Island, Chile. *Mathematical Biosciences and Engineering*, v. 10, p. 1455-1474, 2013.
- CHOWELL, G.; HENGARTNER, N. W.; CASTILLO-CHAVEZ, C.; FENIMORE, P. W.; HYMAN, J. M. The basic reproductive ratio number of Ebola and the effects of public health measures: The cases of Congo and Uganda. *Journal of Theoretical Biology*, v. 229, p. 119-126, 2004b.
- CHOWELL, G.; NISHIURA, H. Quantifying the transmission potential of pandemic influenza. *Physics of Life Reviews*, v. 5, p. 50-77, 2008
- CHOWELL, G.; NISHIURA, H.; BETTENCOURT, L. M. A. Comparative estimation of the reproduction number for pandemic influenza from daily case notification data. *Journal of the Royal Statistical Society Interface*, v. 4, p. 155-166, 2007.
- COELHO, F. C.; CODEÇO, C. T.; GOMEZ, M. G. M. Bayesian framework for parameter estimation in dynamical models. *PloS One*, 6(5), e19616, 2011.
- COWLING, B. J.; HO, L. M.; LEUNG, G. M. Effectiveness of control measures during the SARS epidemic in Beijing: A comparison of the R_t curve and the epidemic curve. *Epidemiology and Infection*, v. 136, p. 562-566, 2008.
- DAGNINO, S. What was the plague of Athens? *Revista Chilena de Infectologia*, v. 28, p. 374-380, 2011.
- DE KOEIJER, A. A.; DIEKMANN, O.; DE JHON, M. C. Calculating the extinction of a reactivating virus, in particular bovine herpes virus. *Mathematical Biosciences*, v. 212, p. 111-131, 2008.
- DERRICK, W. R.; VAN DEN DRIESSCHE, P. A disease transmission model in a nonconstant population. *Journal of Mathematical Biology*, v. 31, p. 495-512, 1993.
- DIEKMANN, O.; HEESTERBEEK, J. A. P. *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*. Nova Iorque: John Wiley, 2000.
- DIEKMANN, O.; HEESTERBEEK, H.; BRITTON, T. *Mathematical Tools for Understanding Infectious Diseases Dynamics*. Princeton: Princeton University Press, 2012.
- DIEKMANN, O.; HEESTERBEEK, J. A. P.; METZ, J. A. J. 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*, v. 28, p. 365- 382, 1990.
- DIETZ, K. Transmission and control of arbovirus diseases. In: LUDWIG D.; COOKE, K. L. (eds.). *Epidemiology*. p. 104-121. Philadelphia: Society for Industrial and Applied Mathematics, 1975.

- DIETZ, K. On the transmission dynamics of HIV. *Mathematical Biosciences*, v. 90, p. 397- 414, 1988.
- DIETZ, K. The estimation of the basic reproduction number for infectious diseases. *Statistical Methods in Medical Research*, v. 2, p. 23-41, 1993.
- DONNELLY, C. A.; GHANI, A. C.; LEUNG, G. M.; HEDLEY, A. J.; FRASER, C.; RILEY, S.; ABU-RADDAD, L. J.; HO, L.-M.; THACH, T.-Q.; CHAU, P.; CHAN, K.-P.; LAM, T.-H.; TSE, L.-Y.; TSANG, T.; LIU, S.-H.; KONG, J. H. B.; LAU, E. M. C.; FERGUSON, N. M.; ANDERSON, R. M. Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. *Lancet*, v. 361, p. 1761-1766, 2003.
- DUBLIN, L. On the true rate of natural increase: As exemplified by the population of the United States. *Journal of the American Statistical Association*, v. 20, p. 305-339, 1925.
- EFRON, B.; TIBSHIRANI, R. Bootstrap method for standard errors, confidence Intervals, and other measures of statistical accuracy. *Statistical Science*, v. 1, p. 54-75, 1986.
- EN'KO, P. D. On the course of epidemics some infectious diseases. *International Journal of Epidemiology*, v. 18, p. 749-755, 1889.
- FARRINGTON, C. P.; KANAAN, M. N.; GAY, N. J. Estimation of the basic reproduction number for infectious diseases from age-stratified serological survey data. *Applied Statistics*, v. 50, p. 1-33, 2001.
- FAVIER, C.; DEGALLIER, N.; ROSA-FREITAS, M. G.; BOULANGER, J. P.; COSTA LIMA, J. R.; LUITGARDS-MOURA, J. F.; MENKÈS, C. E.; MONDET, B.; OLIVEIRA, C; WEIMANN, E. T. S.; TSOURIS, P. Early determination of the reproductive number for vector-borne diseases: the case of dengue in Brazil. *Tropical Medicine and International Health*, v. 11, p. 332-340, 2006.
- FENN, E. *Pox Americana: The Great Smallpox Epidemic of 1775-82*. Nova Iorque: Hill & Wang, 2001.
- FERGUSON, N. M.; CUMMINGS, D. A.; FRASER, C. Strategies for mitigating an influenza pandemic. *Nature*, v. 442, p. 448-452, 2006.
- FERGUSON, N. M.; DONNELLY, C. A.; ANDERSON, R.M. The foot and mouth epidemic in Great Britain: Pattern of spread and impact of interventions, *Science*, v. 292, p. 1155-1160, 2001.
- FERGUSON, N. M.; DONNELLY, C. A.; WOOLHOUSE, M. E. Estimation of the basic reproduction number of BSE: The intensity of transmission in British cattle. *Proceedings of the Royal Society B: Biological Sciences*, v. 266, p. 23-32, 1999.
- FLAHAULT, A.; LETRAIT, S.; BLIN, P.; HAZOUT, S.; MÉNARÉS, J.; VALLERON, A. J. Modelling the 1985 influenza epidemic in France. *Statistics in Medicine*, v. 7, p. 1147-1155, 1988.

- FRASER, C.; RILEY, S.; ANDERSON, R. M.; FERGUSON, N. M. Factors that make an infectious disease outbreak controllable. *Proceedings of the National Academy of Sciences of the United States of America*, v. 101, p. 6146-6151, 2004.
- GANI, S. R.; HUGHES, H.; FLEMING, D.; GRIFFIN, T.; MEDLOCK, K.; LEACH, S. Potential impact of antiviral drug use during influenza pandemic. *Emerging Infectious Disease*, v. 11, p. 1355-1362, 2005.
- GANI, S. R.; TASLIN, A. S.; KADI, A. S. The transmission dynamics of pandemic influenza, A/H1N1 200-2010 in India. *Research Communications*, v. 101, p. 1065-1072, 2011.
- GORDILLO, L. F.; GREENWOOD, P. E. Bimodal epidemic size distributions for near-critical SIR with vaccination. *Bulletin of Mathematical Biology*, v. 70, 589-602, 2007.
- GORDILLO, L. F.; GREENWOOD, P. E. The effect of patterns of infectiousness on epidemic size. *Mathematical Biosciences and Engineering*, v. 3, p. 429-435, 2008.
- GRAVENOR, M. B.; PAPASOZOMENOS, P.; McLEAN, A. R.; NEOPHYTOU, G. A scrapie epidemic in Cyprus. *Epidemiology and Infection*, v. 132, p. 751-760, 2004.
- HAENSCH, S.; BIANUCCI, R.; SIGNOLI, M.; RAJERISON, M.; SCHULTZ, M.; KACKI, S.; VERMUNT, M.; DARLENE, A. W.; DEREK, H.; ACHTMAN, M.; CARNIEL, E.; BRAMANTI, B. Distinct clones of *yersinia pestis* caused the black death. *PloS Pathogens*, v. 6(10), e1001134, 2010.
- HAGMANN, R.; CHARLWOOD, J. D.; GIL, V.; CONCEIÇÃO, F.; do ROSÁRIO, V.; SMITH, T. Malaria and its possible control on the Island of Príncipe. *Malaria Journal*, v. 2 p. 1-9, 2003.
- HAYDON, D. T.; CHASE-TOPPING, M.; SHAW, D. J.; MATTHEWS, L.; FRIAR, J. K.; WILESMITH, J.; WOOLHOUSE, M. E. J. The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth outbreak. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, v. 270, p. 121-127, 2003.
- HEESTERBEEK, J. A. P.; DIETZ, K. The concept of R_0 in epidemic theory. *Statistica Neerlandica*, v. 50, p. 89-110, 1996.
- HEESTERBEEK, J. A. P. A brief history of R_0 and a recipe for its calculation. *Acta Biotheoretica*, v. 50, p. 189-204, 2002.
- HEFFERNAN, J. M.; SMITH, R. J.; WAHL, L. M. Perspectives on the basic reproductive ratio. *Journal of the Royal Society Interface*, v. 2, n. 4, p. 281-293, 2005.
- HENS, N.; SHKEDY, Z.; AERST, M.; FAES, C.; VAN DAMME, P.; BEUTELS, P. *Modeling Infectious Disease Parameters Based on Serological and Social Contac Data: A Modern Statistical Perspective*. Statistics for Biology and Health Series. New York: Springer, 2012.

- HETHCOTE, H. W. Mathematical models for the spread of infectious diseases. In: LUDWIG D.; COOKE, K.L. (eds.). *Epidemiology*, p. 122-131, Philadelphia: Society for Industrial and Applied Mathematics, 1975.
- HETHCOTE, H. W. An immunization model for a heterogeneous population. *Theoretical Population Biology*, v. 14, p. 338-349, 1978.
- HETHCOTE, H. W. The mathematics of infectious diseases. *Siam Review*, v. 42, p. 599-653, 2000.
- HETHCOTE, H. W.; THIEME, H. R. Stability of the endemic equilibrium in epidemic models with subpopulation. *Mathematical Biosciences*, v. 75, p. 205-227, 1985
- HETHCOTE, H. W.; TUDOR, D. W. Integral equation models for endemic infectious diseases. *Journal of Mathematical Biology*, v. 9, p. 37-47, 1980.
- HILL, A. N.; LONGINI, J. I. M. The critical vaccination fraction for heterogeneous epidemic models. *Mathematical Biosciences*, v. 181, p. 85-106, 2003.
- HUANG, Y.; ROSENKRANZ, S. L.; WU, H. Modeling HIV dynamics and antiviral response with consideration of time-varying drug exposures, adherence and phenotypic sensitivity. *Mathematical Biosciences*, v. 184, p. 165-186, 2003.
- HYMAM, J. M.; LI, J. An intuitive formulation for the reproductive number for the spread of diseases in heterogeneous populations. *Mathematical Biosciences*, v. 167, p. 65-86, 2000.
- INABA, H.; NISHIURA, H. The basic reproduction number of an infectious disease in a stable population: The impact of population growth rate on the eradication threshold. *Mathematical Modelling of Natural Phenomena*, v. 3, p. 194-228, 2008.
- JHONSON, N. *Britain and the 1918-1919 Influenza Pandemic. A Dark Epilogue*. Nova Iorque: Routledge, 2006.
- KAPLAN, E. H.; CRAFTC, D. L.; WEIN, L.M. Emergency response to a smallpox attack: The case for mass vaccination. *Proceedings of the National Academy of Sciences*, v. 99, p. 10935-10940, 2002.
- KEELING, M. J. The effects of local spatial structure on epidemiological invasion. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, v. 266, p. 859-867, 1999.
- KEELING, M. J.; ROHANI, P. *Modeling Infectious Diseases in Humans and Animals*. Princeton: Princeton University Press, 2007
- KERMARCK, W. O.; MCKENDRICK, A. G. Contributions to the mathematical theory of epidemics. Part I. *Proceedings of the Royal Society A*, v. 115, p. 700-721, 1927.
- KUCZYNSKI, R. R. *The Balance of Births and Deaths*. Nova Iorque: Macmillan, 1928.

- LAHROUZ, A.; OMARI, L.; KIOUACH, D. Global analysis of a deterministic and stochastic nonlinear SIRS epidemic model. *Nonlinear Analysis Modelling and Control*, v. 16, p. 59-76, 2011.
- LEKONE, P. E.; FINKENSTÄDT, B. F. Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. *Biometrics*, v. 62, p. 1170-1177, 2006.
- LI, G.; JIN, Z. Global stability of an SEI epidemic model. *Chaos, Solitons & Fractals*, v. 21, p. 925-931, 2004.
- LI, G.; JIN, Z. Global stability of a SEIR epidemic model with infectious force in latent, infected and immune period. *Chaos, Solitons & Fractals*, v. 25, p. 1177-1184, 2005.
- LI, J.; BLAKELEY, D.; SMITH, J. R. The failure of R_0 . *Computational and Mathematical Methods in Medicine*, v. 2011, p. 1-17, 2011.
- LI, Y. M.; JOHN, R.; GRAEF, A.; LIANCHENG, W. A.; KARSAL, J. Global dynamics of a SEIR model with varying total population size. *Mathematical Biosciences*, v. 160, p. 191-213, 1999.
- LIPSITCH, M.; COHEN, T.; COOPER, B.; ROBINGS M. A.; MA, S.; JAMES, L.; GOPALA-KRISHNA, G.; CHEW, K. S.; SAMORE, H. M.; FISMAN, D.; MURRAY, M. Transmission dynamics and control of severe acute respiratory syndrome. *Science*, v. 300, p. 1966-1970, 2003.
- LIU, W. M.; HETHCOTE, H. W.; LEVIN, S. A. Dynamical behavior of epidemiological models with nonlinear incidence rates. *Journal of Mathematical Biology*, v. 13, p. 359-380, 1987.
- LLOYD, A. L. Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods. *Proceedings of the Royal Society. Series B: Biological Sciences*, v. 268, p. 985-993, 2001.
- LLOYD, A. L.; MAY, R. M. Spatial heterogeneity in epidemic models. *Journal of Theoretical Biology*, v. 179, p. 1-11, 1996.
- LLOYD, A. L.; MAY, R. M. How viruses spread among computers and people. *Science*, v. 292, p. 1316-1317, 2001.
- LONGINI, J. I. M.; HALLORAN, M. E.; NIZAM, A.; YANG, Y. Containing pandemic influenza with antiviral agents. *American Journal of Epidemiology*, v. 159, p. 623-633, 2004.
- LUZ, P. M.; CODEÇO, C. T.; MASSAD, E.; STRUCHINER, C. J. Uncertainties regarding dengue modeling in Rio de Janeiro. *Memórias do Instituto Oswaldo Cruz*, v. 98, p. 871-878, 2003.
- MA, J.; DUSHOFF, J.; BOLKER, B. M.; EARN, D. J. D. Estimating initial epidemic growth rates. *Bulletin of Mathematical Biology*, v. 76, p. 245-260, 2014.

- MACDONALD, G.; CUELLAR, C. B.; FOLL, V. The dynamics of malaria. *Bulletin of the World Health Organization*, v. 38, p. 743-755, 1968.
- MARGOTTA, R. *An Illustrated History of Medicine*. Londres: Paul Hamlyn, 1968.
- MARK, S.; KLEMPNER, M. D.; DANIEL, S.; SHAPIRO, M. D. Crossing the species barrier - one small step to man, one giant leap to mankind. *The New England Journal of Medicine* v. 350, p. 1171-1172, 2004.
- MATTHEW, J. Estimation and inference of R_0 of an infectious pathogen by a removal method. *Mathematical Biosciences*, v. 198, p. 14-26, 2005.
- MATTHEWS, L., HAYDON, D. T.; SHAW, D. J.; CHASE-TOPPING, M. E.; KEELING, M. J.; WOOLHOUSE, M. E. J. Neighbourhood control policies and the spread of infectious disease. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, v. 266, p. 1659-1666, 2003.
- MATTHEWS, L.; WOOLHOUSE, M. E. J.; HUNTER, N. The basic reproduction number for scrapie. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, v. 266, p. 1085-1090, 1999.
- MILLS, C. E.; ROBINS, J. M.; LIPSITCH, M. Transmissibility of 1918 pandemic influenza. *Nature*, v. 432, p. 904-906, 2004.
- MOLLISON, D. Spatial contact models for ecological and epidemic spread (with Discussion). *Journal of the Royal Statistical Society. Series B (Methodological)*, v. 39, p. 283-326, 1977
- MOLLISON, D. The structure of epidemic models. In: MOLLISON, D. (ed.). *Epidemic Models: Their Structure and Relation to Data*, p. 13-33. Cambridge: Cambridge University Press, 1995.
- MURRAY, J. D. *Mathematical Biology I: An Introduction*, 3^a.ed. Berlim: Springer, 2002.
- NASELL, I. The threshold concept in stochastic epidemic and endemic models. In: MOLLISON, D. (ed.). *Epidemic Models: Their Structure and Relation to Data*, p. 71-83. Cambridge: Cambridge University Press, 1995.
- NISHIURA, H. Mathematical and statistical analyses of the spread of dengue. *Dengue Bulletin*, v. 30, p. 51-67, 2006.
- NISHIURA, H.; CHOWELL, G. The effective reproduction number as a prelude to statistical estimation to model assumptions. In: CHOWELL, G., HYMAN, J. M., BETTENCOURT, L. M. A.; CASTILLO-CHAVEZ, C. (eds.). *Mathematical and Statistical Estimation Approaches in Epidemiology*, p. 103-121. Londres: Springer, 2009.
- NOLD, A. Heterogeneity in disease-transmission modeling. *Mathematical Biosciences*, v. 52, p. 227-240, 1980.

- PUGLIESE, A. Population models for diseases with no recovery. *Journal of Mathematical Biology*, v. 28, p. 65-82, 1990.
- RILEY, T.; FRASER, C.; DONELLY, C. A.; GHANI, A. C.; ABU-RADDAD, L. J.; HEDLEY, A. J.; LEUNG, G. M.; HO, L.-M.; LAM, T.-H.; THACH, T. Q.; CHAU, P.; CHAN, K.-P.; LO, S.-V.; LEUNG P.-Y.; L.; TSANG, T.; HO, W.; LEE, K.-H.; LAU, E. M. C.; FERGUSON, N. M.; ANDERSON, R. M. Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions. *Science*, v. 300, p. 1961-1966, 2003.
- RÖHM, C.; HORIMOTO, T.; KAWAOKA, Y.; SUSS, J.; WEBSTER, R. Do hemagglutinin genes of highly pathogenic avian influenza viruses constitute unique phylogenetic lineages? *Virology*, v. 209, p. 664-670, 1995.
- RUDNICK, S. N.; MILTON, D. K.; 2003. Risk of indoor airborne infection transmission estimated from carbon dioxide concentration. *International Journal of Indoor Environment and Health*, v. 13, p. 237-245, 2003.
- RUSHTON, S.; MAUTNER, A. J. The deterministic model of a simple epidemic for more than one community. *Biometrika*, v. 42, p. 126-132, 1955
- SHARPE, F. R.; LOTKA, A. J. A problem in age-distribution. *Philosophical Magazine*, v. 21, p. 435-438, 1911.
- SIETTOS, I. C.; RUSSO, L. Mathematical modeling of infectious disease dynamics. *Virulence*, v. 4, p. 295-306, 2013.
- SOPER, H. E. The interpretation of periodicity in disease prevalence. *Journal of the Royal Statistical Society*, v. 92, p. 34-73, 1929.
- STEGEMAN, A.; BOUMA, A.; ELBERS, A. R. W.; DE JONG, M. C. M.; NODELIJK, G.; DE KLERK, F.; KOCH, G.; VAN BOVEN, M. Avian influenza A virus (H7N7) epidemic in The Netherlands in 2003: Course of the epidemic and effectiveness of control measures. *Journal of Infectious Diseases*, v. 190, p. 2088-2095, 2004.
- SUTER, R. E. Hippocratic thought: Its relationship to and between Andrew Taylor Still and Sir William Osler. *Medical education*. v.88, p. 10-1243, 1988.
- THIEME, H. R. Epidemic and demographic interaction in the spread of potentially fatal diseases in growing populations. *Mathematical Biosciences*, v. 111, p. 99-130, 1992.
- VAN DEN DRIESSCHE P.; WATMOUGH, J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences*, v. 180, p. 29-48, 2002.
- VAN DEN DRIESSCHE, P.; WATMOUGH, J. Further Notes on the Basic Reproduction Number. In: BRAUER, F.; VAN DEN DRIESSCHE, P.; WU, J. (Eds.). *Mathematical Epidemiology*, p. 159-178. Lecture Notes in Mathematics, Volume 1945. Berlin: Springer, 2008.

- VYNNYCKY, E.; TRINDALL, A.; MANGTANI P. Estimates of the reproduction numbers of Spanish influenza using morbidity data. *International Journal of Epidemiology*, v. 36, p. 881-889, 2007.
- WALLINGA, J.; LIPSITCH, M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences*, v. 274, p. 599-604, 2007.
- WALLINGA, J.; TEUNIS, P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of Epidemiology*, v. 160, p. 509-516, 2004.
- WONHAM, M. J.; BEEK, T.; LEWIS, M. A. An epidemiological model for West Nile virus: Invasion analysis and control applications. *Proceedings of the Royal Society B: Biological Sciences*, v. 271, p. 501-507, 2004.
- WOOLHOUSE, M. E. J.; DYE, C.; ETARD J.-F.; SMITH, T.; CHARLWOOD, J.D.; GARNETT, G. P.; HAGAN, P.; HILL, J. L. K.; NDHLOVU, P. D.; QUINNELL, R. J.; WATTS, C. H.; CHANDIWANA, S. K.; ANDERSON, R. M. Heterogeneities in the transmission of infectious agents: Implications for the design of control programs. *Proceedings of the National Academy of Sciences*, v. 94, p. 338- 342, 1997.
- XIA, Y. C.; GOG, J. R.; GRENFELL, B. T. Semiparametric estimation of the duration of immunity from infectious disease time-series: influenza as a case study. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, v. 54, p. 659-672, 2005.
- YANG, H. M. *Estudos dos Efeitos da Vacinação em Doenças de Transmissão Direta*. Campinas: Editorial Unicamp, 2001.
- YANG, Y.; JONATHAN, D.; SUGIMOTO, M.; HALLORAN, E.; NICOLE, E. CHAO, D. L.; MATRAJT, L. POTTER, G.; KENAH, E.; IRA, M.; LONGINI, J. The transmissibility and control of pandemic influenza A (H1N1) virus. *Science*, v. 326, p. 729-733, 2009.